



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

ICHL26

INTERNATIONAL CONFERENCE ON HISTORICAL LINGUISTICS 2023

Heidelberg, 4–8 Sept.



BOOK OF ABSTRACTS

ICHL 26 – International Conference on Historical Linguistics 2023, Heidelberg, 4–8 Sept.
Book of Abstracts

Texts are reproduced as submitted by the session organisers and/or authors.

Design and Layout: Nicolas M. Jansens (Heidelberg)

© Heidelberg University, 2023

Funded by
 Deutsche
Forschungsgemeinschaft
German Research Foundation

CONTENTS

PAGES

| | |
|-------------|-------------------------------------|
| XII | Welcome |
| XIII | Orientation |
| XIV | Programme |
| | |
| 1 | Abstract no. PL1 – Kiparsky |
| 2 | Abstract no. PL2 – Lahiri |
| 3 | Abstract no. PL3 – Marten |
| 4 | Abstract no. PL4 – Engelberg et al. |
| 5 | Abstract no. PL5 – van Kemenade |
| 6 | Abstract no. PL6 – Vincent et al. |
| 7 | Abstract no. PL7 – Smith |
| | |
| 8 | Workshop no. W1 – Robbeets |
| 12 | Contribution to W1 – Berge |
| 14 | Contribution to W1 – Knapen |
| 15 | Contribution to W1 – Miyamoto |
| 16 | Contribution to W1 – Bradley |
| 18 | Contribution to W1 – Deng |
| 19 | Contribution to W1 – Sidwell |
| 20 | Contribution to W1 – Heggarty |
| 22 | Contribution to W1 – Joseph |
| 23 | Contribution to W1 – Hudson |
| | |
| 24 | Workshop no. W2 – Drinka et al. |
| 27 | Contribution to W2 – Nichols |
| 29 | Contribution to W2 – Andersen |
| 30 | Contribution to W2 – Nijs et al. |
| 32 | Contribution to W2 – Sobolev |
| 34 | Contribution to W2 – Gvozdanović |
| 36 | Contribution to W2 – Sowada |
| 38 | Contribution to W2 – Enrique-Arias |
| 39 | Contribution to W2 – Mesthrie |
| 40 | Contribution to W2 – Salmons |

| | |
|------------|--|
| 41 | Workshop no. W3 – Degaetano-Ortlieb et al. |
| 44 | Contribution to W3 – Al-Laith et al. |
| 46 | Contribution to W3 – Jensen et al. |
| 48 | Contribution to W3 – Maurer et al. |
| 50 | Contribution to W3 – Marr |
| 54 | Contribution to W3 – Schlechtweg |
| 55 | Contribution to W3 – Amaral et al. |
| 57 | Contribution to W3 – Rönchen et al. |
| 59 | Contribution to W3 – Dereza et al. |
| | |
| 61 | Workshop no. W4 – Zehentner & De Cesare |
| 65 | Contribution to W4 – Smirnova |
| 67 | Contribution to W4 – Ceuppens & De Smet |
| 69 | Contribution to W4 – Felser |
| 70 | Contribution to W4 – Wolfsgruber |
| 72 | Contribution to W4 – Ritt & Böhm |
| 75 | Contribution to W4 – Seržant |
| 77 | Contribution to W4 – Haspelmath |
| | |
| 79 | Workshop no. W5 – Kölligan & van Beek |
| 82 | Contribution to W5 – van Beek |
| 84 | Contribution to W5 – Bartolotta |
| 87 | Contribution to W5 – Ginevra |
| 89 | Contribution to W5 – Pompeo |
| 91 | Contribution to W5 – Roth |
| 94 | Contribution to W5 – Zampetta et al. |
| 96 | Contribution to W5 – Kölligan |
| | |
| 98 | Workshop no. W6 – Grestenberger et al. |
| 100 | Contribution to W6 – Calabrese |
| 101 | Contribution to W6 – Alfieri |
| 102 | Contribution to W6 – Hasselbach-Andee |
| 103 | Contribution to W6 – Tan |
| 105 | Contribution to W6 – Werner |
| | |
| 107 | Workshop no. W7 – Bjørn & Kilani |
| 111 | Contribution to W7 – Hansen & Davletshin |
| 113 | Contribution to W7 – Bostoen et al. |

| | |
|------------|--|
| 116 | Contribution to W7 – Souag |
| 118 | Contribution to W7 – Widmer & Sonnenhauser |
| 119 | Contribution to W7 – Wier |
| 121 | Contribution to W7 – Yurayong et al. |
| 124 | Workshop no. W8 – Karim & Gholami |
| 129 | Contribution to W8 – Kreidl |
| 131 | Contribution to W8 – Kim |
| 133 | Contribution to W8 – Mohammadirad |
| 135 | Contribution to W8 – Gholami & Naghshbandi |
| 137 | Contribution to W8 – Suleymanov |
| 140 | Workshop no. W9 – Däbritz |
| 140 | Contribution to W9 – Däbritz |
| 143 | Contribution to W9 – Krasnoukhova et al. |
| 145 | Contribution to W9 – Hengeveld |
| 146 | Contribution to W9 – Creissels |
| 148 | Contribution to W9 – Camilleri |
| 150 | Workshop no. W10 – Baudel et al. |
| 155 | Contribution to W10 – Satō & Bugaeva |
| 157 | Contribution to W10 – Shimabukuro |
| 159 | Contribution to W10 – Kinuhata |
| 161 | Contribution to W10 – Baudel |
| 163 | Contribution to W10 – Majtczak |
| 165 | Workshop no. W11 – Auderset et al. |
| 170 | Contribution to W11 – Božović |
| 172 | Contribution to W11 – Lionnet |
| 175 | Contribution to W11 – Kirby & Pittayaporn |
| 176 | Contribution to W11 – Kirby & Pittayaporn |
| 177 | Contribution to W11 – Arnold |
| 179 | Contribution to W11 – Grimm |
| 181 | Contribution to W11 – Sæbø & Grossman |
| 183 | Contribution to W11 – Perekhvalskaya & Vydrin |
| 186 | Workshop no. W12 – Orqueda & González Saavedra |
| 188 | Contribution to W12 – Mithun |

| | |
|------------|--|
| 190 | Contribution to W12 – Brosig & Dolgor |
| 192 | Contribution to W12 – Ishiyama |
| 195 | Contribution to W12 – Stanković |
| 197 | Contribution to W12 – Næss |
| 199 | Contribution to W12 – Neri & de Vaan |
| 200 | Contribution to W12 – Luján & Ngomo Fernández |
| 202 | Contribution to W12 – Orqueda & Pooth |
| 203 | Workshop no. W13 – Cassarà et al. |
| 207 | Contribution to W13 – Figura |
| 210 | Contribution to W13 – Trips & Rainsford |
| 212 | Contribution to W13 – Cassarà et al. |
| 214 | Contribution to W13 – Kodner |
| 217 | Workshop no. W14 – Jäger et al. |
| 218 | Contribution to W14 – Brigada Villa et al. |
| 219 | Contribution to W14 – Rzymiski |
| 220 | Contribution to W14 – Forkel & Greenhill |
| 221 | Contribution to W14 – Dellert & Blaschke |
| 222 | Contribution to W14 – Blum & List |
| 223 | Contribution to W14 – Mertner & Guzmán Naranjo |
| 224 | Abstract no. 25 – Olgúin-Martínez |
| 226 | Abstract no. 29 – Fonteyn et al. |
| 227 | Abstract no. 30 – Salaberri |
| 229 | Abstract no. 31 – Shamseddinov & Authier |
| 231 | Abstract no. 32 – Bonmann et al. |
| 234 | Abstract no. 35 – Schützler |
| 236 | Abstract no. 36 – Poletto et al. |
| 238 | Abstract no. 37 – Persohn |
| 240 | Abstract no. 38 – Pan |
| 241 | Abstract no. 39 – Long |
| 243 | Abstract no. 40 – Bronikowska |
| 245 | Abstract no. 41 – van Dam |
| 247 | Abstract no. 43 – Maiden |
| 249 | Abstract no. 48 – Rodríguez-Somolinos |
| 251 | Abstract no. 50 – Rosenkvist |
| 253 | Abstract no. 51 – Kisiel & Sobotka |

| | |
|------------|--|
| 255 | Abstract no. 52 – Börjars & Vincent |
| 257 | Abstract no. 55 – Mofidi |
| 258 | Abstract no. 56 – Huang |
| 260 | Abstract no. 57 – Hualde |
| 262 | Abstract no. 58 – Manterola |
| 264 | Abstract no. 59 – Conradie |
| 265 | Abstract no. 61 – Aldridge |
| 267 | Abstract no. 63 – Currie |
| 269 | Abstract no. 65 – Idiatov |
| 271 | Abstract no. 66 – Ongenae |
| 273 | Abstract no. 67 – Hamans |
| 274 | Abstract no. 68 – Dedvukaj |
| 276 | Abstract no. 69 – Pacchiarotti et al. |
| 278 | Abstract no. 70 – Šefčík |
| 280 | Abstract no. 72 – Bru |
| 282 | Abstract no. 73 – Munteanu |
| 284 | Abstract no. 74 – Mirelman |
| 286 | Abstract no. 75 – Bogdanowska-Jakubowska & Bogdanowska |
| 288 | Abstract no. 77 – Västerdal |
| 289 | Abstract no. 78 – Wolfe |
| 290 | Abstract no. 79 – Roth |
| 291 | Abstract no. 80 – Bloom |
| 293 | Abstract no. 84 – Jensen & Schack |
| 295 | Abstract no. 86 – Alfieri & Pozza |
| 297 | Abstract no. 88 – Saiz Sánchez |
| 299 | Abstract no. 89 – Kayenbergh & De Smet |
| 301 | Abstract no. 90 – Westergaard & Boye |
| 303 | Abstract no. 91 – Gobena |
| 304 | Abstract no. 92 – Westergaard |
| 306 | Abstract no. 93 – Boye |
| 307 | Abstract no. 94 – Huback & Fontes Martins |
| 308 | Abstract no. 96 – Shcherbakova et al. |
| 310 | Abstract no. 97 – Chankova |
| 312 | Abstract no. 98 – Friedman |
| 314 | Abstract no. 99 – Mendoza et al. |
| 316 | Abstract no. 100 – Inglese et al. |
| 318 | Abstract no. 102 – Ricquier & Demolin |
| 319 | Abstract no. 103 – Schäfer |

| | |
|------------|---|
| 320 | Abstract no. 104 – Bossuyt & Daveloose |
| 322 | Abstract no. 105 – Esher |
| 324 | Abstract no. 108 – Iezzi |
| 326 | Abstract no. 109 – Salvesen |
| 328 | Abstract no. 111 – Visser |
| 329 | Abstract no. 112 – Brunner |
| 331 | Abstract no. 113 – Feltgen |
| 333 | Abstract no. 114 – Farina et al. |
| 336 | Abstract no. 115 – Stratton |
| 338 | Abstract no. 116 – Leddy-Cecere |
| 340 | Abstract no. 118 – Gosemann |
| 342 | Abstract no. 121 – De Vos |
| 344 | Abstract no. 122 – Concu |
| 346 | Abstract no. 123 – Sternefeld |
| 348 | Abstract no. 124 – Kümmel et al. |
| 350 | Abstract no. 125 – Schulte |
| 351 | Abstract no. 126 – Eythorsson |
| 353 | Abstract no. 127 – Coenen |
| 355 | Abstract no. 128 – Pleyer |
| 357 | Abstract no. 129 – Fromm |
| 359 | Abstract no. 133 – Tan |
| 361 | Abstract no. 135 – Janda & Joseph |
| 363 | Abstract no. 137 – Daveloose |
| 365 | Abstract no. 138 – Rosemeyer et al. |
| 367 | Abstract no. 142 – Voigtmann |
| 369 | Abstract no. 143 – Tikhonov et al. |
| 371 | Abstract no. 144 – Jonjić et al. |
| 373 | Abstract no. 145 – Assenzi |
| 375 | Abstract no. 146 – Burns |
| 377 | Abstract no. 147 – Sims-Williams |
| 378 | Abstract no. 148 – Espíndola Moschner & Rosemeyer |
| 380 | Abstract no. 150 – Flaksman |
| 382 | Abstract no. 152 – Ariztimuño & Salaberri |
| 384 | Abstract no. 153 – Chen |
| 386 | Abstract no. 154 – Di Bartolo |
| 388 | Abstract no. 157 – Hofmann |
| 390 | Abstract no. 158 – Pierce |
| 391 | Abstract no. 164 – Zeng |

- 393** Abstract no. 165 – Tresoldi et al.
396 Abstract no. 166 – Pache
397 Abstract no. 167 – Markopoulos
399 Abstract no. 168 – Franco
401 Abstract no. 169 – Rahman & Banerjee
403 Abstract no. 170 – Reetz
405 Abstract no. 171 – Dücker
407 Abstract no. 173 – Cennamo
409 Abstract no. 174 – Nieder & Tomaschek
410 Abstract no. 175 – Dömötör
412 Abstract no. 176 – Torres-Latorre
414 Abstract no. 178 – Billing & Elgh
416 Abstract no. 180 – Darling et al.
418 Abstract no. 181 – Reinöhl et al.
420 Abstract no. 183 – Elter
422 Abstract no. 184 – Mailhammer & Harvey
424 Abstract no. 186 – Kaye & Maisak
426 Abstract no. 187 – Meyer
428 Abstract no. 188 – Honeybone
430 Abstract no. 189 – Pronk
432 Abstract no. 190 – Reinöhl & Ellison
434 Abstract no. 191 – Bauer
435 Abstract no. 192 – Mous
437 Abstract no. 194 – Egedi
439 Abstract no. 196 – Round et al.
441 Abstract no. 197 – Kozhanov
443 Abstract no. 198 – Lindgren & Tresoldi
446 Abstract no. 199 – Herce & Cathcart
448 Abstract no. 200 – Hirvonen
450 Abstract no. 203 – Björnsdóttir
452 Abstract no. 206 – Petré
454 Abstract no. 207 – Sapp et al.
456 Abstract no. 208 – Cluyse et al.
458 Abstract no. 211 – Meisterernst
460 Abstract no. 212 – Russell
462 Abstract no. 213 – Elens et al.
464 Abstract no. 214 – Gotthard
467 Abstract no. 215 – Hernáiz

| | |
|------------|---|
| 468 | Abstract no. 217 – Olivier |
| 470 | Abstract no. 219 – Ebert |
| 472 | Abstract no. 220 – Sigurðardóttir |
| 474 | Abstract no. 221 – Wichers Schreur |
| 476 | Abstract no. 222 – Gisborne & Truswell |
| 478 | Abstract no. 225 – Lionnet |
| 480 | Abstract no. 227 – Gibson et al. |
| 482 | Abstract no. 228 – Gugán |
| 484 | Abstract no. 230 – Brown & Grollemund |
| 486 | Abstract no. 232 – Gunnink et al. |
| 488 | Abstract no. 233 – McCarley |
| 490 | Abstract no. 234 – Gfeller |
| 492 | Abstract no. 235 – Strauss |
| 494 | Abstract no. 236 – Potochnik |
| 496 | Abstract no. 237 – Capano |
| 498 | Abstract no. 238 – Holopainen |
| 500 | Abstract no. 239 – Halfmann & Korobzow |
| 501 | Abstract no. 240 – Gopal et al. |
| 503 | Abstract no. 242 – Verkerk |
| 505 | Abstract no. 243 – Juge |
| 507 | Abstract no. 244 – Auderset |
| 509 | Abstract no. 245 – Dockum & Lu |
| 511 | Abstract no. 247 – Gelumbeckaite et al. |
| 513 | Abstract no. 248 – Ritt |
| 515 | Abstract no. 251 – Hakimov |
| 516 | Abstract no. 253 – Ulman |
| 517 | Abstract no. 254 – Igartua |
| 519 | Abstract no. 255 – Boyeldieu |
| 525 | Abstract no. 258 – Litvinova |
| 526 | Abstract no. 261 – Klævik-Pettersen |
| 527 | Abstract no. 263 – Santamaria |
| 529 | Abstract no. 264 – Cattafi |
| 531 | Abstract no. 265 – Mounole & Manterola |
| 533 | Abstract no. 266 – Das |
| 536 | Abstract no. 267 – Serangeli |
| 538 | Abstract no. 268 – Rapold |
| 540 | Abstract no. 269 – Wieczorek |
| 542 | Abstract no. 270 – Juge |

| | |
|------------|---|
| 544 | Abstract no. 272 – Pompei |
| 546 | Abstract no. 273 – Caso & Hale |
| 548 | Abstract no. 274 – Dockum & Wang |
| 550 | Abstract no. 275 – Pounder |
| 552 | Abstract no. 278 – Benvenuto & Bichlmeier |
| 554 | Abstract no. 279 – Dinu et al. |
| 556 | Abstract no. 280 – Paterson, R. |
| 558 | Abstract no. 281 – Tiekü |
| 560 | Abstract no. 282 – Paterson, H. |
| 562 | Abstract no. 283 – Riad |
| 564 | Abstract no. 286 – Sitchinava |
| 566 | Abstract no. 290 – Swanenvleugel |
| 568 | Contributions by participant |

WELCOME TO THE 26TH INTERNATIONAL CONFERENCE ON HISTORICAL LINGUISTICS, HEIDELBERG UNIVERSITY 2023

This **jubilee** International Conference on Historical Linguistics (**ICHL**) marks a milestone of 50 years since the inaugural event and the founding of the International Society for Historical Linguistics (**ISHL**) in Edinburgh in 1973. Appropriately, it showcases the highly specialized and multifaceted nature of historical linguistics as a branch of the language sciences, with proven research methods buttressed by keen expertise and continually able to deliver new theoretical insights.

In its mission to advance and support research on historical linguistics, the ISHL has regularly organised biennial conferences, each time in a different city and, if possible, on a different continent. By now, linguists from all continents have joined the ISHL and participate in its conferences. These events have become a major international forum, bringing together historical linguists and specialists in related fields to explore advances in methods of linguistic reconstruction, formal and functional approaches to language change, historical sociolinguistics, computational approaches to historical linguistics, contact and areal linguistics, and interfaces with other disciplines.

This year's conference features fourteen workshops, organized mainly by young scholars, bringing fresh interdisciplinary perspectives to areas of current relevance. Plenary talks will be given by five established authorities in their fields, and two innovative panels, on lexical dynamics and on linguistic models of morphosyntactic change, will critically discuss the achievements and limitations of research in these areas.

The general session represents current research trends and challenges, at times reformulating previous findings. Digital methods also play a major role, in assessing large corpora, answering research questions, and bridging the gap between multifarious language data and the principles established or proposed by linguistic theories.

Thus, while continuing past traditions, the **26th International Conference on Historical Linguistics** stands as a platform for important innovations of interdisciplinarity and digitalisation within the open and inclusive research landscapes of contemporary science.

Prof. Dr. Jadranka Gvozdanović
President of ISHL, Chair of ICHL26

ORIENTATION

The conference begins on Sept. 4th, 2023 with **registration** opening at 8:00 am on the ground floor of the **New University** building (Neue Universität).

► Address: Universitätsplatz 1
69117 Heidelberg

The **opening plenary talk** will take place in the **Alte Aula** on the first floor of the **Old University** building (Alte Universität). Beginning with the coffee break (10:00–10:30) on Sept. 4th, all further events take place in the **New University** building.



Lecture halls in the New University building:

Ground floor: H1, H2, H3, H4, H4a

First floor: H5, H6, H7, H8, H12

Second floor: H12a, A1 = Neue Aula, ES = Former Senate Hall (behind the Neue Aula)

Lunches and coffee breaks are provided for on the New University building's ground floor.

Wi-Fi internet access is available in university buildings via **eduroam** (with appropriate credentials). At various locations in the city, including the New University building, free Wi-Fi is available through **Heidelberg4you** (see: <https://www.heidelberg.de/904453.html>).

PROGRAMME

General overview

| | Monday (D1) | Tuesday (D2) | Wednesday (D3) | Thursday (D4) | Friday (D5) |
|-------------|---------------------------------|----------------------------------|---|--|-----------------------------------|
| 09:00–10:00 | Paul Kiparsky | Aditi Lahiri | Lutz Marten | Ans van Kemenade | John Charles Smith |
| 10:00–10:30 | <i>Coffee break</i> | | | | |
| 10:30–11:00 | W1, General sessions | W2, General sessions | Panel: Stefan Engelberg et al. | Panel: Nigel Vincent et al. | W8, W11, W14, General sessions |
| 11:00–11:30 | | | | | |
| 11:30–12:00 | | | | | |
| 12:00–12:30 | | | | | |
| 12:30–13:30 | <i>Lunch break</i> | | | | |
| 13:30–14:00 | W1, W5, W7, General sessions | W2, W4, W12, General sessions | General sessions | W3, W6, W9, W10, W13, General sessions | W8, W11, W14, General sessions |
| 14:00–14:30 | | | | | |
| 14:30–15:00 | | | | | |
| 15:00–15:30 | | | | | |
| 15:30–16:00 | <i>Break</i> | | | | Farewell reception |
| 16:00–16:30 | W1, W5, W7, General sessions | W2, W4, W12, General sessions | <i>Free afternoon: museums and nature</i> | W3, W6, W9, W10, W13, General sessions | |
| 16:30–17:00 | | | | | |
| 17:00–17:30 | | | | | |
| 17:30–18:00 | | | | | |

DAY 1: MONDAY, SEPTEMBER 4th

D1: General sessions

| | Analytic methods (H1) | Historical morphology (H6) | History of tense and aspect (H12) | Grammatical processes (H5) | Workshops |
|-------------|---|-------------------------------|--------------------------------------|-------------------------------|---|
| 09:00–10:00 | Paul Kiparsky (Location: Alte Aula in the Old University building, afterwards the New University, A1) | | | | |
| 10:00–10:30 | <i>Coffee break</i> | | | | |
| 10:30–11:00 | Mailhammer & Harvey | Maiden | Persohn | Ariztimuño & Salaberri | Workshops: D1, D2, D4, D5 (see general programme) |
| 11:00–11:30 | Juge, 270 | Conradie | Coenen | Börjars & Vincent | |
| 11:30–12:00 | Pleyer et al. | Hamans | Kümmel et al. | Björnsdóttir et al. | |
| 12:00–12:30 | Sims-Williams | Janda & Joseph | Meisterernst | Gotthard | |
| 12:30–13:30 | <i>Lunch break</i> | | | | |
| 13:30–14:00 | Olivier | Alfieri & Pozza | Juge, 243 | Manterola et al. | |
| 14:00–14:30 | Gugán | Visser | Shamseddinov & Authier | Leddy-Cecere | |
| 14:30–15:00 | Gfeller | Fromm | Kaye & Maisak | Tresoldi et al. | |
| 15:00–15:30 | Gopal et al. | Dedvukaj | Petré | Feltgen | |
| 15:30–16:00 | <i>Break</i> | | | | |
| 16:00–16:30 | Saiz Sánchez | | Mofidi | Mounole & Manterola | |
| 16:30–17:00 | Tieku | | Espindola Moschner & Rosemeyer | Cluyse et al. | |
| 17:00–17:30 | | | | Elens et al. | |
| 17:30–18:00 | | | | Paterson, R. | |
| 18:00–18:45 | Welcome reception | | | | |

D1: Workshops

| | W1 – Robbeets (H8) | W5 – Kölligan & van Beek (H4) | W7 – Bjørn & Kilani (H12a) |
|-------------|-----------------------|----------------------------------|-------------------------------|
| 10:00–10:30 | <i>Coffee break</i> | | |
| 10:30–11:00 | Robbeets | | |
| 11:00–11:30 | Berge | | |
| 11:30–12:00 | Knapen | | |
| 12:00–12:30 | Miyamoto | | |
| 12:30–13:30 | <i>Lunch break</i> | | |
| 13:30–14:00 | Bradley | van Beek | Bjørn & Kilani |
| 14:00–14:30 | Deng | Bartolotta | Hansen & Davletshin |
| 14:30–15:00 | Sidwell | Ginevra | Bostoen et al. |
| 15:00–15:30 | Heggarty | Pompeo | Souag |
| 15:30–16:00 | <i>Break</i> | | |
| 16:00–16:30 | Joseph | Roth | Widmer & Sonnenhauser |
| 16:30–17:00 | Hudson | Zampetta et al. | Wier |
| 17:00–17:30 | | Kölligan | Yurayong et al. |
| 17:30–18:00 | | | Discussion |

DAY 2: TUESDAY, SEPTEMBER 5th

D2: General sessions

| | Historical morphology (H5) | Historical phonology (H12a) | Historical lexical semantics (H2) | Interfaces, sem./ synt./infor. struct. (H6) | Prosody, metrics (H8) |
|-------------|-------------------------------|--------------------------------|--------------------------------------|--|--------------------------|
| 09:00–10:00 | Aditi Lahiri (A1 = New Aula) | | | | |
| 10:00–10:30 | Coffee break | | | | |
| 10:30–11:00 | Rahman & Banerjee | Bonmann et al. | Fonteyn et al. | Kisiel & Sobotka | Hualde |
| 11:00–11:30 | Santamaria | Idiatov | Farina et al. | Huang | Västerdal |
| 11:30–12:00 | Torres-Latorre | Ongenae | Stratton | Chankova | Schulte |
| 12:00–12:30 | Darling et al. | Šefčič | Bru | Bossuyt & Daveloose | Pronk |
| 12:30–13:30 | Lunch break | | | | |
| 13:30–14:00 | Meyer | Tan | Friedman | Eyþórsson & Sigurðardóttir | Hofmann |
| 14:00–14:30 | Round et al. | Huback & Fontes Martins | Serangeli | Voigtmann | Auderset |
| 14:30–15:00 | Kozhanov | Lionnet | Concu | Tikhonov et al. | Ritt & Hofmann |
| 15:00–15:30 | Russell | Chen | | Jonjić et al. | Riad |
| 15:30–16:00 | Break | | | | |
| 16:00–16:30 | Sigurðardóttir | Pierce | | Shcherbakova et al. | Caso & Hale |
| 16:30–17:00 | Wichers Schreur | Reinöhl et al. | | Egedi | Boyeldieu |
| 17:00–17:30 | Strauss | Honeybone | | Dockum & Lu | Litvinova |
| 17:30–18:00 | Ulman | Rapold | | Dömötör | |

D2: Workshops

| | W2 – Drinka et al. (H1) | W4 – Zehentner & De Cesare (H4) | W12 – Orqueda & González Saavedra (H12) |
|-------------|----------------------------|------------------------------------|--|
| 10:00–10:30 | Coffee break | | |
| 10:30–11:00 | Drinka et al. | | |
| 11:00–11:30 | Nichols | | |
| 11:30–12:00 | Andersen | | |
| 12:00–12:30 | Nijs et al. | | |
| 12:30–13:30 | Lunch break | | |
| 13:30–14:00 | Sobolev | Smirnova | Mithun |
| 14:00–14:30 | Gvozdanović | Ceuppens & De Smet | Brosig & Dolgor |
| 14:30–15:00 | Sowada | Felser | Ishiyama |
| 15:00–15:30 | Enrique-Arias | Wolfgruber | Stanković |
| 15:30–16:00 | Break | | |
| 16:00–16:30 | Mesthrie | Ritt & Böhm | Næss |
| 16:30–17:00 | Salmons | Seržant | Neri & de Vaan |
| 17:00–17:30 | Discussion | Haspelmath | Luján & Ngomo Fernández |
| 17:30–18:00 | | | Orqueda & Pooth |

DAY 3: WEDNESDAY, SEPTEMBER 6th**D3: General sessions**

| | Interface with pragmatics (H5) | Historical language norms (H4) | Translation (H2) | Periodisation/reconstruction (H1) | Pragmatics, discourse (H6) |
|-------------|--|--|----------------------------|---|--------------------------------------|
| 09:00–10:00 | Lutz Marten (A1 = New Aula) | | | | |
| 10:00–10:30 | <i>Coffee break</i> | | | | |
| 10:30–11:00 | Stefan Engelberg et al. “Empirical approaches to the dynamics of the lexicon” with Annette Klosa-Kückelhaus, Peter Meyer, Samira Ochs, Jan Oliver Rüdiger, Sascha Wolfer | | | | |
| 11:00–11:30 | | | | | |
| 11:30–12:00 | | | | | |
| 12:00–12:30 | | | | | |
| 12:30–13:30 | <i>Lunch break</i> | | | | |
| 13:30–14:00 | Bloom | Currie | Mirelman | Munteanu | Markopoulos |
| 14:00–14:30 | Halfmann & Korobzow | Roth | Flaksman | Ricquier & Demolin | Bogdanowska-Jakubowska & Bogdanowska |
| 14:30–15:00 | Salvesen | Schäfer | | Lindgren & Tresoldi | Boye |
| 15:00–15:30 | Pan | Gelumbeckaitė et al. | | Verkerk et al. | di Bartolo |
| 15:30–16:00 | Westergaard & Boye | Daveloose | | Dinu et al. | Potochnik |
| 16:00–16:30 | <i>Free afternoon: museums and nature</i> | | | | |
| 16:30–17:00 | | | | | |
| 17:00–17:30 | | | | | |
| 17:30–18:00 | | | | | |

DAY 4: THURSDAY, SEPTEMBER 7th

D4: General sessions

| | Historical phonology (H2) | Historical syntax (ES) | Historical morphosyntax (H6) | Grammatical semantics (H7) | History of word order (H5) |
|-------------|--|------------------------|------------------------------|----------------------------|----------------------------|
| 09:00–10:00 | Ans van Kemenade (A1 = New Aula) | | | | |
| 10:00–10:30 | Coffee break | | | | |
| 10:30–11:00 | Nigel Vincent et al. "Linguistic models" with Kasper Boye, Ashwini Deo, Mirjam Fried, George Walkden | | | | |
| 11:00–11:30 | | | | | |
| 11:30–12:00 | | | | | |
| 12:00–12:30 | | | | | |
| 12:30–13:30 | Lunch break | | | | |
| 13:30–14:00 | Holopainen | Mendoza et al. | Esher | Cennamo | Poletto et al. |
| 14:00–14:30 | Hakimov | Inglese et al. | de Vos | Benvenuto & Bichlmeier | Brunner |
| 14:30–15:00 | Paterson, H. | Gosemann | Gibson et al. | Igartua | Reetz |
| 15:00–15:30 | Dockum & Wang | Wolfe | Jensen & Schack | Gobena | Dücker |
| 15:30–16:00 | Break | | | | |
| 16:00–16:30 | Pounder | McCarley | Rosenkvist | Pompei | Ebert et al. |
| 16:30–17:00 | | | | Gisborne & Truswell | Klævik-Pettersen |
| 17:00–17:30 | Business meeting, then conference dinner | | | | |
| 17:30–18:00 | | | | | |

D4: Workshops

| | W6 – Grestenberger et al. (H4) | W9 – Däbritz (H12) | W10 – Baudel et al. (H12a) | W13 – Cassarà et al. (H8) |
|-------------|--------------------------------|---------------------|----------------------------|---------------------------|
| 12:30–13:30 | Lunch break | | | |
| 13:30–14:00 | Grestenberger et al. | Däbritz | Satō & Bugaeva | Cassarà et al. |
| 14:00–14:30 | Calabrese | Krasnoukhova et al. | Shimabukuro | Figura |
| 14:30–15:00 | Alferi | Hengeveld | Kinuhata | Trips & Rainsford |
| 15:00–15:30 | Hasselbach-Andee | Creissels | Baudel | Cassarà et al. |
| 15:30–16:00 | Break | | | |
| 16:00–16:30 | Tan | Camilleri | Majtczak | Kodner |
| 16:30–17:00 | Werner | Discussion | Baudel et al. | Discussion |

| W3 – Degaetano-Ortlieb et al. (H1) | |
|------------------------------------|-----------------|
| 12:30–13:30 | Lunch break |
| 13:30–13:45 | Introduction |
| 13:45–14:10 | Al-Laith et al. |
| 14:10–14:35 | Jenset et al. |
| 14:35–15:00 | Maurer et al. |
| 15:00–15:25 | Marr |
| 15:25–16:00 | Break |
| 16:00–16:25 | Schlechtweg |
| 16:25–16:50 | Amaral et al. |
| 16:50–17:15 | Rönchen et al. |
| 17:15–17:40 | Dereza et al. |
| 17:40–17:50 | Closing remarks |

DAY 5: FRIDAY, SEPTEMBER 8th

D5: General sessions

| | Language contacts (H6) | Internal reconstruction (H7) | Mood and modality (H12) | Reconstruction and periodization (H8) | Language corpora (H5) |
|-------------|------------------------------------|---------------------------------|----------------------------|--|--------------------------|
| 09:00–10:00 | John Charles Smith (A1 = New Aula) | | | | |
| 10:00–10:30 | <i>Coffee break</i> | | | | |
| 10:30–11:00 | Burns | Bauer | Long | Pacchiarotti et al. | Rosemeyer et al. |
| 11:00–11:30 | lezzi | Sapp et al. | Rodríguez-Somolinos | Billing & Elgh | Sitchinava |
| 11:30–12:00 | Elter | Herce & Cathcart | Westergaard | Brown & Grollemund | Schützler |
| 12:00–12:30 | Zeng | Reinöhl & Ellison | Sternefeld | Pache | Bronikowska |
| 12:30–13:30 | <i>Lunch break</i> | | | | |
| 13:30–14:00 | Capano | Kayenbergh & De Smet | Assenzi | Swanenvleugel | Nieder & Tomaschek |
| 14:00–14:30 | Hirvonen | Gunnink et al. | Salaberri | Mous | Cattafi |
| 14:30–15:00 | Das | van Dam | | Hernáiz | Wieczorek |
| 15:00–15:30 | Olguín-Martínez | | | Franco | |
| 15:30–16:00 | Farewell reception | | | | |
| 16:00–16:30 | | | | | |
| 16:30–17:00 | | | | | |
| 17:00–17:30 | | | | | |
| 17:30–18:00 | | | | | |

D5: Workshops

| | W8 – Karim & Gholami (H2) | W11 – Auderset et al. (H12a) | W14 – Jäger et al. (H1) |
|-------------|------------------------------|---------------------------------|----------------------------|
| 10:00–10:30 | <i>Coffee break</i> | | |
| 10:30–11:00 | Karim & Gholami | Božović | Jäger et al. |
| 11:00–11:30 | Kreidl | Lionnet | Brigada Villa et al. |
| 11:30–12:00 | Kim | Kirby & Pittayaporn | Rzyski |
| 12:00–12:30 | Mohammadirad | Arnold | Forkel & Greenhill |
| 12:30–13:30 | <i>Lunch break</i> | | |
| 13:30–14:00 | Gholami & Naghshbandi | Grimm | Dellert & Blaschke |
| 14:00–14:30 | Belelli | Sæbø & Grossman | Blum & List |
| 14:30–15:00 | Suleymanov | Perekhvalskaya & Vydrin | Mertner & Guzmán Naranjo |
| 15:00–15:30 | | Auderset & Dockum | Discussion |
| 15:30–16:00 | Farewell reception | | |
| 16:00–16:30 | | | |

Paul Kiparsky

Word-order cycle

The development from head-final to head-initial syntax in branches of Indo-European, Uralic, Sino-Tibetan, Niger-Congo, and Afro-Asiatic is a notable instance of the problematic phenomenon of convergent long-term “drift”. Less often discussed is the contrary development of head-initial to head-final syntax, documented in Austro-Asiatic and Austronesian, and inferrable by internal reconstruction also elsewhere. I argue that these two opposite “drifts” are phases of a larger word-order cycle, and propose a causal mechanism for it, formally grounded in a system of violable linearization constraints that derives a word order typology:

- (1) 1. XP<HEAD: Heads follow their complements (= generalized subject-predicate order).
 2. OP<XP: Functional heads (operators) precede their complements.
 3. C<XP: Complementizers (subordinating operators) precede their complements.
- (2) HARMONY: If A dominates B, then A and B have the same headedness. (cf. Hawkins 1994)

If these constraints are defined on overt surface syntactic structure above the word level (as required by typological evidence such as the FOFC), they predict an important correlation between word structure and syntactic headedness: languages that have no overt syntactic functional heads, but express functional information by inflecting words, are head-final. The grammaticalization of lexical elements into functional heads, and the further reduction of these to clitics and affixes, then have consequences for syntactic headedness by constraints (1) and (2), which can be leveraged into an explanation for the word order cycle. Formally, “drift” can be seen as the result of learners’ bias at each stage of acquisition for the most probable language that is consistent with what they have already learned, where the probability of a language L is measured by its ranking volume, the proportion of fully ranked constraint systems that generate L.

Aditi Lahiri

Phonological grammars: Pertinacious constraints on change

'Not everything goes': a familiar phrase, applicable to phonological and morpho-phonological variation, observable both in synchronic systems and in change. Variability occurs on all levels – segmental, metrical as well as tonal. Critical sources of variation range from differences in vocal tract sizes, regular phonological alternation followed by the attrition of phonological contexts of regular rules and of course loans, leading to the maximal modification of reanalysis.

The level of variation as well as change, we will claim, is however, severely constrained. The hypothesis entertained is the following: phonological opacity may lead to varying choices for native speakers, and the resulting choice is governed by existing phonological preferences. Evidence that the native system plays a constraining influence comes from detailed examination of texts and poetry from Germanic languages (Dutch, English, German, Norwegian, Swedish) and Bengali. Phonological nonesuches (segmental, quantity and tonal) could change the statistical preferences but at each stage the phonological grammar has a restrictive effect.

Lutz Marten

Historical linguistics and Ubuntu translanguaging: Towards a model of multilingualism, language change and linguistic convergence in the Bantu Linguistic Area

Bantu languages, a group of about 300-400 languages spoken across Central, Eastern, and Southern Africa, have long been recognised as genetically related. However, the internal relation and classification of Bantu languages remains a puzzle. In part, this is because in addition to sharing long-term historical relations, Bantu languages have been in close contact for long periods, and so the languages exhibit features of both a language family and a linguistic area or spread zone. A specific aspect of this overall picture are morphosyntactic convergence effects which result in increased similarity of languages at the centre of the Bantu-speaking area ('centripetal convergence'), obscuring genetically based classification boundaries. The situation is not easy to understand within established models of language relationships employing tree or wave metaphors. In contrast I will explore the conceptual framework of Ubuntu Translanguaging (Makalela 2019) which directly addresses multilingualism and the fluidity of linguistic practices. The talk shows how this model provides a novel way of understanding complex linguistic relationships of the past and their repercussions in the present.

Literature

Makalela, Leketi. 2019. 'Uncovering the universals of *ubuntu* translanguaging in classroom discourses'. *Classroom Discourse* 10 (3-4): 237-51. <https://doi.org/10.1080/19463014.2019.1631198>.

Stefan Engelberg, Annette Klosa-Kückelhaus, Peter Meyer, Samira Ochs,
Jan Oliver Rüdiger and Sascha Wolfer

Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language

The Department of Lexical Studies at the Leibniz-Institut for the German Language has a special research focus on the variation and the dynamics of the lexicon of German. We study lexical borrowing, processes of neology and archaization, variation in lexical syntagmatics, and quantitative distribution patterns in the lexicon. Since our institute is also committed to making digital resources available to the scientific community (corpora, dictionaries, online grammars, research tools), many of our research endeavors are accompanied by digital developments. In pursuit of these goals, the Department of Lexical Studies has made available Internet dictionaries, lexical research platforms, and digital exploration tools.

The talk will show how these tools and platforms are being developed in the context of specific research projects. Since lexical changes and innovations are characterized by a comparatively high speed and by their susceptibility to influences from contact languages, many of our digital applications are designed with these features in mind.

After outlining the research orientation of our department, the talk will present four digital applications as examples and describe their development, their connection to our research, and their potential usefulness to other researchers: (i) a dictionary of neologisms and the process of their corpus-based determination, (ii) a tool for the monitoring of lexical changes in real-time corpora, (iii) an application resulting to gender-linguistic research visualizing lexical change in role nouns and job titles, and (iv) a research platform for exploring lexical borrowings from German into other languages.

Ans van Kemenade

Word order change, architecture and interfaces: Evidence from V2 word orders and their loss in the history of English

Word order change, architecture and interfaces: evidence from V2 word orders and their loss in the history of English

Present-day English stands out in the Germanic context as a language that has quite limited Verb Second (V2) word order, in *wh*-questions (*why did you go there?*) and in exclamatives with initial negatives (*Not one of them did he find useful!*). V2 here stands for a word order X-Vf-subject. Early English has been shown to have a wider range of “V2” word orders, where X is more variable, Vf includes lexical finite verbs, and the subject is often not inverted with Vf, especially when it is a pronoun. This paper presents a detailed case study of the history of V2 word orders, showing that the attested synchronic variation and the pathways of change crucially involve interaction between syntactic constraints, information structure and pragmatics, and prosody, illuminating how syntactic change is subject to pressure from interface conditions.

From the earliest stages, English has more (patterned) variation than we know so far of the other Old Germanic languages. I distinguish three V2-like patterns in OE, which are subject to different constraints and follow different trajectories of change and loss over the Middle and early Modern English periods. One factor that they have in common is that verb fronting of lexical finite verbs was lost over the 16th century, leaving auxiliary fronting only:

- 1) questions, initial negatives and clauses introduced by *then* which show categorical inversion of all types of finite verb and subject. This pattern was partially lost (following adverbs like *then*) over the 16th century and became restricted to auxiliaries in questions and negative-initial clauses;
- 2) X-initial clauses with transitive/unergative intransitive verbs, which mostly show inversion of nominal subjects, but not of pronominal subjects; this pattern was lost over the late 15th century;
- 3) X-initial clauses with unaccusative verbs, which show more inversion of nominal subjects. It can be shown that nominative subjects often occur in low positions in the clause which are essentially object positions. This can but need not represent a typical V2 pattern, and was not lost, living on in present-day English as complex inversion and locative inversion (*In the tank are sitting all of the pots.*)

I present a fine-grained corpus study, based on the on the relevant parsed corpora for OE, ME and EME, further enriched with information about finite verb type, noun type, Information status of the subject, and weight calculations for X, Vf, and subject. This will serve to identify the factors differentially determining the development and loss of V2 word orders (1) and (2) over the Old, Middle and early Modern English periods:

- 4) for initial X: weight, discourse linking, focality, operator status
- 5) for Vf: weight, verb type (auxiliary, unaccusative, transitive/unergative intransitive)
- 6) for subjects: weight, NPtype (bare, quantified/negated, indefinite, definite, demonstrative, proper), Information status (given vs. new)
- 7) syntactic structure, which in OE and early ME allowed for differential positions for nominal and pronominal subjects.

I will argue that the pattern in (1) (when following adverbs like *then*) was lost as auxiliaries were reanalysed as function words over the 16th century, losing primary stress on the stem, leaving the clause-initial prosodic foot unheaded. The pattern in (2) was lost due to the reduction/levelling of discourse-motivated syntactic positions.

Nigel Vincent, George Walkden, Mirjam Fried, Kasper Boye and Ashwini Deo
Linguistic models (with a focus on morphosyntactic change)

The aim of this panel is to consider and evaluate some of the approaches to modelling morphosyntactic change to be found in the current literature. The session will begin with panel members briefly outlining the main properties of the frameworks within which they conduct their own research: generative historical syntax (**George Walkden**), construction grammar (**Mirjam Fried**), functionalism (**Kasper Boye**), formal semantics (**Ashwini Deo**). This will be followed by a discussion among panel members of the comparative strengths and weaknesses of the different models with respect to both our understanding of the mechanisms of morphosyntactic change and the possibility of deploying these frameworks in the reconstruction of earlier historical stages. Discussion will then be opened up to questions from members of the audience.

John Charles Smith

Fifty years of ICHL, 1973–2023

The first International Conference on Historical Linguistics took place during the first week of September 1973 – exactly 50 years ago. In this talk, I shall discuss the intellectual and social history of the Conference, looking at the circumstances which brought it about, its evolution over half a century, and the contribution it has made to the discipline.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

From climate change to language change

Convenor

Martine Robbeets

Max Planck Institute of Geoanthropology, Jena

robbeets@shh.mpg.de

Over the last decades, our Earth has experienced an alarming number of extreme events, such as heatwaves, heavy rainfall, flooding, melt events, drought, forest fires, cyclones, etc. With progressing climate change, such extreme events can be expected to occur more frequently and potentially become more severe (Martin et al. 2021). A new field of study has risen from the ashes of these events: Geoanthropology studies present and past interactions between humans and the Earth system, integrating fields such as Climate science, Earth system science, Ecology, Environmental history, Archaeology, Economics, Law, Anthropology and Political sciences. In our panel, we wish to add linguistics to this list and explore the relevance of Historical Linguistics for the field of Geoanthropology.

How do climate and language connect? The link between the two lies in humans and how they respond to changing conditions and extreme events. Simply put, climate change can affect speaker populations in the following three ways.

(1) *The speaker population declines to extinction*

Disrupting subsistence industries of speakers of endangered languages, climate change is forcing these speakers to assimilate to the language and subsistence strategies of more dominant linguistic groups or to scatter around the globe, thus threatening linguistic survival. For example, as reindeer populations are threatened by climate change, reindeer herders speaking Evenki, a Tungusic language in Northeastern Siberia, are shifting not only to jobs in industry but also to the Russian language.

(2) *The speaker population migrates to a new environment*

By contrast, climate change can also increase linguistic diversity. During the Little Ice Age these Tungusic speakers expanded their territory because colder weather appears to increase reindeer populations (Hudson 2020, Robbeets & Oskolskaya 2022). Moreover, climate change can force populations to move, along with their crops and languages to search for a more viable environment. In such cases, we expect language split between the part of the speech community that stays and the part that leaves, leading to the development of separate daughter languages. The daughter language on the move can either be maintained and interact with contact languages at its new destination, or, alternatively, it can be abandoned, with speakers shifting to a new target language, spoken by a more dominant speech community in the new environment. For example, a large group of Maldivian climate refugees has moved to India or Sri Lanka. Even if the immigrants' language has received substantial influence from Tamil, Hindustani and English, they maintain Dhivehi, spoken in the Maldives, as their native language.

(3) *The speaker population adapts to the changing environment*

Even if certain speech communities manage to stay in place and maintain their native language, they will need to adapt it to the changing local environment (Frainer et al. 2020). This may involve coining new words, losing specific cultural vocabulary, lexical recycling, borrowing from better adapted speakers, etc. Ongoing climate change in Alaska, for instance, created new opportunities for agriculture. In Aleut, the agricultural verbs 'to plant' and 'to sow' are recycled from original hunter-gatherer terminology meaning 'to drop a fishing line' and 'to distribute sea-catch' (Berge 2017).

How can we extrapolate, projecting observable cases of climate-driven language change to reconstruct linguistic prehistory? Geoanthropologists use the designation "Anthropocene" as a unit of geologic time, used to describe the period when human activity started to have a significant impact on our planet's climate and ecosystems. Other suggestions for the starting date being the Industrial Revolution and the invention of the atomic bomb, some researchers argue that the Anthropocene began approximately 8 000 years ago with the development of farming and sedentary cultures (Foley et al. 2013; Smith and Zeder 2013, Renn 2020). This falls within the time frame that can be investigated by applying the traditional historical-comparative linguistic method, the practical cut-off point for this method lying around 10 000 years ago (Comrie 2000; Campbell 2000). It is no coincidence that many of the world's major language families started to disperse around the Neolithic Revolution. For instance, language families such as Bantu (Philipson 2002), Semitic (Diakonoff 1998), Austronesian (Blust 1995, 2013; Pawley 2002; Bellwood & Dizon 2008), Transeurasian (Robbeets et al. 2021), Sino-Tibetan (Sagart et al. 2019, Zhang et al. 2020), Tai-Kadai (Ostapirat 2005), Austroasiatic (Higham 2002, Diffloth 2005, Sidwell and Blench 2011, Sagart 2011, van Driem 2017), Dravidian (Fuller 2002) Arawakan (Aikhenvald 1999), Otomanguean (Kaufman 1990, Brown et al. 2013a/b, 2014a/b) are argued to owe

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

their primary dispersal to the adoption of agriculture by their early speakers. The link between postglacial warming and farming/language dispersals is generally accepted (Richerson et al. 2001, Bellwood 2022: 150) but it remains to be investigated how climate versatility and extreme events in specific regions may have influenced language loss, change and dispersal.

Our panel proposes a wide range of questions stressing the need of case studies that illustrate in what ways climate reshaped individual languages and language families across the world. Is climate change threatening certain languages and accelerating language loss of already endangered languages? Can climate change also have a positive effect on linguistic diversity, leading to the birth of new daughter languages? What is the relation between the reduction of biological, cultural and linguistic diversity through climate change? What is the reason for/ mechanism behind the correlations? Can the conservation of species be expected to lead to the conservation of languages? Can regions that have high biodiversity be linked to the development of linguistic diversity? Can we correlate established periods of climate change in a certain region in prehistory with periods of linguistic dispersal and language loss? Do dated trees of individual language families support such a correlation? Can we extrapolate our understanding of climate-driven language change not only to reconstruct the past but also to predict the future? In what way and to which extent did the emergence of the Anthropocene impact language loss, dispersal and change? What is the influence of extreme events on language diversification? Can the impact of extreme events be modeled, for instance by Dixon's (1997) equilibrium/punctuation model or by Hudson's (2017) adaptive cycle model? Are there case studies that illustrate the impact of extreme events on language change? What is the impact of time on climate-driven language change? Is it reasonable to expect that linguistic diversity will restore at a higher speed than biological diversity? What is the role of climate in proposals like "the Farming/Language Dispersal Hypothesis" (Bellwood & Renfrew 2002), which posits that many of the world's major language families owe their dispersal to the adoption of agriculture by their early speakers?

References

- Bellwood, Peter 2022. *The Five-Million-Year Odyssey. The human journey from ape to agriculture*. Princeton: Princeton University Press.
- Bellwood, Peter & Dizon, Eusebio. 2008. Austronesian cultural origins: Out of Taiwan, via the Batanes Islands, onwards to Western Polynesia. In *Past Human Migrations in East Asia, Matching Archaeology, Linguistics and Genetics*, Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross, Ilia Peiros & Marie Lin (eds), 23–39. London: Routledge.
- Bellwood, Peter & Renfrew, Colin (eds). 2002. *Examining the Farming/Language Dispersal Hypothesis*. Cambridge: McDonald Institute for Archaeological Research.
- Berge, Anna. 2017. Subsistence terms in Unangam Tunuu (Aleut). In *Language Dispersal Beyond Farming*, Robbeets, Martine & Alexander Savelyev (eds.), 47-74. Amsterdam: Benjamins.
- Blust, Robert. 1995. The prehistory of the Austronesian-speaking peoples: a view from language. *Journal of World Prehistory* 9: 453–510.
- Blust, Robert. 2013. *The Austronesian Languages* [Asia-Pacific Linguistics. Open access monographs]. Canberra: Australian National University.
- Brown, Cecil et al. 2013a. The paleobiolinguistics of domesticated chili pepper (*Capsicum* spp.). *Ethnobiology Letters* 4: 1–11.
- Brown, Cecil et al. 2013b. The paleobiolinguistics of domesticated manioc (*Manihot esculenta*). *Ethnobiology Letters* 4: 61–70.
- Brown, Cecil et al. 2014a. The paleobiolinguistics of domesticated maize (*Zea mays* L.). *Ethnobiology Letters* 5: 52–64.
- Brown, Cecil et al. 2014b. The paleobiolinguistics of the common bean (*Phaseolus vulgaris* L.). *Ethnobiology Letters* 5: 104–115.
- Campbell, Lyle. 2000. Time perspective in linguistics. In *Time depth in historical linguistics*, Renfrew, Colin and McMahon, Alice (eds), 3-31. Cambridge : McDonald Institute for Archaeological Research
- Comrie, Bernard 2000 Is there a single time depth cut-off point in historical linguistics? In *Time depth in historical linguistics*, Renfrew, Colin and McMahon, Alice (eds), 34-60. Cambridge : McDonald Institute for Archaeological Research
- de Boer, Elisabeth. 2010. *The Historical Development of Japanese Tone: From Proto-Japanese to the Modern Dialects [and] the Introduction and Adaptation of the Middle Chinese Tones in Japan*. Wiesbaden: Harrassowitz.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

- de Boer, Elisabeth M. 2020. The Classification of the Japonic Languages. *Transeurasian Languages*, ed. M. Robbeets & A. Savelyev, 40–58. Oxford: Oxford University Press.
- Diakonoff, Igor. 1998. The earliest Semitic society. *Journal of Semitic Studies* 43 (2): 209–219.
- Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austroasiatic. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 77–80. London: Routledge Curzon.
- Dixon, Robert M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.
- Foley, Stephen F. et al. 2013. The Palaeoanthropocene: the beginnings of anthropogenic environmental change. *Anthropocene* 3:83–88
- Fortescue, Michael & Edward Vajda. 2022. *Mid-Holocene language connections between Asia and North America*. BSILA 17. Leiden/ Boston: Brill.
- Fuller, Dorian. 2002. An agricultural perspective on Dravidian historical linguistics: Archaeological crop packages, livestock and Dravidian crop vocabulary. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 191–214. Cambridge: McDonald Institute for Archaeological Research.
- Frainer, André et al. 2020. Cultural and linguistic diversities are underappreciated pillars of biodiversity PNAS 117(43): 26539–26543 www.pnas.org/cgi/doi/10.1073/pnas.2019469117
- Francis-Ratté, Alexander T. 2016. Proto-Korean-Japanese: A New Reconstruction of the Common Origin of the Japanese and Korean Languages. Ph.D. dissertation, the Ohio State University.
- Francis-Ratté, Alexander T., and J. Marshall Unger. 2020. Contact between Genetically Related Languages: the Case of Old Korean and Old Japanese. In Robbeets & Savelyev 2020, 705–14.
- Gorenflo, L. J. et al. 2012. Co-occurrence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas PNAS 109: 8032–8037
- Higham, Charles. 2002. Languages and farming dispersal: Austroasiatic languages and rice cultivation. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 223–232. Cambridge: McDonald Institute for Archaeological Research.
- Hudson, Mark J. 2017. Socio-ecological resilience and language dynamics: An adaptive cycle model of long-term language change. *Journal of Language Evolution*, 2018, 1–9. doi: 10.1093/jole/lzy008
- Hudson, Mark J. 2020. Language dispersals and the “Secondary Peoples’ Revolution”:
A historical anthropology of the Transeurasian unity. In: *The Oxford guide to the Transeurasian languages*, Robbeets, Martine & Savelyev, Alexander (eds.), 806–813. Oxford: Oxford University Press.
- Hudson, Mark J. et al. 2012. Prospects and challenges for an archaeology of global climate change. *WIREs Clim Change* 3:313–328. doi: 10.1002/wcc.174
- Kaufman, Terence. 1990. Early Otomanguean homelands and cultures: some premature hypotheses. *University of Pittsburgh Working Papers in Linguistics* 1: 91–136.
- Martin M.A. et al. 2021. Ten new insights in climate science 2021: a horizon scan. *Global Sustainability* 4: e25, 1–20. <https://doi.org/10.1017/sus.2021.25>
- Ostapirat, Weera. 2005. Kra-dai and Austronesians. In *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics*, Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds), 107–131. London: Routledge Curzon.
- Pawley, Andrew. 2002. The Austronesian dispersal: languages, technologies and people. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 251–274. Cambridge: McDonald Institute for Archaeological Research.
- Philipson, David. 2002. Language and farming dispersals in sub-Saharan Africa, with particular reference to the Bantu-speaking peoples. In *Examining the Language/Farming Dispersal Hypothesis*, Peter Bellwood & Colin Renfrew (eds), 177–190. Cambridge: McDonald Institute for Archaeological Research.
- Renn, Jürgen 2020. *The evolution of knowledge: Rethinking science for the Anthropocene*. Princeton: Princeton University Press.
- Richerson, P. J., Boyd, R., & Bettinger, R. L. 2001. Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *American Antiquity*, 66(3): 387–411.
- Robbeets, Martine et al. 2021. Triangulation supports agricultural spread of the Transeurasian languages. *Nature* 599: 616–621. DOI 10.1038/s41586-021-04108-8 <https://www.nature.com/articles/s41586-021-04108-8>

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

- Robbeets, Martine, and Alexander Savelyev (ed.) 2020. *The Oxford Guide to Transeurasian Languages*. Oxford: Oxford University Press.
- Robbeets, Martine & Oskolskaya, Sofia 2022. Proto-Tungusic in Time and Space. In *Tungusic languages: Past and present*, Hölzl, Andreas & Payne, Thomas E. (eds), 263–294. Berlin: Language Sciences Press.
- Romaine, Suzanne & Gorenflo, L. J. 2017. Linguistic diversity of natural UNESCO world heritage sites: bridging the gap between nature and culture. *Biodivers Conserv* (2017) 26:1973–1988. DOI 10.1007/s10531-017-1340-x
- Sagart, Laurent et al. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proc. Natl Acad. Sci.* **116**, 10317–10322 (2019).
- Sagart, Laurent. 2011. How many independent rice vocabularies in East Asia? *Rice* 4 (3): 121–133.
- Sidwell, Paul & Blench, Roger. 2011. The Austroasiatic Urheimat: the Southeastern Riverine Hypothesis. In *Dynamics of Human Diversity: the Case of Mainland Southeast Asia*, Nick Enfield (ed.), 315–343. Canberra: Pacific Linguistics.
- Smith, Bruce D. & Zeder Melinda A. 2013 The onset of the Anthropocene. *Anthropocene* 4:8–13.
- Unger, J. Marshall. 2001. Layers of Words and Volcanic Ash in Japan and Korea. *Journal of Japanese Studies*, vol. 27 no. 1 pp. 81–111.
- Unger, J. Marshall. 2009. *The Role of Contact in the Origins of the Japanese and Korean Languages*. Honolulu: University of Hawai'i Press.
- van Driem, George 2017 The domestications and the domesticators of Asian rice In *Language Dispersal Beyond Farming*, Robbeets, Martine & Alexander Savelyev (eds.), 183-214. Amsterdam: Benjamins.
- Zhang, H. et al. 2020 Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Nature Scientific Reports* 10: 20792 .

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Prehistoric climate changes and their effects on the development of the Eskaleut languages Anne Berge

Prehistoric climate change, population movements, and language contact in the Bering Sea region are intimately connected. The archaeological and paleo-environmental records consistently show cooler climatic periods associated with more abundant marine resources and population expansions, while warming periods correlate with marine instability, region-wide population stresses, decreases, and migrations, as well as evidence of warfare. Although we see this in the smaller climate fluctuations at local levels, the two biggest changes in the past 4000 years coincide with the most important linguistic splits in the Eskaleut language family.

Eskaleut consists of two major branches: Aleut, with a single language spoken today, Unangam Tunuu, and Eskimo, with two major branches, Yupik and Inuit. The age of Proto-Eskaleut is generally put somewhere between 6000 BP and 4000 BP, during the Neoglacial period in the Bering Strait area. Unangam Tunuu (Aleut) split off first, probably via an independent migration ca. 4500 BP, becoming an independent language by ca. 3500 BP (Berge 2018). This timing corresponds almost exactly both with the end of the Neoglacial period and with a massive volcanic eruption that isolated the Eastern Aleutians from the Alaskan mainland and the related culture on Kodiak Island (Maschner 2016), leading to their linguistic differentiation (Berge forthcoming). The eruption caused a catastrophic population crash in central western Alaska, leading to movements from the interior to the coast and significant cultural changes associated with the development of Proto-Yupik culture, although not necessarily language (Tremayne and Brown, 2017).

The warmer period that followed the Neoglacial allowed the spread of whales northward into the Bering Sea, and consequently to the development of the whaling cultures later associated with the Yupik and Inuit peoples on the Siberian coast (Crockford and Frederick 2007). Despite local variations in climate, the next 2000 years were relatively stable and cool (although not glacial), allowing these cultures to flourish, particularly from 2000–1100 BP. From about 1000 BP, the climate warmed significantly, with drastic consequences. In the earliest part of this Medieval Climate Optimum, one of these cultures spread out aggressively from Siberia to Alaska (Mason 2009), precipitating a period of intense societal destabilization in northern coastal Alaska. Around 800 BP, there was a sudden and a very rapid emigration from this part of Alaska and colonization of the northern Canadian arctic to Greenland, a movement associated specifically with the development and spread of Inuit. Although Moss et al. (2007) find no evidence linking this expansion with the start of the climate change, the earlier migration from Siberia does correlate with the change. A concurrent Inuit expansion southward in Alaska precipitated five centuries of tribal wars and population displacements in Yupik areas (Funk 2010). This movement resulted in the arrival of the Yupik language Alutiiq to the Pacific Coast, its replacement of Unangam Tunuu on Kodiak Island (Berge, forthcoming) and the dialect leveling of Unangam Tunuu along the Aleutians (Woodbury 1984).

Climate change is certainly not the only factor in linguistic development. Natural disasters such as the volcanic eruption at the end of the Neoglacial may be a more direct cause of the development of Unangam Tunuu. Other factors include resource depletion as a result of increases in human population, activity, or improvements in technology; and cultural contact through trade, warfare, etc. have all affected the development of the Eskaleut languages. Nevertheless, when climate changes occurred, they acted as significant stressors leading to isolation, migration, or warfare. In this paper, I discuss how important prehistoric climate changes have been on the development of the Eskaleut languages.

References

- Berge, Anna. Forthcoming. Signs of prehistoric language shifts involving Unangam Tunuu (Aleut). *Proceedings of the International Conference on Historic Linguistics 24*. Amsterdam: John Benjamins.
- Berge, Anna. 2018. Re-evaluating the linguistic reconstruction of Proto-Eskimo-Aleut. *Journal of historical linguistics*, 8(2). 231–273.
- Crockford, Susan J. & S. Gay Frederick. 2007. Sea ice expansion in the Bering Sea during the Neoglacial: Evidence from archaeozoology. *The Holocene* 17(6). 699–706.
- Funk, Caroline. 2010. The bow and arrow war days on the Yukon-Kuskokwim Delta of Alaska. *Ethnohistory* 57(4). 523–569. DOI 10.1215/00141801-2010-036
- Mason, Owen K. 2009. Flight from the Bering Strait: Did Siberian Punuk/Thule military cadres conquer Northwest Alaska? In Herbert Maschner, Owen Mason & Robert McGhee (eds.), *The Northern World, AD 900-1400*, 76–130. Salt Lake City: University of Utah Press.
- Moss, Madonna L., Peteet, Dorothy M., and Whitlock, Cathy. 2007. Mid-Holocene culture and climate on the Northwest Coast of North America. In David G. Anderson, Kirk A. Maasch and Daniel H. Sandweiss (eds.) *Climate Change and Cultural Dynamics: A Global Perspective on Mid-Holocene Transitions* 491–529. Elsevier.

References

- Hattori, Shirō. 1999 [1959]. *Nihongo no Keitō* (The lineage of the Japanese language). Tokyo: Iwanami Shoten.
- Iha, Fuyū. 2000 [1910]. P on'kō (Study of the sound p). *Koryūkyū* (Old Ryukyu). (Hokama Shuzen, ed. and revisor.) Iwanami Bunko blue series, N 102.1. Tokyo: Iwanami Shoten. pp.375-386.
- Kiku, Chiyo and Toshizō Takahashi. 2005. *Yoron Hōgen Jiten* (Dictionary of Yoron dialect). Musashino Shoin.
- Li, Chihkai. 2015. *A Reconstruction of Old Okinawan: A corpus-based approach*. Ph.D. dissertation, University of Hawai'i at Manoa.
- Nakamoto, Masachie. 1976. *Ryūkyū Hōgen On'in no Kenkyū* (Study of the Ryukyuan dialect phonology). Tokyo: Hōsei Daigaku Shuppanyoku.
- Tawata, Shin'ichirō. 2010. *Okinawago On'in no Rekishiteki Kenkyū* (Historical Study of the Okinawan language phonology). Hiroshima: Keisuisha.
- Thorpe, Maner L. 1983. *Ryukyuan Language History*. University of Southern California Doctoral Dissertation.
- Uchima, Chokujin and Nohara Mitsuyoshi. 2006. *Okinawa-go Jiten: Naha Hōgen o Chūshin ni* (Dictionary of the Okinawan language: Focusing on the dialect of Naha). Tokyo: Kenkyūsha.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Seals and sea ice: the (possible) climatic background of Amuric influence on Ainu Martijn Knapen

Amuric is a small language family historically and presently spoken on the Lower Amur and Sakhalin Island by the Nivkh people. Since the thirteenth century at least, the language family has shared Sakhalin with the local variety of Ainu (Janhunen 2022a). While this variety extensively borrowed from Amuric (Shiraishi and Tangiku 2022), there is also Amuric linguistic material in Proto-Ainu, the ancestor to the modern Ainu varieties, which implies their interaction predates the arrival of Ainu to Sakhalin. For contact to occur, two languages must have been present in the same location. The most likely location appears to be Hokkaidō, which therefore suggests an early Amuric presence there (Vovin 1993; 2016).

Before Amuric spread to Hokkaidō or even Sakhalin, it was most likely spoken in the southern section of the Amur Basin, near the Ussuri and Sungari, as recent research suggests (cf. Janhunen 2022b; Knapen, in press). The trajectory of its expansion mirrors that of the Okhotsk culture, an archaeological culture that has its origins on the Amur and was present on northern Hokkaidō from 550 AD to 1200 AD. It was characterised by heavy reliance on marine resources and was noticeably distinct from its contemporary neighbours on Hokkaidō, the Epi-Jomon (100BC-550 AD) and Satsumon (600-1200 AD) cultures, the predecessors of later Ainu culture (Hudson 2004). The impetus for the migration of the Okhotsk culture to Hokkaidō may have been a cold period that lasted from 150 AD to 650 AD, which resulted in increased sea ice on the Sea of Okhotsk and with that improved conditions for hunting pinnipeds (Abe et al. 2016). As the bearers of the Okhotsk culture are often suggested to be related to the Nivkh (but also various other modern Northeast Asian ethnic groups) (Zgusta 2015), these climatic conditions could also be tied to the spread of the Amuric language family to Hokkaidō. The main indeterminate here is whether Amuric may be regarded as the language spoken by the bearers of the Okhotsk culture. This problem is approached from the perspective of linguistic palaeontology (Heggarty 2014): by reconstructing terminology suggestive of familiarity with a particular way of life, the homeland of a particular proto-language is inferred. The vocabulary in this case will be delimited to items referring to local marine fauna and the exploitation of such resources. The procedure follows Janhunen's (2016) approach, by not just considering Amuric internal data (cf. Fortescue 2016), but also external data, primarily from the Tungusic languages. To avoid circularity, Ainu data is not included. Additionally, the evidence of Amuric-Ainu contact proposed by Vovin (1993; 2016) is evaluated as well as further connections. This evidence is then set against other hypotheses on the linguistic identity of the Okhotsk culture. Aside from advancing the line of inquiry started by Vovin (1993; 2016), this paper will provide further clarification to the (linguistic) prehistory of northeast Asia and its indigenous peoples, for which written records are scarce, as well as the possibility of climatic factors influencing language dispersal.

References

- Abe, Chiharu, Christian Leipe, Pavel E Tarasov, Stefanie Müller, and Mayke Wagner. 2016. 'Spatio-Temporal Distribution of Hunter-Gatherer Archaeological Sites in the Hokkaido Region (Northern Japan): An Overview'. *The Holocene* 26 (10): 1627–45.
- Fortescue, Michael. 2016. *Comparative Nivkh Dictionary*. Munich: LINCOM.
- Heggarty, Paul. 2010. 'Prehistory through Language and Archaeology'. In *The Routledge Handbook of Historical Linguistics*. Routledge.
- Hudson, Mark J. 2004. 'The Perverse Realities of Change: World System Incorporation and the Okhotsk Culture of Hokkaido'. *Journal of Anthropological Archaeology* 23 (3): 290–308.
- Janhunen, Juha. 2016. 'Reconstructio Externa Linguae Ghiliacorum'. *Studia Orientalia* 117: 3–27.
- Janhunen, Juha A. 2022a. 'Ainu Ethnic Origins'. In *Handbook of the Ainu Language*, edited by Anna Bugaeva, 57–78. De Gruyter.
- Janhunen, Juha A. 2022b. 'Amuric Hydronyms in Manchuria and the Puyō Connection of Ghilyak'. *International Journal of Eurasian Linguistics* 4 (1): 23–40.
- Knapen, Martijn G.T.M. In press. 'Amuric-Tungusic Language Contact and the Amuric Homeland'. In *Agropastoralism and Languages Across Eurasia: Expansion, Exchange, Environment*, edited by Mark Hudson and Martine Robbeets. Archaeology of East Asia. Oxford: BAR Publishing.
- Shiraishi, Hidetoshi, and Itsuji Tangiku. 2022. 'Language Contact in the North'. In *Handbook of the Ainu Language*, edited by Anna Bugaeva, 209–52. De Gruyter.
- Vovin, Alexander. 1993. *A Reconstruction of Proto-Ainu*. Brill's Japanese Studies Library, v. 4. Leiden ; New York: E.J. Brill.
- Vovin, Alexander. 2016. 'ON THE LINGUISTIC PREHISTORY OF HOKKAIDŌ'. *Studia Orientalia*, no. 117: 29–38.
- Zgusta, Richard. 2015. *The Peoples of Northeast Asia through Time: Precolonial Ethnic and Cultural Processes along the Coast between Hokkaido and the Bering Strait*. Leiden: Brill.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Spread of Proto Japanese from Korean Peninsula to Japanese Archipelago influenced by natural environment change

Kazuo Miyamoto

It is believed that there were four stages of spread of early agriculture in North-East Asia (Miyamoto 2014, 2015). The first stage involved the spread of millet agriculture to the Korean Peninsula and to the southern Russian Far East in the middle of the fourth millennium BC. The second stage was the spread of rice agriculture from the Shandong Peninsula to the Liaodong Peninsula at c. 2400 BC. The third stage, in the middle of the second millennium BC, consisted of irrigated agriculture and spread from the Shandong Peninsula via Liaodong Peninsula to the Korean Peninsula. Finally, the fourth stage involved the spread of irrigated agriculture from the southern Korean Peninsula to Northern Kyushu, Japan, beginning about 9th century BC. These four stages were triggered by immigrants due to cooler climate conditions and the development of farming society.

The fourth spread of early agriculture from 9th to 8th century BC is spread of irrigated rice agriculture with rice paddy field from southern Korean Peninsula to Northern Kyushu. This spread was triggered by the immigration from Southern Korea to Northern Kyushu to get new lands for cereal agriculture due to cooler climate conditions (Miyamoto 2016, 2019). The spread direction of irrigated rice agriculture from Korean Peninsula to Northern Kyushu was divided into two phases. The former phase is immigration from Namgang River basin to Karatsu and Itoshima Plains at 9th to 8th century BC. The latter phase is immigration from lower Nagdong River basin to Fukuoka Plains at 7th century BC. Those dual phases accorded to cooler climate conditions (Miyamoto 2016, 2019).

These dual immigrations speaking Proto Japanese in Korean Peninsula spread to Northern Kyushu mixed with Jomon people speaking Jomon Languages. They invented Yayoi culture in Fukuoka Plains based on Mumun culture in southern Korean Peninsula at 6th to 5th century BC. In this time, Yayoi culture people in Fukuoka Plain replaced Proto Japanese from Jomon languages (Miyamoto 2016, 2022).

Yayoi culture originated from Fukuoka plain spread immediately to the whole of western Japan at 6th to 5th century BC. This spread of Yayoi culture is also spread of Proto Japanese replaced from Jomon Languages in the Western Japan. The spread of Yayoi culture with irrigated agriculture was promoted by demographic pressure due to the stable weather conditions.

References

- Miyamoto, K. 2014 Human Dispersal in the Prehistoric Era in East Asia. Nakahashi, T. and Fan, W. ed. *Ancient People of the Central Plains in China*. Kyushu University Press. Fukuoka: 63-83.
- Miyamoto, K. 2015 The Initial Spread of Agriculture into Northeast Asia. *Asian Archaeology* Vol.3: 11-26.
- Miyamoto, K. 2016 Archeological Explanation for the Diffusion Theory of the Japonic and Koreanic Languages. *Japanese Journal of Archaeology* 4 (1): 53-75.
- Miyamoto, K. 2019 The spread of rice agriculture during the Yayoi period: From the Shandong Peninsula to Japanese Archipelago via Korean Peninsula. *Japanese Journal of Archaeology* 6(2): 109-124.
- Miyamoto, K. 2022 The emergence of ‘Trans Eurasian’ language families in Northeast Asia as viewed from archaeological evidence. *Evolutionary Human Sciences* 4, 2022, e3, 1-17, doi:10.1017/ehs.2021.49

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Climate change and the dispersal of Proto-Tibeto-Burman

David Bradley

From a likely origin in the Majiayao Culture of what is now Gansu in China from circa 5.3K YBP, the Proto-Tibeto-Burman (PTB) community migrated and divided rapidly during times of ancient climate change. Their initial agriculture was mainly based on *Setaria* and *Panicum* millet and rice, and their domestic animals were dogs, pigs and *Bos Taurus* cows (Liu & Chen 2012). Etyma for these three crops and three domestic animals are reconstructed for PTB (Bradley 2011, 2016, 2022). Majiayao was a western offshoot of Proto-Sino-Tibetan (PST) Yangshao Culture, which flourished to the east of Majiayao from circa 7K-5K YBP during a period of favourable climate, cultivated the two millets and had domestic dogs and pigs, and later developed into Sinitic Longshan Culture. Domestic taurine cows were introduced from the west circa 5.6K YBP (Brunson et al. 2020), and the PST COW etymon * $\eta w\grave{a}$ supports dating the PTB/Sinitic split to after 5.6K YBP. Rice was first domesticated in the lower Yangtze area by circa 6K YBP and later spread northwest to late Yangshao and early Majiayao cultures (Fuller et al. 2007), with PTB but no earlier PST etyma.

Subsequent PTB migrations were shaped by climate change; firstly, a warm and wet climate from circa 5K YBP, which permitted cultivation of these crops at higher altitudes in eastern Tibet and western Sichuan (d'Alpoim Guedes et al. 2014, 2016). Later periods of cooling climate (Cheung et al. 2019, Chen et al. 2020) perhaps triggered further migrations beyond southwest China, with the Karenic subgroup reaching west Southeast Asia and the Central subgroup reaching northeast South Asia. Ecological changes led to shifts in crops and domestic animals, with contact introducing some new crops and animals. This discussion will trace the lexical outcomes for crop and domestic animal vocabulary and show how archaeologically documented dates for contact-introduced and newly-domesticated crops and animals can assist to date the early phylogeny of PTB.

Two crops arrived from the west circa 4.5K YBP: *Triticum* and *Hordeum*. Unlike *Setaria* and *Panicum*, these can adapt to cooler climate, so their cultivation spread and increased rapidly with cooling climate from circa 4.2K YBP. Rice was also more suitable for some new ecological niches. Two domestic animals also adaptable to cooler climate, sheep and goats, were introduced from the west circa 4.4K YBP (Liu & Chen 2012). The subgroup of PTB which on independent comparative evidence appears to have separated first from PTB, Karenic, lacks cognates of PTB etyma for WHEAT, BARLEY and GOAT; it has a cognate of the PST and PTB etymon for wild BOVID * $\text{ja}\eta$, also present in Sinitic (the later Longshan Culture offshoot of Yangshao Culture in its original area and further east) and in PTB. The cognate means 'goat' in Karenic, while it means 'sheep' in the rest of PTB, and both in Sinitic; the PTB GOAT etymon is * c^hit (Bradley 2022). Thus the split of Karenic from PTB may have preceded 4.5K YBP.

Bos grunniens (yak) was probably domesticated by 3.65K YBP (Jacques et al. 2021) and cultivation of *Hordeum vulgare* var. *nudum*, a variety of barley suitable for cold climate (d'Alpoim-Guedes et al. 2015, Zeng et al. 2015) developed in parts of the area during expansion into higher-altitude environments such as the Karuo Culture, and expanded during the cold climate period from circa 3.5K YBP. The horse was introduced from the west into China circa 3.3K YBP (Liu & Chen 2012). *Fagopyrum* (buckwheat) cultivation started in upland southwest China by circa 3.15K YBP (Xue et al. 2022). These developments are reflected by the distribution of etyma for these crops and animals among TB languages. A YAK etymon has cognates in Eastern and Western TB but not Central TB. Western TB and Eastern TB have distinct BUCKWHEAT etyma; the latter is borrowed into Chinese. The forms for HORSE are loans, with a wide variety of alternative forms, including various similar Eastern TB forms, a completely different Western TB form, also Indic loans in Central TB and some Western TB languages south of the Himalayas, and another form in Karenic languages. Overall, this suggests that the second split within PTB was Central TB, perhaps associated with the 4.2K YBP climate cooling, followed by a later split between Western and Eastern TB associated with the 3.5K YBP climate cooling, after the domestication of the yak but before the introduction of the horse circa 3.3K YBP and before the domestication of buckwheat.

References

- d'Alpoim Guedes, Jade et al. (2014) Moving agriculture onto the Tibetan plateau: The archaeobotanical evidence. *Archaeological and Anthropological Sciences* 6(3): 255–69.
- d'Alpoim Guedes, Jade et al. (2015) Early evidence for the use of wheat and barley as staple crops on the margins of the Tibetan plateau. *Proceedings of the National Academy of Sciences* 112(18): 5625–5630.
- d'Alpoim Guedes, Jade et al. (2016) A 5,500 year model of changing crop niches on the Tibetan plateau. *Current Anthropology* 57(4).
- Bradley, David (2011) Proto-Tibeto-Burman grain crops. *Rice* 4(3-4): 134-141.
- Bradley, David (2016) 《山海经》与汉藏语系中的十二生肖 Chinese calendar animals in *Shanhaijing* and in Sino-Tibetan languages. In David Bradley and Pei Likun (eds.), 《山海经》世界地理与中国远古文明 *Shanhaijing World Geography and Ancient Chinese Civilisation*. Beijing: Foreign Languages Press, p. 7-17, 93-105.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

- Bradley, David (2022) Phylogeny of Tibeto-Burman from plants and animals. In Mark Post, Stephen Morey and Toni Huber (eds.), *Eastern Himalayan Prehistory*. Leiden: Brill, p. 173-210.
- Brunson, Katherine et al. 2020. Zooarchaeology, ancient mtDNA, and radiocarbon dating provide evidence for the emergence of domestic cattle and caprines in the Tao River valley of Gansu Province, northwest China. *Journal of Archaeological Science: Reports* 31: 102262.
- Chen Tingting et al. (2020) Human responses to climate change in the late prehistoric western loess plateau, northwest China. *Radiocarbon* 62(5): 1193-1207.
- Cheung, Christina et al. (2019) Stable isotope and dental caries data reveal abrupt changes in subsistence economy in ancient China in response to global climate change. *PLOSOne* 14(7): e0218943.
- Fuller, Dorian et al. (2007) Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity* 81: 316-331.
- Jacques, Guillaume et al. (2021) Yak domestication: A review of linguistic, archaeological and genetic evidence. *Ethnobiology Letters* 12(1): 103-114.
- Liu Li & Chen Xingchan (2012) *The Archaeology of China from the Late Paleolithic to the Early Bronze Age*. Cambridge: Cambridge University Press.
- Xue Yining et al. (2022) Post-Neolithic broadening of agriculture in Yunnan, China: Archaeobotanical evidence from Haimenkou. *Archaeological Research in Asia* 30: 100364.
- Zeng Xingquan et al. (2015) The draft genome of Tibetan hulless barley reveals adaptation patterns to the high stressful Tibetan plateau (sic). *Proceedings of the National Academy of Sciences* 112(4): 1095-1100.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Climate change reflected in early Sino-Tibetan borrowings for crops and animals Bingcong Deng

The Holocene Climatic Optimum (HCO) occurred in northeastern China around 9500-5000 BP, marked by increased precipitation and temperature (Jia et al. 2016, Liu et al. 2022). Previous studies suggest that the period from 7000 to 5000 BP was characterized by a favorable climate in the Yellow River region (Liu et al. 2022), until the temperature and humidity dropped around 4000 BP (Sun et al. 2019). During this time, there was a significant increase in the spread of rice in northeastern China (d'Alpoim et al. 2015). In the West Liao River basin, the Bronze Age was characterized by a transition of human subsistence strategies as a response to climate change, with an increased reliance on animal husbandry in comparison to millet cultivation (Jia et al. 2016).

This paper aims to investigate the lexical borrowings of crops and animals in northeastern China, which could reflect the climate events linguistically. Emphasizing on the loanwords in northeastern China, two language phyla will be the focus of this study, namely Sino-Tibetan and Transeurasian. Rice cultivation, which was spread during the peak of HCO in northern China, may have led to borrowing of vocabulary related to rice farming from Sino-Tibetan to Transeurasian languages. Similarly, the increased reliance on animal husbandry in the West Liao River Basin could lead to borrowings of animal-related vocabulary from Transeurasian to Sino-Tibetan languages. Based on these premises, the research questions of this paper are: (1) What is the impact of climate change on crops and animals in northeastern China? How is that reflected in prehistoric lexical borrowings? (2) Can climate be seen as an impact of the transmission of the words for crops and animals?

This paper maps the approximate climate situations on the contact settings between Sino-Tibetan and Transeurasian in time and space, in reference to the loanwords to specific contact settings based on a loanword database compiled by the current author. A separate database for loanwords of crops and animals between Sino-Tibetan and language families in the south (e.g., Austronesian, Austroasiatic, and Tai-Kadai) was also collected, for the purpose of comparing the quantity and quality of borrowings that happened in the northeast. Data on archaeological sites and climatic information were collected from previous research.

The preliminary results suggest that (1) climate change correlates with the spread of certain crops and animals, further coinciding with the borrowing date of related lexical items. For instance, the introduction of wheat and barley from Central Asia is mirrored by the lexical borrowings referring to these crops detected in Old Chinese, Tungusic, Japonic and Korenic. This suggests that climate is likely to have played an important role in agricultural lexical borrowings between the two phyla. (2) The lexical borrowings between Sino-Tibetan and languages with a southern origin are larger in size in comparison to loans detected in the north (i.e., with Transeurasian). This difference might also be explained by the climate different between the two geographical regions. This research sheds light on the human response to climate change from a linguistic perspective. By investigating prehistoric lexical borrowings, it shows that climate events are one of the contributing factors to language contact and lexicon change.

References

- d'Alpoim Guedes, J., Jin, G., & Bocinsky, R. K. (2015). The Impact of Climate on the Spread of Rice to Northeastern China: A New Look at the Data from Shandong Province. *PLOS ONE*, 10(6), e0130430. <https://doi.org/10.1371/journal.pone.0130430>
- Jia, X., Sun, Y., Wang, L., Sun, W., Zhao, Z., Lee, H. F., Huang, W., Wu, S., & Lu, H. (2016). The transition of human subsistence strategies in relation to climate change during the Bronze Age in the West Liao River Basin, Northeast China. *The Holocene*, 26(5), 781–789. <https://doi.org/10.1177/0959683615618262>
- Liu, L., Chen, J., Wang, J., Zhao, Y., & Chen, X. (2022). Archaeological evidence for initial migration of Neolithic Proto Sino-Tibetan speakers from Yellow River valley to Tibetan Plateau. *Proceedings of the National Academy of Sciences*, 119(51), e2212006119. <https://doi.org/10.1073/pnas.2212006119>
- Sun, Q., Liu, Y., Wünnemann, B., Peng, Y., Jiang, X., Deng, L., Chen, J., Li, M., & Chen, Z. (2019). Climate as a factor for Neolithic cultural collapses approximately 4000 years BP in China. *Earth-Science Reviews*, 197, 102915. <https://doi.org/10.1016/j.earscirev.2019.102915>

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Austroasiatic dispersal: sea levels and estuarine environments in late Neolithic Mainland SEAsia.

Paul Sidwell

The paper discusses a radical reinterpretation of Austroasiatic (AA) prehistory in the light of sea level changes in Mainland Southeast Asian during the late Neolithic revolution there (circa 4kyBP).

How and when the Austroasiatic language phylum dispersed has been a contentious and difficult problem for a century. Nonetheless, in the past decade a consensus has begun to emerge based on a synthesis of linguistics, archaeology, and genetics called the “two layer hypothesis”. The model holds that AA emerged in northern Indo-China from the fusion of indigenous neolithic forager-farmers with East Asian cereal farmers attracted to the delta environments that facilitate intensive rice cropping.

Historically scholars have proposed the AA homeland in diverse locations (Indo-China, Gangetic India, Eastern India, Central China, Southwest China, etc.), all conceptualizing the dispersal as a problem of determining which overland or down-river routes were taken. Recent proposals (Sidwell 2022, 2020, Rau & Sidwell 2019, etc.) have argued that early AA speakers dispersed out of Northern Vietnam and around the Indo-Chinese coast and beyond to India by coastal navigation.

In this context, we need to consider how conditions differed from the present day. We know that Holocene sea levels peaked at around 2m higher than present some 7kyBP, gradually dropped by 3 metres, and rose again to almost the same peak from 4ky to 3.5kyBP. Many present day delta environments that are intensively cultivated for rice were very different: coastlines were further inland and low islands, coastal marshes and mangroves existed in places where paddy fields dominate today.

It is proposed that early Austroasiatic speakers ventured to seek new favourable estuarine environments rich in opportunities for hunting, fishing, vegeculture and cereal production. However, areas available for paddy farming were much more limited than today and this motivated growing populations to migrate ever further, eventually settling in the Malay peninsula, the Nicobar islands, and the Mahanadi River Delta in Eastern India. As sea levels declined larger delta areas formed, facilitating the rise of more organised societies such as the Davaravati Mon and pre-Angkorian Khmer states. In some areas the attractive coastal areas were overtaken by newer migrants, such that the Aslians in Malaysia and Mundas in India moved inland to rely more on shifting cultivation. The Nicobarese largely abandoned cereal farming in their adopted island home, assimilating culturally to some extent with Austronesians.

While today the greater diversity of AA speakers appears to be reflected in upland and shifting cultivators, this is a reflection of later diversification of those who moved inland. Environmental, cultural and societal change along the coasts and near inland favoured state formation and linguistic assimilation as sea levels fell and stabilised around present values.

References

- Sidwell, Paul. 2022. Austroasiatic dispersal: the AA "water-world" extended: Were the Proto-Austroasiatics coastal migrants? In Mark Alves & Paul Sidwell (eds.) *Papers from the 30th meeting of the Southeast Asian Linguistics Society (2021)*. Honolulu: University of Hawaii Press. Pp. 57-72.
- Sidwell, Paul. 2020. Austroasiatic Dispersal: The Indo-China Maritime Hypothesis. *Indian Linguistics* 81.12:15-32.
- Rau, Felix, and Paul Sidwell. 2019. The Munda Maritime Hypothesis. *Journal of the Southeast Asian Linguistics Society* 12.2:35-57.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Languages, ecology and climate change: Worldwide perspectives and the test-case of the Andes

Paul Heggarty

This workshop raises a series of open questions on climate change and language(s). Here I explore illustrative cases from around the world that offer valuable perspectives on several of these big-picture questions.

Research into the world linguistic panorama, and on how it came to be, has long looked for correlations with ecology, articulated especially through the subsistence strategies practised by different speaker populations. Most far-reaching is the farming/language dispersals hypothesis, invoked to account for how just a few language families expanded so spectacularly — and drove other language lineages, and overall human linguistic diversity, into decline. In effect, this hypothesis looks to climate change as an ultimate driver, for it was only once post-glacial warming took hold that agriculture emerged at all, repeatedly and independently.

The correlation is not so straightforward or immediate, however. Many language families hypothesised as spread by early farmers do not in fact seem to have started expanding until long after agriculture first began in the regions concerned, even by up to a few millennia. Alternative hypotheses stress later phases of intensification, secondary products, or specialisations (e.g., to pastoralism), which themselves may arise in response to ongoing climate changes.

So as this contrast already illustrates, the basic question is not whether links between languages and ecology exist, for they clearly do. Rather, it is about how far those links either point to environmental determinism, or reflect how human societies have responded to their ecological contexts and challenges, to mitigate and even take advantage of them. The contrasting fates of human language lineages through prehistory may in part record failed or successful responses to ecology. The parallels with the debate on the Anthropocene, and when it began, are striking. For our language diversity, too, has over time been transformed (and increasingly destroyed) by our own human impacts. Today's globalization marks an acceleration, but of a linguistic transformation that began many millennia ago.

Language-ecology relationships can differ greatly in causation and scale, however. Some language expansions are hypothesised as driven by one-off, extreme 'punctuation' events, such as the White River Ash volcanic eruptions pushing Athabaskan speakers southwards (Workman 1974), or a possible role for a 'Black Sea deluge' in spreading early Indo-European (see Nichols 2007). Even short-term events, if extreme enough, can have long-term linguistic consequences. Sometimes, humans also induce their own ecological collapses — although of the five cases explored by Diamond (2005), only one led to language extinction (of Greenlandic Norse). It is a different, much broader question how far linguistic fates have been shaped by full-blown climate-change, more gradually over far longer time-scales.

Many of these questions are ideally illustrated in one particular part of the world, where topography and the Tropics conspire to fashion a natural laboratory of ecological extremes and diversity, in immediate proximity. Out of the rainforest of Amazonia, the dry Andes rise rapidly to host the highest farmable lands on Earth, before dropping swiftly away to a coastal desert, but fringed by the superabundant waters of the Pacific. In this microcosm of extreme and fragile ecologies, a pristine civilisation arose, perhaps even before farming, and followed a tumultuous trajectory through both sudden and longer-term climate perturbances. This makes for an ideal test-case in how far language distributions may have been shaped by climate changes, or largely resistant to them, where their speakers ingeniously adapted to attenuate and harness the ecological challenges.

Generally, language distributions align with the stark differences between Amazonia, the Andes, and the Pacific Coast. The first complex societies, on the coast, did not spread their languages beyond the ecological limitations they faced. Major language expansions came only once complex societies in the highlands so transformed the Andean landscape, by terracing and irrigation, as to raise carrying capacity, expand demographically, and take their languages with them. The very name of the major surviving language family of the Americas, Quechua, originally denoted an ecological zone, the *q̄īc̄wa* mid-elevations ideal for cultivating what became the primary staple, maize. The grasslands of the higher puna zone, meanwhile, suited pastoralism better, once camelids like the llama had been domesticated, as well as tuber crops like the potato. These underpinned the Tiwanaku culture around Lake Titicaca, up until its collapse, widely attributed to climate change. Yet this did not efface their language, Puquina, although it did leave it vulnerable to the later power of the Incas, and their mastery of the Andean environment. The Incas even named languages, too, in 'ecological' terms (Itier 2015). They also resettled populations en masse far across their Empire, but often deliberately into regions ecologically similar to their homelands, and taking their Quechua and Aymara languages with them.

References

Diamond, J.M. 2005. *Collapse: How Societies Choose to Fail or Succeed*. London: Allen Lane.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

- Itier, C. 2015. «Quechua» y el sistema inca de denominación de las lenguas. *Mélanges de la Casa de Velázquez. Nouvelle série* (45–1): p.37–56.
- Nichols, J. 2007. Language dispersal from the Black Sea region. In: V. Yanko-Hombach et al. (eds) *The Black Sea Flood Question: Changes in Coastline, Climate and Human Settlement*, p.775–96. New York: Springer. http://doi.org/10.1007/978-1-4020-5302-3_32
- Workman, W.B. 1974. The cultural significance of a volcanic ash which fell in the Upper Yukon Basin about 1,400 years ago. In: S. Raymond & P. Schledermann (eds) *International Conference on the Prehistory and Paleoecology of the Western North American Arctic and Subarctic*, p.239–61. Calgary, Alberta: University of Calgary Press.

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

(Im)mobility, climate, and language: Towards a geoanthropology of the Balkans

Brian Joseph

The ways in which humans interact with the physical environment of the geography within which they live, i.e. their geoanthropology, have ramifications for their language. The peoples of the Balkans offer various case studies showing such geoanthropological effects, focusing on movement, or lack thereof, across different environments, for different reasons, and with different results. Thus, by way of illustrating the range of these geoanthropological interactions with language, I survey here some of these cases, drawing in part on Friedman & Joseph 2023. In particular, I discuss the linguistic correlates of a nomadic versus a sedentary lifestyle for Roma populations in the Balkans, as well as the effects of the “transhumance” of both the Balkan Romance speakers of Aromanian in the central Balkans (especially Albania, Greece, and the Republic of North Macedonia) and the Sarakatsani speakers of Greek who live in northern Greece, Bulgaria (where they are known as Karakačani), and the Republic of North Macedonia (where they are known as Sarakačani), by which whole villages relocate at different altitudes for summer and for winter.

I then draw parallels with similar situations in other parts of the world, looking in particular at the linguistic consequences of nomadic versus settled Bedouin Arabic lifestyles in the Middle East (as discussed in Cadora 1992) and the so-called “vertical” bilingualism in the Caucasus (Nichols 2015), by which people in higher altitude villages know the languages of those lower down the mountain, but those in the lowlands do not bother to learn highland languages.

Based on these case studies, I argue first that the observed effects have largely to do with differential patterns of contact with speakers of other languages brought on by the differential interaction these groups have through their shared geography. Ultimately, therefore, I claim further that there is no specific geoanthropological effect as a *primary* mechanism of language change, but rather that any such effects are secondary, deriving from well-known and well-understood mechanisms of contact-induced change (as outlined in Weinreich 1968, Thomason & Kaufman 1988, Winford 2003, and Matras 2009, among other sources).

In this way, the linguistic effects of geoanthropology are rather like what has been argued for other aspects of the historical development of languages, especially grammaticalisation (Campbell 2001) and exaptation (Joseph 2016), i.e. they are real, yes, but are derivative and thus arguably epiphenomenal, in that they reduce to already well-established patterns of linguistic change.

References

- Cadora, Frederic. 1992. *Bedouin, Village and Urban Arabic*. Leiden: Brill.
- Campbell, Lyle. 2001. What's wrong with grammaticalization? *Language Sciences* 23: 113-161.
- Friedman, Victor A. & Brian D. Joseph. 2023. *The Balkan languages*. Cambridge University Press.
- Joseph, Brian D. 2016. Being exacting about exapting: An exaptation omnibus. *Exaptation and Language Change*, ed. by Muriel Norde & Freek van Velde, 37-55. Amsterdam: John Benjamins.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Nichols, Johanna. 2015. Types of spread zones: Open and closed, horizontal and vertical. *Language Structure and Environment. Social, cultural, and natural factors*, ed. by Randy La Polla & Rik De Busser, 261-286. Amsterdam: John Benjamins.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Weinreich, Uriel. 1968. *Languages in contact: Findings and problems*. The Hague: Mouton.
- Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell

Workshop proposal for the 26th International Conference on Historical Linguistics, University of Heidelberg, 4 to 8 September 2023, organized by Prof. Dr. Jadranka Gvozdanović

Risk, resilience and the ecology of farming/language dispersals

Mark Hudson

The farming/language dispersal hypothesis is ultimately about demography. As farmers have more babies, their population increases and they expand at the expense of hunter-gatherers who have lower fertility. This process is known as the Neolithic Demographic Transition and occurred despite the fact that farming also led to higher mortality through new disease vectors. The basic pattern of farming expansion from centres of domestication is now well-understood. Until recently, archaeological studies of this expansion were based primarily on archaeobotanical and zooarchaeological data relating to the distribution of plants and animals. However, new approaches using isotopic and biomolecular archaeology are now enabling us to study questions of ecological ‘adaptation’ in farming dispersals in more detail. Three aspects are relevant here: the ways in which farmers adapted their crops and domesticated animals to different environments as they expanded into novel territories; the extent to which farmers made use of wild resources such as nuts and fish; and their responses to environmental change over both the short- and long-term. This perspective acknowledges that, while farming was economically more productive than hunter-gathering, it was also associated with high risks. In fact, the most productive peasant economies (such as Late Imperial China) were often associated with the highest level of risk: when something went wrong, it had very serious impacts on the livelihoods of huge numbers of people. Yet another recent approach in Neolithic studies is a greater focus on traces of violence resulting from new methods aimed at identifying cranial trauma. This work has shown that warfare and inter-personal violence were common in Neolithic societies, raising further questions about risk and resilience in early agriculture.

The first part of this presentation will summarise recent research on the cultural and environmental adaptations of early farmers, using examples from Europe, Japan and Island Southeast Asia. The discussion considers how such adaptations worked to enhance risk buffering and resilience. In the process of settling in to a territory, language must have been a key element of social learning, yet new evidence that has become available over the past decade or so shows that while Neolithic farmers expanded in a dynamic fashion, their lifeways were frequently subject to high risk and low resilience. Greater globalisation of food crops and increased exchange and commercialisation of foods were associated with more resilient agropastoral systems in the Bronze Age. This paper will explore the implications of these findings for the farming/language dispersal hypothesis, analysing the spread of Austronesian, Indo-European and Japonic as case studies.

WORKSHOP 2
MACRO-LEVEL SOCIAL MOTIVATIONS FOR LANGUAGE CHANGE:
CONTACT, MIGRATION, AND GLOBALIZATION

Bridget Drinka (bridget.drinka@utsa.edu)
 Terttu Nevalainen (terttu.nevalainen@helsinki.fi)
 Gijsbert Rutten (g.j.rutten@hum.leidenuniv.nl)

In her 1989 article on the role of socio-political forces as motivators of linguistic change, Susan Gal noted that the examination of speakers' micro-level responses to "macrohistorical processes" could provide new insights into the operation of contact as a motivator of change (Gal 1989: 357). In the years since the publication of this work, historical linguists, sociolinguists, and socio-historical linguists have grown increasingly aware of the interface between macro-historical processes and micro-level responses, as witnessed by paths of inquiry such as the following:

- The recognition of the role of ecology in establishing the trajectory of early varieties of African American English (Mufwene's 2001, 2008)
- The identification of population size as a determining factor in the linguistic outcome of contact and the level of complexity of contact varieties (Trudgill 2011), with smaller populations maintaining more complexity (Sinnemäki 2020) but also at times showing largescale areal distributions of complexity (Tallman and Epps 2018)
- The investigation of the role of koineization (Tuten 2003, 2021) and of socio-demographic factors (Sessarego 2019, 2021) under conditions of colonization and contact
- The development of new tools for the quantitative and qualitative analysis of the role of the individual in large-scale language change (Petré and Van de Velde 2018) and the mapping of large-scale and genealogical and geographical trends across time and space (Nichols 2016, 2020; Bickel 2020)
- The analysis of the interactive role of migration and urbanization in Africa and Europe (Mesthrie 2022; Kerswill & Wiese 2022; Wiese 2022; Mufwene 2022)

In this workshop, we propose to bring together scholars whose work focuses on macro-level motivations for linguistic change to explore how socio-political forces—invasion and migration, religious conversion and exclusion, colonization and globalization—have brought populations into contact, and what the micro-level effect on the languages of these speakers has been.

We regard this topic as critical at this moment in history, especially in light of several noteworthy trends:

- Approximately 4% of the world's population are global migrants: in 2020, there were about 281 million migrants in the world. [Migration Policy Institute]; in 2022, those fleeing conflict, violence, and other threats numbered more than 100 million (UNHCR, The UN Refugee Agency). Language contact is a constant among migrating populations.

- In 1945, about one third of the world’s population (approximately 750 million people) lived under colonial rule (United Nations). While this number has diminished greatly in recent years, linguistic effects of colonial rule persist.
- Closely tied to colonization is globalization, defined by Vignouroux and Mufwene (2008: 4) as “the worldwide network of economic interconnectedness and interdependencies.” English and other European languages continue to exert influence in the realm of commerce, academics, and popular culture.

Such macrohistorical pressures continue to leave their mark on the languages of the world today, and on the linguistic choices that each individual speaker makes.

What we hope to accomplish in this workshop is an in-depth examination of the mechanisms through which these and other macro-level processes have influenced the language of speakers.

In order to achieve this goal, we invite submissions focusing on the following research questions or other related issues:

- To what extent are macro-level motivations responsible for the creation of linguistic areas?
- What new methodologies can be employed to map the effects of past macro-level influences? What kinds of remnants of past influence persist, and how can we analyze and interpret these most effectively?
- Do some demographic features turn out to be more influential in contact situations than others? Are some of these features intersectional in their influence?
- What role does contact play in determining the level of complexity in larger or smaller speech communities?
- Is koineization to be found in languages around the world, or only in those which have experienced particular macrohistorical pressures?
- To what extent do changing social hierarchies and political and religious ideologies impact patterns of change?
- What role does prestige play in setting up superstratal influence and roofing effects? How do such factors influence the actual language of speakers? That is, to what extent do these factors illustrate micro-level responses to macro-level processes?
- Are some traditional examples of language change better explained as having been motivated by macrohistorical processes or, more generally, by contact?

REFERENCES

- Arkadiev, Peter, and Francesco Gardani (eds.) 2020. *The complexities of morphology*. Oxford: Oxford University Press
- Baechler, Raffaella and Guido Seiler (eds.). 2016. *Complexity, isolation, and variation*. Berlin / Boston: De Gruyter.
- Bickel, Balthasar. 2020. Large and ancient linguistic areas. In Crevels and Muysken (eds.). 78–101.
- Chappell, Whitney and Bridget Drinka (eds.). 2021. *Spanish socio-historical linguistics: Isolation and contact*. Advances in Historical Sociolinguistics 12. Amsterdam/Philadelphia: Benjamins.

- Crevels, Mily and Pieter Muysken (eds.) 2020. *Language dispersal, diversification, and contact: A global perspective*. Oxford: Oxford University Press.
- Gal, Susan. 1989. Language and political economy. *Annual Review of Anthropology* 18: 345-67.
- Kerswill, Paul, and Heike Wiese (eds.). 2022. *Urban contact dialects and language change: Insights from the Global North and South*. New York: Routledge.
- Mesthrie, Rajend. 2022. Migrants and urban contact sociolinguistics in Africa and Europe. In: Kerswill and Wiese (eds.). 337-343.
- Migration Policy Institute. (<https://www.migrationpolicy.org/>)
- Mufwene, Salikoko. 2001. *The ecology of language evolution*. Cambridge Approaches to Language Contact. Cambridge: Cambridge University Press.
- Mufwene, Salikoko. 2008. *Language evolution: Contact, competition, and change*. London: Continuum.
- Mufwene, Salikoko. 2022. Multilingualism and super-diversity: Some historical and contrastive perspectives. In: Mufwene and Escobar (eds.) 144-171.
- Mufwene, Salikoko, and Anna María Escobar (eds.). 2022. *The Cambridge handbook of language contact. Vol. 2: Multilingualism in population structure*. Cambridge: Cambridge University Press.
- Nichols, Johanna. 2016. Complex edges, transparent frontiers: Grammatical complexity and language spreads. In: Baechler and Seiler (eds.). 116-137.
- Nichols, Johanna. 2020. Dispersal patterns shape areal typology. In: Crevels and Muysken (eds.) 25-43.
- Petré, Peter, and Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94.4: 867-901.
- Sessarego, Sandro. 2019. *Language contact and the making of an Afro-hispanic vernacular: variation and change in the Colombian Chocó*. Cambridge: Cambridge University Press.
- Sessarego, Sandro. 2021. A socio-historical perspective on the origin and evolution of two Afro-Andean vernaculars. In: Chappell and Drinka (eds.). 163-184.
- Sinnemäki, Kaius. 2020. Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics* 6 (2): 1-39.
- Tallman, Adam, and Patience Epps. 2020. Morphological complexity, autonomy, and areality in western Amazonia. In: Arkadiev and Gardani (eds.). 230-263.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Tuten, Donald. 2003. *Koineization in medieval Spanish*. Berlin / New York: Mouton de Gruyter.
- Tuten, Donald. 2021. Complexification of the early modern Spanish address system: A role for koineization? In: Chappell and Drinka (eds.) 17-47.
- United Nations: The United Nations and Decolonization (www.un.org/dppa/decolonization/)
- UNHCR: The UN Refugee Agency (www.unhcr.org/)
- Vigouroux, Cécile, and Salikoko Mufwene (eds.). 2008. *Globalization and language vitality: Perspectives from Africa*. London / New York: Continuum.
- Wiese, Heike. 2022. Urban contact dialects. In: Mufwene and Escobar (eds.) 115-144.

Reconstructing prehistoric sociolinguistics from modern grammatical evidence

Johanna Nichols, UC Berkeley

Though a good deal is known to prehistorians about early centers of population growth and dispersal, and linguistics can identify some grammatical symptoms of sociolinguistic dominance and language shift, four problems remain unsolved. (1) Typology can now identify favored targets of selection in sociolinguistically asymmetrical language contact (e.g. canonical typology, Trudgill 2011), but these have not been applied to determining which language families descend from backwater refugees and which from expanding and sociolinguistically dominant ones. (2) Nor have they been applied to diachronic studies of head-marking, especially polysynthetic, languages with templatic morphosyntax. (3) Where past expansions can be identified, it is usually not known whether that involved spread into (near-)uninhabited land, dominance and absorption of a prior population, or sidelining a previous population with minimal substratal effects. (4) The effects of dense vs. sparse networks and short vs. long connections can now be modeled (Fagyal et al. 2010), but it is also known that very long travels in pre-Neolithic societies were routine (Graeber & Wengrow 2021:173); should ancient mobile groups (and selection in them) be modeled as nodes in sparse populations? as separate small populations (Bickel 2022)? as individuals in large, densely connected populations?

This paper uses case studies of four known or likely centers of expansion to propose answers by pushing back the temporal reach of sociolinguistic reconstruction. Additional theoretical considerations are the patch-and-pump model of first and early settlements (Author in press); staging areas and cost-path modeling (Anderson & Gillam 2001, Anderson et al. 2013) identifying centers and trajectories of spread; relational complexity (Author in press) to identify targets of selection in polysynthetic languages; self-similarity at different levels as an effect of selection (Nichols 2018); improved typological descriptions of features subject to selective pressure (e.g. Authors 2022); and isolation-by-distance modeling to identify centers and peripheries of spreads (cf. Grünthal et al. 2022).

Six variables are traced here across four case studies: (1) The early Columbia Plateau, for which the set of "Penutian" families is shown to be a likely early frontier population preserving a Eurasian-like typology as subsequent immigrants brought or developed a very different typology. (2) The later Columbia Plateau, where postglacial desiccation triggered the various "Penutian" spreads south and west, argued here to have begun in spreads along existing networks with minimal contact effects; (3) the Lower Mississippi Valley, a long-standing staging area (Kaufman 2014). (4) The Altai region (upper Irtysh and Yenisei, Minusinsk Basin, northern Kazakhstan and Mongolia), from which Pre-Uralic, Pre-Turkic, and Pre-Mongolic may have dispersed; the very self-similar Ural-Altaic typology is barely emergent in reconstructed ProtoUralic (c. 4500 BP) and highly developed in Proto-Turkic (c. 2000 BP) and Proto-Mongolic (c. 1000 BP). This evolution points to long-term selection without sociolinguistic dominance. Variables: Harmonic pronoun consonantism; self-similar morphology and syntax (morpheme/word order, head/dependent marking); fixed base lexical valence; high/low causativization (base intransitivity); consistency in finiteness across different clause types; configurational/templatic. All are relatively stable in families, and high frequencies of either polar value (e.g. head-final vs. head-initial) are favored in selection.

Keywords

Centers and peripheries, language spreads, linguistic networks, linguistic selection, sociolinguistic typology.

References

- Anderson, David G., Thaddeus G. Bissett, and Stephen J. Yerka. 2013. The Late-Pleistocene human settlement of interior North America: The role of physiography and sea-level change. *Paleoamerican Odyssey*, ed. K. E. Graf, C.V. Ketron, and M. R. Waters, 183–203. College Station, TX: Center for the Study of the First Americans, Texas A & M University.
- Graeber, David G. and J. Christopher Gillam. 2001. Paleoindian interaction and mating networks: Reply to Moore and Moseley. *American Antiquity* 66:3.530–535.
- Bickel, Balthasar. 2022. Large and ancient linguistic areas. *Language dispersal, diversification, and contact*, ed. Mily Crevels and Pieter Muysken, 78–99. Oxford: Oxford University Press.
- Fagyal, Zsuzsanna, Samarth Swarup, Anna Maria Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and peripheries: Network role in language change. *Lingua* 120:8.2061–2079.
- Graeber, David and David Wengrow. 2021. *The dawn of everything: A new history of humanity*. New York: Farrar, Straus, and Giroux.
- Grünthal, Riho, Volker Heyd, Sampsa Holopainen, Juha Janhunen, Olesya Khanina, Matti Miestamo, Johanna Nichols, Janne Saarikivi, and Kaius Sinnemäki. 2022. Drastic demographic events triggered the Uralic spread. *Diachronica* 39:4.490–524.
- Kaufman, David V. 2014. The Lower Mississippi Valley as a language area. PhD dissertation, University of Kansas.
- Nichols, Johanna. 2018. Non-linguistic conditions for causativization as a linguistic attractor. *Frontiers in Psychology* 8.2356.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.

Macro-changes at the dawn of history: The Slavic Expansion

Henning Andersen, UCLA

The extraordinary expansion of Slavic-speaking territories during the early centuries of our era (300s–700s) has long defied explanation. Slavic scholars have tried to link it with such macro-events as the Little Ice Age in the 500s or the Justinian plague. But these events are chronologically off and would at best explain population displacements and not the attested, vast territorial spread. Besides, there is linguistic evidence of distinct stages in the Expansion and of the role of language contact from its earliest stages (Andersen, To appear); this evidence leaves no doubt that the Expansion resulted from gradual, substantial population growth.

A rational account for this can refer to a macro-event of a different character, the gradual integration of Slavic-speaking populations into early medieval iron-age culture: The gradual adoption of an iron-age tool-kit and the replacement of slash-and-burn agriculture with crop rotation would naturally lead to a better return on hours worked, result in improved social health, and produce growing populations across the Slavic-speaking areas.

In connection with the adoption of iron-age tools it is significant that there is no single Proto-Slavic word for 'blacksmith' but instead a handful of native synonymous neologisms with a geographical distribution that reflects distinct population flows in the Expansion. Furthermore, hundreds of Slavic placenames reflect chronological stages in this development. The earliest stage (i) may be the introduction of industrial iron smelting and manufacturing, archaeologically evidenced in Poland in the 100s–300s. Perhaps simultaneously with this, (ii) iron-making spread across the land as a part-time activity of farmers, likewise part of the archeological record. A later stage (iii) was the gradual specialization of successful local blacksmiths who each supported farming communities in a small area. Stages (i) and (iii) are rather spectacularly reflected in Slavic placenames with geographical distributions that appear independent of that of the appellatives. Eventually, of course, (iv) every village would have its blacksmith.

The Expansion redistributed early dialect differences (Andersen 2020) and formed the background for the development of new isogloss systems across the Slavic-speaking territories.

References

- Andersen, Henning. 2020. On the formation of the Common Slavic *koiné*. Boček et al., 11–42. • To appear. The Slavic Expansion: Streams, springs, and wells. *Scando-Slavica* 2023. • Boček, Vit & Nicolas Jansens & Tomáš Klír (eds.), *New perspectives on the early Slavs and the rise of Slavic: Contact and migration*. Heidelberg. • Lindstedt, Jouko & Elina Salmela. 2020. Migrations and language shifts as components of the Slavic spread. Boček et al. • Noerbach [Nørbach], Lars Christian (ed.). 2002. *Prehistoric and medieval direct iron smelting in Scandinavia and Europe: aspects of technology and society*. Aarhus. • Pleinder, Radomir. 2000. *Iron in archaeology: the European bloomery smelters*. Prague. • Rusek, Jerzy. 1996. *Dzieje nazw zawodów w językach słowiańskich*. Warsaw. • Rybakov, B. A. 1948. *Remeslo drevnej Rusi*. Moscow. • Sedov, V. V. 1994. *Slavjane v drevnosti*. Moscow. • Trubačev, O. N. 1966. *Remeslennaja terminologija v slavjanskih jazykach (etimologija i opyt gruppovoj rekonstrukcii)*. Moscow. • Wells, Peter S. & Katherina Rebay-Salisbury & Colin Haselgrove (eds.). 2018. *The Oxford Handbook of the Iron Age*. Oxford.

An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German

Julie Nijs, Freek Van de Velde, and Huybert Cuyckens

Larger languages in high-contact communities are morphologically less complex and rely more on lexical strategies and word order than smaller languages in close-knit communities (Lupyan & Dale 2010). This study focuses on the West-Germanic languages Dutch, English and German, which are known to have been exposed to different degrees of internal (dialect) contact and external contact (O’Neil 1978; Weerman 2006). Specifically, English has been more exposed to contact than Dutch, which in turn has been more exposed than German. To assess whether degree of contact correlates with morphological as well as syntactic complexity in these languages, we measure morphological and syntactic complexity by the mathematical notion of ‘Kolmogorov complexity’ (Kolmogorov 1968), an information-theoretic approach which defines a string’s complexity in relation to its information content.

The Dutch, English and German texts making up our dataset were taken from the Book of Genesis and the Gospel of Matthew, as they occur in the multilingual parallel EDGeS Diachronic Bible Corpus (Bouma, Coussé, Dijkstra & van der Sijs 2020). A total of 47 texts from different time periods between the 14th and 19th century have been analyzed: 21 for Dutch, 18 for English and 8 for German.

Following Juola (2008) and Ehret (2017), morphological complexity can be calculated after randomly deleting 10% of a text’s orthographic transcribed characters and compressing the file with gzip. The random deletion leads to morphological distortion, in that the number of unique tokens increases, which makes compressibility worse. Texts characterized by a high surface token diversity (as a result of affixal complexity, root-internal alternation or other morphological operations) will be comparatively less affected by distortion, because they already contain a higher amount of unique tokens before distortion. In terms of Kolmogorov complexity, these are the texts that are morphologically more complex. Syntactic complexity can be calculated in the same way, but instead of characters, words are deleted. This leads to a distortion of the word order rules, a higher number of unique lexical n-grams and thus worse compressibility. Texts with strict word order have more structural surface redundancies and will therefore be more affected by distortion, while languages with free word order will be less affected due to their lower number of redundancies. This means that in terms of Kolmogorov complexity rigid word order is considered as more complex.

The morphological complexity ratio is calculated as $\frac{mc}{c}$, where mc is the compressed file size in bytes after morphological distortion, and c is the compressed file size in bytes before distortion. The syntactic complexity ratio or the word order rigidity ratio is calculated as $\frac{sc}{c}$, where sc is the compressed file size in bytes after syntactic distortion, and c is the compressed file size in bytes before distortion. For each text the mean morphological and syntactic complexity was calculated over 1000 iterations, to take the aleatoric effect of the randomization into account.

We have found a significant interaction effect between year and language for the morphological complexity ratio. Morphological simplification happens faster in English compared to Dutch, as expected, but German seems to be more on the side of English, counter to what we expect. Syntactic complexity, then, shows the mirror image. We can thus observe a negative correlation between the morphological and syntactic complexity ratio (Figure 1). The three languages each take up their own space in the graph. Dutch is morphologically the most complex, but syntactically less complex; English is syntactically the most complex, but morphologically less complex; German lies in-between.

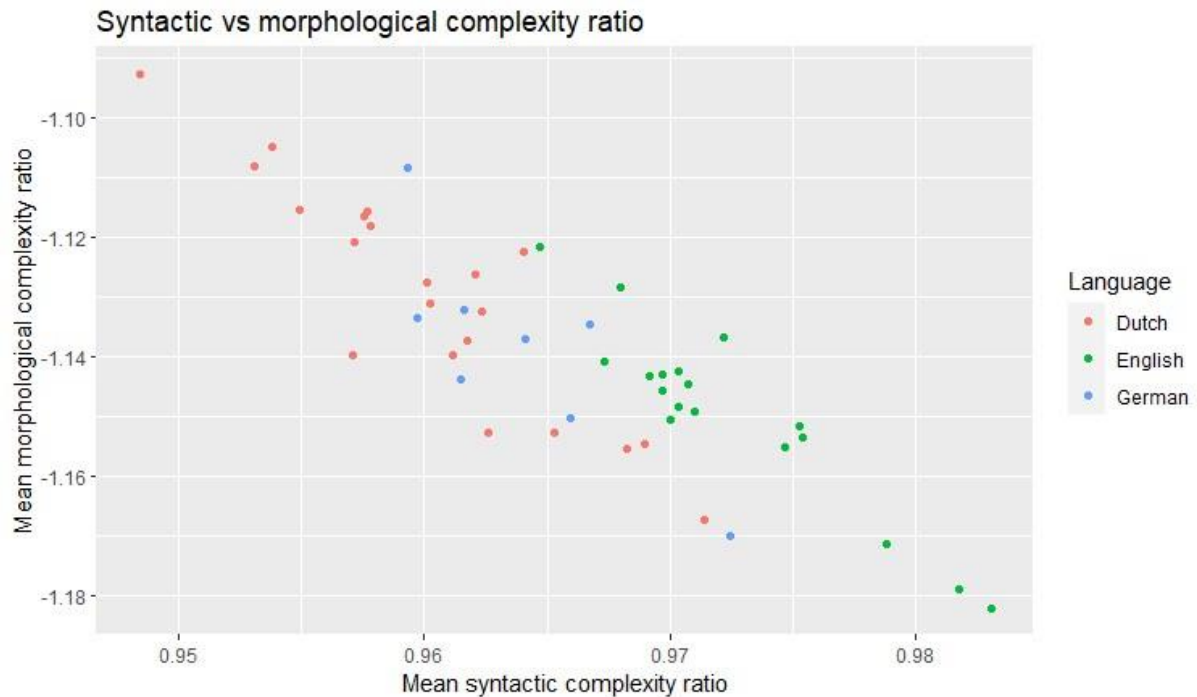


Figure 1: Syntactic vs morphological complexity ratio

References

- Bouma, G., E. Coussé, T. Dijkstra & N. van der Sijs. 2020. The EDGeS diachronic bible corpus. In *Proceedings of the 12th international conference on language resources and evaluation*, 5232-5239. Paris: ELDA.
- Ehret, K. 2017. *An information-theoretic approach to language complexity: variation in naturalistic corpora*. Freiburg: Albert-Ludwig-Universität Freiburg dissertation.
- Juola, P. 2008. Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language complexity: typology, contact, change*, 89-108. Amsterdam: John Benjamins.
- Kolmogorov, A. Ni. 1968. Three approaches to the quantitative definition of information. *International Journal of Computer Mathematics* 2(1-4). 157-168. doi:10.1080/00207166808803030.
- Lupyan, G. & R. Dale. 2010. Language structure is partly determined by social structure. *PLoS One* 5(1).
- O'Neil, W. 1978. The evolution of the Germanic inflectional systems: a study in the causes of language change. *Orbis* 27: 248-286.
- Weerman, F. 2006. It's the economy, stupid: Een vergelijkende blik op *men* en *man*. In: M. Hüning, et al. (eds.), *Nederlands tussen Duits en Engels*. SNL. 19-47.

Contact as a major Motivation for Linguistic Change in the History of Balkan Slavic

Andrey N. Sobolev Philipps University Marburg (originally Institute for Linguistic Studies, Russian Academy of Sciences)

This paper presents theoretical, methodological and practical results obtained in the last decades in the field of comparative-historical Balkan Slavic linguistics. Against the backdrop of the major theoretical issues of Balkan linguistics, e.g. principles of genetic, areal, social or contact determination or restriction in language evolution [Friedman, Joseph 2023], contact is viewed as the major motivation for linguistic change in the history of Balkan Slavic [Sobolev 2019], and is responsible for the creation of the Balkan linguistic area.

The genetic, areal-typological, anthropological and socio-political analysis of Slavic languages in the Balkans reveals divergent and convergent developments which can be interpreted against the background of comparative-historical theory, geolinguistic theory, language contact and Sprachbund theory, including the dialectology of convergent linguistic groups [Sobolev (ed.) 2021]. General mechanisms of genetic splitting and typological merging, borrowing and calquing, language shift, and language and ethnic separation and symbiosis interplayed to make this a truly unique area of Europe and Eurasia.

South Slavic entered the Balkan peninsula as at least two genetically differentiated subgroups (so-called West South Slavic and East South Slavic), but eventually became part of the Balkan linguistic landscape irrespective of this primary genetic subdivision. Due to profound multilingualism, the Balkan linguistic landscape can be viewed as an uninterrupted continuum of closely and distantly related dialects (languages), characterized by an array of isoglosses that run irrespective of “language borders.” Among the most prominent features are the following: identical or similar inventories of affricates as well as palatal consonants; the postpositive definite article; “case loss” and the analytic marking of grammatical relations on the noun; “infinitive loss”; the volitive future tense; the possessive perfect; grammaticalized evidentiality markers; and semantic patterns borrowed from Greek, Latin, and Turkish.

This extreme tendency towards borrowing is well-illustrated by (1):

(1) Golo Bordo dialect of Macedonian [Sobolev & Novik 2013]

| | |
|---------------------------|--------------|
| 'imat d'el'veno | na=d'eʃi |
| have distribute.PPP | PREP=rams.PL |
| '(They) distributed rams' | |

This example illustrates not only the direct material borrowing from Alb. dash [daj] ‘ram’, partially integrated into the morphology of Macedonian, cf. daj SG.INDEF, d'aʃof ~ d'aʃot SG.DEF ‘ram,’ but also the adoption of the Albanian apophonic plural marking a ~ e, that is Alb. dash ~ desh, which is completely alien to Slavic, alongside the affixation of the common Slavic plural marker -i. The inclusion of the preposition na to mark the direct object, following the Balkan Romance model, adds the final touch to this extraordinary and highly redundant amalgamation.

At the same time, some particularly Slavic features persist and appear to act as barriers to language integration: stress shift on proclitics (as in Bulgarian b'ez=žena ‘without a wife’); the category of peripheral case forms as opposed to structural cases; the category of animateness and personness; opposing “short” and “long” forms of adjectives with unclear intrasystemic functions; the absence of any categorial marker for definiteness on any member of the nominal group, that is, of an explicit marker for individualizing, generic, specific, or indefinite meaning;

the category of verbal aspect with the admittedly vague general meaning of terminativity, expressed by a root morpheme or a suffix.

Thus, we see abundant evidence for major structural innovation motivated by contact, leading to a deep qualitative reorganization of Balkan Slavic languages throughout their history. On the other hand, certain inherited characteristics persist which resist these changes and do not spread beyond Slavic to other languages of the Balkan peninsula.

References

- Friedman, Victor A.; Joseph, Brian. The Balkan languages. Cambridge: Cambridge University Press, 2023 (Manuscript).
- Kochev, Ivan (ed.). Balgarski dialekten atlas. Obobshtavasht tom. I–III. Fonetika. Aktsentologiya. Leksika. [Bulgarian dialect atlas. Generalizing volume. I–III. Phonetics. Accentology. Lexicon]. Sofija: Knigoizdatelska kashta „Trud“, 2001.
- Sobolev, Andrey N. Slavic dialects in the Balkans: unified and diverse, recipient and donor. In: Motoki, Nomachi and Danylenko, Andriy (Eds.). Slavic on the Language Map of Europe. Historical and Areal-Typological Dimensions. Berlin; New York: de Gruyter Mouton, 2019. P. 315–346.
- Sobolev, Andrey N. (Ed.). Southeastern European Languages and Cultures in Contact. Between Separation and Symbiosis. Berlin; New York: de Gruyter Mouton, 2021.
- Sobolev, Andrey N. & Aleksandr A. Novik. 2013. Golo Bordo (Gollobordë), Albania. Iz materialov balkanskoj ekspeditsii ran i SPbGU 2008–2010 gg. [Golo Bordo (Gollobordë) Albania. From the materials of the RAS and SPSU expedition of 2008–2010]. Munich: Otto Sagner Verlag, St. Petersburg: Nauka.

Ideology, language choice and language change

Jadranka Gvozdanović (University of Heidelberg)

The tradition of research on ideology and language assumes that ideology, related to ideas, beliefs and opinions, construes underlying patterns of meaning and the corresponding frames of interpretation. These have a bearing on different types of discourse (cf. e.g. Verschueren 2012). The research presented here adds to this by focusing on (a) ideological self-ascription, (b) choice of the intended addressees, and (c) language choice on three levels:

- the macro-level of the extended social group, usually with an intended ethnic, religious or political identity,
- the meso- level of ideological group differentiation within the frame provided by the macro-level, and
- the micro-level of the (Self or Other) ascription by an individual speaker.

The paper discusses these matters based on representative modern-era ideologically conditioned periods of change in Central Europe that systematically addressed ideological (initially religious, later national) issues through language. Firstly, the Reformation and the Catholic Counter-Reformation in the 16th and 17th centuries addressed the transregional macrolevel and promoted spoken languages, for which an adequate choice of language norm was required. This brought about language-ideological considerations and caused significant language shifts mediated by codification.¹ Secondly, the emergence of national ideologies in the 19th century accompanied by a search for the language variety with sufficient cultural weight to represent the nation (such as the language of the 16th century Kralice Bible for Bohemia, or the language of the Baroque poetry of Dubrovnik for Croatia, influenced by the Catholic Counter-Reformation); the accompanying codification brought about major language shifts across the national territories. The third period of change started with the loss of the ideology of supra-national standard languages (particularly Serbo-Croatian, in part paralleled by CzechoSlovak) in Central Europe, which overtly preceded the loss of the political ideology of supranational states. Superficially seen this was a process opposite to the former ones (i.e. linguistic ideology change preceded the corresponding political change of ideology), but in fact the loss of the linguistic supra-national ideology was a proxy for the loss of the political supra-national ideology, officially forbidden by the ruling communist regimes.

These major periods of change were triggered by increase vs. decrease of the macro-level ideological scope (in the latter case, dissolution of the overarching language norm) leading to change implemented by codification.

The contemporary period is mainly characterized by meso-level ideological differentiations systematically expressed either by preserving linguistic conservatism or adopting innovation (i.e., either by rejecting or adopting the forms or orthographies proposed by recent language reforms).

Mass media choose these alternatives to implicitly advertise their political-ideological adherences. This is another example of language ideology as a proxy for political ideology attested both in contemporary Czech (Bermel 2007) and Croatian (Gvozdanović 2010, PetiStanić 2013). It is neither destandardization nor demotization (as defined by Kristiansen & Coupland 2011), but ideologically driven symbolic dissection on the meso-level within the macro-frame of the standard language.

Based on Croatian and Czech examples, this paper traces effects of an increase or decrease of ideological scope on the macro level, showing how changes of political, religious and national ideologies are interrelated with language ideologies that condition the corresponding language changes.

¹ E.g., for Croatian cf. the discussion by Knežević (2007), Gabrić-Bagarić (2010).

References

- Bermel, Neil. 2007. *Linguistic Authority, Language Ideology and Metaphor. The Czech Orthography Wars*. Berlin: Mouton de Gruyter.
- Gabrić-Bagarić, Darija. 2010. Četiri ishodišta hrvatskoga standardnoga jezika. *Fluminensia* 22,1: 149-162.
- Gvozdanić, Jadranka. 2010. Jezik i kulturni identitet Hrvata. *Kroatologija* 1: 39-57.
- Knežević, Sanja. 2007. Nazivi hrvatskoga jezika u dopreporodnim gramatikama. *Croatica et Slavica Jadertina*, 41-69.
- Kristiansen, T. & Coupland, N. 2011 *Standard Languages and Language Standards in a Changing Europe*. Oslo: Novus Forlag.
- Peti-Stantić, Anita i Keith Langston. 2013. *Hrvatsko jezično pitanje danas: identiteti i ideologije*. Zagreb: Srednja Europa.
- Verschueren, Jeff. 2012. Ideology in language use. In: Verschueren, Jeff, *Pragmatic Guidelines for Empirical Research*, 7-20. CUP.

Language use in Alsace from 1914 to 1919. Private texts between official legislation and individual identity construction

Lena Sowada (University of Heidelberg)

This proposition aims to explore the interface between macro-historical processes, such as contact, linguistic prestige and language legislation, and micro-level responses in a GermanFrench corpus of private texts written during World War I by Alsatian soldiers and their families.

For Alsatian society, linguistic and cultural contact has been for centuries a constitutive element of their everyday life. Over the times, Alsatians were confronted several times with changing national languages, different linguistic prestige and varying language legislation. The main question that the paper seeks to explore is: How did macro-level processes such as official language legislation, language ideology and linguistic prestige influence the actual language use of the writers?

The texts, which have not yet been examined, are from two Alsatian families, the Jeandon family from Lapoutroie (*Schnierlach* in German) and the Braun family from Oberhaslach. In total, the corpus includes 162 German and 12 French postcards and letters as well as one French diary written by Auguste Jeandon who, like all Alsatians who had not fled their homeland, fought as a soldier on the side of the German Empire. The authors of the texts can roughly be considered as less-experienced writers as their everyday life before 1914 hardly required any writing practice.

In particular, the paper takes up the following questions raised in the description of the workshop.

The role of contact with regard to the complexity of speech communities is fundamental for this specific linguistic community. Depending on the place of residence, different language laws apply, speakers have different first languages that condition schooling, and religious confession influences language behaviour. Every factor depends on the specific circumstances of the locality and the legislation applied there: e.g. the number of German immigrants, the attitude of the local authority, the acceptance as francophone community (or not) and the self-representation of each writer. Moreover, the corpus shows not only contact between the normative standard varieties of German and French as well as between the Germanic and Romance dialects, but also contact of two different scripts in use in the two cultural spaces.

Social and political hierarchies as well as religious ideologies are crucial for the linguistic choices the authors make. Linguistic legislation in Alsace during World War I depends on the civil and military authorities who do not always follow the same lines. Catholic and protestant churches play an important role in the maintenance of French or in the support of the Germanization of the population. On a macrolevel there is a strong linguistic pressure for the Germanization, especially from the military authorities supported by Protestantism, but on a micro-level the writers show some resistance in the use of German because they maintain French and/or the dialects.

Linguistic prestige and language as a marker for identity construction are of essential importance in the negotiation of language use in this border region. However, the attribution of prestige to a particular language is not the same for all Alsatians, but depends on their specific context. Some Alsatian writers may respond explicitly to the changing political hierarchies expressing their political affiliation in the texts. The only use of French in a German-speaking context, such as keeping a French diary in the German army, could be seen as a political and ideological positioning.

The present proposition does not claim to be representative, but at least, it can contribute for this period to a broader view of the language use subject to the aforementioned constraints. The corpus shows exemplarily how Alsatian writers individually respond to the significant processes of the macro-level and which concrete linguistic forms result from this specific situation.

References

- Antoine, Gérald/Martin, Robert (edd.), *Histoire de la langue française 1880–1914*, Paris, CNRS, 1985.
- Antoine, Gérald/Martin, Robert (edd.), *Histoire de la langue française 1914–1945*, Paris, CNRS, 1995.
- Appel, René/Muysken, Pieter, *Language contact and bilingualism*, London et al., Arnold, 1987.
- Auer, Peter, From codeswitching via language mixing to fused lects. Toward a dynamic typology of bilingual speech, *The international journal of bilingualism* 3:4 (1999), 309–332.
- Auer, Peter/Eastman, Carol M., Code-switching, in: Jaspers, Jürgen/Östman, Jan-Ola/ Verschueren, Jef (edd.), *Society and language use*, Amsterdam/Philadelphia, John Benjamins, 2010, 84–112.
- Bothorel-Witz, Arlette/Philipp, Marthe/Spindler, Sylviane, *Atlas linguistique et ethnographique de l'Alsace*, vol. 2, Paris, Éd. du CNRS, 1984.
- Bullock, Barbara E./Toribio, Almeida. Jacqueline (edd.), *The Cambridge handbook of linguistic codeswitching*, Cambridge et al., Cambridge University Press, 2009.
- Elspaß, Stephan, «Everyday language» in emigrant letters and its implications for language historiography the German case, *Multilingua. Journal of cross-cultural and interlanguage communication* 26:2–3 (2007), 151–165.
- Ernst, Gerhard, Konvergenzen und Divergenzen im Schreiben ungeübter Schreiber (Französisch, Italienisch, Deutsch), in: Patzelt, Carolin/Prifti, Elton (edd.), *Diachrone Varietätenlinguistik. Theorie, Methoden Anwendungen*, Berlin et al., Peter Lang, 2020, 21–38.
- Gardner-Chloros, Penelope, Le code-switching à Strasbourg, in: Salmon, Gilbert-Lucien (ed.), *Le français en Alsace*, Paris/Genève, Champion/Slatkine, 1985, 51–60.
- Grosjean, François, *Parler plusieurs langues. Le monde des bilingues*, Paris, Albin Michel, 2015.
- Huck, Dominique, *Une histoire des langues de l'Alsace*, Strasbourg, La Nuée Bleue, 2015.
- Lévy, Paul, *De la Révolution française à 1918. Histoire linguistique d'Alsace et de Lorraine*, vol. 2, Paris, Les Belles lettres, 1929.
- Matras, Yaron, *Language contact*, Cambridge, University Press, 2009.
- Nevalainen, Terttu/Raumolin-Brunberg, Helena, *Historical sociolinguistics. Language change in Tudor and Stuart England*, London et al., Longman, 2003.
- Pahta, Päivi/Skaffari, Janne/Wright, Laura (edd.), *Multilingual practices in language history. English and beyond*, Boston/Berlin, De Gruyter Mouton, 2018.
- Winter-Froemel, Esme, Entlehnung in der Kommunikation und im Sprachwandel. Theorie und Analysen zum Französischen, Berlin et al., De Gruyter, 2011.

Political influence as a factor in morphosyntactic variation: demonstratives *este* and *aqueste* in medieval Aragonese

Andrés Enrique-Arias
(University of the Balearic Islands)

Keywords: Iberorromance; historical morphosyntax, demonstratives, medieval Aragonese, language contact

Medieval Ibero-Romance languages exhibited variation between short and long variants of the demonstratives (i.e. Spanish *este* and *aqueste*, respectively, both meaning ‘this’). Data collected from notarial documents (cf. Enrique-Arias 2018) shows that in most Ibero-Romance varieties (Galician-Portuguese, Leonese, Castilian, Navarrese) long forms such as *aqueste* were a small minority throughout the Middle Ages and disappeared completely by the 1600s. Catalan is a notable exception: in this language the long forms (i.e. *aquest* as opposed to short form *est*) were almost categorical from the earliest texts and have continued to exist to this day.

This research focuses on the peculiar situation of Aragonese, which experienced a spectacular increase in the frequency of *aqueste* type forms throughout the 14th century to become almost categorical at the beginning of the 15th century. Shortly after, however, the long forms declined rapidly and disappeared in the 16th century.

In this paper I argue that the abrupt changes involving the distribution of *aqueste* type forms in Aragonese legal documents are changes from above that reflect how writers adopted alternating scriptural models –first Catalan, and later Castilian– dictated by the successive power centers that dominated Aragon.

Throughout the Middle Ages the Crown of Aragon, which also included the Principality of Catalonia, was ruled by a Catalan-speaking dynasty and, for the most part, the Royal Chancellery issued its documents in Catalan; thus, long demonstrative forms, which are characteristic of Catalan, became part of the prestigious model adopted by the scribes. This situation changed dramatically after 1412, when the Crown of Aragon was taken over by the Castilian-speaking Trastámara rulers; a few decades later there was a dynastic union with Castile under the Catholic Monarchs (Isabella and Ferdinand) which further increased the political and cultural Castilian influence among the peninsular kingdoms. This political change is reflected in a sudden decrease in the use of the Aragonese long forms and the adoption of Castilian style short forms.

In order to investigate these changes, I analyze a wide corpus of 2500 medieval Ibero-Romance documents (<https://corpuscodea.es/>) as well as other text types, looking at aspects such as the precise geographical distribution of the short and long variants, the realm where legal documents were issued (ecclesiastical, municipal or private), as well as additional texts from different typologies, such as documents of the Royal Chancellery and literary texts.

In sum, this investigation explores the powerful role of political influence in the introduction of contact-induced morphosyntactic structures. Other similar cases in the Iberian Peninsula will be considered, such as the increase of proclisis (Martins 2011, 2015) and prepositional object marking (Paixão de Sousa 2004) in Portuguese due to Castilian influence during the dynastic union with Spain (1580-1640), and the abrupt decrease of these structures once Portugal regained its independence.

Macro sociohistorical forces, contact, convergence and the development of modern linguistic areas: insights from South Africa

Rajend Mesthrie, University of Cape Town.

Mesthrie (2017) makes a case for the development of a robust South African linguistic area (impinging on neighbouring parts of southern Africa). In pre-colonial times (up to the mid-17th C) the autochthonous Khoisan languages formed an important substratum that resulted in the eventual transformation of Southern Bantu languages, especially in their phonologies. Colonisation wrought further changes on the indigenous Bantu languages, firstly via Afrikaans (17th C on) and then English (19th C on). South Africa differs from other heavily colonised, settled and exploited territories (in the senses used by Mufwene 2001) in that indigenous languages survived and remain an essential part of an African multilingualism with official status since 1994. This has opened up new avenues of mutual influence between the living, growing substrata of mainly Bantu languages and the globally and locally prestigious English language. Mesthrie (2017) shows how Afrikaans played an intermediary role – almost as a clearing house – in disseminating features within the emerging linguistic area. The current paper for ICHL 26 will emphasise the role of two dimensions of macrolinguistic relevance: (a) processes of second language learning under socially constrained conditions (notably apartheid) that resulted in distinctly Africanised varieties of English and (b) a more egalitarian multilingualism today showing “third space” effects among younger people comfortable in English and an African language – i.e. showing innovations that go beyond each of the monolingual codes involved in language switching.

References:

Mesthrie, R. 2017. South Africa and Areal Linguistics. In R. Hickey (ed) *The Cambridge Handbook of Areal Linguistics*. Cambridge: CUP.

Mufwene, S. 2001. *The Ecology of Language Evolution*. Cambridge: CUP.

Verticalization and the historical sociolinguistics of language maintenance

Joseph Salmons

University of Wisconsin – Madison

From the present day back as far as we can see into prehistory, migration and colonization have correlated with language shift, where a community abandons its traditional language(s) for another, usually socially dominant one. A new model ties language shift to changes in community structure, laid out in Brown (2022) and built on Warren (1978). Central to the model is that minority-language communities who control their own local institutions and resources tend to maintain their languages; when that control moves to those beyond the community, a process of ‘verticalization’, we see shift to the language(s) of that broader community. The model has been widely tested with immigrant languages in North America and increasingly beyond (Brown 2022, with initial comparative work in Salmons 2022), and it is general enough that it can be applied to almost any setting of contact and shift past or present.

The model has barely been used for deeper historical situations, where evidence is sparser and harder to interpret, though Frey and Salmons (2012) did an initial study of verticalization in Latin-Germanic contact. This presentation explores how verticalization can be generally integrated into historical sociolinguistic research. Warren identifies five “major functions” carried out within communities and/or from beyond them (1978: 9-13): Production–distribution–consumption; socialization; social control; social participation; mutual support. I draw examples from the history of English, especially English-French contact, to probe how these factors correlate with the ultimate maintenance of English. Recent work (e.g., Timofeeva and Ingham 2018) helps us to see how even important institutional roles for French in domains like religion and education did not create the strong and broad vertical patterns which would have led to wholesale shift to French, rather than just powerful language-contact effects on English.

References

- Brown, Joshua R., ed. 2022. *The verticalization model of language shift: The Great Change in American communities*. Oxford: Oxford University Press.
- Frey, Benjamin E. & Joseph Salmons. 2012. Dialect and language contact in emerging Germanic. *Archaeology and Language: Indo-European Studies presented to James P. Mallory*, ed. by Martin E. Huld, Karlene Jones-Bley and Dean Miller, 95-120. *Journal of Indo-European Studies Monograph*, 60.
- Salmons, Joseph. 2022. The last stages of language shift and verticalization: Comparative Upper Midwestern data. *Selected proceedings of the 11th workshop on immigrant languages in the Americas*, ed. by Kelly Biers & Joshua R. Brown, 71-78. Somerville, MA: Cascadilla Press.
- Timofeeva, Olga and Richard Ingham, eds. 2018. *Mechanisms of French contact influence in Middle English: Diffusion and maintenance*. Special issue of *English Language and Linguistics*, 22.
- Warren, Roland L. 1978. *The community in America*. 3rd edition. Chicago: RandMcNally.

Workshop: Computational models of diachronic language change

Organizers:

Stefania Degaetano-Ortlieb*, Lauren Fonteyn+, Pauline Krielke*, Elke Teich*

*Saarland University, +Leiden University

While the study of diachronic language change has long been firmly grounded in corpus data analysis, it seems fair to state that the field has been subject of a ‘computational turn’ over the last decade or so, computational models being increasingly adopted across several research communities, including corpus and computational linguistics, computational social science, digital humanities, and historical linguistics.

The core technique for the investigation of diachronic change are distributional models (DMs). DMs rely on the fact that related meanings occur in similar contexts and allow us to study lexical-semantic change in a data-driven way (e.g. as argued by Sagi et al. 2011), and on a larger scale (e.g. as shown on the Google NGram corpus by Gulordava & Baroni 2011). Besides count-based models (e.g. Hilpert & Saavedra 2017), contextualized word embeddings are increasingly employed for diachronic modeling, as such models are able to encode rich, context-sensitive information on word usage (see Lenci 2018 or Fonteyn et al., 2022 for discussion).

In previous work, DMs have been used to determine laws of semantic change (e.g. Hamilton et al. 2016b, Dubossarsky et al. 2017) as well as develop statistical measures that help detect different types of change (e.g. specification vs. broadening; cultural change vs. linguistic change; Hamilton et al. 2016a, Del Tredici et al. 2019). DMs have also been used to map change in specific (groups of) concepts (e.g. ‘racism’, ‘knowledge’; see Sommerauer & Fokkens 2019 for a discussion). Further studies have suggested ways of improving the models that generate (diachronic) word embeddings to attain these goals (e.g. Rudolph & Blei 2018).

Existing studies and projects focus on capturing and quantifying aspects of semantic change. Yet, over the past decade, DMs have also been shown to be useful to investigate other types of change in language use, including grammatical change. Within the computational and corpus linguistic communities, for example, Bizzoni et al. (2019, 2020) have shown an interdependency between lexical and grammatical changes and Teich et al. (2021) use embeddings to detect (lexico-) grammatical conventionalization (which may lead to grammaticalization). Within diachronic linguistics, the use of distributional models is focused on examining the underlying functions of grammatical structures across time (e.g. Perek 2016, Hilpert and Perek 2015, Gries and Hilpert 2008, Fonteyn 2020, Budts 2020). Specifically targeting historical linguistic questions, Rodda et al. (2019) and Sprugnoli et al. (2020) have shown that computational models are promising for analyzing ancient languages, and McGillivray et al. (2022) highlight the advantages of word embeddings (vs. count-based methods) while also pointing to the challenges and the limitations of these models.

A common concern across these different communities is to better understand the general principles or “laws” of language change and the underlying mechanisms (analogy, priming, processing efficiency, contextual predictability as measured by surprisal, etc.). In the proposed workshop, we will bring together researchers from relevant communities to talk about the unique promises that computational models hold when applied to diachronic data as well as the specific challenges they involve. In doing so, we will identify common ground and explore the most pressing problems and possible solutions. The program of the workshop will include talks by both invited speakers and open call for paper presentations.

Specific questions will concern:

Model utility: How can we capture change in language use beyond lexical-semantic change, e.g. change in grammatical constructions, collocations, phraseology?

Model quality: How can we evaluate computational models of historical language stages in absence of native-speaker ‘gold standards’? To what extent does the quality of historical and diachronic corpora affect the performance of models?

Model analytics: How do we transition from testing the reliability of models to employing them to address previously unanswered research questions on language change? How can we detect and “measure” change? What are suitable analytic procedures to interpret the output of models?

References:

- Bizzoni, Y., Degaetano-Ortlieb, S., Menzel, K., Krielke, P., and Teich, E. (2019). “Grammar and meaning: analysing the topology of diachronic word embeddings”. In Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change, ACL, Florence, Italy, pp. 175–185.
- Bizzoni, Y., Degaetano-Ortlieb, S., Fankhauser, P., and Teich, E. (2020). “Linguistic variation and change in 250 years of English scientific writing: a data-driven approach”. *Frontiers in Artificial Intelligence*, 3.
- Budts, S. (2020). “A connectionist approach to analogy. On the modal meaning of periphrastic do in Early Modern English”. *Corpus Linguistics and Linguistic Theory*, 18(2), pp. 337–364.
- Del Tredici, M., Fernández, R., and Boleda, G. (2019). “Short-term meaning shift: A distributional exploration.” In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL): Human Language Technologies, Minneapolis, Minnesota, USA, pp. 2069–2075.
- Dubossarsky, H., Weinshall, D., and Grossman, E. (2017). “Outta control: laws of semantic change and inherent biases in word representation models”. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, pp. 1136–1145.
- Fonteyn, L. (2020). “What about grammar? Using BERT embeddings to explore functional-semantic shifts of semi-lexical and grammatical constructions.” *Computational Humanities Research CEUR-WS*, pp. 257–268.
- Fonteyn, L., Manjavacas, E., and Budts, S. (2022). “Exploring Morphosyntactic Variation & Change with Distributional Semantic Models”. *Journal of Historical Syntax*, 7(12), pp. 1–41.
- Gries, S. T., and Hilpert, M. (2008). “The identification of stages in diachronic data: variability-based Neighbor Clustering”. *Corpora*, 3(1), pp. 59–81.
- Gulordava, K., and Baroni, M. (2011). “A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus”. In Proceedings of Geometrical Models for Natural Language Semantics (GEMS), EMNLP, Edinburgh, United Kingdom, pp. 67–71.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016a). “Cultural shift or linguistic drift? comparing two computational models of semantic change”. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP, Austin, Texas, USA, pp. 2116–2121.
- Hamilton, W. L., Leskovec, J., and Jurafsky, D. (2016b). “Diachronic word embeddings reveal statistical laws of semantic change”. In Proceedings of Morphosyntactic Variation & Change with DSMs, 54th Annual Meeting of the Association for Computational Linguistics, ACL, Berlin, Germany, pp. 1489–1501.
- Hilpert, M., and Saavedra, D.C. (2020). “Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims”. *Corpus Linguistics and Linguistic Theory*, 16(2), pp. 393–424.
- Hilpert, M. and Perek, F. (2015). “Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts”. *Linguistics Vanguard*, 1(1), pp. 339–350.
- Lenci, A. (2018). “Distributional Models of Word Meaning”. *Annual Review of Linguistics*, 4, pp. 151–171.
- McGillivray, B., Jensen, G.B., Salama, K. (2022). “Investigating patterns of change, stability, and interaction among scientific disciplines using embeddings”. *Nature Humanities and Social Science Communication* 9, 285.

- Perek, F. (2016). “Using distributional semantics to study syntactic productivity in diachrony: a case study”. *Linguistics*, 54(1), pp. 149–188.
- Rodda, M.A., Probert, P., and McGillivray, B. (2019). “Vector space models of Ancient Greek word meaning, and a case study on Homer”. *TAL Traitement Automatique des Langues*, 60(3), pp. 63–87.
- Rudolph, M., and Blei, D. (2018). “Dynamic embeddings for language evolution”. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*, Lyon, France, pp. 1003–1011.
- Sagi, E., Kaufmann, S., and Clark, B. (2011). “Tracing semantic change with Latent Semantic Analysis”. *Current Methods in Historical Semantics*, 73, pp. 161–183.
- Sommerauer, P., and Fokkens, A. (2019). “Conceptual Change and Distributional Semantic Models: An Exploratory Study on Pitfalls and Possibilities”. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, Florence, Italy, pp. 223–233.
- Sprugnoli, R., Moretti, G., and Passarotti, M. (2020). “Building and Comparing Lemma Embeddings for Latin. Classical Latin versus Thomas Aquinas”. *IJCoL. Italian Journal of Computational Linguistics*, 6(6-1), pp. 29–45.
- Teich, E., Fankhauser P., Degaetano-Ortlieb, S., and Bizzoni, Y. (2021). “Less is More/More Diverse: On the Communicative Utility of Linguistic Conventionalization”. *Frontiers in Communication*, 5.

A Diachronic Analysis of Using Sentiment Words in Scandinavian Literary Texts from 1870-1900

Ali Al-Laith, Kirstine Nielsen Degn, Bolette Sandford Pedersen, Daniel Hershovich, Jens Bjerring-Hansen

University of Copenhagen, Denmark

Diachronic corpora, or collections of texts spanning a significant time period, are useful computational linguistics tools for studying language change and evolution. They can be used to investigate changes in vocabulary [1, 2], grammar [3], and usage patterns over time [4]. Additionally, they can be used to analyze the development of different language varieties, and dialects [5, 6]. They can also be used to understand how language is used in different contexts and how language use changes in response to social, cultural, and historical factors [7, 8, 9, 10]. Other potential applications of diachronic corpora in computational linguistics include the creation of language processing tools and systems that consider the historical context in which a text was produced [11].

To track the cultural development in society through literature analysis, one can study the themes and ideas present in the literature over time and look for trends, and changes [12]. This includes examining shifts in how these themes and ideas are presented and changes in the style and form of literature and subjects addressed. It is also essential to consider the social, political, and economic context in which the literature was produced, as these factors can influence the culture and development of society [13]. There are several ways to track the use of emotional language over time in literature [14, 15]. One method is to conduct a content analysis of the text, in which the frequency of emotional words and phrases is counted [16]. Another approach is to use thematic analysis, which involves examining the themes related to emotions in the text and how they are presented [17, 18]. A third option is to employ sentiment analysis, which uses computational tools to analyze the emotional content of the text through natural language processing algorithms or the use of dictionaries or lexicons of emotional words and phrases [19, 20].

Given the large collection of diachronic literary texts that is currently available, we expect to see variations in the usage of sentiment-bearing words in different time periods and in relation to the shifting discussions and themes over time. In this research, we examine the evolution of sentiment words' use in the MEMO corpus, a collection of almost 900 Danish and Norwegian novels from the latter part of the 19th century [21].

A dynamic BERTopic model is a powerful tool for analyzing the evolution of topics in a collection of documents over time. It uses transformers and class-based TF-IDF to identify clusters of words and phrases representing the main topics discussed in the corpus. It also incorporates important words in the topic descriptions for improved interpretability. By tracking the use of sentiment words, the dynamic BERTopic model allows us to gain a deeper understanding of the changes and developments in the discussions over time. To further analyze these patterns, we employ the Danish Sentiment Lexicon (DDS)¹ [22, 23] to identify any changes in the use of sentiment words over time.

This research aims to track the evolution of sentiment towards a specific topic over time and the evolution of which words are used to express sentiment. The goal is to understand how public sentiment or attitudes towards the topic have changed, identify trends and patterns in the way the topic is discussed, and provide historical context that helps explain how the topic has been represented.

Keywords— Sentiment Analysis, Sentiment Lexicon, Topic Modeling, Scandinavian Literature, Diachronic Corpora, Danish Text, Norwegian Text

¹<https://github.com/dsldk/danish-sentiment-lexicon>

References

P. Cassotti, P. Basile, M. de Gemmis, and G. Semeraro, “Analysis of lexical semantic changes in corpora with the diachronic engine.,” in CLiC-it, 2020.

- T. McEnery and A. Wilson, “Diachronic Corpora in the Study of Vocabulary Change: A Review,” *Corpus Linguistics and Linguistic Theory*, vol. 11, no. 2, pp. 203–226, 2015.
- C. Mair, “Tracking ongoing grammatical change and recent diversification in present-day standard english: the complementary role of small and large corpora,” in *The changing face of corpus linguistics*, pp. 355–376, Brill, 2006.
- I. Renau and R. Nazar, “Automatic extraction of lexical patterns from corpora,” in *En EURALEX International Congress: Lexicography and Linguistic Diversity*, pp. 823–830, 2016.
- A. Karjus, R. A. Blythe, S. Kirby, and K. Smith, “Challenges in detecting evolutionary forces in language change using diachronic corpora,” arXiv preprint arXiv:1811.01275, 2018.
- A. Jatowt and K. Duh, “A framework for analyzing semantic change of words across time,” in *IEEE/ACM Joint Conference on Digital Libraries*, pp. 229–238, IEEE, 2014.
- M. Hilpert, “The great temptation: What diachronic corpora do and do not reveal about social change,” *Corpora and the Changing Society: Studies in the Evolution of English*. Amsterdam and Philadelphia: John Benjamins, pp. 3–27, 2020.
- G. M. Alessi and A. Partington, “Modern diachronic corpus-assisted language studies: methodologies for tracking language change over recent time.,” 2020.
- M. Liakata and P. Rayson, “Using Diachronic Corpora to Study Language Change and Evolution: A Review,” *Corpus Linguistics and Linguistic Theory*, vol. 8, no. 2, pp. 227–250, 2012.
- S. Kemmer and E. Zaretsky, “Diachronic Corpus-Based Approaches to the Study of Semantic Change: A Review,” *Corpus Linguistics and Linguistic Theory*, vol. 11, no. 1, pp. 29–62, 2015.
- M. Piotrowski, “Natural language processing for historical texts,” *Synthesis lectures on human language technologies*, vol. 5, no. 2, pp. 1–157, 2012.
- M. L. Jockers and D. Mimno, “Significant themes in 19th-century literature,” *Poetics*, vol. 41, no. 6, pp. 750–769, 2013.
- G. Blix, “The Social Role of Literature in Society,” *Acta Universitatis Upsaliensis*, vol. 11, no. 3, pp. 53–63, 1986.
- J. Petzold and M. Dickinson, “Tracking Emotional Language in Literature over Time: A Corpus-Based Approach,” *Corpus Linguistics and Linguistic Theory*, vol. 8, no. 2, pp. 267–291, 2012.
- T. L. C. Jockers and J. D. Porter, “A Quantitative Approach to Tracking Emotional Language in Literary Texts over Time,” *Literary and Linguistic Computing*, vol. 29, no. 1, pp. 115–132, 2014.
- S. Fiedler and K. Kunz, “The Use of Content Analysis in the Study of Emotional Language in Literary Texts,” *Methods of Empirical Linguistics*, vol. 21, no. 1, pp. 17–38, 2014.
- S. Fiedler and K. Kunz, “Thematic Analysis of Emotional Language in Literary Texts: A Diachronic Corpus-Based Study,” *Corpus Linguistics and Linguistic Theory*, vol. 5, no. 2, pp. 195–224, 2009.
- S. Fiedler and K. Kunz, “Thematic Analysis of Emotional Language in Literary Texts: A Review of Methodological Approaches,” *Methods of Empirical Linguistics*, vol. 18, no. 1, pp. 1–23, 2012.
- M. Müller and A. Panchenko, “Sentiment Analysis of Emotional Language in Literary Texts: A Comparative Study of Machine Learning and Dictionary-Based Approaches,” *Corpus Linguistics and Linguistic Theory*, vol. 7, no. 2, pp. 227–250, 2011.
- S. Fiedler and K. Kunz, “Sentiment Analysis of Emotional Language in Literary Texts: A Review of Methodological Approaches,” *Methods of Empirical Linguistics*, vol. 21, no. 1, pp. 17–38, 2014.
- J. Bjerring-Hansen, R. D. Kristensen-McLachlan, P. Diderichsen, and D. H. Hansen, “Mending fractured texts. a heuristic procedure for correcting OCR data,” 2022.
- S. Nimb, S. Olsen, B. S. Pedersen, and T. Troelsgaard, “A thesaurus-based sentiment lexicon for danish: The danish sentiment lexicon,” in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 2826–2832, 2022.
- B. S. Pedersen, S. Nimb, and S. Olsen, “Dansk betydningsinventar i et datalingvistisk perspektiv.,” *Danske Studier*, 2021.

Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal Nature

Gard Jensen¹, Isabell Landwehr², Barbara McGillivray³ and Stefania Degaetano-Ortlieb²

¹Springer Nature Group, ²Saarland University, ³King's College London and The Alan Turing Institute

We trace the linguistic evolution of English written scientific communication within the journal Nature, one of the world's leading multidisciplinary science journals, published since 1869. Our study applies computational models for diachronic linguistic analysis to investigate the statistical distribution of lexical and lexical-semantic features in a collection consisting of over 230,000 titles and abstracts from articles published in the journal Nature between 1869 and 2022, accessed via the Dimensions database (Hook et al. 2018).

We dynamically model changes in scientific language use over time. This overcomes the limitations of working with raw frequencies which tend to highlight only high-frequency features, disregarding low-frequency items (e.g. Biber and Gray 2016; Moskowich and Crespo 2012; Rissanen et al. 1997; Teich et al. 2016). We compare changes in probability distributions of individual lexical, grammatical, and semantic features with relative entropy as a measure of divergence for entire sets of features (e.g. all lemmas, parts of speech etc.), allowing for a comprehensive coverage of frequency bands. The dynamicity of the model is achieved by sliding over the timeline and continuously comparing adjacent time spans. The more a distribution of a feature changes over time, the higher the divergence will be, indicating changes in use. The sum of all features' divergence at a particular point in time gives an overall estimate of how much current language use is distinct from past practices, i.e. if a large number of features shows an increase in divergence over a time span, this will indicate a period of change. In terms of interpretability of the model, we are not only able to detect periods of change in a data-driven fashion, but can attribute these changes to sets of linguistic features that contribute to them. In addition, drawing on title and abstract embeddings for Nature articles using Google's Universal Sentence Encoder, we measure the trends in similarity between articles over time.

Previous work on the publications of The Royal Society of London (Degaetano-Ortlieb and Teich 2019, Degaetano-Ortlieb 2021) has proven the adaptability of applying dynamic divergence models to investigate change in scientific language use, showing specialisation trends at the lexical level and at the same time grammatical conventionalization trends. Sun et al. (2021) show similar results employing word embeddings methods. Research using embedding technologies applied to the labels of scientific disciplines rather than to the linguistic content has also found evidence for disciplines undergoing a process of global convergence combined with local specialisation (McGillivray et al. 2022). Previous work on Nature (Monastersky and Van Noorden 2019a) has shown specialisation of particular keywords in individual titles and abstracts. Our overarching question is whether these trends can be found for the journal Nature at scale, indicating general mechanisms of change in language use which contribute to the formation of the English scientific register. In addition, we are interested in changes that might be an indication of journal-specific linguistic features, especially considering the leading position of Nature in the scientific research landscape, as well as the journal's shift in focus over time (Monastersky and Van Noorden 2019a). We investigate the following sub-questions: (a) Can we observe similar/diverging diachronic trends between Nature and The Royal Society corpus, i.e. can we detect lexical and lexical-semantic diversification and grammatical conventionalization in Nature? (b) While we would assume similar diverging trends at the lexical level (new discoveries and technical advancement call for new linguistic expressions), do we encounter journal-specific trends at the grammatical and semantic level, and if so, are these disparate trends or do some trends start off in one journal and are picked up later in the other? Here we assume, besides grammatical trends indicating terminology formation processes, also changes in grammatical features that indicate text structuring functions (e.g. introductory linguistic

material such as prepositional phrases or discourse markers) and those that meet expressive needs given extra-linguistic pressures, such as passive voice usage during periods of increased experimental work).

References

- Biber, Douglas & Bethany Gray. 2016. *Grammatical complexity in academic English: Linguistic change in writing*. Studies in English Language. Cambridge, UK: Cambridge University Press.
- Degaetano-Ortlieb, S. (2021). Measuring informativity: The rise of compounds as informationally dense structures in 20th century Scientific English. In Elena Soave and Douglas Biber (eds.), *Corpus Approaches to Register Variation*, chapter 11, John Benjamins Publishing Company, pp. 291-312.
- Degaetano-Ortlieb, Stefania and Teich, Elke. "Toward an optimal code for communication: The case of scientific English" *Corpus Linguistics and Linguistic Theory*, vol. 18, no. 1, 2022, pp. 175-207.
- Hook, D.W., Porter, S.J. and Herzog, C., 2018. Dimensions: building context for search and evaluation. *Frontiers in Research Metrics and Analytics*, 3, p.23.
- McGillivray, B., Jensen, G.B., Salama, K. and Schut, D. 2022. Investigating patterns of change, stability, and interaction among scientific disciplines using embeddings. *Humanities and Social Sciences Communications* 9, 285. <https://doi.org/10.1057/s41599-022-01267-5>
- Monastersky, Richard & Van Noorden, Richard. 2019a. 150 years of Nature: a data graphic charts our evolution. *Nature*: 575, 22-23. <https://doi.org/10.1038/d41586-019-03305-w>
- Monastersky, Richard & Van Noorden, Richard. 2019b. 150 years of Nature: a data graphic charts our evolution. Supplementary information: Methodology. *Nature*: 575. <https://www.nature.com/magazine-assets/d41586-019-03305-w/17345736> (last accessed date: 23 December 2022).
- Moskovich, Isabel & Begona Crespo (eds.). 2012. *Astronomy Playne and simple: The writing of science between 1700 and 1900*. Amsterdam/Philadelphia: John Benjamins.
- Rissanen, Matti, Merja Kytö & Kirsi Heikkonen (eds.). 1997. *English in transition: Corpus-based studies in linguistic variation and genre analysis*. Berlin: Mouton de Gruyter.
- Sun, K., Liu, H. and Xiong, W., 2021. The evolutionary pattern of language in scientific writings: A case study of *Philosophical Transactions of Royal Society (1665–1869)*. *Scientometrics*, 126(2), pp.1695-1724.
- Teich, Elke, Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes & Ekaterina Lapshinova-Koltunski. 2016. The linguistic construal of disciplinarity: A data mining approach using register features. *Journal of the Association for Information Science and Technology (JASIST)* 67(7). 1668–1678.

Quantifying Changes in English Noun Compound Productivity and Meaning

Maximilian Maurer & Chris Jenkins & Filip Miletic & Sabine Schulte im Walde
Universität Stuttgart

Combinations of words are considered to be multi-word expressions (MWEs) if they are semantically idiosyncratic to some degree, i.e., the meaning of the combination is not entirely (or even not at all) predictable from the meanings of the constituents [Sag et al., 2002, Baldwin and Kim, 2010]. MWEs subsume multiple morpho-syntactic types, including noun compounds such as *flea market*, which have been explored extensively and across research disciplines from synchronic perspectives [Reddy et al., 2011, Bell and Schäfer, 2013, Schulte im Walde et al., 2013, Salehi et al., 2014, 2015, Schulte im Walde et al., 2016, Cordeiro et al., 2019, Alipoor and Schulte im Walde, 2020, i.a.], but state-of-the-art studies are lacking large-scale distributional approaches towards diachronic models of noun compound meaning. The current study goes beyond the restricted synchronic concept of compound semantics and provides a novel diachronic perspective on meaning changes and compositionality (i.e., meaning transparency) of English noun compounds. We specifically investigate the diachronic evolution of the productivity of compound constituents relative to their degree of compositionality, relying on an established gold standard dataset with human compositionality ratings by Reddy et al. [2011] and a cleaned version of the English diachronic corpus CCOHA [Alatrash et al., 2020]. Given that type and token frequencies and probabilities, type-token ratios, entropy, etc. represent key concepts in determining quantitative properties of corpora as well as regarding individual word types and co-occurrences, we compute a range of statistical measures to quantify changes in productivity. These include Baayen’s Large Number of Rare Events (LNRE) measures [Baayen, 2001], which have become a standard in statistical estimation of productivity, as well as measures that represent textual constants and therefore smooth the effect of different text lengths. For example, Tweedie and Baayen [1998] showed that with the exception of two measures, K suggested by Yule [1944] and Z suggested by Orlov [1983], all constants systematically change as a function of the text length.

In terms of empirical findings, we hypothesise that the current-language degree of compositionality differs for compounds with high- vs. low-productive constituents [Jurafsky et al., 2001, Hilpert, 2015, i.a.]. That is, we expect to find distinct analogical temporal development patterns for compositional compounds (such as *maple tree*, *prison guard*, *climate change*) in comparison to more idiosyncratic compounds (such as *flea market*, *night owl*, *melting pot*), with regard to modifier as well as head productivity. Our results constitute an important step towards a better understanding of compound semantics over time, as well as a reference point for future work deploying other modeling approaches on the same topic.

References

- Reem Alatrash, Dominik Schlechtweg, Jonas Kuhn, and Sabine Schulte im Walde. CCOHA: Clean Corpus of Historical American English. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6958–6966, Marseille, France, 2020.
- Pegah Alipoor and Sabine Schulte im Walde. Variants of Vector Space Reductions for Predicting the Compositionality of English Noun Compounds. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4379–4387, Marseille, France, 2020.
- R. Harald Baayen. *Word Frequency Distributions*. Kluwer Academic Publishers, 2001.
- Timothy Baldwin and Su Nam Kim. Multiword expressions. In Nitin Indurkha and Fred J. Damerau, editors, *Handbook of Natural Language Processing*, pages 267–292. CRC Press, Boca Raton, USA, 2010.
- Melanie J. Bell and Martin Schäfer. Semantic Transparency: Challenges for Distributional Semantics. In *Proceedings of the IWCS Workshop on Formal Distributional Semantics*, pages 1–10, Potsdam, Germany, 2013.

- Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. Unsupervised Compositionality Prediction of Nominal Compounds. *Computational Linguistics*, 45(1):1–57, 2019.
- Martin Hilpert. From hand-carved to computer-based: Noun-Participle Compounding and the Upward Strengthening Hypothesis. *Cognitive Linguistics*, 26(1):1–36, 2015.
- Daniel Jurafsky, Alan Bell, Michelle Gregory, and William D. Raymond. Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Joan Bybee and Paul Hopper, editors, *Frequency and the Emergence of Linguistic Structure*, *Typological Studies in Language*, pages 229–254. John Benjamins, Amsterdam / Philadelphia, 2001.
- Y. K. Orlov. Ein Modell der Häufigkeitsstruktur des Vokabulars. In *Studies on Zipf's Law*, pages 154–233. Brockmeyer, Bochum, 1983.
- Siva Reddy, Diana McCarthy, and Suresh Manandhar. An Empirical Study on Compositionality in Compound Nouns. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pages 210–218, Chiang Mai, Thailand, 2011.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of the Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico, 2002.
- Bahar Salehi, Paul Cook, and Timothy Baldwin. Using Distributional Similarity of Multi-way Translations to Predict Multiword Expression Compositionality. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 472–481, Gothenburg, Sweden, 2014.
- Bahar Salehi, Paul Cook, and Timothy Baldwin. A Word Embedding Approach to Predicting the Compositionality of Multiword Expressions. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics/Human Language Technologies*, pages 977–983, Denver, Colorado, USA, 2015.
- Sabine Schulte im Walde, Stefan Müller, and Stephen Roller. Exploring Vector Space Models to Predict the Compositionality of German Noun-Noun Compounds. In *Proceedings of the 2nd Joint Conference on Lexical and Computational Semantics*, pages 255–265, Atlanta, GA, USA, 2013.
- Sabine Schulte im Walde, Anna Häddy, and Stefan Bott. The Role of Modifier and Head Properties in Predicting the Compositionality of English and German Noun-Noun Compounds: A Vector-Space Perspective. In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*, pages 148–158, Berlin, Germany, 2016.
- Fiona J. Tweedie and R. Harald Baayen. How Variable May a Constant be? Measures of Lexical Richness in Perspective. *Computers and the Humanities*, 32:323–352, 1998.
- G. Udny Yule. *The Statistical Study of Literary Vocabulary*. Cambridge University Press, 1944.

A computerized investigation of Albanian diachronic phonology

Clayton Marr

The Ohio State University

Computerized forward reconstruction, or CFR (Sims-Williams, 2018), offers an automatic and systematic means of testing hypotheses about the chronology of sound change in a language. While computing the effects of historical sound changes over millennia for thousands of etyma is laborious and extremely time-consuming, this task is accomplished within seconds by a CFR system such as DiaSim, which was created for not only evaluating hypothesized relative chronologies of sound changes, or “diachronic cascades”, but also “debugging them” by reporting statistics on how errors pattern (Marr and Mortensen, 2020). As a test case, past work applied this system to the phonological evolution of Latin into French, and a CFR-enabled “debugging” procedure improved accuracy from a 3.2% baseline for a cascade based on the 1934 received view to 84.9%. In the process, various proposals in the post-1934 literature on French were supported by the fact that they were independently produced as part of a systematic debugging process using DiaSim that was undertaken without reference to them (Marr and Mortensen, 2022), while the endeavor also may have revealed a new regular sound change in Old French, which was ultimately robustly supported by additional data (Marr, 2023b). However, as French boasts both a large corpus since medieval times and extensive past research, the experiment with French was more of a “laboratory run” to test the validity of the approach of debugging a language’s historical phonology via CFR, a prelude to bringing it into the field as an investigative technique.

This paper will bring in CFR to tackle Albanian diachronic phonology, starting with the Latin stratum of the its lexicon. Given the lack or loss of attestation of Albanian before the 15th century and its status as the only surviving member of its branch of Indo-European (Rusakov, 2018), reconstruction of Albanian diachronic phonology, and thus of Proto-Albanian, has always leaned heavily on the outcomes of strata of loanwords in Albanian from better-attested sources (Orel, 2000). Of these, the Latin layer (Çabej, 1962; Bonnet, 1998) is by far the most significant. Latin loanwords are more numerous than inheritance from Proto-Indo-European, Proto-Albanian is dated in relation to the time of contact with Latin, and Albanian diachronic phonology is in a large part an exercise in generalization from analyses of the outcomes of ancient Latin loans (Orel, 2000; Demiraj, 2006; Rusakov, 2017; De Vaan, 2018), though with significant contributions from Albanian historical dialectology (Curtis, 2018) and the other “layers”. Nevertheless, issues do remain that concern the Latin layer of Albanian, such as rival etymologies between imperial-era Latin loans and later Romance loans (Bonnet, 1998), and these have potential implications for the reconstruction of Proto-Albanian, and the greater mysteries of the language’s history within the Balkans (Friedman and Joseph, 2022). Thus, an evaluation and debugging of the received view on Albanian diachronic phonology as applied to its largest single pillar, the Latin stratum, offers both a new approach to an old but still vexing problem, and a step for CFR as an empirical method, between the curated “lab” case of French, and the “field” of understudied languages and language families.

This endeavor will apply DiaSim to CLEA, a dataset compiled in 2020–2022 and to be released with this paper, of 1007 Albanian etyma of ancient Latin origin as asserted by at least one of a set of reputed references (Bonnet, 1998; Orel, 1998, 2000; De Vaan, 2018; Topalli, 2017; Çabej, 1986), and will work from a base cascade representing the views of Orel (2000) and De Vaan (2018). The same debugging process as Marr and Mortensen (2022) will be applied, with accuracy reported for modern Albanian outcomes, and discussion of any systematic patterning of errors and possible solutions proposed.

Keywords: computerized forward reconstruction, diachronic phonology, Albanian, Latin

References

- Anamali, S. and Prifti, K. (2002). *Historia e popullit shqiptar*, volume 1. Botimet Toena. AShSh and IGJL (2006). *Fjalor i gjuhës shqipe*. Online at: <https://fjalorthi.com/Fjalorthi.com>. Bonnet, G. (1998). *Les mots latins de l'albanais*. L'Harmattan.
- Brüch, J. (1922). Lateinische etymologien. *Indogermanische Forschungen*, 40(1):196–247.
- Buflì, G. and Rocchi, L. (2021). A historical-etymological dictionary of turkisms in albanian (1555-1954).
- Çabej, E. (1953). Grupet nd, ng në gjuhën shqipe. *Buletin për Shkencat Shoqërore*, 4.
- Çabej, E. (1962). Zur Charakteristik der lateinischen Lehnwörter im Albanischen. *Revue de Linguistique (București)*, 7:161–199.
- Camarda, D. (1864). *Saggio di grammatologia comparata sulla lingua albanese*. Livorno: Succesore di Egisto Vignozzi.
- Çabej, E. (1986). *Studime gjuhësore*, volume I–VIII. Prishtinë: Rilindja.
- Curtis, M. C. (2012). Slavic-Albanian language contact, convergence, and coexistence. The Ohio State University. Ph.D. dissertation.
- Curtis, M. C. (2018). 98. The dialectology of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1800–1811. De Gruyter Mouton.
- Dashi, B. (2013). Italianismi nella lingua albanese. *Edizioni Nuova Cultura*.
- De Vaan, M. (2004). PIE *e in Albanian. *Die Sprache. Zeitschrift für Sprachwissenschaft*, 44(1):70–85.
- De Vaan, M. (2018). 95. The phonology of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1732–1749. De Gruyter Mouton.
- Demiraj, B. (2018). 100. The evolution of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1812–1815. De Gruyter Mouton.
- Demiraj, B. and Dayan, P. (1997). *Albanische Etymologien: Untersuchungen zum albanischen Erbwortschatz*, volume 134. Rodopi.
- Demiraj, S. (2004). *Gjuhësi Ballkanike*. Akademia e Shkencave e Republikës së Shqipërisë, Instituti i Gjuhësisë dhe i Letërsisë.
- Demiraj, S. (2006). Albanian. In *The Indo-European Languages*. Routledge.
- Ducellier, A. (1981). *La façade maritime de l'Albanie au moyen age, Durazzo et Valona du XIe au XVe siècle*. Greece, Thessaloniki: Institute for Balkan Studies.
- Fine, J. V. and Fine, J. V. A. (1994). *The late medieval Balkans: A critical survey from the late twelfth century to the Ottoman conquest*. University of Michigan Press.
- Friedman, V. A. and Joseph, B. D. (2022). *The Balkan Languages*. Cambridge University Press.
- Gjinari, J. (1989). *Dialektet e gjuhës shqipe*. Tiranë: Akademia e Shkencave e RPS të Shqipërisë.
- Haarmann, H. (1972). *Der lateinische Lehnwortschatz im Albanischen*. Hamburg: Helmut Buske.
- Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. *Loanwords in the world's languages: A comparative handbook*, pages 35–54.
- Helbig, R. (1903). *Die italienischen Elemente im Albanesischen*. JA Barth.
- Huld, M. E. (1979). *An etymological glossary of selected Albanian items*. University of California, Los Angeles.
- Huld, M. E. (1986). Accentual stratification of Ancient Greek loanwords in Albanian. *Zeitschrift für vergleichende Sprachforschung*, 99(2. H):245–253.
- Hyllested, A. and Joseph, B. D. (2022). 13. Albanian. In Olander, T., editor, *The Indo-European Language Family*, pages 227–248. Cambridge University Press.
- Janda, R. D. and Joseph, B. D. (2003). Reconsidering the canons of sound-change. In *Historical Linguistics 2001: Selected Papers from the 15th International Conference on Historical Linguistics, Melbourne, 13-17 August 2001*, volume 237, page 205. John Benjamins Publishing.
- Jokl, N. (1984). *Sprachliche Beiträge zur Paläo-Ethnologie der Balkanhalbinsel: (zur Frage der ältesten griechisch-albanischen Beziehungen)*, volume 29. Austrian Academy of Sciences Press.

- Jorgaqi, K. (2001). Ndikimi i italishtes në letërsinë e vjetër shqipe: shek. XVI-XVII, Botimet Toena, Tiranë.
- Joseph, B. D. (2020). Language contact in the Balkans. *The handbook of language contact*, pages 537–549.
- Kore, M. K. (2013). Similarities between Albanian and Romanian in the entire language sub- systems. *Mediterranean Journal of Social Sciences*, 4(2):175–175.
- Lafe, G. (2000). Der italienische Einfluß auf das Albanische. Zweiter Teil. *Wörterbuch der Italianismen im Albanischen*, in *Ponto-Baltica*, 10:31–120.
- Landi, A. (1989). *Gli elementi latini nella lingua albanese*. Napoli: Edizioni Scientifiche Italiane.
- Marr, C. (2023a). The angevin–albanian element in the albanian lexicon.
- Marr, C. (2023b). A regular velar onset voicing rule for 12th century French. Submitted for publication.
- Marr, C. and Mortensen, D. R. (2020). Computerized forward reconstruction for analysis in diachronic phonology, and Latin to French reflex prediction. In *Proceedings of LT4HALA 2020-1st Workshop on Language Technologies for Historical and Ancient Languages*, pages 28–36.
- Marr, C. and Mortensen, D. R. (2022). Large-scale computerized forward reconstruction yields new perspectives in French diachronic phonology. *Diachronica*.
- Matzinger, J. (2018). 98. The lexicon of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1788–1800. De Gruyter Mouton.
- Mazaudon, M. and Lowe, J. B. (1993). Regularity and exceptions in sound change. In *Annual Conference of the Linguistic Society of Belgium*.
- Meyer, G. (1891). *Etymologisches Wörterbuch der albanesischen Sprache*. Strasbourg: Trübner.
- Meyer-Lübke, W. (1935). *Romanisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.
- Miklosich, F. (1870). *Albanische Forschungen: I. Die slavischen Elemente im Albanischen*. Wien: Karl Gerold's Sohn; Denkschr. Akad. Wien XIX.
- Miklosich, F. (1871). *Albanische Forschungen: II. Die romanischen Elemente im Albanischen*. Wien: Karl Gerold's Sohn; Denkschr. Akad. Wien XIX.
- Newmark, L. (1998). *Albanian-English Dictionary*. Oxford University Press, USA. Online at: [http://www.seelrc.org:8080/albdict/Online Albanian Dictionary](http://www.seelrc.org:8080/albdict/Online%20Albanian%20Dictionary).
- Niedermann, M. (1905). *Contributions à la critique et à l'explication des gloses latines*, volume 1. Attinger.
- Ölberg, H. M. (1972). *Griechisch-albanische Sprachbeziehungen*. Institut für vergleichende Sprachwissenschaft der Universität.
- Orel, V. (1998). *Albanian etymological dictionary*. Brill.
- Orel, V. È. (2000). *A concise historical grammar of the Albanian language: reconstruction of Proto-Albanian*. Brill.
- Pokorny, J. (1959). *Indogermanisches etymologisches Wörterbuch*, volume 1. Bern, München: Francke Verlag.
- Prifti, E. (2012). Aromanian elements of an Albanian argot.
- Rusakov, A. (2017). Albanian. *The Indo-European languages*, pages 552–608.
- Rusakov, A. (2018). 94. The documentation of Albanian. In *Handbook of Comparative and Historical Indo-European Linguistics*, pages 1716–1731. De Gruyter Mouton.
- Schuhardt, H. E. M. (1866-8). *Der Vokalismus des Vulgärlateins I-III*. Leipzig: B. G. Teubner.
- Seliščev, A. M. and Olesch, R. (1931). *Slavjanskoe naselenie v Albanii*. Böhlau.
- Sims-Williams, P. (2018). Mechanising historical phonology. *Transactions of the Philological Society*, 116(3):555–573.
- Svane, G. (1992). *Slavische Lehnwörter im Albanischen*, volume 67. Aarhus Universitetsforlag.
- Thumb, A. (1909). *Altgriechische elemente des albanischen*. *Indogermanische Forschungen*, 26:1–20.
- Topalli, K. (2017). *Fjalor etimologjik i gjuhës shqipe. Qendra e Studimeve Albanologjike, Instituti i Gjuhësisë dhe i Letërsisë*.

Uhlisch, G. (1964). Neugriechische Lehnwörter im Albanischen. Typescript: Berlin.

Vătășescu, C. I. (1997). Vocabularul de origine latină din limba albaneză în comparație cu româna. PhD thesis, Editura Universității din București.

Vicario, F. (1993). L'influsso, lessicale veneto in albanese. *Balkan-Archiv (NF)*, 17(18):1992. Ylli, X. (2000). Das slavische Lehngut im Albanischen: Teil 2: Ortsnamen. Peter Lang International Academic Publishers.

The LSCD Benchmark - A testbed for diachronic word meaning tasks

Dominik Schlechtweg
Universität Stuttgart

Lexical Semantic Change Detection (LSCD) is a field of NLP that studies methods automating the analysis of changes in word meanings over time. In recent years, this field has seen much development in terms of models, datasets and tasks (Schlechtweg et al., 2020). This has made it hard to keep a good overview of the field. Additionally, with the multitude of possible options for preprocessing, data cleaning, dataset versions, model parameter choice or tuning, clustering algorithms, and change measures a shared testbed with common evaluation setup is needed in order to precisely reproduce experimental results. Hence, we present a benchmark repository implementing evaluation procedures for models on most available LSCD datasets. We hope that the resulting benchmark by standardizing the evaluation of LSCD models and providing models with near-SOTA performance can serve as a starting point for researchers to develop and improve models. The benchmark allows for a wide application and testing of models by focusing on multilingual models and their evaluation on several languages.

Models solving the LSCD task often employ sub-models solving other related lexical semantic tasks like Word Sense Induction (WSI, Navigli, 2009) or Word-in-Context (WiC, Pilehvar & Camacho-Collados, 2020). Performance on these tasks can be evaluated separately contributing to optimization of individual model components and to facilitation of error analysis. However, existing data sets for the latter two tasks are usually synchronic, which makes it hard to compare different sub-models and select optimal ones for the LSCD task that requires good performance on diachronic data. Hence, we exploit existing, richly annotated LSCD datasets as evaluation data for WSI and WiC in a diachronic setting. Using the same data sets for evaluation of WSI, WiC and LSCD has the additional advantage that performance on the meta task LSCD can be directly related to performance on the subtasks WSI and WiC, as it can be assumed that performance on the subtasks directly determines performance on the meta task. We aim to stimulate transfer between the fields of WSI, WiC and LSCD by providing a repository allowing for evaluation on all these tasks with shared model components.

References

- Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, Nina Tahmasebi. 2020. SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. Proceedings of the 14th International Workshop on Semantic Evaluation.
- Roberto Navigli. 2009. Word sense disambiguation: a survey. ACM Computing Surveys.
- Mohammad Taher Pilehvar and Jose Camacho-Collados. 2019. WiC: The Word-in-Context Dataset for Evaluating Context-Sensitive Meaning Representations. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Model evaluation for diachronic semantics: A view from Portuguese and Spanish

Amaral, Patrícia*; Hu, Hai**; Tian, Zuoyu*; Kübler, Sandra*

*Indiana University; **Shanghai Jiao Tong University

For research on semantic change that spans over several centuries, assessing the accuracy of embeddings comes with two challenges: (i) native speakers who can provide judgments about meaning are not available, and (ii) historical corpora are often much smaller than contemporary datasets, which raises issues of model accuracy (Hellrich, 2019; Hu et al., 2021). This paper presents the lessons learned from developing intrinsic evaluations to test the quality of distributional models used to investigate semantic change in Medieval Spanish and Portuguese. For Spanish we experimented on a 7 million word corpus (Chronicles corpus, with texts from 13th-16th c.) (Hu et al., 2021) and for Portuguese on a ca. 2,5 million token corpus, CIPM, with texts from 12th-16th c. (Tian et al., 2021).

The lessons learned include the following: 1) We cannot use tests developed for modern languages/corpora off the shelf, since the tests' vocabulary (e.g., capitals of the world, country names and currencies) does not overlap with that of the historical corpus.

We cannot use tests developed for other historical corpora without adaptations since those corpora tend to be restricted to specific domains, which also leads to a lack of overlap in vocabulary.

We need to account for spelling and morphological variation, which are important features of many Medieval corpora. For the historical Spanish corpus, e.g., we had to delete the test “adjective to adverbs” from contemporary Spanish (Cardellino, 2016), which maps an adjective to its corresponding adverb in mente, since the variability of forms of adverbs in Medieval Spanish would have resulted in more than one possible target form, including multi-word expressions (Company and Flores Da'vila, 2014). Instead, we added tests for several types of inflection (verbal morphology, gender and number in adjectives). The morphology tests were generated by using vocabulary based on the frequency counts from the Chronicles corpus. A summary of our analogy test is given in Table 1.

If the corpora are very small, using analogy tests alone may not provide enough information. Our work on the Portuguese corpus shows that using different tests that include a range of relations is important. The tests we created include: word similarity, outlier detection, and coherence assessment (see Table 2 for a summary). The latter is based on Zhao et al. (2018), who proposed a new evaluation method for assessing the quality of domain-specific word embedding models. They assume that the neighbors of a given word embedding should have the same characteristics of that word (e.g. neighbors of drug names should be drug names). In the Portuguese corpus, names of people and places are frequent, thus we can assess coherence by reporting the percentage of neighbors generated for a proper noun that were also proper nouns.

To summarize: Given the importance of register in research on semantic and syntactic change, as well as orthographic and morphological variation in historical corpora, specific tests are required for a proper assessment of distributional models in studies of semantic change. Overall, assessment of word embeddings for historical research must meet the following criteria: appropriateness (corpus vocabulary is taken into account), sustainability (i.e. not requiring extensive expert input), comprehensiveness (tasks target different types of relations, i.e. syntactic, semantic, morphological), and complementarity (avoiding the biases of individual methods).

| Source | Category | Example | #Questions |
|--------|--|-----------------------------------|------------|
| MTS | Morphology nouns: kinship terms | padre madre : hijo hija | 506 |
| | Morphology verbs: third person singular | comer come : ir va | 650 |
| | Morphology verbs: infinitive to participle | saber sabido : tomar tomado | 1190 |
| | Morphology verbs: gerund to participle | sabiendo sabido : tomando tomado | 1190 |
| ours | Morphology adj.: singular to plural | negra negras : rica ricas | 992 |
| | Morphology adj.: singular to plural | negro negros : rico ricos | 992 |
| | Morphology adj.: masc to fem | negro negra : negros negras | 992 |
| | Morphology adj.: masc to fem | negros negras : ricos ricas | 992 |
| | Morphology nouns : singular to plural | casa casas: capilla capillas | 1332 |
| | Morphology/Semantics: antonyms | feliz infeliz : posible imposible | 42 |
| | Semantics: antonyms | cerca lejos : bien mal | 342 |
| Total | | | 9220 |

Table 1: Structure of our analogy test; MTS denotes the analogy test from [Mikolov et al. \(2013\)](#), translated into Spanish.

| Test | Categories | #Questions |
|----------------------|---|------------|
| Analogy Test | nouns: gender; nouns: singular to plural; verbs: 1st person singular to 3rd person singular; verbs: 3rd person singular to 3rd person plural; verbs: infinitive to 3rd person singular; verbs: infinitive to gerund | 2994 |
| Word Similarity | synonymous; related (not synonymous); not related | 97 |
| Outlier Detection | body parts; Christianity; color; food; geography; parts of buildings; titles/professions; war | 512 |
| Coherence Assessment | proper nouns (names of people and places) | 25 |

Table 2: Summary of the benchmark for assessing word embeddings generated for Medieval Portuguese

References

- Cardellino, C. (2016). Spanish Billion Words Corpus and embeddings. Online at <https://crscardellino.github.io/SBWCE/>; retrieved August 2019.
- Company, C. and Flores Dávila, R. (2014). Adverbios en mente. In Company, C., editor, *Sintaxis histórica de la lengua española. Tercera parte: Adverbios, preposiciones y conjunciones. Relaciones interoracionales*, pages 1195–1340. Fondo de Cultura Económica y Universidad Nacional Autónoma de México.
- Hellrich, J. (2019). *Word Embeddings: Reliability and Semantic Change*. PhD thesis, Jena University Language and Information Engineering Lab.
- Hu, H., Amaral, P., and Kübler, S. (2021). Word embeddings and semantic shifts in historical Spanish: Methodological considerations. *Digital Scholarship in the Humanities*, 37(2):441–461.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of International Conference on Learning Representations (ICLR)*, Scottsdale, AZ.
- Tian, Z., Jarrett, D., Escalona Torres, J., and Amaral, P. (2021). BAHP: Benchmark of assessing word embeddings in historical Portuguese. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 113–119, Punta Cana, Dominican Republic (online). Association for Computational Linguistics.
- Zhao, M., Masino, A. J., and Yang, C. C. (2018). A framework for developing and evaluating word embeddings of drug-named entity. In *Proceedings of the BioNLP 2018 workshop*, pages 156–160, Melbourne, Australia. Association for Computational Linguistics.

Using simulated data to evaluate models of Indo-European vocabulary evolution

Philipp Rönchen¹, Oscar Billing¹, and Tilo Wiklund²

¹Department of Linguistics and Philology, Uppsala University, Sweden

²Chief Data Scientist, UAB Sensmetry, Vilnius, Lithuania

In the last two decades the project of using data from the lexicon of modern languages to make inferences about historical language stages, though long envisioned (Hymes 1960, Embleton 1986), has been gaining steam. Gray and Atkinson (2003), Bouckaert et al. (2012) and Chang et al. (2015) use increasingly sophisticated methods to estimate the age of Indo-European, however the results of the earlier studies run counter to the established majority opinion in historical linguistics (Pronk, 2022) and Chang et al.'s methodology gives a different result. This raises the question how different computational models can be validated (see Nakhleh et al. 2005, Ritchie and Ho 2019, Jäger 2019a and 2019b)

Ideally one would like to evaluate computational methods using held-out data sets and test cases in which the correct inferences are known. However, compared to other disciplines like biology, the amount of lexical data available in data bases is very limited and the precise history of most language families in the world is unknown, leaving only a few quite shallow families as potential test cases. Moreover, it is not clear whether the success of a computational model on a language family from one part of the world should generalise to other families, since different evolutionary mechanisms might have operated. To work around the lack of data available for validation, Greenhill et al. (2009), Murawaki (2015) and Bradley (2016) simulate data sets which they use to evaluate computational methods.

We create a large number of simulated data sets to evaluate the inferences of Chang et al. (2015) and Bouckaert et al. (2012) on Indo-European. Our data sets are specifically tailored to the methodologies of Chang et al. and Bouckaert et al. and try to mimic different plausible (though hypothetical) pre-histories of Indo-European, including loan events, a tree topology not too far from the consensus view in historical linguistics, and varying lexical change rates. We employ the computational fact that it is much easier to create realistic models for simulating data than it is to make inferences from existing data (see Kelly and Nicholls 2017 for the difficulties involved in constructing an inference method that allows for loans).

Both Chang et al.'s and Bouckaert et al.'s methodologies fail to correctly infer the age of Indo-European that was used to create our simulated data sets. We believe this warrants more investigation in the validity of different computational models.

References

- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., Gray, R. D., Suchard, M. A., and Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, 337(6097):957–960.
- Bradley, S. (2016). Synthetic language generation and model validation in BEAST2. arXiv preprint arXiv:1607.07931.
- Chang, W., Cathcart, C., Hall, D., and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1):194–244.
- Embleton, S. M. (1986). *Statistics in historical linguistics*, volume 30. Brockmeyer.
- Gray, R. D. and Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426(6965):435–439.
- Greenhill, S. J., Currie, T. E., and Gray, R. D. (2009). Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B: Biological Sciences*, 276(1665):2299–2306.
- Hymes, D. H. (1960). Lexicostatistics so far. *Current anthropology*, 1(1):3–44.
- Jäger, G. (2019a). Computational historical linguistics. *Theoretical Linguistics*, 45(3-4):151–182.
- Jäger, G. (2019b). Model evaluation in computational historical linguistics. *Theoretical Linguistics*, 45(3-4):299–307.

- Kelly, L. J. and Nicholls, G. K. (2017). Lateral transfer in stochastic dollo models. *The Annals of Applied Statistics*, 11(2):1146–1168.
- Murawaki, Y. (2015). Spatial structure of evolutionary models of dialects in contact. *Plos one*, 10(7):e0134335.
- Nakhleh, L., Warnow, T., Ringe, D., and Evans, S. N. (2005). A comparison of phylogenetic reconstruction methods on an IE dataset. *Transactions of the Philological Society*, 3(2):171–192.
- Pronk, T. (2022). Indo-European secondary products terminology and the dating of Proto-Indo-Anatolian. *Journal of Indo-European Studies*, 49(1&2):141–170.
- Ritchie, A. M. and Ho, S. Y. (2019). Influence of the tree prior and sampling scale on Bayesian phylogenetic estimates of the origin times of language families. *Journal of Language Evolution*, 4(2):108–123.

Evaluating historical word embeddings: strategies, challenges and pitfalls

Oksana Dereza, Theodorus Fransen and John P. McCrae
University of Galway, Insight Centre for Data Analytics

When it comes to the quantitative evaluation of word embeddings, there are two main strategies: extrinsic, i.e. using pre-trained embeddings as input vectors in a downstream ML task, such as language modelling, and intrinsic, i.e. through analogy and similarity tasks that require special datasets ([Bakarov, 2018](#)).

Extrinsic evaluation

Language modelling seems to be the easiest way to evaluate historical word embeddings, since it is language independent, scalable and does not require dataset creation. Hypothetically, using pre-trained embeddings must lower the perplexity of a language model, even if these embeddings were trained on a different period of the same language. However, language modelling, as well as the majority of modern NLP tasks, is not very relevant to historical linguistics, so we might want to find a better downstream task or turn to intrinsic evaluation.

Intrinsic evaluation

There are two major tasks used for intrinsic evaluation of word embeddings: similarity and analogy. The **similarity task** consists in comparing similarity scores of two words yielded by an embedding model to those calculated based on experts' judgment. We did not explore this option, because it requires too much manual work by definition. The **analogy task** is simply asking an embedding model “What is to **a**' as **b** is to **b'**?”, and expecting **a** as an answer. Analogy datasets can be created automatically or semi-automatically if there exists a comprehensive historical dictionary of a language in question in machine readable format or a WordNet.

Traditionally, analogy datasets are based on pairwise semantic proportion and therefore every question has a single correct answer. Given the high level of variation in historical languages, such a strict definition of a correct answer seems unjustified. Therefore, in our Early Irish analogy dataset we follow the authors of [BATS \(Gladkova et al., 2016\)](#) providing several correct answers for each analogy question and evaluating the performance with set-based metrics, such as an average of vector offset over multiple pairs (3CosAvg).

Our dataset consists of 4 parts: morphological variation and spelling variation subsets were automatically extracted from [eDIL \(eDIL, 2019\)](#), while synonym and antonym subsets are translations of correspondent BATS parts proofread by 4 expert evaluators. However, the scores that Early Irish embedding models achieved on the analogy dataset were low enough to be statistically insignificant. Such a failure may be a result of the following problems:

The highest inter-annotator agreement score (Cohen's kappa) between experts was 0.339, which reflects the level of disagreement in the field of historical Irish linguistics. It concerns such fundamental questions as “What is a word? Where does it begin and end? What is a normalised spelling of a word at a particular stage of the language history?”, which was discussed in ([Doyle et al., 2018](#)) and ([Doyle et al., 2019](#)) regarding tokenisation. It is arguable that it might be true for historical linguistics in general.

There is a lack of standardisation in different resources for the same historical language. For example, ~65% of morphological and spelling variation subsets, retrieved from eDIL, were not present in the whole Early Irish corpus retrieved from CELT (CELT, 1997), on which the biggest model was trained. As for synonym and antonym subsets, ~30% are missing in the corpus. Although our embedding models used subword information and were able to handle unknown words, such a discrepancy between the corpus,

on which they were trained, and the historical dictionary, which became the source for the evaluation dataset, seriously affected the performance. This discrepancy originates from different linguistic views and editorial policies used by different text editors, publishers and resource developers throughout time.

References

- [Bakarov, A. \(2018\). A Survey of Word Embeddings Evaluation Methods. ArXiv:1801.09536 \[Cs\]. <http://arxiv.org/abs/1801.09536>](https://arxiv.org/abs/1801.09536)
- [Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. \(2016\). Enriching Word Vectors with Subword Information. ArXiv:1607.04606 \[Cs\]. <http://arxiv.org/abs/1607.04606>](https://arxiv.org/abs/1607.04606)
- [CELT: Corpus of Electronic Texts. \(1997\). University College Cork. <http://www.ucc.ie/celt>](http://www.ucc.ie/celt)
- [Doyle, A., McCrae, J. P., & Downey, C. \(2018\). Preservation of Original Orthography in the Construction of an Old Irish Corpus. Proceedings of the LREC 2018 Workshop “CCURL2018 – Sustaining Knowledge Diversity in the Digital Age”, 67–70.](https://www.lrec-conf.org/proceedings/lrec2018/workshop_ccurl2018/papers/67-70.pdf)
- [Doyle, A., McCrae, J. P., & Downey, C. \(2019\). A Character-Level LSTM Network Model for Tokenizing the Old Irish text of the Würzburg Glosses on the Pauline Epistles. Proceedings of the Celtic Language Technology Workshop, 70–79. <https://www.aclweb.org/anthology/W19-6910>](https://www.aclweb.org/anthology/W19-6910)
- [EDIL 2019: An Electronic Dictionary of the Irish Language, based on the Contributions to a Dictionary of the Irish Language \(Dublin: Royal Irish Academy, 1913-1976\). \(2019\). \[www.dil.ie\]\(http://www.dil.ie\)](http://www.dil.ie)
- [Gladkova, A., Drozd, A., & Matsuoka, S. \(2016\). Analogy-based detection of morphological and semantic relations with word embeddings: What works and what doesn't. Proceedings of the NAACL Student Research Workshop, 8–15. <https://doi.org/10.18653/v1/N16-2002>](https://doi.org/10.18653/v1/N16-2002)
- [Kutuzov, A., Øvrelid, L., Szymanski, T., & Velldal, E. \(2018\). Diachronic word embeddings and semantic shifts: A survey. ArXiv:1806.03537 \[Cs\]. <http://arxiv.org/abs/1806.03537>](https://arxiv.org/abs/1806.03537)
- [Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. \(2020\). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. ArXiv:2007.11464 \[Cs\]. <http://arxiv.org/abs/2007.11464>](https://arxiv.org/abs/2007.11464)

Workshop ‘Ambiguity (avoidance) as a factor in language change’

Eva Zehentner (University of Zurich), eva.zehentner@es.uzh.ch
Ilaria De Cesare (Universität Potsdam), decesare@uni-potsdam.de

There is general agreement on the fact that ambiguity is abundant in language, and is present in all linguistic domains. That is, as Felser (2017: 271) points out, “strings of human speech sounds (or strings of graphemes) may be compatible – at least, temporarily – with more than one possible phonological, morphological, syntactic, semantic or pragmatic representation”. In the linguistic literature, most attention so far has been paid to lexical and structural ambiguities, such as e.g. garden path sentences or PP-attachment. The role of ambiguity in language change has been extensively discussed in the literature, specifically its role as a crucial factor in both syntactic and semantic reanalysis (e.g. Evans & Wilkins 2000; Traugott 2012; Traugott & Trousdale 2013; Denison 2017; Felser 2017). Ambiguous instances are here often considered as ‘bridging contexts’, triggering new analyses of a structure or word if ambiguity is prevalent enough (cf. Winter-Froemel 2021: 12-14, positing a threshold of 50% of ambiguous contexts needed to instigate reanalysis). However, these assumptions have been criticised as ‘logically flawed’ (De Smet 2009: 1728), as ambiguity is the result rather than the motivation of reanalysis – in order for ambiguities to arise, the target interpretation must already be given (De Smet 2009: 1729). Despite calling for caution in attributing causal force to ambiguity and viewing it as the ‘spark’ of change, De Smet & Markey (2021: 21) nevertheless show that ambiguity can act as ‘fuel’ in the diffusion of innovations, as they may “spread more easily to contexts where the innovation is less conspicuous”.

A related, yet slightly different question is the role of ambiguity avoidance as a cognitive pressure in language use and change (e.g. Stefanowitsch 2021): Studies into syntactic phenomena such as differential object marking (Fedzechinka et al. 2012; Lemmolo 2013; Levshina 2020; Tal et al. 2022), argument structure (e.g. Flack 2007; De Swart et al. 2008; Lamers & de Swart 2012 or MacWhinney et al. 2014; Kittilä et al. 2011; Kulikov et al. 2006), as well as e.g. Temperley (2003) on relative clauses suggest that there is a cross-linguistic tendency for ambiguity to be resolved, with ‘trade-offs’ between disambiguation strategies being frequently observed. A well-known example for such trade-offs is the history of English argument structure, where the decreasing disambiguation power of case marking appears to correlate with an increasing reliance on constituent order for identifying ‘who did what to whom’. Explorations of the role of ambiguity avoidance in morphotactic histories of languages (e.g. Baumann et al. 2019) as well as lexical and morphological homophony avoidance (e.g. Baermann 2011; Munteanu 2021) suggest that similar tendencies could be given in other domains.

Although such explanations are intuitively appealing and seem to hold in specific cases, they have not been extensively tested against empirical data, and existing results are somewhat inconclusive (cf. e.g. Fedzechinka et al. 2012; Levshina 2021; Zehentner 2021; De Cesare & Demske 2022). Furthermore, the cross-linguistic commonness of ambiguity as discussed by Wasow (2015) or Piantadosi et al. (2012) as well as the synchronic evidence (e.g. Ferreira 2006, 2008; Ferreira & Dell 2000; Ferreira et al. 2005; Roland et al. 2006, among others) call into question whether ambiguity avoidance can be considered a general, strong, potentially universal and stable cognitive mechanism in the first place, or whether it may only come into play in relatively restricted areas and under specific circumstances.

The present workshop aims to provide a platform for discussing both the role of ambiguity in language change as well as the role of ambiguity avoidance as a cognitive pressure triggering and/or shaping diachronic change. In particular, the workshop addresses the following questions (among potential other issues):

- Do ambiguous (bridging) contexts serve as (a) triggers of change, (b) the fuel for change, facilitating or accelerating change once an innovation has emerged, or (c) instead as the result of change (cf. De Smet 2009; De Smet & Markey 2021)?
- Does ambiguity (avoidance) affect different linguistic levels in similar ways, or are there differences between e.g. ambiguity effects in phonology versus syntax, or between the role of structural versus semantic/ lexical ambiguity (e.g. Winter-Froemel 2021)?
- Which changes are particularly prone to be affected by ambiguity (avoidance)? Is there a difference between linguistic phenomena at the same linguistic level?
- How does ambiguity affect comprehension and (how) are comprehension effects reflected in production?
- What is the role of the (individual) speaker or hearer regarding language change and ambiguity and how conscious/ unconscious are these processes?
- Are both ambiguity and ambiguity avoidance stable and quasi-universal factors in language change or does their effect interact with other factors impacting language change, both social and cognitive ones? For example, how does mode interact with ambiguity?
- How does ambiguity relate to vagueness and/or fuzziness, and where does this distinction come into play in regard to language change (cf. e.g. Denison 2017)?

While we welcome more theoretically-focused contributions of such issues to the workshop, one main goal for the workshop is to discuss the role of ambiguity (avoidance) on the basis of empirical data, as well as possible methodological challenges. That is, we particularly invite empirical (corpus-linguistic or other) contributions on the impact of ambiguity in change, aiming to also address questions such as:

- How can we operationalise ambiguity (avoidance) in historical data? Which methodological problems may arise in data extraction and analysis in diachronic studies of ambiguity?
- What are possible options to assess the impact of ambiguity(avoidance) on specific changes, and how can we empirically address the question of causality in particular instances of change (cf. Winter-Froemel 2021; Zehentner 2021)?

Finally, the workshop intends to cover an as broad as possible range of languages and time-depths beyond Germanic/ Indo-European languages, and is neither restricted in terms of linguistic level of analysis nor regarding theoretical framework.

References

- Baerman, Matthew. 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 47(1). 1-29. <https://doi.org/10.1017/S0022226710000022>.
- Baumann, Andreas, Christina Prömer & Nikolaus Ritt. 2019. Word form shapes are selected to be morphotactically indicative. *Folia Linguistica* 53(40, 1): 129-151. <https://doi.org/10.1515/flih-2019-0007>.
- De Cesare, Ilaria & Ulrike Demske. 2022. Ambiguity avoidance as an efficiency strategy driving word order change. *The 25th International Conference on Historical Linguistics, University of Oxford*.
- Denison, David. 2017. Ambiguity and vagueness in historical change. In Marianne Hundt, Sandra Mollin & Simone Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*, 292-318. <https://doi.org/10.1017/9781316091746.013>.
- De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119: 1728-1755. <https://doi.org/10.1016/j.lingua.2009.03.001>.
- De Smet, Hendrik & Marie-Anne Markey. 2021. The spark or the fuel? On the role of ambiguity in language change. *Journal of Historical Syntax* 5: 1-24 (Special Issue: Whither reanalysis?). <https://doi.org/10.18148/hs/2021.v5i32-39.144>.
- De Swart, Peter, Monique Lamers & Sander Lestrade. 2008. Animacy, argument structure, and argument encoding. *Lingua* 118(2): 131-140. <https://doi.org/10.1016/j.lingua.2007.02.009>.

- Evans, Nicholas & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76(3): 546-592.
- Fedzechkina, Maryia, Florian Jaeger & Elissa Newport. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109(44): 17897-17902. <https://doi.org/10.1073/pnas.121577610>.
- Felser, Claudia. 2017. Syntactic ambiguity in real-time language processing and diachronic change. In Marianne Hundt, Sandra Mollin & Simone Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives*, 271-291. <https://doi.org/10.1017/9781316091746.012>.
- Ferreira, Victor. 2006. Avoid ambiguity! (If you can). *CRL Technical Reports* 18: 3-13.
- Ferreira, Victor. 2008. Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation: Advances in Research and Theory* 49: 209-246. [https://doi.org/10.1016/S0079-7421\(08\)00006-6](https://doi.org/10.1016/S0079-7421(08)00006-6).
- Ferreira, Victor & Gary Dell. 2000. The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40: 296-340. <https://doi.org/10.1006/cogp.1999.0730>.
- Ferreira, Victor, Robert Slevc & Erin Rogers. 2005. How do speakers avoid ambiguous linguistic expressions? *Cognition* 96: 263-284. <https://doi.org/10.1016/j.cognition.2004.09.002>.
- Flack, Kathryn. 2007. Ambiguity avoidance as contrast preservation: Case and word order freezing in Japanese. In Leah Bateman, Adam Werle, Michael O'Keefe & Ehren Reilly (eds.), *UMass Occasional Papers in Linguistics 32: Papers in Optimality Theory III*. Amherst: GLSA, 57-88. <https://doi.org/doi:10.7282/T3HQ42T3>.
- Iemmolo, Giorgio. 2013. Symmetric and asymmetric alternations in direct object encoding. *STUF Language Typology and Universals* 66(4): 378-403. <https://doi.org/10.1524/stuf.2013.0019>.
- Kittilä, Seppo, Katja Västi & Jussi Ylikoski (eds.). 2011. *Case, animacy and semantic roles*. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.99>.
- Kulikov, Leonid, Andrej Malchukov & Peter de Swart (eds.). 2006. *Case, valency and transitivity*. Amsterdam: Benjamins. <https://doi.org/10.1075/slcs.77>.
- Lamers, Monique & Peter de Swart (eds.). 2012. *Case, word order and prominence*. Dordrecht: Springer. <https://doi.org/10.1007/978-94-007-1463-2>.
- Levshina, Natalia. 2020. Communicative efficiency and differential case marking: A reverse engineering approach. *Linguistics Vanguard* 7(s3): 20190087. <https://doi.org/10.1515/lingvan-2019-0087>.
- Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Front. Psychol.* 12: 648200. <https://10.3389/fpsyg.2021.648200>.
- MacWhinney, Brian, Andrej Malchukov & Edith Moravcsik (eds.). 2014. *Competing motivations in grammar and usage*. Oxford: OUP. <https://doi.org/10.1093/acprof:oso/9780198709848.001.0001>.
- Munteanu, Andrei. 2021. Homophony avoidance in the grammar: Russian nominal allomorphy. *Phonology* 38(3): 401-435. <https://doi.org/10.1017/S0952675721000257>.
- Piantadosi, Steven, Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122: 280-291. <https://doi.org/10.1016/j.cognition.2011.10.004>.
- Roland, Douglas, Jeffrey Elman & Victor Ferreira. 2006. Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98: 245-272. <https://doi.org/10.1016/j.cognition.2004.11.008>.
- Stefanowitsch, Anatol (ed.). 2021. Ambiguity (non-)avoidance in English. Special Issue: *Zeitschrift für Anglistik und Amerikanistik* 69(3). <https://doi.org/10.1515/zaa-2021>.
- Tal, Shira, Kenny Smith, Jennifer Culbertson, Eitan Grossman & Inbal Arnon. 2022. The impact of information structure on the emergence of differential object marking: An experimental study. *Cognitive Science* 46(3): e13119. <https://doi.org/10.1111/cogs.13119>.
- Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79(3): 464-484.

- Traugott, Elizabeth. 2012. On the persistence of ambiguous linguistic contexts over time: Implications for corpus research on micro-changes. *Language and Computers* 75: 231-246. https://doi.org/10.1163/9789401207713_019.
- Traugott, Elizabeth & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199679898.001.0001>.
- Wasow, Thomas. 2015. Ambiguity avoidance is overrated. In Susanne Winkler (ed.). *Ambiguity: Language and communication*. Berlin: De Gruyter, 29-48. <https://doi.org/10.1515/9783110403589-003>.
- Winter-Froemel, Esme. 2021. Reinvestigating ambiguity and frequency in reanalysis: A two-step methodology for corpus-linguistic analyses based on bridging use exposure. *Journal of Historical Syntax* 5: 1-24 (Special Issue: Whither reanalysis?). <https://doi.org/10.18148/hs/2021.v5i32-39.143>.
- Zehentner, Eva. 2021. Ambiguity avoidance as a factor in the rise of the English dative alternation. *Cognitive Linguistics* 33(1): 3-33. <https://doi.org/10.5167/uzh-216295>.

The role of ambiguity at different stages of diachronic change

Elena Smirnova (University of Neuchâtel)

This paper addresses the question of the relationship between ambiguity and reanalysis in syntactic change: Is ambiguity the prerequisite or the result of reanalysis (cf. De Smet 2009; De Smet & Markey 2021). It will be argued that there is no established chronological order between the two phenomena, and that ambiguity may both trigger change as well as result from change. Furthermore, attention will be paid to the notion of the (bridging) context. It will be shown that ambiguous interpretations often arise from paradigmatic analogical relations, and not from contextually induced pragmatic implicatures, as frequently assumed in the literature.

In Smirnova et al. (2019), we argued that ambiguity is crucial for the initiation of change. Using the grammaticalization of the German passive auxiliary *werden* ‘become’, we demonstrated how atypical and infrequent combinations triggered change due to their inherent ambiguity. At the same time, the initial reinterpretation of ambiguous combinations of *werden* with past participles of activity verbs resulted in a cascade of semantic reinterpretation processes that affected first combinations with accomplishment verbs and then with achievement verbs, the process we called diffusion. Importantly, those combinations had existed before and they only became ambiguous after the reinterpretation of *werden* and activity verbs had taken place. That is, we observe a “chain” of ambiguous contexts, where the resolution of one ambiguous interpretation triggers the next ambiguous context.

The present study will focus on German deictic adverbial elements *hin-/her-P* such as *heraus*, *hinaus* etc. In present-day German, they display multiple structural ambiguities and may be considered as adverbs, as verb particles as well as parts of circumpositions, see (1) – (3):

- ambiguous between adverb & verb particle
- (1) *Plötzlich kam er **heraus**.*
Suddenly he came out.
- ambiguous between adverb, part of the circumposition, & verb particle
- (2) *Plötzlich kam er **aus dem Wald heraus**.*
Suddenly he came out of the forest.
- non-ambiguous verb particle
- (3) *Das Buch kommt mit einer Startauflage von 30 Tausend Exemplaren **heraus**.*
The book gets out with an initial print run of 30 thousand copies.

Though speakers of present-day German do not seem to face interpretation problems, as the structural ambiguity is not always tied with semantic ambiguity, alternative spelling variants in the corpus (e.g. *heraus gekommen* vs. *herausgekommen*) suggest that some speakers/writers are unsure as to the categorial status of these elements.

The present corpus study is based on the data from 1600 to present-day German (DTA-Kernkorpus & DWDS-Kernkorpus). Five pairs of *hin-/her*-adverbs (*hinaus–heraus*, *hinein–herein*, *hinauf–herauf*, *hinab–herab*, *hinunter–herunter*) will be analyzed with respect to their combinatorial potential. It will be shown that diachronically, ambiguity resides in local contexts of individual elements and only sometimes leads to reanalysis, which may but not need to be coupled with semantic reinterpretation. When the reanalysis takes place in one context, it is likely that ambiguity will arise in one or several other related contexts. That is, ambiguous contexts trigger further ambiguous contexts. Similar to the grammaticalization case of *werden* mentioned above, the ambiguity “chain”

relies heavily on paradigmatic analogical relations.

Methodologically, the paper will discuss some problems that arise when dealing with diachronic corpus data, namely the problem of detecting ambiguous contexts in the data.

References

De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119: 1728-1755.

De Smet, Hendrik & Marie-Anne Markey. 2021. The spark or the fuel? On the role of ambiguity in language change. *Journal of Historical Syntax* 5: 1-24.

Smirnova, Elena & Robert Mailhammer, Susanne Flach. 2019. The role of atypical combinations in the grammaticalization of passives in German and English. *Diachronica* 36:3, 384–416.

Losing one's senses: causes of obsolescence in lexical semantics

Hilke Ceuppens & Hendrik De Smet (KU Leuven)

While the general mechanisms of semantic extension are relatively well-understood (e.g. Geeraerts 1997; Traugott & Dasher 2005), the reverse side of the process – semantic loss – has been theorized less thoroughly. The present paper proposes one recurrent pattern of change that gives rise to obsolescence in lexical semantics.

Theoretical background: The proposed mechanism hinges on two assumptions. First, it is hypothesized that polysemy can be sustained as long as senses can be distinguished in usage. Typically, distinguishability is safeguarded by contextual clues: as long as different senses occupy different contextual niches (e.g. specific genres, specific collocational schemas, etc.) they do not give rise to ambiguity, so that polysemy at word-level is sustainable. This predicts that contextual overlap between senses is what gives rise to obsolescence. Contextual overlap is particularly likely to arise when semantic extension is caused by contextually-driven inferences as opposed to, for instance, metaphor. Second, the different senses of a word are linked into radial networks, organized around one or more core senses, from which peripheral senses are derived (e.g. Evans 2005). It can therefore be predicted that the loss of a core sense will affect any peripheral senses derived from (and synchronically motivated by) it. In combination, then, the emergence of a new sense through contextual implicatures is likely to threaten the source sense of the change, which in turn is likely to threaten any other senses derived from the core sense. This results in a cascade of obsolescing senses.

Empirical evidence: The proposed mechanism is supported here through a number of case studies on English evaluative adjectives. For these, radial networks are first proposed based on the relevant entries in the Oxford English Dictionary. Next, the diachronic predictions of the above model – particularly, the predicted sequence of semantic extensions and losses – is tested against diachronic corpus evidence, drawing on data from Early English Books Online.

By way of example, the adjective *strange* initially had a core sense 'foreign' (1) that motivated various derived senses, including 'unfamiliar' (2), 'unfriendly, uncomplying' (3). However, through pragmatic implicature the sense 'unfamiliar' gave rise to new extensions 'exceptional, abnormal' (4). Contextual overlap with the original core sense 'foreign' caused loss of the latter, which was accompanied by the loss of other extensions from it, particularly the sense 'unfriendly, uncomplying'.

- (1) your excellent renome shyneth as well in *strange* regions as with in the royaume of england (1472, EEBO) ['foreign']
- (2) than was no cocko / betwene the eest and west to laye wronge egges / within a *straunge* nest
['unfamiliar']
- (3) ffor i trowe i was neuer *straunge* to doo for you / that laye in my power (1481, EEBO)
['unfriendly, uncomplying']
- (4) to whom did hadde maruailous and *strange* aduentures (1532, EEBO) ['exceptional, abnormal']

The mechanism proposed and documented here offers one recurrent and internally driven scenario for semantic obsolescence and explains, at least in part, why some polysemies are sustained over long periods, while others are diachronically unstable.

References

Evans, V. 2005. The meaning of time: Polysemy, the lexicon and conceptual structure. *Journal*

of Linguistics 41: 33-75.

Geeraerts, D. 1997. *Diachronic prototype semantics: A contribution to historical lexicology*.
Oxford: Oxford University Press.

Structural ambiguity in language comprehension and production

Claudia Felser (University of Potsdam)

Ambiguity has often been argued to play a role in language change, but the psycholinguistic mechanisms and cognitive constraints that might facilitate ambiguity-related change are as yet poorly understood. Here I will discuss structural ambiguity from the perspectives of real-time sentence comprehension and production.

During sentence comprehension, local syntactic ambiguities as in the garden-path sentence *The log floated down the river sank* can give rise to processing difficulty. The initial omission of disambiguating sentence material (*The log that was floated...*) may lead comprehenders to misanalyse the first part of the sentence. Coming across the disambiguating word or phrase (here, the verb *sank*) later on will disrupt comprehension and trigger computationally costly structural and semantic revision processes. Garden-path effects reflect comprehenders' tendency to parse locally ambiguous strings of words as if they were unambiguous. This allows for processing to be fast and incremental but carries the risk of computing erroneous analyses. Erroneous local parsing decisions may be licensed by the grammar (as in the case of garden-path sentences) or not. In the latter case, maintaining rather than correcting an unlicensed analysis may sometimes be the more resource-friendly option, especially if the analysis is structurally economical and does not result in misinterpretation. Note that parsing errors can also occur if the input is unambiguous, and that misanalyses that find their way into the grammar may result in more rather than less ambiguity (compare De Smet, 2009).

While ambiguity can create problems for language comprehension, it should not normally be a problem for speakers or writers as the message to be conveyed is perfectly clear to them. Avoiding to produce structural ambiguities may be motivated by audience design considerations, however. A speaker/writer seeking to avoid ambiguity would have to (i) be aware of which syntactic encoding variant of the message to be conveyed is ambiguous and likely to cause comprehension difficulties, and (ii) decide in favour of an unambiguous structural variant even if this variant is not the easiest one to produce. As real-time language production is incremental with limited planning scope, and subject to cognitive and memory-related constraints (MacDonald, 2013), this kind of audience design is more likely to be applied during writing than during speaking. Evidence for speakers' choosing to avoid syntactic ambiguity is indeed relatively scarce (Ferreira, 2008; Wasow, 2015).

In summary, while ambiguities tend to be resolved unconsciously during comprehension, avoiding structural ambiguity in language production would seem to require potentially costly, and possibly conscious, planning. Psycholinguistic models that propose a tight link between real-time production and comprehension (Gambi & Pickering, 2017) might offer amore integrative perspective on ambiguity avoidance, however.

References

- De Smet, H. (2009). Analysing reanalysis. *Lingua* 119, 1728-1755.
- Ferreira, V.S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation: Advances in Research and Theory* 49, 209-246.
- Gambi, C., & Pickering, M.J. (2017). Models linking production and comprehension. In E.M.Fernández & H.S. Cairns (eds.). *The handbook of psycholinguistics* (pp. 157-181). Wiley.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology* 4:226. doi: 10.3389/fpsyg.2013.00226
- Wasow, T. (2015). Ambiguity avoidance is overrated. In S. Winkler (ed.), *Ambiguity: language and communication* (pp. 29-47). De Gruyter.

Text-type specific conventions, subordinate environments and ambiguity (avoidance) in Medieval Spanish passive *se*-constructions

Anne Wolfsgruber (Humboldt University Berlin)

The Latin reflexive pronoun *se* has knowingly developed into a middle marker in the Romance languages and can mark today a wide variety of constructions. In Spanish, it serves e.g. as a marker of (non-)oppositional middles, anticausatives, passives and impersonal active constructions (cf. Sansò 2011, De Benito Moreno 2022 among many others; for a typology of *Middle Voice Systems* cf. Inglese 2021). It is known that already Medieval Spanish texts exhibit constructions that are to be classified as passive *se*-constructions. In these, a theme subject (in this case *el pan* ‘the bread’ in (1)) agrees with the verb (cf. Lapesa 1950, Monge 1955, Ricós 1995, Bogard 2006 among others).

- (1) verán por los ojos cómo se gana el pan
 see.FUT.3PL by the eyes how REFL earn.PRS.3SG the bread
 ‘They will see by the eyes how bread is earned.’ [*Cid*, 90]

However, from about 1250 onwards, structures are found that move away from the passive interpretation on formal grounds, e.g. constructions in which no theme subject is expressed anymore. This creates more and more ambiguous environments that open up the way for what should rather be analyzed as impersonal active *se*-constructions (cf. Giacalone Ramat & Sansò 2011 for an extensive study on similar developments in Old Italian).

It has been posited in the literature that the development of impersonal active *se*-constructions involves a reanalysis of the *se_{passive}* V subject structure to *se_{imp}* V object_(former subject) structure (cf. Bassols de Climent 1948, Monge 1955, Detges & Waltereit 2002, Martins 2005 among others), i.e. the fact that on the surface level, the post-verbal subject of a passive *se*-sentence occupies a position shared (on the surface level) by the object of transitive SVO configurations, seems to provide crucial, ambiguous grounds to fuel a reanalysis which then leads to visible changes, e.g. instances in which the lexical subject is dropped, not readily identifiable or not necessary anymore.

An analysis of legal texts (CORDE, 1250-1400 C.E.) reveals new insights on passive *se*-constructions on several levels: On a general, textual level, these texts exhibit a style that tries to avoid ambiguity in that lexical DPs are often further specified by the use of relative clauses (cf. Temperley 2003) to make clear e.g. which legal party is being referred to, whose belongings are at stake, etc.

- (2) Et aquellas cosas que se pueden uender [...]
 and those things COMP REFL can.PRS.3SG sell.PTCP
 ‘And these things which can be sold’ [*Fuero de Soria*, p.161]

Interestingly, it appears that precisely this text-type specific tendency seems to favor the use of passive *se*-constructions. This is reflected in that between 80% and up to 90% of all passive *se*-sentences of the quantitatively analyzed legal texts are found in subordinate constructions as shown in (2)). This is much higher than the occurrence of passive *se*-constructions in subordinate environments in e.g. scientific texts (showing greater variance, 30% to 60% depending on the text). The novel data are significant because the high text-type specific occurrence of passive *se* in subordinate structures and specifically relative *que* bring two important ingredients for the further development of impersonal active *se* to the table:

- i) *se* is forced into a preverbal position in these subordinate contexts – as opposed to main clause configurations where at this stage *se* could also be found in post-verbal positions cf. Fontana 1993, Bouzouita 2008, MacKenzie 2019 i.a. – thus fixing the *se* + verb linearization in

passive *se*-constructions which in turn has been viewed to be crucial for an SVO (re-)interpretation.

- ii) in structures like lexical DP + *que* + *se* + verb, the connection between the lexical DP as the subject of a passive sentence is weakened because the DP is positioned outside the subordinate structure that contains the *que* + *se* + verb complex. This syntactic configuration seems to be connected to a higher frequency of elliptic structures in later legal texts.

References

- [CORDE] Real Academia Española. *Corpus Diacrónico del Español*. <https://www.rae.es/banco-de-datos/corde>
- Bassols de Climent, Mariano. 1948. *Sintaxis histórica de la lengua latina*. Vol. 2. Madrid: C.S.I.C.
- Bogard, Sergio. 2006. El clítico *se*. Valores y evolución. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española*, Primera parte: La frase verbal, 753-870. México D.F.: Universidad nacional autónoma de México. Fondo de cultura económica.
- Bouzouita, Miriam. 2008. At the syntax-pragmatics interface: Clitics in the history of Spanish. In Robin Cooper and Ruth Kempson (eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*, 221-263. London: College Publications.
- De Benito Moreno, Carlota (2022): *The Middle Voice and Connected Constructions in Ibero-Romance A variationist and dialectal account*. Amsterdam: John Benjamins.
- Detges, Ulrich and Waltereit, Richard. 2002. Grammaticalization vs. reanalysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft* 21: 151–195.
- Fontana, Josep. 1993. Phrase structure and the syntax of clitics in the history of Spanish. PhD dissertation. Philadelphia: University of Pennsylvania.
- Giacalone Ramat, Anna and Sansò, Andrea. 2011. From Passive to impersonal: A case study from Italian and its implications. In Andrej Malchukov and Anna Siewierska (eds), *Impersonal constructions. A cross-linguistic perspective*, 189-228. Amsterdam: John Benjamins.
- Inglese, Guglielmo. 2021. Towards a typology of middle voice systems. *Linguistic Typology* 26(3): 489- 531.
- Lapesa, Rafael. 1950. *Historia de la lengua española. Segunda edición corregida y aumentada*. Madrid: Escelicer.
- Mackenzie, Ian. 2019. *Language structure, variation and change: the case of Old Spanish syntax*. Cham: Palgrave Macmillan.
- Martins, Ana Maria. 2005. Passive and impersonal *se* in the history of Portuguese. In Claus D. Pusch, Johannes Kabatek and Wolfgang Raible (eds.), *Romanische Korpuslinguistik. Korpora und diachrone Sprachwissenschaft*, 411-429. Tübingen: Narr.
- Monge, Félix. 1955. Las frases pronominales con sentido impersonal en español. <http://ifc.dpz.es/recursos/publicaciones/01/64/1monge.pdf> (09.09.2022).
- Ricós, Amparo. 1995. *Uso, función y evolución de las construcciones pasivas en español medieval. (Estudio de ser + participio y se + forma verbal)*. València: Universitat de València.
- Sansò, Andrea. 2011. Grammaticalization and prototype effects. A history of the agentive reflexive passive in Italian. *Folia Linguistica Historica* 32: 219-252.
- Temperley, David. 2003. Ambiguity avoidance in English relative clauses. *Language* 79(3): 464-484.

Sound changes tend to reduce morphotactic ambiguity

Nikolaus Ritt & Irene Böhm (University of Vienna)

Our paper discusses ambiguity in the semiotic relation between phonotactic shapes and morphotactic structures. We hypothesize that such ambiguity is dispreferred because it impedes the processing and the acquisition of morphological regularities (Korecky-Kröll et al. 2014; Post et al. 2008), and that it might, therefore, be a significant factor in the actuation and implementation of phonological changes.

To test that hypothesis, we investigated three English sound changes and asked whether they reduced or increased the morphotactic ambiguity of the phonotactic shapes they affected. To measure morphotactic ambiguity, we used appropriate corpora (such as the EEBO, the PPCME, the PPCEME, and the LAEME Corpus) to establish type and token frequencies of word forms with pre- and post- change shapes. Then we determined the proportions of morphologically simple and complex items among word shapes before and after the changes. Our prediction was that the changes should significantly skew the distribution of complex vs. simple items among words with the same phonotactic shapes, so that some word form shapes would become increasingly indicative of morphotactic complexity and others of simplicity.

The sound changes we investigated were (a) the Middle English lenition (or voicing) of final /s/ in noun plurals (ModE *stone*[z] < OE *stan+a*[s]), genitives (ModE *man*[z] < OE *monn+e*[s]), and third person present indicatives (ModE *sin*[z] < Northern ME *sinne*[s]; Ringe 2003); (b) Early Middle English Open Syllable Lengthening (MEOSL), which lengthened short non-high vowels in open disyllables of words regularly if they became monosyllabic (EME /makə/ > /ma:kə/ > /ma:k/ ‘make’), but only rarely in disyllables whose second syllable remained stable (EME /bodi/ > */bo:di/ ‘body’; Mailhammer, Kruger & Makiyama 2015, Minkova & Lefkowitz 2020); as well as (c) the (sporadic) devoicing of past tense /d/ after sonorants in forms such as *spoilt* or *burnt* (Lahiri 2009; Weřna 2009).

The findings from all three studies provided considerable support for our hypothesis. (a) The lenition of plural /s/ significantly reduced the morphotactic ambiguity of forms in which the plural morpheme surfaced as /z/ (i.e., after vowels and sonorants). After the change, the vast majority of these items were complex, while forms ending in sonorants or vowels followed by /s/ were predominantly simple (Baumann, Prömer & Ritt 2019). (b) MEOSL and its failure to affect open disyllables had the combined effect that disyllabic wordforms with heavy first syllables became increasingly indicative of morphologically complex words, while disyllables with light first syllables strongly signalled morphologically simple words (Matzinger & Ritt 2021). Finally, (c) the irregular past tense forms produced by the devoicing of final /d/ after sonorants were – at least for a while – slightly less ambiguous than their regular competitors, since these shared the shapes of many simple items ending in voiced /d/ (such as *wind*, *round*, *build*, or *bold*; Baumann, Prömer & Ritt 2019).

Our findings suggest that sound change tends to reduce morphotactic ambiguities and to be blocked where its implementation would increase them. Our paper describes our methods and our findings in greater detail, and relates our study to extant research on morphotactics (Dressler & Dziubalska-Kořaczyk 2006, 2010; Baumann & Kaźmierski 2018), on the way in which sound changes interact with the frequency of phonotactic patterns in the lexicon and in use (Wedel 2006; Blevins 2009; Kelley & Tucker 2017), and on principles underlying the way in which languages exploit the design space of phonotactically well-formed sound patterns for building actual words and word forms (Tamariz 2004, 2008; Vitevich 2005; Reali & Griffiths 2009; Monaghan et al. 2014; Pierrehumbert 2016; Dautriche et al. 2017).

References

- Baumann, Andreas, and Kamil Kaźmierski. 2018. "Assessing the Effect of Ambiguity in Compositionality Signaling on the Processing of Diphones." *Language Sciences* 67 (May): 14–32. <https://doi.org/10.1016/j.langsci.2018.03.006>.
- Baumann, Andreas, Christina Prömer, and Nikolaus Ritt. 2019. "Word Form Shapes Are Selected to Be Morphotactically Indicative." *Folia Linguistica* 40 (1): 129–51. <https://doi.org/10.1515/flih-2019-0007>.
- Blevins, Juliette. 2009. "Structure-Preserving Sound Change: A Look at Unstressed Vowel Syncope in Austronesian." In *Austronesian Historical Linguistics and Culture History: A Festschrift for Bob Blust*, edited by Alexander Adelaar and Andrew Pawley, 33–49. Canberra: Pacific Linguistics.
- Dautriche, Isabelle, Kyle Mahowald, Edward Gibson, and Steven T. Piantadosi. 2017. "Wordform similarity increases with semantic similarity: An analysis of 100 languages". *Cognitive Science* 41 (8): 2149–2169. <https://doi.org/10.1111/cogs.12453>.
- Davies, Mark. 2017. "Early English books online (EEBO)." <https://www.english-corpora.org/eebo/>
- Dressler, Wolfgang U, and Katarzyna Dziubalska-Kořaczyk. 2006. "Proposing Morphotactics." *Wiener Linguistische Gazette* 73 (1–19): 108–9.
- Dressler, Wolfgang U, Katarzyna Dziubalska-Kořaczyk, and Lina Pestal. 2010. "Change and variation in morpho-tactics." *Folia Linguistica Historica*, 44: 51–67. <https://doi.org/10.1515/flih.2010.003>.
- Kelley, Matthew C., and Benjamin V. Tucker. 2017. "The Effects of Phonotactic Probability on Auditory Recognition of Pseudo-Words." *The Journal of the Acoustical Society of America* 141 (5): 4038. <https://doi.org/10.1121/1.4989319>.
- Korecky-Kröll, Katharina, Wolfgang U. Dressler, Eva Maria Freiberger, Eva Reinisch, Karlheinz Mörth, and Gary Libben. 2014. "Morphotactic and phonotactic processing in German-speaking adults". *Language Sciences* 46: 48-58. <https://doi.org/10.1016/j.langsci.2014.06.006>.
- Kroch, Anthony, and Ann Taylor. 2000. "Penn-Helsinki Parsed corpus of Middle English (PPCME)". Philadelphia, PA: University of Pennsylvania.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2004. "Penn-Helsinki Parsed corpus of Early Modern English (PPCEME)". Philadelphia, PA: University of Pennsylvania.
- Lahiri, Aditi. 2009. "The Dental Preterites in the History of English". In *The nature of the word: Studies in honor of Paul Kiparsky*, edited by Kristin Hanson and Sharon Inkelas, 507-525. Cambridge, MA: Mit Press.
- Laing, Margaret. 2013. "A Linguistic Atlas of Early Middle English (LAEME), 1150-1325." Edinburgh: The University of Edinburgh.
- Mailhammer, Robert, William W. Kruger, and Alexander Makiyama. 2015. "Type Frequency Influences Phonological Generalizations: Eliminating Stressed Open Syllables with Short Vowels in West Germanic." *Journal of Germanic Linguistics* 2(3): 205–37. <https://doi.org/10.1017/S1470542715000069>.
- Matzinger, Theresa, and Nikolaus Ritt. 2022. "Phonotactically probable word shapes represent attractors in the cultural evolution of sound patterns". *Cognitive Linguistics* 33 (2): 415-446.
- Minkova, Donka, and Michael Lefkowitz. 2020. "Middle English Open Syllable Lengthening (MEOSL) or Middle English Compensatory Lengthening (MECL)?" *English Language and Linguistics*, no. December 2015: 1–26. <https://doi.org/10.1017/S1360674319000522>.
- Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen, and Simon Kirby. 2014. "How arbitrary is language?". *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 1–12. <https://doi.org/10.1098/rstb.2013.0299>.
- Post, Brechtje, William D. Marslen-Wilson, Billi Randall, and Lorraine K. Tyler. 2008. "The processing of English regular inflections: Phonological cues to morphological structure". *Cognition* 109: 1-17. <https://doi.org/10.1016/j.cognition.2008.06.011>.

- Pierrehumbert, Janet. 2016. "Phonological Representation: Beyond Abstract Versus Episodic." *Annual Review of Linguistics* 2: 33–52.
- Reali, Florencia, and Thomas L Griffiths. 2009. "The Evolution of Frequency Distributions: Relating Regularization to Inductive Biases through Iterated Learning." *Cognition* 111 (3): 317–28. <https://doi.org/10.1016/j.cognition.2009.02.012>.
- Ringe, Don. 2003. "Internal reconstruction". In *The handbook of historical linguistics*, edited by Brian D. Joseph and Richard D. Janda, 244–261. Malden, MA: Blackwell Pub.
- Tamariz, Monica. 2004. *Exploring the adaptive structure of the mental lexicon*. University of Edinburgh PhD thesis. <http://www.isrl.uiuc.edu/~amag/langev/paper/tamariz05phd.html>.
- Tamariz, Monica. 2008. "Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon". *The Mental Lexicon* 3(2): 259–278. <https://doi.org/10.1075/ml.3.2.05tam>.
- Vitevitch, Michael S. 2005. "Increases in Phonotactic Probability Facilitate Spoken Nonword Repetition." *Journal of Memory and Language* 52 (2): 193–204.
- Wedel, Andrew. 2006. "Exemplar Models, Evolution and Language Change." *The Linguistic Review* 23 (3): 247–74.
- Wełna, Jerzy. 2009. "The post-sonorant devoicing of [d] in the past/past participle forms of weak verbs (sent, spend, etc.)" In *Pe laurer of oure Englische tonge*, edited by Marcin Krygier and Liliana Sikorska, 21–34. Frankfurt: Peter Lang.

Ambiguity avoidance and DOM

Ilja Seržant (University of Potsdam)

Since a transitive clause has two arguments (A and P), it must be ensured that the hearer will be able to discern which of the arguments should be interpreted as A and P, respectively. Moreover, other potential misinterpretations, such as one NP modifying the other NP – if both are adjacent to each other – or both NPs being coordinated (without a conjunction), should be excluded. There are many ways in which ambiguity avoidance may be implemented in a particular language or even in a particular sentence, with flagging being one of them:

(1) Ambiguity avoidance of P flagging (economy subsumed)

In a transitive clause, the A and the P argument must be sufficiently disambiguated, e.g. by word order, agreement, voice, world knowledge, and it is only if they are not that there is dedicated P flagging.

A number of researchers have argued that there is only little or no evidence for (Aor P) flagging systems being driven by ambiguity avoidance as defined in (1) cross-linguistically (*inter alia*, Aissen 2003; Malchukov 2008; various papers in de Hoop & de Swart 2008). Levshina (2021) shows on the basis of the large-scale AUTOTYP database that there is no statistically significant effect of ambiguity avoidance observable for flagging because there are only very few languages in which flagging is primarily driven by ambiguity avoidance. Sometimes even in these languages, ambiguity avoidance does not serve the purpose of ambiguity avoidance between A and P alone: a function inherited from the source construction and often some ongoing conventionalization of the most frequent ambiguity avoidance patterns override the discriminatory function to various extents. Having said this, it has been repeatedly suggested that flagging might also serve the ambiguity avoidance, especially if A and P have similarly ranked input (cf., *inter alia*, Comrie 1978, 1989; Dixon 1994; Silverstein 1976; Kibrik 1997). Bossong (1985: 117) even assumed that the emergence of DOM is primarily due to ambiguity avoidance.

In this paper, I will provide qualitative evidence for the claim that ambiguity avoidance does operate across genealogically and areally diverse DOM systems. At the same time, I will also argue that its impact is mostly weakened by other competing processes to which it is subordinate, the effect being that there is only marginal evidence for it in the synchronic distribution.

References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3). 435–483. DOI:10.1023/A:1024109008573
- Bossong, Georg. 1985. Markierung von Aktantenfunktionen im Guaraní. In Frans Plank (ed.), *Relational typology*, 1–29. Berlin, New York: Mouton de Gruyter. DOI:10.1515/9783110848731.1
- Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*, 329–394. Austin: University of Texas Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. 2nd edn. Chicago: University of Chicago Press.
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press. DOI:10.1017/CBO9780511611896
- de Hoop, Helen & Peter de Swart (eds.). 2009. *Differential subject marking*. Dordrecht: Springer. DOI:10.1007/978-1-4020-6497-5
- Kibrik, Aleksandr E. 1997. Beyond subject and object: Toward a comprehensive relational typology. *Linguistic Typology* 1(3). 279–346. DOI:10.1515/lity. 1997.1.3.279
- Levshina, N. 2021. Communicative efficiency and differential case marking: Areverse-

- engineering approach. *Linguistics Vanguard*, 7(s3): 20190087. doi:10.1515/lingvan-2019-0087.
- Malchukov, Andrej L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118(2). 203–221. DOI:10.1016/j.lingua.2007.02.005
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Ambiguity avoidance vs. expectation sensitivity as functional factors in language change and language structures: Beyond argument marking

Martin Haspelmath (University of Leipzig)

There is a long tradition of invoking ambiguity avoidance as a functional factor in explaining the rise of differential argument marking (e.g. Caldwell (1856: 271), who suggested that special accusative marking in Dravidian is employed “in order to avoid misapprehension”). But more recently, some authors have contrasted anti-ambiguity as a motivating factor with “predictability-based marking” or “expectation sensitivity” (e.g. Haspelmath 2019: §8; Tal et al. 2022: §1.2; see also Zehentner 2022 for discussion).

In this presentation, I will revisit the debate, also making reference to Grice’s “Avoid ambiguity” maxim and recent psycholinguistic perspectives such as Wasow (2015), as well as the recent typological perspective of Seržant (2019). My critique of the anti-ambiguity explanation will start out from a discussion of the concepts of ambiguity, polysemy, and indeterminacy (= vagueness), which are not often kept apart clearly. Especially in (lexical) semantic-map research (e.g. Georgakopoulos & Polis 2021), “polysemy” (which should be the same as ambiguity) is often conflated with indeterminacy. But indeterminacy is of course rampant in language structures, and it could not be otherwise because there is no way to specify every aspect of meaning that might conceivably be interesting.

On the empirical side, I will extend the discussion of diachronic motivations and pathways from argument marking to other kinds of differential coding, such as alienable vs. inalienable contrasts (e.g. Koptjevskaja-Tamm 1996), independent vs. dependent possessor forms (e.g. Michaelis 2019), causative vs. anticausative marking (e.g. Haspelmath 2016; Inglese 2022), and plurative vs. singulative marking (e.g. Grimm 2018). I will argue that in all these systematic differential-coding situations, expectation-sensitivity provides a good explanation of the typological patterns and their diachronic motivations, while ambiguity avoidance is often irrelevant. This is a very indirect argument in favour of anti-ambiguity explanations, but since the understanding of diachronic change typically relies on indirect inferences, these considerations seem highly relevant to the broader picture.

References

- Caldwell, Robert. 1856. *A comparative grammar of the Dravidian or South-Indian family of languages*. London: Harrison.
- Georgakopoulos, Thanasis & Polis, Stéphane. 2021. Lexical diachronic semantic maps: Mapping the evolution of time-related lexemes. *Journal of Historical Linguistics* 11(3). 367–420. (doi:10.1075/jhl.19018.geo)
- Grimm, Scott. 2018. Grammatical number and the scale of individuation. *Language* 94(3). 527–574. (doi:10.1353/lan.2018.0035)
- Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2). 33–63. (doi:10.1515/linpo-2016-0009)
- Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF – Language Typology and Universals* 72(3). 313–334.
- Inglese, Guglielmo. 2022. Cross-linguistic sources of anticausative markers. *Linguistic Typology at the Crossroads* 2(2). 127–186. (doi:10.6092/issn.2785-0943/14224)
- Koptjevskaja-Tamm, Maria. 1996. Possessive noun phrases in Maltese: Alienability, iconicity and grammaticalization. *Rivista di Linguistica* 8(1). 245–274.
- Michaelis, Susanne Maria. 2019. Support from creole languages for functional adaptation in grammar: Dependent and independent possessive person-forms. In Schmidtke-Bode, Karsten & Levshina, Natalia & Michaelis, Susanne Maria & Seržant, Ilja A. (eds.), *Explanation in*

- typology*, 179–201. Berlin: Language Science Press. (<http://langsci-press.org/catalog/book/220>)
- Seržant, Ilja A. 2019. Weak universal forces: The discriminatory function of case in differential object marking systems. In Schmidtke-Bode, Karsten & Levshina, Natalia & Michaelis, Susanne Maria & Seržant, Ilja A. (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 149–178. Berlin: Language Science Press. (<http://langsci-press.org/catalog/book/220>)
- Tal, Shira & Smith, Kenny & Culbertson, Jennifer & Grossman, Eitan & Arnon, Inbal. 2022. The Impact of Information structure on the emergence of differential object marking: An experimental study. *Cognitive Science* 46(3). e13119. (doi:10.1111/cogs.13119)
- Wasow, Thomas. 2015. Ambiguity avoidance is overrated. In Winkler, Susanne (ed.), *Ambiguity: Language and communication*, 21–51. Berlin: De Gruyter. (doi:10.1515/9783110403589) (<https://www.degruyter.com/view/product/448294>)
- Zehentner, Eva. 2022. Ambiguity avoidance as a factor in the rise of the English dative alternation. *Cognitive Linguistics* 33(1). 3–33. (doi:10.1515/cog-2021-0018)

Workshop: Conceptual metaphors in a comparative and diachronic perspective

Convenors

Daniel Kölligan, Julius-Maximilians-Universität Würzburg, Institut für Altertumswissenschaften, Lehrstuhl für Vergleichende Sprachwissenschaft, Oswald-Külpe-Weg 84, D-97074 Würzburg, email: daniel.koelligan@uni-wuerzburg.de (corresponding author)

Lucien van Beek, University of Leiden, Faculty of Humanities, Leiden University Centre for Linguistics, Reuvensplaats 3-4, 2311 BE Leiden, The Netherlands, email: l.c.van.beek@hum.leidenuniv.nl

Keywords

cognitivism, conceptual metaphor, diachrony, etymology

Abstract

In cognitive linguistics, the term "conceptual metaphor", or "cognitive metaphor", refers to the understanding of one idea, or conceptual domain, in terms of another. The source domain is usually more concrete, relating to basic human experience and perception, the target domain is usually more abstract, e.g. time conceptualized in terms of space ("the days ahead of us" like "the road ahead of us"). The research paradigm took off with Lakoff & Johnson (1980), who investigated the persistent use of metaphorical language in all areas of human experience, such as love conceptualized as a journey.

Among the cases studied most in cognitive linguistics literature is the concept of anger conceptualized as a hot fluid in a container, e.g., "You make my blood boil", "He's just letting off steam" (cf. e.g. Lakoff 1987: 380ff., Kövecses 1986, Kövecses 1998, Stefanowitsch 2006:92, etc.). This specific metaphor has been claimed not to be a universal based on the general physiology of humans, but rather a historically contingent feature of languages and cultures influenced by the Ancient Greek Hippocratic theory of humors that was further developed in early modern Europe (Geeraerts & Grondelaers 1995). While the basic point that the history of ideas must be taken into account in cognitive studies is well made (as e.g. in studies like Brock 2013 on Greek political imagery), Geeraerts and Grondelaers did not discuss similar metaphors in non-European languages such as Sanskrit and in pre-Hippocratic European traditions, e.g. Classical Greek and Latin. Indeed, early evidence for anger as a pressurized fluid may be found, for instance in the etymology of Latin *furor* 'anger' (Kölligan 2020).

Conceptual metaphors have now been applied to (and described for) most languages spoken today, and also to some for the ancient and medieval languages (e.g., Cairns 2016; Forte 2018; Horn 2016; Zanker 2019 for Greek; Short 2013 for Latin; Izdebska 2016 on Old English). At the same time, attempts are being made to compare metaphors from different Indo-European languages, and to re-construct specific metaphors for Proto-Indo-European (e.g., van Beek 2017 on metaphors for 'law' and 'justice', Bartolotta 2018 on the deixis of past and future events based either on absolute positioning or one relative to ego, Johnson 2019 et al. on metaphors for 'succeed, be successful' based on the notion of a motion forward as in Lat. *mihi succedit* etc.). The analysis of conceptual metaphors may also support or undermine specific etymological reconstructions (Kölligan 2022).

The workshop invites papers discussing the applicability of conceptual metaphor theory to historical language data, asking what is universal and what is historically contingent, whether and how conceptual metaphors may help us in judging etymologies, and inviting cross-linguistic and diachronic comparisons.

References

- Bartolotta, Annamaria (2018). Spatio-temporal deixis and cognitive models in early Indo-European. *Cognitive Linguistics* 29, 1–44.
- Beek, Lucien van (2017). Die Bildersprache des Rechts im Indogermanischen: Griechisch ἰθεῖα δίκη und δίκην βλάπτειν. In H. Bichlmeier and A. Opfermann (eds), *Das Menschenbild bei den Indogermanen. Akten der Arbeitstagung der Indogermanischen Gesellschaft in Halle (31. März – 2. April 2011)*, 129–150. Hamburg: Baar.
- Brock, Roger (2013). *Greek Political Imagery from Homer to Aristotle*. London & New York: Bloomsbury Academic.
- Cairns, Douglas (2016). Clothed in Shamelessness, Shrouded in Grief: The Role of “Garment” Metaphors in Ancient Greek Concepts of Emotion. In Fanfani, G., M. Harlow, M.-L. Nosch (eds), *Spinning Fates and the Song of the Loom: The Use of Textiles, Clothing and Cloth Production as Metaphor, Symbol and Narrative Device in Greek and Latin Literature*, 25–41. Oxford: Oxbow.
- Forte, Alexander (2018). The Cognitive Linguistics of Homeric Surprise. In: P. Meineck et al. (eds), *The Routledge Handbook of Classics and Cognitive Theory*, 39–58. London/New York.
- Geeraerts, Dirk & Stefan Grondelaers (1995). Looking back at anger: Cultural traditions and metaphorical patterns. In John R. Taylor & Robert E. MacLaury (eds.), *Language and the Cognitive Construal of the World*, 153-179. Berlin, New York: De Gruyter Mouton.
- Horn, Fabian (2016). ‘Building in the Deep’: Notes on a Metaphor for Mental Activity and the Metaphorical Concept of Mind in Early Greek Epic. *Greece and Rome* 63, 163–174.
- Izdebska, Daria (2016). Metaphors of weapons and armour through time. In: Anderson, W. et al. (eds.), *Mapping English Metaphor Through Time*, 221-242. Oxford: OUP.
- Johnson, Cynthia A., Peter Alexander Kerkhof, Leonid Kulikov, Esther Le Mair, und Jóhanna Barðdal. „Argument Structure, Conceptual Metaphor and Semantic Change: How to Succeed in Indo-European without Really Trying“. *Diachronica* 36, Nr. 4 (2019): 463–508. <https://doi.org/10.1075/dia.00014.bar>.
- Kölligan, Daniel (2020). „Seething anger: Latin furor“. In: L. Repanšek, H. Bichlmeier, and V. Sadovski (eds.), *vácāmsi miśrā kṛṇavāmahai*. Proceedings of the international conference of the Society for Indo-European Studies and IWoBA XII, Ljubljana 4–7 June 2019, 397–412. Hamburg: Baar.
- Kölligan, Daniel (2022). “Conceptual metaphors and historical linguistics – some case studies”. Talk at the Università Cattolica del Sacro Cuore, Milan. March 22, 2022.
- Kövecses, Zoltán (1986). Metaphors of anger, pride and love: a lexical approach to the structure of concepts. *Pragmatics & beyond*, 7,8. Amsterdam. Benjamins.
- Kövecses, Zoltán. „Are there any emotion-specific metaphors?“ In *Speaking of Emotions. Conceptualization and Expression*, ed. by Angeliki Athanasiadou & Elzbieta Tabaskowska, 127–51. Berlin / New York: Mouton de Gruyter.
- Lakoff, George and Mark Johnson (1980). *Metaphors we Live by*. Chicago: University of Chicago Press.
- Lakoff, George (1987). *Woman, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.

Marcinkowska-Rosół, Maria and Sellmer, Sven (2021). *The Mind as Container: A Study of a Metaphor in Homer and Hesiod with a Parallel Analysis of the Sanskrit Epics*. Mnemosyne pre-publication, <https://doi.org/10.1163/1568525X-bja10076> (last visited 2022-09-19).

Short, William (2013). 'Transmission' Accomplished? Latin's Alimentary Metaphors of Communication. *American Journal of Philology* 134, 247–275.

Stefanowitsch, Anatol (2006). "Words and their metaphors: A corpus-based approach". In Anatol Stefanowitsch and S. T. Gries (eds.), *Corpus-based approaches to metaphor and metonymy*, 61–105. Berlin, New York: Mouton de Gruyter.

Zanker, Andreas T. (2019). *Metaphor in Homer: Time, Speech and Thought*. Cambridge: CUP.

Accepted speakers (including the convenors)

1. van Beek, Lucien
2. Bartolotta, Annamaria
3. Ginevra, Riccardo
4. Pompeo, Flavia
5. Roth, Theresa
6. Zampetta, Silvia *et al.*
7. Kölligan, Daniel

Abstracts

Lucien van Beek

Universiteit Leiden, Leiden University Centre for Linguistics

L.van.Beek@hum.leidenuniv.nl

Clouds or Arrows? Conceptual Metaphors and the Etymology of Homeric Greek *kertoméō* ‘to mock; taunt’

The description of metaphors in Homer has recently been given a new impulse with, among other things, the work of Cairns (e.g. 2016) and the publication of Zanker (2019), who discusses conceptual metaphors for Time, Speech and Thought. In this paper I will illustrate how our improved synchronic understanding of conceptual metaphors in Homeric Greek may help us find and judge etymologies, using the verb κερτομέω ‘to taunt; mock’ as a case in point.

The etymology of κερτομέω, a verb referring to a speech act (on the lexical meaning cf., *inter alia*, Hooker 1986; Clarke 2001; Lloyd 2004), has been the object of a continued scholarly debate. However, a definitive conclusion still hasn’t been reached, and etymological dictionaries remain cautious (*GEW*: “expressives Wort strittiger Herkunft”; *DELG*: “mot expressif sans étymologie établie”; *EDG* claims Pre-Greek origin). Almost all previous proposals assume that the second part -τομέω has the same root as τέμνω ‘to cut’. As for the first part κερ-, Jones (1989) and Clarke (2001) proposed that it reflects κῆρ ‘heart’, an idea that was current already in the ancient lexicographical tradition. Against this, Perpillou (1986) and García Ramón (2007) derive κερ- from a verbal root PIE *ker- ‘to cut’, following an older proposal by Prellwitz.

In my paper, I will first discuss the problems with these existing etymological analyses. After that, I suggest two new ways to make sense of κερτομέω by analyzing it as a verb phrase reflecting a conceptual metaphor:

- (1) κῆρ τετμεῖν ‘to reach the heart’; I will argue that this could reflect WORDS ARE ARROWS (Durante 1958; recently Zanker 2019: 125-131);
- (2) κῆρ *τομεῖν ‘to obscure/cloud the heart’: -τομέω would be an inherited PIE causative *tomH-éje/o- ‘to cloud, cover with darkness’ from the verbal root *temH- ‘get dark’. This phrase would reflect the metaphor GRIEF IS A COVER (Cairns 2016).

Next, I will show how Homeric poets may have consciously employed the above-mentioned metaphors in some contexts where κερτομέω or a related word is used. Finally, I will consider potential evidence for conceptual metaphors involving descendants of *temH- ‘get dark’ in other ancient Indo-European languages. On this basis, I will make a choice between the two possible reconstructions.

Literature

- Cairns, Douglas. 2016. "Clothed in shamelessness, shrouded in grief: the role of "garment" metaphors in Ancient Greek concepts of emotion", in G. Fanfani, M. Harlow and M.-L. Nosch (eds.), *Spinning Fates and the Song of the Loom: The Use of Textiles, Clothing and Cloth Production as Metaphor, Symbol, and Narrative*, 25–41. Oxford.
- Clarke, Michael. 2001. "'Heart-cutting talk': Homeric κερτομέω and related words." *CQ* 51, 329-338.
- DELG : Chantraine, P. 1968–1980. *Dictionnaire étymologique de la langue grecque: histoire des mots*. Paris.
- Durante, Marcello. 1958. "Epea pteroenta. La parola come 'cammino' in immagini greche e vediche." *Rendiconti. Classe di Scienze morali, storiche e filologiche* 13, 3-14.
- EDG : Beekes, R.S.P. 2010. *Etymological Dictionary of Greek* (2 vols). Leiden.
- García Ramón, José-Luis. 2007. "Atilatein *cortumiō* 'Geländeausschnitt', Idg. **kr-tomh*₁-ó- *(Schnitt) schneidend', *contemnō* 'schmähe' und griechisch κέρτομος 'schmähend', κερτομέω 'schmähe'." *Aevum antiquum* 7 (NS), 285-298.
- GEW : Frisk, H. 1960–1972. *Griechisches etymologisches Wörterbuch* (3 vols). Heidelberg.
- Hooker, J.T. 1986. "A Residual Problem in *Iliad* 24." *Classical Quarterly* 36, 32–37.
- Jones, P.V. 1989. "*Iliad* 24.649: another solution." *CQ* 39, 247-250.
- Lloyd, Michael. 2004. "The Politeness of Achilles: Off-Record Conversation Strategies in Homer and the Meaning of *kertomia*." *Journal of Hellenic Studies* 124, 75-89.
- Perpillou, Jean-Louis. 1986. "De « couper » à « insulter »." In: A. Etter (ed.), *O-o-pe-ro-si. Festschrift für Ernst Risch zum 70. Geburtstag*, 72-84. Berlin.
- Zanker, Andreas. 2019. *Metaphor in Homer: Time, Speech and Thought*. Cambridge.

Annamaria Bartolotta, Università di Palermo

annamaria.bartolotta@unipa.it

The right-left conceptual mapping in a comparative and diachronic perspective

This paper investigates the right-left conceptualization of space in ancient Indo-European languages. In a crosslinguistic perspective, RIGHT and LEFT terms can be recruited to designate cardinal directions (Hertz 1909: 567; Lloyd 1962: 59; Brown 1983: 136). These terms turn out to be associated respectively to *east* and *west* in languages such as Vedic Sanskrit, Hittite, and Homeric Greek. However, the interpretation of such metaphorical mapping from the source domain to the target domain is still an open question. This is also due to some unresolved inconsistencies between etymology and semantic developments emerged in the reconstruction of the Indo-European roots of these terms since the earlier studies of Grimm. The German linguist ascribed the origin of the spatial uses of RIGHT and LEFT to the orientation of the observer's body (1848: 981). The question is further complicated by the unclear origin of linguistic metaphors for positive and negative valence, through an associative mapping from the concrete right-left space to the abstract emotional concepts of 'goodness' and 'badness'. The mental spatial schema is indeed activated to represent such concepts by means of the well-known Good is Right and Left is Bad conceptual mapping (cf. Casasanto 2009; 2014). From a strictly linguistic perspective, a strong asymmetry has been observed between RIGHT and LEFT terms. More specifically, while the RIGHT terms of most Indo-European languages derive from one and the same root **deks-* (Walde 1930: 784; Pokorny 1959: 190), the LEFT terms cannot be traced back to one common ancestor (cf. Buck 1949: 865). Traditionally, such an asymmetry has been ascribed to cultural conventions (cf. Van Leeuwen-Turnovcová 1990), which, however, would ultimately reflect the original embodied asymmetry within the hand domain (cf. Meillet 1906 [1982]: 290; Cuillandre 1947; Heesterman 1959: 256; Giannakis 2019: 256-257). Yet, from an etymological perspective, it has been shown how the words for RIGHT and LEFT derive from lexical roots that are not primarily related to the sides of the body (cf. Foolen 2019: 145), thus challenging an embodied origin of these mental metaphors. Now, recent studies on Indo-European spatial Frames of Reference (FoRs) have revealed that RIGHT and LEFT terms could be used within an absolute or geocentric FoR (Bartolotta 2022). Such results might shed light on the transfer pattern from the concrete domain of spatial regions to the abstract domain of right-left dimensions. Indeed, although it is widely assumed that the human body is the main source domain for the linguistic conceptualization of the entire domain of spatial relations, and that, accordingly, hands are the conceptual source for RIGHT and LEFT polarity (Heine 1997: 49; cf. Bickel 1994: 32), the analysis of the data from a comparative and diachronic perspective seems to suggest a different path of this conceptual metaphor. More specifically, the textual analysis of the RigVeda and the Homeric poems, aside from supporting pieces of evidence derived from Hittite oracle and ritual texts (cf. Ünal 1978; Puhvel 1983; Sakuma 2009) and the Umbrian Tabulae Iguvinae (Prosdocimi 1979; 2015; Untermann 2000: 475), suggests that the extension to hands is the result of a conceptual metaphor which goes from cosmogony (involving the concrete movements of the sun) to the body (cf. Kuiper 1970: 128; Gonda 1972: 8; Abrams & Primack 2001: 1769), thus proving that the metaphoric mapping between body-parts and other domains is not unidirectional (cf. Sinha & Jensen de López 2000: 24; Yu 2008: 408).

References

- Abrams, N. E., Primack J. R. (2001), *Cosmology and 21st-century Culture*, «Science» 293 (5536): 1769-1770.
- Bartolotta, A. (2022), *Spatial Cognition and Frames of Reference in Indo-European*, in D. Romagno, F. Rovai, M. Bianconi, M. Capano (eds.), *Variation, Contact, and Reconstruction in the Ancient Indo-European Languages. Between Linguistics and Philology*, Brill, Leiden/Boston, pp. 179-209.
- Bickel, B. (1994), *Mapping Operations in Spatial Deixis and the Typology of Reference Frames*, Working Paper n. 31, Cognitive Anthropology Research Group, Max Planck Institute for Psycholinguistics, Nijmegen.
- Brown, C.H. (1983), *Where Do Cardinal Direction Terms Come From?*, «Anthropological Linguistics» 25 (2): 121-161.
- Buck, C.D. (1949), *A dictionary of selected synonyms in the principal Indo-European languages*, University of Chicago Press, Chicago/London.
- Casasanto, D. (2009), *Embodiment of Abstract Concepts: Good and Bad in Right-and Left-Handers*, «Journal of Experimental Psychology: General» 138 (3): 351-367.
- Casasanto, D. (2014), *Experiential origins of mental metaphors: Language, culture, and the body*, in M. J. Landau, M. D. Robinson, B. P. Meier (Eds.), *The power of metaphor: Examining its influence on social life* (Vol. 155), American Psychological Association, Washington, DC.
- Cuillandre, J. (1943), *La droite et la gauche dans les poèmes homériques en concordance avec la doctrine pythagoricienne et la tradition celtique*, Imprimeries Réunies, Rennes.
- Foolen, A. (2019), *The value of left and right*, in J. L. Mackenzie, L. Alba-Juez (Eds.), *Emotion in Discourse*, John Benjamins, Amsterdam/Philadelphia, pp. 139-158.
- Giannakis, G. K. (2019), *The east/west and right/left dualism and the rise of some taboos in ancient Greek language and culture*, in G.K. Giannakis, C. Charalambakis, F. Montanari, A. Rengakos (Eds.), *Studies in Greek Lexicography*, Walter de Gruyter, Berlin/New York, pp. 233-262.
- Gonda, J. (1972), *The Significance of the Right Hand and the Right Side in Vedic Ritual*, «Religion» 2 (1): 1-23.
- Grimm, J. (1848), *Geschichte der Deutschen Sprache*, Zweiter Band, in der Weidmannschen Buchhandlung, Leipzig.
- Heine, B. (1997), *Cognitive Foundations of Grammar*, Oxford University Press, New York/Oxford.
- Hertz, R. (1909), *La prééminence de la main droite: Étude sur la polarité religieuse*, «Revue Philosophique de la France et de l'Étranger» 68: 553-580.
- Heesterman, J.C. (1959), *Reflections on the significance of the "Dákṣiṇā"*, «Indo-Iranian Journal» 3 (4): 241-258.
- Kuiper, F.B.J. (1970), *Cosmogony and Conception: A Query*, «History of Religions» 10 (2): 91-138.
- van Leeuwen-Turnovcová, J. (1990), *Rechts und Links in Europa: ein Beitrag zur Semantik und Symbolik der Geschlechterpolarität*, Otto Harrassowitz, Wiesbaden.
- Lloyd, G.E.R. (1962), *Right and Left in Greek Philosophy*, «The Journal of Hellenic Studies» 82: 56-66.
- Meillet, A. (1906 [1982]), *Quelques hypothèses sur des interdictions de vocabulaire dans les langues indo-européennes*, in A. Meillet (Ed.), *Linguistique historique et linguistique générale*, Paris, Champion, pp. 281-291.
- Pokorny, J. (1959), *Indogermanisches Etymologisches Wörterbuch*, I, Francke Verlag, Bern/München.
- Prodocimi, A. L. (1979), *Umbria: Asisium, Parte III (Note e commenti)*. Rivista di Epigrafia Italica, in «Studi Etruschi» 47: 376-379.
- Prodocimi, A. L. (2015), *Le Tavole Iguvine. Preliminari all'interpretazione. La testualità: Fatti e metodi*, II. Olschki, Firenze.
- Puhvel, J. (1983), *Homeric Questions and Hittite Answers*, «The American Journal of Philology» 104 (3): 217-227.
- Sakuma, Y. (2009), *Hethitische Vogelorkeltexte*, Ph.D. dissertation, Julius Maximilian University, Würzburg.

- Sinha C. & K. Jensen de López (2000), *Language, culture, and the embodiment of spatial cognition*, «Cognitive Linguistics» 11 (1/2): 17-41.
- Ünal, A. (1978), *Ein Orakeltext über die Intrigen am hethitischen Hof*, Carl Winter Universitätsverlag, Heidelberg.
- Untermann, J. (2000), *Wörterbuch des Oskisch-Umbrischen*. Handbuch der Italischen Dialekte. Band III. Carl Winter Universitätsverlag, Heidelberg.
- Walde, A. (1930), *Vergleichendes Wörterbuch der indogermanischen Sprachen. Herausgegeben und bearbeitet von Julius Pokorny*. 1. Band, Walter de Gruyter, Berlin/Leipzig.
- Yu, N. (2008), *The Relationship Between Metaphor, Body and Culture*, in R. M. Frank, R. Dirven, T. Ziemke, E. Bernárdez (Eds.), *Body, Language and Mind. Sociocultural Situatedness* (Vol. 2), Mouton de Gruyter, Berlin, pp. 387-408.

Riccardo Ginevra

Università del Sacro Cuore, Milan / Center for Hellenic Studies, Harvard

riccardo.ginevra@unicatt.it / rginevra@chs.harvard.edu

Indo-European Poetics meets Cognitive Linguistics: an integrated approach to the comparative reconstruction of metaphoric and metonymic expressions

The development of Conceptual Metaphor Theory (cf. Lakoff and Johnson 1980) and Cognitive Linguistics in general (cf. Croft and Cruse 2004), as well as their application to Historical Linguistics (cf. Sweetser 1990:23-48; Winters, Tissari and Allan 2011), have greatly improved our understanding of how figurative expressions like metaphors and metonymies work. Even though cognitive-linguistic notions have been successfully employed in the study of figurative language within single ancient IE traditions (e.g., Latin: Short 2008; 2013; Fedriani 2016; Kölligan 2020; Ancient Greek: Pagán Cánovas 2011; Forte 2019; Zanker 2019; Vedic Sanskrit: Jurewicz 2010), correspondences between traditional formulaic phrases attested in several Indo-European traditions have been traditionally investigated exclusively through the lens of Historical Linguistics and Comparative Indo-European Poetics (on which see, e.g., Watkins 1995 and García Ramón 2021).

Aim of the presentation is to argue that, as proposed in Ginevra (2019, 2021a and 2021b), uniting Comparative Indo-European Poetics and Cognitive Linguistics might not only be possible, but also of great use to both disciplines. To this end, after reconstructing – on the basis of evidence from several IE languages – an inherited system of figurative expressions involving the conceptualization of LIFE and DEATH, this reconstructed system will be interpreted as a reflex of two basic metaphoric and metonymic processes that have long been discussed within Cognitive Linguistics:

- on the one hand, these IE traditional expressions will be shown to instantiate Lakoff's (1993:222–223) “Event Structure Conceptual Metaphor”, according to which STATES are mapped onto LOCATIONS, CHANGES onto MOVEMENTS, and CAUSES onto FORCES;
- on the other hand, this reconstructed system will be argued to reflect a so-called “complex event Idealized Cognitive Model” (Kövecses and Radden 1998:51) of the state TO BE ALIVE, i.e., as an event involving several distinct subevents that are habitually more or less co-present in the life of a human being.

The identification of these two (likely universal) processes of human cognition as the basic principles underlying this formulaic system of Indo-European heritage will be argued to be of fundamental importance for the investigation of further issues of Indo-European etymology and historical semantics, especially if combined with other well established notions of Cognitive Linguistics (e.g., image schemata).

References

- Croft, W., and D. A. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Fedriani, C. 2016. Ontological and Orientational Metaphors in Latin: Evidence from the Semantics of Feelings and Emotions. In W. M. Short (ed.), *Embodiment in Latin Semantics*, 115-139. Amsterdam/Philadelphia: John Benjamins.

- Forte, A. 2019. The Cognitive Linguistics of Homeric Surprise. In: P. Meineck, W. M. Short and J. Devereaux (eds.), *The Routledge Handbook of Classics and Cognitive Theory*, 39-58. New York: Routledge.
- Kölligan, D. 2020. Seething anger: Latin *furor*. In: L. Repanšek, H. Bichlmeier, and V. Sadovski (eds.), *vácāmsi miśráḥ kṛṇāvāmahai. Proceedings of the international conference of the Society for Indo-European Studies and IWoBA XII, Ljubljana 4–7 June 2019*, 397–412. Hamburg: Baar.
- García Ramón, J. L. 2021. Poética, léxico, figuras: fraseología y lengua poética indoeuropea. In: L. Galván (ed.), *Mimesis, acción, ficción: Contextos y consecuencias de la «Poética» de Aristóteles*. Reichenberger.
- Ginevra, R. 2019 (2020). Indo-European Cosmology and Poetics: Cosmic Merisms in Comparative and Cognitive Perspective. *Archivio Glottologico Italiano* 104.1.5–17. Ginevra, R. 2021a. Metaphor, metonymy, and myth: Persephone's death-like journey in the *Homeric Hymn to Demeter* in the light of Greek phraseology, Indo-European poetics, and Cognitive Linguistics. In: I Rizzato et al. (eds.), *Variations on Metaphor*, 181–211. Newcastle upon Tyne: Cambridge Scholars.
- Ginevra, R. 2021b. Reconstructing Indo-European Metaphors and Metonymies: a Cognitive Linguistic Approach to Comparative Poetics. *AIQN-Linguistica* 10.163–181. Jurewicz, J. 2010. *Fire and Cognition in the Ṛgveda*. Warszawa: Elipsa.
- Kövecses, Z., and G. Radden. 1998. Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics* 9.1.37-78.
- Lakoff, G., and M. Johnson. 1980. *Metaphors We Live by*. Chicago: University of Chicago Press. Lakoff, G. 1993. The contemporary theory of metaphor. In: A. Ortony (ed.), *Metaphor and thought*, 202-251. Cambridge: Cambridge University Press.
- Pagán Cánovas, C. 2011. The Genesis of the Arrows of Love: Diachronic Conceptual Integration in Greek Mythology. *American Journal of Philology* 132.553-79.
- Short, W. M. 2008. Thinking places, placing thoughts: Spatial metaphors of mental activity in Roman culture. *I Quaderni del Ramo d'Oro on-line* 1.106-129.
- Short, W. M. 2013. "Transmission" Accomplished? Latin's Alimentary Metaphors of Communication. *American Journal of Philology* 134.247-75.
- Sweetser, E. 1990. *From Etymology to Pragmatics. Metaphorical and cultural aspects of semantic Structure*. Cambridge: Cambridge University Press.
- Watkins, C. 1995. *How to Kill a Dragon: Aspects of Indo-European Poetics*. New York/Oxford: Oxford University Press.
- Winters, M. E., H. Tissari, and K. Allan. 2011. *Historical Cognitive Linguistics*. Berlin/New York: De Gruyter Mouton.
- Zanker, A. T. 2019. *Metaphor in Homer. Time, Speech, and Thought*. Cambridge: Cambridge University Press.

Flavia Pompeo

University of Rome, “La Sapienza”

flavia.pompeo@uniroma1.it

New meanings and old constructions: the conceptualization of ‘fearing’ and ‘protecting’ in Old Persian in comparison with other Indo-Iranian languages

In the ancient Indo-Iranian languages, the argument structures of verbs meaning ‘to fear’ and ‘to protect’ respectively show interesting similarities. In this regard, Old Persian data, despite the scarcity of the evidence, are particularly interesting.

This paper focuses on the Old Persian verb *tars-* ‘to fear, to respect’, drawing on the tools of both comparative-historical and cognitive linguistics. In particular, two issues will be taken into account: 1. the meaning and the etymology of the verb *tars-*; 2. its argument structure. Both aspects can be adequately explained by adopting the perspective of cognitive linguistics and, in particular, by considering metaphorical and metonymic processes.

As far as the first point is concerned, the verb *tars-* is the only emotion verb attested in Old Persian where it means ‘to fear, to respect’ with a strong political connotation. This meaning is the end point of a semantic change that originates from the Indo-European root **tres-* ‘to tremble (with fear)’ (cf. *LIV*² 650-651). This “composite” meaning has been reconstructed on the basis of the plurality of meanings attested in Indo-European languages, among which ‘to be afraid’/‘to fear’ is the most frequent meaning, followed by ‘to tremble, to shake’ and, more rarely, ‘to flee (in fear)’.

In particular, the relationship between the meaning ‘to tremble, to shake’ and that of ‘to be afraid, to fear’ will be considered. The conceptualization underlying the semantic shift from ‘to tremble (with fear)’ to ‘to fear/to be afraid’ – that is, from the more concrete source domain to the abstract target domain – will be explained as an essentially metonymic process, according to Kövecses (1998: 148-149 and further works) and Radden (e.g. Radden 1998, in Athanasiadou and Tabakowska). Furthermore, the analysis will confirm the need to study human emotions according to an “integrated” perspective that takes into account both the biological-cognitive and the socio-cultural aspects.

The second aspect worthy of attention is the construction of *tars-*. Here again, drawing on one of the fundamentals of cognitive linguistics, it will be assumed that the choice of a specific linguistic expression reflects a particular conceptualization of a given event, in other words it is motivated by cognitive factors. In particular, it will be shown that, despite the intervening semantic change, the Old Persian verb *tars-* has retained the original construction with the ablative (and the preposition *hacā* ‘from’), which was common to ancient Indo-Iranian languages (Vedic, Avestan, and Old Persian). Interestingly, in this linguistic group, the construction with the ablative (with or without a preposition) is shared by *verba timendi* and verbs meaning ‘to protect’. In Old Persian the construction is exactly the same for both verbs (see example 1 for Old Persian *tars-* and example 2 for Old Persian *pā-*).

(1) *iyam dahyāuš Pārsa (...) hacā anīyanā naī tṛsati*

‘This country Persia (...) does not fear anybody else’ (DPd 6-7, 11-12).

(2) *utā imām dahyāuṃ Auramazdā pātu hacā haināyā*

‘And may Auramazdā protect this country from the (enemy) army’ (DPd 15-17)

Finally, for the construction shared by *verba timendi* and verbs meaning ‘to protect’, a common meaning will be proposed that includes both the semantic component “cause” (a metaphorical

extension from “origin/source”) and the component “distance”, both typical of the Indo-European ablative. The shared meaning can be formulated as follows: “X trembles with fear/fears in relation to an entity that is a potential source of danger and that must be kept at a distance”.

References

- Kövecses, Zoltán (1988). “Are there any emotion-specific metaphors?”. In: *Speaking of Emotions. Conceptualization and Expression*, ed. by Angeliki Athanasiadou & Elzbieta Tabaskowska, 127-151. Berlin / New York: Mouton de Gruyter.
- Kövecses, Zoltan (1990). *Emotion Concepts*. New York et al.: Springer-Verlag.
- Kövecses, Zoltan (2000). *Metaphor and Emotion. Language, Culture, and Body in Human Feeling*. Cambridge: CUP, Paris: Maison des Sciences de l’Homme.
- LIV² = Rix, Helmut and Kümmel, Martin (2001). *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstambildungen*. Zweite, erweiterte und verbesserte Auflage. Wiesbaden: Reichert.
- Radden, Günter (1998). “The conceptualisation of emotional causality by means of prepositional phrases”. In: *Speaking of Emotions. Conceptualization and Expression*, ed. by Angeliki Athanasiadou & Elzbieta Tabaskowska, 273-294. Berlin / New York: Mouton de Gruyter.

Theresa Roth

Philipps-Universität Marburg

rotht@staff.uni-marburg.de

Etymologies and emotions: Historical linguistics as a key to emotion categories

The problem of researching the history of emotions

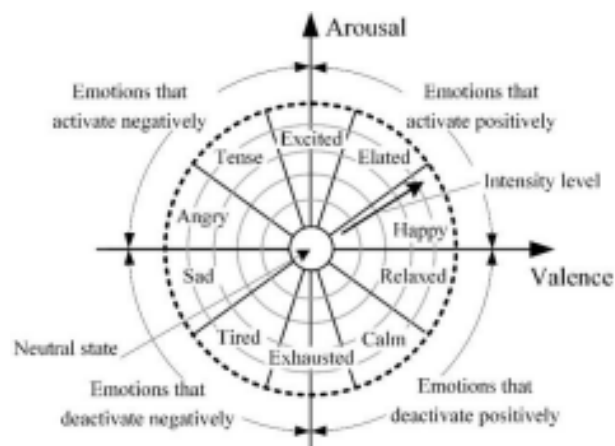
One of the major issues in researching the history of emotions has been the question whether our ancestors did feel completely different or indeed very much the same as we do today. The positions voting for the latter may be labelled, on the one hand, “universalist” – claiming a set of emotions as “basic” according to the allegedly universal evaluation of facial expressions (see Ekman / Friesen 1971) – or, on the other hand, “presentist”, assuming that the “emotion modules” of our brains (e.g., for mate finding or hunting animals) developed in the Paleolithic period and haven’t changed as much since then (cf. Cosmides / Tooby, “Evolutionary Psychology: A Primer.”)

Constructionist theories and the role of language

However, these theories fail to account for the role of concept knowledge in *doing emotion*, that is the fact that we make sense of otherwise unspecific or ambiguous sensations and perceptions only by mapping them on emotion concepts, crucially represented in and acquired by language (for a detailed account on “emotional compounds” see Lindquist *et al.* 2015). This psychological constructionism corresponds with the social constructionist theory which claims emotions of former cultures to be accessible for us mainly via their – culturally shaped and historically embedded – conceptualizations. Researching textual sources of historical communities thus allows “to uncover systems of feeling (...); the emotions that they value, devalue, or ignore; the nature of the affective bonds between people that they recognize; and the modes of emotional expression that they expect, encourage, tolerate, and deplore” (Rosenwein 2010: 11).

Assessing emotional properties

Russel’s circumplex model of emotions arranges emotion concepts according to two main parameters forming the axes of *affective valence* (if a sensation is experienced as positive or negative) and *arousal* (the degree to which an experience or sensation causes neural or physical activity). It might prove to be a helpful tool in assessing historical emotions – but the valence and degree of activation inhering a specific emotion might not always be obvious from the linguistic expression or context.



Russel's circumplex model of emotions

Evaluating the emotion words of a historical community of speakers and especially considering etymologies and conceptual metaphors could, in my view, provide an important key to both of these parameters. As an example, we might have a look at two of the Hittite expressions for “fear” revealing via their etymologies quite opposite degrees of activation (cf. Beckman 2022: 176):

pittuliya-, “to be constricted; anxious, anguished” (CHD P: 366–367; EDHIL 680–681) cf. *pittula-*, “loop, knot” (CHD P: 365–366)

lahlahḫiya-, “to be agitated; to worry” (CHD L–N: 10–12; HED 4: 10–12) cf. *lahlahḫeškenu-* (CHD L–N: 12) with horses as object, “to work them up, cause them to run”

Other aspects which can be “extracted” from conceptual metaphors represented in language (but also enacted in ritual performance) refer to the valence of a sensation, a component prominently featured in conceptual metaphor theory: GOOD = up, present, light, sweet, in order, at rest; BAD = down, absent, dark, disordered, unsettled.

The goal of my paper is thus *not* to discuss how conceptual metaphors might be helpful in finding or judging etymologies but, quite the other way around, to raise the question if and to what extent historical linguistics might provide keys to approach emotional concepts via parameters like valence and activation. Doing so would help us to better delimit several “types” of emotions belonging to one conceptual cluster, to identify their correlations with specific situations or members of the community for whom they are considered appropriate – and to finally create more fine-grained maps of a community’s emotional repertoires.

References

- Barrett, Lisa Feldman / Russell, J. A. 2015. *The psychological construction of emotion*. New York, NY, US: Guilford Press.
- Beckman, Gary. 2022. Hittite emotion terms. In: Sonik, K. / Steinert, U. (eds.): *The Routledge Handbook of Emotions in the Ancient Near East*. London / New York: Routledge, 171–196.
- CHD: *Chicago Hittite Dictionary = The Hittite Dictionary of the Oriental Institute of the University of Chicago*. Güterbock, Hans G./Hoffner, Harry A. Chicago: The Oriental Institute of the University of Chicago. 1980–. Cosmides, Leda / Tooby, John. 1997. *Evolutionary Psychology: A Primer*. <https://www.cep.ucsb.edu/primer.html> Ekman, Paul / Friesen, Wallace V. 1971. Constants across Cultures in the Face and Emotion. In: *Journal of Personality and Social Psychology* 17,2, 124–39.
- Goschke, T. / Dreisbach, G. 2010. Kognitiv-affektive Neurowissenschaft: Emotionale Modulation des Denkens, Erinnerns und Handelns. In: Wittchen U. / J.Hoyer (eds.): *Klinische Psychologie und Psychotherapie*. 2nd edition. Berlin: Springer.
- HED: *Hittite Etymological Dictionary*. Puhvel, Jaan. Berlin/Amsterdam/New-York: Mouton Publishers. 1984–. Kövecses, Zoltán. 1986. *Metaphors of anger, pride and love: a lexical approach to the structure of concepts*. Pragmatics & beyond, 7,8. Amsterdam. Benjamins.
- Lakoff, George / Johnson, Mark. 1980. *Metaphors we Live by*. Chicago: University of Chicago Press.
- Lindquist, Kirsten A. et al. 2015. The role of language in emotion: predictions from psychological constructionism. In: *Frontiers in Psychology* 6, 1–17.

- Rosenwein, Barbara H. 2010. Problems and Methods in the History of Emotions. In: *Passions in Context* I, 1–32.
- Russell, J.A. 1980. A circumplex model of affect. In: *Journal of Personality and Social Psychology* 39, 1161– 1178.
- Schwarz-Friesel, Monika. 2013. *Sprache und Emotion*. 2nd edition. Tübingen / Basel: A. Francke Verlag UTB.

Silvia Zampetta⁺, Erica Biagetti^{*}, Nicolò de Rossi^{*}, Martina Giuliani[#], Chiara Zanchi^{*}, Silvia Luraghi^{*}

^{*}University of Pavia, ⁺University of Bologna, [#]University of Pavia / Bergamo

Chiara Zanchi <chiara.zanchi01@unipv.it>

***Calidum hoc est!* Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin**

This paper focusses on patterns of polysemy of the Sanskrit (Skt.), Ancient Greek (AG), and Latin (Lat.) temperature terms carving up the subdomains of HOT and COLD in these Indo-European (IE) languages. Lexical typology has recently granted much attention to what, since Koptjevskaja-Tamm (2015; but see earlier Plank 2003), has been named ‘linguistics of temperature’. The domain of TEMPERATURE is a good viewpoint to study the link among natural phenomena, human body, and cognition: we experience and evaluate temperature primarily through our bodies. Also, the perception of temperature is scalar and relative: different degrees of heat and coldness can represent good or bad experiences for humans. Since Lakoff/Johnson (1980), metaphorical extensions of temperature terms have been pointed out, mostly discussing the linguistic expression of emotions: positive and negative INTENSE EMOTIONS (e.g. LOVE, HATE and ANGER) are mapped onto HEAT, whereas LACK OF EMOTIONS/CONTROLLED EMOTIONS are conceptualized as COLD/LACK OF HEAT (Lakoff et al. 1991, Radden 2000, Kövecses 1995, Lorenzetti 2009, Coschignano 2021). Other axiologically negative emotions (e.g. DISLIKE, FEAR, and SADNESS) are mapped onto the COLD subdomain (Barcelona 1986, Lakoff et al. 1991, Apresjan 1997, Kövecses 2005, Zhong/Leonardelli 2008). Little research is available on temperature terms in ancient IE languages. What has been done is hardly framed within a cognitive linguistic framework (Fruyt 2013), lies outside the field of linguistics (on *tapas*- ‘heat’ in Skt. literature, see, e.g., Kaelber 1976, 1979) or regards a single conceptual metaphor in a single language (Kölligan 2020). This paper is a first step to fill this gap. By adopting an onomasiological perspective, we will investigate and compare the metaphorical extensions of the terms covering the subdomains of HOT and COLD in Skt., AG, and Lat. with one another and with other not necessarily related languages. Relevant lexical items will be manually extracted from reference dictionaries. Our data will show that metaphorical extensions of temperature terms go beyond the domain of EMOTIONS: e.g., in AG *thermèn epì psukhroîsi kardían êkheis* ‘a hot spirit in a cold business’ (S. Ant. 88), COLD is used to mean USELESSNESS and INEFFECTIVENESS. Moreover, despite stemming from the same bodily metonymy (specifically, THE PHYSIOLOGICAL EFFECTS OF AN EMOTION ARE THE EMOTION ITSELF), temperature terms can develop axiologically opposite shifted meanings. See e.g. the subdomain of HEAT in Lat: in *Reperiamus aliquid calidi conducibilis consili* ‘Let’s find a useful nice and warm plan!’ (Pl. Epid 256), *calidus* instantiates the metaphor HOT IS PRODUCTIVITY, whereas in *Calidum hoc est! etsi procul abest, urit male* ‘This is a burning matter! Though it is far away, it terribly smells like burning’ (Most. 609a), *calidum* and *urit* show a metaphorical shift based on which HIGH INTEREST/ DANGER are conceptualized as HEAT. As pointed out for other languages (e.g., Ameka 2015 described HEAT as associated to BLACK MAGIC), some semantic extensions of temperature terms are noticeably culture-dependent: this is the case of Skt. *tapas*- ‘heat’ > ‘religious austerity, bodily mortification’, *śītala*- ‘cold’ > ‘free from passion, calm, gentle’. The data of this paper will contribute to enriching the semantic annotation contained in three comparable WordNets for ancient IE languages (Biagetti et al. 2021), which, providing etymological information, will allow scholars to investigate whether Skt., AG, and Lat. cognate words lexicalize comparable arrays of concepts. Containing information on periodization(s) and genre(s), and distinguishing literal and non-literal meanings, WordNets will also allow tracking the development of metaphorical meanings over time and across genres.

References

- Ameka Felix K. 2015. Hard sun, hot weather, skin pain. In M. Koptjevskaja-Tamm (ed.), Apresjan, Valentina Ju. 1997. Emotion Metaphors and Cross-Linguistic Conceptualization of Emotions. *Cuadernos de filología inglesa* 612: 179-195.
- Barcelona Sánchez, Antonio. 1986. On the Concept of Depression in American English: A Cognitive Approach. *Revista Canaria de Estudios Ingleses*, 12, 7-35.
- Biagetti, Erica, Chiara Zanchi and William Michael Short. 2021. Toward the creation of WordNets for ancient Indo-European languages. In *Proceedings of the 11th Global Wordnet Conference, University of South Africa (UNISA)*, 258–266.
- Coschignano, Serena. 2021. The semantic network of temperature: Non-sensory domains accessed with metaphorical extensions of the Italian adjectives caldo and freddo. *Review of cognitive linguistics* 19(1): 232-258.
- Fruyt, Michèle. 2013. Temperature and Cognition in Latin. *Revue de Linguistique Latine du Centre Alfred Ernout. De Lingua Latina* 9, 1-34.
- Kaelber, Walter O. 1976. “Tapas”, Birth, and Spiritual Rebirth in the Veda. *History of Religions* 15(4), 343-386.
- Kaelber, Walter O. 1979. Tapas and Purification in Early Hinduism. *Numen* 26(2), 192-214.
- Koptjevskaja-Tamm, Maria (ed.). 2015. *The linguistics of temperature*. Amsterdam/Philadelphia: Benjamins.
- Kölligan, Daniel. 2020. “Seething anger: Latin furor”. In L. Repanšek, H. Bichlmeier, and V. Sadovski (eds). *vácamsi mišrá kṛṇavāmahai, Proceedings of the international conference of the Society for Indo-European Studies and IWoba XII, Ljubljana 4-7 June, 2019*, 397-412. Hamburg: Baar.
- Kövecses, Zoltán. 1995. Anger: Its language, conceptualization, and physiology in the light of cross-cultural evidence. In John R. Taylor & Robert E. MacLaury (eds.), *Language and the Cognitive Construal of the World*. Berlin: Mouton de Gruyter. 181–196.
- Kövecses, Zoltán. 2005. *Metaphor in Culture: Universality and Variation*. Cambridge: Cambridge University Press.
- Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.
- Lakoff, George, Jane Espenson and Alan Schwartz. 1991. *The Master Metaphor List*. Berkeley: Cognitive Linguistics Group, University of California, Berkeley.
- Lorenzetti, Maria Ivana. 2009. “That girl is hot, her dress is so cool, and I’m just chilling out now”: Emergent metaphorical usages of temperature terms in English and Italian. Paper presented at the *Corpus-Based Approaches to Figurative Language* Colloquium in Liverpool, UK, July 2009.
- Plank, Frans. 2003. Temperature talk: The basics. Paper presented at the *Workshop on Lexical Typology at the ALT conference in Cagliari*, September 2003.
http://ling.unikonstanz.de/pages/home/plank/for_download/unpublished/09_FP_TemperatureBasics_2003.pdf
- Radden, Günter. 2000. How metonymic are metaphors? In Antonio Barcelona (ed.), *Metaphor and Metonymy at the Crossroads. Topics in English Linguistics* 30. Berlin/New York: Mouton de Gruyter. 93-108.
- Zhong, Chen-Bo, and Geoffrey J. Leonardelli 2008. Cold and Lonely Does Social Exclusion Literally Feel Cold? *Psychological science* 19, 838-42.

Daniel Kölligan

Julius-Maximilians-Universität Würzburg, Institut für Altertumswissenschaften,
Lehrstuhl für Vergleichende Sprachwissenschaft, Oswald-Külpe-Weg 84, D-97074 Würzburg

daniel.koelligan@uni-wuerzburg.de

Conceptual metaphors and etymology

With the development of conceptual metaphor (CM) theory within the research paradigm of cognitive linguistics since the 1980s (cf. Lakoff & Johnson 1980) and the attribution of the pervasive character of CMs as permeating all registers of language and not being restricted to poetic language vel sim., their use and development is relevant for language diachrony in general. This applies both to independently recurring CMs found across languages and times, which might be due to general features of human experience (cf. e.g. Cairns 2016 and Kölligan 2017 on ‘nakedness’ and ‘shame’), and to CMs contingent upon culturally and historically unique circumstances (which one might call a “locality constraint”, e.g., the development of political metaphors such as “the state is a body” [“body politic”] and “the state is a ship” in ancient Greece, cf. Brock 2013).

This paper will examine a set of well-known CMs and their application to etymological research mostly in Greek and Latin arguing that

(a) the CM ‘anger is a hot fluid in a container’ may not only explain the case of Lat. *furor* ‘anger’ (cf. Kölligan 2020), but also, e.g., that of Greek *σκυδαίνω/σκύζομαι* ‘be angry’ and *σκύζα* ‘lust, heat’, which have not received a satisfactory explanation so far (cf. Beekes 2010: 1360) and which may be related to the Proto-Indo-European root **skeud-* ‘to impel’, and allow to connect the superficially divergent roots **ǵʰer-* ‘to shine’ (Lith. *žėriù*) and **ǵʰerH-* ‘to be angry’ (Ved. *hṛṇite*);

(b) that the CM ‘the soul is a sea’ occurs not only in Lat. *tranquillus* ‘calm’ (cf. Kölligan 2022), Lat. *aequus, aequor* (cf. *aequo animo* ‘with a calm mind’) and Greek *γαλήνη* ‘calm of the sea/of the soul’, but also in derivatives of PIE **sem-* ‘one (and the same)’ developing the meaning ‘even, calm’ (cf. Mlr. *sám*; Germ. *sanft*);

(c) that the CM ‘success is reaching the end of the path’ (cf. Lakoff 1993: 222) underlies the diachronic development of verbs originally meaning ‘to stretch’, ‘to move (towards)’, etc., denoting (agent-oriented) ability such as Greek *δύναμαι* (cf. Kölligan 2021), Toch. *cämp* and German *gelingen* ‘to succeed’ next to *(ge-)langen* ‘to suffice, to reach’.

References

- Beekes, Robert Stephen Paul. 2010. *Etymological dictionary of Greek*. 2 vols. Leiden: Brill.
- Brock, Roger. 2013. *Greek political imagery from Homer to Aristotle*. London; New York: Bloomsbury.
- Cairns, Douglas. 2016. “Clothed in Shamelessness, Shrouded in Grief: The Role of “Garment” Metaphors in Ancient Greek Concepts of Emotion”. In: Fanfani, G., M. Harlow, M.-L. Nosch (eds.), *Spinning Fates and the Song of the Loom: The Use of Textiles, Clothing and Cloth Production as Metaphor, Symbol and Narrative Device in Greek and Latin Literature*. Oxford: Oxbow, 25–41.
- Horn, Fabian. 2016. “‘Building in the Deep’: Notes on a Metaphor for Mental Activity and the Metaphorical Concept of Mind in Early Greek Epic.” *Greece and Rome* 63, 163–174.
- Kölligan, Daniel. 2017. “PIE *h₂eǵd- ‘to reveal’ and its descendants”. In: David M. Goldstein / Stephanie W. Jamison / Brent Vine (eds.), *Proceedings of the 27th Annual UCLA Indo-European Conference*, Bremen: Hempen, 61–78.
- Kölligan, Daniel. 2020. “Seething anger: Latin furor”. In: L. Repanšek, H. Bichlmeier, and V. Sadovski (eds.), *vácāmsi miśrākṛṇavāmahai*. Proceedings of the international conference of the Society for Indo-European Studies and IWoBA XII, Ljubljana 4–7 June 2019. Hamburg: Baar, 397–412.
- Kölligan, Daniel. 2021. “Getting there? Greek δύναμαι ‘be able’”. In: *Synchrony and Diachrony of Ancient Greek*, ed. by Georgios K. Giannakis, Luz Conti, Jesús de la Villa, and Raquel Fornieles, 151–62. Berlin: De Gruyter. DOI: 10.1515/9783110719192.
- Kölligan, Daniel. 2022. “Full and calm - Notes on Greek ἡσυχος and Latin tranquillus”. *Acta Linguistica Petropolitana*. Vol. 18.1. *Colloquia Classica et Indogermanica VII. Miscellanea in honorem Nikolai N. Kazansky septuagenarii*. pp. 165–186. DOI:10.30842/alp23065737181165186.
- Lakoff, George, & Mark Johnson. 1980. *Metaphors we live by*. Chicago: Univ. of Chicago Press.
- Lakoff, George. 1993. “The contemporary theory of metaphor.” In: A. Ortony (ed.), *Metaphor and thought*. Cambridge: Cambridge University Press, 202–251.

WS 6: Categorizers in diachrony

ICHL 26, Sept. 4-8, University of Heidelberg

Convenors:

Laura Grestenberger, Austrian Academy of Sciences (laura.grestenberger@oeaw.ac.at)

Iris Kamil, University of Edinburgh (iris.kamil@ed.ac.uk)

Viktoria Reiter, University of Vienna (viktoria.reiter@univie.ac.at)

Description:

Although the form, meaning, and ontological status of “categorizing” (“stem-forming”) morphology have received some attention in the typological and theoretical literature on word classes (e.g., Vogel & Comrie 2000, Baker 2003, Knobloch & Schaefer 2005, Panagiotidis 2011), its diachrony remains understudied: It is unclear how and why new categorizers arise historically and what “mechanisms” of change are responsible for the rise of new categorization devices. Do new categorizers arise due to semantic bleaching/grammaticalization (e.g., nominal diminutives > nominalizers), reanalysis of functional heads in the context of decategorial (“secondary”) derivation (nominalizers > verbalizers, e.g., Grestenberger forthcoming), the need for “compensation of phonological reduction” (Haspelmath 1995), or is there no uniform diachronic path that gives rise to these grammatical categories?

The goal of this workshop is to discuss the diachrony of categorizing morphology with the aim of establishing cross-linguistic regularities and generalizations concerning the rise, function, and development of nominal, verbal, and adjectival stem-forming morphology. Examples include the reanalysis of nominalizers as verbalizers, (1), of adjectivizers as verbalizers, (2), or of adjectivizers as participial affixes, (3), but also a variety of phenomena usually classified as “grammaticalization” (e.g., the reanalysis of nominal second compound members as nominal or adjectival suffixes).

1. $n \rightarrow v$: Ancient Greek [*basil-eú*]_{n-s} ‘king’: [[**basil-eú*]_{n-j}]_{v-ō} ‘am/act as king’ → Modern Greek *stóx-os* ‘target’ [[*stox*]_{n-év}]_{v-o} ‘to aim at’; Pre-Proto-Algonquian [**api*]_{v-hm}]_n ‘sitting place, seat’, **net*-[[*api*]_{v-hm}]_{n-ena-n} ‘where we sit; our sitting place’ → Proto-Algonquian **net*-[[*api*]_{v-hm}]_{v7-ena-n} (*ma-hi*) ‘we sit over there’ (Oxford 2014: 14-15)
2. $a \rightarrow v$: Gm. *Kraft* ‘strength’: [[*kräft*]_{n-ig}]_a ‘with strength, strong’; [[[*kräft*]_{n-ig}]_{a-en}]_{v/T[-fin]} ‘to strengthen’ → *Pein* ‘pain’: [[[*pein*]_{n-ig}]_{v-en}]_{T[-fin]} ‘to torture’ (**pein-ig* ‘painful’)
3. $a \rightarrow v/ptcp$: Sanskrit *ásva-* ‘horse’: [[*asv*]_{n-in}]_{a-} ‘possessing horses’ → √*yaj* ‘sacrifice’: [*yāj-in*]_{ptcp-} ‘sacrificing’

The papers in this workshop bring specific predictions from different theoretical approaches to bear on these issues and adduce novel empirical arguments from a variety of different language families to the debate. The contributions will address (and go beyond) the following issues:

- What role do morphological reanalysis and resegmentation, especially mechanisms such as “**affix telescoping**” (Haspelmath 1995) play in the establishment of new categorizers, and what is the role of “phonological erosion” or loss of phonological material in these processes?
- How does categorization interact with morphosyntactic features such as number or classifier morphology and gender (on *n*) or Aktionsart on *v*? Which diachronic generalizations as to these interactions are possible? For example, in Distributed Morphology, roots only receive their categorization in the course of the syntactic derivation by combining with the categorizing heads *v* (verbalizers), *n* (nominalizers), and *a* (adjectivizers or “stativizers”). Categorization is thus fundamentally syntactic, and the extent to which categorizers are also associated with syntactico-semantic “content” such as definiteness (in the nominal domain) or Aktionsart (in the verbal domain) is debated (Panagiotidis et al. 2017). In (broadly) lexicalist approaches, on

the other hand, “stem classes” or “conjugational classes” are treated as properties of words and hence, the lexicon. These approaches also differ in how conjugational class elements such as “theme vowels” are treated both from a synchronic and from a diachronic perspective (cf., e.g., Calabrese 2019, Bertocci & Pinzin 2020), and with respect to the analysis of change in classifier systems and their connection to (noun class) categorization (e.g., Craig 1986).

- Are there unambiguous diagnostics for distinguishing between categorizing morphology and derivational morphology in the more technical sense, that is, category-changing morphology with specific (argument- and event-structure changing) functions, e.g., agent noun- and verbal abstract-forming morphology in the nominal domain or causativizing and applicativizing morphology in the verbal domain? Empirical and conceptual arguments in favor of separating “low” categorizing morphology from “higher” functional, category-changing projections (e.g., Himmelmann 2005, Marantz 1997, Borer 2015; Panagiotidis et al. 2017) have not yet been connected to the diachrony of these entities in a systematic way.
- What role does language acquisition play in the diachronic development of categorizing morphology? For example, syntactic change has been argued to proceed via “**upwards reanalysis**” (Roberts & Roussou 2003) of lexical projections as higher functional projections, and this is compatible with L1 acquisition evidence of how children acquire, for example, epistemic modal verbs by overextending their functional domain “upwards” (Cournane 2014). Does this overextension parallel the changes we see in the historical record? That is, is categorizer change inherently directional?

References

- Baker, M. C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: CUP.
- Bertocci, D. & Pinzin, F. 2020. Two kinds of verbal roots in Latin: Evidence from thematic vowels and word-formation processes. *Lingue antiche e moderne* 9. 23–56.
- Borer, H. 2015. The category of roots. In *The syntax of roots and the roots of syntax*, 112–148. Oxford: OUP.
- Calabrese, A. 2019. Morpho-phonological investigations: A theory of PF. From syntax to phonology in Sanskrit and Italian verbal systems. Ms., University of Connecticut.
- Cournane, A. 2014. In search of L1 evidence for diachronic reanalysis: Mapping modal verbs. *Language Acquisition*, 21(1): 103–117.
- Craig, C. G. 1986. Jacaltec noun classifiers: A study in grammaticalization. *Lingua* 70: 241–284.
- Grestenberger, L. Forthcoming. The diachrony of verbalizers in Indo-European: Where does *v* come from? *Journal of Historical Syntax*.
- Haspelmath, M. 1995. The growth of affixes in morphological reanalysis. *Yearbook of Morphology* 1994, 1–29. Dordrecht: Kluwer.
- Himmelmann, N. 2005. Gram, construction, and class formation. In Knobloch & Schaefer 2005, 79–94.
- Knobloch, C. & B. Schaefer (eds.). 2005. *Wortarten und Grammatikalisierung. Perspektiven in System und Erwerb*, 79–94. Berlin: de Gruyter.
- Marantz, A. 1997. No escape from syntax: Don’t try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2), 201–225.
- Oxford, W. 2014. Microparameters of agreement: A diachronic perspective on Algonquian verb inflection. University of Toronto PhD thesis.
- Panagiotidis, P. 2011. *Categorial features: A generative theory of word class categories*. Cambridge: CUP.
- Panagiotidis, P., V. Spyropoulos & A. Revithiadou. 2017. Little *v* as a categorizing verbal head: Evidence from Greek. In *The verbal domain*, 29–48. Oxford: OUP.
- Roberts, I. & A. Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: CUP.
- Vogel, P. & B. Comrie (eds.). 2000. *Approaches to the Typology of Word Classes*. Berlin: de Gruyter.

Contributions

1. Inflectional vocalic pieces in Latin verbal morphology: a synchronic and diachronic analysis

Andrea Calabrese

University of Connecticut

This paper will look at the historical development of reconstructed VP-shell and actional/aspectual formatives from Proto-Indo-European (PIE) into Latin. Thus, on the one hand, it will look at the outcomes of formatives such as **-eye-* characteristic of causatives, the **-ye* of denominatives, the **-eh₁-* characteristic of statives, and, on the other, at the outcomes of actional/aspectual formatives like **-e*, and **-ye-*. These formatives developed into the Latin root-adjacent vocalic pieces *-ā-*, *-ē-*, *-ě-*, *-ĩ-*, *-ī-*. The pieces *ā-*, *ē-*, *ī-*, developed from VP-shell elements. Thus, the *-ā-* conjugation developed mostly from denominatives in **-ye-* whose bases were the nominal stems of the *-ā-* (<**-eh₂-*) declension: */-ā-/* < **-eh₂-ye* (with loss of the intervocalic glide, subsequent merging of the vowel sequence and eventual reanalysis of the resulting piece as a *v⁰-derivative*): e.g., *curāmus* ‘cure’ (cf. *curā* ‘cure’). The *-ē-* conjugation developed mostly from the stative suffix *-ē-* (<**-eh₁-*) and from causatives in **-eye-* (with *o*-grade of root): */-ē-/* < **-eh₁-*: e.g., *sedēmus* ‘we sit’ (<**sed-eh₁-*; cf. *sīdo*, **si-sd-* ‘I sit down’), */-ē-/* < **-eyē-*, e.g., *monēmus* ‘we warn’ (<**mon-eye-*). The *-ī-* conjugation developed mostly from denominatives in **-ye-*, */-ī-/* < **-denominative *-yē-*, e.g., *fīnīmus* ‘limit’ (cf. *fīnis* ‘end’), but also from original stems in **-ye-*: *venīmus* ‘come’ (<**g^wen-ye-*). *ě-*, *-ī-*. The pieces */-ě-/* and */-ĩ-/* developed from actional/aspectual **-e*, **-ye* (*legimus* < **leg^v-e* ‘collect’; *capio* < **kap-ye-ti* ‘takes’).

I will argue against recent proposals by Bertocci and Pinzin (2020, 2021), who hypothesize that all these elements preserved their functional status in their development from PIE to Latin so that */-ā-/* and */-ī-/* are functional elements in the VP shell whereas */-ě-/* and */-ĩ-/* (as well as */-ē-/* in Bertocci and Pinzin’s analysis) are actional/aspectual markers. In contrast, I will support Aronoff’s (1994) original hypothesis that all root-adjacent vocalic pieces in Latin are simply ornamental elements. I will show how Latin root-adjacent vocalic pieces lost semantic specificity and were bleached in meaning due to their disparate etymological sources; for example, */-ā-/* did not develop only from the denominative sequence **-eh₂-ye* but also from de-adjectival factitive with the suffix **-h₂*: *novare* ‘to renew’ from *novus*, *nova*, *novum* ‘new’, and even possibly from a root-final laryngeal as in the case of primary verbs in */-ā-/*, which do not have a clear etymology: *amāre* ‘to (make) love’, *arāre* ‘to plow’, *volāre* ‘to fly’, *cubāre* ‘lie down’, *flagrāre* ‘to glow’ (note the semantic inhomogeneity of these verbs, which can be transitive, intransitive and also unaccusative). I will propose that this bleaching finally led to a major reanalysis of Latin morphophonology. Inflectional consonantal pieces were reinterpreted as exponents of functional nodes and inflectional vocalic pieces as exponents of ornamental nodes. This will lead to a radical theoretical simplification of Latin verbal morphology.

The analysis of the development of the PIE formatives into Latin will require a detailed investigation of the morphosyntactic structure of the PIE verbal forms and specifically of the PIE VP-shell. The original status and the development of the *v⁰-formatives* will be of crucial importance in the analysis. It will be shown that they don’t need to be phonologically overt. The consequences of this fact will be explored.

References

- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge, MA: MIT Press.
 Bertocci, Davide & Francesco Pinzin. 2020. Two kinds of verbal roots in Latin: Evidence from thematic vowels and word-formation processes. *Lingue antiche e moderne* 9: 23-56.

Bertocci, Davide & Francesco Pinzin. 2021. Morphology in action: some issues in the formation of the Latin perfect. In *Linguisticae Dissertationes. Current Perspectives on Latin Grammar, Lexicon and Pragmatics. Selected Papers from the 20th International Colloquium on Latin Linguistics (Las Palmas de Gran Canaria, Spain, June 17-21, 2019)*, 13-32. Madrid: Ediciones Clásicas.

2. On adjectivalizers in Rig-Vedic Sanskrit

Luca Alfieri

Università Telematica G. Marconi

This talk focuses on “adjectivalizers” in Rig-Vedic Sanskrit. The basic idea is that any study on “categorizers” cannot but set up from a clear definition of the lexical categories of the described language (noun, verb, adjective, etc.). Still, the definition of these categories in RV Sanskrit is far from trivial, especially when it comes to the adjective.

It is well-known that many languages lack adjectives (Dixon & Aikhenvald 2004). However, it is also well-known that the criteria whereby a language is said to “have” or “lack” adjectives are problematic, if not inconsistent (Dryer 1997, Croft 2001: 67ff., Haspelmath 2012). The best proof for the inconsistency comes from the paradox of *inconsistent category assignment* that is, the situation in which a same language is classified as “without” or “with” adjectives by different scholars on the basis of almost the same empirical data. The definition of the adjectival class in Sanskrit perfectly exemplifies the paradox. Indian native grammar ignores the adjective class (Pontillo & Candotti 2011). Traditional European grammars of Sanskrit usually teach that Sanskrit indeed “has” adjectives, but these adjectives are not as sharply distinguished from nouns as Latin adjectives. Speyer (1896), followed by Joshi (1967) and Bhat (1994), claimed that Sanskrit is a language “without” adjectives or “with noun-like adjectives” that is, with adjectives totally merged with nouns. Alfieri claimed that in RV Sanskrit can better be seen as a language “with verb-like adjectives” or with quality concepts merged with verbal roots in the lexicon, since the most typical Quality Predicate is a verbal form (e.g. *módate* ‘is delighted’) or, at least, a derived adjective built on a verbal root and added to an optional copula (e.g. *tapús (asti)* ‘is hot’ < *tap-* ‘heat, become hot’, see Alfieri 2020); and since the most typical Quality Modifier is not a simple adjective, as in Latin; it rather is a derived adjective built on a verbal root of quality or nearly quality meaning (e.g. *śub^hrā-* ‘beautiful’ < *śub^h-* ‘beautify’, see Alfieri 2016, 2021).

The methodology whereby the last conclusion was reached is relevant for our topic. In Alfieri (2016, 2021) a sample of 51 hymns of RV was gathered and all the Quality Modifiers in the sample were collected: on 1003 “adjectives” therein found, 42.6% are deverbal adjectives such as *tapú-* and *śub^hrā-* (see above), 24.8% are compound adjectives (that is, the *bahuvrīhi* type termed by Indian grammarians) such as *hiraṇya-pāṇi-* ‘having gold hands’, 13.7% are prefixed adjectives such as *su-vīra-* ‘heroic’ < *vīrá-* ‘hero’, 9.8% are denominative adjectives such as *pītriya-* ‘paternal’ < *pīṭ-* ‘father’, 7.8% are simple adjectives such as *kṛṣṇá-* ‘black’, and 2.1% are prepositional adjectives such as *paramá-* ‘most distant’ < *pārā* ‘away’. In the talk the corpus in Alfieri (2016, 2020, 2021) is taken up and further elaborated upon, by discussing all the affixes that convert nouns, verbal roots and preposition into adjectives. The aim of the research is: a) to provide a corpus-based description of the different adjectivalizers in RV Sanskrit; b) to show that a typologically informed definition of the adjective class can contribute to our understanding of adjectival-forming morphology in RV Sanskrit and its PIE origin.

References

Alfieri, Luca. 2021. Parts of speech comparative concepts and Indo-European linguistics. In *Linguistic categories, language description and linguistic typology*, 313-366. Amsterdam: Benjamins.

- , 2020. The lexicalization of the adjective as an innovative feature in the Indo-European Family. *Poznań Studies in Contemporary Linguistics* 56(3): 379-412.
- , 2016. The typological definition of the (apparently historical) notion of root. *Archivio Glottologico Italiano*, 102(1): 129-169.
- Bhat, Shankara D.N. 1994. *Adjectival Categories*. Amsterdam: Benjamins.
- Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Dryer, Matthews. 1997. Are grammatical relations universal? In *Essays on Language Function and Language Type dedicated to T. Givón*, 115-144. Amsterdam: Benjamins.
- Dixon, Robert M.W. & Aikhenvald Alexandra Y. (eds.), *Adjective Classes: A Cross-linguistic Typology*, 1-50. Oxford: Oxford University Press.
- Haspelmath, Martin. 2012. How to compare major word-classes across languages. In *Theories of Everything in honor of Edward Keenan* (UCLA Working Papers in Linguistics 17), 109-130.
- Joshi, Shivaram D. 1967. Adjectives and Substantives as a Single Class in the Parts of Speech. *Journal of the University of Poona, Humanities Section* 25: 19-30.
- Pontillo, Tiziana & Candotti, Maria P. 2011. Discriminare tra aggettivo e sostantivo: appunti sulla tradizione pāṇiniana. *Atti del Sodalizio Glottologico Milanese* (n.s.) 6: 66-84.
- Speyer [~ Speijer], Jakob S. 1974 [1896¹]. *Vedische und Sanskrit Syntax*. Graz: Akademische Druck.

3. One or All: The Development of Singulatives to Collectives in Semitic

Rebecca Hasselbach-Andee
University of Chicago

Semitic languages generally have two genders, masculine and feminine. Masculine nouns are unmarked (as in Classical Arabic *ʔibn-* ‘son’) while feminine nouns are marked by either *-t* or its allomorph *-at* (as in Classical Arabic *bint-* ‘daughter’ and *madīnat-* ‘city’). This distinction of gender and gender marking is found in all major branches of Semitic and can be reconstructed to the proto language without difficulty. There is evidence, however, that the feminine marker *-(a)t* did not originate as gender marker in the nominal system of Semitic. As argued in Hasselbach (2014ab), the morpheme *-(a)t* has various other functions in Semitic languages, such as marking abstracts, singulatives, and collectives, to name the most frequent functions (Hasselbach 2014b: 331). In the same article it was suggested based on comparison with other, less frequent Semitic feminine markers, that the original function of the morpheme might have been the marking of singulatives (Hasselbach 2014b: 342) – although the function to mark abstracts must have developed early on in the history of the language family since it is attested in all major branches.

The third function of *-(a)t*, the marking of collectives, seemingly contradicts the proposed reconstruction of the morpheme as originally marking singulatives. In the articles from 2014, it was hypothetically proposed that the use of *-(a)t* with collectives might have arisen through the use of the morpheme with numerals, but at that point there was no satisfactory explanation for this phenomenon. In this talk I would like to reconsider the semantic and syntactic constructions that might have caused the development of a morpheme that marked singulatives into one that can also mark collectives. The marking of collectives clearly seems to be secondary since this function only occurs in West Semitic languages (Semitic has two major branches, East Semitic, which includes Akkadian and Eblaite, and West Semitic, which includes all other Semitic languages). We can also trace a similar development with a less common Semitic feminine marker, *-ay*, which also has the function to mark collectives besides marking feminine gender and abstracts (Hasselbach 2014b: 335).

The methodology used for this investigation will be based on Typology and Historical Linguistics in order to explore the diachronic processes that led to the seemingly contradictory functions

of *-(a)t* in Semitic, and to find potential cross-linguistic parallels. The same morpheme also developed into the marker of the 3rd feminine singular on perfect verbs. The investigation of sources for third person verbal markers and use of these forms might shed additional light on the question.

There is surprisingly little literature on this topic and no detailed explanatory framework that could account for the developments in Semitic. This talk intends to fill this gap in our understanding of the diachronic processes involved in the functional developments of feminine markers, both from a Semitic and cross-linguistic perspective (Corbett 1991), and to provide such an explanatory framework.

References

- Corbett, G.G. 1991. *Gender*. Cambridge: Cambridge University Press.
 Hasselbach, R. 2014a. Agreement and the development of gender in Semitic (pt. I). *ZDMG* 164: 33-64.
 Hasselbach, R. 2014b. Agreement and the development of gender in Semitic (pt. II). *ZDMG* 164: 319-344.

4. ‘Inalienable’ nominalisers across Meto

Tamisha Tan

Harvard University & Nanyang Technological University

1. Overview The Meto dialect continuum (Austronesian: West Timor) displays several characteristics typical of Central/Eastern Indonesian languages (Klamer 2002; Blust 2009), including subject marking, possessor suffixes, and a distinction between alienable and inalienable nouns. This paper investigates an understudied morphosyntactic property of these languages in the form of the idiosyncratically distributed nominal suffixes *-k*, *-ʔ*, and *-n*, which obligatorily occur on certain bound roots to create alienable nouns.

(1) Bound nominal forms across Meto (from Edwards 2021)

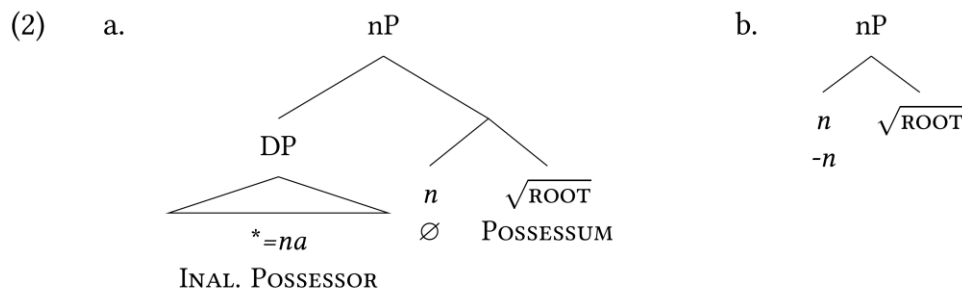
PMP **haRəzan* ‘ladder’ > *era-ʔ*, *era-k* [Amarasi] ‘stairs’, *ela-k* [Molo] ‘ladder’

PMP **rəbaq* ‘collapse’ > *refe-k* [Ro’is Amarasi], *kefa-n* [Kotos Amarasi] ‘ravine, cliff, gap’

PMP **letay* ‘above’ > *k-nete-ʔ* ‘hill’ [Kotos Amarasi], *nete-n* ‘mountain range’ [Molo]

This paper proposes that i) these suffixes originated from the diachronic Spec-Head reanalysis of inalienable possessors into *n* head categorisers; and ii) their innovation facilitated the aggressive resegmentation of etymologically **C#* nouns as the combination of a *V#* root and *-C* nominaliser, which has given rise to the illusion of synchronic subtractive morphology across a number of contexts in the Meto languages.

2. From possessor to *n* Meto inalienable possessor morphology descends from the PMP genitive enclitics (1sg **=ku* > *-k*, 2sg **=mu* > *-m*, 3sg **=na* > *-n*), which originally instantiated pronominal arguments that co-indexed both inalienable and alienable possessors. Following Alexiadou (2003); Ritter & Rosen (2010), I assume inalienable nouns allow the merger of a possessor DP into Spec, nP without needing a mediating PossP, and propose that this specifier was where these enclitics were originally merged as pronominal DPs (2a). Synchronically, several Meto nouns admit both inalienable or alienable possession as determined by semantic context; e.g. *au sisi* ‘my meat (from animals; to eat)’ vs. *au sisi-k* ‘my (own) flesh’ (cf. den Dikken 2015 on Hungarian). Given this variability in usage (and as certain nouns would have obligated 3sg/pl possessors e.g. edges, slopes), I posit that these arguments were grammaticalised into *n* head categorisers (2b) via Spec-Head reanalysis (Simpson & Wu 2002; cf. van Gelderen’s 2004 Head Preference Principle). This change was accompanied by vowel syncope and sporadic consonant reduction (1sg **=ku* > *-k/-ʔ*).



3. ‘Subtractive’ Morphology Edwards (2017, 2020) claims that Meto languages synchronically exhibit C# subtraction as a process which i) derives verbs from nouns (3a) and ii) is obligatory on the first element of nominal compounds (3b). I propose that these phenomena actually involve lexical items which have been reanalysed as roots + an overt *n* head (*-n/k/?*), even where the final consonant is etymological (3a).

- (3) a. PMP **quzan* ‘rain’ > *uran* ‘rain’ → *na-ura* ‘(it) rains’ [Amarasi]
 b. PMP **muntay* ‘citrus tree’ > *muke-ʔ* ‘citrus’ → *muke kase-l* ‘foreign citrus’ [Amfo’an]

This resegmentation predicts the absence of these nominalisers in verbs (i.e. $\sqrt{\text{ura}}$ + *v* ‘to rain’). Further, independent prosodic evidence from metathesis and stress assignment (Mooney 2021; Tan 2021b) and the absence of C# deletion in verbal compounds supports analysing constructions like (3b) as $\sqrt{\text{root}}$ + $\sqrt{\text{root}}$ compounds sharing a single categorising head (Harðarson 2017; Fenger & Harðarson 2019; Tan 2021a) whose allomorphy (here *-l*) is controlled by the final root. That these C# are synchronically segmentable is supported by the presence of *-ʔ* in several nominalising circumfixes in Amarasi (*m(a)- ... -ʔ* and *ʔ- ... -ʔ*), and the productivity of the deverbal nominaliser *-k/?* in the closely-related Rote languages (Tamelan 2021)

References

- Alexiadou, Artemis. 2003. Some notes on the structure of alienable and inalienable possessors. In Martine Coene & Yves D’hulst (eds.), *From NP to DP. Volume 2: The expression of possession in noun phrases*, 167–188. Amsterdam: John Benjamins.
- Blust, Robert. 2009. The position of the languages of eastern Indonesia: A reply to Donohue and Grimes. *Oceanic Linguistics* 36–77.
- den Dikken, Marcel. 2015. The morphosyntax of (in)alienably possessed noun phrases: The Hungarian contribution. In Katalin É. Kiss, Balázs Surányi & Éva Dékány (eds.), *Approaches to Hungarian. Volume 14: Papers from the 2013 Piliscsaba conference*, 121–145. Amsterdam: John Benjamins.
- Edwards, Owen. 2017. *Metathesis and unmetathesis: Parallelism and complementarity in Amarasi, Timor*. Canberra: Australian National University dissertation.
- Edwards, Owen. 2020. *Metathesis and unmetathesis in Amarasi*. Berlin: Language Science Press.
- Edwards, Owen. 2021. *Rote-Meto comparative dictionary*. Canberra: ANU Press.
- Fenger, Paula & Gísli R. Harðarson. 2019. One classy number: Linking morphemes in Dutch and German. In Mariam Asatryan, Yixiao Song & Ayana Whitmal (eds.), *NELS 50: Proceedings of the Fiftieth Annual Meeting of the North East Linguistic Society*, Amherst, MA: GLSA.
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins.
- Harðarson, Gísli R. 2017. *Cycling through grammar: On compounds, noun phrases and domains*. Storrs, CT: University of Connecticut dissertation.
- Klamer, Marian. 2002. Typical features of Austronesian languages in central/eastern Indonesia. *Oceanic Linguistics* 41(2). 363–383.
- Mooney, Kate. 2021. *Metathesis is Deletion and Feature Spreading: Evidence from Uab Meto*. Unpublished ms., New York, NY: New York University.
- Ritter, Elizabeth & Sara Thomas Rosen. 2010. Possessors as external arguments: Evidence from Blackfoot. In J. Randolph Valentine & Monica Macaulay (eds.), *Papers of the 42nd Algonquian Conference*, 211–229.
- Simpson, Andrew & Zoe Wu. 2002. Agreement, shells, and focus. *Language* 287–313.

- Tamelan, Thersia Magdalena. 2021. *A grammar of Dela: an Austronesian language of Rote, eastern Indonesia*: Canberra: Australian National University dissertation.
- Tan, Tamisha L. 2021a. Syntax-Prosody mismatches and the duality of Vedic Devatā Dvandva. In Rachel Soo, Una Y. Chow & Sander Nederveen (eds.), *Proceedings of WCCFL 38*, 427–438. Somerville, MA: Cascadilla Proceedings Project.
- Tan, Tamisha L. 2021b. Three ways to serialise verbs in Amarasi. Talk given at the *Workshop on Field Work, Language Documentation, and Linguistic Theory in Indonesia*, University of Delaware; Atma Jaya Catholic University of Indonesia; and Tokyo University of Foreign Studies.

5. When verbal complexes become nouns via infinitive nominalization: A parallel to the verbal domain or category-individual?

Martina Werner
Austrian Academy of Sciences

Nominalized infinitives (NIs, such as *(das) Gehen* ‘walking’, *(das) Abschneiden* ‘cutting off’) are the most frequent deverbal nominalization patterns for abstract nouns in present-day German (PDG; cf. Blume 2004, Werner 2020), the NIs are involved in different constructions in PDG, e.g. in light-verb constructions (such as *ins Rollen kommen* ‘to get going’) or in the progressive (such as *Sie ist am Arbeiten* ‘she is working’). In PDG, the stems of infinitive nominalization come from simplex, prefix, and particle verbs and the NI does not have any morphological restrictions (1a). This is not the case for other nominalization patterns such as *-ung*-nominals (1a’), which originally only accepted only simplex verbs as bases but now also combine with prefix and particle verbs (for the diachronic details, see Demske 2000, Iordăchioaia/Werner 2019).

(1a) *(das) (An-)chatten* ‘(the) chatting’

(1a’) **Chattung*

(1b) *(das) Freunde-Anchatten* ‘(the) chatting with friends’

(1b’) ?*Freunde-Anchattung*

(1c) *(das) ständig-die-Freunde-Anchatten* ‘(the) constantly-chatting-with friends’

(1c’) **Ständig-die-Freunde_{AKK}-Anchattung*

(1d) *(das) Chatten der Freunde_{GEN}* ‘(the) chatting of friends’

(1d’) **Chattung der Freunde*

While NIs can be formed from phrases containing a verb and arguments or modifiers (cf. 1b-d), this is not the case for *-ung*-nouns (cf. 1b’-c’) although both patterns form abstract nouns in PDG. In addition, only the NI, but no *-ung*-nouns can nominalize verbal complexes which is shown in (2-5).

(2a) *(das) Gegessen-Haben* lit. ‘(the) having eaten’, i.e., ‘the fact that one has eaten’

(2a’) **Gegessen-Habung/-Haberei*

(3a) *(das) Akzeptiert-Sein* ‘(the) being accepted’

(3a’) **Akzeptiert-Seiung/-Seierei*

(4a) *(das) Akzeptiert-Worden-Sein* lit. ‘(the) having been accepted’

(4a’) **Akzeptiert-Worden-Seiung/Seierei*

(5a) *(das) Schlafen-Müssen* lit. ‘the having-[to]-sleep’

(5a’) **Schlafen-Müssung/Müsserei*

Here we see that converted NIs contain perfect, passive and modal auxiliaries while *-ung*-nouns are restricted in PDG (more details in Iordăchioaia/Werner 2019). But also other derivational affixes like *-erei* do also not allow for auxiliary nominalization despite formally non-restricted productivity (cf. **Gegessen-Haberei*, **Akzeptiert-Seierei*).

NIs in Old and Middle High German were typically conversions from simplex verbs, while prefix and particle verbs followed later (Werner 2020). In this light, the talk aims to answer the question of

how the NIs developed the ability to nominalize verbal complexes or, in other words, to what extent inflectional verbal categories can be integrated into nominals (or, vice versa, Grestenberger 2022). It will be asked if a certain logic can be identified as to whether some verbal categories (e.g., tense) are nominalized before others (e.g., modality or mood). This is of special interest because research on grammaticalization has identified sequences in which verbal categories develop, e.g. that aspect develops before tense (see e.g. Leiss 1992) and that verbal periphrases encoding tense developed from predicative constructions containing adjectives (see e.g. Bybee et al. 1994: 61ff). However, in such a view, potential restrictions of such a conceivable development, i.e., whether some categories do not participate in integration into nominals, are not automatically excluded. Questions regarding the degree to which there is a logic behind the nominalization of verbal categories provide important answers regarding the architecture of verbal categories, of the potential and the limits of nominalizability, and of a better derivation-inflection divide, since verbal categories are only allowed within the pattern of NIs, but not within that of derivation (see 2–5).

By taking a look at the sequence of category changes involved, the data-based talk (corpora: DTA/DWDS, Austrian Media Corpus) shows infinitive nominalization exactly follows the well-known principle of grammaticalization research, namely that aspectual or temporal forms develop first, while modal forms come last. In other words, the development of verbal categories in the nominal domain directly seems to reflect or follow the logic of grammaticalization of the verbal categories in the verbal domain. Despite these parallels however, there are also some differences between the infinitives of the nominal and verbal domain, especially w.r.t. modal verbs. In the light of different kinds of modality (deontic, reportative, epistemic), the talk identifies category-specific restrictions of infinitive nominalization, which have not been described in the literature so far. Furthermore, it aims to explain why the detected restrictions of nominal category change are inherently of stable nature by pointing to results from syntax, semantics. and language philosophy.

References

- Blume, K. (2004): *Nominalisierte Infinitive*. Tübingen: Niemeyer.
- Bybee, J. / Perkins, r. & Pagliuca, W. (1994): *The Evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.
- Demske, U. (2000): Zur Geschichte der *-ung*-Nominalisierung im Deutschen. Ein Wandel morphologischer Produktivität. In: *Beiträge zur Geschichte der deutschen Sprache und Literatur* 122, 365-411.
- Grestenberger, L. (2022): Reanalysis² and the directionality of “change events”: Bringing morphological change into the fold. PPT-slides, talk held at U Konstanz 11/24/2022.
- Iordăchioaia, G. / Werner, M. (2019): Categorical Shift via Aspect and Gender Change in Deverbal Nouns. In: Heyvaert, L. / Cuyckens, H. / Hartmann, S. (eds.): Special issue on Categorical shifts: from description to theory and back again. *Language Sciences* 73, 62-76.
- Leiss, E. (1992): *Die Verbalkategorien des Deutschen*. Berlin & New York: de Gruyter.
- Werner, M. (2020): Korpuslinguistische Perspektiven auf die sprachhistorische Entwicklung der nominalisierten Infinitive im Deutschen. *Linguistik online* 102,2, 155-181.

W7 - Interactions at the dawn of history: Methods and results in prehistoric contact linguistics

Organizers:

Rasmus G. Bjørn (Max Planck Institute of Geoanthropology, Jena)
bjorn@shh.mpg.de

Marwan Kilani (University of Basel - Swiss National Science Foundation)
marwan.kilani@unibas.ch

Workshop Description

It is well known that the elements of a language acquired through contact preserve traces of the past socio-cultural interactions of the communities that used it. This observation is particularly interesting when dealing with pre- and proto-historic realities, because it implies that these elements can be used to build bridges between languages and between language families, which in turn can be extremely useful in contextualizing such languages and family, in highlighting their positions in cross-linguistic networks, and in better locating them in relation with other languages, and thus both in space and time.

These linguistic concepts have been known for decades. However, recent developments in ancient genetics have introduced completely novel frameworks for investigating contacts between human populations in the past (Haak et al. 2015, Allentoft et al. 2015), which in turn have stimulated new, fresh debates about the possibility to combine ancient genetics, archaeology, and historical linguistics for the study of pre- and proto-historic realities.

As a result, new increasingly robust and sophisticated reconstructions of the social ecology of whole language families are being formulated (Sagart et al. 2019, Robbeets et al. 2021, Narasimhan et al. 2019, Rocha & Fehn 2016), and historical linguistics has witnessed a renewed interest in issues of contacts between pre- and proto-historic speech communities (and proto-languages). This new trend is well represented by various research projects on these topics that have been launched in the past few years, such as the recent ERC project by Guus Kroonen and his team, based in Leiden, which focuses on language contacts in prehistoric Europe in the context of Indo-European linguistics. It is also worth noting that this renewed interest is not limited to Europe and the Indo-European language family, but extends beyond it: good examples touching on different regions are the ongoing project of Wolfgang Behr based at the university of Zurich on pre- and proto-historic Wanderwörter in Central and East Asia, the recently concluded project by Federico Giusfredi on language contacts in pre-/proto-historic Anatolia, the recently (2022) launched project by Koen Bostoen at Ghent University on prehistoric contacts between Bantu and Khoisan languages, or the also recently (2022) launched project by Marwan Kilani at the university of Basel on linguistic interactions and Wanderwörter in Bronze Age Egypt and the Levant, just to name but a few.

These projects (and the work of several other scholars) are opening new avenues of research, are making new data available, and are suggesting new methodological approaches. Nevertheless, the work is far from over. On the contrary, the research developed in recent years has already yielded fruitful linguistic and historical insights, but it has also raised new questions and new methodological needs. First and foremost, there are theoretical questions that need to be discussed. While research on language contacts in modern languages has a long and established tradition, the systematic study of linguistic contacts in ancient languages

is still in its infancy, especially outside the Indo-European reality. Moreover, while the analytical frameworks developed to explore contacts in modern languages are undoubtedly valuable, the nature of the available evidence for ancient and proto-languages raises unique questions that require specific theoretical and methodological approaches to be answered satisfactorily. The fact that the data attesting prehistoric contact situations is usually limited and often difficult to substantiate by the comparative method alone, makes the need for solid, commonly agreed means to assess the veracity of hypotheses even more pressing. Moreover, the question of if and how linguistic data can be correlated with archeological and genetic evidence is becoming increasingly relevant, and sound discipline-specific methodologies (in our case, on the linguistic side) are a crucial basis for a constructive interdisciplinary dialogue.

It is thus clear that the question of language contacts and language interactions in pre- and proto-historic societies can be approached in multiple different ways, which we believe makes it an ideal topic for a conference such as this one.

First and foremost, we are aiming at gathering contributions that address methodological issues and offer new approaches to tackle them. We aim to have a good representation of research that focuses on non-European regions and/or deals with non-Indo-European languages, as we believe that a broader scope is essential to identify patterns and specificities. Discussions of specific case studies (whether based on single language-to-language interactions, or involving large geographical areas or *longue durée* approaches) is also welcomed and encouraged: good theory can only be developed on the basis of a careful and systematic investigation of real cases.

As mentioned above, several projects have emerged in recent years that aim to explore contact phenomena from different angles, often using interdisciplinary approaches that combine linguistic data with archeological and genetic evidence. Papers arising from such projects or presenting interim or final results are also welcomed.

We welcome discussions of contact phenomena touching on any linguistic level (phonology, morphology, lexicon, etc.), and we are especially interested in realities involving multiple languages. In this respect, we are particularly interested in contributions dealing with Wanderwörter that permeate several languages and distinct language families. Recent scholarship (Boutkan & Kossmann 2001, de Vaan 2008, Antonov & Jacques 2011, Haynie et al. 2014, Piispanen 2020, Peyrot 2016, Bjørn 2020, 2022, etc.) has focused on the specificities of Wanderwörter, highlighting how Wanderwörter are like breadcrumbs attesting ancient (and often pre- and proto-historic) networks of interlinguistic and intercultural interactions. Furthermore, Wanderwörter are characterized by two features that make them particularly interesting for the study of pre- and proto-historic contacts, namely their datability and their multiple interfaces. These two features can provide crucial insights into the historical and cultural contexts in which the words were transferred, thus making Wanderwörter a valuable tool for the investigation and contextualization of ancient interactions, of the participating speech communities, and of the history of the items they denote. Therefore, we believe that the analysis of Wanderwörter provides a very attractive topic for this conference.

Finally, we believe that there are several other types of language contact phenomena that deserve renewed scrutiny in light of recent and emerging research on prehistory, including but not limited to calques (e.g. Puhvel 1993), areal phenomena (e.g. Peyrot 2019), and extinct substrate languages (e.g. Lubotsky 2001). Papers focusing on these topics are also welcomed.

References

- Allentoft, Morten E., Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B. Damgaard, et al. 2015. ‘Population Genomics of Bronze Age Eurasia’. *Nature* 522 (7555): 167–72.
- Antonov, A., and Guillaume Jacques. 2010. ‘Turkic Kümüş “silver” and the Lambdaism vs Sigmatism Debate’. *Turkic Languages*.
- Bjørn, Rasmus G. 2020. ‘Nouns and Foreign Numerals: Anatolian “Four” and the Development of the PIE Decimal System’. In *Dispersals and Diversification: Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, edited by Matilde Serangeli and Thomas Olander, 54–76. Brill.
- . 2022. ‘Indo-European Loanwords and Exchange in Bronze Age Central and East Asia: Six New Perspectives on Prehistoric Exchange in the Eastern Steppe Zone’. *Evolutionary Human Sciences* 4: e23. <https://doi.org/10.1017/ehs.2022.16>.
- Boutkan, Dirk, and Maarten Kossmann. 2001. ‘On the Etymology of “Silver”’. *NOWELE* 38 (1): 3–15.
- Haak, Wolfgang, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, et al. 2015. ‘Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe’. *Nature* 522 (7555): 207–11.
- Haynie, Hannah, Claire Bower, Patience Epps, Jane Hill, and Patrick McConvell. 2014. ‘Wanderwörter in Languages of the Americas and Australia’. *Ampersand* 1: 1–18.
- Lubotsky, Alexander. 2001. ‘The Indo-Iranian Substratum’. In *Early Contacts Between Uralic and Indo-European: Linguistic and Archaeological Considerations: Papers Presented at an International Symposium Held at the Tvärminne Research Station of the University of Helsinki, 8-10 January, 1999*, edited by Christian Carpelan, Asko Parpola, and Petteri Koskikallio, 301–17. *Suomalais-Ugrilaisen Seuran Toimituksia* 242. Helsinki: Suomalais-Ugrilainen Seura.
- Mooijaart, M. A., M. J. van der Wal, and Michiel de Vaan, eds. 2008. ‘On Wanderwörter and Substrate Words in Etymological Research’. In *Yesterday’s Words: Contemporary, Current and Future Lexicography*, 199–207. Newcastle: Cambridge Scholars Pub.
- Narasimhan, Vagheesh M., Nick Patterson, Priya Moorjani, Nadin Rohland, Rebecca Bernardos, Swapan Mallick, Iosif Lazaridis, et al. 2019. ‘The Formation of Human Populations in South and Central Asia’. *Science*, 327–35.
- Peyrot, Michaël. 2017. ‘Language Contact in Central Asia: On the Etymology of Tocharian B Yolo “Bad”’. In *Etymology and the European Lexicon: Proceedings of the 14th Fachtagung Der Indogermanischen Gesellschaft, 17-22 September 2012, Copenhagen*, edited by Bjarne Simmelkjaer Sandgaard Hansen. Wiesbaden: Reichert Verlag.
- . 2019. ‘The Deviant Typological Profile of the Tocharian Branch of Indo-European May Be Due to Uralic Substrate Influence’. *Indo-European Linguistics* 7 (1): 72–121.
- Piispanen Peter Sauli. 2020. ‘An Ancient East Asian Wanderwort’. *Acta Orientalia Academiae Scientiarum Hungaricae* 73 (4): 567–84.
- Puhvel, Jaan. 1993. ‘A Hittite Calque in the Iliad’. *Historische Sprachforschung / Historical Linguistics* 106 (1): 36–38.
- Robbeets, Martine, Remco Bouckaert, Matthew Conte, Alexander Savelyev, Tao Li, Deog-Im An, Ken-ichi Shinoda, et al. 2021. ‘Triangulation Supports Agricultural Spread of the Transeurasian Languages’. *Nature* 599 (7886): 616–21.
- Rocha, Jorge, and Anne-Maria Fehn. 2016. ‘Genetics and Demographic History of the Bantu’. In *ELS*, 1–9. John Wiley & Sons, Ltd.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, and Johann-Mattis List. 2019. ‘Dated Language Phylogenies Shed Light on

the Ancestry of Sino-Tibetan'. *Proceedings of the National Academy of Sciences* 116 (21): 10317–22.

Abstracts

1. Tracing borrowings in and out of proto-Nahuatl

Magnus Pharao Hansen (University of Copenhagen)

Albert Davletshin (Universidad Veracruzana)

The Nahuan languages are group of closely related languages spoken in Mexico and El Salvador, which form a well-defined sub-branch within the southern branch of the Uto-Aztecan language family. They are the only Uto-Aztecan languages that form part of the Mesoamerican linguistic area. Proto-Nahuatl displays assimilation to the languages of Mesoamerica in all aspects of linguistic structure including phonology, morphology, and syntax. Surprisingly, the lexicon does not appear to have been affected to the same degree, as most of the core vocabulary of Proto-Nahuatl can be traced back to Proto-Uto-Aztecan. Identification of borrowings between Mesoamerican languages has played an important role in studying prehistoric processes of the area. For example the word cacao, proposed by Campbell and Kaufman (1976) as borrowings from a Mixe-Zoquean language that was widely diffused within the region has been seen as significant argument for the identification of the Olmec culture as Mixe-Zoque speaking. However, Dakin and Wichmann (2001) later argued that the word ‘cacao’ might have been of Uto-Aztecan origins and suggested that Nahua speakers had an early presence and a dominant role in trade networks in Mesoamerica (Dakin 2003). This argument was rejected by Terrence Kaufman and John Justeson (2007, 2009) who maintained that prior to the rise to dominance of Nahuan speaking peoples in the Post-Classic period, Proto-Nahuatl was primarily a recipient of borrowings from other Mesoamerican languages. They proposed a number of additional borrowings from Mesoamerican languages into proto-Nahuatl, from Mayan, Tepehua-Totonacan, and Mixe-Zoquean languages. These proposals of borrowings into Proto-Nahuatl have been used to locate the place of origin of Nahuan languages in the North-Eastern periphery of Mesoamerica during the Classic Period rather than in North Western Mexico closer to the other Uto-Aztecan languages (e.g. Beekman & Christensen 2003), or whether the proto-Nahua community was already located within central Mexico as argued by Dakin (2003). This challenges us to find out whether Proto-Nahuatl was indeed mainly a recipient language in Mesoamerica, or perhaps also a donor.

Ongoing reconstruction work on proto-Nahuatl and the intermediary stages of Uto-Aztecan shows that many loans identified by Kaufman and Justeson can equally well be seen as inherited from proto-Uto-Aztecan, suggesting the opposite direction of borrowing. However, Proto-Tepehua-Totonacan and Proto-Mixe-Zoquean lexicons have demonstrated cases where there are viable reconstructions in both language families, making it a hard to determine the direction of borrowing. Therefore, there is a pressing need to develop methodologies to assess and evaluate the overall probability of the different borrowing scenarios involving the Mesoamerican languages. The paper describes the challenges involved and suggests some avenues for developing an approach to this challenge.

References

- Beekman, C.S. and Christensen, A.F., 2003. Controlling for doubt and uncertainty through multiple lines of evidence: A new look at the Mesoamerican Nahua migrations. *Journal of Archaeological Method and Theory*, 10(2), pp.111-164.
- Campbell, L. and Kaufman, T., 1976. A linguistic look at the Olmecs. *American Antiquity*, 41(1), pp.80-89.
- Campbell, L., Kaufman, T. and Smith-Stark, T.C., 1986. Meso-America as a linguistic area. *Language*, pp.530-570.

- Dakin, K., 2003. Uto-Aztecan In The Linguistic Stratigraphy Of Mesoamerica. *Language contacts in prehistory: studies in stratigraphy*, 239
- Dakin, K. and Wichmann, S., 2000. Cacao and chocolate: A Uto-Aztecan perspective. *Ancient Mesoamerica*, 11(1), pp.55-75.
- Kaufman, T. and Justeson, J., 2007. The history of the word for cacao in ancient Mesoamerica. *Ancient Mesoamerica*, 18(2), pp.193-237.
- Kaufman, T. and Justeson, J., 2009. Historical linguistics and pre-Columbian Mesoamerica. *Ancient Mesoamerica*, 20(2), pp.221-231.

2. Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction

Koen Bostoen, (UGent Centre for Bantu Studies, Ghent University, Belgium)

Lorenzo Maselli, (UGent Centre for Bantu Studies, Ghent University, Belgium; Université de Mons, Belgium)

Sara Pacchiarotti, (UGent Centre for Bantu Studies, Ghent University, Belgium)

and Jean-Pierre Donzo (UGent Centre for Bantu Studies, Ghent University, Belgium; ISP Gombe, Kinshasa, DRC)

Rainforest Hunter-Gatherer (RHG) communities in Central Africa, also known as *Batwa* or “Pygmies” and commonly seen as the descendants of the region’s earliest modern humans, are thought to have abandoned – in times unknown – their own ancestral languages for different Bantu, Central Sudanic or Ubangi languages. As there are no written records of those putative ancestral RHG languages, two main research strategies have been pursued in the search of a potentially shared prehistoric RHG substrate: (i) lexical comparison aimed at identifying traces of ancestral pre-shift vocabulary (Carpaneto and Geremi 1992; Bahuchet 1993; Hideaki and Ichikawa 2003; Terashima 2003; Demolin 2021) and (ii) phonological and morphological features distinguishing the varieties spoken by RHG from those spoken by food-producing populations (Hulstaert 1948; Schebesta 1952; Vorbichler 1964, 1967, 1968; Hulstaert 1978; Möhlig 1981; Motingea Mangulu 1994, 2010, 2021). Although it is challenging to recover Central Africa's pre-Bantu linguistic landscape, recent historical-comparative research focusing on languages of Bantu speech communities which may have incorporated ancestral RHG suggests that linguistic diversity among autochthonous RHG before they shifted to Bantu languages might have been high (Pacchiarotti and Bostoen 2021). In this talk, we focus on a specific phonological feature possibly diagnostic of RHG substrate, namely the simplification of NC clusters (where N= nasal and C = oral stop) in favor of the oral stop (e.g. /ŋg/ > /g/). This sound shift, which is quite rare in Bantu, has recently been observed in some newly documented RHG Bantu languages spoken in the southern fringes of the Congo rainforest, more specifically in the Mai-Ndombe province of the Democratic Republic of the Congo (DRC). In languages of the West-Coastal Bantu (WCB) branch spoken in and to the southeast of the Mai-Ndombe, the simplification of clusters of nasal and oral stops is also widespread, but always in favor of the nasal (e.g. /ŋg/ > /ŋ/). In other RHG Bantu languages of the Mai-Ndombe and in geographically more distant RHG communities to the north and west, there is no such simplification.

This phenomenon is of particular interest for at least three reasons. First, unlike the Bantu dissimilatory sound change known as Kwanyama’s Rule, whereby a NC cluster is reduced to C in C(onsonant)² position if the word contains another NC cluster in C1 position (e.g. *ŋgàndú > ŋgàdú), the simplification phenomenon in selected Mai-Ndombe RHG varieties happens independently of the nature of C1. This type of change is very uncommon in Bantu and contrasts with the type of NC cluster simplification found in neighboring WCB languages. Second, the same NC cluster simplification has been reported in other Bantu languages spoken further north in the Congo rainforest by RHG communities (Chabiron et al. 2013). Third, often times /d/ as the simplified outcome of *nd is realized as [d] in selected Mai-Ndombe RHG varieties. These also attest an abundance of retroflex flaps [ɾ] elsewhere uncommon in Bantu. Apparently, RHG communities speaking the Central Sudanic language Efe also show the retroflex realization of [d] and [ɾ] which is a phonetic feature not shared by non-RHG communities speaking Efe and closely related Central Sudanic varieties such as Mamvu and Mangbetu (Vorbichler 1967, 1968). Besides RHG varieties, a couple of apparently non-RHG Bantu varieties in the Mai-Ndombe also attest a phonemic nasal retroflex /ŋ/ (historically originating in C2 *n and *nd), a unique case in Africa to the best of our knowledge (Maselli et al. 2022).

We will provide a systematic account of the distribution of this unconditioned NC cluster reduction in newly and previously documented RHG Bantu languages in order to assess the historical implications of this possible substrate feature. We will also assess to what extent retroflexion should indeed be considered as an additional substrate feature.

References

- Bahuchet, Serge. 1993. History of the Inhabitants of the Central Africa Rainforest : Perspectives from Comparative Linguistics. In C. M. Hladik, A. Hladik, O.F. Linares, H. Pagezy, A. Semple & M. Hadley (eds.), *Tropical forests, people and food: Biocultural interactions and applications to development* (Man and the Biosphere 13), 37-54. Paris: UNESCO Paris & The Parthenon Publishing Group.
- Carpaneto, Giuseppe M. & Francesco P. Geremi. 1992. Diversity of mammals and traditional hunting in central African rain forests. *Agriculture, Ecosystems & Environment* 40(1). 335-354.
- Chabiron, Clotilde, Silvia Gally & Didier Demolin. 2013. Les parlers pygmées du bassin équatorial du Congo. *Géolinguistique* 14. 125-144.
- Demolin, Didier. 2021. The languages of the Ituri forest Pygmies: Contact and historical perspectives. Paper presented at the Diedrich Westermann-Workshop, Humboldt-Universität zu Berlin.
- Hideaki, Terashima & Mitsuo Ichikawa. 2003. A comparative ethnobotany of the Mbuti and Efe hunter-gatherers in the Ituri forest, Democratic Republic of the Congo. *African Study Monographs* 24(1/2). 1-168.
- Hulstaert, Gustaaf. 1948. Le dialecte des pygmoïdes Batswa de l'Equateur. *Africa* 18(1). 21-28.
- Hulstaert, Gustaaf. 1978. Notes sur la langue des Bafotó (Zaire). *Anthropos* 73. 113-132.
- Maselli, Lorenzo, Veronique Delvaux, Jean-Pierre Donzo, Sara Pacchiarotti & Koen Bostoen. 2022. Retroflex sounds in the Mai-Ndombe (DRC): the case of nasals in North Boma B82 and Nunu B822. Paper presented at the Conference Bantu 9, Malawi University of Science and Technology. . Paper presented at the Bantu 9, Malawi University of Science and Technology.
- Möhlig, Wilhelm J.G. 1981. Stratification in the History of the Bantu Languages. *Sprache und Geschichte in Afrika*. 251-317.
- Motingea Mangulu, André. 1994. Notes sur le Parler des Pygmées d'Itendo (Zone de Kiri/Maindombe). *Annales Aequatoria* 15. 341-382.
- Motingea Mangulu, André. 2010. *Aspects des parlers minoritaires des lacs Tumba et Inongo: Contribution à l'histoire de contact des langues dans le bassin central congolais*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Studies.
- Motingea Mangulu, André. 2021. Vers une écologie linguistique des anciens chasseurs cueilleurs du bassin central congolais: Traces d'une proto-population ancestrale homogène. *Anthropos* 116(1). 137-143.
- Pacchiarotti, Sara & Koen Bostoen. 2021. Final vowel loss in Lower Kasai Bantu (DRC) as a contact-induced change. *Journal of Language Contact* 14. 437-474.
- Schebesta, Paul Joachim. 1952. *Les pygmées du Congo Belge*. Bruxelles: Institut Royal Colonial Belge.
- Terashima, Hideaki. 2003. Names, Use and Attributes of Plants and Animals among the Ituri Forest Foragers: A Comparative Ethnobotanical and Ethnozoological Study. *African Study Monographs* 28. 7-24.
- Vorbichler, Anton. 1964. Die sprachlichen Beziehungen zwischen den Waldnegern und Pygmäen in der Republik Kongo-Léo. *Ethnologie*. Musée de l'Homme, Univ. de Paris.

- Vorbichler, Anton. 1967. Erzählungen in der Mamvu-Sprache, mit einer vergleichenden Einführung in die Phonologie und das Verbalsystem. *Afrika und Ubersee* 50. 244–278.
- Vorbichler, Anton. 1968. Erzählungen in der Mamvu-Sprache, mit einer vergleichenden Einführung in die Phonologie und das Verbalsystem. *Afrika und Ubersee* 51. 2–29, 116–144.

3. Prehistoric language contact in Berber

Lameen Souag (LACITO - Centre National de la Recherche Scientifique)

In Nichols' (1992) terminology, North Africa and the Sahara constitute a classic spread zone, dominated throughout most of recorded history by a single indigenous language family: Berber. Any analysis of prehistoric language contact around the Mediterranean accordingly needs to take Berber into account. The surprising homogeneity of this family seems to reflect a history of repeated levelling events, facilitated by interregional trade and by high mobility in pastoralist regions (Souag 2017). The contact history of Berber is likewise overshadowed by the influence of major expansions into the region, with successive layers of Punic, Latin, Arabic, and Romance loanwords predominating even in regions neither Carthage nor Rome ever dreamed of ruling (Vycichl 1952; Múrcia Sánchez 2011; Kossmann 2013; Blažek 2014). Berber languages have nevertheless preserved a certain number of clues to what preceded these eras of centralisation.

A number of candidate prehistoric borrowings are pan-Berber. A couple of Egyptian borrowings are prominent in date palm terminology, reflecting the westward expansion of this agriculturally vital species (Kossmann 2002; Vycichl 1991); some localised words shared with Nubian, such as 'onion', may reflect a similar contact scenario (Vycichl 1961; Kossmann & Jakobi *fc*). The numerals 5-9 are evidently Semitic in origin, but equally evidently reflect contact with a stage of Semitic more conservative than Punic or even Arabic. The names of several metals, such as iron and silver (Boutkan & Kossmann 1999), are well-known *Wanderwörter* whose precise source presents difficulties but must be rather early; a comparable situation is found for equine terminology. Efforts to identify Berber roots for "proto-Mediterranean" substrate terms in languages of the northern Mediterranean (Chaker 2013; Argiolas 2020) largely appear unconvincing, but suggest some promising directions for further research.

Aside from prehistoric contact between Berber and other families, the increasing attention paid towards intra-Berber variation opens up the difficult but interesting possibility of exploring prehistoric substrata within North Africa itself. The most promising case so far involves the Tuareg of the central and southern Sahara, where a number of phonologically anomalous terms with no good Berber source are concentrated in the domains of hunting and farming (Kossmann 2005). Analysis of kinship terminology suggests that this reflects a substratum with similarities to modern Songhay, whose speakers would have a substantial influence on Tuareg social structure. Much less can be said for the present about other areas, but in the Fezzan a few words seem like potential candidates for remnants of a Saharan substratum, while the sharply divergent vocabulary of Zenaga is unlikely to be explicable solely in terms of an early split.

The time is ripe for reexamining prehistoric loans in Berber: more comparative data is available on Berber languages than ever before, and our understanding of the historical phonology of Berber has advanced significantly in recent decades (Prasse 2003; van Putten 2019; Kossmann 2020). This talk will therefore seek to present a new synthesis, sifting better candidates from proposals that need to be abandoned and suggesting new possibilities.

References

Argiolas, Valeria. 2020. Un état des études sur l'action du substrat/adstrat « libyque » en Sarde : problèmes et méthodes. *Asinag*. Institut Royal de la Culture Amazighe (IRCAM) (15). 31–61.

- Blažek, Václav. 2014. Phoenician/Punic loans in Berber languages and their role in chronology of Berber. *Folia Orientalia* 51. 275–293.
- Boutkan, Dirk & Maarten Kossmann. 1999. Some Berber parallels of European substratum words. *Journal of Indo-European studies* 27(1–2). 87–100.
- Chaker, Salem. 2013. Olivier/Olive : Note linguistique. *Encyclopédie berbère*. Éditions Peeters (35). 5749–5760. <https://doi.org/10.4000/encyclopedieberbere.2814>.
- Kossmann, Maarten. 2002. Deux emprunts à l'égyptien ancien en berbère. In Kamal Naït-Zerrad (ed.), *Articles de linguistique berbère : Mémorial Werner Vycichl*, 245–252. Paris: L'Harmattan.
- Kossmann, Maarten. 2005. *Berber Loanwords in Hausa* (Berber Studies v. 12). Köln: Rüdiger Köppe.
- Kossmann, Maarten. 2013. *The Arabic Influence on Northern Berber*. Leiden: Brill.
- Kossmann, Maarten. 2020. Proto-Berber phonological reconstruction: An update. *Linguistique et Langues Africaines* 6. 11–42.
- Kossmann, Maarten & Angelika Jakobi. fc. On alleged Berber borrowings into Nubian. *Sprache und Geschichte in Afrika* 18.
- Múrcia Sánchez, Carles. 2011. Que sait-on de la langue des Maures? Distribution géographique et situation sociolinguistique des langues en Afrique Proconsulaire. In C. Ruiz Darasse & E. R. Luján (eds.), *Contacts linguistiques dans l'Occident méditerranéen antique*, 103–126. Madrid.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago ; London: University of Chicago Press.
- Prasse, Karl-G. 2003. La vocalisation du protoberbère. In *Selected Comparative-Historical Afrasian Linguistic Studies in Memory of Igor M. Diakonoff* (Lincom Studies in Afroasiatic Linguistics 14), 41–54. Munich: LINCOM.
- Putten, Marijn van. 2019. Introducción al estudio diacrónico del bereber. In José Juan Batista Rodríguez (ed.), *Estudios sobre toponimia canaria prehispánica*, 225–307. La Laguna: Academia Canaria De La Lengua.
- Souag, Lameen. 2017. La diffusion en berbère : Reconcilier les modèles. In *Diffusion: Implantation, Affinités, Convergence* (Mémoires de La Société Linguistique de Paris (Nouvelle Série) 24), 83–107. Louvain: Peeters.
- Vycichl, Werner. 1952. Punischer Spracheinfluss im Berberischen. *Journal of Near Eastern Studies* 11(3). 198–204.
- Vycichl, Werner. 1961. Berber Words in Nubian. *Kush* 9.
- Vycichl, Werner. 1991. Vocabulary, African Contacts with Autochthonous Coptic. (Ed.) Aziz Suryal Atiya. *Coptic Encyclopedia*. New York: Macmillan.

4. Linguistic convergence in the Ancient Near East

Paul Widmer (University of Zurich)

Barbara Sonnehauser (University of Zurich)

Applying advanced methods (sBayes, Ranacher et al. 2021) Efrat-Kowalsky et al. (in rev.) found evidence for an amount of similarity between the unrelated ancient Near East languages Hurrian and Sumerian that cannot be accounted for by contact, universal preferences, or inheritance. The authors suggest that Hurrian and Sumerian might be the last survivors of an earlier area which was wiped out by later spreads of Semitic and Indo-European, or alternatively, Hurrian and Sumerian reflect an ancient global distribution which is different from today's.

We follow up on this promising approach and assess the impact of language sampling on the results. To do this, we expand the language sample by adding languages from the same region (e.g. Ancient Greek, Classical Armenian, Old and Middle Iranian varieties) and ancient and medieval varieties from Europe that were not part of the original sample. We apply the same methodology as Efrat-Kowalsky et al. (in rev.) and find that Hurrian and Sumerian still display similarity that cannot be explained by genealogy or universal preferences. However, the algorithm identifies two Indo-European languages, Middle Persian and Classical Armenian, that are assigned to the same cluster as Hurrian and Sumerian. We suggest that the similarity between these languages is best explained by areal convergence, a signal formerly not captured because of the restricted language sample.

References

- Efrat-Kowalsky, Nour et al. (in rev.). “Oldest attested languages in the Near East reveal deep transformations in the distribution of linguistic features”. In: Scientific Reports. in revision.
- Ranacher, Peter et al. (2021). “Contact-tracing in cultural evolution: a Bayesian mixture model to detect geographic areas of language contact”. In: Journal of The Royal Society Interface 18.181, p. 20201031. doi: 10.1098/rsif.2020.1031.

5. Language Contact in the Ancient Caucasus: the View from Kartvelian

Thomas Wier (Free University of Tbilisi)

The Kartvelian language family has been spoken in the southwestern portion of the Caucasus since at least the Middle Bronze Age (Tuite 2004), and as a consequence constitutes a particularly rich window onto language contacts throughout the region. Unlike almost all other autochthonous Caucasian languages, the written attestation of such contacts also extends back to the fifth century A.D. in the Old Georgian corpus. This allows us to trace with much greater precision than with most other regional languages how and when such contacts occurred. In this talk, I will provide a survey of Kartvelian's contacts with other language families, and discuss how these contacts elucidate various aspects of the phonological history of Kartvelian and other language families.

We might divide the set of lexical contacts of Kartvelian into five main sets: (1) intra-Kartvelian; (2) Indo-European, (3) Nakh-Daghestanian, (4) Semitic and (5) all other non-Indo-European. The first category of contacts has occurred continuously since the protolanguage's first phylogenetic differentiation, but is visible in texts mostly as borrowings from the Zan languages (Megrelian and Laz) into Georgian and, later, Svan, and Georgian into all the other branches. The second consists of a vast and diverse array of direct loans in various periods from Greek, Indo-Iranian, Armenian, Hittite and other often indeterminate but likely Indo-European sources. The third consists of loans (often with fossilized remnants of gender markers) from Lezgian, Tsezic, Nakh, and Avar-Andic languages or other languages likely from the Nakh-Daghestanian family (Xalilov 1993). The fourth consists of loans primarily from Akkadian and Aramaic, but also a distinct residue of loans from unclear Semitic sources. Although it is likely that Kartvelian has/had been in contact with Abkhaz-Adyghean and Hurro-Urartian languages since remote antiquity, demonstrable evidence of direct ancient lexical loans from these sources is surprisingly limited. (Loans from Abkhaz into Megrelian are ubiquitous however.)

Such loans both within and without Kartvelian not only provide a picture of who Kartvelian speakers were in contact with, it also provides data that allow us to understand the internal phonological development of the family. By careful comparison of dates of first attestation with attested forms in donor languages, we can begin to build a picture of when certain phonological shifts occurred over time. Thus the Megrelian shift that raised *a to /o/ (Gamkrelidze & Machavariani 1965; Fähnrich 2007) must have occurred after *pat- was borrowed into Greek as *Φᾶσις*, but before Greek's own well-documented rule of assibilation; this Zanism became the modern town of *Poti*. This narrows down the Megrelian sound-shift to around the late first millennium BC. And because some Svan words like *čönčx* 'skeleton, face' borrowed from Megrelian *čončxi* undergo umlaut, we can date Svan's umlaut rule to a period after Megrelian's raising rule. Likewise, the Megrelian rule lowering *e to /a/ and epenthesizing nasal obstruents in accented syllables must have occurred after a loaning event: *mankana* 'machine, device' from Greek *μηχανή*. We also see direct evidence for the loss of a laryngeal in Kartvelian: Kartvelian **hezo* 'courtyard' from Ugaritic *h̄zr* 'courtyard'. Such ancient contacts in other words reveal not just the lexical but also the structural history of the family.

References

- Fähnrich, H. 2007. *Kartwelisches Etymologisches Wörterbuch*. Brill.
 Gamkrelidze, T.V. and Machavariani, G.I., 1965. *The System of Sonants and Ablaut in Kartvelian Languages: A Typology of Common Kartvelian Structure*. Tbilisi: Metsniereba Publishers (in Georgian and Russian).

- Tuite, K., 2004. 'Early Georgian'. *Cambridge Encyclopedia of the World's Ancient Languages*, pp.967-987.
- Xalilov, Madžid Š. 1993. *Gruzinsko-dagestanske jazykovye kontakty (na materiale avarsko-cezskix i nekotoryx lezginskix jazykov)* [Georgian-Daghestanian linguistic contacts (on the basis of Avar-Tsezic and some Lezgian languages)]. Doctoral dissertation. Makhachkala: Russian Academy of Sciences, Daghestan Scientific Center.

6. An archaeolinguistic approach to Indianisation and Sinicisation of languages in Eastern Eurasia

Chingduang Yurayong (University of Helsinki)
et alii (see after the abstract)

Linguistic processes of Indianisation and Sinicisation are two areal phenomena visible across respective contact zones in Eastern Eurasia. They occurred from late prehistory onwards with major watersheds marked by the introduction of writing systems and vocabularies, giving expression to cosmopolitan modes of rulership, religion and trade. Linguistic Indianisation and Sinicisation have consequent correlation in the material record, which over the same period reflects multiple trajectories of state formation and subsequent transnational history. Processes of linguistic and cultural spread have been extensively studied for the individual regions (e.g. Smith 1999; Byington 2013; Carter et al. 2021; Huang & Kang 2022), but fewer transregional comparisons have been conducted (e.g. Lieberman 2003, 2009; Manguin et al. 2011).

Adapting principles of the Wave Theory (Schmidt 1872) to Güldemann's (2008) model of linguistic area, our study delineates the spatially variegated degrees of Indianisation and Sinicisation as they extend into Southeast and Northeast Asia from first millennia BCE to CE. Combining evidence from linguistics, archaeology and history we examine whether the degrees of contact-induced outcomes decrease relative to geographical distance from their areal hotbeds as is the case for Western Lingnan Sprachbund (Szeto & Yurayong 2022). We predict geographic radially being complicated by maritime polities. Our study is further informed by Watkins' (2000: xxii) parallels between language as a cognitive nonmaterial culture, and artefacts as material culture. We hypothesize that certain categories, such as loanwords/Wanderwörter in language, and prestige items for trades in material culture, represent more superficial layers of their respective fields which travel further, while categories including typological profiles, toponyms and artefacts reflective of local subsistence patterns constitute deeper layers which travel less far. Such variegation will become complicated with the adoption of cosmopolitan signification systems by early states giving rise to multiple sub-areal hotbeds which together form larger core circles of contact. The data are visualised cartographically with ArcGIS programme by illustrating four categories of evidence: 1) ancient epigraphs in which historical discourse on contact events with the Indic and Sinitic civilisations were attested, 2) ancient Indic and Sinitic-styled architecture, 3) sites where traded goods as traces of the maritime silk road have been found, and 4) language communities reconstructed through historical records and their present-day distribution. The first three sets of data are interpreted as presence or absence of evidence, while the linguistic data can be further quantified by scores aggregated from degrees of Indic or Sinitic loanwords and typological convergence which each language datapoint shows. The anticipated results will show that toponyms, ancient epigraphs and architecture such as ancient commanderies can be used to draw boundaries of the areal hotbeds which falls under a direct contact with the source of influence, while shared typological tendencies lying in human cognition can extend further to the core circles with lower contact intensity. Meanwhile, loanwords/Wanderwörter and traded goods such as glass beads can spread beyond the core circles towards the peripheries where contact influence is not necessarily direct but transmitted through intermediators. For instance, traces of native Sinitic epigraphs and commanderies are located as far as to Liaodong Peninsula in the north and Northern Vietnam in the south, marking boundaries between the areal hotbeds and core circles in which the degree of linguistic Sinicisation observed in Koreanic and Vietic is high (Eom 2015; Alves 2022), while the degree of Sinicisation gradually decreases towards the peripheries as it was largely transmitted secondarily through Koreanic to Japonic (Yurayong & Szeto 2020) and through Vietic to Chamic (Thurgood 1999).

The current study puts Eastern Eurasia in the current trend of a cross-disciplinary approach to prehistoric contact and its outcomes by illustrating more quantifiable data illustration and analysis methods which can facilitate estimation of degrees of contact intensity in different times and spaces.

References

- Alves, Mark. 2022. Lexical evidence of the Vietic household before and after language contact with Sinitic. In Trang Phan, John Phan & Mark Alves (eds.), *Vietnamese Linguistics: A State of the Field*, 15-58. Honolulu, HI: University of Hawai'i Press.
- Byington, Mark E. (ed.). 2013. *The Han Commanderies in Early Korean History*. Cambridge, MA: Korea Institute, Harvard University.
- Carter, Alison Kyra, Laure Dussubieux, Miriam T. Stark, & H. Albert Gilg. 2021. Angkor Borei and Protohistoric Trade Networks: A View from the Glass and Stone Bead Assemblage. *Asian Perspectives* 60(1). 32-70.
- Eom, Ik-Sang. 2015. 2,200 years of language contact between Korean and Chinese. In William S-Y. Wang & Chaofen Sun (eds.), *The Oxford Handbook of Chinese Linguistics*, 226-235. Oxford: Oxford University Press.
- Güldemann, Tom. 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern Sub-Saharan Africa. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 151-185. Cambridge: Cambridge University Press.
- Huang, Chin-Hao & David C. Kang. 2022. State Formation in Korea and Japan, 400–800 CE: Emulation and Learning, Not Bellicist Competition. *International Organization* 76(1). 1-31.
- Lieberman, Victor. 2003. *Strange parallels: Vol. 1 Integration on the Mainland: Southeast Asia in Global Context, c. 800–1830*. Cambridge: Cambridge University Press.
- Lieberman, Victor. 2009. *Strange Parallels: Vol. 2 Mainland Mirrors: Europe, Japan, China, South Asia, and the Islands: Southeast Asia in Global Context, C. 800-1830*. Cambridge: Cambridge University Press.
- Manguin, Pierre-Yves, A. Mani & Geoff Wade (eds.). 2011. *Early Interactions between South and Southeast Asia: Reflections on Cross-Cultural Exchange*. Singapore: Institute of Southeast Asian Studies.
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: H. Böhlau.
- Smith, Monica L. 1999. “Indianization” from the Indian point of view: Trade and cultural contacts with Southeast Asia in the early first millennium C.E.. *Journal of the Economic and Social History of the Orient* 42(1). 1-26.
- Szeto, Pui Yiu & Chingduang Yurayong. 2022. Establishing a Sprachbund in the Western Lingnan region: Conceptual and methodological issues. *Folia Linguistica* 56(1). 25-55.
- Thurgood, Graham. 1999. *From ancient Cham to modern dialects: two thousand years of language contact and change*. Honolulu, HI: University of Hawai'i press.
- Watkins, Calvert. 2000. *The American Heritage Dictionary of Indo-European Roots*. Boston, MA: Houghton Mifflin Co.
- Yurayong, Chingduang & Pui Yiu Szeto. 2020. Altaicization and De-Altaicization of Japonic and Koreanic. *International Journal of Eurasian Linguistics* 2(1). 108-148.

Co-authors and responsibility areas

Historical linguistics

1. Chingduang Yurayong (University of Helsinki) – chingduang.yurayong@helsinki.fi > Lexical Indianisation of Suvarṇabhūmi; Typological Sinicisation of Korea

2. Sami Honkasalo (University of Helsinki) – sami.honkasalo@helsinki.fi > Typological Indianisation of Suvarṇabhūmi; Typological Sinicisation of Japan
3. Pui Yiu Szeto (Ca' Foscari University of Venice) – puiyu.szeto@unive.it > Typological and lexical Sinicisation of Lingnan
4. Charlotte Ye Xu (Nanjing University) – charlottexuye@gmail.com > Lexical Sinicisation of Korea and Japan

Historiography

5. Andrew Logie (University of Helsinki) – andrew.logie@helsinki.fi > Sinicisation of Korea, Japan and Vietnam
6. Utain Wongsathit (Silpakorn University) – tainzgree@yahoo.com > Indianisation of Northeast India and Myanmar
7. Sombat Mangmeesuksiri (Silpakorn University) – sombat69@hotmail.com > Indianisation of Thailand and Laos
8. Kangvol Khatshima (Silpakorn University) – kangvol@gmail.com > Indianisation of Cambodia and Southern Vietnam

Archaeology

1. Youngchan Oh (Ewha Womans University) – ycoh@ewha.ac.kr > Sinicisation of Wiman Joseon (195-108 BC) and Triệu Đà of Nam Viet (203-111 BC)
2. Ohyoung Kwon (Seoul National University) – koy1108@snu.ac.kr > Sinicisation of Baekje (200-600 AD) and Champa Lâm Ấp (192-629 AD)
3. Junyoung Park (Seoul National University) – zkdwlsk0111@snu.ac.kr > Trade of glass beads in India and Southeast Asia
4. Cheunsoo Park (Kyungpook National University) – chspark@knu.ac.kr > Trade of glass artefacts and potteries in Central Asia and Northeast Asia

“Filling in the diachronic gaps: the view of Old Iranian from the present”

Convenors:

Shuan Osman Karim, Goethe University of Frankfurt

Saloumeh Gholami, Goethe University of Frankfurt

Corresponding Author:

Shuan Osman Karim (karim.56@osu.edu)

Description

Research into the prehistory of Iranian languages is a field doubly blessed: (1) there is a fairly large corpus of Old Avestan dating back between the 1st and 2nd millennia BCE and a small corpus of Old Persian dating back as far as the 6th century BCE (Skjærvø, 2017, 471). Because of the corpora, much is known about Old Iranian, and Old Iranian has played an important role in the reconstruction of Proto-Indo-European (PIE). (2) The modern languages of the greater Iranian world are diverse and numerous, preserving features of PIE already lost in the extant Old Iranian texts (e.g., the retention of PIE laryngeals in New Iranian languages following Kümmel, 2014). Despite these archaisms, many of these languages have changed radically and independently along what Stilo (2008) has deemed the reduction and innovation axes. They have lost case and innovated it anew. According to a proposal by Karim, they have lost gender in all but a few facets of the grammar and renovated it anew (Karim, 2021, ch2 and ch4). These radical transformations lead to the inevitable question: what would our picture of Old Iranian be without the extant Old Iranian texts, and to what extent does our reliance on Old Iranian bias our analysis of New Iranian languages? None of the New Iranian languages is the direct descendant of any of the Middle or Old Iranian languages except for New Persian (< Middle Persian < Old Persian following Korn, 2017, 609).

Additional issues affecting the historical analysis of Iranian languages are that Iranian populations were largely nomadic in their early history, and there has been massive borrowing between genetically related languages (Korn, 2017, 611). This situation invokes the analogy of the Rubik’s cube: As each group migrates to a new region, its contact languages change, and those languages undergo sprachbund-like shared changes, “mirror[ing] the multilingual situation of the vast majority of speakers of Ir. languages in past and present times” (Korn, 2017, 611). The existence of many phonological convergences due to borrowing suggests that Iranian historical linguists should prefer morphological innovation over regular sound change. Korn (2019, 268) uses morphological isoglosses to develop the current best understanding of the genealogy of Iranian, following Clackson's (2007) assertion that “It is now generally agreed among linguists that the most certain sub-groups are constructed on the basis of unique shared morphological innovations.” This runs contrary to the typical methods of historical linguists that begin with sound change because of Neo-Grammarians regularity; “[s]ound change I, in so far as it takes place mechanically, takes place according to laws that admit no exception” (Zosthoff and Brugmann, 1878, apud Hock & Joseph, 1996). Recently work by Gholami has suggested that phonological changes cannot be dismissed a priori despite the difficulty in establishing cognacy. Additionally, it is hard to compare constructions across the Iranian languages because the pioneering work on many varieties was conducted by scholars with little to no linguistic training. The ultimate result is inconsistent and innovative terminology being used to refer to

well-understood linguistic concepts. For instance, there are at least four terms for definite articles: “definite” (Mackenzie, 1961; MacKenzie, 1966; Mahmoudveysi & Bailey, 2013; Mahmoudveysi, Bailey, Paul, & Haig, 2012; Opengîn, 2016, etc.), “demarcative” (McKinnon, 2011), “determinative” (Windfuhr, 2012), and “deictic” (Windfuhr, 1991) appear in the literature (Karim, 2021, 217); three terms for applicatives: “applicatives” (Karim & Salehi, 2022), “placeholder constructions” (Jügel, 2016), and “absolute prepositions” (Mackenzie, 1961); and there is idiosyncratic terminology for adjectives, possessives, etc.

These issues, migration and borrowing, combined with a lack of documentation and inconsistent terminology, make the study of the genealogical relationships between the New Iranian languages opaque. Originally, the Iranian languages were divided into four geographical distinctions Northwestern, Southwestern, Northeastern, and Southeastern (Schmitt, 1989). These designations were fraught from the beginning, with Northwestern languages like Balochi spoken in the far southeast of the greater Iranian world and Ossetian (NE) spoken in the far northwest. The geographic designation, long-recognized as inadequate, was most recently challenged by Korn, who proposes a Central Iranian core with Bactrian, Sogdian, and Parthian (traditionally NE, NE, and NW) along with the entire Northwestern group (Korn, 2016, 2019). The rest of the Iranian languages form peripheral groups that resist further subcategorization.

In this workshop, we do not make any prescriptions as to historical approaches. Comparative, socio-historical, and computational approaches are to be given equal consideration, as well as multidimensional analyses that combine multiple approaches. The goal of this workshop is to reexamine the validity of previous approaches and established methods as applied to the diachronic study of Iranian languages and, when necessary, to develop new approaches that address the difficulties presented by the unique socio-linguistic situation in the greater Iranian world.

Papers presented in this workshop will focus on:

- Establishing cognacy despite massive borrowing from genetically related languages
- The significance of isoglosses (phonological, morphological, syntactic, and semantic)
- Relation models within the Iranian family
- Waves of contact and migration across time and space in the Iranian world
- The reciprocal influence between Iranian and non-Iranian minority languages
- Innovative methods in historical reconstruction.

Languages represented:

This workshop favors submissions that feature data from and analyses of endangered, minoritized, and understudied languages or those spoken by displaced peoples. Submissions are welcome from all languages with a presence in the greater Iranian world regardless of their genealogy, i.e., papers on Iranian, Neo-Aramaic, Dravidian, Armenian, Turkic, etc. are welcome as long as the paper’s aims match the goals of the workshop.

- Anonby, C. van der W. (2019). Kumzari. In G. Khan & G. Haig (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 625–658). Berlin, Boston: De Gruyter.
- Anonby, E., & Taheri-Ardali, M. (2019). Bakhtiari. In G. Khan & G. Haig (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 445–480). Berlin, Boston: De Gruyter.
- Borjian, H. (2015). Judeo-Iranian Languages. In L. Khan & A. D. Rubin (Eds.), *Handbook of Jewish Languages* (pp. 234–297). Leiden; Boston: Brill.
- Bulut, C. (2019). Iraqi-Turkic. In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 354–384). Berlin: De Gruyter.
- Clackson, J. (2007). *Indo-European Linguistics. An Introduction*. Cambridge: Cambridge University Press.
- Doron, E., & Khan, G. (2012). The typology of morphological ergativity in Neo-Aramaic. *Lingua*, 122(3), 225–240. <https://doi.org/10.1016/j.lingua.2011.11.008>
- Gholami, S. (2011). Definite Articles in Bactrian. In A. Korn, G. Haig, S. Karimi, & P. Samvelian (Eds.), *Topics in Iranian Linguistics (Beiträge zur iranistik 34)* (pp. 11–22). Wiesbaden: Reichert.
- Gholami, S. (2014). Selected features of Bactrian grammar.
- Gholami, S. (2018a). *Endangered Iranian Languages*. Wiesbaden: Reichert Verlag.
- Gholami, S. (2018b). Pronominal clitics in Zoroastrian Dari (Behdīnī) of Kerman. In S. Gholami (Ed.), *Endangered Iranian Languages* (pp. 111–122). Wiesbaden: Reichert.
- Haig, G. (2019a). Northern Kurdish (Kurmanji). In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 106–158). Berlin; Boston: Walter de Gruyter.
- Haig, G. (2019b). The Iranian languages of northern Iraq. In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 267–304). Berlin; Boston: Walter de Gruyter.
- Hock, H. H., & Joseph, B. D. (1996). *Language history, language change, and language relationship : an introduction to historical and comparative linguistics. Trends in linguistics. Studies and monographs*. Berlin ; SE - xv, 602 pages : illustrations ; 24 cm: Mouton de Gruyter. Retrieved from http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=007302059&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA
- Jügel, T. (2016). Enclitic Pronouns in Middle Persian and the Placeholder Construction. In M. J. Dehaghi (Ed.), *Quarterly Journal of Language and Inscription 1/1 [1396 h.š.], dedicated to Professor Mansour Shaki* (pp. 41–63). Tehran.
- Karim, S. O. (2021). *The synchrony and diachrony of New Western Iranian nominal morphosyntax*. the Ohio State University.
- Karim, S. O. (2022). The Ezafe and the Article. In S. Karimi (Ed.), *Advances in Iranian Linguistics II*. Amsterdam: John Benjamins.
- Karim, S. O., & Salehi, A. (2022). An applicative analysis of Soranî “absolute prepositions.” In S. Pacchiarotti & F. Zúñiga (Eds.), *Applicative Morphology: Neglected Syntactic and Non-syntactic Functions [TiLSM 373]* (pp. 263–298). Berlin: De Gruyter Mouton.

- Key, G. (2008). Differential object marking in a Medieval Persian text. In D. L. Stilo, S. Karimi, & V. Samiiian (Eds.), *Aspects of Iranian Linguistics* (pp. 227–248). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Khan, G. (2007). The north-eastern neo-aramaic dialects. *Journal of Semitic Studies*, 52(1), 1–20. <https://doi.org/10.1093/jss/fgl034>
- Khan, G. (2019). the Neo-Aramaic Dialects and their Historical Background. *The Syriac World*, 266–289. <https://doi.org/10.4324/9781315708195-17>
- Korn, A. (2003). Balochi and the Concept of North-Western Iranian. In C. Jahani & A. Korn (Eds.), *The Baloch and their Neighbours: Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times* (pp. 49–60). Wiesbaden: Reichert.
- Korn, A. (2008). A New Locative Case in Turkmenistan Balochi*. *Iran and the Caucasus*, 12, 83–100.
- Korn, A. (2011). Pronouns as Verbs, Verbs as Pronouns: Demonstratives and the Copula in Iranian. In A. Korn, G. Haig, S. Karimi, & P. Samvelian (Eds.), *Topics in Iranian Linguistics (Beiträge zur iranistik 34)* (pp. 53–70). Wiesbaden: Reichert.
- Korn, A. (2016). A partial tree of Central Iranian. *Indogermanische Forschungen*, 121(1), 401–434. <https://doi.org/10.1515/if-2016-0021>
- Korn, A. (2017). The evolution of Iranian. In J. S. Klein, B. D. Joseph, & M. Fritz (Eds.), *Handbook of comparative and historical Indo-European linguistics Band 1* (pp. 608–624). Berlin: De Gruyter Mouton.
- Korn, A. (2019). Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics*, 9(2), 239–281.
- Korn, A., & Olsen, B. A. (2012). On Armenian -agin: additional evidence for a third West Middle Iranian dialect? *Munchener Studien Zur Sprachwissenschaft*, 66(2), 201–220.
- Kümmel, M. J. (2014). The development of laryngeals in Indo-Iranian. In *The Sound of Indo-European*. Opava.
- Mackenzie, D. N. (1961). *Kurdish dialect, studies 1-2. studies 1-2*. London; New York: Oxford University Press.
- Mackenzie, D. N. (1966). *The dialect of Awroman (Hawraman-i Luhon): Grammatical sketch, texts, and vocabulary*. Kobenhavn: Kommissionaer : Munksgaard.
- Mackenzie, D. N. (1999a). Bājalānī (1956). In Daniel Nathan Mackenzie, C. G. Cereti, & L. Paul (Eds.), *Iranica Diversa I* (pp. 409–426). Roma: Istituto italiano per l’Africa e l’Oriente. https://doi.org/10.1163/1573-3912_islam_SIM_1003 LK - <https://osu.on.worldcat.org/oclc/5862823479>
- Mackenzie, D. N. (1999b). Pseudoprotokurtica (1963). In C. G. Cereti & L. Paul (Eds.), *Iranica Diversa II* (pp. 402–402). Roma: Istituto italiano per l’Africa e l’Oriente.
- Mahmoudveysi, P., & Bailey, D. (2013). *The Gorani language of Zarda, a village of West Iran*.
- Mahmoudveysi, P., & Bailey, D. (2019). Hawrāmī of western Iran. In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 533–568). Berlin; Boston: Walter de Gruyter.
- Mahmoudveysi, P., Bailey, D., Paul, L., & Haig, G. (2012). *The Gorani Language of Gawraju, a village of West Iran: texts, grammar, and lexicon*. Wiesbaden: Reichert.
- McKinnon, C. (2011). LORI LANGUAGE i. LORI DIALECTS. In *Encyclopædia Iranica*. Retrieved from <https://www.iranicaonline.org/articles/lori-dialects>

- Noorlander, P. M., & Stilo, D. (2015). On the Convergence of Verbal Systems of Aramaic and its Neighbours. Part I: Present-Based Paradigms. In G. Khan & L. Napiorkowska (Eds.), *Neo-Aramaic in its linguistic context*. Piscataway, NJ: Gorgias Press.
- Opengîn, E. (2016). *The Mukri variety of Central Kurdish : grammar, texts, and lexicon*. Reichert Verlag.
- Öpengin, E. (2019). Accounting for the combinations of clitic and person markers in Central Kurdish. In S. Gündoğdu, E. Öpengin, G. Haig, & E. Anonby (Eds.), *Current Issues in Kurdish Linguistics*. Bamberg: University of Bamberg Press.
- Paul, L. (1998). The position of Zazaki among West Iranian languages. *Proceedings of the Third European Conference of Iranian Studies: Held in Cambridge, 11th to 15th, 1995: Part 1 Old and Middle Iranian Studies (Beiträge Zur Iranistik 17)*, 163–177.
- Paul, L. (2003). The Position of Balochi Among the Western Iranian Languages: The Verbal System. In C. Jahani, A. Korn, & P. Titus (Eds.), *The Baloch and Their Neighbors: Ethnic and Linguistic Contact in Balochistan in Historical and Modern Times* (pp. 61–71). Wiesbaden: Reichert.
- Paul, L. (2019). Persian. In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 569–624). Berlin; Boston: Walter de Gruyter.
- Schmitt, R. (1989). *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.
- Sims-Williams, N. (2011). Differential Object Marking in Bactrian. In A. Korn, G. Haig, S. Karimi, & P. Samvelian (Eds.), *Topics in Iranian Linguistics (Beiträge zur Iranistik 34)* (pp. 23–38). Wiesbaden: Reichert.
- Skjærvø, P. O. (2017). The documentation of Iranian. In J. S. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of comparative and historical Indo-European linguistics Band 1* (pp. 471–480). Berlin: De Gruyter Mouton.
- Stilo, D. (2008). Case In Iranian: From Reduction and Loss to Innovation and Renewal. In A. Spencer & A. L. Malchukov (Eds.), *The Oxford Handbook of Case* (pp. 700–715). Oxford; New York: Oxford University Press.
- Stilo, D. (2018a). Dikin Marāqei Tati of Alamut: an undocumented conservative Tati language. In S. Gholami (Ed.), *Endangered Iranian Languages* (pp. 41–70). Wiesbaden: Reichert.
- Stilo, D. (2018b). The Caspian region and South Azerbaijan: Caspian and Tatic. In G. Haig & G. Khan (Eds.), *The languages and linguistics of western Asia: An areal perspective* (pp. 659–829). Berlin, Boston: Walter de Gruyter.
- Stilo, D., & Noorlander, P. M. (2015). On the Convergence of Verbal Systems of Aramaic and its Neighbors. Part II: Past Paradigms Derived from Present Equivalents. In G. Khan & L. Napiorkowska (Eds.), *Neo-Aramaic in its linguistic context*. Piscataway, NJ: Gorgias Press.
- Wendtland, A. (2011). The Emergence and Development of the Sogdian Perfect. In A. Korn, G. Haig, S. Karimi, & P. Samvelian (Eds.), *Topics in Iranian Linguistics (Beiträge zur Iranistik 34)* (pp. 39–52). Wiesbaden: Reichert.
- Werner, B. (2018). Forms and Meanings of the Ezafe in Zazaki. In S. Gholami (Ed.), *Endangered Iranian Languages* (pp. 71–91). Wiesbaden: Reichert Verlag.
- Windfuhr, G. (1991). CENTRAL DIALECTS. In *Encyclopaedia Iranica* (pp. V, 3, 242–252). Retrieved from <https://www.iranicaonline.org/articles/central-dialects>
- Windfuhr, G. L. (2012). FĀRS viii. Dialects. In *Encyclopædia Iranica* (p. Vol. IX, Fasc. 4, 362-373). Retrieved from <https://iranicaonline.org/articles/fars-viii>

Bactrian influence on local languages of Eastern Afghanistan

Julian Kreidl, Indiana University Bloomington

While Bactrian has no modern descendants, it has left its traces in local languages of Eastern Afghanistan. Unlike potential Bactrian loanwords in Persian (e.g. Lurje, Yakubovich 2017) or Tocharian (e.g. Tremblay 2005: 435-437), the issue of Bactrian influence on Pamir languages or Pashto have received less or no attention. In many cases, due to the relatively close genetic relationship of the involved languages, differentiating shared inherited features and borrowings is a difficult task. For example, it is hard to tell whether Pashto *walwár* ‘bride price’ should be considered a genuine outcome of **wadū-bāra-* or a loan from Bactrian (ολοβαρο) because both would be possible phonologically (Cheung 2015: 57). But I argue that the situation is different regarding, for example, Pashto *γunǰ*, *γwunǰ* ‘bag’ and Bactrian *γωνζο*, *γονζο* ‘bag, sack’. Sims-Williams 2007: 207 derives the Bactrian term from **gaunīčiya-* (cf Sanskrit *goṇī-*, Gandhari *goni*, Khotanese *gūñā-*). While the Pashto word may in origin also go back to **gaunī-čiya-*, the Pashto form is puzzling because one would rather expect **γinj*. In the sequence **-auCī-* (as in **gaunī-*), the final **ī* would lead to umlaut of the preceding vowel, as in Pashto *wína* < **win* (+ secondary *-a*) < **wauni-* < *wahuni-* ‘blood’. Old Iranian **č*, on the other hand, should yield *j* [dz], not *ǰ* [dʒ], in Pashto. While there is occasional umlaut also in Bactrian, it is due to a lack of examples unclear if this also affects lexemes of the shape **-auCī-*. Old Iranian **č* yields, depending on the environment, *σ* or *ζ* in Bactrian. Both the Graeco-Bactrian Sigma and the Zeta represent more than one phoneme, and without keeping in mind the etymology, *γωνζο* could be interpreted as [γo:ndz], [γo:nʒ] or [γo:ndʒ]. But the front vowel following **č* in **gaunī-čiya-* makes it likely that *ζ* stood for either [ʒ] or [dʒ] here, represented in Pashto *γ(w)unǰ*, a loan from Bactrian.

A Bactrian feature of a different kind which spread into other local languages is the lambdacism **d > *δ > l*. It is found in Munji, Yidgha, Pashto and the Nuristani language Prasun (Kreidl 2021: 176-184). While this makes identifying Bactrian loanwords even harder in languages which participated in the lambdacism, it is, on the other hand, facilitating the search for Bactrianisms in closely related languages which did not. Therefore, I suggest that, e.g., Wakhi *liv*, *liw* ‘cannibal giant; crazy’ and Sanglechi *lēw* ‘demon; madman’ (Steblin-Kamenskij 1999: 225, Morgenstierne 1973: 401), cautiously considered loanwords from Munji by Morgenstierne *ibid*, should be taken as borrowings from Bactrian, a language far more prestigious than Munji. Similarly, Wakhi *məlúng* ‘middle’ < **madana-ka-* and *vul* ‘smell’ < **bauda-* (Steblin-Kamenskij 1999: 237, 383) may likewise be from Bactrian.

In my contribution, I plan on presenting further evidence for Bactrian loanwords in the Pamir languages and Pashto, as well as Nuristani and Dardic, shedding light on the complex relationship of the Eastern Iranian languages to each other.

References

Cheung, Johnny. 2015. “The linguistic and geographic position of Pashto”, In: de Chiara, Matteo; Grassi, Evelin (Eds). *Iranian Languages and Literatures of Central Asia*, 29-65. Leuven: Peeters.

- Kreidl, Julian. 2021. "Lambdacism and the development of Old Iranian *t in Pashto", *Iran and the Caucasus*, 25/2, 175-193.
- Lurje, Pavel; Yakubovich, Ilya. 2017. "The Myth of Sogdian Lambdacism", In: Team „Turfanforschung“ (Hrsg). *Zur lichten Heimat : Studien zu Manichäismus, Iranistik und Zentralasienkunde im Gedenken an Werner Sundermann*, 319-341. Wiesbaden: Harrassowitz.
- Morgenstierne, Georg. 1973. *Indo-Iranian Frontier Languages, Vol. II: Iranian Pamir Languages: Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi*. 2nd ed. Oslo: Universitetsforlaget.
- Sims-Williams, Nicholas. 2007. *Bactrian Documents from Northern Afghanistan II: Letters and Buddhist Texts*. London: The Nour Foundation.
- Steblin-Kamenskij, I. M. 1999. *Ètimologičeskij slovar' vaxanskogo jazyka*. Sankt-Peterburg: Peterburgskoe Vostokovedenie.
- Tremblay, Xavier. 2005. "Irano-Tocharica et Tocharo-Iranica", *Bulletin of the School of Oriental and African Studies*, 68/3, 421-449.

Steppe Iranian in the *longue durée*: contact, relative chronology, and internal reconstruction

Ronald I. Kim, Adam Mickiewicz University n Poznań

For over a millennium, from c. 900 BC until the early centuries AD, the Eurasian steppe zone from the Pontic region to eastern Central Asia was home to numerous Iranian-speaking peoples whose names and movements are known from the testimony of neighboring civilizations, primarily Greco-Roman and Persian. These historical records, along with the rich archaeological evidence of burial sites from the Danube to the Altai, revealed that groups continuously migrated — generally from east to west and often over considerable distances—linguistic variation. It thus comes as little surprise that despite enormous advances in Iranian philology over the past 50 years, our knowledge of the linguistic history of Steppe Iranian has not progressed greatly beyond that of such seminal works as Abaev (a949) or Harmatta (a970).

Recent developments herald a welcome change, however, such as the appearance of two new studies of Iranian loanwords in Tocharian (Dragoni 2022, Bernard 2023). As the source of the earliest such Iraniana borrowings in Tocharian, Bernard posits an “Old Steppe Iranian” spoken in the Altai region and Dzhungaria, though a location in eastern Central Asia is also imaginable. Of the features ascribed to this “Old Steppe Iranian,” the appearance of [l] for OIr. *r before dentals (cf. TB *melte* ‘pile’, TA *malto* ‘in first place’ ← OIr. *marda- ‘head, top of the body’; TB *speltke*, TA *ratäk* ‘army’ ← OIr. *rata-ka- ‘line, formation’) contrast with Ossetic, where OIr. *r is usually retained except before *l or *y and apocope preceded syncope (Cheung 2002:69-85). Otherwise it exhibits few innovations, corresponding to the meager Scythian evidence (Mayerhofer 2006). The one alleged defining trait of Scythian, the shift of OIr. [ð] > [l] in the name Παράλαται < OIr. *para-dāta-, is not in fact probative but could simply represent an attempt by Greek speakers to render the voiced interdental fricative [ð]; this would square with recent arguments against lambdacism in Sogdian (Lurje & Yakubovich 2017).

It is only from the Sarmatian period that the defining phonological changes ancestral to Ossetic such as voicing of intervocalic stops or palatalization of *ti > *dʲ > [dz] vel sim. are reflected in the extensive onomastic material (see most recently Palunčić 2019). Importantly, the ethnonym Ἀορσοί, whose etymological connection with Oss. D *ors*, I *urs* ‘white’ and OIr. *aruša- has long been debated, confirms the early syncope of *u required by D *ford*, I *furd* ‘great river’ < OIr. *paruta- and D *mex*, I *mix* ‘stake’ < OIr. *mayuṣṣa-.

The absence of connected texts greatly hampers investigation of morphological developments, so that e.g. although the collective suffix *-tā- is known from ethnonyms recorded as far back as Herodotus (Sauromatai/Syrmatai, Thyssagetai, Massagetai, Iaxamatai/Ixibatai), one cannot know when it became generalized as the productive plural formant. Here it is historical-comparative investigation of Ossetic grammar that can offer some guide to the chronology of prehistoric changes. For instance, the Oss. Periphrastic future in -ʒVn- (e.g. D *cær-ʒæn-æ̃n*, I *cær-ʒyn-æ̃n* ‘I will live’) must have its origin in nominal compounds *X-čānāh ‘desiring X_N’ (whence deverbal ‘(be) wanting to X_V’ > ‘X_V-FUT’; Kim fthc. A); given the derivational isolation of *-čānāh- in Iranian, this construction must have evolved already in OIr. Times. Another example

is the Oss. Transitive preterite, which with Christol (1990: 43-4) goes back to a periphrasis of past participle + *dā- ‘make.’ As simplex *dā- ‘put’ was already becoming rare in OIr., this construction is likely to be an innovation of the late Sarmation or early Alanic period (Kim fthc. B); the formal resemblance to the Germanic dental (weak) preterite is suggestive and raises the possibility of contact-induced change, but extralinguistic evidence for sufficiently early contacts is so far lacking.

Abbreviations: D = Digor; I = Iron; OIr. = Old Iranian; Oss. = Ossetic; TA, TB = Tocharian A, B.

References:

- Abaev, V. I. [Абаев, В. И.]. 1949. *Осеминский язык и фольклор*. Том I. Москва/Ленинградъ Издательство Академии Наук СССР.
- Bernard, Chams. 2022. *Like Dust on the Silk Road. An Investigation of the Earliest Iranian Loanwords and of Possible BMAC Borrowings in Tocharian*. Dissertation, Leiden University.
- Cheung Johnny. 2002. *Studies in the Historical Development of the Ossetic Vocalism*. Wiesbaden: Reichert.
- Christol, Alain. 1990. Introduction à l’ossète: éléments de grammaire compare. LALIES: *Actes des sessions de linguistique et de littérature* 8, 7-50.
- Dragoni, Federico. 2022. *Watañi lāntam. Khotanese and Tumshuqese loanwords in Tocharian*. Dissertation, Leiden University.
- Harmatta, J[ános]. 1970. *Studies in the History and languages of the Sarmatians*. (Acta Universitatis de Attila Jósef Nominatae, Acta Antiqua et Archaeologica, XIII.) Szeged: Szegedi Nyomda.
- Kim, Ronald I. Forthcoming a. The Indo-Iranian background of the Ossetic future. Submitted to *Folia Linguistica Historica*.
- Kim, Ronald I. Forthcoming b. The Ossetic transitive preterite: typology, evolution, contact.
- Lurje, Pavel and Ilya Yakubovich. 2017. The myth of Sogdian lambdacism. In *Zur lichten Heimat: Studien zu Manichäismus, Iranistik und Zentralasienkunde im Gedenken an Werner Sundermann*, hrsg. Von einem Team Turfanforschung”. Wiesbaden: Harrassowitz, 319-41.
- Mayrhofer, Manfred. 2006. *Einiges zu den Skythen, ihrer Sprache, ihrem Nachleben*. (Sitzungsberichte der Österreichischen Akademie der Wissenschaften, Philosophisch-historische Klasse, 742. Veröffentlichungen der Iranischen Kommission, 36.) Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Palunčić, Filip. 2019. Ossetic historical phonology and North-Eastern Iranian anthroponomastics from the North Pontic region, 1st-5th c. CE. In Pavel B. Lurie (ed.), *Proceedings of the Eighth European Conference of Iranian Studies*. Vol. I: *Studies on Pre-Islamic Iran*. St. Petersburg: State Hermitage Publishers, 311-29.

Remarks on the category of copula in Gorani dialects

Masoud Mohammadirad, Cambridge University

Keywords: copula | verbalization | person | reanalysis | syncretism

Gorani dialects show considerable variation in the formation and derivation of the present copula paradigm. This paper examines these variations across 10 Gorani dialects. The material was gathered from available grammatical descriptions, and a recent questionnaire developed for studying morphosyntactic and phonological variation within Kurdish. Unlike most modern Iranian languages, the copula paradigm in Gorani consists of the element *(a)n-* to which person forms are added. This element can be reconstructed as an erstwhile 3SG *-n* preceded by the stem *ha-*. It will be argued that the paradigm of the enclitic copula in modern dialects is the result of the reanalysis of morphologically coded 3SG inflection as part of the stem, in line with the trend in historical change cross-linguistically (Watkins 1962; Koch 1995). This paradigm is generally attested in modern dialects, except for Gawrajui which has replicated the Kurdish pattern of enclitic copula.

| (1) | Orthotone copula | | |
|-----|------------------|-----------------|------------------|
| | | Before | After reanalysis |
| | 1SG. | * <i>ha-ā</i> | <i>han-ā</i> |
| | 2SG. | * <i>ha-ī</i> | <i>han-ī</i> |
| | 3SG.M | * <i>ha-n</i> | <i>han-∅</i> |
| | 3SG.F | * <i>ha-n-a</i> | <i>han-a</i> |
| | 1PL. | * <i>ha-mē</i> | <i>han-mē</i> |
| | 2PL. | * <i>ha-dē</i> | <i>han-dē</i> |
| | 3PL. | * <i>ha-ē</i> | <i>han-ē</i> |

Another source of variation concerns the derivation of the copula paradigm. Most Gorani dialects are characterised by deriving certain cells of the copula paradigm, most notably third person and 1SG, from the demonstrative pronouns, a profile which was probably developed under long-standing contact with Semitic languages, e.g., Neo-Aramaic (Khan 2022). In some dialects 1PL and 2PL are derived from the paradigm of oblique clitics.

Yet another source of variation is the assimilation of the enclitic copula paradigm to that of the verbal person suffixes of present tense verbs. Here, the dialects are distributed on a continuum, where one end is characterised by a four-way distinction of person forms in the two paradigms (attested in Kandulai), whereas the other end is distinguished by the complete verbalization of the copula paradigm (attested in Gawrajui).

3rd person and 1st person > 3rd person and 1SG > 3rd person only > 3SG and 2PL > no distinction

References

- Khan, Geoffrey. 2022. The change in the grammatical category of the copula in North-Eastern Neo-Aramaic: *Journal of Historical Linguistics*. John Benjamins Publishing Company 12(3). 446–475. <https://doi.org/10.1075/jhl.21019.kha>.
- Koch, Harold. 1995. The Creation of Morphological Zeroes. In Booij, Geert and Jaap van Marle (eds.) *Yearbook of Morphology 1994*, pp. 31–71. Dordrecht: Kluwer.
- Watkins, Calvert. 1962. *Indo-European Origins of the Celtic Verb*. 1. The Sigmatic Aorist. Dublin: Dublin Institute for Advanced Studies.

Polyptoton for the purpose of emphasizing within Iranian languages

Saloumeh Gholami, Goethe University Frankfurt

Zaniar Naghshbandi, University of Kurdistan

Polyptoton, basically defined as the repetition of a word in different inflected forms, is originally a rhetorical stylistic device that appears usually in literary genres. The figure is therefore similar to the *figura etymologica*. Polyptoton was a common facet of Latin and Greek poetry, however, modern literature shows also examples of this structure.

There is a special type of polyptoton in different periods of Iranian languages, where an inflected verb is used with another word sharing the same root. Both elements are used in a sentence for the purpose of emphasizing an emotion or idea and highlighting a deeper meaning in the text.

In his article on “Maʿūl-e moṭlaq dar zabān-e Fārsī” [absolute object in Persian language], Molayi (2002) presents some of these constructions in early New Persian texts under the title of absolute object and criticizes the scholars who consider it as an Arabic influence on Persian.

It seems that the examples of this kind of polyptoton are attested at least in one New Iranian spoken language. In their article, Karimi and Naghshbandi (2011) discuss Emphatic Progressive Verbal Constructions in Hawrami. In Hawrami, there is a special type of polyptoton forming progressive aspect and emphasizing the verb. The construction is composed of two conjoined parts: the infinitive plus present continuous, past continuous, and simple past verbs; nothing but agreement clitics (either subject-referring clitics in ergative constructions or object-referring clitics in non-ergative constructions) can separate these two parts:

1. ʔæmən wetiaj mæ-s-u
 I to sleep IPRF-sleep.PRS-1SG
 “I am sleeping” or “I am on the edge of falling asleep”
2. ʔemæ sipatæke=man æs-e=ne
 we clothes=1PL buy.PST-3PL=tobe.1PL
 ʔistæ ʃordəj=ʃan mæ-ʃor-me
 now to wash=3PL IMPRF-wash.PST-1PL
 “We have bought the clothes. Now we are washing them.”

However, it is worth mentioning that varieties of Hawrami differ slightly as to how they form the first constituent of this specific construction.

Drawing on data taken from Avesta, Old Persian, Middle Persian and the Pavei variety of Hawrami, this presentation seeks to examine the specific type of polyptoton within these languages. An important question arises: whether the emphatic progressive verbal constructions in Hawrami can be viewed as an archaic feature that originally goes back to the Old Iranian period?

Bibliography:

Karimi, Yadegar, Naghshbandi, Zaniar (2011): “Emphatic Progressive Verbal Constructions in Hawrami.” In *Language Research*. 2. 83-100.

Molayi, Changiz (2002): “Maf’ūl-e moṭlaq dar zabān-e Fārsī” [absolute object in Persian language].” In: *Našrīe-ye dāneškade-ye adabīyāt va ’lūm-e ensānī*. 185.95-102

Semantic Shift and Morphosyntactic Convergence of Tense-Aspect-Mood Categories in Alazan Persian

Murad Suleymanov, Université Paris Sciences et Lettres (PSL) – École Pratique des Hautes Études (EPHE) – Institut des langues rares ILARA) / Institut des langues et civilisations orientales (INALCO) / UMR 7192 “Proche-Orient–Caucase : Langues, Archéologie, Cultures”, Paris

“Southwestern” Iranian languages spoken in the Caucasus have long been known to be represented solely by Tat varieties (Grjunberg 1963, Hacıyev 2009, Authier 2012, Suleymanov 2020). A field mission undertaken in summer 2021 in the Alazan Valley, in the very north of the Republic of Azerbaijan, revealed a hitherto undescribed Iranian variety spoken in the area. Unlike Tat, which, albeit closely related to Persian, is not mutually intelligible with it and shows significant grammatical differences, the Iranian variety of the Alazan Valley can be safely classified as a New Persian dialect. The speech community inhabits half a dozen villages scattered across the Districts of Balakən and Qax (and possibly also found in neighbouring Georgia) and claims descent from late nineteenth- and early twentieth-century immigrants from Persia. There are at least two distinct but mutually intelligible sub-varieties of Alazan Persian (one per district), and the villages maintain active contact with one another.

All Alazan Persian (henceforth AlzP) speakers in Balakən and Qax are bilingual in Azeri, the majority language belonging to the Turkic family and the official language of Azerbaijan. Although there is some tendency for syntactic restructuring as a result of contact, e.g. gradual loss of prepositions (more so than in Tehran Persian), personal clitics reduced to possessive function only, partial suppletion of the paradigms of the verbs *bidān* (cognate of Standard Persian *budan* ‘to be’) and *šidān* (cognate of Standard Persian *šodan* ‘to become’), AlzP does not show novel contact-induced tense-aspect-mood (TAM) categories as do some other Turkic-influenced “Farsic” varieties (Soper 1987). This may certainly be due to a shorter period of contact in comparison to Azeri–Tat and Uzbek–Tajik contact situations.

Instead, AlzP demonstrates different patterns of morphosyntactic convergence of inherited grammatical TAM categories across the two varieties, as seen in (1–2).

- | | |
|---|---|
| (1) Balakən sub-variety | (2) Qax sub-variety |
| <p>a. <i>mān kitab bu-xun-um.</i> I book IPFV-read₁-1SG ‘I am reading a book. / I read books.’</p> <p>b. <i>ägär xeyli gäp bi-zān-um</i> if much word IPFV-hit₁-1SG <i>mān=ä jārīmā bu-kun-id.</i> I=DDO fine IPFV-do₁-3 ‘If I talk too much, he is (definitely) going to fine me.’</p> <p>c. <i>ägär vaxt=im bi-šid</i> if time=POSS:1SG IPFV-be₁.3 <i>kitab=ä mu-xun-um.</i> book=DDO EVT-read₁-1SG ‘If I (hypothetically) have time, I will read the book.’</p> | <p>a. <i>nun=mun=ä mu-xor-än.</i> bread=POSS:1PL=DDO IPFV-eat₁-3PL ‘They eat / are eating our bread.’</p> <p>b. <i>umru borun bə-riz-id.</i> today rain MOD-flow₁-3 ‘Today it is going to rain.’</p> <p>c. <i>ayri bi-šin-äd</i> separate MOD-sit₁-3 <i>ayri mi-šin-äd.</i> separate IPFV-sit₁-3 ‘If he lives apart, he lives apart (and if he does not live apart, he lives with us).’</p> |

The field data illustrates both varieties having a definite/prospective future (1b & 2b), which contrasts with an indefinite/hypothetical future (1c & 2c, glossed as EVT for “eventual”). The prospective category is identical with the subjunctive (shown in conditional contexts in the examples but found elsewhere in the same form), both having the form <bi- + present stem>. In addition, in the Balakən sub-variety, this same category has extended into the present domain (1a), marginalizing the inherited present-future construction <mi- + present stem> into the domain of indefinite/hypothetical future. The Qax sub-variety shows both present and future uses of <mi- + present stem>, similar to Standard Persian, but the latter use is only limited to indefinite/hypothetical future. The typologically common phenomenon of presents grammaticalizing into modal categories such as subjunctives or futures, is not rare in West Asia, including the Iranian Plateau and the South Caucasus (Haspelmath 1998). The eventual vs. prospective future split exists, notably, in most Tat varieties, and, similarly to the Balakən sub-variety of AlzP, in all of them the old present (cognate of the Persian <mi- + present stem> construction) today acts mainly

as a future tense. Cases of subjunctives developing into futures are not uncommon either, with Latin being a notable example (Clackson & Horrocks 2011: 24–25).

In the case of AlzP, the processes by which the constructions <mi- + present stem> and <bi- + present stem> have come to be aligned as they presently are deserve an analysis with a focus on diachrony.

The stability of <mi- + present stem> as a hypothetical future construction is unsurprising given that the semantics it originally conveyed in Persian had largely ceased to be associated strictly with progressivity by the late nineteenth century and became generalised as the gnomic present, yielding also an indefinite/hypothetical future reading (both referring to what “generally expected to happen”). The Balakən sub-variety preserves the latter use while Qax sub-variety preserves both.

The behaviour of <bi- + present stem> is less obvious. Lenepveu-Hotz (2014) traces the development of the verbal prefix bi- from being a mood-independent marker of rhematicity to becoming a modal (subjunctive) marker, change which she dates to the late nineteenth / early twentieth century, i.e. to the time when modern AlzP speakers claim their ancestors left Persia. The rhematic property of bi- could thus quite easily account for the development of <bi- + present stem> (originally conveying a focal action / state in the present) into a prospective category, especially in light of similar semantics being attested in Classical Persian (Jahani 2008: 160) and found in modern languages of the Central Iranian Plateau as a “close future” (Korn 2020: 479, Tāheri 2021). Furthermore, AlzP, or at least its Qax sub-variety, seems to have reinterpreted the focal nature of bi- as “perfective”, which is Haspelmath (1998: 55) considers a common property of futures and subjunctives, allowing it to extend <bi- + present stem> to both the prospective/definite future and subjunctive domains.

The remarkable use of <bi- + present stem> for the general present tense in the Balakən sub-variety represents perhaps a slightly different development process. One can hypothesise that bi- never developed into an aspectual marker in this sub-variety and remained purely focal. When the function of <mi- + present stem> as a present category starting weakening and the construction started drifting towards marking the indefinite/hypothetical future (at it happened in Tat), there arose a need to replace it with a more semantically dynamic construction, and a rhematic construction <bi- + present stem> made for a good replacement. In this respect, the semantic distinction between the present, the prospective future and the subjunctive was irrelevant, and the situation, at least by the time of the earliest speakers' arrival in the Alazan Valley, resembled very much that of pre-modern Persian.

The scope of this paper is limited to presenting and briefly analysing (including within a broader regional context) preliminary data from a peculiar variety of Persian developing outside of its traditional area. A separate study aimed at tracing the origin of AlzP and the movement of its earlier speakers could offer additional clues regarding these changes.

References

- Authier, Gilles (2012). *Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est*. Beiträge zur Iranistik 36 / Bibliothèque iranienne 76. Wiesbaden: Reichert.
- Clackson & Horrocks (2011). *The Blackwell History of the Latin Language*. Chichester & Malden (MA): Wiley-Blackwell.
- Grjunberg, Aleksandr (1963). *Jazyk severoazerbajdžanskix tatov* [The Language of the Tats of Northern Azerbaijan]. Leningrad: Izdatel'stvo Akademii nauk SSSR.
- Hacıyev, Maqsud (2009). *Azərbaycan tatlarının dili / Tatların tarixi-etnoqrafik öçerki* [The Language of Azerbaijani Tats / A Historical and Ethnographic Sketch of Tats]. Baku: Mütərcim.
- Haspelmath, Martin (1998). “The semantic development of old presents: New futures and subjunctives without grammaticalization”. In: *Diachronica* 15, pp. 29–62.
- Jahani, Carina (2008). “Expressions of future in Classical and Modern New Persian”. In: Karimi, Simin & Samiian, Vida & Stilo, Donald, eds. *Aspects of Iranian Linguistics*. Newcastle: Cambridge Scholars Publishing, pp. 155–176.
- Korn, Agnes (2020). “Grammaticalization and reanalysis in Iranian”. In: Bisang, Walter & Malchukov, Andrej. *Grammaticalization Scenarios: Cross-linguistic variation and universal tendencies 1: Grammaticalization Scenarios from Europe and Asia*. Comparative Handbooks of Linguistics 4. Berlin/Boston: de Gruyter, pp. 465–498.

- Lenepveu-Hotz (2018). “Evolution of the subjunctive in New Persian (10th–20th): From disappearance to reappearance”, *Folia Linguistica Historica* 39/2, pp. 421–440.
- Soper, John (1987). “Loan Syntax in Turkic and Iranian: The Verb Systems of Tajik, Uzbek, and Qashqai”. Ph.D. dissertation, University of California at Los Angeles.
- Suleymanov, Murad (2020). *A Grammar of Şirvan Tat*. Beiträge zur Iranistik 46. Wiesbaden: Reichert.
- Tāheri, Esfandiār (2021). “Dar bāre-ye sāxt va kār kard-e sāxthā-e āyandasāz dar jowšaqāni” [On the form and function of the future in Jowshaqani]. In: *Journal of Researches in Linguistics* 13/1, pp. 21–40.

**“Your birch-bark bag has something” –
Grammaticalization and diachrony of locative, existential and possessive predication**

Chris Lasse Däbritz (University of Hamburg)
chris.lasse.daebritz@uni-hamburg.de

It is widely known that locative, existential and possessive predications are closely related in many languages of the world (see Lyons 1967, Clark 1978, Freeze 1992, Hengeveld 1992, Koch 2012). In what follows, I conceive locative and existential predications as expressing the temporary presence or absence of a figure (a.k.a. theme, pivot) in a ground (a.k.a. location, coda), their difference lying in perspectivization (Hengeveld 1992: 94–100; Creissels 2019: 37). The prototypical instances of locative and existential predications are clauses like (1a) and (1b), respectively. In turn, predications which either express the permanent presence/absence of a certain referent (1c) or lack a specified location (1d) represent a different, though often formally similar, type of predication. Following Koch (2012), I call the former *bounded existentials* and the latter *generic existentials*. Possessive predication expresses an asymmetric and usually unidirectional relation of two entities, the possessor and the possessee, whereby the possessee belongs to the possessor (1e).

- | | | |
|------|---------------------------------|-----------------------|
| (1a) | The book is on the table. | (LOCATIVE) |
| (1b) | There is a book on the table. | (EXISTENTIAL) |
| (1c) | There are many lions in Africa. | (BOUNDED EXISTENTIAL) |
| (1d) | There are many unhappy people. | (GENERIC EXISTENTIAL) |
| (1e) | Bill has a book. | (POSSESSIVE) |

Given the overlap of the functional domains expressed, it is not surprising that many languages use similar or even the same linguistic structures to express the predications of type (1a) to (1e) (Hengeveld 1992: Ch. 5.1.3; Heine 1997: Ch. 2). Whereas this can be described on a synchronic level from various perspectives, it has also diachronic implications given that languages evolve during time and linguistic structures may spread from one functional domain to another.

Within the realm of possessive predication, the grammaticalization of so-called *habeo*-verbs is a classical instance. They often have their lexical source in verbs like *get*, *grab*, *take*, *obtain*, *hold*, *carry* or alike, as e.g. the Dullay (< Eastern Cushitic < Afro-Asiatic) verbal root *-sheeg-* ‘have; carry on one’s head or shoulder’ or the Khanty (< Uralic) verb *taj-* ‘have; hold; carry’ (Heine 1997: 47–48; Honti 2008: 172). Additionally, as shown by Koch (2012: 572–575) and Creissels (2019: 70–76), *habeo*-verbs can appear in existential clauses, like in Greek (< Indo-European) (2); the distinguishing criterion of a possessive (2a) and existential (2b) reading is the locative coding of the “possessor” in (2b). The Mansi (< Uralic) example (3) shows a sentence, structurally ambiguous between the two readings, but the semantics of the “possessor” rather favour an existential reading.

- | | | | | | |
|------|-------------------------------------|----------|-----|--------------|--------------|
| (2a) | Ta | chōriá | den | échoun | dáskalous. |
| | the | villages | NEG | have.PRS.3PL | teachers.ACC |
| | ‘The villages don’t have teachers.’ | | | | |

(2b) *Den eiche dáskalous sta chōriá.*
 NEG have.PST.3SG teachers.ACC in.the villages
 ‘There were no teachers in the villages.’
 (Greek (< Indo-European); Creissels 2019: 71)

(3) *Pajp-ən matər o:nsi-i.*
 birchbark.bag-POSS.2SG something have-PRS.3SG
 ‘There is something in your birch-bark bag.’ ~
 ?‘Your birch-bark bag has something.’
 (Mansi (< Uralic); Kannisto & Liimola 1956. OUDB Northern Mansi Corpus. Text ID 1235, 211)

Besides that, existential predications of the type (1b) show a wide variation of potential source structures, as shown by Creissels (2019). E.g., Icelandic (< Indo-European) shows a construction, which formally resembles identificational clauses (4a). In Nganasan (< Uralic), a similar construction seems to have developed further on the grammaticalization pathway: Existential clauses are formed with the existential verb *təisʹa*, lexicalized from the combination of the demonstrative stem *tə-* and the copula verb *isʹa* (Wagner-Nagy 2019: 354; example 4b).

(4a) *Það eru mys í baðkerinu.*
 that are mice in bathtub
 ‘There are mice in the bathtub.’ (lit. ‘That are mice in the bathtub.’)
 (Icelandic (< Indo-European); Creissels 2019: 79)

(4b) *tahariábə təndə siiti bəŋgüʔtiə təi-čü.*
 now there two burrow EX-AOR.3SG
 ‘Now, there are two burrows.’ (< lit. ‘Now, that is two burrows there.’)
 (Nganasan (< Uralic); Wagner-Nagy 2019: 355)

Finally, Hengeveld (1992: 238–240), Newman (2002) and Ameka & Levinson (2007), among others, account for the grammaticalization of posture verbs like *stand*, *sit*, *lie* as copula elements in locative and existential predication. As a case in point, Mbay (< Nilo-Saharan) uses, among others, the posture verb *tən* ‘lie’ in existential clauses (5).

(5) *mbētē li-i lā tən.*
 book POSS-you LOC lying
 ‘Here is your book.’
 (Mbay (< Nilo-Saharan); Newman 2002: 10, cit. from Keegan 1997: 76)

This non-exhaustive sketch already shows that many synchronically observed overlaps in the realm of locative, existential and possessive predication are connected to diachronic developments. Having in mind the similar, if not identical, underlying semantic structure of the discussed predication types, this does not surprise.

The aim of this workshop is to bring together researchers working on various aspects of the named functional domain and to discuss the role of diachrony and grammaticalization processes within it. Therefore, contributions may take any theoretical perspective and deal with single languages or work cross-linguistically, granted that they somehow acknowledge the diachronic perspective of the conference. Additionally, it is desirable that the presented work relates to the

theoretical understanding of locative, existential and possessive predication. Finally, the contributions to this workshop shall not interfere with eventual contributions to the SLE workshop on core and periphery in locative and existential predication. The accepted abstracts cover various aspects of the discussed domains. Two of them are more theoretical in nature, whereas the other two are rather case studies dealing with Semitic and Indo-European languages, respectively. All of them discuss relevant co-expression patterns and aspects of their diachronic development; one abstract additionally targets negative structures in the discussed domains.

References:

- Ameka, Felix K. & Levinson, Stephen C. 2007. Introduction: The typology and semantics of locative predicates: posturals, positionals, and other beasts. *Linguistics* 45(5/6), 847–871.
- Clark, Eve 1978. Locationals: Existential, locative and possessive constructions. In: Clark, Eve & Greenberg, Joseph (eds.). *Universals of human language. Vol. 4. Syntax*, 85–126. Stanford [Ca.]: Stanford University Press.
- Creissels, Denis 2019. Inverse-locational predication in typological perspective. *Italian Journal of Linguistics* 31(2), 37–106.
- Freeze, Ray 1992. Existentials and other locatives. *Language* 68(3), 553–595.
- Heine, Bernd 1997. *Possession. Cognitive sources, forces and grammaticalization*. Cambridge: Cambridge University Press.
- Hengeveld, Kees 1992. *Non-verbal predication: Theory, typology, diachrony*. (Functional Grammar Series 15). Berlin & New York: Mouton de Gruyter.
- Honti, László 2008. Habere ‘po-uralski’ [Habere ‘in Uralic’]. *Linguistic Uralica* 44(3), 161–177.
- Kannisto, Artturi & Liimola, Matti 1956. *OUDB Northern Mansi Corpus*. Text ID 1239. Ed. by Janda, Gwen Eva. <http://www.oudb.gwi.uni-muenchen.de/?cit=1239>. (Accessed on 2022-10-01).
- Keegan, John M. 1997. *A reference grammar of Mbay*. Munich & Newcastle: Lincom Europa.
- Koch, Peter 2012. Location, existence and possession: A constructional-typological exploration. *Linguistics* 50(3), 533–603.
- Lyons, John 1967. A note on possessive, existential and locative sentences. *Foundations of Language* 3(4), 390–396.
- Newman, John 2002. A cross-linguistic overview of the posture verbs ‘sit’, ‘stand’, and ‘lie’. In: Newman, John (ed.) 2002. *The linguistics of sitting, standing, and lying*. (Typological Studies in Language 51), 1–24. Amsterdam & Philadelphia: John Benjamins.
- Wagner-Nagy, Beáta 2019. *A Grammar of Nganasan*. Leiden: Brill.

Negated but similar - Negation in the domains of locative, existential, and possessive predication: The case of Indo-European.

(Olga Krasnoukhova & Shahar Shirtz & Annemarie Verkerk;

o.krasnoukhova@hum.leidenuniv.nl, shahar.shirtz@gmail.com, annemarie.verkerk@uni-saarland.de)

The relationship between the domains of predicative possession, predicate location, and existence has been explored by many scholars. These relationships have often been argued for based on similarities in the structural coding means (i.e., type of copula, indexation, or flagging) deployed in affirmative clauses across these domains. Here, we ask to what degree does this relationship extend to the patterns in which these domains are negated. This is motivated by the well-known finding that negation in these domains shows rich and complex synchronic and diachronic patterns, both in individual languages and cross-linguistically (e.g., Croft 1991, Veselinova 2014, Van der Auwera & Krasnoukhova 2020, Shirtz, Talamo, & Verkerk 2021, Verkerk & Shirtz 2022).

To do this, we focus on the expression of negation in the three target domains across the Indo-European language family, a diverse family with a large amount of data available throughout most of its branches. We explore the variety of ways in which each domain is negated in the languages of our sample, illustrating the typological wealth of negation patterns across the three domains and the intra-linguistic variation in negation patterns within and across domains. We use this to explore the similarities and differences in negation patterns in the three domains across Indo-European and its branches, thus measuring the degree to which negation patterns support the purported grammatical relationship between predicative possession, predicate location, and existence.

To illustrate this variation, consider the Hindi clause in (1), expressing predicative possession with the copula *hai* indexing the possessed and the possessor flagged by *ke* Genitive + *pa:s* ‘near’. The same coding means are deployed also in clauses expressing predicate location, which differ in the relative order of ‘cats’ and ‘book’. The clause in (1) and its predicate location counterpart are both negated by the standard Hindi negation marker, *nahĩ*. The negation of Hindi existentials, however, may also be signaled by *nahĩ* functioning as a negative existential copula, without *hai* (Bashir 2006). This, then, illustrates the difference in negation patterns across domains.

Hindi (Indo-Aryan; own knowledge)

(1) *billi:jõ=ke pa:s kita:b nahĩ hai* ‘the cats don’t have the book’
 cat.PL=GEN near book NEG COP.PRS.3SG

The Odia negative copula *nah-* is used to negate clauses across all three domains, illustrated in (2a-b). In the past tense, however, the negation marker *nɔ* is deployed, followed by the past tense copula *tʰa*, culminating in a tense/aspect-based split of copular negation that is common across Indo-Iranian. English illustrates another pattern of variation, where existential and possessive predication may be negated by the indefinite negator *no* as in (3a), but also by the English negated auxiliary construction *do + not* as in (3b), or (rarely) by both patterns as in (3c). The strategies found in (3b-c), however, are not available in English existentials and predicate location.

Odia (Indo-Aryan; Neukom & Patnaik: 2003: 343-344; edited glosses)

(2a) *tɔmɔ-rɔ kɔ:nɔ kɔnca lɔnka nah-ĩ ki* ‘Don’t you have green chili?’
 2.POL-GEN QUANT green chili COP.NEG-3SG Q

- (2b) *set^{hi}-re kehi nah-anti* ‘There is no one in it’
 there-LOC anyone COP.NEG-3PL

English (Germanic; COCA (Davies 2012))

- (3) (a) *We have no car* (b) *We don't have a car* (c) *I don't have no car*

In this study, we focus on the emergence of within-family splits in the negation strategies of locative, existential, and possessive predicates. We identify splits of different nature 1) splits affecting all three domains equally (e.g., those based on tense-aspect), 2) splits between domains, such that possession and/or location and/or existence are negated in different ways, and 3) complex combinations of 1) and 2). We typologize the different diachronic processes that give rise to such splits, shedding light on sources of both semantic, lexical, and syntactic innovation that shape the expression of locative, existential, and possessive predicates.

References:

- Auwera, J. V. D., & Krasnoukhova, O. 2020. The typology of negation. In Déprez, V., & Espinal, M. T. (eds). *The Oxford handbook of negation*, 91-116. Oxford: Oxford University Press.
- Shirtz, S., Talamo, L., & Verkerk, A. 2021. The Evolutionary Dynamics of Negative Existentials in Indo-European. *Frontiers in Communication*, 6, 661862.
- Verkerk A. & Shirtz S. 2022. Negative existentials in Indo-European: a typological and diachronic overview. In: Veselinova, Ljuba & Hamari, Arja (eds.). 2022. *The Negative Existential Cycle*. (Research on Comparative Grammar 2). Berlin: Language Science Press. DOI: 10.5281/zenodo.6306474
- Bashir, E. 2006. Change in progress: Negation in Hindi and Urdu. In Rajendra Singh (ed.) *The yearbook of SA languages and linguistics 2006*: 3-29. Berlin: Mouton de Gruyter.
- Croft, W. 1991. The evolution of negation. *Journal of linguistics*, 27(1), 1-27.
- Davies, M. 2012. Corpus of Contemporary American English (<https://www.english-corpora.org/coca/>)
- Neukom, L. & Patnaik, M. 2003. *A grammar of Oriya*. Zürich: Seminar für Allgemeine Sprachwissenschaft.
- Veselinova, L. 2014. The negative existential cycle revisited. *Linguistics* 52(6) 1327-1389.

The development of locative, existential and possessive predication from a functional perspective

(Kees Hengeveld; P.C.Hengeveld@uva.nl)

This paper discusses various diachronic pathways of development of locative, existential and possessive predication using the framework of Functional Discourse Grammar (Hengeveld & Mackenzie 2008) and drawing on earlier work on the topic by the author (Hengeveld 1992). The focus is on two different aspects. The first concerns the diachronic development of the constructions involved as a whole, the second concerns the diachronic development of the copular element used within these constructions, if any.

As regards the first aspect, I will show that locative, existential, and possessive constructions may express meanings other than their original ones within the domain under study. Table 1 shows the distribution of constructions over meanings. It clearly shows that possessive meaning is most often parasitic on constructions that not are possessive in origin, locative meaning least often, with (locative-)existential meaning occupying an intermediate position. The paper will provide the empirical data that support Table 1.

| Construction Meaning | Lexical | Pseudo-transitive | Propriative | Predicative quantifier | Locative | Existential |
|------------------------|---------|-------------------|-------------|------------------------|----------|-------------|
| Locative | + | | | | | |
| (Locative-)Existential | + | + | + | + | | |
| Possessive | + | + | + | + | + | + |

Table 1. Constructions versus meanings

As regards the second aspect, the paper discusses the development of the copular element in the different types of predication. This copular element may have its origin in a locative, possessive, perception, or existential predicate of a lexical nature. Table 2 shows how these types of predicate enter the different construction types. Interestingly, it is the lexical possessive predicate that enters the widest range of construction types. Again, the paper will present the empirical data on which Table 2 is based.

| Origin of Copula Construction | Locative predicate | Perception predicate | Possessive predicate | Existential predicate |
|-------------------------------|--------------------|----------------------|----------------------|-----------------------|
| Existential | | | + | + |
| Pseudo-transitive | | + | + | |
| Locative | + | | | |

Table 2. Distribution of copula of different origins across different construction types

Combining the data in the two tables, it seems that the conclusion may be that possessive meaning is expressed drawing in the widest possible range of construction types, while at the same time lexical possessive predicates are an important source for the creation of copular elements in languages.

References

- Hengeveld, Kees (1992), *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin: Mouton de Gruyter.
- Hengeveld, Kees & Mackenzie, J. Lachlan (2008), *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.

‘Be/have’ verbs in historical perspective

(Denis Creissels; Denis.Creissels@univ-lyon2.fr)

A ‘be/have’ verb is a verb expressing possession in clauses such as English *John has a car*, in which the coding of the Possessor and the Possessee is similar to that of A and P in typical transitive clauses, but also used intransitively as a copula in plain-locational predication (i.e., in clauses such as English *John is in his office*),¹ sometimes also in nominal and/or adjectival predication (*John is a teacher*, *John is tall*). The following examples (from Li & Navarro 2015: 86, 89, 93) illustrate such a configuration in Kikuyu (Bantu), an AVP/SV language in which subjects are obligatorily indexed by means of a verbal prefix.²

- (1) *mũ-ti-rĩ* *arimũ.*
 SBJ:2PL-NEG-**be/have** PL.teachers(2)
 ‘You are not teachers.’ (nominal predication)

- (2) *tũ-rĩ* *a-rũaru.*
 SBJ:1PL-**be/have** cl2-sick
 ‘We are sick.’ (adjectival predication)

- (3) *i-bera* *rĩ-rĩ* *gĩ-kombe-inĩ.*
 SG-pear(5) SBJ:cl5-**be/have** SG-cup(7)-LOC
 ‘The pear is in the cup.’ (plain-locational predication)

- (4) *tũ-rĩ* *n-gari.*
 SBJ:1PL-**be/have** SG-car(9)
 ‘We have a car.’ (possessive predication)

- (5) *ha-rĩ* *benjũ* *metha-inĩ.*
 SBJ:cl16-**be/have** SG.pencil(9) SG.table(9)-LOC
 ‘There is a pencil on the table.’ (inverse-locational predication)

Most of the languages that have a ‘be/have’ verb are spoken in Mainland South East Asia. In this area, according to Chappell & Lü (2022), ‘be/have’ verbs are mainly found in Tibeto-Burman (Jingpho, Tujia, and several languages belonging to the Lolo-Burmese, Qiangic and Karenic branches of Tibeto-Burman), but also in two Austroasiatic languages (Bugan and Mang), in one Hmongic language (Yanghao), in three Sinitic languages (Hainan Southern Min, Linxia and Dabu Hakka), and in four varieties of Bai (a language whose classification as a Sinitic language or a highly sinicized Tibeto-Burman language is unclear).

Outside of Mainland South East Asia, this configuration is attested in a few languages of the Ghana-Togo region in West Africa: Akan (Kwa; Boadi 1971, Redden & Owusu 1995), Nkonya (Kwa; Reineke 1972) and Lama (Gur; Simnara 2019).

The other languages for which I have been able to find mentions of the existence of a ‘be/have’ verb show no areal clustering:

- Indonesian (Austronesian; Sneddon 1996),

¹ On plain-locational predication, as opposed to inverse-locational predication, see Creissels (2019).

² The role played by the subject index of class 16 in the inverse-locational clause (5) is comparable to that of *there* in the English equivalent of this clause.

- Diu Indo-Portuguese (Creole; Cardoso 2009),
- Gulf Pidgin Arabic (Bakir 2014),
- Iatmul and Manambu (two closely related Papuan languages; Jendraschek 2012, Aikhenvald 2008),
- Kikuyu (Bantu; Li & Navarro 2015).

In the presentation I would like to submit for the Workshop “Grammaticalization and diachrony of locative, existential and possessive predication”, I show that, for at least some of the languages listed above, there is solid evidence that the emergence of a ‘be/have’ verb resulted from one of the following scenarios:

- ‘have’ verb > existential predicator > locational copula
- copula used in possessive clauses of the type ‘At Possessor is Possessee’ > ‘have’ verb
- copula used in possessive clauses of the type ‘Possessor is with Possessee’ > ‘have’ verb

References

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Bakir, Murtadha J. 2014. The multifunctionality of fii in Gulf Pidgin Arabic. *Journal of Pidgin and Creole Languages* 29:2. 410-436.
- Boadi, Lawrence A. 1971. Existential sentences in Akan. *Foundations of language* 7. 19-29.
- Cardoso, Hugo C. 2009. *The Indo-Portuguese language of Diu*. Doctoral dissertation. University of Amsterdam.
- Chappell, Hilary & Shanshan Lü. 2022. A semantic typology of location, existence, possession and copular verbs: areal patterns of polysemy in Mainland East and Southeast Asia. *Linguistics* 60(1). 1–82.
- Creissels, Denis. 2019. Inverse locational predication in typological perspective. *Italian Journal of Linguistics* 31(2). 38-106.
- Jendraschek, Gerd 2012. *A grammar of Iatmul*. Habilitation thesis. University of Regensburg.
- Li, Jessica & Anaí Navarro. 2015. Copular clauses. In Englebretson, Robert (Ed.), with Wambüi Mūringo Wa-Ngatho (Language Consultant). *A basic sketch grammar of Gĩkũyũ*. Rice Working Papers in Linguistics, Vol. 6, Special Issue.
- Redden, J.E. & Owusu, N. 1995. *Twi basic course*. Hippocrene.
- Reineke, Brigitte. 1972. *The structure of the Nkonya language*. Leipzig : Verlag Enzyklopädie.
- Simnara, Komi. 2019. *L’expression des émotions en lama (Gur, Togo)*. Doctoral dissertation. University of Paris.
- Sneddon, James Neil 1996. *Indonesian: a comprehensive grammar*. London: Routledge.

Parallels in the development from locative and existential predications to possessive structures in Arabic and Hebrew

(Maris Camilleri, mcamil@essex.ac.uk)

This work takes as its starting point claims made in the typological and grammaticalisation literature then blends these with statements and analyses that stem from theoretical syntactic perspectives, with the aim at a reconstruction of Arabic and colloquial Hebrew possessive structures, meant to provide an analysis of the varied steps in the trajectory. Highlighting what led to the grammaticalisation of what synchronically appears to be a transitive have-like possessive structures in languages that do not possess a quintessential lexical ‘have’ predicate of the type that characterises Romance and Germanic possessive structures, the study will provide an answer to the question how be possessive predications mould into have ones, having themselves already stemmed from other clausal structures. Possessives in Arabic have developed out of a (predicative) locative structure Comrie (1991); Heine (1997), while according to Berman (1978), the Hebrew possessive structure is a development out of existentials. A synchronic analytical difference which characterises the two possessive structures is the following: The theoretical Arabic literature appears to have caught up with claims in Stassen (2009) that Arabic clausal possessives display a have-Drift that has led to their transitive have-like nature. Hallman (2020) has argued that Arabic possessives can be classified as be and have types, further mentioning that the latter is a development of the former, in line with a number of claims in the literature, e.g. Benveniste (1966). In the Hebrew syntactic literature, in contrast, possessives such as (1) are analysed distinctly, even if the varied strands in the literature agree on their diachronic origin as existentials.

- (1) yeš le-dani harbe sfarim
 EXIST to-Dani many books
 Dani has many books.

The claim put forward here is that the above Hebrew structure can best be characterised as a transitive have structure as Shlonsky (1987) analyses it. However, that is not all. The full picture is such that structures such as (2) are also available. In the analysis to be presented here, these structures are treated as be predicates on a par with Arabic counterparts. These are hypothesised to have functioned as precursors of the have structures in (1), even if the availability of such structures is not given much exposure in the literature.

- (2) le-dani sfarim harbe
 to-Dani books many
 Dani has many books.

Key to the development in the structures across the two systems is the earlier development of a P that bleaches into a CASE marker, in which *la* in Hebrew develops as a DATIVE marker (Borer and Grodzinsky, 1986), while collectively, the locative Ps *ʕand* ‘at’, *maʕ* ‘with’ and *la* ‘to’ grammaticalise as dependent markers that identify their erstwhile complement as the possessor NP. In both instances, a possessor grammaticalises as the SUBJ of a BE possessive predication. The main difference is that in Hebrew it is a NP, while in Arabic, it is a PP, parallel to ‘to’ + NP structures in English. This stage in the development constitutes a be predication; one that in the case of Arabic is merely a semantic development out of an inverted locative predicative structure. In both languages, it is a zero element that predicates of these structures. It may have been for this reason that by time we then observe the development of a pseudo-verbal HAVE predication moulding itself, as the BE possessive structures in both systems shift and develop into a HAVE structure. While Arabic reaches this stage via a dependent-to-head marking shift, Hebrew makes use of the existential structure, with the change involving a remapping between the grammatical functions/relations and the different thematic arguments involved.

References

- Benveniste, E. (1966). *Problèmes de linguistique générale*. Gallimard, Paris.
- Berman, R. A. (1978). *Modern Hebrew Structure*. University Publishing Projects, Tel Aviv.
- Borer, H. and Grodzinsky, Y. (1986). Syntactic cliticization and lexical cliticization: The case of Hebrew dative clitics. In *Syntax and Semantics, 19*, pages 175–217. Academic Press, New York.
- Comrie, B. (1991). On the importance of Arabic for general linguistic theory. In Comrie, B. and Eid, M., editors, *Perspectives on Arabic Linguistics III*, pages 3–30. John Benjamins, Amsterdam/Philadelphia.
- Hallman, P. (2020). Head and dependent marking in clausal possession. *Linguistic Inquiry*, 4:1–30.
- Heine, B. (1997). *Possession: Sources, forces and grammaticalization*. Cambridge University Press, Cambridge.
- Shlonsky, U. (1987). *Null and displaced subjects*. PhD thesis, Massachusetts Institute of Technology.
- Stassen, L. (2009). *Predicative Possession*. Oxford University Press, Oxford.

[W10] – The (Pre)History of the Languages of Japan – Current issues and prospects
Thursday September 7th – 13h30-17h

Organisers:

Étienne Baudel – École des hautes études en sciences sociales (Paris) (etienne.baudel@gmail.com)

Aleksandra Jarosz – Nicolaus Copernicus University in Toruń (aljarosz@umk.pl)

Georg Orlandi – Osaka University (georg.orlandi@gmail.com)

In spite of its tenacious reputation of being a monolingual country, Japan is actually home to a variety of languages that reflects a rich and complex linguistic history. Although this diversity is now finally starting to be acknowledged and protected, most of the minority languages of Japan are now severely endangered (Moseley 2009).

Thus, this workshop aims at embracing this diversity and at fostering multiple and interdisciplinary approaches to the questions regarding the linguistic (pre)history of Japan.

In this perspective, we will try to bring together researchers of diverse backgrounds and expertise, and to stimulate a discussion about the interactions of various approaches and scales of consideration.

Context

Over the past decades, a lot of research has been conducted on the history of the languages of Japan, and substantial advances have been made on the interaction between archaeological, genetic and linguistic data (for instance, Lee and Hasegawa 2011; Jarosz et al. 2022).

However, the hypothesis of a possible relatedness of the Japonic language family with any other neighbouring language families (and most notably with Koreanic) remains controversial (see for instance Vovin 2010 vs Robbeets 2005).

On an inner Japonic level, since the seminal works of Kindaichi Haruhiko and Hattori Shirō, the past decades have seen a spectacular surge of dialectology, which allowed new discoveries regarding the inner classification of Japonic (Kibe et al. 2021; Igarashi 2021), even though the classification of some famous “language islands” such as Hachijō are still a matter of debate (see Kupchik 2011: 7; vs Pellard 2015: 15 or 2018: 2).

In the meantime, Japan also saw important development in sociolinguistics (Heinrich and Ōhara, 2019; Asahi et al. 2022), which allowed to observe a lot of recent and ongoing language shifts, and especially the importance of new language contacts (ex. Long 2018).

On a philological scale, the numerous studies conducted recently on Eastern Old Japanese (Kupchik 2011; Vovin 2021), on Old Okinawan (Tawata 2010; Lin 2015; Serafim and Shinzato 2021), and the publication of the Oxford-NINJAL corpus of Old Japanese (NINJAL 2020) have dramatically transformed the access to ancient language data.

Similarly, a lot of progress has been made on the reconstructions of proto-languages, following the works of Martin (1987) and Thorpe (1983), and their revisions by Miyake (2003), Shimabukuro (2007) and Frellesvig and Whitman (2008). Intermediary proto-languages have also started to be reconstructed, for instance Proto-North-Ryukyuan (Lawrence 2009) and Proto-South-Ryukyuan (Jarosz 2019). However, in this perspective, one can but lament the lack of a proper etymological dictionary of the Japonic languages, since, sadly, Alexander Vovin could not complete his ambitious project during his lifetime.

On another note, a lot has also been uncovered on the (pre)history of Ainu languages since Vovin’s seminal work (1993), but a lot of questions still remain. Most notably, there is still no consensus regarding the origin of the Ainu, and Ainu can still not be classified as anything but an isolate. In parallel, however, the question of the contact and loans between Ainu and Japonic varieties has become a very active field of research (e.g.: Vovin 2009; Kupchik 2021).

Finally, based on Supalla's works on the linguistic history of the American Sign Language, Japanese Sign Language also has recently become an object of historical and comparative research (Nakamura 2006; Sasaki 2007; Kanda and Osugi 2011). Since that research, the critically endangered indigenous sign languages of Japan such as Amami Sign Language and Miyakubo Sign Language are also gaining rising attention (Kanda and Kimura 2016). However, there is still a lot to be discovered on the origin and evolutions of sign languages in Japan.

Research questions and goals:

Our workshop aims at studying the history and prehistory of all indigenous languages of Japan. Those include discussions on the proto-languages, as well as the ancient and modern forms of all the following:

- mainland Japanese varieties
- Hachijō language
- Ryukyuan languages
- Ainu languages
- ‘contact languages’, such as Bonin English and Ogasawara Japanese
- sign languages: Japanese Sign Language, Amami Sign Language, Miyakubo Sign Language

Furthermore, we wish to study those languages from several perspectives. Thus, we welcome contribution propositions that may discuss (but need not be limited to):

- the prehistory of the languages of Japan and of their speakers
- the history of those languages
- the changes in the “linguistic ecology” of Japan
- the ongoing changes in the synchrony of the languages of Japan
- the languages of Japan outside of Japan, as heritage or migrant languages (for instance, in Hawai'i, in South America, etc.)
- the implementation of recent concepts and of new technologies to the historical linguistics of those languages

References

- Alonso de la Fuente, José Andrés. 2012. *The Ainu languages: traditional reconstruction, Eurasian areal linguistics, and diachronic (holistic) typology*. PhD dissertation, Universidad del País Vasco – Euskal Herriko Unibertsitatea.
- Asahi Yoshiyuki, Usami Mayumi and Inoue Fumio. 2022. *Handbook of Japanese Sociolinguistics*. Berlin: Mouton de Gruyter.
- Bugaeva, Anna (ed.). 2022. *Handbook of the Ainu language*. Berlin: De Gruyter Mouton.
- de Boer, Elisabeth. 2020. The classification of the Japonic languages. In: Robbeets, Martine and Savelyev Alexander (eds.), *The Oxford Guide to the Transeurasian languages*. Oxford: Oxford University Press. 40-58.
- Frellesvig, Bjarke. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press.
- Frellesvig, Bjarke and Whitman John. 2008. *Proto-Japanese: Issues and Prospects*. Amsterdam: John Benjamins Publishing Company.
- Frellesvig, Bjarke; Shibatani Masayoshi and Smith, John Clark (eds.). 2007. *Current issues in the History and Structure of Japanese*. Tokyo: Kuroshio Publishers.
- Frellesvig, Bjarke; Kinsui Satoshi and Whitman, John. 2017. *Handbook of Historical Japanese Linguistics*. Berlin: Mouton de Gruyter.
- Heinrich, Patrick and Ōhara Yumiko. *Routledge Handbook of Japanese Sociolinguistics*. London: Routledge.
- Igarashi, Yōsuke. 2021. Bunki-gaku-teki shuhō ni motodzuuta nichiryū shogo no keitō bunrui no kokoromi [An Attempt to Systematic Classify Japanese and Ryukyu Languages Based on Cladding Techniques]. In: Hayashi Yuka; Kinuhata Tomohida; Kibe Nobuko (eds). *Fīrudo to bunken kara miru nichiryū shogo no keitō to rekishi* [Lineage and History of Japanese and Ryukyu Languages Seen from Fields and Documents]. Tokyo: Kaitakusha.
- Miyake, Marc. 2003. *Old Japanese: A phonetic reconstruction*. London: Routledge Curzon.
- Jarosz, Aleksandra. 2019. Fine details of a larger picture: Proto-Sakishima *au and *ao diphthongs. *Folia Linguistica Historica*, 40 (2), 401-437.
- Jarosz, Aleksandra et al. 2022. Demography, trade and state power: a tripartite model of medieval farming/language dispersals in the Ryukyu Islands. *Evolutionary Human Sciences*, 4. Cambridge: Cambridge University Press.
- Kanda Kazuyuki and Osugi Yutaka. 2011. Database of historical changes in Japanese signs from 1901–2011. Paper presented at the International Conference on Historical Linguistics, Osaka, Japan.
- Kanda Kazuyuki and Kimura Tsutomu. 2016. Ritō ni okeru rōsha hōmusain no dēta shūshū ni kansuru chōsa hōkoku [Survey Report on Data Collection of Deaf People's Home Signs on Remote Islands]. *Hyūman'intāfēsu gakkai kenkyūkai hōkoku-shū*. 18. 1. 45-48.
- Kibe Nobuko; Hayashi Yuka; Kinuhata Tomohide. 2021. *Fīrudo to bunken kara miru Nichiryū shogo no keitō to rekishi* [The phylogeny and history of the Japonic languages seen from the field and from philological records]. Tokyo: Kaitakusha.
- Kubozono Haruo. 2016. Language contact and accent changes in Japanese. *The Journal of the Acoustical Society of America*, 140 (4). 3050-3051.
- Kupchik, John. 2011. *A Grammar of the Eastern Old Japanese dialects*. PhD dissertation, Mānoa: University of Hawai'i.

- Kupchik, John. 2021. Ainu Loanwords in Hachijō. In; Kupchik, John; Alonso de la Fuente, José Andrés and Miyake, Marc Hideo. *Studies in Asian Historical Linguistics, Philology and Beyond*. Leiden: Brill.
- Lawrence, Wayne. 2009. Kita Ryūkyū sogo no meishi onchō – Shiron [The tone system of Proto-Northern-Ryukyuan nouns – A proposal]. In: *Okinawa bunka*, 43, 2, 106. 102-186.
- Lee, Sean & Hasegawa Toshikazu. 2011. Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B* 278.3662-69.
- Lin Chihkai. 2015. *A Reconstruction of Old Okinawan: a Corpus-Based Approach*. PhD dissertation, Mānoa: University of Hawai'i.
- Long, Daniel. 2018. *Ogasawara Shotō no kongō gengo no rekishi to kōzō : Nihon ganrai no tabunka kyōsei shakai de okita gengo sesshoku* [History and structure of mixed languages in the Ogasawara Islands: Language contact that occurred in Japan's original multicultural society]. Tokyo: Hitsuji shobō.
- Martin, Samuel Elmo. 1987. *The Japanese Language through Time*. New Haven: Yale University Press.
- Matsuoka Kazumi; Yano Haiko. 2017. Ehime-ken Oshima Miyakubo-cho no shuwa: airand sign [Island Signs: The Sign Language of Miyakubo in Ehime Prefecture]. *Kagaku*, 87 (5). 415–417.
- Matsuoka Kazumi; Taira Eiji; Yano Haiko. 2014. Ehime ken ōshima ni okeru birejissain (shuwa hōgen) no hozon oyobi gengo-gaku-teki bunseki no tame no dētabēsu no kōchiku [Construction of a database for the preservation and linguistic analysis of village signs (sign language dialects) in Ōshima, Ehime prefecture]. *Fukutake zaidan*.
- Moseley, Christopher (ed.) and Nicolas, Alexander (cartographer). 2009. *Atlas of the world's languages in danger*, 3rd edition. Paris: Unesco.
- Nakamura, Karen. 2006. *Deaf in Japan: signing and the politics of identity*. Ithaca: Cornell University Press.
- National Institute for Japanese Language and Linguistics (2020) “Oxford-NINJAL Corpus of Old Japanese” <http://oncoj.ninjal.ac.jp/>
- Osada Toshiki (ed.). 2020. *Nihongo "kigen" ron no rekishi to tenbō: Nihongo no kigen wa dono yō ni ronjirarete kita ka* [Debating the origins of the Japanese language: history and outlook]. Tokyo: Sanseidō.
- Osada Toshiki and Vovin, Alexander (eds.). 2003. *Nihongo Keitōron no genzai Perspectives on the Origins of the Japanese Language*. Tokyo: International Research Center for Japanese Studies.
- Osterkamp, Sven. Foreign Sources on the History of the Japanese Language: Problems and Prospects. In: *Acta Asiatica: Bulletin of the Institute of Eastern Culture*, 111 (Recent Trends in Japanese Language Studies).75–93.
- Osugi Yutaka; Supalla, Ted; Webb, Rebecca. 1999. The use of word elicitation to identify distinctive gestural systems on Amami Island. *Sign Language & Linguistics*, 2, 1. 87-112.
- Pellard, Thomas. 2015. The linguistic archaeology of the Ryukyu Islands. In: Heinrich, Patrick; Miyara Shinshō; Shimoji Michinori. *Handbook of the Ryukyuan languages: History, structure, and use*. Berlin: De Gruyter Mouton. 13–37.
- Pellard, Thomas. 2018. The comparative study of the Japonic languages. Presentation given in *Approaches to endangered languages in Japan and Northeast Asia: Description, documentation and revitalization*. Tachikawa: National Institute for Japanese Language and Linguistics.
- Robbeets, Martine. 2005. *Is Japanese related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden: Harrassowitz.

- Sasaki Daisuke. 2007. “The lexicons of Japanese Sign Language and Taiwan Sign Language: A preliminary comparative study of handshape differences. In: Quinto-Pozos, David. *Sign Languages in Contact*. Washington DC: Gallaudet. 123-150.
- Serafim, Leon and Shinzato Rumiko. 2021. *The Language of the Old-Okinawan Omoro Sōshi*. Leiden: Brill.
- Shibatani Masayoshi. 1990. *The Languages of Japan*. Cambridge: Cambridge University Press.
- Shimabukuro Moriyo. 2007. *The Accental History of the Japanese and Ryukyuan Languages: a Reconstruction*. London: Global Oriental.
- Takubo Yukinori; Whitman, John; Hirako Tatsuya. *Ryūkyū shogo to Kodai Nihongo: Nichiryū sogo no saiken ni mukete* [Ryukyuan and premodern Japanese: Toward the reconstruction of proto-Japanese-Ryukyuan]. Tokyo: Kuroshio.
- Tawata Shin'ichirō. 2010. *Okinawa-go on'in no rekishi-teki kenkyū* [A Historical Study of Okinawan Phonology]. Hiroshima: Keisuisha.
- Thorpe, Maner Lawton. 1983. *Ryukyuan Language History*. University of Southern California: PhD dissertation.
- Unger, James Marshall. *The Role of Contact in the Origins of the Japanese and Korean Languages*. 2009. Honolulu: University of Hawai'i Press.
- Vovin, Alexander. 1993. *A Reconstruction of Proto-Ainu*. Leiden: Brill.
- Vovin, Alexander. 2009. Man'yōshū to fudoki ni mi rareru fushigina kotoba to jōdai nipponrettō ni okeru ainukotoba no bunpu [Strange Words in the Man'yōshū and the Fudoki and the Distribution of the Ainu Language in the Japanese Islands in Prehistory]. Kyoto: International Research Center for the Japanese Studies.
- Vovin, Alexander. 2010. *Koreo-Japonica: A re-evaluation of a common genetic origin*. Honolulu: University of Hawai'i Press.
- Vovin, Alexander. 2021a. ‘Names of Large Exotic Animals and the Urheimat of Japonic’. *International Journal of Eurasian Linguistics* 2.1.
- Vovin, Alexander. 2021b. *The Eastern Old Japanese Corpus and Dictionary*. Leiden: Brill.
- Whitman, John B. 2012. *The relationship between Japanese and Korean*. In: Tranter, David N (ed.). *The Languages of Japan and Korea*. London: Routledge, 24-38

On stative/active intransitive split within tripartite alignment: A case of Kuril Ainu

Tomomi Satō (Hokkaido University) and Anna Bugaeva (Tokyo University of Science)

Ainu, the only non-Japonic language of Japan, was gradually pushed from Honshū to the north so that “northern Hokkaidō was occupied by ethnic Ainu by c. 1000 CE, southern Sakhalin by c. 1300 CE, and the Kurile Islands... as late as c. 1500–1600 CE.” (Janhunen 2022: 63)

This paper focuses on the least documented Kuril variety of Ainu, which disappeared in the early 20th century without any substantial texts left. Using both published and archival Kuril Ainu materials, we attempt to reconstruct its system of organizing grammatical relations.

Just like Hokkaidō Ainu, Kuril Ainu shows mixed alignment in verbal indexing being nominative-accusative in 1SG, neutral in 2nd and 3rd SG/PL, and tripartite in 1PL exclusive and inclusive. We assume that like in Hokkaidō Ainu, 1PL inclusive in Kuril Ainu is marked on the verb by *an-* for the transitive subject (A), *-an* for the intransitive subject (S), and *i-* for the object (O), which presumably also have a number of other functions conventionally gathered under the ‘4th person’ label, for example, the impersonal (‘(some)one, people’), 2nd person honorific, and logophoric functions. However, unlike any other Ainu variety, Kuril Ainu demonstrates an additional stative/active intransitive split within the 4th person by marking the subject of stative predicates (So) with *i-*, which is the object marker, and the subject of agentive predicates (Sa) with the regular intransitive subject marker *-an*.

- (1) *i-okay hi* {4.O-exist.PL Q} ‘Is someone alive?’ (KS #312)
i-omke wa {4.O-cough FIN} ‘Someone coughed.’ (KS #426, #462)
i-mokor-ci wa {4.O-sleep-PL FIN} ‘People slept.’ (KS #1097, #1099)
i-merayke {4.O-be.cold} ‘Someone felt cold.’ (Krasheninnikov 1755-II: 187)
i-mos wa {4.O-wake FIN} ‘Someone woke up.’ (Dybowski 1891: 29)
i-ru wa {4.O-melt FIN} ‘Something melted.’ (Dybowski 1891: 29)
i-wor-osma {4.O-water-enter} ‘Something sank.’ (Dybowski 1891: 33)
tanto i-pirka {today 4.O-be.good} ‘Today (the weather) is good.’ (Torii 1903: 131)
- (2) *sattek ek-an* {be.thin come.SG-4.S} ‘Someone came on foot.’ (KS #295)
ironno-an {catch.prey-4.S} ‘Someone caught prey.’ (KS #525)
kunne-ipe-an {be.dark-have.meal-4.S} ‘Someone had dinner.’ (KS #709)
hekirpa-an {turn.around-4.S} ‘Someone turned around.’ (KS #743)
ma-an {swim-4.S} ‘Someone swam.’ (KS #890)
as-an {stand.SG-4.S} ‘Someone stood.’ (KS #1118)
terke-an {jump-4.S} ‘Someone jumped.’ (KS #1125)

Semantics-driven intransitive splits are not unusual (cf. Old Japanese in Yanagida & Whitman 2009), but, to our knowledge, they have hardly ever been reported for a language with tripartite alignment, which is heavy enough by overdistinguishing grammatical relations. Unsurprisingly, the stative/active distinction has eventually been lost in most Ainu dialects.

References

Dybowski, Benedykt. 1891. *Słowniki narzecz Ludów Kamczackich* [Dictionaries of the people of Kamchatka]. Rozprawy Akademii umiejętności, Wydział Fylologiczny, Serya II, Tom I. Krakow: 53-119.

Janhunen, Juna. 2022. Ainu ethnic origins. In Anna Bugaeva (ed.), *Handbook of the Ainu language* (Handbooks of Japanese Language and Linguistics 12). 57–78. Berlin: Mouton de Gruyter.

Krasheninnikov, Stepan Petrovich. 1755. *Описание земли Камчатки* [Description of Kamchatka land], volumes I and II. St. Peterburg: Imperial Academy of Sciences.

KS Voznesensky, Ilya Gavrilovich (unpublished manuscript, 1844?) *Kuril'skie Slova* [Kuril words]. St. Petersburg Branch of the Archive of the Russian Academy of Sciences, f.53, op.1, #43.

Torii, Ryūzō. 1903. *Chishima Ainu* [The Kuril Ainu]. Tokyo: Yoshikawa kōbunkan.

Yanagida, Yuko & John Whitman. 2009. Alignment and Word Order in Old Japanese. *Journal of East Asia Asian Linguistics* 18, 101-144.

Debuccalization of *p in the Naha dialect of the Ryukyuan language

Moriyo Shimabukuro (University of the Ryukyus)

The purpose of this paper is to examine debuccalization in the Naha dialect (hereafter shortened to ‘Naha’) of the Ryukyuan Okinawan language, and to show how this change affected the language’s phonological system.

Concerning the historical development of Ryukyuan phonology, Iha (2000 [1910]), Hattori (1999 [1959]), Nakamoto (1976), and Thorpe (1983) claimed that *p* turned into *h*. Past studies also show that the *p* was \emptyset at some point before becoming *h*. This change can be attested by historical documents written in Korean and Chinese (Tawata 2010 and Li 2015).

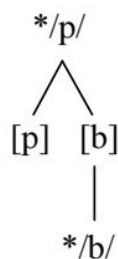
It is important to note, however, that this change did not necessarily occur in all dialects of the language – some still retained *p*. The word for ‘nose’ in Yoron dialect is *pana* (Kiku and Takahashi 2005), for example, while the same word is *hana* in Naha (Uchima and Nohara 2006). The same phenomenon of *h* and *p* can be traced to *p in Proto Ryukyuan (PR).

As this paper demonstrates, *p* did not in fact completely shift to *h*, but rather a split, i.e., *p* splitting to *p* and *h*. Indeed, there are examples where *p* still exists, even in those dialects in which the change is said to have occurred. We know that *p* exists because native speakers recognize *p* in [kampatʃi] ‘scar’, i.e., /kanpaci/, not /kanhaci/, and minimal pairs such as one in below exists in Naha.

(1) sampin ‘jasmine tea’: sammin ‘calculation’

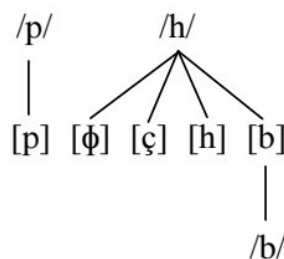
In Naha *p* and *h* are phonemic. The former can be realized as only [p]. Conversely, *h* can be [ϕ], [ç], [h], or [b]. The diagram below shows the relation between the phonemes and their allophones in both PR and Naha. (The diagram includes b to show that the sound [b] is an allophone of not only b, but also h in Naha.)

(2) Proto-Ryukyuan



>

Modern Naha dialect



Based on the distribution of the sounds [p], [ϕ], [ç], and [h] in the modern Naha phonological system, we hypothesize that there were three stages of changes in the complex structure of Naha’s development. A spirantization of *p* to [ϕ] occurred first. When followed by the vowel *i* a palatalization of *p* occurred, turning *p* into the sound [ç]. The occurrence of [ϕ] was eventually limited to the environment followed by *u*, and [h] to the environment with non-high vowels. As a result, in addition to *p*, the phoneme *h* was also established.

References

- Hattori, Shirō. 1999 [1959]. *Nihongo no Keitō* (The lineage of the Japanese language). Tokyo: Iwanami Shoten.
- Iha, Fuyū. 2000 [1910]. P on'kō (Study of the sound p). *Koryūkyū* (Old Ryukyu). (Hokama Shuzen, ed. and revisor.) Iwanami Bunko blue series, N 102.1. Tokyo: Iwanami Shoten. pp.375-386.
- Kiku, Chiyo and Toshizō Takahashi. 2005. *Yoron Hōgen Jiten* (Dictionary of Yoron dialect). Musashino Shoin.
- Li, Chihkai. 2015. *A Reconstruction of Old Okinawan: A corpus-based approach*. Ph.D. dissertation, University of Hawai'i at Manoa.
- Nakamoto, Masachie. 1976. *Ryūkyū Hōgen On'in no Kenkyū* (Study of the Ryukyuan dialect phonology). Tokyo: Hōsei Daigaku Shuppankyoku.
- Tawata, Shin'ichirō. 2010. *Okinawago On'in no Rekishiteki Kenkyū* (Historical Study of the Okinawan language phonology). Hiroshima: Keisuisha.
- Thorpe, Maner L. 1983. *Ryukyuan Language History*. University of Southern California Doctoral Dissertation.
- Uchima, Chokujin and Nohara Mitsuyoshi. 2006. *Okinawa-go Jiten: Naha Hōgen o Chūshin ni* (Dictionary of the Okinawan language: Focusing on the dialect of Naha). Tokyo: Kenkyūsha.

Reconstructing the Proto-Japonic demonstrative system

Tomohide Kinuhata (Fukuoka University)

The demonstrative system of Old Japanese (OJ) was significantly different from that of Modern Japanese (ModJ). While OJ *ko-* referred to proximate objects as with ModJ *ko-*, distal demonstrative pronouns, presumably *ka-*, were rarely used. *So-* was solely used as an anaphoric pronoun, not referring to a medial object deictically. (See Hashimoto 1966, Kinuhata 2022.) These distributions amount to the difference depicted in Figures 1 and 2.

Fig. 1: OJ

| | deic. | anaph. |
|-------|-----------------|------------|
| prox. | <i>ko-</i> | <i>so-</i> |
| dis. | (<i>ka-?</i>) | |

Fig 2: ModJ

| | deic. | anaph. |
|-------|------------|---------------|
| prox. | <i>ko-</i> | <i>so-/a-</i> |
| med. | <i>so-</i> | |
| dis. | <i>a-</i> | |

Fig 3: PM-PR

| | deic. | anaph. |
|-------|--------------|-------------|
| prox. | * <i>ko-</i> | * <i>o-</i> |
| dis. | * <i>ka-</i> | |

Recently, Kinuhata and Hayashi (2018) hypothesized a demonstrative system similar to OJ for Proto-Miyakoan (PM) based on the data from their Shinzato and Karimata dialects. Moreover, Kinuhata (2021) found the use of this system in the Irabu dialect of Miyakoan and discusses its origin in Proto-Ryukyuan (PR) (Fig. 3). While the above semantic resemblance is attractive, morphological issues remain in reconstructing a Proto-Japonic demonstrative system.

In reconstructing Proto-Old Japanese (POJ), one must consider anaphoric *si-* and demonstrative adverbs, i.e., proximate *ka-* and anaphoric *sika-*. The anaphoric adverb *sika-* evidently consists of the anaphoric pronoun *si-* and the adverb *ka-*. Comparing the anaphoric pronouns *so-* and *si-*, it is more probable for *si-* to be older than *so-* because 1) the former had constituted the anaphoric adverb *sika-* in OJ and 2) the formation of *so-* can be explained by an analogical extension of **ko-* to **si-*. That is, the vowel of **ko-*, i.e., /o2/, was adapted to *si-* to create a new anaphoric pronoun *so-*. This process later created an anaphoric adverb *sa-* in Early Middle Japanese (EMJ) (Okazaki 2010), adapting the vowel of the demonstrative adverb *ka-*, i.e., /a/, to the anaphoric pronoun *so-*. Thus, we can assume at least three distinct morphemes for POJ, as in Fig. 4.

The demonstrative adverbs of Ryukyuan languages widely attest *ka-* for deictic use and *a-* for anaphoric use (cf. Nakamoto 1983, Uchima 1984). Since the proximate adverb *ka-* has the cognate in OJ, i.e., OJ proximate *ka-*, it traces back to Proto-Japonic (PJ) **ka-*. Given the proximate adverb **ka-* in PJ, the nominal **ko-* and adverbial **ka-* opposition could have analogically extended to the o- and a- in the anaphoric use. Therefore, we can consider the anaphoric morpheme o- and a- in Ryukyuan languages as later innovation, like *so-* and *sa-* in Japanese. Instead of postulating them, this presentation proposes reconstructing **e-* for the anaphoric use in Proto-Ryukyuan. Though **e-* does not have many reflexes in modern Ryukyuan languages, it surfaces as *isii-* (anaphoric adverbial with instrumental =*sii*) in the Irabu dialect of Ryukyuan (Tomihama 2013). This reconstruction leads us to posit four distinct morphemes given in Fig. 5 for PR.

Fig. 4: POJ

| | deic. | anaph. |
|------|---------------------------------|--------------|
| noun | * <i>ko-</i> , (* <i>ka-?</i>) | * <i>si-</i> |
| adv. | * <i>ka-</i> | |

Fig. 5: PR revised

| | deic. | anaph. |
|------|-----------------------------|-------------|
| noun | * <i>ko-</i> , * <i>ka-</i> | * <i>e-</i> |
| adv. | * <i>ka-</i> | |

Fig. 6: PJ

| | deic. | anaph. |
|------|-----------------------------|--------------|
| noun | * <i>ko-</i> , * <i>ka-</i> | * <i>se-</i> |
| adv. | * <i>ka-</i> | |

Comparing the two reconstructed demonstrative systems in Figures 4 and 5 still leaves the questions regarding 1) what is the origin of /s/ in the anaphoric pronoun of POJ and 2) whether the distal pronoun *ka-* can trace back to Proto-Japonic (PJ). I will discuss in the presentation that the /s/ goes back to PJ and the rare attestation of distal *ka-* is due to the problem of source materials in OJ. These assumptions lead us to conclude that the PJ demonstrative system has four distinct morphemes, given in Fig. 6.

References

- Hashimoto, Shiro (1966) “Kodaigo no shizitaikai: Jōdai o chūshin ni [The demonstrative system of Old Japanese: Focusing on early Old Japanese],” *Kokugo Kokubun*, 35(6): 329–341.
- Kinuhata, Tomohide (2021) “Ryūkyūshogo to Jōdai Nihongo kara mita sogo no shizi taikai shiron [A preliminary study on the proto-system of demonstratives viewed from Ryukyuan languages and Old Japanese],” in Hayashi, Yuka, Tomohide Kinuhata, and Nobuko Kibe eds. *Firudo to bunken kara miru Nichiryūshogo no keitō to rekishi*, pp. 190–213, Tokyo: Kaitakusha.
- Kinuhata, Tomohide (2022) “Jōdaigo shizishi no shizi tokuchō: bunken o mochiita kansatsu dēta no kōchiku [Deictic property of demonstratives in Old Japanese: a method to construct observational data using written texts],” *Jinbun Ronsō: Fukuoka University*, 54(3): 629–670.
- Kinuhata, Tomohide and Yuka Hayashi (2018) “On the anaphoric use of demonstratives in Miyakoan,” in Shin Fukuda, Mary Kim and Mee-Jeong Park eds. *Japanese/Korean Linguistics 25*, pp. 35–47, Stanford: CSLI.
- Nakamoto, Masachie (1983) *Ryūkyū goishi no kenkyū* [A study on the history of Ryukyuan lexicon], Tokyo: San-ichi Shobo Publishing Inc.
- Okazaki, Tomoko (2010) *Nihongo shizishi no rekishiteki kenkyū* [Historical study of Japanese demonstratives], Tokyo: Hituzi Syobo.
- Tomihama, Sadayoshi (2013) *Miyako Irabu hōgen ziten* [Dictionary of the Miyako Irabu dialect], Okinawa: The Okinawa Times.
- Uchima, Chokuji (1984) *Ryūkyūhōgen bunnō no kenkyū* [A study of Ryukyuan grammar], Tokyo: Kasama Shoin.

Reconsidering the classification of Hachijō: A glimpse from historical phonology

Étienne Baudel (École des hautes études en sciences sociales)

Hachijō (locally simply called 島言葉 *Shima-kotoba* ‘island speech’) is an endangered minority language of Japan, originally spoken in the southern part of the Izu archipelago (primarily on the three islands of Hachijō, Kojima and Aogashima). Like most Japonic varieties, it was long considered a dialect of Japanese; however, the dominant view among specialists is now to treat it as a separate language. Being now critically endangered, Hachijō was included in 2009 in the online version of UNESCO’s *Atlas of the world’s languages in danger* (Moseley, 2009), alongside Ainu and six Ryukyuan languages.

On the other hand, the term ‘Eastern Old Japanese’ (EOJ) serves as collective term to refer to several dialects of Old Japanese that are primarily attested in the *Man’yōshū* (books 14 and 20), and in a few other minor sources (see Vovin, 2021). EOJ is usually considered as a ‘dialect continuum’ (Vovin, 2021:27) within Old Japanese, and, according to some, a few of those dialects might be divergent enough from Western Old Japanese (WOJ) to be considered a ‘separate branch of the Japanese subgroup of the Japonic language family’ (Kupchik, 2011: 6).

The classification of Hachijō within the Japonic language family has been a topic for discussion since at least the beginning of the Meiji period, when several phonological, morphological and lexical similarities were noted (first by Dickins and Satow, 1878: 464) between Hachijō and EOJ. Based on those resemblances, the idea that Hachijō could be a living descendant of EOJ gradually became somewhat widespread, see for instance: Tachibana & Tōjō (1934:45), Hirayama (1965), Hattori (1968), Ōshima (1975:52), Kaneda (2011:154), Kaneda & Holda (2018:1), Kupchik (2011:6; 2016).

According to this interpretation, Hachijō is usually supposed to be the sole descendant of EOJ, which would place it on its own branch within modern Japanese varieties. However, other specialists consider instead that EOJ could also be the mother language of other modern Eastern varieties of Japanese, in addition to Hachijō (see, for instance, de Boer, 2020:28).

Finally, Hachijō was also compared more recently with other Japonic branches, such as north-eastern Japanese dialects (especially Tōhoku, Akiyamagō and Toshima varieties), Kyūshū dialects and Ryukyuan languages; and, based on these comparisons, some specialists estimate that most of the similarities between EOJ and Hachijō are, in fact, most likely to be due to shared archaism rather than to shared innovation. Thus, according to them, there is not enough evidence yet to assert whether there is a direct genetic relationship between them (see for instance Pellard, 2018:2).

Thus, this talk aims at taking a closer look at the most recent descriptions of Eastern Old Japanese data (developed most notably by Kupchik, 2011 and Vovin, 2021) and at the most comprehensive Hachijō data (compiled in Baudel, forthcoming), in order to examine arguments for the classification of Hachijō.

Due to time limitation, this talk will focus solely on arguments from historical phonology, and mostly to the treatment of proto-Japonic vowels and glides.

References

- Baudel, Étienne (forthcoming). *Shima-kotoba: A synchronic and diachronic study of the Hachijō language*. PhD dissertation, Paris: École des hautes études en sciences sociales.
- De Boer, Elisabeth. 2020. The classification of the Japonic languages, in Robbeets Martine & Savelyev Alexander (eds), *The Oxford Guide to the Transeurasian languages*. Oxford: Oxford University Press. 40-58.
- Dickins, Frederick Victor & Satow, Ernest. 1878. Notes of a visit to Hachijō in 1878. *Transactions of the Asiatic Society of Japan*, 6. 435–77.
- Hattori Shirō. 1968. Hachijō-jima hōgen ni tsuite. *Kotoba no Uchū*, 3. 92–95.
- Hirayama Teruo. 1965. Hachijōjima, Kojima, Aogashima hōgen. *Izu shotō hōgen no kenkyū*. 168–209. Tōkyō: Meiji Shoin.
- Kaneda Akihiro. 2011. Hachijō hōgen: kodai tōgoku hōgen no nagori. In Kurebito Megumi (ed.) *Nihon no kiki gengo: gengo, hōgen no tayōsei to dokujisei*. 153-169. Sapporo: Hokkaidō Daigaku Shuppankai.
- Kaneda, Akihiro & Holda, Martin. 2018. ‘Hachijō, South Ryukyuan languages and East-Northeast Japan dialects from the viewpoint of the concentric circle theory of dialect divergence’, NINJAL presentation, NINJAL International Symposium
- Kupchik, John. 2011. *A Grammar of the Eastern Old Japanese dialects*, PhD dissertation, Mānoa: University of Hawai‘i.
- Kupchik, John. 2016. On the etymology of the Eastern Japanese word tego. *Journal of the American Oriental Society*, 136. 733-743.
- Mase Yoshio. Ikite iru Azuma-uta no gohō. *Gengo seikatsu*, 342, 34-40, Tōkyō: Chikuma Shobō.
- Moseley, Christopher (ed.). 2009. *Atlas of the world’s Languages in danger*. Paris: UNESCO.
- Ōshima Ichirō. 1975. Hachijō hōgen. *Shin Nihongo Kōza*, 3. 41–60.
- Pellard, Thomas. 2018. The comparative study of the Japonic languages. Approaches to endangered languages in *Japan and Northeast Asia: Description, documentation and revitalization*. Tachikawa: National Institute for Japanese Language and Linguistics.
- Tachibana Shōichi & Tōjō Misao. 1934. Honshū tōbu no hōgen: Izu shotō hōgen. *Kokugo kagaku kōza* VII, 42–55. Tōkyō: Meiji Shoin.
- Vovin, Alexander. 2018. On the etymology of the name of Mt. Fuji. In: Vovin, Alexander & McClure William (eds.), *Studies in Japanese/Korean historical and theoretical linguistics and beyond. A Festschrift for John B. Whitman on the Occasion of his Sixtieth Birthday*. Folkestone & Leiden: Global Oriental/Brill. 80–89.
- Vovin, Alexander. 2021. *The Eastern Old Japanese Corpus and Dictionary*. Leiden: Brill.

[Foreword to discussion]

Old, Middle and New: Periodisation as a back-burnered topic in the diachronic research of Japanese

Tomasz Majtczak (Jagiellonian University in Kraków)

The proposed contribution to the workshop aims to raise the question of the periodisation of the history of the Japanese language, with the focus on its literate (written) phase.

There are two aspects to this problem, which – although seemingly independent of each other – are nevertheless partly connected: the division itself and the terminology.

The former aspect, that of establishing the time boundaries between periods, is a most basic element of any diachronic description, and yet most historical linguists of Japanese appear to settle for adopting the socio-political periodisation as it comes (see e.g. Frellesvig 2010, Miyake 2020; also Martin 1987, under the somewhat distanced heading “Periods discussed by Japanese grammarians”, and Calvetti 1999, with a longer elaboration). This is hardly satisfying or even acceptable in linguistic research.

Terminology, on the other hand, can be regarded as purely arbitrary and conventional, as well as language-bound, but certain names of periods – even if this is not fully intended – do suggest some stronger connection between particular stages of language development (cf. e.g. *Old Japanese / Early Middle Japanese / Late Middle Japanese* in Frellesvig 2010 versus *Old Japanese / Late Old Japanese / Middle Japanese* in Takeuchi 1999 – referring to the same three time spans).

In both cases the decision about the diagnostic features and their selection are of course of paramount importance, but for Japanese they are mostly left unmentioned.

The paper is not to propose any coherent and ultimate solution to the indicated problems, but rather to spark off a debate over this neglected point of diachronic study of Japanese. The widespread periodisation based on the socio-political history will be presented, its disadvantages discussed and compared with some other – far less popular but usually much better substantiated – propositions available in the relevant specialist literature (as e.g. Rickmeyer 2017 and Narrog 1999, with certain modifications and specifications in Majtczak 2016 and especially in Osterkamp 2021 on the one hand, or Satō 2001 on the other). A very interesting and desirable side effect of this paper might be a parallel consideration of the Ryukyuan languages and of the division of their history into periods.

References

- Calvetti, Paolo. 1999. *Introduzione alla storia della lingua giapponese*. Napoli: Istituto Universitario Orientale.
- Frellesvig, Bjarke. 2010. *A history of the Japanese language*. Cambridge: Cambridge University Press.
- Majtczak, Tomasz. 2016. *The inflectional system of Classical Japanese*. Kraków: Księgarnia Akademicka.
- Martin, Samuel E. 1987. *The Japanese language through time*. New Haven – London: Yale University Press.
- Miyake, Marc. 2020. Historical sources and periodization of the Japonic and Koreanic languages. — *The Oxford guide to the Transeurasian languages*. Edited by Martine Robbeets, Alexander Savelyev. Oxford: Oxford University Press, 9–21.
- Narrog, Heiko. 1999. *Japanische Verbflexive und flektierbare Verbalsuffixe*. Wiesbaden: Harrassowitz Verlag.
- Osterkamp, Sven. 2021. The last days of Old Japanese. Early Heian gloss texts and the periodization of Japanese language history. — *Studies in Asian historical linguistics, philology and beyond. Festschrift presented to Alexander V. Vovin in honor of his 60th birthday*. Edited by John Kupchik, José Andrés Alonso de la Fuente, Marc Hideo Miyake. Leiden – Boston: Brill, 5–25.
- Rickmeyer, Jens. 2017. *Einführung in das Klassische Japanisch anhand der Gedichtanthologie «Hyakuninshū»*. München: Iudicium Verlag.
- Satō, Takeyoshi 佐藤武義 (ed.). 2001. *Gaisetsu Nihongo no rekishi*. Tōkyō: Asakura Shoten.
- Takeuchi, Lone. 1999. *The structure and history of Japanese. From Yamatokotoba to Nihongo*. London – New York: Longman.

The diachrony of tone: connecting the field

Sandra Auderset (MPI for Evolutionary Anthropology),
Rikker Dockum (Swarthmore College), Ryan Gehrman (Payap University)

Tone, that is the use of pitch to distinguish lexical and/or grammatical forms, is an integral feature of many—possibly a majority of—languages across the world (Yip 2002). Despite this, tonal phenomena are conspicuously absent from most studies on language change, so that interest and progress in the understanding of the origins and evolution of suprasegmental contrasts lags behind that of segmental contrasts (Janda & Joseph 2003, Dockum 2019, Campbell 2021).

Nevertheless, starting in the latter half of the 20th century, steady progress has been made in the investigation of tonogenesis, i.e. in the emergence of tonal contrasts. This research has identified various pathways for a language to acquire tonal contrasts from segments. Haudricourt's (1954) model of tonogenesis in the so-called *Sinospheric Tonbund* (Matisoff 2001) connects the emergence of tonal contrasts with originally segmental material and processes of simplification of syllable structure. Similar progressions, from segmental contrast to tonal, can be observed in other languages and language families, such as Athabaskan (Kingston 2005), Mayan (see discussion in Bennett 2016, 497-499), Uto-Aztecan (Manaster-Ramer 1986; Guion et al. 2010), Punjabi/Northwest Indo-Aryan (Baart 2014; Evans et al. 2018), Malagasy (Howe 2017), and Afrikaans (Coetzee et al. 2018), among others. Prosodic contrasts can also give rise to tones, as in Cushitic (Kießling 2004). Formerly predictable stress patterns, for which pitch has become salient, became unpredictable in conjunction with syllable or word structure reduction. Tonal contrasts can also be acquired through contact and bilingualism of a non-tonal language with a tonal one. Such developments have been observed in languages such as Southern Qiang (Evans 2001) and Mal (L-Thongkum & Intajamornrak 2008), for example.

There are, however, language families in which tonality has such a long history that its origins might not ever be discoverable. This is the case in Otomanguean, where all the daughter families are reconstructed as tonal, e.g. proto-Mixtecan (Longacre 1957) and proto-Mixtec (Dürr 1990), proto-Chinantec (Rensch 1968), proto-Oto-Pamean (Bartholomew 1965), proto-Mazatec (Gudschinsky 1958; Kirk 1966), proto-Popolocan (Gudschinsky 1959), proto-Zapotec (Benton 2001), and proto-Chatino (Campbell 2013). It is thus assumed that proto-Otomanguean also had tonal contrasts (Rensch 1976, Kaufman forthcoming). Proto-Niger-Congo has also been reconstructed with tonal contrasts (Hyman 2016). Tone change *per se* is much less well studied than tonogenesis, and often not addressed even in language families with old tone systems. This can be at least partially attributed to impressionistic statements on the volatility of tones (Ratliff 2015; Cahill 2011; Beam de Azcona 2007; Morey 2005; Dürr 1990, among others), leading to the assumption that tones play at best a minor role in unraveling the history of a language family.

There is thus a considerable gap in the field of historical linguistics when it comes to the diachronic study of tones. A welcome exception to this is the recent collected volume on tone neutralization and phonetic tone change, Kubozono and Giriko 2018, and see also the overview by Yang & Xu (2019) of existing tone change work in Asia. This gap also applies to computer-assisted

methods, such as automatic alignment and cognate detection (List et al. 2018), and quantitative methods, such as Bayesian phylogenetics (Greenhill et al. 2020), which have gained traction in the field over the past two decades. Studies using such methodologies have been applied to few language families with tonal contrasts (e.g. Sagart et al. 2019 and Zhang et al. 2019, both on Sino-Tibetan) and none have addressed tone, despite evidence of historical tone categories having significant phylogenetic signal (Dockum 2019).

Workshop content and goals

As a result of the issues described above, comparatively few linguists focus on the diachronic study of tone. Individual specialists tend to sort themselves into regional and language family niches, leaving the field fragmented with little dialogue or cross-pollination between interested scholars. Given that the diachronic study of tone is in need of intensified research, the absence of exchange between scholars creates a further impediment to progress in this area.

This workshop brings together linguists from different regions and language families who work on tone diachrony and initiating an ongoing dialogue. Our goal is to form and strengthen collaborations between participants and attendants to advance this research area in the future. Presentations address topics including but not limited to:

- phonological environments that condition the emergence of tone contrasts or tone changes in existing tones;
- morphosyntactic patterns involving the innovation of new tone contrasts or changes to existing tone contrasts;
- underlying articulatory, acoustic, and perceptual mechanisms of tonogenesis and/or tone change;
- methodological considerations in the study of tone diachrony, e.g. the comparability of tonal systems in the absence of detailed phonetic studies, and the creation of reusable datasets and databases;
- addressing similarities and differences, both theoretically and empirically, in the study of tonal and segmental change;
- the contribution of tone to our understanding of the linguistic past, including subgrouping and classification in a language family, explaining historical contact phenomena between languages and language families, etc.;
- the relationship of historical tone studies with language documentation and description of tonal languages and language families;
- descriptions of tone change in under-described languages

References

Baart, Joan. 2014. Tone and stress in north-west Indo-Aryan: A survey. In Caspers et al. (eds.), *Above and Beyond the Segments: Experimental Linguistics and Phonetics*, 1-13. Amsterdam & Philadelphia: John Benjamins Publishing Company.

- Beam de Azcona, Rosemary G. 2007. Problems in Zapotec tone reconstruction. In *Annual Meeting of the Berkeley Linguistics Society*, 3–15.
- Bennett, Ryan. 2016. Mayan phonology. *Language and Linguistics Compass* 10(10), 469–514.
- Benton, Joe. 2001. *A reconstruction of the tone system of Proto Zapotec*. Unpublished manuscript.
- Cahill, Michael. 2011. *Tonal diversity in languages of Papua New Guinea*. SIL Electronic Working Papers.
- Campbell, Eric W. 2013. The internal diversification and subgrouping of Chatino. *International Journal of American Linguistics*, 79(3), 395–420.
- Campbell, Eric W. 2021. Why is tone change still poorly understood, and how might documentation of less-studied tone languages help? In Patience Epps, Danny Law & Na'ama Pat-El (eds.), *Historical Linguistics and Endangered Languages: Exploring Diversity in Language Change*. New York & London: Routledge.
- Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66, 185-216.
- Dockum, Rikker. 2019. *The Tonal Comparative Method: Tai tone in diachronic perspective*. New Haven: Yale University PhD thesis.
- Dürr, Michael. 1990. A preliminary reconstruction of the Proto-Mixtec tonal system. *Indiana* 11: 19-61.
- Evans, Jonathan. 2001. Contact-induced tonogenesis in Southern Qiang. *Language and Linguistics* 2(2), 63-110.
- Evans, Jonathan, Wen-Chi Yeh & Rukmini Kulkarni. 2018. Acoustics of tone in Indian Punjabi. *Transactions of the Philological Society* 116(3), 509-528.
- Greenhill, Simon, Paul Heggarty, Russell D. Gray. 2020. Bayesian phylolinguistics. In Janda, Richard, Brian D. Joseph & Barbara S. Vance (eds.), *The Handbook of Historical Linguistics*, Vol. 2. pp. 226-253. Wiley Blackwell
- Gudschinsky, Sarah C. 1958. Mazatec dialect history: a study in miniature. *Language* 34(4), 469–481.
- Gudschinsky, Sarah C. 1959. *Proto-Popotecan: a comparative study of Popolocan and Mixtecan*. Baltimore: Waverly Press.
- Guion, Susan G., Jonathan D. Amith, Christopher S. Doty & Irina A. Shport. 2010. Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language. *Journal of Phonetics* 38(2), 137-166.
- Haudricourt, André-Georges. 1954. De l'origine des tons en Viêtnamien. [On the origin of tones in Vietnamese]. *Journal Asiatique* 242, 69-82.
- Howe, Penelope Jane. 2017. *Tonogenesis in Central Dialects of Malagasy: Acoustic and Perceptual Evidence with Implications for Synchronic Mechanisms of Sound Change*. Houston: Rice University PhD thesis.

- Hyman, Larry. 2016. On Reconstructing Tone in Proto-Niger-Congo. *UC Berkeley PhonLab Annual Report*, 12.
- Janda, Richard D. & Brian D. Joseph. 2003. On language, change, and language change – or, of history, linguistics, and historical linguistics. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, pp. 3-180. Blackwell.
- Kaufman, Terrence. forthcoming. Comparative Oto-Mangean grammar research: Phonology, aspect-mood marking, valency changers, nominalizers on verbs, numerals, pronouns, deictics, interrogatives, adpositionoids, noun classifiers, noun inflexion, compounds, word order, and diversification model. In Wichmann, Søren (ed.), *Languages and linguistics of Mexico and Northern Central America: a comprehensive guide*. De Gruyter Mouton.
- Kießling, Roland. 2004. Tonogenesis in Southern Cushitic (Common West Rift). In Rose-Juliet Anyanwu (ed.), *Stress and Tone – the African Experience*, 141-163. (Frankfurter Afrikanistische Blätter 15). Köln: Rüdiger Köppe Verlag.
- Kingston, John. 2005. The phonetics of Athabaskan tonogenesis. In Hargus, Sharon & Karen Rice (eds.), *Athabaskan Prosody*, 137-184. (Current Issues in Linguistic Theory 269). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Kirk, Paul L.. 1966. Proto-Mazatec phonology. PhD dissertation, University of Washington.
- Kubozono, Haruo & Mikio Giriko. 2018. *Tonal change and neutralization*. Berlin, Boston: De Gruyter Mouton.
- L-Thongkum, Theraphan & Chommanad Intajamornrak. 2008. Tonal evolution induced by language contact: A case study of the T'in (Lua') language of Nan province, northern Thailand. *Mon-Khmer Studies* 38, 57-68.
- List, Johann-Mattis, Mary Walworth, Simon J. Greenhill, Tiago Tresoldi, & Robert Forkel. 2018. Sequence comparison in computational historical linguistics. *Journal of Language Evolution* 3(2), 130–144.
- Longacre, Robert E. 1957. *Proto-Mixtecan*. Indiana University.
- Manaster-Ramer, Alexis. 1986. Genesis of Hopi tones. *International Journal of American Linguistics* 52(2), 154-160.
- Matisoff, James. 2001. Genetic versus contact relationship: Prosodic diffusibility in South-East Asian languages. In Alexandra Y. Aikhenvald & R.M.W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, 291–327. Oxford: Oxford University Press.
- Morey, Stephen. 2005. Tonal change in the Tai languages of Northeast India. *Linguistics of the Tibeto-Burman Area*, 28(2), 139–202.
- Ratliff, Martha. 2015. Tonoexodus, tonogenesis, and tone change. In Honeybone, Patrick & Joseph Salmons (eds.). *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press. pp. 245-261.
- Rensch, Calvin R. 1968. *Proto Chinantec phonology*. Papeles de la Chinantla 6. México, D.F: Museo Nacional de Antropología.

- Rensch, Calvin R. 1976. *Comparative Otomanguean Phonology*. Bloomington: Indiana University Press.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, & Johann-Mattis List. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences of the United States of America* 116(21), 10317-10322.
- Yang, Cathryn and Yi Xu. 2019 Cross-linguistic trends in tone change: A review of tone change studies in East and Southeast Asia. *Diachronica* 36(3), 417-459.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
- Zhang, Menghan, Shi Yan, Wuyun Pan & Li Jin. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569, 112–115.
-

Tone, stress and length interactions in Central Neo-Štokavian

Đorđe Božović (University of Belgrade)

This paper discusses a set of hitherto unobserved or underdescribed prosodic phenomena in the Central Neo-Štokavian (sub)dialect of Serbo-Croatian (SCr). SCr has inherited and innovated in various ways a relatively complex pitch-accent system from Proto-Slavic, based on the interaction of tone, length and stress (ictus). Tone and stress interactions are particularly complex in the Neo-Štokavian dialect, where an innovative Stress Retraction rule (SRR) has operated to produce a rather skewed distribution of pitch contours. Namely, as a result of SRR, stress coincides with a H tone in Neo-Štokavian only if the syllable bearing H is word-initial. Otherwise, stress is assigned to the syllable immediately preceding H, thus producing a rising contour tone (L*H). Standard SCr is based on this prosodic system (cf. Lehiste & Ivić 1986, Inkelas & Zec 1988, Zec & Zsiga 2009).

The most innovative central group of Neo-Štokavian dialects (spoken in Bosnia and adjacent areas), however, tends to differ prosodically from standard SCr in a number of ways. First of all, they are characterized by a length-based qualitative vowel reduction, affecting all tonic and posttonic syllables. When disyllabic words with a rising pitch on the initial syllable are affected by the reduction, they will surface as monosyllables with a rising pitch (e.g. *kōnji* ‘horses’ > *kōnj*, *dóđi* ‘come-IMP’ > *dóđ*), a situation dispreferred in standard SCr. In addition, in a number of polysyllabic words, an innovative rising pitch appears on the initial syllable instead of the etymological falling one (e.g. *májka* ‘mother’ for the etymological *májka*, *kárta* ‘card’ for the etymological *kârta*, etc.), thereby effectively manifesting tone reversal.

In this paper, I examine more closely the diachronic evolution of the Neo-Štokavian prosodic system, in the context of tone-stress interaction and co-evolution over time. I show how Central Neo-Štokavian prosodic innovations represent a series of repair strategies for the Neo-Štokavian skewed distribution of pitch contours, that was brought about by the SRR. As a consequence, Central Neo-Štokavian features as a transitory idiom between different tone-stress interaction types (from tone governing stress, as in standard SCr, to dominantly stress governing tone), but also from a relatively complex and unstable standard SCr pitch-accent system to a more stable, but still typologically awkward, rising-contour initial stress system, with word-initial L tone attracting stress while avoiding the lexical H altogether. Therefore it doesn’t fit easily into the general typology of tone and stress interactions (de Lacy 2002), where systems such as Central Neo-Štokavian, in which stress tends to be attracted to L tone while simultaneously H tone is ignored, are explicitly excluded. In that sense, the innovative Central Neo-Štokavian data discussed here may contribute to a more fine-grained understanding of the possible tone-stress interaction types, but also of the exact mechanisms and motivations for tonal change and transition between different prosodic systems.

References

- de Lacy, P. (2002). The interaction of tone and stress in Optimality Theory. *Phonology* 19(1): 1–32.

Inkelas, S. & D. Zec (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language* 64(2): 227–248.

Lehiste, I. & P. Ivić (1986). *Word and Sentence Prosody in Serbocroatian*. MIT Press.

Zec, D. & E. Zsiga (2009). Interaction of tone and stress in Standard Serbian: phonological and phonetic evidence. In W. Browne et al. (eds), *Formal Approaches to Slavic Linguistics 18: The Cornell Meeting*. Ann Arbor: Michigan Slavic Publications, 537–555.

Accent and tone: the double origin of the Paicî tone system

Florian Lionnet (Princeton University)

Introduction. Tone mostly originates from the phonologization of redundant F0 differences caused by segmental laryngeal features – i.e., from non-prosodic features (Michaud and Sands 2020, a.o.). Languages where tone originated from prosodic features (e.g. accent) appear to be the exception rather than the rule (e.g., Scandinavian languages, cf. Kingston 2011). What has yet to be fully documented is what happens when an accentual language develops a tonal contrast from laryngeal features: is the accent system kept alongside the new tone contrast? Is it lost? Do both prosodic systems merge?

In this paper, I present the typologically interesting tone system of Paicî (Oceanic, New Caledonia), which illustrates the latter outcome: tonogenesis occurred in an accentual language and caused the former accent system to be reinterpreted as tonal.

Paicî data – Tone. Paicî is one of five tonal languages of New Caledonia, all Oceanic (Grace 1955; Haudricourt 1968, 1971; Rivierre 1993, 2001). These languages are famous for being the only Oceanic languages that have developed tone without any external influence. They are thus particularly interesting for what they might reveal about the historical development of tone systems.

Paicî has arguably the most complex tone system in New Caledonia (Rivierre 1974, 1993, 2001). The tonal inventory itself is simple, with only two tonemes, H(igh) and L(ow), as shown in (1).

- (1) *i* ‘to cry’ *i* ‘louse’
pádi ‘to thrash’ *pàdi* ‘to divide’

This tonal contrast results from the transphonologization of a former aspiration contrast on plosives and voicing contrast on sonorants (Rivierre 1993, 2001), e.g. *tii* ‘to strip bark’ (cf. Nemi *t^{hi}-*) vs. *tìi* ‘letter, book’ (cf. Nemi *tii*).

Paicî data – Downstep. More complex is the behavior and history of downstep in Paicî. Two types of downstep are attested: a metrically conditioned downstep systematically marking the boundary between the first two L-toned bimoraic feet within a prosodic word (2), and underlying downstep found in about 20 tonal enclitics (3) (cf. Rivierre 1974, Lionnet 2022).

- (2) /tèèpàà/ → (tèè)[↓](pàà) ‘to arrive’
(3) /gò =[↓]i bwə / → gò =[↓]i bwə ‘on the banyan tree’
on =DET banyan

The Paicî downstep has many typologically rare properties (Lionnet 2022): (i) it affects only L, and is incompatible with H; (ii) it is mostly autonomous from lexical tone; (iii) it is culminative (i.e. there cannot be more than one) within the prosodic word; (iv) it is (partly) metrically conditioned; and (v) it is realized utterance-initially. Properties (ii)-(v) give it a strong accentual flavor, giving the impression of two parallel prosodic systems in Paicî : a H vs. L tonal contrast, and a downstep-based accent.

Comparative data and diachronic hypothesis: Comparative evidence from neighboring (non-tonal) Xârâcùù strongly suggests that downstep in Paicî was indeed originally an accentual system (Rivierre 1978). It is not accentual anymore in contemporary Paicî, as can be seen from the fact that it lacks obligatoriness, one of the definitional criteria of accent (cf. Hyman 2006) – indeed it is never found with H-toned words, i.e. one third of the lexicon. It can even be shown to interact with the tonal system – at least in stating distributional constraints, e.g. “no downstep on words carrying a H tone”.

The complexity of the Paicî tone system is thus the result of its double historical origin: accentual and tonal. Tonogenesis innovated a H tone in a downstep-marked accentual system. Consequently, the innovative H-toned words were, so to speak, removed from the “regular” accent system, while the rest of the lexicon maintained its former accentual behavior, only reinterpreted as involving a L tone, as a consequence of which downstep was reinterpreted as a property of L-toned words. This double origin offers a simple explanation for the typologically rare features of the Paicî downstep listed above.

References

- Grace, George W. 1955. Notes: Paicî. Field notes. University of Hawai‘i at Mānoa Digital Media Collection.
- Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257.
- Haudricourt, André. 1968. La langue de Gomen et la langue de Touho en Nouvelle Calédonie. *Bulletin de la Société de Linguistique de Paris* 63(1). 218–235.
- Haudricourt, André Georges. 1971. New Caledonia and the Loyalty Islands. In Thomas A. Sebeok (ed.), *Linguistics in Oceania* (Current Trends in Linguistics 8), 359–396. Berlin: Mouton de Gruyter.
- Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23(2): 225–257. DOI: <https://doi.org/10.1017/S0952675706000893>.
- Kingston, J. (2011). Tonogenesis. M. van Oostendorp, C. J. Ewen, E. Hume, and K. Rice (eds.) *Blackwell Companion to Phonology*, v. 4, (pp. 2304-2334). Oxford, UK: Blackwell Publishing.
- Lionnet, Florian. 2022. Tone and downstep in Paicî (Oceanic, New Caledonia). *Phonological Data and Analysis* 4(1): 1-47. DOI: <https://doi.org/10.3765/pda.v4art1.45>
- Michaud, Alexis and Bonny Sands. 2020. Tonogenesis. In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Rivierre, Jean-Claude. 1974. Tons et segments du discours en paicî (Nouvelle-Calédonie). *Bulletin de la Société de Linguistique de Paris* 69(1). 325–340.
- Rivierre, Jean-Claude. 1978. Accents, tons et inversion tonale en Nouvelle-Calédonie. *Bulletin de la Société de Linguistique de Paris* 73(1). 415–43.
- Rivierre, Jean-Claude. 1993. Tonogenesis in New Caledonia. *Oceanic Linguistics Special Publications* 24. 155–173.

Rivierre, Jean-Claude. 2001. Tonogenesis and evolution of the tonal systems in New Caledonia: The example of Cèmuhî. In Shigeki Kaji (ed.), *Proceedings of the symposium Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Japanese accentology, and other topics*, 23–42. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa (ILCAA).

Tone and voicing in Cao Bằng Tai: implications for tonal evolution and change

James Kirby (Institute for Phonetics and Speech Processing, LMU Munich)

Pittayawat Pittayaporn (Southeast Asian Linguistics Research Unit, Chulalongkorn University)

This study examines the phonetic variation in the realization of lexical tones and onset consonants in the Tai dialect of Cao Bằng province (CBT), Vietnam [tyz]. As a rare specimen of a language in the late stages of a tone split, CBT is important for our general understanding of tonal diachrony, because it provides us a chance to study up close how phonetic realization and phonological patterning interact in phonologization.

At the time the binary register split, CBT had a four-way contrast between aspirated $*p^h$ -, unaspirated $*p$ -, modal $*b$ - and implosive $*ɓ$ - onsets, in addition to voiced and voiceless fricatives and sonorants ($*f$ - $*v$ - $*m$ - $*^hm$ -). Subsequently, tones following $*b$ -, $*m$ - and $*v$ - were lowered in pitch; the historically modally voiced stops $*b$ - became breathy-voiced $/b̤$ -, while implosive $*ɓ$ - became modal voiced $/b$ -; and the voicing contrast in sonorants was ostensibly lost ($*^hm$ - and $*m$ - $>$ $/m$ -). This resulted in a system where six tones are found in syllables headed by sonorants, while in the obstruent sub-system, only high-register tones (1, 3, 5) are found after $/b$ - p - p^h -/ and only low-register tones (2, 4, 6) are found after $/v$ -/ and $/b̤$ -/ (Hoàng Văn Ma 1997; Pittayaporn 2009).

The details of the process, however, remain murky. Previous work (Hoàng Văn Ma 1997; Pittayaporn & Kirby 2017) indicated that older CBT speakers may still realize $*b$ as breathy-voiced $[b̤]$, while younger speakers may produce a devoiced $[p̥]$, consistent with tonogenetic models which ascribe a central role to voice quality (Thurgood 2002). However, acoustic-phonetic and electroglottographic data from a more recent age- and gender-stratified sample of 19 speakers revealed a linguistically homogeneous speech community, in which $*b > /b̤$ - appears to have merged acoustically with $/p^h$ -. This means that tonal register is no longer predictable after $/p^h$ -, illustrating a further step in the evolution of how tone splits evolve over time.

In addition, we observed some unexpected differences in the pitch trajectories following different onset types. For syllables bearing high-register tones (1, 3, and 5), we found a marked tendency for raised f_0 following $/p^h$ -/ compared to $/b$ -/ and (in some cases) $/p$ -/, but $/m$ -/ was found to pattern with $/p^h$ -/ in this respect. For syllables bearing low-register tones (2, 4, and 6), we observed a less marked tendency for raised f_0 following $/b̤$ -/ $>$ $[p^h]$ -/ and $/v$ -/ relative to $/m$ -. These findings suggest that the historically voiceless sonorants still retain at least some of the acoustic properties they presumably shared with voiceless plosives prior to conditioning the tone split. In other words, in an important sense there are still two kinds of sonorants in CBT, illustrating that the phonetic specification of tonal categories may be much richer than the apparent phonological patterns suggest.

These findings show that the timing of the interrelated processes of tonal register split and neutralization of laryngeal contrast may be more complex than previously assumed: while historical sonorant voicing may well have conditioned the initial split (L-Thongkum 1997), it does not appear that the sonorant voicing merger must necessarily be completed prior to the devoicing of originally voiced obstruents. Our study of CBT thus provides new insights into the internal complexities of the tonogenetic process, reminding us that the temporal ordering implied by stages in diachronic models cannot be taken too literally, and illustrating how careful analysis of synchronic acoustic-phonetic variability can contribute to our understanding of the diachrony of tone change.

References

- Hoàng Văn Ma. 1997. The sound system of the Tày language of Cao Bằng province, Vietnam. In Jerold A. Edmondson & David B. Solnit (eds.), *Comparative Kadai: the Tai branch*, 221–231. Dallas, TX: Summer Institute of Linguistics.
- L-Thongkum, Theraphan. 1997. Implications of the retention of proto-voiced plosives and fricatives in the Dai Tho language of Yunnan province for a theory of tonal development and tai language classification. In Jerold A. Edmondson & David B. Solnit (eds.), *Comparative Kadai: the Tai branch*, 191–219. Dallas, TX: Summer Institute of Linguistics.
- Pittayaporn, Pittayawat. 2009. *The phonology of proto-Tai*. Cornell University PhD dissertation. <http://ecommons.cornell.edu/handle/1813/13855>. (24 April, 2015)
- Pittayaporn, Pittayawat & James Kirby. 2017. Laryngeal contrasts in the Tai dialect of Cao Bằng. *Journal of the International Phonetic Association* 47(1). 65–85. <https://doi.org/10.1017/S0025100316000293>.
- Thurgood, Graham. 2002. Vietnamese tone: revising the model and the analysis. *Diachronica* 19(2). 333–363.
-

Tone splits from vowel height in the Austronesian language of Raja Ampat

Laura Arnold (University of Edinburgh)

Tone is very rare in Austronesian languages: of the 1,200 or so languages in the family, only around two dozen are described as tonal (Blust 2013: 657–659; Himmelmann & Kaufmann 2020: 371–372). Austronesian languages are thus rarely included in typological overviews of and theoretical discussions about tone. In this contribution to the workshop, I seek to boost the mainstream visibility of the diachrony of tone in the Austronesian family, by discussing a cross-linguistically rare sound change that has occurred in multiple Austronesian languages spoken in the Raja Ampat archipelago: splits in tone systems conditioned by vowel height.

Raja Ampat lies just off the northwest tip of the island of New Guinea. It is home to six Austronesian languages, all of which belong to the understudied South Halmahera-West New Guinea subbranch of the family. All six languages are tonal. The tone systems are typologically diverse: from Ambel, which has a single underlying tone contrasting with toneless syllables in a system that is culminative but not obligatory (Arnold 2018a); through languages which distinguish two (Ma'ya, Salawati, Biga) or three (Batta) underlying tones on word final syllables (van der Leeden 1993; Remijsen 2001; Arnold 2021); to Matbat, with six tones that can occur anywhere in the word (Remijsen 2001, 2007).

Tone splits conditioned by vowel height have occurred at least twice in Raja Ampat, in the ancestors of Ma'ya and Ambel. In an ancestor of Ma'ya, reconstructed *High tone split: it remained High on syllables with close vowel nuclei, and merged with Rise elsewhere (Arnold 2018b). In proto-Ambel, toneless syllables remained toneless if the vowel was close, otherwise merged with High tone (Arnold 2020). Preliminary investigations suggest that similar splits may have occurred several more times in the Raja Ampat languages.

In this talk, I will exemplify the Ma'ya and Ambel splits, and touch on two points of theoretical significance. First, only a handful of other tone changes conditioned by vowel quality have so far been attested worldwide (Kingston 2011; Köhnlein & van Oostendorp 2017; Michaud & Sands 2020). The Ma'ya and Ambel splits thus contribute to the growing body of evidence demonstrating that, contrary to what some have claimed (e.g. Hombert 1977; Hombert et al. 1979), vowel height can and does condition diachronic tonal developments. Second, in all attested cases thus far, syllables with close vowel nuclei develop higher tones; the Ambel split, in which non-close vowel nuclei developed High tone, has not previously been attested.

I conclude this talk by discussing a phonetic mechanism that may explain the unusually frequent tone changes conditioned by vowel height in Raja Ampat: the phonologisation of differences in the intrinsic fundamental frequency (IF0) of vowels in these languages. IF0 is a near-universal phenomenon in which, all else being equal, close vowels are produced with a higher F0 than open vowels; cross-linguistically, the mean difference in IF0 between close and open vowels is 1.65 semitones (Whalen & Levitt 1995). Recent production data, however, suggests that IF0 differences in the Raja Ampat languages are much larger than average: in Salawati and Biga, the mean difference is as large as 2.8 semitones in some contexts (Arnold et al. submitted). As

well as describing the phonological environments that condition tone changes in these understudied languages, this talk will therefore also provide a potential articulatory explanation for these changes, thus deepening our theoretical understanding of tonal diachrony more generally.

References

- Arnold, Laura. 2018a. Lexical tone in Mentyo Ambel. *Oceanic Linguistics* 57(1): 199-220.
- Arnold, Laura, 2018b. A preliminary archaeology of tone in Raja Ampat. In Antoinette Schapper (ed.), *Contact and substrate in the languages of Wallacea, part 2*, 7–39.
- Arnold, Laura, 2020. Highs and lows: Towards reconstructing the word-prosodic system of proto-Ambel. *Transactions of the Philological Society*, 118: 141–158.
- Arnold, Laura. 2021. The tonal phonologies of three undocumented Raja Ampat languages. Paper presented at the 15th International Conference on Austronesian Linguistics, Palacký University, Olomouc (online), 28 June–2 July, 2021.
- Blust, Robert. 2013. *The Austronesian Languages*, revised edition. (Asia-Pacific Linguistics Open Access Monographs). Canberra: Research School of Pacific and Asian Studies, The Australian National University.
- Himmelman, Nikolaus P. and Daniel Kaufmann. 2020. Austronesia. In: Carlos Gussenhoven and Aoju Chen (eds.), *The Oxford handbook of language prosody*, 370–383. Oxford: Oxford University Press.
- Hombert, Jean-Marie. 1977. Development of tones from vowel height? *Journal of Phonetics* 5: 9-16.
- Hombert, Jean-Marie, John Ohala, and William G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55(1):37–58.
- Kingston, John. 2011. Tonogenesis. In: Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume and Keren Rice (eds.), *The Blackwell companion to phonology*, vol. IV, 2304–2333. Wiley-Blackwell: Malden, MA.
- Köhnlein, Björn and Marc van Oostendorp. 2017. Introduction. In: Wolfgang Kehrein, Björn Köhnlein, Paul Boersma and Marc van Oostendorp (eds.), *Segmental structure and tone* (Linguistische Arbeiten 552), 1–10. Berlin: de Gruyter.
- van der Leeden, A. C. 1993. *Ma'ya: A language study. A: Phonology*. Jakarta: Lipi-Rul.
- Michaud, Alexis and Bonnie Sands. 2020. Tonogenesis. In: Mark Aronoff (ed), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.
- Remijsen, Bert. 2001. *Word-prosodic systems of Raja Ampat languages*. Utrecht: LOT.
- Remijsen, Bert. 2007. Lexical tone in Magey Matbat. In: Vincent J. van Heuven and Ellen van Zenten (eds.), *Prosody in Indonesian languages*, 9–34. Utrecht: LOT.
- Whalen, D. H. and Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23:349–66.
-

A diachronic study of grammatical tone in northwestern Bantu

Nadine Grimm (University of Rochester)

Tone in Bantu languages: There is considerable variation among the around 500 Bantu languages with respect to tone systems, ranging from fully tonal languages (e.g. Kikuyu) to non-tonal languages (e.g. Swahili) and a “bewildering variety of intermediate types” (Clements & Goldsmith 1984). Meeussen (1967) reconstructs Proto-Bantu with two tonemes: H and L, which are associated with vowels and syllabic nasals. Synchronically, Bantu languages vary a great deal in the nature of their tonal systems and the functions that tone encodes, despite some common themes such as tonal inventories and processes or functional domains. While there is a decent amount of reconstruction accomplished for lexical tone in Bantu (Meeussen 1967, Hyman & Katamba 1990, Marlo & Odden 2019) and, more broadly, Niger-Congo (Hyman 2016a), grammatical tone and its historical changes is vastly understudied.

Grammatical tone in northwestern Bantu: Northwestern (NW) Bantu languages show significant differences from PB and synchronic Bantu languages of other areas (South, East, and West). They are often reported to distinguish L, H and Ø TBUs and exhibit similar tonal processes (floating tones, high tone spreading). However, “the nature of these [tonal] systems as a whole is not well understood” (Odden & Bickmore, 2014: 3). Reasons for that include the fact that i) distinctive tones are still left out of some descriptions or only described for certain parts of the grammar, ii) there is a bias towards eastern Bantu languages in tone descriptions, which have very different tonal systems, and iii) we are still lacking an investigative framework to collect and compare tonal data, with the result that data is not complete and/or comparable (but see Marlo 2013).

In this talk, I investigate an important tonal feature that is common in NW Bantu, namely grammatical H tones in the verb phrase. These floating H tones share similar targets across languages of the area (subject markers, the finite verb, phrase-medial verb position, object noun class prefixes) and surface under similar conditions, often determined by certain tense/aspect/mood categories. Tonal phenomena of H tones in this domain have been described in the literature under differing terms in different Bantu areas, including “metatony” in NW Bantu (Hyman & Lionnet 2011), a “conjoint/disjoint distinction” in eastern and southern Bantu (Hyman 2016b), and “tone-case” in western Bantu (Kavari et al. 2012, Van der Wal 2015). It is, however, unclear how they developed historically. Based on data from the literature on NW Bantu languages (e.g. Bakweri, Basaa, Abo, Mpiemo, Kwasio, Eton) and my own fieldwork on Gyeli (Grimm 2021), I propose that tonal change in NW Bantu is largely driven by changes in phonology (segmental loss and constraints on syllable numbers) and interrelated morphological complexity. In languages where grammatical tone is a co-phonology of segmental material, e.g. a tense marker, its functional load is relatively weak and it is often not clear what tone contributes to the meaning or functional category. In contrast, when segmental material erodes and only the tone survives, tone takes on a higher functional load and may develop into entirely tonal paradigms to distinguish functional categories. Such an extreme case is found in Gyeli (Grimm, forthcoming).

Advances in the understanding of GT and its historical dimension will not only shed light on patterns in tone system changes, but also contribute to current unknowns, such as quantifying a language as to the degree to which it employs grammatical tone (Rolle 2018), identifying the exact range of grammatical functions tone can encode, areas of grammar where tone carries more grammatical functions than in others, and identifying grammatical categories and sub-categories that are expected to be marked tonally in Bantu languages.

References

- Clements, G.N. & John Goldsmith. 1984. *Autosegmental studies in Bantu tone: Introduction*. Dordrecht: Foris.
- Grimm, Nadine. 2021. *A grammar of Gyeli*. Berlin: Language Science Press.
- Grimm, Nadine. Forthcoming. Grammatical tone in Gyeli: the interaction between phonology, grammatical functions, and lexical tone. *Phonology. Special Issue: Theoretical approaches to grammatical tone*.
- Hyman, Larry M. & Francis X. Katamba. 1990. Spurious high-tone extensions in Luganda. *South African Journal of African Languages* 10, 142-158.
- Hyman, Larry M. & Florian Lionnet. 2011. Metatony in Abo (Bankon), A42. *UC Berkeley Phonology Lab Annual Report*. pp. 168-182.
- Hyman, Larry M. 2016a. On reconstructing tone in Proto-Niger-Congo. *UC Berkeley PhonLab Annual Report*, 12.
- Hyman, Larry M. 2016b. Disentangling conjoint, disjoint, metatony, tone cases, augments, prosody, and focus in Bantu. In Van der Wal, Jenneke and Larry M. Hyman (Eds.) *The Conjoint/Disjoint Alternation in Bantu*. Berlin: De Gruyter Mouton, pp. 100-121.
- Kavari, Jekura, Lutz Marten and Jenneke van der Wal. 2012. Tone cases in Otjiherero: head-complement relations, linear order, and information structure. *Africana Linguistica* 18: 315-353.
- Marlo, Michael R. 2013. Verb tone in Bantu languages: micro-typological patterns and research methods. *Africana Linguistica* 19: 137-234.
- Marlo, Michael R. & David Odden. 2019. Tone. In Van de Velde, Mark et al. (Eds.) *The Bantu languages*. London: Routledge, pp. 150-171.
- Meeussen, A. E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3: 79–121.
- Odden, David & Lee Bickmore. 2014. Melodic tone in Bantu. *Africana Linguistica* 20: 3-13.
- Rolle, Nicholas R. 2018. *Grammatical Tone: Typology and Theory*. PhD Dissertation, UC Berkeley.
- Van der Wal, Jenneke. 2015. Bantu syntax. Oxford Handbooks Online.
-

A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis

Lilja Sæbø (University of Oxford), Steven Moran (University of Neuchâtel),
Eitan Grossman (Hebrew University of Jerusalem)

Tonogenesis has become a topic of increasing interest, including numerous descriptions of tonogenetic events across a variety of language families (Haudricourt 1954, Arnold 2020, Hyslop 2009) with much recent work towards models (Dockum 2019, Gehrman 2022) and typologies of tonogenesis (Hyman 2018, Hyslop submitted). To aid in these endeavors, we have constructed a preliminary database of tonogenetic events (DTE), which aims to organize the documentation that exists on tonogenesis into a format that is easy to filter, search, and compare.

The DTE currently describes 229 tonogenetic events from 90 language varieties in 26 language families across five macro-areas (North America, Africa, Asia, Europe and Papunesia). Along with language variety metadata, the database includes information about the triggering context of each tonogenetic event, the resultant tone, and a description of the effect that the tonogenetic event had on pitch. The triggering contexts have been divided into five main groups: onset, coda, nucleus (i.e. tonogenesis triggered by a feature of the vowel itself, such as for instance height (Ratliff 2015:253), stress and word type (i.e. tonogenesis triggered by the loss or gain of a syllable). Thus, the coding of the DTE provides a general classification of tonogenetic events reported in the literature, making it straightforward to see what kind of triggering contexts result in what kind of tones.

Our goal is to provide an open source and easily accessible resource that can be harnessed to ask, and answer, typological questions related to tonogenesis. Towards this aim, we have so far investigated two areas of tonogenesis that we will report on in this talk. The first is the relationship between the triggering context and the resultant tone, and the second is an areal classification of the distribution of different types of tonogenesis.

As to the first issue, our data largely concurs with trends already reported in the literature, such as the fact that voiced onsets tend to give rise to lower tones than voiceless onsets (Kingston 2011, Hombert, Ohala and Ewan 1979). While some triggers almost always have a given effect on the pitch, there were other triggers that could have various different effects. An example is voiceless aspirated and unaspirated stops, where there is no clear trend as to what trigger will give the highest tone. Regarding areality, the DTE for example shows that among the languages in the sample from Asia, it is very common to have undergone a two step tonogenetic process similar to that described for Vietnamese by Haudricourt (1954). That is, in the first step the coda consonants first create two or more different contour tones, and in a second step these tones are doubled by a second series of tonogenetic events based on the voicing/aspiration of the onset, generally creating a high and a low register. In our data, this kind of tonogenesis is not found outside of Asia. Other areal trends include the fact that the most common context for tonogenesis in North America is codas, while word-type tonogenesis is the most common in Europe. In this talk, we will discuss both topics in greater detail.

In accordance with principles for open cross-linguistic typological research put forth in the Cross-Linguistic Linked Data (CLLD) project (Forkel et al. 2018), we are making the DTE available as a CLDF data set. This makes its contents interoperable with Glottolog and numerous other typological databases (e.g., WALS, PHOIBLE, ASJP), so that new types of questions can be asked about tonogenesis. For example, one can investigate whether there are any correlations between the current phonological system of a language and the types of tonogenesis that it has undergone. Thus, in line with the goals of this workshop, we aim to bring together linguists and data sources to generate dialogue and collaboration to shed light on the diachronic study of tone and the processes of tonogenesis.

References

- Arnold, L. 2020. Highs and lows: A previously unattested tone split from vowel height in Metnyo Ambel. *Transactions of the Philological Society*, 118, 141–158.
- Dockum, R. 2019. The Tonal Comparative Method: Tai Tone in Historical Perspective. PhD thesis, Yale University.
- Forkel, R., et al. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data*, 5:180205.
- Gehrmann, R. 2022. Desegmentalization: Towards a Common Framework for the Modeling of Tonogenesis and Registrogenesis in Mainland Southeast Asia with Case Studies from Austroasiatic. PhD thesis, The University of Edinburgh.
- Haudricourt, A. G. 1954. De l'origine des tons en vietnamien. *Journal Asiatique*, 242, 69–82. English translation by Marc Brunelle: The origin of tones in Vietnamese.
- Hombert, J. M, Ohala, J. J, and Ewan, W. G (1979). Phonetic explanation for the development of tones. *Language*, 55(1), 37–58.
- Hyman, L. 2018. Towards a Typology of Tone System Changes. In: H. Kubozono and M. Giriko (Eds.), *Tonal Change and Neutralization*, 203–222.
- Hyslop, G. 2009. Kurtöp Tone: A tonogenetic case study. *Lingua*, 119, 827–845.
- Hyslop, G. Submitted. Toward a Typology of Tonogenesis: refining the model.
- Kingston, J. (2011). Tonogenesis. In: Oostendorp, M., Hume, E. V., and Ewen, C. J. eds. *The Blackwell Companion to Phonology*. John Wiley & Sons.
- Ratliff, M. 2015. Tonoexodus, Tonogenesis, And Tone Change. In: Honeybone, P., Salmons, J eds., *The Oxford Handbook Of Historical Phonology*. Oxford University Press. 245–261.
-

Tonal density and its correlation with the types of tonal systems: Diachronic aspects

Elena Perekhval'skaya (Institute for Linguistic Studies, Russian Academy of Sciences),

Valentin Vydrin (INALCO-LLACAN)

Tonal systems are usually compared by the number of tones, by the character of tones (level or modulated), and by their function (lexical or grammatical). Following Gussenhoven (2004: 34), we suggest to compare these systems also by the criterion of tonal density, and introduce a Tonal Density Index (TDI), which equals the number of tonemes per 100 syllables. In order to make the calculation of the tonal density possible, it is necessary to define some key notions: toneme, tonal domain, marked tone and default tone, tonal and toneless syllables.

Toneme as a meaningful tone, i.e. a tone (or tonal contour, in a larger sense) which can distinguish lexical and/or grammatical meanings. With some reserves, toneme can be compared with what is phoneme in the segmental phonology. For certain types of languages, distinction between tone and toneme is marginal, but for some others, it is of a primary importance. This term was introduced by Pike (1948), it reappears sporadically (Welmers 1959; Hyman & Leben 2021), however, we are unaware of any serious attempt to elaborate this notion and make it work.

The **tonal domain** is a segmental chain on which a toneme is realized. It may vary from zero (for floating tones) to a long sequence of syllables; its length is language-specific. In some languages, a segmental chain is entirely subdivided into tonal domains; in some others, certain segments can remain outside tonal domains.

With respect to the **marked and default tones**, it is important to distinguish between unmarked tones which can be still regarded as tonemes and those which should be interpreted as absence of tones (or zero tones).

Toneless is a syllable (or mora) to which no toneme is assigned at the underlying level. Toneless syllables are found even in some languages with very high tonal density.

When this approach is applied to tonal languages, it turns out that tonal systems can be roughly subdivided into three major types, and these types correlate with the tonal density: **omnisyllabic** type (the TDI close to 100); **tonemic** type (the TDI between 50 and 90); **privative** type, i.e. languages with marked and zero tones (the TDI is below 50); **pitch-accent** type (the TDI is below 30).

Distribution of tonal languages by these types follows a clusterization model, and existence of some intermediate (hybrid) types can be envisaged.

With respect to the tonal types, diachronic evolution of tonal systems can follow various patterns.

1) Emergence of tones in an originally toneless language (or loss of tones). This case can be illustrated by Tibetic languages going back to the atonal Old Tibetan which has split around 9-10 century. An atonal modern language Amdo (TDI = 0) retains voiced consonants as well as certain onset consonant clusters (Makley et al. 1999). In Utsang (Lhasa Tibetan), the tonogenesis resulted from the devoicing of originally voiced consonants and the simplification of the onset consonant clusters (Huang Bufan 1995). According to Jäschke (2018) and Tournadre & Dorje

(2003), Utsang has two meaningful tones (i.e. tonemes): high (H) and low (L). Their superficial realizations depend on the segmental structure of a syllable, the main factors being the syllable weight, and on the extension of the tonal domain which can be equal to one or two syllables. TDI for the Lhasa Tibetan is 51.

2) Mobility between the tonal types. For example, in the Baltic group (Daugavet 2012; Kushnir 2018), the Lithuanian Aukštaitian is a typical pitch-accent language. It has two tonemes (falling and rising) and a mobile stress (an accentuated syllable can occupy any position in a word-form). Only accentuated heavy syllables (i.e., containing long vowels, or diphthongs, or diphthongoids) carry meaningful tones. Light syllables (stressed or unstressed) and unstressed heavy syllables cannot carry meaningful tones. The TDI of this variety is 30.

The Old Latvian had a system close to the Lithuanian, but the situation in the modern literary Latvian is different. It has a word-initial stress and three tonemes: rising/high, falling and rising-falling, the latter also includes an interrupting phonation. Every heavy syllable (both stressed or unstressed) carries a meaningful tone, and light syllables are toneless. Because of the historical loss of short vowels and the subsequent syllabic contraction, two or more heavy (and, subsequently, tonal) syllables can appear in one wordform. As a result, the Latvian language has evolved from the pitch-accent type toward the omnisyllabic type, and its TDI equals 43; it can be regarded as a hybrid type.

3) Change of tonal density within the limits of one tonal type. For example, both Bambara (Manding < Western Mande, TDI = 66-67.5) and Kakabe (Mokole < Western Mande, TDI = 62.5) are tonemic languages with similar tonal systems. The decrease of TDI in Kakabe can be explained by a couple of innovations, such as: definitive loss of tones by light postpositions (in Bambara, this rule is facultative); reinforcement of the rule of the phrase-final high tone lowering if preceded by a low tone (in Bambara, this rule is also facultative); partial loss of tones by personal pronouns (Vydrina 2017).

References

- Daugavet, Anna. 2012. *Latyšskie i litovskie slogovyje aktsenty: registr i kontur* (Латышские и литовские слоговые акценты: регистр и контур) [Lettish and Lithuanian syllabic accents: Register and contour]. In. St. Petersburg.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge University Press.
- Huang Bufan. 1995. Conditions for tonogenesis and tone split in Tibetan dialects. *Linguistics of the Tibeto-Burman area* 18(1). 43–62.
- Hyman, Larry M. & William R. Leben. 2021. Tone systems. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody*, 45–65. Oxford. <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780198832232.001.0001/oxfordhb-9780198832232-e-6>.
- Jäschke, Heinrich August. 2018. *Tibetan grammar*. London: Dalton house.
- Kushnir, Yuriy. 2018. *Prosodic Patterns in Lithuanian Morphology*. Leipzig: Universität Leipzig Ph.D. dissertation.

- Makley, Charlene, Keith Dede, Hua Kan & Wang Qingshan. 1999. The Amdo dialect of Labrang. *Linguistics of the Tibeto-Burman area* 22(1). 97–127.
- Pike, Kenneth Lee. 1948. *Tone Languages: A Technique for Determining the Number and Type of Pitch Contrasts. Language, with studies in tonemic substitution and fusion* (University of Michigan Publications. Linguistics 4). Ann Arbor.
- Tournadre, Nicolas & Sangda Dorje. 2003. *Manual of Standard Tibetan. Language and civilization*. Ithaca, NY: Snow Lion Publ.
- Vydrina, Alexandra. 2017. *A corpus-based description of Kakabe, a Western Mande language: prosody in grammar*. Paris: INALCO Ph.D. dissertation.
- Welmers, William E. 1959. Tonemics, morphotonemics, and tonal morphemes. *General Linguistics* 5. 1–9.

Workshop 12

From and Towards Demonstratives: Grammaticalization Processes and Beyond

Verónica Orqueda (Pontificia Universidad Católica de Chile)

Berta González Saavedra (Universidad Complutense de Madrid)

Demonstratives are generally seen as deictic elements, which are primarily used to point to a referent, focusing the hearer's attention on an entity (Diessel 1999). However, their nature, their inner possible classifications, and their grammaticalization processes from and towards such a category have long been topics of debate. With respect to the sources of demonstratives, there is a well-known discussion regarding whether demonstratives can or cannot develop from lexical sources. Thus, Heine et al. (2020: 421) claim that “there are at least three main lexical sources that may lead to the emergence of demonstrative categories. But these sources do not seem to exhaust the range of pathways”, contra Diessel (2006: 481), who believes that “demonstratives are so old that their roots are not etymologically analyzable”.

As for the grammaticalization processes that start with demonstratives, it has been noted that demonstratives can develop into complementizers, conjunctions, copulas, definite articles, focus, third person pronouns, relatives and subordinators, among others. As Diessel (1999) shows, the targets may vary according to the syntactic classification of the source demonstratives. As well, demonstratives are not restricted to one single path of grammaticalization. Among examples of different targets that stem from the same source, there is the case of Latin *ille*, which develops both as a definite article (*el*) and as a third person personal pronoun (*él*) in Spanish (see e.g., Giusti 2001, Roca 2009, and van Gelderen 2011), probably depending on the different contexts.

Particularly, the connection between demonstratives and personal pronouns through grammaticalization processes is still a field of fruitful discussions, and one may wonder whether demonstratives may develop as 1st, 2nd, and 3rd person pronouns: there is plenty of evidence of 3rd person pronouns derived from demonstratives, (see e.g., Heine and Kuteva 2002), while there is no evidence of 1st person pronouns, and scarce evidence of 2nd person pronouns, as is the case of *anata* in Japanese (distal demonstrative > 2nd sg. person pronoun, see Ishiyama 2012 and Ishiyama 2019).

Regarding grammaticalization processes within the category of demonstratives, there is also an ongoing debate on whether exophoric uses (this is, in speech act situations) necessarily precede or not anaphoric or discursive uses. This debate has a direct implication to the question of unidirectionality of grammaticalization (see, e.g., Stavinschi 2012).

Lastly, recent cognitive investigations on the selection and use of demonstratives can shed light of possible explanations for the development of demonstratives. Thus, for instance, Peeters et al. (2021), among others, show that the selection of specific demonstratives may be determined by the communicational situation and the perception of the speaker-addressee relationship, and not only by the proximity or distance of the object. Such synchronic observations may lead one to wonder what cognitive factors are behind the grammaticalization of demonstratives towards new functions.

The purpose of this workshop is to invite scholars working on different aspects of the grammaticalization of demonstratives and from diverse theoretical frameworks, in order to jointly elaborate

a more complete map of possible developments of demonstratives and their related aspects that have taken place or are still taking place in languages of the world. As suggestions, some topics that will be welcomed are:

- New proposals for the origin of demonstratives
- New proposals of grammaticalization processes from demonstratives
- Processes with more than one result, e.g. Lat. *ille* > Sp. *él* and *el*.
- Cognitive processes involved in the grammaticalization of demonstratives from cross-linguistic perspectives.
 - New approaches from diverse linguistic areas (sociolinguistics, pragmatics, among others) that help us understand the processes involved in the grammaticalization of demonstratives.

REFERENCES

- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. (2006). “Demonstratives, joint attention, and the emergence of grammar”. *Cognitive Linguistics* 17(4). 463–489.
- van Gelderen, E. (2011). *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.
- Giusti, G. (2001). “The birth of a functional category: From Latin ILLE to the Romance article and personal pronouns”. In G. Cinque & G. Salvi (eds.), *Current studies in Italian syntax: Essays offered to Lorenzo Renzi*, 157–171. Amsterdam: North-Holland.
- Heine, B. and T. Kuteva. (2002). *World Lexicon of Grammaticalization*, Cambridge, CUP.
- Heine, B., T. Kuteva, H. Long, H. Narrog and F. Wu. (2020). “Where do demonstratives come from?” *STUF* 73(3). 403-434.
- Ishiyama, O. (2019). *Diachrony of personal pronouns in Japanese*. John Benjamins Publishing Company.
- Ishiyama, O. (2012). “The diachronic relationship between demonstratives and first/second person pronouns”. *Journal of Historical Pragmatics* 13(1). 50–71.
- Peeters, D., E. Krahmer, and A. Maes. (2021). “A Conceptual Framework for the Study of Demonstrative Reference”. *Psychonomic Bulletin & Review* 28. 409–433.
- Roca, F. (2009). “The Left Periphery of Nominal Constructions and the Evolution of Definite Determiners in Romance”. In J. Rafel (ed.) *Diachronic Linguistics*, 497–553. Girona: Documenta Universitaria.
- Stavinschi, A. (2012). “On the development of the Romance demonstrative systems: Historical remarks and theoretical conclusions”. *Diachronica* 29(1). 72-97.

Contributions

Further Pathways Towards Demonstratives

Marianne Mithun (University of California, Santa Barbara)

Demonstratives are generally seen as deictic elements, which are primarily used to point to a referent. In a series of important works, Diessel (2006 and more) observes that though demonstratives are ubiquitous cross-linguistically, we rarely see evidence of the expected pathways of grammaticalization which underlie them. If demonstratives were indeed impervious to change, we would expect them to match across related languages, apart from regular sound changes. But such is often not the case. More recently, Heine et al (2020) have uncovered some pathways by which demonstratives have developed in certain languages from locative adverbs, positional verbs, and classifiers. They point out that while in many cases the developments involve processes of renewal, whereby original demonstratives are reinforced by additional elements, in some others demonstratives may not have been part of the source construction at all.

Demonstratives are especially pervasive in speech in languages of the Iroquoian family, indigenous to eastern North America. They are used as in other languages as pronouns, but they also occur in a wide array of other constructions, including pervasive conventionalized discourse structures. Yet they are not fully cognate across the languages.

Those in some of the languages show partial similarities which indicate development from shared communicative strategies. Basic proximal demonstrative pronouns for ‘this, this one’ include Mohawk *kí:ken*, Oneida *kaʔika*´, Onondaga *né:kə*, Cayuga *né:kyə*, Seneca *nə:kə:*, and Tuscarora *kyè:ní:kə:*. (The Mohawk digraph *en* is a nasalized vowel ɛ ; spelling has otherwise been regularized to show correspondences.) Distal demonstratives include Mohawk *thí:ken*, Oneida *thika*´, Onondaga *thó:kə*, Cayuga *thó:kyə*, Seneca *hi:kə*, and Tuscarora *hè:ní:kə:*.

All combine two of the sources described by Heine et al, locative adverbials and lexical verbs, but not via processes of renewal. Dialectal alternants in Mohawk provide a clue. In place of *kí:ken* ‘this’, some speakers use *ken*´ *i:ken*, and for *thí:ken* ‘that’, they use *tho*´ *i:ken*. Mohawk *kèn:* ‘*en* or *ken*’ is the proximal locative adverbial ‘here’, and *tho* is the distal locative adverbial ‘there’. The word *i:ken* is a verb consisting of the neuter pronominal prefix *ka-* and a verb stem *-i* which occurs only in certain fixed constructions. The sequence *a+i* fuses to the nasalized vowel. Verbs must be at least disyllabic, so the initial *i* is prothetic. The sources are thus ‘here it is’ and ‘there it is’. The languages have used different initial demonstratives in this construction.

Tuscarora has another proximal demonstrative of interest: *kyé:nə:* ‘this one’. It appears in such constructions as ‘Drink **this**’, ‘Suck on **this**’, ‘Hold **this**’, ‘Take **this** into the house’, ‘Cook **this one**’, etc., as well as ‘**This** is the tree’ and ‘**Here** is a treat for you’. Its source is a well-formed verb ‘I am holding it’: *k-yenə-*: 1SG.AGT-hold-STATIVE. This same demonstrative is the first element of the basic proximal demonstrative *kyè:ní:kə:* ‘this one’ (*kyé:nə:*´ *i:kə:*).

Comparison of demonstratives in Northern Iroquoian languages thus suggests certain pathways of development foreseen by Heine et al., in some cases from locative adverbial demonstratives

plus verbs, but not via renewal, and in others directly from verbs on their own. The developments still reflect well-known processes of grammaticalization: content extension, desemanticization, decategorialization, loss of lexical autonomy (fusion), and substance erosion.

REFERENCES

- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17.4: 463-489.
- Heine, Bernd, Tania Kuteva, Haiping Long, Heiko Narrog, and Fang WU. 2020. Where do demonstratives come from? *STUF* 73.3: 403-434.

From spatial noun to medial demonstrative: the case of Khalkha Mongolian

Benjamin Brosig (University of Bern)

Guntsetseg Dolgor (Ludwig-Maximilians-University)

In this presentation, we discuss the development of two attributive/nominalized spatial nouns into addressee-centered demonstratives in Khalkha Mongolian as a so far unattested path of grammaticalization. Common Mongolic had an opposition of a demonstrative *ene, signaling closeness, to a demonstrative *tere that signaled distance. These forms consisted of the stems *e- and *te- and not fully transparent subsequent locative elements (cf. locative adverbs *ende/*tende, adjectival similatives *eyimü/*teyimü), but the stems themselves cannot be traced back any further. This basic demonstrative system has been retained in the Central Mongolic varieties Buryat, Khorchin or Oirat. However, the Khalkha branch has four demonstratives (as already suggest, but not elaborated on, by Janhunen 2012: 131-2, Guntsetseg 2016: 37-9, Brosig et al. 2018: 76):

Table 1: Demonstratives (for attributive or argument use) of Khalkha Mongolian

| | speaker | addressee |
|----------|------------|---|
| close to | <i>en</i> | <i>naa-d(-ax)</i> (NOUN)= <i>čín</i> |
| far from | <i>ter</i> | <i>caa-d(-ax)</i> (NOUN)= <i>čín</i> |

Structurally the stems *naa-* ‘near side of’ and *caa-* ‘remote side of’ belong to the class of “spatial nouns”, a specific word class in Mongolic that can be grouped as distantly related to the class of regular nominals (substantives, adjectives, personal pronouns, numerals) in that it shares some historical and synchronous morphology. Spatial nouns lack the nominative, but inflect for idiosyncractic locative and prolative suffixes. They also allow for ablatives (formed by attaching the regular ablative suffix to the locative) and form attributives or nominative argument forms through *-d*, which can then be turned into non-nominative or plural argument forms by adding *-x* and a case suffix (cf. Janhunen 2012: 121-5).

Table 2: The paradigm of Khalkha spatial nouns in juxtaposed with a sub-part of the paradigm of substantives

| | Spatial paradigm | | Corresponding noun forms | |
|----------------------------|-----------------------|------------------------------------|-------------------------------|---|
| Locative | <i>naa-n</i> | ‘on the close side [of]’ | <i>zam-[i]d</i> | way-DAT |
| Locative Ablative | + <i>naa-n-aas</i> | ‘from the close side [of]’ | <i>zam-aas</i> | way-ABL |
| Prolative | <i>naa-[ɣ]uur</i> | ‘along the close side [of]’ | <i>[zam-aar]</i> | way-INS (with possible prolative interpretation) |
| Allative | <i>naa-š</i> | ‘towards the close side [of]’ | <i>[zam-ruu]</i> | way-ALL |
| Nominative- Attributive | <i>naa-d</i> [N/Ø] | ‘the N/one on the close side [of]’ | <i>zam,</i> <i>zam-iiŋ</i> | way (argument), way-GEN |

| | | | | |
|------------|----------------------------|-------------------------------------|------------------|---|
| Case forms | <i>naa-d-[a]x-</i> CASE | ‘the N-CASE on the close side [of]’ | <i>zam-iij-x</i> | way-GEN-NMLZ (‘the one belonging to the way) |
|------------|----------------------------|-------------------------------------|------------------|---|

Regular spatial nouns are mostly used as adverbials or postpositions, as in (1). The shift of the deictic origo to second person and with it the qualitative change to a second person-centered demonstrative has taken place in the presence of a postposed clitic such as the second person singular form =*čin* that indicates that the entity in question is ‘on the near (front) side of you’, which usually means within the perception of the addressee, as in (2), or ‘on the remote (back) side of you’, i.e. not only remote from the addressee, but also beyond her sight.

- (1) *tern-ees naa-n yuu=č med[-ex=güi.]*
 DEM.DIST-ABL this.side.of-LOC what=LIM.FOC know-FUT.PTCP=NEG
 ‘I don’t know anything beyond that [point in time] (i.e. that is closer to the present).’
- (2) *naa-d=čin kod=güi.*
 this.side.of-ATTR=2POSS code=NEG
 ‘That [which, seen from my perspective, is on the near side of you] doesn’t have a code.’

In corpus data, the new demonstratives are most well-established in argument function in free conversational data (*en*: 1032, *ter*: 793, *naad*: 226, *caad*: 26, discounting non-spatial uses), while they are basically absent in newspaper texts (e.g., for argument usage, *ene*: 4536, *naad*: 9).

Typologically, this change is distinct from developments of adverbs like ‘here’/‘there’ to demonstratives (Kuteva et al. 2019: 229-32, 430-1) since spatial nouns code relations between two entities that are not intrinsically linked to the interlocutors.

REFERENCES

Brosig, Benjamin, Foongha Yap & Kathleen Ahrens. 2019. Assertion, assumption and presupposition: an account of the erstwhile nominalizer YUM in Khalkha Mongolian. *Studies in Language* 43(4): 896–940.

Guntsetseg, Dolgor. 2016. *Differential Case Marking in Mongolian*. Wiesbaden: Harrassowitz.

Janhunen, Juha. 2012. *Mongolian*. Amsterdam: Benjamins.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

On the Development of Demonstratives into Personal Pronouns

Osamu Ishiyama (Soka University of America)

It is well known that crosslinguistically demonstratives are the major source of third person pronouns. Previous studies show that the close relationship between demonstratives and third person pronouns is due to their functional similarity and that many languages use demonstratives in place of third person pronouns (e.g. Bhat 2004; Diessel 1999; Siewierska 2004). However, questions regarding the tenuous link between demonstratives and first/second person pronouns have received little attention. For example, ‘How uncommon is it actually for demonstratives to develop into first/second person pronouns?’ and ‘What are some of the reasons for that?’ In this study, I first present the results of a crosslinguistic survey based on a representative sample of 100 languages. The present study shows that demonstratives gave rise to first/second person pronouns in a clear manner only in three languages. I then propose some reasons as to why demonstratives rarely develop into first/second person pronouns.

There are only three languages in the sample (Basque, Japanese, Malagasy) with first and/or second person pronouns that show a clear link with demonstratives. In some Eastern varieties of Basque, the second person singular *ori* is derived from the medial demonstrative *hori* (Trask 2003: 150), and in Japanese, one of the second person pronouns comes from the demonstrative adverb *anata* ‘that way (distant from both you and me)’ (Ishiyama 2012, 2019). Malagasy shows a more extensive connection between demonstratives and personal pronouns. Garvey (1964: 40-41) states that Malagasy demonstratives are composed of the demonstrative prefix *i-* and the locative stems (e.g. *iti* ‘this (very near)’, *iú* ‘that (near)’, *ítsi* ‘that (not far)’), and that this characteristic is shared by all Malagasy independent personal pronouns (with the exception of one of the first person singular forms) which consist of the same demonstrative prefix *i-* and the pronoun stems, as in *izáu/iànáu/ízi* ‘first/second/third person singular’, respectively. There are several languages in the sample that optionally use demonstratives for the speaker and addressee as contextual substitutes for first/second person pronouns. This contextual use occurs predominantly in a typologically similar languages, particularly in East and Southeast Asia.

I argue that the tenuous link between demonstratives and first/second person pronouns is due to their functional dissimilarities. The basic function of demonstratives is to (i) indicate the location of a referent in relation to the deictic center and (ii) coordinate the interlocutors’ joint attentional focus (Diessel 2003, 2006). (i) may lead to the use of demonstratives for the speaker/addressee, but within the scope of the original demonstrative function (i.e. spatial semantics), thus providing little need for demonstratives to become first/second person pronouns (cf. Ishiyama 2012, 2019). For (ii), the referent of first/second person pronouns is generally presupposed and readily accessible to all relevant parties. That is, the joint attentional focus is in most instances taken for granted for first/second person pronouns. The nature of deictic force involving demonstratives on the one hand and first/second person pronouns on the other is also quite distinct. First/second person pronouns assume less stability of referents than demonstratives in the speaker-addressee interaction, that is, the referent of first/second person pronouns is ‘more shifting’. For demonstratives to become first/second person pronouns, it is necessary to lose the two basic functions and acquire qualitatively different one: i.e. losing the function to achieve joint attention and gaining the ability to be used repeatedly for presupposed referents regardless of the spatial relationship that holds between a referent and the deictic center.

REFERENCES

- Bhat, D.N.S. 2004. *Pronouns*. Oxford: Oxford University Press.
- Cooke, Joseph Robinson. 1968. *Pronominal reference in Thai, Burmese, and Vietnamese*. Berkeley, CA: University of California Press.
- Diessel, Holger. 1999. *Demonstratives: Form, function, and grammaticalization*. Amsterdam: John Benjamins.

- Diessel, Holger. 2003. The relationship between demonstratives and interrogatives. *Studies in Language* 27: 635-655.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4): 463-489.
- Garvey, Catherine J. 1964. *A Sketch of Malagasy grammar*. Washington, D.C.: United States Department of Health, Education and Welfare.
- Ishiyama, Osamu. 2012. The diachronic relationship between demonstratives and first/second person pronouns. *Journal of Historical Pragmatics* 13(1). 50–71.
- Ishiyama, Osamu. 2019. *Diachrony of personal pronouns in Japanese: A functional and cross-linguistic perspective*. Amsterdam: John Benjamins.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Trask, R.L. 2003. The noun phrase: Nouns, determiners and modifiers; pronouns and names. In Hualde, José Ignacio & Jon Ortiz de Urbina (eds.), *A Grammar of Basque*, 113-170. Berlin: Mouton.

LANGUAGES

Abkhaz (Northwest Caucasian, Northwest Caucasian; Georgia), Acoma (Keresan, Keresan; United States), Alambak (Sepik Hill, Sepik; Papua New Guinea), Amele (Madang, Trans-New Guinea; Papua New Guinea), Apurinã (Purus, Arakawan; Brazil), Arabic (Egyptian) (Semitic, Afro-Asiatic; Egypt), Arapesh (Mountain) (Kombio-Arapesh, Torricelli; Papua New Guinea), Asmat (Asmat-Kamoro, Trans-New Guinea; Indonesia), Bagirmi (Bongo-Bagirmi, Central Sudanic; Chad), Barasano (Tucanoan, Tucanoan; Colombia), Basque (Basque, Basque; France/Spain), Bengali (Indic, Indo-European; Bangladesh, India), Berber (Middle Atlas) (Berber, Afro-Asiatic; Morocco), Bunuba (Bunuban, Bunuban; Australia), Burmese (Burmese-Lolo, Sino-Tibetan; Myanmar), Burushaski (Burushaski, Burushaski; Pakistan), Canela-Krahô (Ge-Kaingang, Macro-Ge; Brazil), Chamorro (Chamorro, Austronesian; Guam), Chinantec (Plantla) (Chinantecan, Oto-Manguean; Mexico), Chukchi (Northern Chukotka-Kamchatkan, Chukotka-Kamchatkan; Russia), Cree (Plains) (Algonquian, Algic; Canada), Daga (Dagan, Dagan; Papua New Guinea), Dani (Lower Grand Valley) (Dani, Trans-New Guinea; Indonesia), Drehu (Oceanic, Austronesian; New Caledonia), Dyirbal (Northern Pama-Nyungan, Pama-Nyungan; Australia), English (Germanic, Indo-European; United Kingdom), Fijian (Oceanic, Austronesian; Fiji), Finnish (Finnic, Uralic; Finland), French (Romance, Indo-European; France), Georgian (Kartvelian, Kartvelian; Georgia), German (Germanic, Indo-European; Germany), Gooniyandi (Bunuban, Bunuban; Australia), Grebo (Kru, Niger-Congo; Liberia), Greek (Greek, Indo-European; Greece), Greenlandic (West) (Eskimo, Eskimo-Aleut; Greenland), Guaraní (Tupi-Guarani, Tupian; Paraguay), Hausa (West Chadic, Afro-Asiatic; Niger, Nigeria), Hebrew (Modern) (Semitic, Afro-Asiatic; Israel), Hindi (Indic, Indo-European; India), Hixkaryana (Cariban, Cariban; Brazil), Hmong Njua (Hmong-Mien, Hmong-Mien; China), Imonda (Border, Border; Papua New Guinea), Indonesian (Malayo-Sumbawan, Austronesian; Indonesia), Ingush (Nakh, Nakh-Daghestanian; Russia), Jakaltek (Mayan, Mayan; Guatemala), Japanese (Japanese, Japanese; Japan), Kannada (Southern Dravidian, Dravidian; India), Kayah Li (Eastern) (Karen, Sino-Tibetan; Myanmar, Thailand), Kayardild (Tangkic, Tangkic; Australia), Kewa (Engan, Trans-New Guinea; Papua New Guinea), Khalkha (Mongolic, Altaic; Mongolia), Khoekhoe (Khoek-Kwadi, Khoek-Kwadi; Namibia), Kiowa (Kiowa-Tanoan, Kiowa-Tanoan; United States), Koasati (Muskogean, Muskogean; United States), Kobon (Madang, Trans-New Guinea; Papua New Guinea), Korean (Korean, Korean; Korea), Koyra Chiini (Songhay, Songhay; Mali), Kutenai (Kutenai, Kutenai; Canada, United States), Kyuquot (Southern Wakashan, Washakan; Canada), Lakhota (Core Siouan, Siouan; United States), Lango (Nilotic, Eastern Sudanic; Uganda), Lavukaleve (Lavukaleve, Solomons East Papuan; Solomon Islands), Lezgian (Lezgitic, Nakh-Daghestanian; Azerbaijan, Russia), Luvale (Bantoid, Niger-Congo; Angola), Madurese (Malayo-Sumbawan, Austronesian; Indonesia), Malagasy (Barito, Austronesian; Madagascar), Mandarin (Chinese, Sino-Tibetan; China), Mangarrayi (Mangarrayi, Mangarrayi-Maran; Australia), Mapuche (Araucanian, Araucanian; Chile), Marathi (Indic, Indo-European; India), Maricopa (Yuman, Hokan; United States), Martuthunira (Western Pama-Nyungan, Pama-Nyungan; Australia), Maung (Iwaidjan, Iwaidjan; Australia), Maybrat (North-

Central Bird's Head, West Papuan; Indonesia), Meithei (Kuki-Chin, Sino-Tibetan; India), Mixtec (Chalcatongo) (Mixtecan, Oto-Manguean; Mexico), Ngiti (Lendu, Central Sudanic; DR of the Congo), Ngiyambaa (Southeastern Pama-Nyungan, Pama-Nyungan; Australia), Nkore-Kiga (Bantoid, Niger-Congo; Uganda), Nunggubuyu (Nunggubuyu, Guwinyguan; Australia), Oneida (Northern Iroquoian, Iroquoian; United States), Oromo (Harar) (Lowland East Cushitic, Afro-Asiatic; Ethiopia), Persian (Iranian, Indo-European; Iran), Pirahã (Mura, Mura; Brazil), Pitjantjatjara (West Pama-Nyungan, Pama-Nyungan; Australia), Quechua (Imbabura) (Quechuan, Quechuan; Ecuador), Rapanui (Oceanic, Austronesian; Chile), Russian (Slavic, Indo-European; Russia), Samoan (Oceanic, Austronesian; Samoa), Sango (Ubangi, Niger-Congo; Central African Republic), Sanuma (Yanomam, Yanomam; Brazil, Venezuela), Semelai (Aslian, Austro-Asiatic; Malaysia), Slave (Athapaskan, Na-Dene; Canada), Spanish (Romance, Indo-European; Spain), Supyire (Gur, Niger-Congo; Mali), Swahili (Bantoid, Niger-Congo; Tanzania), Tagalog (Greater Central Philippine, Austronesian; Philippines), Thai (Kam-Tai, Tai-Kadai; Thailand), Tibetan (Bodic, Sino-Tibetan; China), Tiwi (Tiwian, Tiwian; Australia), Tukang Besi (Celebic, Austronesian; Indonesia), Turkish (Turkic, Altaic; Turkey), Una (Mek, Trans-New Guinea; Indonesia), Vietnamese (Viet-Muong, Austro-Asiatic; Vietnam), Warao (Warao, Warao; Venezuela), Wari' (Chapacura-Wanham, Chapacura-Wanham; Brazil), Wichí (Matacoan, Matacoan; Bolivia, Argentina), Wichita (Caddoan, Caddoan; United States), Yagua (Peba-Yaguan, Peba-Yaguan; Peru), Yaqui (Cahita, Uto-Aztecan; Mexico), Yoruba (Defoid, Niger-Congo; Benin, Nigeria), Zulu (Bantoid, Niger-Congo; South Africa)

Types of contexts inducing the grammaticalization of demonstratives into definite articles – the case of a language without articles

Branimir Stanković (University of Niš)

We hypothesize that there are certain **types of contexts** that are mostly responsible for initiating the grammaticalization process(es) of demonstratives from spatial, deictic elements into discourse-relevant anaphorics thru context-induced reinterpretation (Heine, Claudi & Hünemeyer 1991), in which the use of these items is necessary for obtaining the intended definite interpretation, and not simply for reasons of disambiguating between the available indefinite and definite interpretation of bare NPs. This idea is based on the situation in Serbo-Croatian, a language lacking the categories of definite and indefinite articles, but in which the use of demonstratives is mandatory in the following types of contexts.

i. **cardinal numbers and partitivity**. Discourse-old cardinal number phrases (1) and partitive phrases (2) must be marked for definiteness, as the bare phrases unambiguously yield the indefinite interpretation. This is achieved by the use of demonstratives:

(1) Belić je napisao [dva rada o dijalektima južne
Belić AUX write-PAST.SG.MASC two papers on dialects-INST.PL Southern-GEN.SG
Srbije]_i. On u {[dva rada]_{*i,j}}/{[ta dva rada]_{i,*j}} objašnjava da...
Serbia-GEN.SG he in two papers that-PAUCAL.MASC two papers explains that
“Belić wrote two papers on the dialects of South Serbia. In {some two papers / those two papers} he explains that...”

(2) Belić piše o [delu reči]_i. {[Deo reči]_{*i,j} / [Taj deo reči]_{i,*j}} je...
Belić writes about part-LOC.SG word-GEN.PL part word-GEN.PL that part word-GEN.PL is
“Belić writes about a part of the words. {A part of the words / That part of the words is...}”

ii. **discourse-old indefinite specific pronominal referents**. The discourse status of previously introduced indefinite specific pronominal referents must be signaled with the use of demonstratives; otherwise, the indefinite pronouns remain unambiguously indefinite:

(3) Neko_i je napisao rad. {[Neko]_{*i,k} / [Taj neko]_{i,*k}} je Belić.
someone AUX write-PAST.SG.MASC paper someone that someone is Belić
“Someone wrote a paper. {Someone / That someone} is Belić.”

iii. **temporal constructions**. A series of temporal genitive constructions consist of a mandatory “determiner” and a noun denoting time period sequence (considering Meillet’s (1912) broad notion of grammaticalization, which includes the evolution of grammatical constructions):

(4) ove godine / tog jutra /
this-GEN.SG.FEM year-GEN.SG.FEM that-GEN.SG.NEUT morning-GEN.SG.NEUT
onog dana
that-GEN.SG.MASC day-GEN.SG.MASC
“this year / that morning / that day”

Although unidirectional in its nature (Greenberg 1978; Lyons 1977; Heine, Claudi & Hünemeyer 1991; Hawkins 1994; Diessel 1999), the proposed hypothesis does not negate the possibility of the reverse grammaticalization pattern, as shown by Frajzyngier (1996) for Chadic and Stavinschi (2012) for Romance languages. As a matter of fact, the presented Serbo-Croatian demonstratives are a result of a diachronic integration of the initial deictic items *ovъ*, *tъ* and *onъ* with the anaphoric pronoun *i/jъ* (*ovъ* / *tъ* / *onъ* + *jъ* > *ovъjъ* / *tъjъ* / *onъjъ* > *ovaj* / *taj* / *onaj*). Eventually, the anaphoric item *i/jъ* entirely disappeared from the

language, leaving its traces throughout the pronominal system and in the category of definite adjectival aspect.

REFERENCES

Diessel H. 1999. *Demonstratives*; **Frajzyngier** Z. 1996. “On Sources of Demonstratives and Anaphors”; **Greenberg** J.H. 1978. “How does a language acquire gender markers?”; **Hawkins**, J.A. 1994. *A performance theory of order and constituency*; **Heine**, B. U. **Claudi** & F. **Hünemeyer** 1991. *Grammaticalization: a conceptual framework*; **Lyons** J. 1977. *Semantics* 2; **Meillet**, A. 1912. “L’evolution des formes grammaticales”; **Stavinschi**, A. 2012. “On the development of the Romance demonstrative systems”.

Demonstratives taking over discourse: the grammaticalisation of deictic clitics in Äiwoo

Åshild Næss (University of Oslo)

The Oceanic language Äiwoo (Solomon Islands) has deictic particles *le* (PROX)/*lä* (DIST) and enclitics =*Ce* (PROX)/=*Câ* (DIST) which are extremely frequent in discourse, and which frequently occur together, ‘bracketing’ a word or constituent:

- (1) a. **Lâ** kâlikâli nugo=**ngâ** i-po-pâko=to.
 DIST sweet.potato POSS:FOOD.1MIN=DIST PFV-cook-good=now
 ‘My sweet potatoes are cooked.’
- b. **Lâ** maniok=**kâ lâ** ki-to=**kâ**.
 DIST manioc=DIST DIST IPFV-be=DIST
 ‘That’s manioc there.’

The distribution and function of this construction is complex and not easily summarised. It can occur with almost any type of constituent, and with multiple constituents within the same clause, as in (1b). It has functions related to focus and information structure, but the bracketed constituent is not necessarily the focused constituent; it can equally well be the presupposed part of the clause. It is often found in constructions involving a notionally subordinate clause, but the bracketed constituent can be either the subordinate or the main clause. It can be used to indicate that a clause has a topic-comment structure, as opposed to forming part of the presupposition, as in the pair *nelo lâ lägä=kâ* (sea DIST dry=DIST) ‘the tide (topic) was low (comment)’ vs. *lä nelo lägä=kâ* (DIST sea dry=DIST) ‘at low tide’ (background information for a further assertion). In short, the construction just seems to indicate that a particular sequence **forms an interpretationally relevant unit with respect to the surrounding discourse** – it provides a cue to the overall structure of the utterance rather than indicating a specific function (Næss 2021).

There are no historical records that would provide evidence of how this unusual situation has arisen. However, what we know about the grammaticalisation of demonstrative forms in other Oceanic languages may provide clues to the pathways that have led to the Äiwoo construction: among other things, Oceanic languages use demonstratives as phrase demarcation devices, as markers of topic, and to mark notionally subordinate clauses (Moyse-Faurie 1997, Næss and Hovdhaugen 2011, Bril 2010, François ms.). The process known as insubordination, where formally subordinate clauses take on independent uses, might help account for the range of environments in which the construction is used; Evans (2007) notes that typical functions of insubordinated clauses include focus constructions and discourse contrasts, and that «in a number of languages, insubordinated clauses have what at first sight seem to be a bewilderingly wide range of functions» (Evans 2007: 423). Mithun (2008) moreover notes that markers of syntactic dependency can be extended to discourse level, with the function of indicating a relationship to the larger context, which is precisely what the Äiwoo deictics seem to do. I propose that demonstrative forms are particularly suited to taking on such a function, as the core function of demonstratives is to «coordinate the interlocutors’ joint attentional focus» (Diessel 2006, cf. also Evans et al. 2018), i.e. to make sure that the hearer is attending to the same object or concept as the speaker. A construction the function of which is to guide the hearer towards correctly identifying the syntactic and information-structural makeup of an utterance would seem to be a natural extension of this attention-coordination function.

REFERENCES

- Bril, Isabelle. 2010. Informational and referential hierarchy: clause-linking strategies in Austronesian-Oceanic languages. In Bril (ed.), *Clause linking and clause hierarchy: syntax and pragmatics*, 269-311. Amsterdam: John Benjamins.
- Diessel, Holger. 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17:4, 463-489.
- Evans, Nicholas. 2007. Insubordination and its uses. In Nikolaeva (ed.), *Finiteness: theoretical and empirical foundations*, 366-431. Oxford: OUP.
- Evans, Nicholas, Henrik Bergqvist, and Lila San Roque. 2018. The grammar of engagement I: framework and initial exemplification. *Language and Cognition* 10, 110-140.
- François, Alexandre, ms. From deictics to clause linkers: discourse deixis and clause backgrounding strategies in the languages of north Vanuatu.
- Mithun, Marianne. 2008. The extension of dependency beyond the sentence. *Language* 84:1, 69-119.
- Moyse-Faurie, Claire. 1997. *Grammaire du futunien*. Nouméa: CDP Nouvelle-Calédonie.
- Næss, Åshild. 2021. Deictics, syntax and information structure in Äiwoo. Paper presented at the 15th International Conference on Austronesian Linguistics.
- Næss, Åshild, and Even Hovdhaugen. 2011. *A grammar of Vaeakau-Taumako*. Berlin: De Gruyter Mouton.

Origin and development of the Albanian demonstratives

Sergio Neri (University of Basel)

Michiel de Vaan (University of Basel)

Albanian has a binary demonstrative pronoun system, with proximal *ky* m., *kjo* f. 'this' contrasting with distal *ai* m., *ajo* f. 'that'. The distal pronouns double as personal pronouns of the third person. In the Old Albanian period (16th to 18th century), the system was basically the same, with the addition of the then still productive neuter gender.

As opposed to the relative ease with which the synchronic morphology can be described, the grammaticalization path leading up to both pronouns is not yet completely understood. Both demonstratives arose from the composition of two deictic elements. The second of which (nominative masc. *-i/-y*, f. *-jo*) is the same for both pronouns and probably continues the PIE demonstrative **so*, **to*-, although the nominative singular forms have not been fully explained yet. The origin of the first elements *k-* resp. *a-* is disputed. In proximal *k-*, scholars have recognized PIE deictic **k-*, PIE interrogative **kw-*, or Romance *(ek)ku-*; distal *a-* has been argued to contain, for instance, PIE **so-u-* 'that one' or PIE **h₂eu-* 'yon'.

In our talk we will first sketch the morphology and the syntactic behaviour of the two demonstrative pronouns in Old Albanian, and then proceed to a re-evaluation of the etymological scenario's that may explain the rise and grammaticalization of *ky* and *ai*.

REFERENCES

- Demiraj, Shaban. 1993. *Historische Grammatik der albanischen Sprache*. Vienna.
 Matzinger, Joachim & Stefan Schumacher. 2018. The morphology of Albanian. In: *Handbook of Comparative and Historical Indo-European Linguistics*, vol. 3, 1749–1771.
 Topalli, Kolec. 2011. *Gramatikë historike e gjuhës shqipe*. Tirana.

From demonstratives to articles in the Celtic languages

Eugenio Luján (Universidad Complutense de Madrid)

Esteban Ngomo Fernández (Universidad Complutense de Madrid)

It is generally agreed on that the definite articles in medieval Celtic languages (Old Irish *in(t)*, Old Welsh *ir*, Old Breton *an*) must have originated from earlier demonstratives in **sindo-*, *-ā* (GOI §467). It is clear, in fact, that they are etymologically related to certain demonstratives in ancient Celtic languages, such as Gaulish *sinde* and *(s)indas* (Lambert 1994: 66). They would thus have followed a most frequent, well-known grammaticalization path DEMONSTRATIVE > DEFINITE ARTICLE (see, e.g., Heine – Kuteva 2004: 109-111).

Old Irish definite articles, however, display an interesting peculiarity – they can co-occur with indefinite referents (GOI §470). This has been variously explained (GOI §470, Ronan 2004) and Goldstein (2022) has recently proposed that they accompany referents that are the focal center of the discourse and also noun phrases that are a signal to the addressee to retrieve mental representation of the referent, which would be in line with Dryer’s (2014) reference hierarchy. In his analyses Goldstein has also applied Löbner’s (1985:298–299) distinction between pragmatically and semantically definite referents to identify the different types of definiteness that can trigger the use of the definite articles in Old Irish.

Our goal in this paper is to try to shed light on the grammaticalization processes that have led to the development of the definite articles of the Celtic languages. Although still quite limited in number, there are now more extant texts in continental Celtic languages, and they provide very interesting linguistic information. We have, therefore, collected and surveyed all the occurrences of demonstratives in the those languages: Celtiberian *so-* and *sto-* (Wodtke 2000: 338-334, Jordán 2019: 230-233, De Hoz in press), the various Gaulish forms (Lambert 1994: 66, Delamarre 2003), and maybe infixed Lepontic *-so-* (Lambert 1994: 66, *LexLep*, s.u. *tošokote*). We have analyzed all those instances in their context in relation to Dryer’s hierarchy and following Löbner’s frame and they appear to display an array of different uses. This allows for a comparison with the distribution of the definite article in Old Irish investigated by Goldstein (2022) and provides additional evidence for refining our understanding of the processes involved in the grammaticalization of Celtic definite articles.

REFERENCES

- De Hoz, Javier, in press: *Historia lingüística de la Península Ibérica en la Antigüedad*, Madrid, C.S.I.C.
- Delamarre, Xavier 2003: *Dictionnaire de la langue gauloise*, Paris, Errance.
- Dryer, Matthew 2014: “Competing methods for uncovering linguistic diversity. The case of definite and indefinite articles (Commentary on Davis, Gillon, and Matthewson)”, *Language* 90.4: 232–249.
- GOI = Thurneysen, Rudolph 1946: *A Grammar of Old Irish*, Dublin, Dublin Institute for Advanced Studies.
- Goldstein, David 2022: “The Old Irish article”, *Journal of Celtic Linguistics* 23.
- Heine, Bernd – Tania Kuteva 2004: *World Lexicon of Grammaticalization*, Cambridge, Cambridge University Press.
- Jordán, Carlos 2019: *Lengua y epigrafía celtibéricas*, Zaragoza, Prensas de la Universidad de Zaragoza.
- Lambert, Pierre-Yves 1994: *La langue gauloise*, Paris, Errance.
- LexLep* = David Stifter et alii (eds.), *Lexicon Leponticum* (<https://lexlep.univie.ac.at/>).
- Ronan, Patricia (2004). “Co n-accae in fer and Functional Grammar”, *Journal of Celtic Linguistics* 8, 133–147.

Wodtko, Dagmar 2000: *Wörterbuch der keltiberischen Inschriften* (vol. V.1 of *Monumenta Linguarum Hispanicarum*, ed. by Jürgen Untermann), Wiesbaden, Dr. Ludwig Reichert Verlag.

Latin *ecce*: arguments in favor of its development from a PIE demonstrative

Verónica Orqueda (Pontificia Universidad Católica de Chile)

Roland Pooth (Independent researcher)

The Latin particle *ecce* has been extensively studied from different perspectives and, yet, it is not easy to reach a consensus as regards its etymology, due to the obscurity of the morphological formation, the scarcity of cognates in related languages, and the diverse uses found already in Archaic Latin (and its continuing stages). As alternative etymologies, it has been claimed that *ecce* derives from the combination of demonstrative roots (e.g. Dunkel 2014: 2011, de Vaan 2008: 185, and Fruyt 2011: 750). Also, it has also been claimed that it may derive from an imperative verb form (Julia 2020). Among the different proposals, the one that considers the demonstrative origin is probably the most widely accepted. However, attempts to connect this morphological formation with an explanation at other linguistic levels: syntax, discourse-pragmatic (e.g. why preceding accusatives? Or with which meaning exactly?) have not been convincing. The aim of this talk is to argue in favor of a derivation from PIE **h₁éd=k'e*, by offering a proposal for its original meaning and for its derivation towards Latin *ecce*, considering the data found in Archaic Latin.

To reach our goal, we classify the total number of cases with *ecce* in Plautus (as representative of Archaic Latin) according to the three possible syntactic contexts in which it appears: with no syntactic integrity to its context (type a), preceding a pronoun (type b1) or a noun phrase (type b2), and preceding a sentence (c). Results show that the most frequent use is 1st person pronoun, this is type (b1), where all referents are evidently animate. Following Diessel's (1999) analysis of demonstratives as elements that call joint attention, the possibility of considering type (b1) as most ancient let us claim that the etymology **h₁éd=k'e* can be explained as the combination of an exophoric demonstrative with ablative marking and the clitic of a here-deictic exophoric demonstrative. This univerbation may have had a meaning close to 'from that', which would explain the original combination with accusatives (*ecce me* 'from that towards me'). Such an original form-meaning construct would be an appropriate starting point for a later grammaticalization process towards an interjection or a discourse marker, as it is better interpreted when preceding sentences and in cases of no integrity with its syntactic context.

REFERENCES

- De Vaan, M. 2008. *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.
- Diessel, H. 1999. *Demonstratives. Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Dunkel, G. 2014. *Lexikon der indogermanischen Partikeln und Pronominal-stämme*. Heidelberg: Universitätsverlag Winter.
- Fruyt, M. 2011. Grammaticalization in Latin. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New perspectives on historical Latin syntax*. Volume 4: Complex sentences, grammaticalization, typology. Berlin & New York: Mouton de Gruyter.
- Julia, M.A. 2020 About Latin *Ecce* 'Behold! Lo! See! There!' and Some Ancient Languages Presentatives. *International Journal of Language and Linguistics*. Vol. 8, No. 1, pp. 17-23.

New methods for old languages: the comparability of data

W13, ICHL26, Heidelberg

Alessia Cassarà (Universität Stuttgart); Lena Kaltenbach (Universität Mannheim);
Mariapaola Piccione (Universität Stuttgart); Tara Struik (Universität Mannheim)

Corresponding convenor: tara.struik@uni-mannheim.de

While historical linguistics is traditionally known to suffer from a “bad data problem” (Labov 1994: 11), the field has seen a surge in the development of (annotated) data collections and computational tools to trace quantitative changes throughout the history of languages, allowing researchers to get more out of the (often sparse) data than ever before. This availability of data opens up many new avenues for research, in particular in explaining the cognitive mechanisms behind language change. In this workshop we want to bring together researchers working in different disciplines to discuss novel empirical methods that allow us to investigate the relation between the structural changes we observe in historical texts and the factors which arguably led to these changes. We aim to do this by focusing on a) how historical corpus data can be related to **models of language learning**, b) **contemporary psycholinguistic models** and c) how we can deal with the **heterogeneity of historical data** in relation to these models.

Historical linguists have discussed the link between historical change and changes in the input and have proposed models that make use of psycholinguistic explanations, especially in terms of language acquisition (e.g. Lightfoot 1999, 2017). However, a challenge for the study of the role of language acquisition in language change is that there is no direct access to the input for past stages of languages. Approximating the input by using corpora of child-directed speech (e.g. CHILDES) for contemporary languages has resulted in the development of learning models, which may also be informative for the historical stages. For instance, Yang’s (2016) Tolerance Principle has been shown to work effectively with small amounts of data, making it very attractive for historical work (Kodner 2020, 2022; Drescher and Lahiri 2022, Ringe and Yang 2022, Trips and Rainsford 2022). However, the application of such models on corpus data requires careful consideration of how the data obtained from corpora can be compared to the input a child received (cf. Trips and Rainsford 2022 for discussion). One potential solution is to compare the frequencies of the most common verbs in a corpus to the most common verbs in a sample of child-directed speech, as Kodner (2019) demonstrates that there is a substantial overlap.

From both a psycholinguistic and historical linguistic perspective the relationship between language change and mechanisms of language processing has only rarely been explicitly addressed (for exceptions, see Jäger & Rosenbach 2008; de Smet & de Velde 2017; see also the contributions in Hundt et al. 2017 and the ongoing work by the the DFG Research Unit SILPAC (FOR 5157)). Notably, some authors have recently pointed to the importance of cross-linguistic and within-language structural priming and syntactic adaptation for studies

of (contact-induced) language change (e.g. Pickering & Garrod 2017; Kaan & Chun 2018; Kootstra & Şahin 2018; Kootstra & Muysken 2019). Effects of priming may be observable in historical corpora in the form of persistence of linguistic forms (see Ecay and Tamminga 2017; also Gries 2005; Szmrecsanyi 2006). From a Uniformitarian perspective (see Bergs 2012, Walkden 2019 for discussion), it follows that psycholinguistic processes active in language change should not differ fundamentally across languages or language stages. Methodologically, changes observed in diachrony could in principle also be elicited in psycholinguistic experiments and the results and methods of psycholinguistic experiments could inform historical corpus analyses.

Applying psycholinguistic methods and learning models to historical data also requires us to think critically about the nature of our data and how informative they are about the actual linguistic environment in which language acquisition and change takes place. Historical corpora may be heterogeneous in nature, consisting of many different genres (e.g. legal prose, narrative verse, etc.), which may not all be equally representative of a language user's input. Some types of text, e.g. theatrical texts, conversation manuals, direct speech in verse narratives, etc. have been argued to be particularly close to spoken language in the past (Ernst 1980, Ayres-Bennett 2000); also, it has been shown that language change does not proceed at the same rate in all text genres (Whitt 2018). However, it is not clear whether a restrictive approach to selecting corpus texts is preferable to one which instead draws on as much data as possible, using statistical techniques to evaluate the effect of genre. A further open question is the extent to which the writers of historical texts are themselves influenced by mechanisms such as priming, whether it is self-priming within a single text, between the two writers in private correspondence or even between two languages in translations. Similarly, it is not always clear what the impact of the linguistic background of individual authors is on the output – are they, for instance, monolinguals, early bilinguals, or possibly late bilinguals writing in their first language or late bilinguals writing in their second language?

In this workshop, we aim to compare different types of historical corpus data not only with each other, but also with the input to language acquirers and with data elicited in psycholinguistic experiments in order to develop novel methodologies bringing the fields of historical linguistics, psycholinguistics and language acquisition closer together. We invite contributions which answer or relate to the following research questions and topics:

- How can models of learnability be applied to historical data?
- What are the psycholinguistic processes behind historical language change?
- Which insights does historical linguistics provide for the study of these psycholinguistic processes?
- Which methods and resources are the best to use if we want to relate historical data to language learner input and which are best for researching the relationship between experimental data and historical data?
- Which additional data types/methodologies can contribute to bridging the gap between the disciplines of historical linguistics, acquisition studies and psycholinguistics, e.g. artificial language studies, longitudinal studies, computational models of language change, etc.?

- How can insights from historical sociolinguistics and philology contribute to a better understanding of the heterogeneity of historical corpus data and the linguistic background of individual authors?
- To what extent are the writers of historical texts themselves influenced by mechanisms of language processing, such as intra- and interindividual priming in monolingual and bilingual situations? How can we use notions such as persistence in historical corpora to tap into the cognitive processes behind the text production of medieval authors?

References

- Ayres-Bennett, W. 2000. Voices from the past: Sources of seventeenth-century spoken French. *Romanische Forschungen*, 112, 323–348.
- Bergs, A. 2012. The Uniformitarian Principle and the risk of anachronisms in language and social history. In J. M. Hernández-Campoy & J. C. Conde-Silvestre (Eds.), *The handbook of historical sociolinguistics* (1st ed., pp. 83–101). <https://doi.org/10.1002/9781118257227.ch5>
- De Smet, H., & Van de Velde, F. 2017. Experimenting on the past: A case study on changing analysability in English ly-adverbs. *English Language & Linguistics*, 21(2), 317–340. <https://doi.org/10.1017/S1360674317000168>
- Dresher, B. E., & Lahiri, A. 2022. The foot in the history of English: Challenges to metrical coherence. In B. Los, C. Cowie, P. Honeybone, & G. Trousdale (Eds.), *English historical linguistics: Change in structure and meaning: Papers from the XXth ICEHL* (pp. 42–59). John Benjamins Publishing Company.
- Ecay, A., & Tamminga, M. 2017. Persistence as a diagnostic of grammatical status: The case of Middle English negation. In E. Mathieu & Truswell, B. (Eds.), *Micro-change and Macro-change in diachronic syntax* (pp. 202–215). Oxford University Press. <https://doi.org/10.1093/oso/9780198747840.003.0013>
- Ernst, G. 1980. Prolegomena zu einer Geschichte des gesprochenen Französisch. In H. Stimm (Ed.), *Zur Geschichte des gesprochenen Französisch und zur Sprachlenkung im Gegenwartsfranzösischen* (pp. 1–14). Steiner.
- Gries, S. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.
- Hundt, M., Mollin, S., & Pfenninger, S. (Eds.). 2017. *The changing English language: Psycholinguistic perspectives*. Cambridge University Press.
- Jäger, G., & Rosenbach, A. 2008. Priming as a testing ground for historical linguists? - A reply to Chang, Eckardt, and Traugott. *Theoretical Linguistics*, 34(2), 143–156.
- Kaan, E., & Chun, E. 2018. Priming and adaptation in native speakers and second-language learners. *Bilingualism: Language and Cognition*, 21(2), 228–242. <https://doi.org/10.1017/S1366728916001231>
- Kodner, J. 2019. Estimating child linguistic experience from historical corpora. *Glossa. A Journal of General Linguistics*, 4(1), 122.
- Kodner, J. 2020. *Language acquisition in the past* [Doctoral dissertation]. University of Pennsylvania.
- Kodner, J. (2022). What learning Latin verbal morphology tells us about morphological theory. *Natural Language and Linguistic Theory*. <https://doi.org/10.1007/s11049-022-09553-2>

- Kootstra, G. J., & Muysken, P. 2019. Structural priming, levels of awareness, and agency in contact-induced language change. *Languages*, 4(3), 65. <https://doi.org/10.3390/languages4030065>
- Kootstra, G. J., & Şahin, H. 2018. Crosslinguistic structural priming as a mechanism of contact-induced language change: Evidence from Papiamentu-Dutch bilinguals in Aruba and the Netherlands. *Language*, 94(4), 902–930.
- Labov, W. 1994. *Principles of linguistic change, Volume 1: Internal factors*. Blackwell.
- Lightfoot, D. 1999. *The development of language: Acquisition, change, and evolution*. Blackwell.
- Lightfoot, D. 2017. Acquisition and learnability. In A. Ledgeway & I. Roberts (Eds.), *The Cambridge handbook of historical syntax* (pp. 381–400). Cambridge University Press.
- Pickering, M., & Garrod, S. 2017. Priming and language change. In M. Hundt, S. Mollin, & S. Pfenninger (Eds.), *The changing English language: Psycholinguistic perspectives* (pp. 173–190). Cambridge University Press.
- Ringe, D., & Yang, C. 2022. The threshold of productivity and the ‘irregularization’ of verbs in Early Modern English. In B. Los, C. Cowie, P. Honeybone, & G. Trousdale (Eds.), *English historical linguistics: Change in structure and meaning: Papers from the XXth ICEHL* (pp. 91–112). John Benjamins Publishing Company.
- Szmrecsanyi, B. 2006. *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Mouton de Gruyter.
- Trips, C., & Rainsford, T. 2022. Tolerating subject-Experiencers? Yang’s Tolerance Principle applied to psych verbs under contact in Middle English. *Journal of Historical Syntax*, 6(12), 1–43.
- Walkden, G. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics*, 4(1), 1. <https://doi.org/10.5334/gjgl.888>
- Whitt, R. J. (Ed.). 2018. *Diachronic corpora, genre, and language change*. John Benjamins Publishing Company.
- Yang, C. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press.

Dative Experiencer Psych Verbs in (Old) French

Lisa Figura, lisa.figura@uni-hamburg.de
Universität Hamburg

Psych Verbs (PV) are verbs that express mental or emotional concepts and have an argument bearing the theta role Experiencer. PV represent a phenomenon that has been discussed for decades (cf. Hirsch 2018 for an overview) as they are a heterogeneous verb class with an unstable argument structure and different syntactic constructions. There are large differences within the (Romance) languages, synchronically as well as diachronically. According to Belletti/Rizzi (1988), the Dative Experiencer Psych Verbs (Dat.Exp.PV) form one of three classes of PV. This class is characterised by the fact that the dative objects can occupy a preverbal position. This non-canonical use is possible in Latin and Romance Languages, e.g. in Spanish, as in (1):

- (1) **A Ana** le **gusta** el chocolate.
 DAT.EXP CL.DAT like.3SG the chocolate
 “Anna likes chocolate.”

In Old French, the preverbal Dat.Exp can also be found (cf. Mathieu 2006: 2), as in (2):

- (2) Et bien set qu’ **a sa mere p_{le}st** que [...]
 And well know.3SG that DAT.EXP like.3SG
 “And she knows well that it is her mother's will that [...]”

In Modern French, this structure is ungrammatical: the Dat.Exp can occupy the preverbal position only by topicalisation or dislocation with doubling (cf. Fischer 2019), as in (3):

- (3) **A Marie,** la musique classique lui **plaît.**
 DAT.EXP the music classic PRN.DAT like.3SG
 “Marie likes classic music.”

In Old French, however, the non-canonical use of Dat.Exp.PV is common (cf. Mathieu 2006), which raises the question why it has disappeared.

The talk opens a new perspective in the debate on Dat.Exp.PV linking language change to principles of language acquisition. My hypothesis is twofold, assuming two parallel developments of Dat.Exp.PV that together lead to its gradual loss – with a few exceptions where the Experiencer is grammaticalised in object position. The first part of my hypothesis is based on markedness in terms of structural complexity: it is assumed that the structure of Dat.Exp.PV is more complex than non-PV verb classes and not acquired easily. The second part of my hypothesis is based on computational efficiency and the assumption that during first language acquisition, rules and exceptions are organised to optimise linguistic

processing. Assuming the Tolerance Principle (cf. Yang 2016), it is argued that preverbal Dat.Exp have not been acquired as a productive rule due to the amount of exceptions to this rule.

On the one hand, it is argued that the Old French PV did not undergo the expected developments towards intransitivity and stativity, which are generally considered to be the properties of less marked argument and event structures (cf. Van Gelderen 2014, 2019; Batllori et al. 2019). Preverbal Dat.Exp can be regarded as a marked input because of the irregular theta-role mapping (cf. Scontras et al. 2015). Language acquisition research shows that marked options are acquired later (cf. Roberts 2007, among others) and Schmitz (2006) argues that dative case is more difficult to be acquired than other cases. On the other hand, I will show that the fixation of French word order from OV to VO during the 12th century led to a low frequency of Dat.Exp in subject position. I will argue that this is the reason why Dat.Exp grammaticalised in object position. This process will be explained by referring to the Tolerance Principle, which has already been applied to Middle English PV and their argument structure (cf. Trips/Rainsford 2022). I will suggest that Old French language learners could not maintain a productive rule which provided a preverbal and a postverbal position for Dat.Exp. Since preverbal Dat.Exp were not as frequent in the PLD as postverbal Dat.Exp, the learners hypothesised as the productive rule for Dat.Exp only the postverbal position.

These hypotheses will be tested by examining two Old French corpora (*MCVF-PPCHF* and the *Nouveau Corpus d'Amsterdam*). Frequencies of both preverbal and postverbal Dat.Exp will be gathered and their argument and event structure will be analysed. A first pretest analysing the PV *plaire* in the *MCVF-PPCHF* showed 268 occurrences of this verb with a Dat.Exp, of which 115 are preverbal and 153 are postverbal. Further results – also concerning the event and argument structure of different Dat.Exp – will be presented in the talk.

References

- Batllori, Montserrat / Gibert-Sotelo, Elisabeth / Pujol, Isabel (2019): “Changes in the argument and event structure of psych verbs in the history of Spanish”, in Bouzouita, Miriam / Breitbarth, Anne /
- Belletti, Adriana / Rizzi, Luigi (1988): „Psych-Verbs and θ -Theory“. *Natural Language & Linguistic Theory*, 6(3), 291–352.
- Fischer, Susann (2019): „On oblique experiencers in psych verb constructions“, in Pomino, Natascha (ed.): *Proceedings of the VII Nereus International Workshop “Morphosyntactic and semantic aspects of the DP in Romance and beyond”*. Arbeitspapier 131. Fachbereich Linguistik, Universität Konstanz 2019, 109-128.
- Hirsch, Nils (2018): *German Psych-Verbs: Insights from a Decompositional Perspective*. Dissertation, Humboldt-Universität zu Berlin, Berlin.

- Kroch, Anthony / Santorini, Beatrice (2021): Penn-BFM Parsed Corpus of Historical French, version 1.0. <https://github.com/beatrice57/mcvf-plus-ppchf>
- Martineau, France, Paul Hirschbühler, Anthony Kroch, and Yves Charles Morin (2021): MCVF Corpus, parsed, version 2.0. <https://github.com/beatrice57/mcvf-plus-ppchf>
- Mathieu, Eric (2006): “Quirky Subjects in Old French“. *Studia Linguistica* 60(3): 282 – 312.
- Roberts, Ian (2007): *Diachronic Syntax*. Oxford: Oxford University Press.
- Schmitz, Katrin (2006): „Indirect objects and dative case in monolingual German and bilingual German/Romance language acquisition“, in Hole, Daniel / Meinunger, Andre / Abraham, Werner (eds.): *Datives and Other Cases*. Amsterdam/Philadelphia: John Benjamins., 239-268.
- Scontras, Gregory / Fuchs, Zuzanna / Polinsky, Maria (2015): „Heritage language and linguistic theory“, *Frontiers in Psychology*. 6:1545.
- Stein, Achim et al. (2006). *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca. 1150-1350)*, établi par Anthonij Dees (Amsterdam 1987), remanié par Achim Stein, Pierre Kunstmann et Martin-D. Gleßgen. Stuttgart: Institut für Linguistik/Romanistik, version 3.
- Trips, Carola / Rainsford, Thomas (2022): “Tolerating subject-experiencers? Yang’s Tolerance Principle applied to psych verbs under contact in Middle English”, *Journal of Historical Syntax*, Vol. 6, No. 12. 1-43.
- Van Gelderen, Elly (2019): “Stability and Change in Intransitive Argument Structure“, *Open Linguistics*, vol. 5, no. 1: 217-232. Wolfgruber, Anne C. (2017): “On, Se and related valency alternations in Medieval French”, *Lingvisticae Investigationes*, 40: 59-80.
- Van Gelderen, Elly (2014): “Changes in Psych-verbs: A reanalysis of little v.”, *Journal of Catalan Linguistics*, 13: 99-122.
- Yang, Charles D. (2016): *The price of linguistic productivity*. MIT Press.

How to use Yang's Principles to model acquisition in diachrony The case of psych verbs

Carola Trips¹ & Tom Rainsford²

¹University of Mannheim, ²University of Stuttgart

Any study looking at acquisition in the past must infer aspects of the input to which children are exposed from written texts, yet it is clear that they are not equivalent. As a model of learnability, Yang's (2016) Tolerance and Sufficiency Principles are good candidates for the study of the acquisition of productive rules in historical data, and they have been applied in a number of recent studies (e.g. Kodner 2019, 2020, 2022; Drescher and Lahiri 2022, Ringe and Yang 2022). The model provides a simple but effective algorithm for predicting the point at which language learners will conclude that the number of lexical items belonging to a given class and providing positive evidence for a particular rule is sufficient to conclude that the rule is productive, barring a small number of exceptions which must be memorized. On the basis of child-directed speech data from the CHILDES corpus (MacWhinney 2000), Yang (2016) has shown that the Tolerance Principle is effective in modelling the acquisition of productive rules such as the use of the *-ed* past-tense marker and its corollary, the Sufficiency Principle, is well-suited to the acquisition of argument structure, such as modelling the subset of ditransitive verbs showing double object constructions in modern English (see also Kodner 2019). Not only has the Sufficiency Principle been shown to be effective in correctly predicting the course of acquisition from small amounts of data, similar to those to which a child would be exposed and to the limited data available to historical linguists, the calculation only requires two parameters to be estimated: the total number of lexical items within the class to which the learner is exposed (henceforth N) and the number of these lexical items to which the rule in question can be applied (henceforth M).

However, applying the Sufficiency Principle to historical data brings a number of unique problems not present in the child-directed speech data examined by Yang. In a recent study of the acquisition of psych verbs in Middle English, Trips and Rainsford (2022) identify three central issues: First is the class size problem: how is it possible to estimate the number of lexical items in a particular class (N), in this case psych verbs, from heterogeneous historical corpora? Second is the attestation problem: what is the best way to estimate the positive evidence for a given rule (M), in this case, the use of a subject-EXPERIENCER argument, from historical data? Third is the data compatibility problem: to what extent is data from historical texts comparable to child-directed speech data?

In the present article, we re-examine the validity of the assumptions made by the authors to address these problems. First, contrasting the psych verbs attested in sections M3 (1350-1420) and M4 (1420-1500) of the *Penn-Helsinki Parsed Corpus of Middle English* (PPCME2) and those attested in modern English child-directed speech from the CHILDES corpus, we show that there is broad semantic equivalence between the most frequent verbs in historical texts and those found in child-directed speech, confirming that basing Sufficiency Principle calculations on a "frequency-trimmed" subset of verbs from historical corpora is the best approach to ensure data comparability (see Kodner 2019). Second,

contrary to Trips and Rainsford (2022), we advocate using corpus data in addition to lexicographical resources to address the attestation problem, showing that this prevents the analysis being affected by hapax constructions recorded in historical dictionaries which are very unlikely to have formed part of the learner's input. We conclude by suggesting a new template for researchers working with models of learnability in diachrony, in which a comparison with modern child-directed speech data forms an essential guide to the correct interpretation of the historical data.

References

- Dresher, B. Elan, and Aditi Lahiri. (2022) “The Foot in the History of English: Challenges to Metrical Coherence.” In *English Historical Linguistics: Change in Structure and Meaning: Papers from the XXth ICEHL*, edited by Bettelou Los, Claire Cowie, Patrick Honeybone, and Graeme Trousdale, 42–59. Current Issues in Linguistic Theory, volume 358. Amsterdam ;Philadelphia: John Benjamins Publishing Company.
- Kodner, Jordan. (2019). “Estimating Child Linguistic Experience from Historical Corpora.” *Glossa. A Journal of General Linguistics* 4, no. 1: 122.
- Kodner, Jordan. (2020). “Language Acquisition in the Past.” Doctoral dissertation, University of Pennsylvania.
- Kodner, Jordan. (2022). “What Learning Latin Verbal Morphology Tells Us about Morphological Theory.” *Natural Language and Linguistic Theory*. <https://doi.org/10.1007/s11049-022-09553-2>.
- Kroch, Anthony, and Ann Taylor, (eds). 2000. *The Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2)*. Philadelphia: University of Pennsylvania.
- B. MacWhinney. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates; CHILDES -- NICHD HD082736.
- Ringe, Don, and Charles Yang. (2022). “The Threshold of Productivity and the ‘Irregularization’ of Verbs in Early Modern English.” In *English Historical Linguistics: Change in Structure and Meaning: Papers from the XXth ICEHL*, edited by Bettelou Los, Claire Cowie, Patrick Honeybone, and Graeme Trousdale, 91–112. Current Issues in Linguistic Theory, volume 358. Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Trips, Carola, and Tom Rainsford. (2022). “Tolerating Subject-Experiencers? Yang’s Tolerance Principle Applied to Psych Verbs under Contact in Middle English” Volume 6, Article 12: 1–43.
- Yang, Charles. (2016). *The Price of Linguistic Productivity*. Cambridge, MA: MIT Press.

**Marked vs. unmarked unaccusativity with alternating verbs:
Linking diachronic and experimental data.**

Alessia Cassarà¹, Achim Stein¹, Chantal van Dijk^{1,2}, Holger Hopp²
¹University of Stuttgart, ²TU Braunschweig

In this talk, we discuss how psycholinguistic studies can help in determining what constitutes a verb class and how this class changes over time. We focus on alternating verbs that can occur in transitive (causative) as well as intransitive (anticausative, unaccusative, inchoative) structures, such as Italian *rompere* ‘to break’. In languages like French or Italian, unaccusative verbs can be either be marked by a reflexive pronoun (R), as in *La chaise se casse* (‘the chair breaks’), or not (U), as in *La temperature augmente* (‘the temperature raises’). Although exhibiting different morphological properties, both marked and unmarked unaccusatives are said to share the same event structure, i.e. the absence of external causation and a subject with non-agentive properties (e.g., Dowty 1979, Jackendoff 1987, Levin & Rappaport Hovav 1995).

Many of the verbs which are marked with the reflexive pronoun in Modern French used to be unmarked in Old French (ex: *fondre* ‘to melt’ in Old French became *se fondre* ‘to melt’ in Modern French), suggesting that change has occurred in this verb class. Auxiliary selection is a second diagnostic that changed, e.g. from OF ‘be’ (*l’eve estoit refroidie* ‘the water has become cold’) to ModF ‘have’ (*l’eau avait refroidie*). We aim to determine what triggered this change from Old to Modern French. Specifically, we investigate whether their shared syntactic structure (presence or absence of the reflexive marker) or semantic factors (shared event structure) play a more decisive role.

We use psycholinguistic methods to address this question in experiments targeting these typical UA properties. In line with previous language processing and priming work by, e.g., Felser (2017) and Kootstra & Muysken (2019), we assume that instances of historical change can also be elicited in synchronic experimental conditions (“change in the lab”) and that the factors causing these changes should also elicit strong priming effects. In this sense, we believe that the use of psycholinguistic methods can shed light on the mechanisms underlying language change.

Several authors have pointed out that Old French displayed a situation similar to that found in Modern standard Italian (e.g. Gougenheim, 1973). In Old French as well as in Italian the diagnostics provide a well-defined definition of unaccusatives, as opposed to Modern French. Therefore we present two experiments on Italian, and plan similar experiments for Modern French in order to parallel the historical change from Old to Modern French.

In a first experiment, we tested the hypothesis that alternating verbs, when primed in their unaccusative form (e.g. *The vase broke*), elicit more unaccusative target structures than when primed with their transitive counterparts (e.g. *The wind broke the vase*). The results of a priming task with 48 native speakers of Italian show evidence of UA priming when primes and targets share the same verbs.

In an ongoing experiment, we explore whether UA priming obtains even if primes and targets contain different verbs. If we observe priming effects, they can be associated either with the syntactic (surface) structure (i.e. the presence or absence of a reflexive marker) or with the semantic (event) structure (i.e. the absence of external causation).

To disentangle the two factors, we designed a follow-up experiment, where the same production task is carried out with cross-lexical prime-target items involving different types of unaccusative verbs (reflexive marked unaccusatives, such as *la sedia si rompe* ‘the chair breaks’ vs. unmarked unaccusative verbs, such as *la pentola bolle* ‘the kettle boils’). Such an experiment will provide evidence as to whether syntactic overlap between prime and target (identical marking) leads to more priming than semantic overlap only (different marking). On the basis of these findings, we will discuss possible links between processing experiments and diachronic change. For instance, if the findings of the experiment reveal that syntactic structure has a stronger priming effect than event structure, we would expect change to occur with verbs that share the same surface structure (either reflexive-marked or unmarked). Conversely, if event structure exhibits a stronger effect, we expect change to have affected both unmarked and marked verbs simultaneously, by virtue of sharing the same semantic properties. More generally, we explore the hypothesis whether priming effects of event structure independent of syntactic overlap suggest that historical change occurs with reference to verb classes that are defined semantically (by shared event structure), rather than syntactically (by presence or absence of the reflexive marker).

References

- Dowty, David. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Felser, Claudia. 2017. Syntactic Ambiguity in Real-Time Language Processing and Diachronic Change. In Hundt, Marianne & Mollin, Sandra & Pfenninger, Simone (eds.), *The Changing English Language: Psycholinguistic Perspectives*, 271–291. Cambridge: Cambridge University Press.
- Gougenheim, Georges. 1973. *Grammaire de la langue française du XVIIe siècle*. Paris.
- Jackendoff, Ray. 1987. The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18. 369–411.
- Kootstra, Gerrit Jan & Muysken, Pieter. 2019. Structural Priming, Levels of Awareness, and Agency in Contact-Induced Language Change. *Languages*. 4(3). 65.
- Legendre, Géraldine. 1989. *Unaccusativity in French*. *Lingua* 79. 95–164.
- Levin, Beth & Rappaport Hovav, Malka. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.

Language Acquisition and a Process-Centered View of Language Change

Jordan Kodner
Stony Brook University

I argue that the actuation of a diverse range of diachronic phenomena in phonology, morphology, and syntax can be subsumed under the process of generalization learning during child language acquisition. These include a secondary split in 20th century Menominee and instance of phonemicization by phonological ‘rule reversal’ in Middle High German (Richter, 2021), the sporadic ‘irregularization’ of Early Modern English past tense forms (Ringe and Yang, 2022), the analogical extension of minority inflectional patterns at the expense of statistically predominant patterns in Late Latin past participles (Kodner, 2022) and Iranian Armenian aorists (Kodner and Dolatian, in prep), ‘Dative Sickness’ ongoing in Icelandic morphosyntax (Nowenstein et al., 2020), and the proliferation of the to-dative construction (Kodner, 2020) and argument structure change for psych-verbs (Trips and Rainsford, 2022) in Middle English. This has broad implications for how we conceptualize language change: an ontology of effects in language change will not line up with an ontology of processes. An approach to the study of change which focuses on *processes or mechanisms* over outcomes and effects stands to bring clarity to a confusing tangle of descriptive phenomena.

The model of generalization learning applied in these studies centers on the Tolerance Principle (TP; Yang, 2016), which provides an exact threshold for the number of exceptions that a linguistic generalization over some scope can tolerate if it is to be entered into a learner’s grammar. Over-regularizations, among the most common innovations in child productions (e.g., Xu and Pinker, 1995; Mayol, 2007) can result from a learner’s calculation over their limited linguistic experience: A TP calculation that would fail over an adult’s lexicon succeeds (perhaps transiently) for the learner, leading to innovation. It is applicable across generalization learning in phonology, morphology, and syntax because it separates the algorithmic aspect of acquisition from the representations over which generalizations are formed (Payne and Yang, 2023), thus a wide range of changes to the grammar may be subsumed under this single mechanism.

In every case investigated here, the TP calculated over acquisition-like samples (Nagy and Anderson, 1984; Yang, 2016; Kodner, 2019) from available corpora reveals patterns of (non-)productivity that are not evident from post-hoc statistical analysis. For example, the TP determines that the statistically predominant Latin participle patterns *-tus* and short *-itus* were actually unproductive. Indeed, they retracted or died out, consistent with this result. But, *-ūtus*, which often supplanted them in Romance, is calculated to be productive within its scope despite its rarity. Thus, this analogical extension works out quantitatively as a standard, albeit fortuitous, case of learner over-regularization. How an innovation like this progresses to language change requires additional population-level mechanisms:

Of course, individual childhood innovations do not entail population-level change, nor is every change child-driven (e.g., Labov, 1994, 2007; Stanford, 2015). Combining insights from competing grammars (Kroch, 1994), with the sociolinguistics of peer-oriented early childhood interaction (e.g., Roberts and Labov, 1995; Nardy et al., 2014; Loukatou et al., 2021), and experimentation on regularization and matching of variable input by children and adults (e.g., Hudson Kam and Newport, 2005; Newport, 2020; Austin et al., 2022), the quantitative predictions of the TP can be extended to model change in the face of population-level variation (Sneller et al., 2019; Kodner and Richter, 2020). This yields insights into why these innovations may progress through actuation and gain a foothold in a population while others may not. This in turn provides a means for distinguishing instances of child-driven from adult-driven change in cases where direct observation is no longer possible.

This work demonstrates that a single mechanism, over-generalization during language acquisition, unites several disparate effects ranging from cases of phonemicization to changes in argument structure. An approach to language change centering the mechanisms or processes (generalization learning, category learning, specific processes of phonetic perception (e.g., Ohala et al., 1981) and production, online syntactic processing, more broadly child- and adult-driven changes, etc.) reconceptualizes the problem space in a way that cross-cuts and reduces traditional taxonomies of effects (analogical leveling, extension, phonemicization, secondary splits, grammaticalization, bleaching, etc.) and opens the door for new insights into when, why, and how language change occurs.

References

- Austin, A. C., Schuler, K. D., Furlong, S., and Newport, E. (2022). Learning a language from inconsistent input: Regularization in child and adult learners. *LLD*, 18(3).
- Hudson Kam, C. L. and Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195.
- Kodner, J. (2019). Estimating child linguistic experience from historical corpora. *Glossa: a journal of general linguistics*, 4(1).
- Kodner, J. (2020). *Language Acquisition in the Past*. PhD thesis, University of Pennsylvania.
- Kodner, J. (2022). What learning Latin verbal morphology tells us about morphological theory. *Natural Language & Linguistic Theory*.
- Kodner, J. and Richter, C. (2020). Transparent /ai/-raising as a contact phenomenon. In Purse, R. and White, Y., editors, *PWPL: NWAV*, volume 25, pages 61–70.
- Kroch, A. (1994). Morphosyntactic variation. In *Proceedings of the thirtieth annual meeting of the Chicago Linguistic Society*, volume 2, pages 180–201.
- Labov, W. (1994). *Principles of linguistic change, volume 1: Internal factors*. John Wiley & Sons.
- Labov, W. (2007). Transmission and diffusion. *Language*, 83(2):344–387.
- Loukatou, G., Scaff, C., Demuth, K., Cristia, A., and Havron, N. (2021). Child-directed and overheard input from different speakers in two distinct cultures.
- Mayol, L. (2007). Acquisition of irregular patterns in Spanish verbal morphology. In *Proceedings of the twelfth ESSLLI Student Session*, pages 1–11, Dublin.
- Nagy, W. E. and Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, pages 304–330.
- Nardy, A., Chevrot, J.-P., and Barbu, S. (2014). Sociolinguistic convergence and social interactions within a group of preschoolers: A longitudinal study. *LVC*, 26(3).
- Newport, E. L. (2020). Children and adults as language learners: Rules, variation, and maturational change. *Topics in Cognitive Science*, 12(1):153–169.
- Nowenstein, I., Sigurjónsdóttir, S., Yang, C., Ingason, A., Wallenberg, J., et al. (2020). The meaning of case: Morphosyntactic bootstrapping and icelandic datives. In *45th Annual Boston University Conference on Language Development*. Cascadilla Press.
- Ohala, J. J., Masek, C. S., Hendrick, R. A., and Miller, M. F. (1981). The listener as a source of sound change. In *Proceedings of the Chicago Linguistics Society*.
- Payne, S. and Yang, C. (2023). Making good on BADS: Commentary on dressler et al. 2020. *Italian Journal of Linguistics*.
- Richter, C. (2021). *Alternation-Sensitive Phoneme Learning: Implications for Children's Development and Language Change*. PhD thesis, University of Pennsylvania.

- Ringe, D. and Yang, C. (2022). The threshold of productivity and the ‘irregularization’ of verbs in Early Modern English.
- Roberts, J. and Labov, W. (1995). Learning to talk Philadelphian: Acquisition of short a by preschool children. *Language Variation and Change*, 7(1):101–112.
- Sneller, B., Fruehwald, J., and Yang, C. (2019). Using the Tolerance Principle to predict phonological change. *Language Variation and Change*, 31(1):1–20.
- Stanford, J. N. (2015). Language acquisition and language change. *The Routledge Handbook of Historical Linguistics*, page 466.
- Trips, C. and Rainsford, T. (2022). Tolerating subject-experiencers? Yang’s Tolerance Principle applied to psych verbs under contact in Middle English. *JHS*, 6(12).
- Xu, F. and Pinker, S. (1995). Weird past tense forms. *Journal of Child Language*, 22(3):531–556.
- Yang, C. (2016). *The Price of Linguistic Productivity*. MIT Press, Cambridge, MA.

1 Introduction by the Organizers

Gerhard Jäger¹, Robert Forkel², Johann-Mattis List^{2,3}

¹ Institute of Linguistics, University of Tübingen

² Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

³ Chair of Multilingual Computational Linguistics, University of Passau

Computational approaches play an increasingly important role in mainstream historical linguistics. Along with these contributions, we note an increased need for standards which drive the curation and sharing of data in historical linguistics (annotated texts, wordlists, collections of structural data, information on phylogenies, etc.). While there have been attempts towards standardization in the past, most prominently reflected in the Cross-Linguistic Data Formats initiative (Forkel et al. 2018), which has been adopted by several teams working on computational and quantitative approaches in the field of historical linguistics, there are still many types of data for which no standards and examples of best practice exist, although they serve frequently as input or output of studies in historical linguistics (e.g. language phylogenies as collected in Greenhill's (2022) "Phlorest collection"). Considering in addition that many new data collections have been published lately (Dellert et al. 2020, List et al. 2022, Kaiping and Klamer 2018), it seems about time to consolidate and discuss which methods we have at our disposal in order to explore highly standardized collections of cross-linguistic data.

The workshop intends to bring together scholars from three different backgrounds: those who work actively on the development of new standards for cross-linguistic data in historical linguistics in particular and comparative linguistics in general, those who design new methods and workflows to explore and exploit standardized data, and those who conduct full-scale analyses of standardized data in order to address concrete scientific problems. The contributions to the workshop can be assigned to one of three key topics: (1) Standards for Cross-Linguistic Data in Historical Linguistics, (2) Methods and Analyses for the Exploitation of Standardized Cross-Linguistic Data, and (3) Research Questions Requiring New/Better Data. Contributions related to key topic (1) present existing standards for linguistic data that have not yet been introduced in historical linguistics, propose new standards for those cases in which standards are lacking, or discuss the role that standards could or should play in historical linguistics (their use, their limits). Contributions to key topic (2) present new methods by which standardized cross-linguistic data can be explored as well as new full-fledged analyses in which specific research questions are addressed by means of workflows that involve standardized cross-linguistic datasets. Contributions to key topic (3) initiate broader discussions on particular research questions that cannot yet be solved but might be solved in the future if sufficiently standardized cross-linguistic data would be available.

Greenhill, Simon J. (2022): Phlorest. <https://github.com/phlorest>

Dellert, Johannes, Thora Daneyko, Alla Münch, Alina Ladygina, Armin Buch, Natalie Clarius, Ilja Grigorjew, et al. 2020. "NorthEuraLex: A Wide-Coverage Lexical Database of Northern Eurasia." *Language Resources & Evaluation* 54: 273–301. <https://doi.org/10.1007/s10579-019-09480-6>.

Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymiski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping, and Russell D. Gray. 2018. "Cross-Linguistic Data Formats, Advancing Data Sharing and Re-Use in Comparative Linguistics." *Scientific Data* 5 (180205): 1–10.

Kaiping, Gereon A., and Marian Klamer. 2018. "LexiRumah: An Online Lexical Database of the Lesser Sunda Islands." *PLOS ONE* 13 (10): 1–29. <https://doi.org/10.1371/journal.pone.0205250>.

List, Johann-Mattis, Robert Forkel, Simon J. Greenhill, Christoph Rzymiski, Johannes Englisch, and Russell D. Gray. 2022. "Lexibank, a Public Repository of Standardized Wordlists with Computed Phonological and Lexical Features." *Scientific Data* 9 (316): 1–31.

2 Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages

Luca Brigada Villa^{1,2}, Erica Biagetti², Chiara Zanchi², Silvia Luraghi²

¹ University of Bergamo

² University of Pavia

Over the past few decades, historical linguistic research has been enriched with the creation of treebanks for several ancient languages. Most developers have adopted the same annotation schemes employed for treebanks of modern languages, often choosing between the two de facto standards of the Penn Treebank phrase-structure format and the Prague Dependency Treebank (PDT) format. The PROIEL scheme (<https://dev.syntacticus.org/proiel.html>), which integrates Dependency Grammar with elements of Lexical Functional Grammar and was originally designed for a parallel treebank of translations of the Gospels in old Indo-European (IE) languages, has been applied to several other texts and is nowadays regarded as a further standard for the annotation of historical IE languages (Eckhoff et al. 2018). The multiplication of projects has led to an ever-growing number of historical treebanks that are incompatible with one another. As a result, new treebanks are created for languages that have already been annotated, but according to a different formalism from the one adopted by the authors. Recently, the annotation scheme designed within the Universal Dependency initiative (UD; Nivre et al. 2016, <https://universaldependencies.org>) has established itself as the standard for dependency annotation. As it favors comparative research, several constituency and dependency treebanks of ancient languages have been converted to UD (notably, we have no knowledge of dependency treebanks being converted to the Penn scheme), and others are now being developed according to this scheme. Yet the achievement of a comparable dataset for historical languages is still hampered by problems related to:

a) coverage and balance of each sub-corpus, b) errors caused by the conversion process, and c) the absence of sufficiently clear and adequate guidelines for the annotation of historical languages.

In this paper, we present the state of the art, some issues and possible solutions to obtain corpora as representative as possible of historical languages. In order not to contribute to the flourishing of individual initiatives, we will open a UD working group dedicated to the annotation of such languages in UD: Universal Dependency for Historical Languages (UD4HL). In this group, we plan to address the following issues with the community. First, tools designed to convert the treebanks to the UD format, such as UDConverter (<https://github.com/thorunna/UDConverter>) and proiel-cli (<https://github.com/proiel/proiel-cli>), need to be further improved to produce cleaner outputs. Second, we aim to stimulate a revision process of both converted and native UD treebanks that tackles one construction type at a time (cf. Brigada Villa et al. 2022, Biagetti et al. 2022): this will make it possible to fix errors caused by the conversion and to provide accurate and consistent guidelines for the annotation of new texts. Finally, the conllu format employed by UD features a MISC (miscellaneous) field that can be enriched with information that is not strictly syntactic but useful for studies on the syntax of historical languages, and is currently underexploited. We propose to add various types of information, such as e.g., metrical information for poetic texts or semantic information regarding the animacy of verbal arguments (PROIEL that had such information in its native format, but this has not been included in the UD converted treebanks). Findings and conclusions reached within the working group will be presented at the conference.

Biagetti, Erica, Chiara Zanchi and Francesco Mambrini. Universal Homeric Dependencies? Towards a complete and updated UD treebank of the Homeric poems. Delbrück Symposium on Indo-European Syntax. Università di Verona, November 9-12 2022.

Brigada Villa, Luca, Erica Biagetti and Chiara Zanchi (2022). Annotating “Absolute” Preverbs in the Homeric and Vedic Treebanks. Proceedings of the Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022; Marseille, June 25, 2022).

Eckhoff, Hanne, Bech, Kristin, Eide, Kristine, Bouma, Gerlof, Haug, Dag T. T., Haugen, Odd E. & Jøhndal, Marius. 2018b. The PROIEL treebank family: A standard for early attestations of Indo-European languages. *Language Resources and Evaluation* 52(1): 29–65.

Nivre, Joakim, De Marneffe, Marie-Catherine, Ginter, Filip, Goldberg, Yoav, Hajic, Jan, Manning, Christopher D., McDonald, Ryan, et al. 2016. Universal Dependencies V1: A Multilingual Treebank Collection. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 1659–66.

3 From Old Data to Fresh Phylogenies — A Linguistic Data Journey in the Times of CLDF

Christoph Rzymiski¹

¹ Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

Historical linguistics involves the study of language change over time, and is often aided by the use of cross-linguistic data. Cross-Linguistic Data Formats (CLDF, Forkel et al. 2018) provides a standardized way to represent and exchange such data, while *cldfbench* (Forkel & List 2020) is a workflow model that facilitates the management and analysis of CLDF data. In this study, we demonstrate how CLDF and *cldfbench* can be used to tackle commonplace tasks in historical linguistics, such as analyzing word lists to identify cognates and building phylogenies. By using CLDF as both input and output, we aim to show how these tools can help streamline the process of working with cross-linguistic data in historical linguistics, from the initial stage of collecting data from “old sources” (i.e., physical sources such as dictionaries and language documentation materials) to the final stage of constructing phylogenies that represent the relationships between languages.

We will demonstrate how to automatically compute cognates (List 2018, List 2021) in word lists using resources such as Concepticon (List 2022) and Glottolog (Hammarström 2022), and how to use these lists as input for BEAST (Bouckaert et al. 2014) to compute phylogenies. Since *cldfbench* supports a workflow that involves using “raw” source data and converting it to one or more CLDF datasets with the help of custom configurations and/or additional Python code, we aim to showcase how this can be utilized to prepare datasets for individual research questions. CLDF, *cldfbench*, and the aforementioned workflows can help researchers to efficiently process and analyze large amounts of data, and facilitate the integration of data from multiple sources.

Overall, our goal is to demonstrate the utility of CLDF and *CldfBench* for researchers in the field of historical linguistics, and to encourage their adoption as standard tools for handling cross-linguistic data. By showcasing innovative approaches to working with standardized cross-linguistic data, we hope to inspire new ideas and perspectives on how to build fresh phylogenies from “old data”.

- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M., Rambaut, A., & Drummond, A. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4), e1003537.
- Forkel, R., List, J.M., Greenhill, S., Rzymiski, C., Bank, S., Cysouw, M., Hammarstrom, H., Haspelmath, M., Kaiping, G., & Gray, R. (2018). Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Scientific data*, 5(1), 1–10.
- Forkel, R., & List, J.M. (2020). CLDFBench. Give your Cross-Linguistic data a lift. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation* (pp. 6997-7004). European Language Resources Association (ELRA).
- Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (2022). CLLD Glottolog 4.7. Max Planck Institute for Evolutionary Anthropology. <https://glottolog.org>.
- List, J. M., Walworth, M., Greenhill, S. J., Tresoldi, T., & Forkel, R. (2018). Sequence comparison in computational historical linguistics. *Journal of Language Evolution*, 3(2), 130-144.
- List, J.M., & Forkel, R. (2021). *LingPy*. A Python library for historical linguistics. Version 2.6.9. URL: <https://lingpy.org>, DOI: <https://zenodo.org/badge/latestdoi/5137/lingpy/lingpy>. With contributions by Greenhill, Simon, Tresoldi, Tiago, Christoph Rzymiski, Gereon Kaiping, Steven Moran, Peter Bouda, Johannes Dellert, Taraka Rama, Frank Nagel. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- List, J.M., Tjuka, A., Rzymiski, C., Greenhill, S., & Forkel, R. (2022). CLLD Concepticon 3.0.0. Max Planck Institute for Evolutionary Anthropology. <https://concepticon.cldd.org>.

4 Phlorest: A Database of Consistent and Reusable Language Phylogenies

Robert Forkel¹, Simon Greenhill²

¹ Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

² School of Psychology, University of Auckland, Auckland

The last few decades have seen the publication of many language phylogenies. These phylogenies have proven to be incredibly powerful tools for making inferences about language relationships (e.g. Gray, Drummond, and Greenhill 2009; Kolipakam et al. 2018; Remco R. Bouckaert, Bowers, and Atkinson 2018; Chang et al. 2015; Greenhill et al. 2022), or as a backbone for testing hypotheses about language change (e.g. Dunn et al. 2011), linguistic reconstructions (e.g. Carling and Cathcart 2021), and evolutionary processes (e.g. Greenhill et al. 2017). Often the results of these phylogenetic studies are repurposed by other researchers to test other hypotheses Watts et al. (2016). Or the results themselves are controversial e.g. witness the arguments about the age of Indo-European Chang et al. (2015) or the debates about language universals Dryer (2011).

We therefore need good ways for researchers to obtain, inspect, compare them, and reuse these phylogenies. However, to date this re-use is hard, often requiring detailed phylogenetic knowledge to identify the relevant files, understand their formats, and extract the critical information. Phlorest is a database of published language phylogenies that aims to standardise the outputs of these analyses to make them Findable, Accessible, Interoperable, and Reusable (Wilkinson et al. 2016). Phlorest collects published language phylogenies into a single database in a consistent and easily usable format (CLDF, Forkel et al. 2018). Currently, Phlorest contains 42 phylogenies, covering a total of 4266 varieties from 2172 languages. Each analysis is preprocessed to a consistent format, providing a summary tree and a posterior tree sample, linked where possible to the raw data. Each taxon in the analysis is mapped to catalogues like Glottolog (<https://glottolog.org>) and D-PLACE (<https://d-place.org/>) so that users can readily identify which languages were included in each analysis.

In this talk we will present Phlorest and discuss the benefits it provides. First, phlorest enables replicability and reuse of these trees. Second, having these phylogenies aligned in time and space enables us to compare patterns and processes across the globe. Third, phlorest allows us to scale up to bigger questions by combining trees into super trees. Finally, phlorest allows us to highlight interesting big picture findings from historical linguistics to the wider public, providing a highly visible resource that brings this research to a wider audience.

- Bouckaert, Remco R., Claire Bowers, and Quentin D. Atkinson. 2018. "The Origin and Expansion of Pama-Nyungan Languages Across Australia." *Nature Ecology & Evolution*. <https://doi.org/10.1038/s41559-018-0489-3>.
- Bouckaert, Remco R, et al. 2012. "Mapping the Origins and Expansion of the Indo-European Language Family." *Science* 337 (6097): 957–60. <https://doi.org/10.1126/science.1219669>.
- Carling, Gerd, and Chundra Cathcart. 2021. "Reconstructing the Evolution of Indo-European Grammar." *Language* 97 (3): 561–98. <https://doi.org/10.1353/lan.2021.0047>.
- Chang, Will, Chundra Cathcart, David Hall, and Andrew Garrett. 2015. "Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis." *Language* 91 (1): 194–244. <https://doi.org/10.1353/lan.2015.0005>.
- Dryer, Matthew S. 2011. "The Evidence for Word Order Correlations." *Linguistic Typology* 15: 335–80. <https://doi.org/10.1515/LITY.2011.024>.
- Dunn, Michael, Simon J. Greenhill, S. C. Levinson, and Russell D. Gray. 2011. "Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals." *Nature* 473 (7345): 79–82. <https://doi.org/10.1038/nature09923>.
- Forkel, Robert, et al. 2018. "Cross-Linguistic Data Formats, Advancing Data Sharing and Re-Use in Comparative Linguistics." *Scientific Data* 5 (1): 180205. <https://doi.org/10.1038/sdata.2018.205>.
- Gray, Russell D., Alexei J Drummond, and Simon J. Greenhill. 2009. "Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement." *Science* 323 (5913): 479–83. <https://doi.org/10.1126/science.1166858>.
- Greenhill, Simon J., Hannah J. Haynie, Robert M Ross, Angela M. Chira, Johann-Mattis List, Lyle Campbell, Carlos A. Botero, and Russell Gray. 2022. "A Recent Northern Origin for the Uto-Aztecan Family," August. <https://doi.org/10.31235/osf.io/k598j>.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. "Evolutionary Dynamics of Language Systems." *Proceedings of the National Academy of Sciences*, 201700388. <https://doi.org/10.1073/pnas.1700388114>.
- Kolipakam, Vishnupriya, Fiona M Jordan, Michael Dunn, Simon J Greenhill, Remco Bouckaert, Russell D Gray, and Annemarie Verkerk. 2018. "A Bayesian Phylogenetic Study of the Dravidian Language Family." *Royal Society Open Science* 5: 171504. <https://doi.org/10.1098/rsos.171504>.
- Levy, Roger, and H Daumé III. 2011. "Computational Methods Are Invaluable for Typology, but the Models Must Match the Questions." *Linguistic Typology* 15: 393–99. <https://doi.org/10.1515/LITY.2011.026>.
- Watts, Joseph, Oliver Sheehan, Quentin D. Atkinson, Joseph Bulbulia, and Russell D. Gray. 2016. "Ritual Human Sacrifice Promoted and Sustained the Evolution of Stratified Societies." *Nature* 532 (7598): 228–31. <https://doi.org/10.1038/nature17159>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

5 Configurable Language-Specific Tokenization for CLDF Databases

Johannes Dellert¹, Verena Blaschke²

¹ Institute of Linguistics, University of Tübingen

² Center for Information and Language Processing, LMU Munich

In any workflow for computational historical linguistics, tokenization of IPA sequences is a crucial pre-processing step, as it shapes the alignments which provide the input of algorithms for cognate detection and proto-form reconstruction. This is also true for EtInEn (Dellert 2019), our forthcoming integrated development environment for etymological theories. An EtInEn project can be created from any CLDF database such as the ones that have been aggregated and unified by the Lexibank initiative (List et al. 2022). Whereas the tools for preparing CLDF databases (Forkel & List 2020) encourage the application of a uniform tokenization across all languages in a dataset, our view is that in many contexts, it is more natural to tokenize phonetic sequences in ways that differ between languages. To provide a simple example, many geminates in Italian need to be aligned to consonant clusters in other Romance languages (e.g. *notte* vs. Romanian *noapte* “night”), which is much easier if they are tokenized into two instances of the same consonant, whereas geminates in Swedish are best treated as cognate to their shortened counterparts in other Germanic languages.

To provide comprehensive support for such cases, EtInEn includes configurable language-specific tokenizers as an additional abstraction layer that allows to reshape forms after the import, and also serves as a generic way to bridge phonetic surface forms and the underlying forms that historical linguists are primarily interested in. Each tokenizer is defined by a token alphabet which is used for greedy tokenization, a list of allophone sets which can be used to abstract over irrelevant subphonemic distinctions, and a list of non-IPA symbols that are defined in terms of phonetic features. The initial state of each tokenizer is based on an analysis of the tokens used by the imported CLDF database. Tokenizer definitions are stored in a human-editable plain-text format which we would like to propose as a new standard.

In EtInEn, tokenizer definitions are manipulated through a graphical editor in which the potential tokens for each language are arranged in the familiar layout of consonant and vowel charts, enhanced by additional panels for diphthongs and tones. Currently defined tokens are highlighted, and allophone sets are summarized under their canonical symbols. Basic edit operations serve to group several sounds into an allophone set, and to join or split a multi-symbol sequence, such as a diphthong or a sound with a coarticulation. More complex operations support workflows for parallel configuration of multiple tokenizers.

Additional non-IPA symbols can be given semantics in terms of a combination of phonetic features, and declared to be part of the token set for any language. On the representational level, this provides the option to use non-IPA symbols for form display, whereas underlyingly, the system will interpret the symbols in terms of their features. On the conceptual level, underspecified definitions provide support for metasympols. In addition to some predefined metasympols (such as V for vowels and C for consonants), the user can assign additional symbols to arbitrary classes of sounds. These are then available throughout EtInEn for various purposes, such as concisely representing the conditioning environments for a soundlaw, or summarizing the probabilistic output of an automated reconstruction module.

In addition to configurable tokenizers, EtInEn provides the option to define form-specific tokenization overrides, allowing to substitute the result of automated tokenization with any sequence over the current token alphabet for the relevant language. This is currently our strategy for handling otherwise challenging phenomena such as metathesis or root-pattern morphology, which we normalize into alignable and concatenative representations. This forms a bridge to existing standards for representing morphology in the CLDF framework (e.g. Schweikhard & List 2020), which currently only support the annotation of morpheme boundaries in terms of simple splits in phonetic IPA sequences.

Dellert, Johannes (2019): “Interactive Etymological Inference via Statistical Relational Learning.” Workshop on Computer-Assisted Language Comparison at SLE-2019.

Forkel, Robert and Johann-Mattis List (2020): “CLDFBench. Give your Cross-Linguistic data a lift.” Proceedings of LREC 2020, 6997-7004.

List, Johann-Mattis, Robert Forkel, S. J. Greenhill, Christoph Rzymiski, Johannes Englisch & Russell Gray (2022): “Lexibank, A public repository of standardized wordlists with computed phonological and lexical features.” *Scientific Data* 9.316, 1-31.

Schweikhard, Nathanael E. and Johann-Mattis List (2020): “Developing an annotation framework for word formation processes in comparative linguistics.” *SKASE Journal of Theoretical Linguistics* 17(1), 2-26.

6 A computational evaluation of regularly recurring sound correspondences

Frederic Blum¹, Johann-Mattis List^{1, 2}

¹ Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

² Chair of Multilingual Computational Linguistics, University of Passau

Regularly recurring sound correspondences are the main tools of the comparative method (Anttila 1972; Lass 1997). The cognate judgements which are based on these correspondences are also used in the phylogenetic approaches to historical linguistics that have received widespread attention in recent years (Greenhill et al. 2020). However, regularity is often more an intuitive notion than a quantified evaluation, and irregularity is argued to be more common than expected from the Neogrammarian hypothesis (Durie & Ross 1996; Labov 1981). Given the recent development of computational methods in historical linguistics and the availability of cross-linguistic comparative formats (Forkel et al. 2018; List 2019), we are now able to improve our workflows in this regard.

We provide a computational machinery that can be used as a means to improve the annotation of cognates in a standardized data set. For this, we focus on a quantitative measure for assessing the regularity of sound correspondences across cognates. This can, for example, be used to compare the results of different automated methods of cognate judgements and alignments, or to identify possible errors in expert cognate annotations. Our workflow proceeds in four stages. In the first stage, we carry out a phonetic alignment analysis (List et al. 2018) of all cognate sets in a standardized wordlist. In the second stage, we preprocess the phonetic alignments by excluding spurious alignment sites (columns in a multiple phonetic alignment). In the third stage, we search for recurring correspondences across our aligned cognate sets and determine potentially regular correspondence patterns. In a fourth stage, we score the overall regularity of the individual cognate sets in our data by counting how many sites in the alignments can be represented by recurring (regular) correspondence patterns, and how many are unique.

In the talk, we showcase the functionality of this workflow using data from the Pano-Tacanan language family. We will focus on two key issues: the automated detection of potential false positive cognate judgements, as well as the detection of potential false negatives. Potential false positives are identified as words in a cognate set with very low regularity in the correspondence patterns across the data set. For the detection of potential false negatives, we compare two different sets of cognate annotations of the same data. If no second expert annotation is available, the first annotation can be compared to an automated judgement of cognacy (List 2019). We identify all cognate words above a custom regularity threshold that are assigned different cognacy in the first set of annotations, but are part of the same cognate set in the second annotation. We show how different thresholds influence the results and discuss possible further applications and developments of this workflow.

Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics*. New York: The Macmillan Company.

Durie, Mark & Malcolm Ross. 1996. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. New York, Oxford: Oxford University Press.

Forkel, Robert, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymiski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-linguistic data formats, advancing data sharing and re-use in comparative linguistics. *Scientific Data* 5(1). <https://doi.org/10.1038/sdata.2018.205>.

Greenhill, Simon J., Paul Heggarty & Russell D. Gray. 2020. Bayesian phylolinguistics. In Barbara S. Vance Richard D. Janda Brian D. Joseph (ed.), *The Handbook of Historical Linguistics*, chap. 11, 226–253. Wiley. <https://doi.org/10.1002/9781118732168.ch11>.

Labov, William. 1981. Resolving the Neogrammarian Controversy. *Language* 57(2). 267–308. <https://doi.org/10.2307/413692>. Lass, Roger. 1997. *Historical linguistics and language change*. Cambridge: Cambridge University Press.

List, Johann-Mattis. 2019. Automatic inference of sound correspondence patterns across multiple languages. *Computational Linguistics* 45(1). 137–161. https://doi.org/10.1162/coli_a_00344.

List, Johann-Mattis, Mary Walworth, Simon J. Greenhill, Tiago Tresoldi & Robert Forkel. 2018. Sequence comparison in computational historical linguistics. *Journal of Language Evolution* 3(2). 130–144. <https://doi.org/10.1093/jole/lzy006>.

7 Exploring the Geographical Distribution of Missing Data Using Approximate Gaussian Processes

Miri Mertner¹, Matías Guzmán Naranjo¹

¹ Institute of Linguistics, University of Tübingen

Gaussian processes (GPs) have several qualities that make them well-suited to spatial statistics, as they allow us to add non-linear effects to a model in a flexible way (see e.g. McElreath, 2020, Chapter 14, for an explanatory example). A GP essentially estimates the effect that every observation has on every other observation in the form of a covariance matrix, which can then be used, for example, as a predictor in a model. In linguistic typology, they have been used as a way to control for spatial autocorrelation between languages, as well as for inferring probable ranges of contact between languages (Guzmán Naranjo & Mertner, 2022). However, they can be prohibitively slow to use with large datasets, such as the global sample of languages included in WALS or Glottolog. Therefore, in order to use them on such large datasets, an approximation of the GP is required.

One of the cases in which a large dataset is necessary to make meaningful inferences is in the exploration of the distribution of missing data in linguistic databases such as WALS (Dryer & Haspelmath, 2013) and ASJP (Wichmann et al., 2022). Using approximate GPs implemented in the programming language Stan (Stan Development Team), the present study will focus on uncovering areal biases in the distribution of missing linguistic data. Geographical and social correlates which could help explain the causal factors behind a higher or lower density of missing data in a particular area will also be tested, such as landscape roughness, climate, and population size. A better understanding of the factors which lead to geographical imbalances in the distribution of missing data could, among other things, improve our ability to impute missing data as part of statistical modelling work.

Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info>, accessed on 2023-01-01.)

Guzmán Naranjo, Matías and Mertner, Miri. 2022. Estimating areal effects in typology: a case study of African phoneme inventories. *Linguistic Typology*. <https://doi.org/10.1515/lingty-2022-0037>

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>

Stan Development Team. 2022. *Stan Modeling Language Users Guide and Reference Manual*, 2.31. <https://mc-stan.org>

Wichmann, Søren, Eric W. Holman, and Cecil H. Brown (eds.). 2022. *The ASJP Database* (version 20).

The areality of the consecutive pattern in Mesoamerican languages

It is clear that clause-linkage strategies can be transferred across family lines with or without substance. The theoretical importance of exploring this domain has been highlighted by various typological studies (e.g. Schmidtke-Bode 2009: 202–203; Martowicz 2011: 327; Hetterle 2015: 269). However, we are just beginning to learn about the myriad ways this can happen. The present paper analyzes the areality of the consecutive construction, a clause-linkage pattern that has received little attention cross-linguistically.

The term consecutive refers to constructions in which only the first clause shows the formal characteristics of an independent clause, and the following clause or clauses are characterized by a reduction or lack of verbal inflection, and/or by the use of a verbal form called the CONSECUTIVE (Creissels et al. 2008:140; Vydrin 2020: 85). In (1), the temporal frame of the discourse is initially anchored with the past tense marker *-a-*, and the second clause appears with the consecutive marker *ka-*.

Manda (Atlantic-Congo/Bantu)

- | | | |
|-----|--|---|
| (1) | <i>va-a-l-ili,</i> 3PL.SBJ-PST-eat-PST | <i>va-ka-wok-a...</i> 3PL.SBJ-CONS-depart-FV |
| | ‘They ate, and then they went from there...’ (Bernander 2017: 196) | |

While the consecutive is common in many African languages, this construction is also attested in other areas of the world, such as Australia (Jones 2011: 270), and Oceanic (Lynch 1978: 50). This suggests that the consecutive is not a unique African phenomenon. Intriguingly, the consecutive pattern appears in various parts of the world in areal clusters. If neighboring languages have similar rare patterns for expressing temporal subsequence, it is statistically unlikely that these languages have undergone such a developmental process independently of one another. Accordingly, the parallelisms are not easily explained by chance. They cannot be explained as a common inheritance, because the languages are not all genetically related. The most likely explanation is language contact, because the languages are spoken in the same geographical region, but it is difficult to see how such fundamental but abstract patterns could be transferred from one language to another.

The question is: How could such deeply-integrated grammatical systems be transferred, usually without substance? Here this question is explored in four genealogically Mesoamerican unrelated languages: Huasteca Nahuatl, Papantla Totonac, San Gabriel Huastec, and Uxpanapa Chinantec.

Based on a number of intra-genetic variance analyses, systematically informed by what is known from social/cultural history, it is proposed that Huasteca Nahuatl served as the source. The consecutive pattern in Huasteca Nahuatl has different functions. It is used for indicating temporal subsequence, motion-cum-purpose, tail-head linkage, afterthoughts, and commands. Intriguingly, while some neighboring languages have copied some of these functions from Huasteca Nahuatl, others have also copied some of these functions, and developed others.

Although some details must remain an object of speculation, close comparison of the systems for marking the consecutive in Huasteca Nahuatl, Papantla Totonac, San Gabriel Huastec, and Uxpanapa Chinantec provides a glimpse of some ways in which areal concentrations might develop.

References

- Bernander, Rasmus. 2017. *Grammar and grammaticalization in Manda. An analysis of the wider TAM domain in a Tanzanian Bantu language*. Helsinki: University of Helsinki doctoral dissertation.
- Creissels, Denis, Gerrit Dimmendaal, Zygmunt Frajzynger, & Christa König. 2008. Africa as a morpho-syntactic area. In Bernd Heine & Derek Nurse (eds.), *Africa as a morpho-syntactic area. A linguistic geography of Africa*, 86-150. Cambridge: Cambridge University Press.
- Hetterle Katja. 2015. *Adverbial clauses in cross-linguistic perspective*. Berlin/Boston: De Gruyter Mouton.
- Jones, Barbara. 2011. *A grammar of Wangkajunga: A language of the Great Sandy Desert of north Western Australia*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Lynch, John. 1978. *A grammar of Lenakel*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Martowicz, Anna. 2011. *The origin and functioning of circumstantial clause linkers: A cross-linguistic study*. Edinburgh: University of Edinburgh doctoral dissertation.
- Schmidtke-Bode, Karsten. 2009. *A typology of purpose clauses*. Amsterdam/Philadelphia: John Benjamins.
- Vydrin, Valentin. 2020. Clause chaining in Bambara. In Nina Pawlak & Izabela Will (eds.), *West African languages: Linguistic theory and communication*, 79-98. University of Warsaw Press.

From ecological to lexical diversity: measuring vocabulary richness in historical corpora

The question whether and how we can measure lexical diversity has long been a pertinent one in Linguistics and related disciplines. Attempts have been made to estimate the vocabulary size of (average speakers of) a particular language (at different ages) (e.g. Ellegård 1960, Brysbaert et al. 2016, Segbers & Schroeder 2017), and many studies in (Diachronic) Construction Grammar are concerned with estimating the number of unique lexical items that may occur in particular morphosyntactic structures for different individuals or across time (e.g. Schmid & Mantlik 2015; Perek 2018). To address these questions, researchers often resort to corpus research, using quantitative measures that rely on type and token frequency and/or hapax legomena, such as (variations on) Mean Word Frequency (MWF) and Type-Token Ratio (TTR) (see Tweedie & Baayen 1998), and realized/potential/expanding productivity (Baayen 2009).

However, in historical corpora, unique character strings cannot always be equated to unique words. This may be due to spelling variation or OCR errors (e.g. the Modern English character <f> is often mistaken for <f> or <l>; thus *strength* <frength> can also be represented by <frength> and <lrength>). Because neither OCR errors nor non-standard spelling variations are entirely systematic, reducing such variation through corpus pre-processing can be challenging.

As a solution, we propose an approach originally developed to estimate ecological diversity (Chao et al. 2019) called the attribute diversity framework, which distinguishes categorical diversity from functional diversity. We define ‘categorical diversity’ as the number of unique ‘items’ (i.e. unique character strings) in a text, and ‘functional diversity’ as a measure that also takes into account the distributional similarity of these items. Operationalizing this distributional similarity by means of word embeddings generated with the historically pre-trained language model MacBERTh (Manjavacas & Fonteyn 2022), we demonstrate that:

- (i) Functional diversity estimates are affected to a much lesser extent by spelling inconsistencies and OCR errors than categorical diversity.
- (ii) Given two sets of unique word types, set A{*dog, bird, rabbit*} and set B{*progesterone, remember, blue*}, the approach also captures the higher functional-semantic diversity of set B.

As a concrete case study to demonstrate the theoretical and practical advantages of discussing ‘vocabulary richness’ or lexical diversity in terms of attribute diversity, we use the diachronic ARCHER corpus (version 3.2) and discuss diachronic changes in and differences between texts from different genres and by different authors in terms of categorical as well as functional diversity.

Baayen, R. H. 2009. Corpus linguistics in morphology: Morphological productivity, in: A. Lüdeling, M. Kytö (eds.), *Corpus Linguistics: An International Handbook*, 899–919. Berlin: De Gruyter Mouton.

Brysbaert, M., M. Stevens, P. Mandera & E. Keuleers. 2016. How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant’s age. *Frontiers in Psychology* 7.

Chao, A., C. Chiu, S. Villéger, I. Sun, S. Thorn, Y. Lin, J. Chiang & W. B. Sherwin. 2019. An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. *Ecological Monographs* 89.

Ellegård, A. 1960. Estimating vocabulary size. *WORD* 16: 219–244.

Manjavacas, E. & L. Fonteyn. 2022. Adapting vs. Pre-training Language Models for Historical Languages. *Journal of Data Mining & Digital Humanities* jdmdh:9152.

Schmid, H.-J. & A. Mantlik. 2015. Entrenchment in historical corpora? reconstructing dead authors’ minds from their usage profiles. *Anglia* 133: 583–623.

Segbers, J. & S. Schroeder. 2017. How many words do children know? A corpus-based estimation of children’s total vocabulary size. *Language Testing* 34: 297–320.

Perek, F. 2018. Recent change in the productivity and schematicity of the way -construction: A distributional semantic analysis, *Corpus Linguistics and Linguistic Theory* 14: 65–97.

Tweedie, F. J. & R. H. Baayen. 1998. How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities* 32: 323–352.

Towards an account of the emergence, evolution and variability of emphatic negative coordination in Indo-European, part 2: A diachronic perspective

Iker Salaberri
(University of the Basque Country)

Keywords: <negation, clause linkage, Indo-European, historical and comparative linguistics>

Correlative constructions that both negate and coordinate, such as Armenian *oč'... oč'...*, English *neither... nor...* and Hindi *nə... nə...* are widespread in Indo-European (IE) languages (Salaberri 2022: 679). These so-called *emphatic negative coordination* (ENC) (Haspelmath 2007) constructions can be reconstructed for the proto-language (Pokorny 1959: 756–757). Even though comprehensive studies on this topic have been recently published (Badiou-Monferran 2004, Liljegren 2016: 349–350, Gajić 2018, Briceño 2019: 123–127, Van der Auwera et al. 2021, Van der Auwera & Koohkan 2022, among others), it is unclear exactly how this clause linkage strategy emerged in the proto-language, how it developed in the various IE branches, and what led to the considerable variation found in the modern languages.

As the second (diachronic) part of a larger study on ENC constructions, this paper analyzes four features: (i) the diachronic origin of ENC markers; (ii) the syntactic complexity of the elements coordinated in ENC constructions, i.e., whether these are phrases, clauses or sentences; (iii) the degree of independence of ENC markers, i.e., whether they suffice on their own or must be accompanied by reinforcing elements such as standard negators and coordinating conjunctions; (iv) the number of ENC strategies in each language. A sample of 240 IE languages with data from reference grammars and dedicated publications is analyzed by means of Cramer's V (Cramer 1946) test, whereby the aforementioned four features are treated as dependent variables and controlled through Theil's uncertainty coefficient (Theil 1966). The results of correlation analysis are used to trace the diachronic evolution of ENC constructions and form a typologically informed hypothesis concerning how these constructions originated in the proto-language.

The results suggest the existence of a four-way typology of ENC constructions: the most frequent type involves correlative ENC markers of the type *nV... nV...* (1a), which link all kinds of coordinands and are attested in most historical stages. However, there are also languages with only non-correlative ENC constructions (1b), languages where ENC markers must be obligatorily reinforced and those where the original *nV... nV...* pattern has been replaced by innovative forms. In addition, some languages display multiple ENC constructions, the choice of which sometimes depends on the syntactic complexity of the coordinands.

- (1) a. *Unio kule bôlben na, amio*
3SG.also open.PP speak.3SG.H.FUT ENCM 1SG.also
na jene charbô na
ENCM know.PP leave.1SG.FUT NEG
,Neither would he tell me straight nor was I willing to let it go without knowing' (Thompson 2012: 302)

(Bengali)

- b. *Ni frithalim-se rucai na-mmebuil*
NEG expect-1SG shame nor-disgrace
,I expect neither shame nor disgrace' (Thurneysen 1946: 540)

(Early Irish)

The data likewise suggest that, much like other negative elements, ENC markers bleach over time and are regularly reinforced and renovated. Accordingly, it is argued that ENC markers of the kind *nV.. nV..* must have originated in the proto-language from the univerbation of negator and conjunction or another kind of emphatic element. They have been subsequently reinforced and renovated in different ways in different IE languages, therefore the variation observable nowadays.

Abbreviations

| | |
|------|---------------------------------------|
| 1/3 | 1st/3rd person |
| ENC | emphatic negative coordination |
| ENCM | emphatic negative coordination marker |
| FUT | future |
| H | honorific |
| NEG | negator |
| PP | perfective participle |
| SG | singular |

References

- Badiou-Monferran, Claire (2004), Négation et coordination en français classique: le morphème *ni* dans tous ses états, *Langue française* 143, 69–92.
- Briceño, Juan E. (2019), *Negation, indefinites, and polarity in early Greek and Indo-Iranian: a typological and comparative approach*, Madrid: Universidad Complutense PhD dissertation.
- Cramer, Harold (1946), *Mathematical methods of statistics*, Princeton: Princeton University Press.
- Gajić, Jovanna (2018), *Niti – real negative coordination inside a strict negative concord language?*, in A. van Alem, A. Ionova and C. Pots (eds.), *Proceedings of the 26th Conference of the Student Organization of Linguistics in Europe*, Leiden: Leiden University Center for Linguistics, 1–19.
- Haspelmath, Martin (2007), Coordination, in T. Shopen (ed.), *Language typology and syntactic description*, Vol. 2: *Complex constructions*, 2nd edn., Cambridge: Cambridge University Press, 1–51.
- Liljegren, Henrik (2016), *A grammar of Palula*, Berlin: Language Science Press.
- Pokorny, Julius (1959), *Indogermanisches etymologisches Wörterbuch*, Vol. 1, Bern / Munich: Francke Verlag.
- Salaberri, Iker (2022), A cross-linguistic study of emphatic negative coordination, *Studies in Language* 46(3), 647–717.
- Theil, Henri (1966), *Applied economic forecasting*, Amsterdam / Chicago: North Holland Publishing Company, Rand McNally.
- Thompson, Hanne-Ruth (2012), *Bengali*, London: School of Oriental and African Studies.
- Thurneysen, Rudolf (1946), *A grammar of Old Irish*, 2nd edn., Dublin: Dublin Institute for Advanced Studies.
- Van der Auwera, Johan, Motoki Nomachi & Olga Krasnoukhova (2021), Connective negation and negative concord in Balto-Slavic, in P. Arkadiev, J. Pakerys, Inesa Šeškauskienė and V. Žeimantienė (eds.), *Studies in Baltic and other languages: a Festschrift for Aexel Holvoet on the occasion of his 65th birthday*, Vilnius: Vilnius University Press, 45–66.
- Van der Auwera, Johan & Sepideh Koohkan (2022), Extending the typology: negative concord and connective negation in Persian, *Linguistic Typology at the Crossroads* 2(1), 1–36.

Contact-driven grammaticalization and drift of new terminal tenses from go-periphrasis in Azeri and Kryz (East Caucasian)

Northern Azerbaijan is home to a dozen East Caucasian languages. High levels of bilingualism in Azeri have prompted many subtle typological shifts in their morphosyntax. As a result of recent and language-particular individual innovations, the rich Tense-Aspect-Mood system of most East Caucasian languages mostly overlaps with the Azeri system. However, unlike Turkic, the East Caucasian languages retain inherited ergative morphology, realized as ergative case marking on the Agent of transitive verbs, and most of them show gender/number agreement of the S or P argument on verbs.

Reminiscent of but different from the ‘aspectual compositions’ seen in other Turkic languages, which express actionality by using posture and movement verbs (Johanson 2021), spoken Azeri juxtaposes a perfective (witnessed) past tense form of the verb ‘go’ with the conjugated lexical verb in order to express a ‘recent perfect’ tense, often glossed by native speakers with the adverbials ‘already’ or ‘quickly’ (ex. 1a&b). The construction is biclausal and clearly grammaticalized, as seen in ex. 2 with a transitive verb.

A previously undescribed copy-construction is found in Kryz, spoken in the Quba region. Kryz has two past perfective tense, the aorist and the perfect, matching the the witnessed and unwitnessed perfective past tenses of Azeri (ex. 3). Verbs do not agree in person but in gender/number (human masculine / feminine (including animals and many inanimates) / neuter / human plural) with either the Single argument or the Patient of intransitive or transitive verbs respectively. In Kryz the ‘go’ periphrasis has fused in the aorist tense into a synthetic paradigm for which elicitation of parallel expanded analytic forms is not available for the unverbated feminine form (ex. 4b), which shows the selection of an archaic ending (-*d* instead of -*d-u* in the current aorist paradigm), while the human plural form (ex. 4d) shows haplology (-*cib+yip*- => -*cip*-). With a transitive lexical verb, the ergative case marking is unchanged (ex. 5 a,b&c).

The Azeri construction is available with most TAMs and persons. In Kryz likewise, the verb ‘go’ can be used with personal clitics and various tense and moods. But in the derived terminal present tense of a transitive verb (ex. 6a&b), suffixed agreement markers of a transitive lexical verb are replaced by a default (neuter/non-human plural) ending. In other combinations, like the derived terminal imperative, and all forms with personal enclitics, the auxiliary has to be preceded by a typical ‘bounder’, which is a lexicalized short stem of the verb ‘go out’, meaning ‘away’ (ex. 7a&b).

The semantics and pragmatic use of this new ‘iamitive’ (?) perfect remain to be clarified, but the two constructions, while sharing a similar starting point in Azeri and Kryz (a focus on the endpoint of a process) show both parallel and different outcomes: like the Azeri construction, the Kryz equivalent gained autonomy in spreading to other TAMs, but the originally intransitive auxiliary, whether unverbated or not, assumed the valency and gender/number agreement of the lexical verb, and became more or less fused in all gender-marked third person forms of the new ‘immediate perfect’ and ‘immediate present’ paradigms (ex. 4, 5 and 6).

Reference: Johanson, Lars. (2021). Postverbal Constructions. In *Turkic*, pp. 597-617). Cambridge: Cambridge University Press.

Examples :

- 1.a *qaç-di-m get-di-m.* run-WPST-1SG go-WPST-1SG
‘I have already escaped.’
- 1.b *qoyun qaç-di get-di.* sheep run-WPST(3) go-WPST(3)
‘The sheep has already escaped.’
2. *qoyun ye-di-m get-di.* sheep eat-WPST-1SG go-WPST(3)
‘I have already eaten the sheep.’
3. *eb-ili-r şayal-bi ula-cib / ula-ca-b.* wolf-OBL-ERG child-PL.NOM eat.PF-AOR.HPL eat.PF-PERF-HPL
‘The wolf ate / has eaten the children.’
- 4.a *gada k’ul-ca şaxhircixhid < *şaxhir-d yixh-id* boy(NOM) house-OBL.IN (M)arrive.PF-AOR.M go.PF-AOR.M
‘The boy has already arrived home.’
- 4.b *riş k’ul-ca şaxhurcipdu < *şaxhur-d(ıt) yip-du* girl(NOM) house-OBL.IN F.arrive.PF-AOR(F) go.PF-AOR.F
‘The girl has already arrived home.’
- 4.c *vul-bi şaxhircixhic < *şaxhr-ic yixh-ic.* sheep-NPL.NOM (M)arrive.PF-AOR.N(PL) go.PF-AOR.N(PL)
‘The boy has already arrived home.’
- 4.d *şayal-bi k’ul-ca şaxhurcipcib < *şaxhur-cib+yipcib. (HAPLOLOGY)* child-PL.NOM house-OBL.IN F.arrive.PF-AOR.HPL+go.PF-AOR.HPL
‘The children have already arrived home.’
- 5.a *riş-ir fu ulacixhic < *ula-c yixh-ic.* girl-ERG bread(N) eat.PF-AOR.N (M/N)go.PF-AOR.N
‘The girl has already eaten the bread.’
- 5.b *gada-r bicah ulacipdu < *ula-d yip-du.* boy-ERG pilav(F) eat.PF-AOR+F.go.PF-AOR.N
‘The boy has already/quickly eaten the plov.’
- 5.c *eb-ili-r şayal-bi ula-cipcib < *ula-cib+yipcib. (HAPLOLOGY)* wolf-OBL-ERG child-PL.NOM eat.PF-AOR.HPL+eat.PF-PERF-HPL
‘The wolf quickly ate the children.’
- 6.a *a-n-ir vul-bi haluca ula-c çe-re.* DIST-H-ERG sheep-NPL(NOM) on_the_spot eat.PF-AOR.N go.IPF-PRS(NPL)
‘She has already eaten the sheep on the spot.’
- 6.b *eb-il-ir şayal-bi ula-c ça-ba-re-b.* wolf-OBL-ERG child-PL.NOM eat.PF-AOR.N go.IPF-HPL-PRES-HPL
‘The wolf has already eaten the children.’
- 7.a *bicah ulats’-ryu=zın ğabç’ çuryu. = Az. qoyunu yeyirəm gedir.* pilav(F) eat.IPF-PRES.F=1SG F.go_out.PF go.IPF.PRES.F
‘I hurry up eating the plov.’
- 7.b *bicah seyil ğabç’ yip-i = Az. plovu ye getsin!* pilav(F) eat.IMPER F.go_out.PF F.go.PF-OPT
‘Hurry up eating the plov !’

26th International Conference on Historical Linguistics, 4 to 8 September 2023

Towards a New Reconstruction of the Proto-Yeniseian Sound System

SVENJA BONMANN, SIMON FRIES, LAURA GÜNTHER, NATALIE KOROBZOW, EUGEN HILL

Universität zu Köln and Julius-Maximilians-Universität Würzburg

The Yeniseian (also Yeniseic, abbreviated: Yen.) language family consists of six members: Ket, Yugh, Pumpokol, Arin, Assan and Kott. Apart from Ket, which has very few competent speakers by now, all Yeniseian languages are extinct. Ket, Yugh and Kott are reasonably well-attested (cf. Werner 1997a,b, 1998; Vajda 2004; Georg 2007; Kotorova & Nefedov 2015), whereas our limited knowledge of Arin, Assan and Pumpokol is based on fragmentary word-lists of the 18th and 19th centuries (cf. Werner 2005). Lexical correspondences and striking similarities in the gender, case or verb systems (cf. I. Verner 1969; G. Verner 1988; G. Starostin 1995) bear witness to the genetic relatedness of these six languages; this shared grammatical and lexical basis must be explained as inheritance from a common ancestor, Proto-Yeniseian (PY). It is commonly assumed that the PY homeland was probably situated near the headwaters of the Yenisey, the Ob or the Irtyš rivers (to judge by toponymic data, see Dul'zon 1959a,b; Maloletko 1992), and that PY was spoken some 2000-2500 years ago (cf. e.g., Werner 2005: 15, Fortescue & Vajda 2022: 238-240, 277).

Hitherto, the reconstruction of Proto-Yeniseian has been mainly pursued with macro-comparative premises in mind. Particularly worthy of mention is the hypothetical “Dene-Yeniseian” language family linking the Old and the New World, cf. e.g., Trombetti (1923: 486, 511), Collins (1954: 35-36), Fortescue (1998), Ruhlen (1998), Vajda (2010a,b, 2019), Fortescue & Vajda (2022); this concept is but an abbreviated version of the much larger “Dene-Caucasian” macro-family (cf. S. Starostin 1982, 1984). Note, however, that the evidence presented so far is considered insufficient to prove beyond doubt the existence of these putative families (cf. Fortescue & Vajda 2022: 244, “an increasing body of comparative linguistic data supports the genealogical unity of Na-Dene and Yeniseian, though the totality of this evidence is still insufficient to conclusively demonstrate Dene-Yeniseian as a proven family”).

In this talk, we present correspondence sets involving word-initial consonants. The data can be used for a systematic application of the comparative method, implying both the reconstruction of proto-phonemes and subsequent phylogenetic research questions (intrafamilial subdivisions among the Yeniseian languages according to shared phonological innovations). We will demonstrate the rigor of the comparative method with a bottom-up approach, focusing here on but one aspect of the recoverable grammatical system of PY, namely phonology. In doing this, we strictly limit our efforts on Yeniseian data alone.¹

Key findings include (I) the discovery of an isogloss which separates Ket, Yugh and Pumpokol from Kott, Assan and Arin in terms of word-initial labial and dental plosives (voiced in the former group, voiceless in the other) and (II) the inference that Proto-Yeniseian probably had a two-layered system of plain voiced and plain voiceless stops. We do not see evidence for the postulation of lateral affricates and aspirated stops (unlike macro-comparatively inspired reconstructions of PY). In addition, there are three correspondences of sibilants and uvulars, respectively, but we cannot, as yet, plausibly posit proto-phonemes in these cases.

¹ Cf. Janhunen's (2020: 166) assessment of previous reconstructive attempts: “The Proto-Yeniseic reconstruction of Sergei Starostin (1982 with later versions) [...] is teleologically oriented towards external comparisons and would need to be redone with a stricter comparative methodology.”

26th International Conference on Historical Linguistics, 4 to 8 September 2023

REFERENCES

- Collins, Henry B. 1954. “*Arctic Area. Indigenous Period*”, (Program of the History of America 1/2). Mexico: Comisión de Historia del Instituto Panamericano de Geografía e Historia.
- Dul’zon, Andrej Petrovič. 1959a. “Geografičeskie nazvaniija Zapodnoj Sibiri kak istočnik eč drevnej istorii [Toponyms of Western Siberia as a source on its ancient history]”, *Nekotorye voprosy drevnej istorii Zapodnoj Sibiri* (Serija “Izučaj svoj kraj“, Tomskij oblastnoj kraevedčeskij muzej 3): 24-28. Tomsk.
- Dul’zon, A. P. 1959b. “Ketskie toponimy Zapodnoj Sibiri [Ket toponyms of Western Siberia]”, *Učēnye zapiski, Tomskij gos. ped. institut* 18: 91-111. Tomsk.
- Fortescue, Michael. 1998. “*Language Relations Across Bering Strait: Reappraising the Archaeological and Linguistic Evidence*”, London, New York: Cassell.
- Fortescue, Michael and Edward Vajda. 2022. “*Mid-Holocene Language Connections between Asia and North America*”, Brill’s Studies in the Indigenous Languages of the Americas, Vol. 17. Leiden-Boston: Brill.
- Georg, Stefan. 2007. “*A Descriptive Grammar of Ket (Yenisei-Ostyak). Part 1: Introduction, Phonology, Morphology*”, Languages of Asia Series, Vol. 1, Folkestone: Global Oriental.
- Janhunen, Juha. 2020. “Towards an Understanding of the Lexical History of the Yeniseic Languages”, *International Journal of Eurasian Linguistics* 2 (1), 166-171.
- Kotorova, Elizaveta and Andrej Nefedov (Eds.). 2015. “*Comprehensive Dictionary of Ket with Russian, German, and English translations*”, 2 volumes, München: Lincom Europa.
- Maloletko, Aleksej Mixajlovič. 1992. “*Paleotoponimika*”, Tomsk: Tomskij gos. universitet.
- Ruhlen, Merritt. 1998. “The Origin of the Na-Dene”, *Proceedings of the National Academy of Sciences* 95: 13, 994-6. Washington, DC.
- Starostin, Georgij Sergeevič. 1995. “Rekonstrukcija praenisejskoj glagol’noj sistemy [Reconstruction of the Proto-Yeniseian verb system]”, *Moskovskij lingvističeskij žurnal* 1: 129-173.
- Starostin, Sergej Anatol’evič. 1982. “Praenisejskaja rekonstrukcija i vnešnie svjazi enisejskix jazykov [Proto-Yeniseian reconstruction and the external relations of the Yeniseian languages]”, *Ketskij sbornik. Antropologija, etnografija, mifologija, lingvistika*: 144-237. Leningrad: Nauka.
- Starostin, S. A. 1984. “Gipoteza o genetičeskix svjazjax sinotibetskix jazykov s enisejskimi i severnokavkazskimi jazykami [The hypothesis of a genetic link between Sino-Tibetan, Yeniseian, and North-Caucasian languages]”, *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka* 4: 19-38. Moskva.
- Trombetti, Alfredo. 1923. “*Elementi di glottologia*”, Bologna: Nicola Zanichelli.
- Vajda, Edward J. 2004. “*Ket*”. Languages of the World/Materials 204. München: Lincom.
- Vajda, E. J. 2010a. “A Siberian link with Na-Dene Languages”, *Anthropological Papers of the University of Alaska*, Volume 5, New Series: 31-99.
- Vajda, E. J. 2010b. “Yeniseian, Na-Dene, and Historical Linguistics”, *Anthropological Papers of the University of Alaska*, Volume 5, New Series: 100-118.

26th International Conference on Historical Linguistics, 4 to 8 September 2023

Vajda, E. J. 2019. “Morphology in Dene-Yeniseian Languages”, Oxford Research Encyclopedia of Linguistics, DOI: <https://doi.org/10.1093/acrefore/9780199384655.013.631>

Verner, Izol'da Georgievna. 1969. “Voprosy imennoj klassifikacii v sovremennyx enisejskix dialektax [Questions of noun classification in modern Yeniseian languages]”, *Voprosy filologii (Učënye zapiski, Omskij gos. ped. institut 52)*: 170-179.

Verner, Genrix Kasparovič [= Werner, Heinrich]. “Opyt rekonstrukcii obščeenisejskoj deklinacionnoj modeli [Reconstructing the Common Yeniseian declension system]”, *Voprosy jazykoznanija 5*: 84-95.

Werner, Heinrich [= Verner, G. K.]. 1997a. “*Das Jugische (Sym-Ketische)*”, Wiesbaden: Harrassowitz.

Werner, H. 1997b. “*Die ketische Sprache*”, Wiesbaden: Harrassowitz.

Werner, H. 1998. “*Abriß der kottischen Grammatik*”, Wiesbaden: Harrassowitz.

Werner, H. 2005. “*Die Jenissej-Sprachen des 18. Jahrhunderts*”, Wiesbaden: Harrassowitz.

Third-person verb inflection in Shakespeare's dramatic texts

This paper investigates the frequencies of the Early Modern English (EModE) verbal suffixes *-(e)th* and *-(e)s* in a corpus of dramatic texts by William Shakespeare. The two forms are allomorphs of the {3rd p. sg.} morpheme, as shown in (1) and (2):

- (1) *Whence cometh this alarum, and the noise?* [1 Henry VI, 1.6]
- (2) *My dearest love, Duncan comes here tonight* [Macbeth, 1.5]

In EModE, the northern suffix *-(e)s* gradually replaced southern *-(e)th* (Barber 1997; Nevalainen 2006). It is likely that the process was partly governed by stylistic factors, with a higher proportion of southern forms retained in formal registers for longer (cf. Görlach 1991; Barber 1997; Cowie 2012). With regard to Shakespeare, Barber (1997) states that *-(e)th* is rarely used in ‘comic or low-life prose scenes’, and Taylor (1972) finds that later plays favour the incoming variant more strongly. Further, the traditional variant correlates with the verbs *do* and *have* (perhaps also *say*) and with stem-final sibilants (as in *surpasseth* or *reacheth*), in which case the affix *-(e)th* functions much like the present-day allomorph [-ɪz].

Using the *Shakespeare First Folio Corpus* hosted by UCREL at Lancaster University, the study takes a multifactorial approach to the alternation of 3rd-person-singular inflectional suffixes in 36 plays by William Shakespeare. A total number of $n = 10,322$ valid tokens is inspected, clustering in $n = 986$ different lemmas. Datapoints were manually coded for PLAY (i.e. name of the play), LEMMA, CATEGORY (comedy, tragedy, history) TIME (i.e. time when the play was finished), FREQUENCY (lemma frequency in a large EModE reference corpus), SIBILANT (i.e. stem-final sibilance) and PATTERN (the syllable structure and stress pattern of the stem), as well as the dependent variable, AFFIX. Variation was then analysed with a mixed-effects binary logistic regression model (with random intercepts for PLAY and LEMMA) using Bayesian estimation as implemented in the R-package *brms* (Bürkner 2021).

As expected, the overall percentage of conservative southern allomorphs is very low. Histories and (to a lesser extent) tragedies are characterised by somewhat higher percentages of *-(e)th*, which suggests that those categories roughly correspond to a more elevated, formal style. There is a surprisingly substantial effect of TIME, with later plays leaning more strongly towards the incoming variant. The FREQUENCY of verb types (or lemmas), on the other hand, hardly plays a role – an effect that was expected based on the exceptional behaviour of high-frequency verbs like *have* and *do*. Finally, a stem-final sibilant makes the selection of the traditional ending considerably more likely.

This paper thus confirms and elaborates several findings concerning a central morphophonemic variable of EModE grammar, based on a unified quantitative analysis. While most results seem plausible, a (self-)critical stance will be taken towards aspects that are at present difficult to measure and quantify. This concerns a more fine-grained analysis of the social dynamics of Shakespeare's plays beyond the rough approximation provided by the three genre categories (comedies, tragedies and histories), as well as the potentially important difference between verse and prose passages (cf. Lass 1999).

References

- Barber, Charles. 1997. *Early Modern English*. Edinburgh: Edinburgh University Press.
- Bürkner, Paul-Christian. 2021. *brms: Bayesian regression models using 'Stan'*. R-package version 2.16.1. <https://cran.r-project.org/web/packages/brms/brms.pdf> [Accessed 2021-09-04]
- Cowie, Claire. 2012. Early Modern English: Morphology. In: Alexander Bergs & Laurel J. Brinton (eds.), *English historical linguistics: An international handbook*. Berlin: De Gruyter Mouton. 604–620.
- Görlach, Manfred. 1991. *Introduction to Early Modern English*. Cambridge: Cambridge University Press.
- Lass, Roger. 1999. *Phonology and morphology*. In: Lass, Roger (ed.), *The Cambridge History of the English Language*. Vol. III. 1476–1776. Cambridge: Cambridge University Press. 23–155.
- Nevalainen, Terttu. 2006. *An Introduction to Early Modern English*. Edinburgh: Edinburgh University Press.
- Samuels, Michael L. 1972 *Linguistic evolution with special reference to English*. Cambridge: Cambridge University Press.
- Shakespeare First Folio Corpus*. <https://cqpweb.lancs.ac.uk/shakff/> [accessed 2022-12-08]
- Taylor, Estelle W. 1976. Shakespeare's use of *eth* and *es* endings of verbs in the First Folio. *CLA Journal* 19(4). 437–457.
- University Centre for Computer Corpus Research on Language (UCREL) at Lancaster University*. <https://ucrel.lancs.ac.uk/> [accessed 2022-12-08]

Learning how to count -- a treebank analysis of V2 word order in two Medieval Romance languages through time

As a central issue in syntactic analysis (Greenberg 1963), the analysis of word order has seen a renewed interest with the development of syntactically annotated treebanks (e.g. Liu 2010). This is equally the case in diachronic research, where both PENN and UD annotation systems have led to significant corpora production. Automatic extraction of word order in annotated datasets supports massive comparison across languages, and through time. Caution is however advisable in that results from automatic extraction can provide misleading results. The research identifies, through the conduct of a novel quantitative analysis on the gradual loss of V2 through time, particular configurations that require separate assessments for the results to be reflective of actual V2 word-order. The take-home message is that quantitative data are most valuable only when their investigations is informed by a qualitative analysis of the phenomenon at hand.

We present the result of a comparative analysis of V2 word order in two Medieval Romance varieties, French and Venetian (e.g. Wolfe 2018 and references therein). The protocol relies on a calibrated corpus to enhance comparability of results. The corpus is calibrated for each language with one text per century over the 14th, 15th and 16th century, at temporal intervals of about a hundred years. They are prose texts belonging to a non-literary genre of legal texts that contain traces of dialogal exchanges, and have been found through preliminary investigations to yield less conservative rates of use of changing variables (Larrivée 2022). The annotated versions of the texts are analysed for position of the finite verb in main and subordinates, using parallel extraction queries from the fine-grained PENN annotation set which is sensitive to phrase-structure. The extraction process however raises two types of methodological questions:

- Some configurations relating to a given word order need a separate assessment;
- Some configurations relating to a given word order should be set aside entirely.

On the first point, early Venetian displays an unexpected pattern by which there seems to be more V2 in subordinates than in main clauses. This makes sense once one realizes that this is due to a nearly categorical use of pre-verbal subjects in embedded clauses, irrespective of V-type (transitives, unergatives, unaccusatives). On the contrary, in main clauses we find that subject can be both pre and post-verbal, in between the auxiliary and the past participle, thereby attesting the expected asymmetry between main and embedded clauses typical of a V2 language.

The second point is illustrated by the surprising frequency of V1 word order. As both Venetian and French are expected to go from a V2 system to a SVO word order, the high proportions of V1 is troubling. Again, a qualitative examination of the data shows that the surprising proportions are due to two configurations that should be set aside: (i) coordinated subjectless clauses inside a sentence; (ii) relative clauses.

We conclude that: (a) despite the apparent prevalence of V2 in embedded clauses and V1 in main clauses, Old Venetian is still a totally regular V2 language in the early XIV c., with a clear asymmetry in the subject/verb position between main and embedded clauses; (b) The assessment of default assertive word order requires methodological and analytical decisions about what to count, and what not to.

Time permitting, we will also focus on the diachronic pattern of loss of the V2 property in the two languages. The refined data allow us to better quantify the rate of V2 and its diachronic demise.

References

- Greenberg, J. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- Larrivé, Pierre. 2022. « Is Medieval French diglossic? New evidence on remnant V2 and register ». Ora Matushansky, Laurent Roussarie, Michela Russo, Elena Soare, Sophie Wauquier (Eds), *Romance Languages and Linguistic Theory (RLLT series: 17)*, *Isogloss* 8,2, 1-16.
- Liu, Haitao. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120,6, 1567-1578.
- Wolfe, Sam. 2018. *Verb Second in Medieval Romance*. Oxford: Oxford University Press.

When ‘still’ comes to signal a near past

In unrelated and geographically non-adjacent languages we find uses of expressions meaning ‘still’ as signals of a near past; the examples in (1, 2) are illustrations.

- (1) Western Dani (Barclay 2008: 440)

At awo wa-gaarak.
3SG still come-PST.3SG
‘He has just come.’

- (2) Gitsxan (Anouki 2021: 69)

K’ay hlis bax=hl gimxdi-’y win ’witxw ha ’w-i’y ky’oots.
still finish run=CONN sister-POSS.1SG SUBORD arrive go_home-1SG yesterday
‘My sister had just finished running when I came home yesterday.’

In my talk, I address such uses in seven languages from four continents: Bende (Bantu), Datooga (Nilotic), Gitsxan/Nisga’a (Tsimshian), Kekchí (Maya), Tunisian Arabic (Afro-Asiatic), and Western Dani (Trans New Guinea).

I first take a synchronic perspective and show that –as far as can be judged from the available data– the relevant constructions share several semanto-pragmatic characteristics. First, the notion of proximity they signal can relate to intervals other than the time of speech, as can be observed in (2), and in (3) below. That is, we are not dealing with tense in the sense of an ordering between topic time and utterance (Klein 1994). Secondly, the aspectual viewpoint is fully contained in the situation’s post-time (anterior aspect a.k.a. “perfect”). Closely related, in coherent discourse ‘still’-as-near-past is consistently found in backgrounded clauses, such as in (3).

- (3) Tunisian Arabic (Afro-Asiatic, Singer 1984: 651)

Kun-t ānā māzil-t kīf bdī-t n-umgud fī tarf
COP.PFV-1SG 1SG still-1SG when/how begin.PFV-1SG 1SG-chew.IPFV in piece
il-lham haḍāya u-zarṣt-i rā-hi
ART.DEF-meat(M) PROX.SG.M and-molar(F)-POSS.1SG PRESTT-3SG.F
tnaṭr-it tanṭīra waḥd-a.
slip_out.PFV-3SG.F slip_out.NMLZ(F) one-SG.F
‘I had just begun chewing on the piece of meat when all of the sudden my molar tooth came flying out.’

Elaborating on this comparison, I discuss the etymologies of the expressions involved and sketch out two major diachronic pathways leading from ‘still’ plus anterior aspect to a near past. In the first scenario, the notion of persistence (< ‘still’) is projected from the runtime of a situation itself onto the post-time portion (< anterior aspect) of the time span characterized by its occurrence: ‘still in the post-arrival period’ > ‘have just arrived’. In the second scenario, first proposed by Anouki (2021) for Gitsxan *k’ay*, the link between the two uses lies in a left-adjacent runtime of the situation: ‘the (now completed) arrival has taken until now’ > ‘have just arrived’. A variation of this theme, in the form of a more indirect link, is found in Kekchí, where *toj* as a near past signal can be traced back to a restrictive ‘not until’ function of the same item. Both, the restrictive use and *toj* as ‘still’ likely share a common ancestor in delimitative ‘until’ (cf. Kockelman 2020).

In throwing a comparative and diachronic light on this hitherto understudied phenomenon, my talk thus contributes to our understanding of the multifarious histories of “phasal polarity” (van Baar 1997) expressions.

References

- van Baar, Theodorus M. 1997. *Phasal polarity*. Dordrecht: Foris.
- Aonuki, Yurika. 2021. Unifying the ‘still’ and ‘just’ readings of k’ay in Gitksan. In Daniel K. E. Reisinger & Marianne Huijsmans (eds.), *Proceedings of the 37th West Coast Conference on Formal Linguistics*, 68–78. Somerville: Cascadilla Proceedings Project.
- Barclay, Peter. 2008. *A grammar of Western Dani*. Munich: Lincom Europa.
- Klein, Wolfgang. 1994. *Time in language*. London: Routledge.
- Kockelman, Paul. 2020. Dual operators and their doubling, in Q’eqchi’ (Maya). *International Journal of American Linguistics* 86. 449–484.
- Singer, Hans R. 1984. *Grammatik der arabischen Mundart der Medina von Tunis*. Berlin: de Gruyter.

***ille ego* and Recognitional Use of Demonstratives**

26th International Conference on Historical Linguistics at the University of Heidelberg

Abstract

Demonstrative pronouns in all languages can be divided into proximal, medial and distal categories (with “medial” being optional sometimes) based on distance from speaker. On the other hand, three types of deixis can be distinguished: person (*I, you & (s)he*), place (*here & there*) and time deixis (*now & then*). And it is self-evident that the referent of distal demonstratives is distant from the speaker or deictic centre (“*I*” in a conversation). However, certain “mismatched” expressions do exist, e.g. Latin *ille ego* ‘I am that one/That I’ (cf. Thesaurus Linguae Latinae s.v. “2. EGO” p.275–276); Greek ὄδ’ ἐκεῖνος ἐγώ· ‘here that one am I’ (Sophocles Oedipus Coloneus line 138); Vedic Sanskrit *ahám só asmi* ‘I am he’ (Rigveda 1.105.7) and *só asmi* ‘that one am I’ (Atharvaveda 6.123.3), Classical Sanskrit *asāv aham* ‘that one am I’ (Bhāgavata-Purāṇa 10.85.17); German *Ich bin derjenige* ‘I am that one’ (Samuel Lutz in 1736). These examples point to the recognitional use of demonstratives, because all the other pragmatic functions, i.e. exophoric, anaphoric and discourse deictic, can be excluded (Diessel 1999:93–105). But according to Diessel (1999:93), “recognitional use is restricted to adnominal demonstratives”, which is clearly not the case in the examples above. Therefore, this paper first offers a philological and comparative study of the “*ille ego*”-type sentences in Latin, Greek and Sanskrit materials, and then tries to contextualise this rare but real usage of demonstratives.

References

- Diessel, Holger. 1999. *Demonstratives. Form, function, and grammaticalization*. Amsterdam: Benjamins.
- Goedegebuure, Petra. 2014. *The Hittite demonstratives. Studies in deixis, topics and focus*. Wiesbaden: Harrassowitz
- Pinkster, Harm. 2015. *The Oxford Latin syntax 1: The simple clause*. Oxford: Oxford University Press.
- Jamison, Stephanie. 1992. Vedic “*sá figé*”: An inherited sentence connective? *Historische Sprachforschung* 105 (2): 213–239.

Separate clause source and initial-to-medial pathway: Formation of Chinese epistemic adverbial and sentence connective *chéngrán*

Modern Chinese *chéngrán* may be used as an epistemic adverbial (shortened as an EA) meaning ‘no doubt’ (see (1a)) and a sentence connective (shortened as a SC; see (1b)).

- (1) a *Tā hěn ài nà jǐ zhī xiǎoyā, xiǎoyā yě chéngrán kě'ài.*
 he very love DEM several CLS little.duck little.duck also no.doubt lovely
 ‘He loves those little ducks, and they are **no doubt** lovely.’

- b *Chéngrán huàjù yǒu yúlè de gōngnéng, dànshì tā bù yīng línwéi*
 although drama have recreation NOMZ function but it NEG should reduce.to

fàn hòu tiándiǎn.
 dinner after sweet

‘The drama has its recreational functions, **but** it should not be reduced to the functions equivalent to dinners after dinners.’

Diachronic investigations reveal that both an EA *chéngrán* and a SC *chéngrán* developed from a separated clause *chéngrán* ‘(it is) quite right’ that appeared in the initial position of another clause in Ancient Chinese.

- (xx) *Jiēyú yuē: “Wú bú xǔ yě.” Qī yuē: “Chéngrán, bú rú qù zhī.”*
 Jieyu say I NEG agree.to FP wife say quite.right. NEG match leave DEM
 ‘Jieyu said: I don’t agree to it. His wife said: **(It is) quite right**. We had better leave the place.’
 (*Tàipíng Yùlǎn*, 983 CE)

A separate clause *chéngrán* had both an epistemic function and a linking-clause function (see (2)). Its epistemic function followed the hypothetical initial-to-medial pathway argued by Long et al. (2022), and developed into a clause-medial EA (see (3)).

- (3) a *Gài Yì zhī shū, chéngrán shì jié jìng jīng wēi.*
 generally Book.of.Changes NOMZ book no.doubt COP clean clear exquisite subtle
 ‘Generally the *Book of Changes* is **no doubt** a book of cleanness, clearness, exquisiteness, and subtlety.’ (*Zhūzǐ Yùlèi*, 1270 CE)

Its linking-clause function developed into a conventionalized SC, and following the hypothetical initial-to-medial pathway, it may also be used in a clause-medial position (see (4)).

- (4) *Zhè zhuāng dōngxī chéngrán bù kě shīluò, dàn yǎnxià wǒmenzhè yī qún*
 DEM CLS thing although NEG may lose but currently we DEM one group

rén duànduàn méi gè huíqù de lǐ.
 people absolutely have.not CLS return NOMZ reason

‘We can’t lose this thing, **but** currently we have such a group of people, and it makes absolutely no sense for us to return to the site.’ (*Érnǚ Yīngxióng Zhuàn*, early 19th century)

This study endeavors to establish a hypothetical source construction of separate clauses for the formation of some SCs that is largely neglected by Traugott (2022) and others, and further explains why some SCs may occupy a clause-medial position in the other languages; see English clause-medial SCs *however* in (5a) and *therefore* in (5b).

- (5) a ... A.H.Q. Malta confirmed that one Hurricane had been slightly damaged. This **however** would appear to have been in combat with Bf110s ... (BNC)
- b This necessarily entails longer term assistance in comparatively stable situations. We **therefore** particularly value our partnership with SCF through TRANSAID... (BNC)

References:

- Long, Haiping, Francesco Ursini, Bernd Heine, and Yaohua Luo. 2022. Grammatical changes of Modern Chinese sentence adverbial *guoran* ‘it really happens’: A hypothesized external-to-medial pathway. *Australian Journal of Linguistics* 42(3): 1–25.
- Traugott, Elizabeth Closs. 2022. *Discourse Structuring Markers in English*. Amsterdam/Philadelphia: Benjamins.

Middle Polish adverb-like predicates ending in *-a* compared to other adverbial and adjectival predicates – corpus-based approach

In Polish, the characteristic form of the predicate in sentences with an infinitival or clausal subject is the form of an adverb or an adjective in a neuter gender (e.g. *Niebezpiecznie było tam iść*. 'It was dangerous [ADV] to go there'; *To dziwne, że wrócił*. 'It's strange [ADJ.N] that he came back'). In the Middle Polish language (16th-18th centuries) this image was more complicated. Firstly, there was a large set of adverbial derivatives with two variants: ending in *-o* and in *-e* (e.g. *niebezpieczno* – *niebezpiecznie* 'dangerously', *dziwno* – *dziwnie* 'strangely'). Originally, in the predicative function there were adverbs ending in *-o*, but adverbs ending in *-e* were also used by analogy.

Secondly, in constructions of this type, the feminine form of the adjective was also used as a predicate (e.g. *niebezpieczna* 'dangerous', *dziwna* 'strange'). This resulted from the simplification of the nominal group containing a feminine noun *rzecz* 'thing' and an adjective in the feminine form demanded by a noun. After eliminating the semantically empty noun, the adjectival forms took over the function of the predicate (*Niebezpieczna rzecz tam iść*. 'It's a dangerous thing to go there.' → *Niebezpieczna tam iść*. 'It's dangerous [ADJ.F] to go there.'). They also began to undergo a process of adverbialization, which, however, did not fully occur (hereinafter I refer to them as "adverb-like predicates ending in *-a*").

The study whose results will be presented aims to show the functioning of adverb-like predicates ending in *-a* among other adverbial and adjectival predicates. The data for the analysis come from the Electronic Corpus of 17th- and 18th-century Polish Texts, a 25M corpus annotated morphosyntactically, collecting texts of various themes, genres and styles (<https://korba.edu.pl>). For the purposes of the study, a dozen or so predicates ending in *-a* with a high frequency in the corpus, belonging to various semantic groups and having different syntactic requirements (connecting with an infinitival or clausal subject) were selected. Each of these predicates has been juxtaposed with synonymous adverbial predicates in *-o* and in *-e*, and with a predicate in the form of a neuter adjective (e.g. *dziwna* [ADJ.F] – *dziwno* [ADV] – *dziwnie* [ADV] – *dziwne* [ADJ.N]).

The study is both quantitative (shows the frequency of using predicates ending in *-a* in texts in comparison to other types of predicates) and qualitative (captures the differences in meaning between particular types of constructions). Although constructions containing particular types of predicates seem to be fully synonymous, it can be assumed that there were some factors determining the choice of one of them, e.g. the style of the text in which the given construction was used.

The large time range of the corpus makes it possible to trace the changes that the relations between the particular types of the discussed construction have undergone over the course of two centuries. The data obtained from the corpus also allow us to speculate on the reasons for the displacement of constructions with adverb-like forms ending in *-a* by other types of predicates.

Keywords: historical syntax, corpus research, adjectives, adverbs

References

Bronikowska R., 2021. Unfinished "verbization" process – the development of predicative constructions with an adjective of the feminine gender in the 17th and 18th centuries in the light of corpus data, "Polonica", 41(1), pp. 97-110. <https://doi.org/10.17651/POLON.41.7>

- Cyran W., 1967. Przymiotniki polskie. Budowa słowotwórcza [Polish adverbs. Word formation structure], Łódź: Łódzkie Towarzystwo Naukowe.
- Gruszczyński W., Adamiec D., Bronikowska R., Kieraś W., Modrzejewski E., Wieczorek A., Woliński M., 2022. The Electronic Corpus of 17th- and 18th-century Polish Texts, "Language Resources and Evaluation", 56, pp. 309–332. <https://link.springer.com/article/10.1007/s10579-021-09549-1>
- Kałkowska A., 1973. Składniowe właściwości przymiotników kwalifikujących [Syntactic properties of qualifying adjectives], "Język Polski", 2, pp. 95-104.
- Przepiórkowski A., Patejuk, A., 2020. Predicative Adverbs and Adjectives with Infinitival Subjects: A Corpus Investigation, "Studies in Polish Linguistics", 15(3), pp. 129-150. <https://doi.org/10.4467/23005920SPL.20.006.12978>
- Przepiórkowski A., Patejuk A., 2021. Predicative Adverbs: Evidence from Polish, "Linguistic Inquiry", 52(4), pp. 835–851. https://doi.org/10.1162/ling_a_00394
- Przepiórkowski A., Patejuk, A., 2018. Predicative Constructions with Infinitival and Clausal Subjects in Polish. In: Miriam Butt and Tracy Holloway King (eds.), The Proceedings of the LFG'18 Conference, Stanford, CA: CSLI Publications, pp. 304–324. <https://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2018/lfg2018-patejuk-przepiorkowski.pdf>
- Sobotka P., 2018. Powstawanie słowiańskich przysłówków w świetle danych etymologicznych i historycznoskładniowych (interpretacja funkcjonalna) [Formation of Slavic adverbs in the light of etymological and historical syntactic data (functional interpretation)], "Z Polskich Studiów Sławistycznych", 13(2), pp. 315-324.
- Sobotka P., 2019. Zmiany zależności składniowych północnosłowiańskich przymiotników odmiany prostej [The dependency changes of the North-Slavic simple declension adjectives], "Prace Filologiczne" LXXII, pp. 333–344.
- Szober S., 1931. Skąd powstała przysłówkowa forma orzecznika w zdaniach z bezokolicznikiem w podmiocie? [Where did the adverbial form of the predicate come from in sentences with the infinitive in the subject?], "Język Polski" XVI(3), pp. 85-87.
- Śmiech W., 1957. O polskich przysłówkach odprzymiotnikowych na -o, -e [About Polish adjective adverbs in -o, -e], "Rozprawy Komisji Językowej Łódzkiego Towarzystwa Naukowego" V, pp. 61-76.

Internal subgrouping of Northern Naga based on Bayesian phylogenetic analysis

Kellen Parker van Dam - Universität Zürich, La Trobe University

This paper presents the results of a large-scale study of the internal subgrouping of Northern Naga using Bayesian inference on lexical data sets. Data are analysed from approximately 140 central Sal doculects covering approximately 750 concepts. Concepts cover many of the most stable etyma (Matisoff 2009) as well as many culturally-relevant concepts such as those described in Matisoff (1978). Further concepts were incorporated based on frequency of inclusion in the literature.

The data used in the study include both published and unpublished wordlists as well as newly elicited lexical data. Efforts were made to include all extant published source material on the languages which provided sufficient lexical data, including descriptions from as far back as the early 19th century. Lexemes were then hand-coded for cognacy based on regular sound correspondences, as determined by newly done historical reconstructions at various levels in the family. As a lack of archaic written data precludes estimations of time depth, with the oldest sources still too recent for proper clock calibration, analysis was done in MRBAYE (Huelsenbeck et al 2001) using Markov Chain Monte Carlo sampling.

The resulting phylogeny confirms a clear north-south split in the family (van Dam, forthcoming), with the northern branch corresponding directly to the Tangsa-Nocte subgroup, and an additional primary split among the southern branch roughly corresponding to the political boundary between India and Myanmar. Lower level subgrouping also confirms many of the previous judgements on subgrouping found in the literature based on both lexical and non-lexical features (Morey 2015; van Dam 2018).

Keywords

Bayesian phylogeny, Tibeto-Burman, Sal, Northern Naga

References

- Huelsenbeck, J.P., and F. Ronquist. 2001. *MRBAYES: Bayesian inference of phylogeny*. *Bioinformatics* 17:754-755.
- Matisoff, J.A., 2009. *Stable roots in Sino-Tibetan/Tibeto-Burman*. *Senri ethnological studies*, 75, pp.291-318.

— 1978. *Variational Semantics in Tibeto-Burman: The "organic" Approach to Linguistic Comparison*. Institute for the Study of Human Issue : Philadelphia. (Wolfenden Society on Tibeto-Burman Linguistics. Occasional papers, v. 6).

Morey, S., 2015. *The internal diversity of Tangsa: vocabulary and morphosyntax*. Language and culture in Northeast India and beyond, p.23.

van Dam, K.P., 2018. *The tone system of Tangsa-Nocte and related Northern Naga varieties*. PhD thesis. La Trobe University

A morphological freeloader: Ibero-Romance *caber*

Modern Spanish *caber* ‘fit, be containable in’ stands out for its almost unique kinds of allomorphy, mainly apparent in the lexical root. Relatively regular verbs such as *barrer* ‘sweep’ or *recibir* ‘receive’, have invariant lexical roots and their future and conditional tense forms that share a thematic vowel with the infinitive (from which the future and conditional are historically derived). In contrast, *caber* has the allomorph *cab-* only in the present indicative and imperfect indicative, the future and the conditional, the infinitive, and past participle; but it has *quep-* throughout the present subjunctive and in the first person singular present indicative, and *cup-* in the preterite and in both series of imperfect subjunctives; moreover, unlike most verbs, its future and conditional (*cabrá*, *cabría*) do not share a thematic vowel with the infinitive (*caber*).

Historically, the allomorphy seen in *caber* is unremarkable, presenting the expected effects of certain regular early Romance phonological and morphological changes. What is remarkable is that this allomorphy has survived intact when other inherited allomorphy of the relevant kinds was analogically eliminated during the Middle Ages (cf. Maiden 2018:50-53; Rini 2020a,b). Modern Spanish retains only about a dozen really irregular verbs and these are semantically basic, very high frequency, lexemes, such as ‘be’, ‘have’ (auxiliary), ‘have’, ‘come’, ‘go’, ‘say’, ‘do’, ‘know’, ‘want’, ‘can’, ‘give’, ‘bring’, ‘put’ and/or are the basis of a larger series of derived verbs. *Caber* is a ‘stow-away’ in this company: it has significantly lower frequency than the other irregular verbs, and is not part of any derivational ‘family’. It should, therefore, have lost its allomorphy to the otherwise general tendency for ‘levelling’ at some time in the Middle Ages.

The historical morphology of *caber* has been the subject of recent studies by Rini (2020a,b), on which I draw here, while also critically adjusting his analysis. Adapting and extending Rini’s idea of ‘analogical retention’ (also Rini 2001), which he applied to only part of the data for *caber* (Rini 2020a:744f.; 2020b:120-122), I argue that the survival of all the irregularities in *caber* is wholly and strictly dependent on the model of the very basic, high frequency, verb *saber* ‘know’, to which *caber* happens to have emerged in early Romance as inflexionally identical in every detail except for the initial consonant and for the the root of the first person singular present indicative form (*quepo* vs *sé*). Thus *saber* has *sab-* in the present indicative and imperfect indicative, the future and the conditional, the infinitive, and past participle, but *sep-* throughout the present subjunctive and in the first person singular present indicative, and *sup-* in the preterite and in both series of imperfect subjunctives; unlike most verbs, its future and conditional (*sabr , sabr a*) do not display the thematic vowel of the infinitive (*caber*).

I show in detail how—not only in the history of Spanish, but in other Ibero-Romance dialects—*caber* has been repeatedly and minutely sensitive to the morphology of *saber* (except in the 1SG present indicative, where the two verbs have been different *ab antiquo* and have correspondingly diverged chaotically).

This ‘parasitic’ diachronic behaviour of *caber* will be argued to be a matter of pure morphology which cannot be explained in functional terms: *caber* and *saber* are utterly different in lexical meaning, argument structure, and syntactic frame. The paradigmatic distributional pattern involved is, equally, arbitrary and idiosyncratic. I argue that the observed diachronic behaviour of these verbs presupposes speakers’ ability to abstract paradigmatic distributional patterns of allomorphy from the inflexional paradigms of individual lexemes even when that allomorphy is idiosyncratically and almost uniquely associated with a particular lexical meaning. I shall compare the significance of such diachronic data to other ‘morphomic’ patterns, such as the English *-ceive* - *-ception* series cited by Aronoff (1994), or the idiosyncratic allomorphy of semantically disparate Spanish verbs in *-ducir*, and assess the relevance of such facts for Maiden’s recent discussion (Maiden 2021) of the minimum conditions necessary for the diachronic emergence of ‘morphomic’ patterns.

References

Aronoff, Mark (1994). *Morphology By Itself*. Cambridge MA: MIT Press.

Maiden, Martin (2018). *The Romance Verb. Morphomic Dstructure and Diachrony*. Oxford: OUP.

Maiden, Martin (2021). ‘The morphome’. *Annual Review of Linguistics* 7:89-108.

Rini, Joel (2001). ‘The extraordinary survival of Spanish “veía”’: another facet of analogy revealed’. *Hispanic Review* 69:501–525.

Rini, Joel (2020a) ‘Spanish *quepo*: the untold story’, *Zeitschrift für romanische Philologie* 136:730–748

Rini, Joel (2020b). ‘A morphological factor in the history of the irregular future (and conditional) of Spanish’. *Studia Neophilologica* 92:111–123.

**From inference to hearsay: the development of the French parentheticals
à ce qu'il paraît, comme il paraît, il paraît, paraît-il**

Amalia Rodríguez-Somolinos
Universidad Complutense de Madrid

The construction *il paraît que* appeared around 1650 with a meaning 'it is obvious that'. It expressed an inference based either on visual perception or on reasoning. We will call it *il paraît que1*. In this first stage of the evolution, *il paraît que1* made an intersubjective reading possible. It concerned knowledge in general; it is an inference that anybody could have established. In Classical French and until the 20th century, *il paraît que* evolved towards greater subjectivity. It was speaker-oriented, which means it was more related to the knowledge of the speaker in particular. We will refer to this use as *il paraît que2*. It expresses inference by reasoning, based on the speaker's knowledge or on clues that are not always easily specified. In this respect, *il paraît que2* is semantically very close to Modern French *il semble que* 'it seems that' 'it would appear that'. *il paraît que2* functioned as a downtoner: it often conveyed doubt or uncertainty. In the course of the 19th century *il paraît que2*, which conveyed inference from reasoning, evolved towards *il paraît que3*, used as a hearsay evidential marker. This is the modern use, which has coexisted with the previous inferential use throughout the 19th century, up until the beginning of the 20th century. From 1925 on, *il paraît que* has only been used as a reportive evidential.

The structure *il paraît que p* has had, since the 18th century, several parenthetical variants, which can occur in initial or final position. The oldest, *à ce qu'il paraît*, is first attested in Frantext in 1755. The parenthetical can be translated in Modern French as *semble-t-il*, 'seemingly', 'it seems'. It is equivalent to *il paraît que2*, which was in use at that time. The assertion *p*, which is in the scope of *à ce qu'il paraît*, conveys the speaker's personal opinion. It expresses an analysis based on reasoning, although the speaker does not completely adhere to his own conclusion.

The remaining parenthetical variants date from the 19th century. According to Frantext, *il paraît* appeared in 1840, and *comme il paraît* appeared in 1854. In the beginning, they both had the meaning of 'so it seems' and were equivalent to *il paraît que2*. Finally, *paraît-il* is attested from 1859 and is used as a hearsay evidential. It is equivalent to *il paraît que3*, which was already in use at that time.

In Contemporary French, *il paraît que p* has two parenthetical variants: *il paraît*, *paraît-il*. Both are mostly reportive evidentials with a meaning 'apparently' 'from what I hear'. Besides, *il paraît* has an inferential use as a syntactically independent answer that is very frequent, with a meaning 'so it seems'. By analogy, in Contemporary French, the parentheticals *paraît*, *il paraît* have recently developed an inferential use in final position, with a meaning similar to French *on dirait* (*so it seems, it would seem so*).

From this overview, we see that, historically, the parenthetical variants are very close to *il paraît que*. The parentheticals *à ce qu'il paraît*, *il paraît*, *comme il paraît*, *paraît-il* follow the semantic evolution of the construction with a *que*-clause in parallel. They have evolved from an inferential to a reportive use. The only exception would be the recent inferential use of the parentheticals *il paraît*, *paraît* in final position.

References

- Dendale, Patrick & van Bogaert, Julie. 2007. A semantic description of French lexical evidential markers and the classification of evidentials. *Rivista di Linguistica* 19:1.65-89.
- López-Couso, María José and Méndez-Naya, Belén. 2014. From clause to pragmatic marker. A

- study of the development of like-parentheticals in American English, *Journal of Historical Pragmatics* 15:1, 36–61.
- Nuyts, Jan. 2001. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33:3.383-400.
- Nuyts, Jan. 2012. Notions of (inter)subjectivity. *English Text Construction* 5. 53-76.
- Pietrandrea, Paola. 2007. The grammatical nature of some epistemic-evidential adverbs in spoken Italian. *Rivista di Linguistica* 19:1.39-64.
- Rossari, Corinne. 2012. Valeur évidentielle et/ou modale de *faut croire, on dirait* et *paraît*. *Langue française*, 173, p. 65-82.
- Squartini, Mario. 2004. Disentangling evidentiality and epistemic modality in Romance. *Lingua* 114.873-895.
- Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* 57,33-65.
- Urmson, J. O. 1952. “Parenthetical verbs”, *Mind* 61, 480-496.
- Schneider, Stefan Glikman, Julie Avanzi, Mathieu eds. 2015. *Parenthetical verbs*. De Gruyter.
- Willems, Dominique. & Blanche-Benveniste, Claire. 2010. Verbes ‘faibles’ et verbes à valeur épistémique en français parlé: *il me semble, il paraît, j’ai l’impression, on dirait, je dirais*. *Actes du XXVe Congrès International de Linguistique et de Philologie Romane* ed. by Maria Iliescu, Heidi Siller-Runggaldier & Paul Danler, 565-576. Berlin & New York: De Gruyter, vol. IV.

Structural ambiguity and reanalysis – the case of Swedish *fortsatt*

In Swedish, the past participle *fortsatt*, derived from the verb *fortsätta* 'to continue', is often used as a clause adverbial with the meaning 'still'.

- (1) Situationen i Göteborgs Hamn är **fortsatt** inte bra. (GP 2009)
 'The situation in Gothenburg's harbour is still not good.'

This usage of *fortsatt* has developed during the 20th century, and can therefore be studied in great detail in electronic corpora. A careful investigation of this change can also contribute to a general understanding of the conditions for structural ambiguity, which is a significant factor in language change. This discussion is not new – Wurzel (1997) refers to Hermann Paul and Georg von der Gabelentz – but it is still dynamic and relevant. Several linguists claim that structural ambiguity is a prerequisite for reanalysis:

To summarize, the conditions necessary for reanalysis to take place are that a subset of the tokens of a particular constructional type must be open to the possibility of multiple structural analyses, where one potential analysis is the old one (applicable to all tokens) and the other potential analysis is the new one (applicable to a subset). (Harris & Campbell 1995:72)

Roughly, reanalysis can be defined as a process that changes the actual (underlying) linguistic structure without necessarily affecting the visible or audible surface manifestation of that structure. So it presupposes ambiguity in linguistic structures [...]. (Burrige & Bergs 2017:108)

See also Hopper & Traugott (2003), Fischer (2007), Denison (2017), Weiß (2021), and many others. However, while structural ambiguity often is recognised as an absolute requirement for reanalysis, the question of whether the frequency of ambiguous constructions may be relevant is rarely – if ever – mentioned.

In this talk, I present a detailed study of the development of *fortsatt*. I show that the change proceeds in two separate reanalyses. In order to evaluate whether the relative frequency of ambiguous structures have any effect on the conditions for change, I introduce a quantitative measure of ambiguity (*ambiguity index* or *ambix*). It is demonstrated that both of the reanalyses of *fortsatt* take off when about 80 % of the respective source constructions are structurally ambiguous. The conclusion is that the changes of *fortsatt* are facilitated by a large amount of structural ambiguity in the investigated texts, and that the ambiguity index drops when the new constructions are introduced and established. Testing whether this is a general tendency in this type of language change, I also utilize a recent study by Delsing (2022), in which he investigates two similar reanalyses (the changes of Swedish *mycket* and *litet* from adjectives to quantifiers). It is shown that these changes seem to corroborate the conclusion that the relative frequency of ambiguous structures is of vital importance in the initial stages of structural reanalysis, and that the establishment of the new structures coincides with a gradual drop of the ambiguity index.

In the light of these findings, the question of the relation between structural ambiguity and reanalysis is further discussed.

Literature

Burridge, Kathryn & Bergs, Alexander. 2017. *Understanding language change*. New York: Routledge.

Denison, David. 2017. Ambiguity and vagueness in historical change. In: Hundt, Marianne, Mollin, Sandra & Pfenninger, Simone E. (ed.): *The Changing English Language*. Cambridge: Cambridge University Press. P. 292–318.

Heine, Bernd. 2002. On the role of context in grammaticalization. In: Wischer, Ilse & Diewald, Gabriele (ed.): *New reflections on grammaticalization*. Amsterdam: John Benjamins. P. 83–101.

Nevalainen, Terttu. 2015. Descriptive adequacy of the S-curve model in diachronic studies of language change. *Studies in Variation, Contacts and Change in English* 16. *Can We Predict Linguistic Change?* Edited by Christina Sanchez-Stockhammer <https://varieng.helsinki.fi/series/volumes/16/index.html>

Delsing, Lars-Olof. 2022. From ‘big’ to ‘much’: On the grammaticalization of two gradable adjectives in Swedish. In: Larsson, Ida & Petzell, Erik M. (ed.): *Morphosyntactic change in Late Modern Swedish*. Berlin: Language Science Press. P. 319–337.

Fischer, Olga. 2007. *Morphosyntactic change*. Oxford: Oxford University Press.

Harris, Alice C., & Campbell, Lyle. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Hopper, Paul J. & Traugott, Elizabeth Closs. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Weiß, Helmut. 2021. Reanalysis involving rebracketing and relabeling: a special type. *Journal of Historical Syntax* 5:1–26. <https://ojs.ub.uni-konstanz.de/hs/index.php/hs>

Wurzel, Wolfgang Ullrich. 1997. Grammatical ambiguity and language change. In: Hickey, Raymond & Puppel, Stanislav (red.): *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak on his 60th Birthday*, Berlin, New York: De Gruyter Mouton. S. 1125–1137.

Elektronisk corpora

Kungliga biblioteket, Svenska dagstidningar: <https://tidningar.kb.se/>

Språkbanken, Korp: <https://spraakbanken.gu.se/korp/#?corpus=&cqp=%5B%5D>.

The functional interpretation of semantic and syntactic shifts in the domain of North Slavic “conversive” preposition-pronominal constructions

Following a dynamic and functional approach (cf. Givón 2015), this paper presents the development of grammatical functions of preposition-pronominal constructions (PPCs) in North Slavic, cf. preposition+pronoun as in Pol. *przy tym* ‘lit. by this; at the same time; moreover’, *przy czym* ‘lit. by what; at the same time’. The study of this topic in Slavic languages has been so far patchy (Rysová 2017, Danlos, Rysová, Rysová & Stede 2018, Kisiel & Sobotka 2022) and has not investigated the difference in diachronic progress from one base pattern toward the state observed in the modern languages. The talk intends to fill in this gap by delivering a functional explanation for the diverse results of the grammaticalization based on the same underlying PPC. Particularly, the paper targets two-element symmetric (preposition + interrogative pronoun vs demonstrative pronoun) grammaticalized structures across grammatical classes (adverbs, relatives, conjunctions, connectors, discourse markers) to interpret differences between grammaticalization paths of PPCs with *conversive pronouns* ‘this’ and ‘what’, e.g. Rus. *potomu* ‘lit. after this; that’s why; because’ vs *počemu* ‘lit. after what; for what reason; why; so’, Pol. *dlatego* ‘lit. for this; that’s why; therefore’ vs *dlaczego* ‘lit. for what; why’, Cz. *přesto* ‘lit. through this; even though; nevertheless’ vs *přes co* ‘lit. through what; what’.

The first part of the paper presents the grammaticalization chain of PPC with an interrogative pronoun, see ORus. (1a-b). This chain is shorter and more predictable than the chain with a demonstrative pronoun, see ORus. (2a-b). It seems also shared by most North Slavic languages: the path of Polish *dlaczego* ‘lit. for what’ or Upper Sorbian *čehodla* ‘lit. what for’ resembles the one of Russian *počemu* ‘lit. for what’ or *začem* ‘lit. beyond what’, regardless the difference in the prepositional element.

- (1) a. *po čemu nareklъ jestъ xristosъ*
PREP INTER call.PST.3SG be.AUX.PRS.3SG Christ.NP
‘For what reason is he called *Christ*’. (ŽivAndJur)
- b. *počemu že ty nazyvaešъ tu zemlju*
INTER PTCL 2SG call.PRS.2SG DEM land.ACC.SG
‘Why are you the one who calls this land?’ (Arx.Str. I 228)
- (2) a. *kažetsja potomu, i žalъ emu menja*
it seems CONN CONJ pity.ACC.SG 3SG.DAT 1SG.GEN
‘It seems that this is why he feels sorry for me.’ (Av.Ž. 52)
- b. *A starca obvinilъ, potomuperedъ
CONJ old.GEN.SG accuse.PST.3SG CONJ PREP
knjazemъ na srokъ ne stalъ.
prince.INS.SG PREP time.ACC.SG NEG stand.PST.3SG
‘and he blamed the old man because he did not appear in time before
the prince.’ (Arx.Str. I 48)*

The second part of the paper focuses on patterns and conditions of change in PPCs with demonstrative pronouns, which show a greater variety in East and West Slavic languages, see e.g. Cz. discourse marker *nadto* ‘moreover’ vs Rus. unlexicalized *nad to* ‘over that’ etc. Also, the functions of PPCs with demonstrative pronouns display a high level of variation between the languages, e.g. Pol. *potem* as an adverb vs Cz. *potom* as an adverb, conjunction, connector, and discourse marker.

Reference

- Danlos, Laurence, Kateřina Rysová, Magdaléna Rysová & Manfred Stede. 2018. Primary and Secondary Discourse Connectives: Definitions and Lexicons. *Dialogue & Discourse* 9(1). 50–78.
- Givón, Talmy. 2015. *The diachrony of grammar*, vol. 1–2. Amsterdam–Philadelphia: John Benjamins Publishing Company.
- Kisiel, Anna & Piotr Sobotka. 2022. The paths of grammaticalization of North Slavic connectors. An interface point of Slavic, Greek and Latin. In Imke Mendoza & Sandra Birzer (eds.), *Diachronic Slavonic Syntax: Traces of Latin, Greek and Church Slavonic in Slavonic Syntax* (Trends in Linguistics. Studies and Monographs 348), 11–36. Berlin–Boston: De Gruyter Mouton.
- Rysová, Magdaléna. 2017. Discourse connectives: From historical origin to present-day development. In Katrin Menzel, Ekaterina Lapshinova-Koltunski & Kerstin Kunz (eds.), *New perspectives on cohesion and coherence. Implications for translation* (Translation and Multilingual Natural Language Processing 6), 11–34. Berlin: Language Science Press.

Auxiliary, light or lexical: the history of GO verbs

The development of lexical verbs into auxiliaries and light verbs is a well studied topic (for an excellent summary of the literature, see Butt (2010)). It has been assumed that light verb is an intermediate diachronic stage in the development from lexical to auxiliary verb (see Hook (1991) and Hopper and Traugott (1993: 108–112), though the latter uses the more restricted term ‘vector verb’). This view has been challenged by Butt (2010), Butt and Geuder (2001), Butt and Lahiri (2013), who argue that the light verb and the auxiliary are independent developments from the lexical original (the challenge is recognised by Hopper and Traugott (2003: 111–114)).

We agree that cross-linguistic data do not support an analysis of the development in terms of a linear trajectory from lexical to light verb to auxiliary. However, in this paper, we use the GO verbs of Germanic and Romance to show that the picture that emerges is more complex than a bifurcation from the lexical verb into an auxiliary and a light verb use.

With the term GO verb we understand a motion verb that is neutral in the sense that it does not make reference to path or manner of motion (Fanego 2012). It is important to recognise that these verbs in the two language families have a range of origins. The English *go* comes from a verb meaning ‘walk’, as does the French *aller*, but the Romance *v*-forms go back to Latin *vadere* ‘rush, advance’, cognate with English *wade*, and the *i*-forms to Latin *ire*, cognate with the Old Eng past *eode*. In most modern varieties of Romance and Germanic there is a (suppletive) GO verb which has developed auxiliary and/or light uses, but there are interesting similarities and differences in use both between and within the two language families. For instance, we show that Dutch, French and Catalan have developed both an auxiliary and a light verb use, whereas Swedish has a light GO verb, but no auxiliary use and the Romanian GO verb *merge* has not developed any non-lexical uses. In French and Dutch the auxiliary use with an infinitive is future oriented, whereas in Catalan GO + INF indicates past. In Italian, GO + PAST PARTICIPLE can be used for a type of passive, though with special semantic restrictions, and in both Swedish and Sicilian mirative uses have developed.

In this paper, we use the GO verbs of the two language families to argue that the data is best captured in terms of a network of uses, which can in turn be represented in a semantic map (compare Lichtenberk 1991).

References

- Butt, Miriam. 2010. The light verb jungle: still hacking away. In *Complex predicates. Cross-linguistic perspectives on event structure*, ed. Mengistu Amberber, Brett Baker, and Mark Harvey, 48–78. Cambridge: Cambridge University Press.
- Butt, Miriam, and Wilhelm Geuder. 2001. On the (semi)lexical status of light verbs. In *Semi-lexical categories: On the content of function words and the function of content words*, ed. Norbert Corver and Henk van Riemsdijk, 323–370. Berlin: Mouton de Gruyter.
- Butt, Miriam, and Aditi Lahiri. 2013. Diachronic pertinacity of light verbs. *Lingua* 135:7–29.
- Fanego, Teresa. 2012. Motion events in English: the emergence and diachrony of manner salience from Old English to Late Modern English. *Folia Linguistica Historica* 33:29–85.
- Hook, Peter Edwin. 1991. The emergence of perfective aspect in Indo-Aryan languages. In *Approaches to grammaticalization*, ed. Elizabeth Traugott and Bernd Heine, 59–89. Amsterdam: John Benjamins.
- Hopper, Paul J, and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hopper, Paul J, and Elizabeth Closs Traugott. 2003. *Grammaticalization (2nd ed)*. Cambridge: Cambridge University Press.
- Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67:475–509.

Competition in the aspect-mood domain: The standardization of a diachronic data set of New Persian

Diachronic studies of the verbal categories in New Persian (10th to 20th centuries) have been conducted either with traditional methods of extraction and analysis of the examples (e.g., Lazard 1963; Natel-Khanlari 1986; Ahmadi-Givi 2001) or through limited sampled data (e.g., Lenepveu-Hotz 2012). This contribution aims at filling the gap of corpus-based studies in this field by means of a discussion of a newly designed data set, concentrated on the aspect-mood markers of New Persian. The ultimate goal is to follow the diachronic changes in the functions that are expressed by these markers. The data set includes 77,000 verb tokens, sampled from 55 texts across eleven centuries (5 per century). In each text, 1400 verbs were extracted from two different batches (700 consecutive verbs from each one), and all of these verbs were labelled for TAM categories as well as their morphological structure, clause type, presence of negation, and event type (stative/dynamic). In total, four inflectional markers as well as four major periphrastic constructions coding aspect-mood categories are attested (see Figure 1).

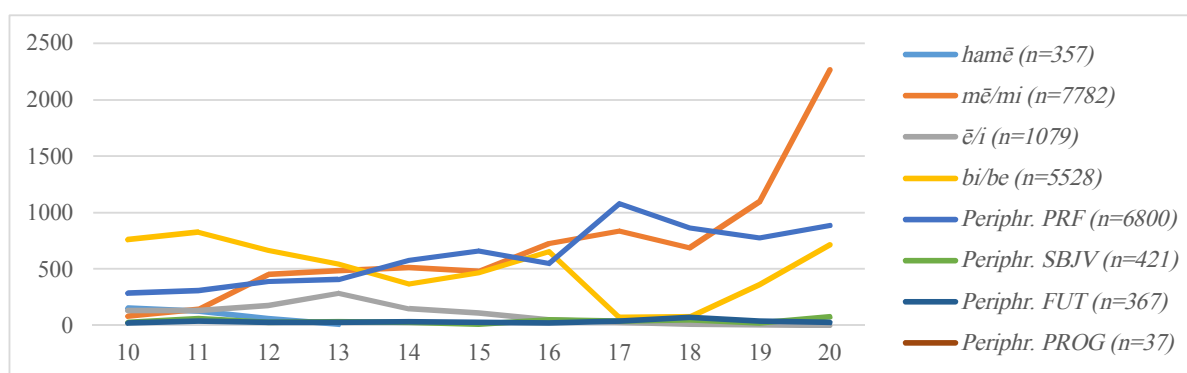


Figure 1. The token frequency of aspect-mood markers in the data set (by century)

While the token frequency is generally genre-based, the proportions of each function expressed by the markers and their changing trends throughout are significant, revealing extensive competition between the markers as well as zero-marking strategies.

In the course of labelling, a number of standardization issues were considered. Firstly, the progressive category was divided into two subcategories (durative and focalized), following Bertinetto et al. (2000), in order to trace the imperfective marker's development more efficiently. Secondly, the future and irrealis were considered as extended interpretations of the general imperfective marker. Thirdly, three stative verbs expressing BE, HAVE, and SHOULD were observed to generally resist aspect-mood inflectional marking, and their exclusion from the analyses showed a more straightforward picture of the generalization of the developing markers. The data set is still being expanded and revised as required by the research topics which are being addressed, but the preliminary results are promising.

References:

- Ahmadi-Givi, Hassan. 2001. *Dastur-e tārix-i-e fe'l* [The historical grammar of verb]. Tehran: Ghatreh.
- Bertinetto, Pier Marco, Karen H. Ebert & Casper de Groot. 2000. The progressive in Europe. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe*, 517–558. Berlin & New York: Mouton de Gruyter.
- Lazard, Gilbert. 1963. *La langue des plus anciens monuments de la prose persan*. Paris: Librairie C. Klincksieck.
- Lenepveu-Hotz, Agnès. 2012. *Etude diachronique du système verbal persan (X^e-XVII^e siècles): d'un équilibre à l'autre?*. Paris: Ecole Pratique des Hautes Études PhD dissertation.
- Natel-Khanlari, Parviz. 1986. *Tārix-e zabān-e Fārsi* [The history of Persian]. Tehran: Ferdows.

The Expression of Negation in Sabde Minyag

The Minyag language (木雅語) (Qiangic, Tibeto-Burman; ISO 639-3: mvm) is a severely endangered language that is spoken by a small group of Tibetan people in the Kangding and Ya'an (雅安) Counties in the western Sichuan province of China (Huang 1985). Most Tibeto-Burman languages in Western Sichuan retain the negative prefix *ma-/mə-* which is reconstructed for Proto-Tibeto-Burman **ma-*, but Sabde Minyag (沙德木雅語) contains either preverbal or postverbal morphological negators *ŋə*, *mə*, and *tɛi*, some of which assimilate to the vowel pattern of adjacent verb stems or directional prefixes. The Minyag negation is more integrated in the verbal morphology; it can undergo fusion with other morphological categories, and it can appear between other inflectional suffixes. Sabde Minyag consists of two negative types— standard vs. non-standard negation. Unlike the negatives in standard negation, only the *tɛ-type* negators are predominantly used for non-standard negation. Consider the following examples of negative structures:

• Standard Negation

(1a) *ʔɛtsí kʰə́=ji tɛ́ kʰə́-tɛv=ri ŋu-və=ti.*
 this dog=ERG house DIR-watch=NMLZ NEG-do=GNO:IMM
 ‘This dog does not watch the house.’

(1b) *məŋæ=yæ mǎŋí-ni kə́ hɛ-ndzi-ŋə-pi*
 Minyag=POSS people-PL:ERG fish DIR-eat-NEG-IMPV.3
rí tə-ló hɛrɪŋv?
 reason one-CL INTRO
 ‘Why don’t the Minyag people like eating fish?’

(2a) *momo=ji mætətɛə yú-mu-tɛʰə-si.*
 mother=ERG lunch DIR-NEG-drink-PFV.3:HIN
 My mother did not drink (have) lunch.

(ab) *ŋi kəŋú tɛ́ tɛ-lə nə qʰə́-tə-mv-sə.*
 1sg:ERG before house one-CL even DIR-buy-NEG-PFV.1sg:HIN
 ‘I did not buy a house (even though I was rich before).’

• Non-Standard Negation

(3a) *kʰə́pʰí nə-tɛə-vi, tɛə=kʰú xu!*
 beg DIR-PROH-do house=LOC go:IMP
 ‘Don’t beg! Go home!’

(3b) *zi qʰo-mv-səʰ=ʰv, ndó hɛ-ndzi=ri tɛə-ndə.*
 pig DIR-NEG-raise=LNK meat DIR-eat=NMLZ NEG-have
 ‘If (you) did not raise the pig, you would not have meat to eat.’

Negation reveals neutralization of tense-aspect distinctions where several negators drop their realis/irrealis temporal disparity in negative contexts. Furthermore, the skewing postverbal negation in Sabde Minyag reflects a recent grammaticalization of a post-head negative particle out of a negative-auxiliary verb combination, which is in close relation to the development of sentence-final aspectual-evidential auxiliaries. The rise of postverbal negation reveals a morphosyntactic mutation from auxiliary negators to morphological negators in diachrony.

References

- Bradley, David. 1979. *Proto-Loloish*. London & Malmö: Curzon Press.
- Chafe, Wallace. 1995. The realis-irrealis distinction in Caddo, the Northern Iroquoian language and English. In Joan Bybee & Suzanne Fleischman (eds.), *Modality in Grammar and Discourse*, 349-366. Amsterdam: John Benjamins.
- Croft, William. 1991. The evolution of negation. *Journal of Linguistics* 27.1: 1-27.
- Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17: 79-106.
- DeLancey, Scott. 2015. The origins of postverbal negation in Kuki-Chin. *North East Indian Linguistics* 7, 203-212. Canberra, Australian National University: Asia-Pacific Linguistic Open Access.
- Dryer, Matthew S. 1988. Universals of negative position. In Michael Hammond, Edith A. Moravcsik, and Jessica Wirth (eds.), *Studies in Syntactic Typology*, 93-124. Amsterdam: John Benjamins.
- Dryer, Matthew S. 2008. Word order in Tibeto-Burman languages. *Linguistics of the Tibeto-Burman Area* 31: 1-88.
- Dryer, Matthew S. 2013. Negative morphemes. In Dryer and Haspelmath (eds.), <<http://wals.info/chapter/112>>.
- Givón, Talmy. 1978. Negation in language: Pragmatics, function, ontology. In Peter Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*, 69-112. New York: Academic Press.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), *Universals of Language*, 40-70. Cambridge, MA: MIT Press.
- Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.
- Huang, Bufan 黃布凡. 1985. Muyayu Gaikuang 木雅語概況 [An Introduction to the Minyag Language], *Minzu Yuwen 民族語文* [Minority Languages in China] 3: 62-77.
- LaPolla, Randy J. 2017. Overview of Sino-Tibetan morphosyntax. In Graham Thurgood and Randy J. LaPolla (eds.), *Sino-Tibetan Languages (2nd)*, 40-69. London & New York: Routledge.
- Master, Alfred. 1946. The zero negative in Dravidian. *Transactions of the Philological Society* 4: 137-155.
- Miestamo, Matti. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin, Germany: Mouton de Gruyter.
- Payne, John R. 1985. Negation. In Timothy Shopen (ed.), *Language Typology and Syntactic Description (vol. 1)*, 197-242. New York: The Cambridge University Press.
- Post, Mark W. 2015. Sino-Tibetan negation and the case of Galo: Explaining a distributional oddity in diachronic terms. *Language and Linguistics* 16.3: 431-464.
- Takumi, Ikeda. 2022. Negation in Mu-nya. In Hayashi Norihiko & Ikeda Takumi (eds.), *Grammatical Phenomena on Sino-Tibetan Languages* 5: 105-115.
- van der Auwera, Johan, and Olga Krasnoukhova. 2020. The typology of negation. In Viviane Deprez and M. Teresa Espinal (eds.), *The Oxford Handbook of Negation*, 91-116. New York: The Oxford University Press.

The diachrony of Basque accentuation: comparative method and internal reconstruction

In this presentation I address two main issues that remain open in the reconstruction of Common Basque accentuation: The origin of a tonal accent contrast in a small Central area and accentogenesis in Common Basque.

Regarding accentuation and from a diachronic point of view, we can distinguish two major areas in the Basque-speaking territory, a Western-Central area and an Eastern area, where the boundary between the two roughly coincides with the course of the Bidasoa river (Hualde 2022). The facts are particularly complex in the Western-Central area, where we find many differences among local varieties regarding both the accentual rules and the phonetic realization of accentual prominence. What serves as unifying factor in the Western-Central area is that, to the extent that local varieties have contrastive accent (i.e. exceptions to the general accent rule), the same words and classes of both words tend to show marked accentuation (e.g. all plurals, certain compounds and derived words, certain borrowings). Although for the most part there is now agreement regarding the diachronic developments that have given rise to the different accentual systems found within Western-Central Basque (see Hualde 2003), an issue where opposite hypotheses have been put forward in recent work is the origin of a contrast between high-toned and low-toned accents in a small Central area (Goizueta and surrounding area, mostly in Navarre). Whereas in some work this has been claimed to have resulted from a relatively recent dialectal development (Hualde 2007, Egurtzegi & Elordieta 2022), in other work the tonal accent contrast is reconstructed for at least Proto-Western-Central Basque (Hualde 2012, 2022). Here I show that in order to explain the interdialectal correspondences that we find, we need to postulate two classes of accented words in Proto-Western-Central Basque, in addition to a larger unaccented class. This interdialectal comparative evidence has not been sufficiently considered before. The interdialectal correspondences that we find in words with bisyllabic stems are the following ([+1] = initial accent, [+2] peninitial accent):

| | Western | Central | Goizueta | Reconstr. | Example |
|---|------------|---------|-----------|------------|----------------------|
| 1 | Unaccented | [+2] | [+2 High] | Unaccented | gizon ‘man’ |
| 2 | [+2] | [+1] | [+2 Low] | [+2 Low] | gizòn-ak ‘(the) men’ |
| 3 | [+1] | [+1] | [+1 Low] | [+1 Low] | màlko ‘tear’ |
| 4 | Unaccented | [+2] | [+1 High] | [+1 High] | lóre ‘flower’ |

Correspondence 2 is obtained when items in the majority class (= correspondence 1) bear certain inflectional suffixes, including the plural. These correspondences are most straightforwardly explained by postulating that the Proto-system had a class of unaccented words (like Western Basque) and two smaller classes of accented words, of different origins (compounds vs loanwords), most likely differentiated by their tonal melody (like in present-day Goizueta). I will consider in detail the diachronic evolution from the reconstructed proto-system to each of the modern Western-Central systems.

A second issue that is explored in this presentation is that of accentogenesis in Common Basque, which has consequences for the exact diachronic link between Western-Central and Eastern accentual systems. Jacobsen 1975 [2022] notices that marked accentuation arising from the contraction of vowel sequences is found in both Western-Central and Eastern areas and claims that this is the original source of accent in Common Basque. Here I will demonstrate that these contractions are all relatively recent developments. All prior prosodic distinctions were lost in Eastern Basque as a consequence of a shift of the accent from the initial or pen-initial syllable to the penultimate (Michelena 1977, Hualde 2007), but later contractions of vowel sequences have sometimes resulted in convergent developments between dialects.

References

- Egurtzegi, Ander & Gorka Elordieta. 2022. A history of the Basque prosodic systems. *Diachronica* [online first] doi.org/10.1075/dia.20066.egu
- Hualde, José Ignacio. 2003. From phrase-final to post-initial accent in Western Basque. In Paula Fikkert & Haike Jacobs (eds.), *Development in prosodic systems*, 249-281. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110894530.249>
- Hualde, José Ignacio. 2007. Historical convergence and divergence in Basque accentuation. In Carlos Gussenhoven & Tomas Riad (eds.), *Tones and tunes*, vol. 1, 291-322. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110207569.291>
- Hualde, José Ignacio. 2012. Two Basque accentual systems and the notion of pitch-accent language. *Lingua* 122: 1335-1351. <https://doi.org/10.1016/j.lingua.2012.05.003>
- Hualde, José Ignacio. 2022. Accent shift and the reconstruction of Old Common Basque accentuation. Junko Ito, Haruo Kubozono & Armin Mester (eds.), *Prosody and prosodic interfaces*, pp. 29-62. Oxford: Oxford Univ. Press.
- Jacobsen, William H., Jr. 1975. Historical implications of the Western Basque Tonal Accent. University of Nevada Reno, ms. Published 2022, *ASJU*, 54 (1-2), 283-303. <https://doi.org/10.1387/asju.23472>.
- Michelena, Luis. 1977. *Fonética histórica vasca*, 2nd ed. Donostia-San Sebastián: Diputación Foral de Guipúzcoa.

The history of the Basque pronoun *zuek* ‘you.all’ in relation to similar Romance developments

1. Introduction

Historical Basque has the following personal pronoun system: 1sg *ni*, 2sg familiar *hi*, 2sg formal *zu*, 1pl *gu* ~ *guk*, 2pl *zuek* (~ *zuok*). Although essentially this system is found since the first extensive documentation of the language (in the 16th century), leaving aside secondary developments, a somewhat different and simpler system can be uncontroversially reconstructed for an earlier stage from strong morphological evidence (Azkue 1923-1925: §641, Trask 1997: 196): 1sg *ni*, 2sg *hi*, 1p *gu*, 2pl *zu*. That is, *zu* ‘2sg formal’ used to be ‘2pl’ and the modern 2pl form *zuek* is a more recent creation; the less successful—it is no longer in use in Modern Basque—1pl *guk* is likewise an innovative form. As stated by Alberdi (1995: 280), “[i]t is impossible to date the proposed evolution for *zu*: 2pl > 2pl & 2sg formal > 2sg formal”; however, in what the emergence of *guk* and *zuek* is concerned, this paper offers some anchors that may contribute to the establishment of the temporal and geographical axes in which this innovation occurred.

2. Issues regarding the emergence of *zuek*

The emergence of new 1pl and 2pl forms sketched above mirrors that from Latin to Spanish, Catalan and Occitan, and Romance influence on Basque is generally assumed. In order to determine the details of this influence the following points need to be borne in mind:

- a) Whereas Romance *nosotros*, *vosotros*, *nosaltres*, *vosaltres*... is a grammaticalization of ‘we/you others’, the Basque pronouns *guk*, *zuek* involves the affixation of a demonstrative.
- b) The 1pl form *guk* is actually attested (in a Basque gloss to an 11th c. Latin document) before Sp. *nosotros* ‘we’ and *vosotros* ‘you.pl’ became grammaticalized (12th-15th c., de Jonge & Nieuwenhuysen 2012: 249-250). If Romance influence is to be invoked, other sources need to be considered.
- c) Unlike in Romance languages, in Basque the innovative second plural *zuek* has given rise to distinct agreement marking on verbs, e.g. *zu zara* ‘you.sg are.sg’ vs *zuek zarete* ‘you.pl are.pl’, with the exception of a few Western dialects, where *zara* has remained ambiguous as for number up to the 20th c.

3. The grammaticalization of *zuek* and the historical development of its Romance counterpart

Language internal and external facts may serve to the establishment of the temporal as well as the geographical axes of the emergence of the new Basque pronouns.

- a) *Zuek* is most plausibly the continuator of the collocation **zu hek*, with the demonstrative used as an emphatic element (Manterola 2015: 340). The grammaticalization of this distal demonstrative *hek* ‘those’ is to be considered under the more general panorama of the emergence of the Basque nominal definite inflection, e. g. *lagun hek* ‘those friends’ > *lagunek* ‘the friends’. The closest Romance parallel would be early plural forms like *vos mesmos* ‘yourselves’ (which were outcompeted by *vosotros* in Romance).
- b) The pronoun *zuek* is present in all Basque dialects; it therefore qualifies as a feature of the old Basque medieval koine as propounded by Mitxelena (1981). Both the emergence of the definite inflection and the process of “koinefication” are generally believed to have started/occurred around the 6th-8th c.
- c) According to Gomila (2022), the new 2pl Romance pronouns spread westward in the Iberian Romances from the Occitan/Catalan area, with early examples in the Kingdom of Navarre (12th-13th c.).

4. Conclusion

This paper argues that the emergence of the new pronoun forms may therefore be set within a time span around the 10th century, and points to the Eastern Basque area (Navarre) in contact with Occitan as its origin. This complies with the historical description that also the new verbal forms spread westward within Basque, and fits with observations beyond (pro)nominal inflection, such as the fact that the ground zero of a well defined layer of old Romance lexical borrowings shares similar times and geography. On an additional note, the development of unambiguous 2pl agreement verbal forms in Basque (unlike in Romance) is explained as resulting from certain properties of Basque verbal morphology, which allowed for analogical transfer of the number distinction from 3rd person forms.

References

- Alberdi, Xabier. 1995. The development of the Basque system of terms of address and the allocutive conjugation. In J. I. Hualde, J. A. Lakarra & R. L. Trask (eds.), *Towards a History of the Basque Language*, 275-293. Amsterdam, Philadelphia: John Benjamins.
- Azkue, Resurrección M^a. 1923-1925. *Morfología vasca*. Euskaltzaindia: Bilbo.
- de Jonge, Bob & Dorien Nieuwenhujsen. 2012. Forms of address. In J. I. Hualde, A. Olarrea & E. O'Rourke (eds.), *The Handbook of Hispanic Linguistics*, 247-262. Malden, MA: Wiley-Blackwell.
- Gomila, Marina. 2022. *Los pronombres nos(otros) y vos(otros) en castellano medieval: variación y cambio en el espacio geográfico peninsular*. Doctoral dissertation, Universitat de les Illes Balears.
- Manterola, Julen. 2015. *Towards a history of Basque morphology: articles and demonstratives*. Doctoral dissertation, UPV/EHU.
- Mitxelena, Koldo. 1981. Lengua común y dialectos vascos. *Anuario del Seminario de Filología Vasca "Julio Urquijo"* 15, 289-313.
- Trask, Robert Lawrence. 1997. *The history of Basque*. Londres, New York: Routledge.

The Afrikaans auxiliary *het* [hɛt] ‘have’ has undergone a development from its full form as a Dutch finite dialectal form of *hebben* ‘to have’ to a clitic in verb-second context and an inflectional ending (alternating with the full form) in clause-final position, e.g.

Sy't ([sɔit]) *gewen, maar hy kon ook gewen het* ([hɛt] or [ɛt])
 she.have.AUX win.PST.PTCP but he can.PRET also win.PST.PTCP have.AUX
 ‘She won, but he could also have won.’

The common clause-final [ɛt] pronunciation of *het* (the written form) is reflected in the spelling used in the Kaaps variety of Afrikaans, viz. *gedoenit* ‘have done’, with a glide (spelt *r*) inserted after long vowels and diphthongs in Kaaps in particular, e.g. *gegeerit* ‘have given’ and *gehourit* ‘have kept’.

Het, which is predominantly an auxiliary but also serves as main verb (meaning ‘to have, possess’), is one of the ten most frequent lexical items in the language. Its status as an inflectional ending when used clause-finally, however, depends on the fact that, unlike all other auxiliaries, it is completely inseparable from the past participle it governs, even in infinitival phrases – which always require an infinitive after the particle *te* ‘to’:

*Om <gister> te geslaag <*gister> het <gister>, was nie maklik nie.*
 COMP <yesterday> to succeed.PST.PTCP <yest.> have <yest.> was NEG₁ easy NEG₂
 ‘To have succeeded yesterday was not easy.’

Through the reanalysis of past participle + *het* as verb + ending, univerbation has taken place and the “new” verb – a periphrastic perfect functioning as past tense – now follows *te* ‘to’ in its entirety.

The purpose of the paper is to provide supporting data for the various phases in the development outlined above. Several interrelated factors in the earlier history of Afrikaans contributed to this development. The elimination of the inherited clause-final variant order of auxiliary + past participle, as in Dutch *hebben geslaagd* ‘have succeeded’, assured auxiliaries of a fixed position *after* the past participle. The collocation of past participle with *het*, in particular, was probably strengthened by the rise in frequency of *het* as auxiliary. This, in turn, was brought about (i) by the replacement of *is* ‘to be’ by *het* ‘to have’ as auxiliary of unaccusative verbs, and (ii) by the increased usage of the periphrastic perfect as general past tense after the demise of the synthetic preterite used in this function. The use of the perfect, again, was facilitated by the across the board regularisation and deflection of inherited past participles, e.g. *gesproken* > *gespreek* for the strong verb *spreek* ‘speak’ and *gewerkt* > *gewerk* for the regular verb ‘work’. Furthermore, the loss of participial suffixes signalled the removal of an important impediment to univerbation.

In sum, the replacement of the synthetic preterite, which is still ongoing in the case of the modal auxiliaries *sou* ‘would’, *moes* ‘had to’, *kon* ‘could’ and *wou* ‘wanted to’, by the periphrastic perfect with *het*, is followed closely by the univerbation of past participle + *het*, whereby a new synthetic tense form is created.

Reconstructing Proto-Austronesian Interrogative Pronouns

Blust (2009/2013) very astutely notices that interrogative pronouns for ‘who’ are morphologically complex in a great number of Formosan and Philippine languages. Specifically, this involves attachment of a reflex of the personal nominative case marker *si to a base *ma*. Note that this base is also found in the forms for ‘what’.

| | | |
|--------|------------|-------------|
| (1) | <u>WHO</u> | <u>WHAT</u> |
| Thao | ti-ma | nu-ma |
| Bunun | si-ma | ma-az |
| Amis | ci-ma | ma-an |
| Paiwan | ti-ma | nu-ma |
| Truku | i-ma | ma-nu |

However, he does not reconstruct *sima ‘who’ to Proto-Austronesian. He opts instead for the form *ima and proposes that the forms in (1) are the result of an innovation which added *si to the original *ima: *si-ima > *sima.

Blust is certainly correct in reconstructing interrogative pronouns with incorporated case markers or determiners, but the exact forms he chooses introduce problems when it comes to accounting for synchronic variation. First, both *ima and *sima are reflected only in Nuclear Austronesian (NucAn; Ross 2009) languages and not in the more conservative languages Rukai, Tsou, and Puyuma.

| | | |
|----------------|------------|-------------|
| (2) | <u>WHO</u> | <u>WHAT</u> |
| Tanan Rukai | a-nu | ma-nu |
| Tsou | si-a | cu-ma |
| Nanwang Puyuma | manay | manay |

Puyuma is particularly revealing, since ‘who’ and ‘what’ are distinguished only by their case marking, adding the personal nominative yields *i manay* ‘who’ and adding the common noun nominative produces *a manay* ‘what’, when the pronouns function as a subject. If the pronoun surfaces in object position, then it is preceded by an object case marker. Tsou also presents an interesting case. The *si-* in the form for ‘who’ is one of several nominative case markers, which each encode the referent’s visibility and distance from the speaker. This *si-* is probably cognate with the personal nominative marker *si* in NucAn languages, but in Tsou it still retains more functions of the demonstrative it grammaticalized from and is not related to person marking. In contrast to this, the Rukai form for ‘what’ clearly shows object marking, object pronominal forms being prefixed with a syllable beginning with a nasal consonant, e.g. *mo-so-a* ‘ACC-you-ACC’. Assuming that the *a-* in the Rukai form for ‘who’ is also a determiner cognate with the Puyuma common noun nominative marker, it can be seen that all of the forms in (2) for ‘who’ are marked with a determiner, typically marking nominative case, while the forms for ‘what’ are generally marked like objects. From this, it can be concluded that PAn interrogative pronouns can be reconstructed as having incorporated subject and object case marking.

I reconstruct the Rukai forms to PAn: *a-nu ‘who’ and *ma-nu ‘what’. These are in turn formed from the attachment of the determiner *a to ‘who’ and the object marker *ma- to the base *nu, which can be reconstructed as an indefinite pronoun. This makes it possible to construct a paradigm of interrogative pronouns including two additional forms: *i-nu ‘where’ < LOC *i + INDEF *nu; and *na-nu ‘which’. From these, the paradigms in both (1) and (2) can be derived. Tsou innovated new forms by adding its own case markers to the PAn forms and then deleting the final syllable: *si-anu > *sia*, *cu-manu > *cuma*, assuming that *cu-* reflects an archaic object case marker in Tsou. Truncation of the form for ‘what’ led to the reanalysis of *ma* as the indefinite pronoun found in all of the forms in (1). In Puyuma, ‘who’ and ‘what’ merged in favor of ‘what’. Truncation did not take place in this language, since there was no morphological incorporation of case marking. The ‘what’ form in Truku directly reflects PAn *manu. Regarding *i-ma* ‘who’, this can be explained in terms of the same rule as the other NucAn forms, i.e. *i* is the nominative marker for personal names in this language, as it is in Puyuma. The other languages reflect the truncated form of ‘what’, which combines with a case marker, nominative personal marking for ‘who’ and object common noun marking for ‘what’. On this analysis, the forms of interrogative pronouns in Formosan languages are explained straightforwardly in terms of a general process of attaching a determiner/case marker to an indefinite pronoun.

References

- Blust, Robert. 2009/2013. *The Austronesian Languages*. Canberra: Research School of Pacific and Asian Studies (2nd edn. Collage of Asia and the Pacific), Australian National University.
- Ross, Malcolm. 2009. Proto Austronesian Verbal Morphology: A reappraisal. In Alexander Adelaar and Andrew Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A festschrift for Robert Blust*, 295-326. Canberra: Pacific Linguistics.

The emergence of a Welsh biblical literary standard and the evidence of early modern manuscript sermons

Oliver Currie, University of Ljubljana

The 1588 Welsh Bible, as revised in 1620, is generally regarded as having formed the basis of modern standard literary Welsh (Lewis 1987; Jones 1998; Robert 2011), however, the process of standardization itself has not been systematically researched. Manuscript sermons can provide an interesting insight into the adoption of an emerging biblical literary standard, as we have a large, underresearched body of comparable texts based on the Bible, but where we find significant linguistic variation between individual preachers. Welsh preachers faced a linguistic dilemma: they could follow the linguistic model of the Welsh Bible, which they read to their congregations week in week out, or use a more colloquial and dialectal language closer to that which they themselves and their congregations spoke.

Today, there is a considerable divergence between standard literary Welsh and colloquial Welsh. By comparing the 1588 and 1620 Bibles to contemporary text types in a more informal register, such as Slander case records and popular drama, we can see that many of the points of divergence between standard literary Welsh and colloquial Welsh had emerged at least as early as the Bible translations and are likely to have been cemented by the Bible translations. This paper examines sociolinguistic variation in a self-compiled corpus of over 50 Welsh language autograph manuscript sermons from the late 16th to the early 18th century in 14 manuscripts by 14 different preachers, focusing on a selection of morphological and morphosyntactic variables where there the 1620 Bible diverged from contemporary more popular texts (e.g. the verbal endings 1SG PRES/FUT *-af/-a*, 1SG PAST *-ais/-es*, general 3PL *-nt/-n*, 3SG IMPF *-ai/-e*, the nominal plural ending *-au/-e*, the third person plural pronoun *hwy/nhwy*, and the retention vs. omission of the preverbal particles *a* and *y*) as well as diatopic variables where the Bible used supralocal as opposed to dialectal variants (e.g. the 3SG M simplex personal pronoun – Biblical supralocal *ef* vs. Northern local/dialectal *fo*).

The paper will first, by way of background, suggest a possible explanation of how the language of the Welsh Bible came to diverge from colloquial usage, then examine the extent to which individual preachers used biblical as opposed to more colloquial or dialectal variants, and finally discuss how the data can contribute to our understanding the development of a Welsh literary standard. While we can observe a progressive adoption of linguistic features consistent with the 1620 Welsh Bible – mid and late 17th century preachers use more biblical features than early 17th century preachers – there is significant synchronic and diachronic variation throughout the 17th century. This simultaneous norm convergence, on the one hand, and variation, on the other, reflects a key characteristic of the emergent standardisation of Welsh in the 17th century: it involved organic convergence to the language of an authoritative and widely diffused text, the Bible, but without a planned or centrally coordinated implementation process, comparable to what Joseph (1987, 60) has termed *circumstantial* as opposed to *engineered standardization* or what Deumert (2004, 3) has described as standardization without “deliberate intervention”. The lack of a deliberate implementation process – promotion of a standard or formal education in Welsh – not only meant that there was less pressure to conform to a standard, but also that the standard itself was not rigidly defined. In this respect, the emerging Welsh biblical literary standard in the 17th century appears to be a standard with fuzzy boundaries (Ammon 2003; Brown 2020). Variation in the adoption of a linguistic model is to be expected not only because of the agency of individual writers who can choose to follow it to varying degrees, but also because the linguistic features of a potential model text differ in how easy they are to adopt because of their variable salience or variable proximity to colloquial or dialectal usage.

References

- Ammon, Ulrich. 2003. "On the social forces that determine what is standard in a language and on conditions of successful implementation." *Sociolinguistica* 17 (1): 1-10.
- Brown, Josh. 2020. "Language history from below: Standardization and koineization in Renaissance Italy." *Journal of Historical Sociolinguistics* 6 (1): 1-28.
- Deumert, Ana. 2004. *Language Standardization and Language Change. The dynamics of Cape Dutch*. Vol. 19. *Impact: Studies in Language and Society*. Amsterdam: John Benjamins.
- Jones, Mari C. 1998. *Language obsolescence and revitalization. Linguistic change in two socially contrasting Welsh communities*. Oxford: Clarendon Press.
- Joseph, John E. 1987. *Eloquence and Power. The Rise of Language Standards and Standard Languages*. London: Frances Pinter.
- Lewis, E. Glyn. 1987. "Attitude to the planned development of Welsh." *International Journal of the Sociology of Language* 66: 11-26.
- Robert, Elen. 2011. "Standardness and the Welsh language." In *Standard Languages and Language Standards in a Changing Europe*, edited by Tore Kristiansen and Nikolas Coupland, 135-141. Oslo: Novus Press.

Vowel reduction to /i/ in functional morphemes in Northern Sub-Saharan Africa

I demonstrate that in many languages of Northern Sub-Saharan Africa vowel qualities of functional morphemes tend to be neutralized through raising, fronting and unrounding towards /i/, similarly to what has been described by Idiatov (2020:65) for the TAM and polarity markers of Greater Manding languages. Vowel reduction to /i/ in functional morphemes can be argued to be an areal phenomenon in Northern Sub-Saharan Africa, as it is attested in various language families of the area and appears to be absent in the genetically related languages spoken outside of this area. An important gap in the relevant area is formed by the Central African interior vowel zone (cf. Rolle et al. 2020), presumably formed under the influence of Chadic languages.

Even though most languages of Northern Sub-Saharan Africa are tonal, such functional morphemes, both affixes and various functional words, can be safely construed as prosodically weak thanks to the fact that typically they are affected by a whole range of concomitant lenition and neutralization processes. Neutralization through raising in prosodically weak positions can be analyzed as a type of vowel reduction process comparable to reduction through centralization. As demonstrated by Kapatsinski et al. (2020:31) reduction through raising, although well-attested in Romance and Slavic languages, is cross-linguistically much less common than reduction through centralization (contra earlier studies by Crosswhite 2001 and Barnes 2006). Interestingly, besides being typologically uncommon, reduction towards /i/ in the languages of Northern Sub-Saharan Africa seems to target primarily functional morphemes. At the same time, in prosodically weak positions within lexical morphemes where vowel reduction is also not uncommon in the region (usually driven by the phenomenon of stem-initial prominence; cf. Lionnet & Hyman 2018:652–55; Idiatov & Van de Velde 2021:93-94), it appears to proceed through the typologically more common processes of shortening, devoicing, unrounding and centralization.

Finally, I argue that recognizing the existence of an areal tendency to reduction to /i/ in functional morphemes in large parts of Northern Sub-Saharan Africa also allows us to offer a principled solution for two types of reconstruction-related issues. First, it can help us to make a principled choice in those cases where multiple, but only slightly formally divergent cognate sets and reconstructions have been proposed for a given functional morpheme, such as the reconstruction of the class 13 nominal prefix reconstructed as **ti-* for Proto-Benue-Congo by De Wolf (1985) but as **tɔ-* for Proto-Bantu, one of its major branches, by Meeussen (1967). Second, it can guide us in the search for the most plausible lexical source of a given functional morpheme, as in the case of the future (“potential”) auxiliary *sí* ~ *sé* in Mandinka that Creissels (2020) relates to the Mandinka verb *sé* ‘reach; overcome’ ignoring the possibility of another lexical source, the verb **sá* meaning ‘come’, that is more plausible both typologically and comparatively but absent as a lexical verb in Mandinka itself.

References:

- Barnes, Jonathan. 2006. *Strength and weakness at the interface: positional neutralization in phonetics and phonology*. Berlin: De Gruyter.
- Creissels, Denis. 2020. Grammaticalization in Manding languages. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization scenarios. Areal patterns and cross-linguistic variation*, 695–727. Berlin: De Gruyter.
- Crosswhite, Katherine. 2001. *Vowel reduction in Optimality theory*. New York & London: Routledge.
- De Wolf, Paul Polydoor. 1971. *The noun class system of Proto-Benue-Congo*. The Hague and Paris: Mouton.
- Idiatov, Dmitry. 2020. Perfective marking conditioned by transitivity status in Western Mande: constructional competition, specialization and merger. *Diachronica* 37(1). 43–82. doi: 10.1075/dia.18050.idi.
- Idiatov, Dmitry & Mark L.O. Van de Velde. 2021. The lexical distribution of labial-velar stops is a window into the linguistic prehistory of Northern Sub-Saharan Africa. *Language* 97(1). 72–107. doi: 10.1353/lan.2021.0013.
- Kapatsinski, Vsevolod, Shelece Easterday & Joan Bybee. 2020. Vowel reduction: A usage-based perspective. *Italian Journal of Linguistics* 32(1). 19–44. doi: 10.26346/1120-2726-146.
- Lionnet, Florian & Larry M. Hyman. 2018. Current issues in African phonology. In Tom Güldemann (ed.), *The languages and linguistics of Africa*, 602–708. Berlin: De Gruyter Mouton.
- Meeussen, Achille E. 1967. Bantu grammatical reconstruction. *Africana Linguistica* 3. 80–121.
- Rolle, Nicholas, Florian Lionnet & Matthew Faytak. 2020. Areal patterns in the vowel systems of the Macro-Sudan Belt. *Linguistic Typology* 24(1). 113–179. doi: 10.1515/lingty-2019-0028.

Towards a Diachronic Account of P-lability in Latin: The Semantic Extension of the Active Intransitive as an Anticausative Strategy in Latin

Lability refers to the phenomenon in which verbs can show valency alternation without a formal change in verbal morphology (Kulikov & Lavidas 2014). Patient-preserving lability (abbr. P-lability) refers to the transitive-intransitive alternation with the preservation of the patient-like argument (e.g. tr. I open the door vs. intr. the door opens). In Latin, P-lability is used to express the causative-anticausative alternation (Cennamo & al. 2015, Gianollo 2014, Pinkster 2015). A case in point is the verb *verto* ('to turn into'), which can be used transitively in its causative sense (1a) and intransitively in its anticausative sense (1b).

- (1) (a) ne ea quae rei publicae
 that not DEM.ACC.N.PL. REL.ACC.N.PL. republic:GEN.F.SG.
 causa egerit in suam
 for the sake of do:SBJ.PRF.3SG.ACT. in POSS.3SG.ACC.F.SG.
 contumeliam **vertat.**
 insult:ACC.FS.G turn:SBJ.PRS.3SG.ACT.
 "In order that he (sc. Pompey) does not turn what he did for the sake of the republic into insult towards him"
 (Caesar, *Commentarii belli civilis*, I, 8, 2; first century BC)
- (b) ut detrimentum in bonum **verteret**
 so that damage:NOM.N.SG. in good:ACC.N.SG. turn:SBJ.PST.3SG.ACT.
 "So that the damage turns into something good"
 (Caesar, *Commentarii belli civilis*, III, 73, 6; first century BC)

The number of verbs displaying P-lability was limited in Early and Classical Latin, but heavily increased in Late Latin. Alternative strategies to express the anticausative of originally transitive verbs were the mediopassive (-r morphology) (2a) and the reflexive strategy (reflexive pronoun + active voice) (2b) (Cennamo & al. 2015, Feltenius 1977, Gianollo 2014, Pinkster 2015).

- (2) (a) id enim et in pus **vertitur**
 DEM:NOM.N.SG. indeed also in pus:ACC.N.SG. turn:IND.PRS.3SG.MPASS.
 "It (sc. the wound) turns also into pus"
 (Celsus, *De medicina*, V, 26; first century AD)
- (b) cum terra in aquam **se**
 when earth:NOM.F.SG. in water:ACC.F.SG. REFL.ACC.
vertit
 turn:IND.PRS.3SG.ACT.
 "When earth turns into water"
 (Cicero, *De natura deorum*, III, 31; first century BC)

This paper aims to clarify in which way the unmarked intransitive grammaticalized to a generalized anticausative strategy in Latin. By means of corpus research, we discuss the following factors: *Aktionsart*, verbal class, agentivity of the anticausative subject, causalness value (= [causative uses]/[anticausative + causative uses]; see Haspelmath 2014 and Heidinger 2015), date and register of the texts.

Keywords

Latin, lability, anticausative, diathesis, *Aktionsart*, diachrony, functional typology

Bibliography

- Cennamo, M., Eythórsson, T., & Barðdal, J. 2015. Semantic and (morpho)syntactic constraints on anticausativization: evidence from Latin and Old Norse-Icelandic. *Linguistics*, 53(4), 677–729. Berlin: De Gruyter.
- Feltenius, L. 1977. Intransitivizations in Latin. Stockholm: Almqvist & Wiksell.
- Gianollo, C. 2014. Labile verbs in Late Latin. *Linguistics*, 52(4), 945-1002. Berlin: De Gruyter.
- Haspelmath, M. 1993. More on the typology of inchoative/causative verb alternations. Bernard Comrie and Maria Polinsky (eds.): *Causatives and transitivity*, 87-120. Amsterdam: Benjamins.
- Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., & Bamyacı, E. 2014. Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics*, 50(3), 587-625. Cambridge: Cambridge University Press.
- Heidinger, S. 2015. Causalness and the encoding of the causative–anticausative alternation in French and Spanish. *Journal of Linguistics*, 51(3), 562-594. Cambridge: Cambridge University Press.
- Kulikov, L., & Lavidas, N. 2014. Special issue: typology of labile verbs: focus on diachrony: Introduction. *Linguistics*, 52(4), 871-877. Berlin: De Gruyter.
- Pinkster, H. 2015. *Oxford Latin Syntax. Volume I: The Simple Clause*. Oxford: Oxford University Press.

Camiel Hamans

A revolution in the history of affix-formation

This paper wants to show a revolutionary change in the origin of affixes in English: from a syntagmatic process to a paradigmatic one.

Traditionally, suffixes are considered to be the result of grammaticalization (Kastovsky 2009: 327), which runs from free form through part of a compound to an affix (Trips 2009), as for example in

| | | | |
|-----|--------------------------|--------------------------|---|
| (1) | OE free form/noun hād | ME Compound child hōd | contemporary English derivative/suffix childhood |
|-----|--------------------------|--------------------------|---|

Kastovsky (2006: 152) describes a similar change: “The suffix *-ly* goes back to OE *-lic*, which was an independent noun meaning ‘body, form’. Thus, *-lic*-formations started out as nominal compounds, but then developed an adjectival function (...). An OE formation *cildlic* was therefore structurally parallel to its Modern English equivalent *childlike*. From such adjectives, adverbs could be formed by adding the suffix *-e*, e.g. *cildlice*. In Early Middle English this *-e* was lost, and the suffix *-ly* also adopted an adverbial function as in *slowly*, *royally*, besides continuing to act as an adjective forming suffix, cf. *manly*, *princely* etc.”

Since part of the development of these affixes goes through a syntagmatic process, compounding, this word-formation process is here called syntagmatic following Marchand’s (1969²) footsteps. Essential to syntagmatic processes is that they make use of morphemes or words. In contemporary English, however, one finds numerous affixes or affix-like segments, whose origin is by no means a morpheme. These affixes are the result of what Zwicky (2010) calls libfixation. Libfixes are non-morphemic suffix-like word fragments that are ‘liberated’ from a longer formation and that can be productively used to form new paradigms of words. Examples are:

| | | | | | |
|-----|--|-----|---|-----|--|
| (2) | -dar from radar gaydar jewdar humordar | (4) | -which from sandwich fishwich hamwich veggiewich | (6) | glut- from gluten glutamine glutamate glutaminase |
| (3) | -gasm from orgasm wargasm nerdgasm shoegasm | (5) | Franken- from Frankenstein frankenfood frankenplant frankenscene | (7) | heli- from helicopter heliport helibus helipad |

Although the liberated segments, libfixes or splinters, in (4), (5) and possibly also in (7) might be described as a result of reinterpretation by naïve language users, such an explanation is impossible in the cases (2), (3) and (6). Nevertheless both groups of libfixes appear to be productive and are on their way to become affixes.

In this presentation, the following aspects will be discussed:

- systematicity of libfixation
- the problem or role of conscious word formation in language change
- the role of the paradigm and/of frequency in the origin of libfixes
- the relation between blending and libfixation
- the role of the model in productivity
- the difference between syntagmatic and paradigmatic affix-formation, which also discusses whether or not intermediate stages of affixoids are involved
- the consequences of paradigmatic affix-formation for the unidirectionality of (de)grammaticalization processes

Reanalyzing the Historical Constructions of Albanian Prepositions

Lindon Dedvukaj, The Ohio State University

Background: The Albanian language is traditionally divided between the two main dialects of Gheg to the geographic north and Tosk (Standard Albanian) to the geographic south. Previous linguistics analyses and etymological reconstructions have not considered the Malsia Madhe dialect of Albanian, a historically and geographically isolated region spread throughout northwestern Albania and southwestern Montenegro (Shkurtaj 1967, 1974, 1975). Malsia Madhe Albanian (Malsia) reveals many preservations of Proto-Albanian (4th-6th century CE) prepositional forms, which have undergone substantial changes in Modern Gheg and Tosk Albanian. This new data allows for a reanalysis of the historical formation of Modern Albanian prepositions.

Analysis: The Standard Albanian prefix /n-/ is derived from an early preposition /ën/ ‘in’ (see Schumacher and Matzinger 2013: 40). Only in the Malsia Madhe region has /en/ been preserved as a preposition *en* /en/ ‘in’ < Proto Indo-European *h₁en ‘in’. In all other Albanian dialectal regions this has developed into a prefix /n-/ (see S. Demiraj 1971: 237-38), which underwent a morpho-phonological homorganic nasal assimilation change when attaching to a stem. The underlying nasal /n/ of the prefix assimilates to the place of articulation of the following stop (e.g. n + b > mb). Old Tosk and Gheg Albanian /ënbuʃ/ ‘to fill’ > Modern Tosk *mbush*, whereas in Malsia it is *bush*.

The Modern Tosk and Gheg preposition *në* ‘in’ is derived from Old Albanian (15th-16th CE) *ndë*. The Old Albanian preposition is derived from a Proto-Albanian form of **en-da*/**en-ta* (Orel 1998: 35, 284). Malsia appears to preserve the Proto form as *en tã* /en tã:/ ‘inside of it, within’. Malsia does not just preserve the preposition *en* ‘in’ from Proto-Albanian, but also an older form of the demonstrative *tã* /tã:/ ‘it’. The Modern Tosk equivalent *të* /tə/ is cited by Kortland (2010) as being derived from an earlier **tom*/**tãm* (Chronology can be posited as **tom* > *tã* > *të*).

The creation of the Modern Tosk preposition *nga* /ŋga/ ‘where, where from’ is described in Forston (2010: 455) as “... *nga* is shortened from **ën-ka* ‘where, from where’, with **ën* ‘in’ added to the old relative adverb *ka* ‘where, from where’ ...” In Malsia both *en* ‘in’, and the adverb *ka* ‘where’ are preserved. It is also well known that Gheg and Tosk differ in the phonological form of the preposition *mbas* ~ *pas* ‘after, behind’ (see Beci 2002: 21- 45 in Klein et. al. 2018: 1803). Tosk *mbas* is formed from the process of *en* + *pas*, whereas Malsia and some Gheg areas have preserved *pas* without the addition of the prefixal *en*.

The reconstructed prepositional forms are not always clear. Tosk *mbi* ‘on, upon’ is reconstructed back to a Proto-Albanian **ambi*. (see Orel 1998: 250-51). Dialectal variants/Proto forms of Modern Tosk *nga*, *mbas* and *mbrapa* that are cited (*ka* ~ *nga*, *pas* ~ *mbas*, *prapa* ~ *mbrapa*), are reconstructed as **en* + PREP./ADV. Because the Malsia form of the preposition *pi* ‘over, above’ (Tosk *mbi*) is unknown, the Proto form is reconstructed as **ambi*, solely based on other IE cognates and a misunderstanding of internal developments within Albanian. The Modern Tosk preposition *mbi* can be reconstructed as **en* + *pi* rather than **ambi*. The Malsia dialectal variant of *pi* ‘over, above’ can also be found outside of Albanian. The Messapian language of the southeastern Italian peninsula (7th-2nd centuries BCE) contains many cognates with Modern Albanian. Matzinger (2019: 88-89) cites Messapian *pi-* with the meaning of ‘on, thereon’ comparable to Vedic Sanskrit *pi-* and Ancient Greek π . Malsia *pi* likely reveals another unknown cognate between Albanian and Messapian.

Conclusion: The Albanian language contains many prepositions whose etymologies have remained somewhat obscure. The Malsia Madhe Albanian dialect opens an avenue for cross comparison to the Old/Modern Tosk and Gheg dialects, revealing many interesting factors that warrant a reanalysis of the historical reconstructions and the Proto Albanian prepositional forms.

References

- Demiraj, Shaban. 1971. *Morfologjia E Gjuhës Së Sotme Shqipe*. Prishtinë, Kosovë. Rilindja Prishtinë.
- Fortson, Benjamin W. IV. 2010. *Indo European Language and Culture*. An Introduction Second Edition. Malden, MA. Wiley-Blackwell.
- Gjinari, Jorgji. 1989. *Dialektet e Gjuhës Shqipe*. Akademia e Shkencave e RPS të Shqipërisë. Instituti I Gjuhësisë dhe I Letërsisë. Tiranë, Albania.
- Klein, J., Joseph, B., & Fritz, M. “edit.” 2018. *Handbook of Comparative and Historical Indo-European Linguistics*. Vol. 3. Berlin/Boston. Walter de Gruyter GmbH.
- Kortlandt, Frederick. “PIE *s in Albanian”. In A.A. Barentsen, B.M. Groen and R. Sprenger (eds.), *Studies in Slavic and General Linguistics*. 10: *Dutch studies in South Slavic and Balkan linguistics*. 1987, 219-226.
- Matzinger, Joachim. 2019. *Messapisch. Kurzgrammatiken indogermanischer Sprachen und Sprachstufen*. Wiesbaden. Reichert Verlag.
- Orel, Vladimir. 1998. *Albanian Etymological Dictionary*. Leiden, Netherlands. Brill.
- Schumacher, Stefan and Matzinger, Joachim. 2013. *Die Verben des Altalbanischen. Belegwörterbuch, Vorgeschichte und Etymologie*. Wiesbaden. Harrassowitz
- Shkurtaj, Gjovalin. 1967. *E folmja e Kastratit*. In Studimet Filologjike 2. Tiranë, Albania.
- . 1974. *E folmja e Hotit*. In Dialektologjia Shqiptare II. Tiranë, Albania.
- . 1975. *E folmja e Kelmendit*. In Dialektologjia Shqiptare II. Tiranë, Albania.

**Uncovering lost paths in the Congo rainforest:
A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu**

Sara Pacchiarotti¹, Natalia Chousou-Polydouri², Jean-Pierre Donzo^{1,3}, Guy Kouarata¹, Lorenzo Maselli^{1,4}, and Koen Bostoen¹

¹UGent Centre for Bantu Studies, Ghent University, Belgium

²Department of Comparative Language Science, University of Zurich, Switzerland

³ISP Gombe, Kinshasa, DRC

⁴Service de Métrologie et Sciences du Langage, Laboratoire de Phonétique, Université Mons, Belgium

The Bantu Expansion is the initial spread of the Bantu languages and the communities speaking them over large parts of Central, Eastern, and Southern Africa from a homeland located in the borderland between present-day Nigeria and Cameroon. This major linguistic, cultural, and demographic process in Late Holocene Africa stands out in three respects: its vastness, its rapidness, and its predominantly longitudinal orientation (Bostoen 2018). A central question about this expansion is whether and how the first Bantu-speaking populations migrated through and settled in the Congo rainforest. While previous studies suggest that the movement of Bantu-speaking people might have been favored by a climate-induced forest reduction in the Sangha River Interval (SRI) around 2500 BP (Bostoen *et al.* 2015; Grollemund *et al.* 2015), a recent study suggests that this migration through the rainforest happened well before that period and was not facilitated by SRI savanna corridors (Koile *et al.* 2022).

Although the specific quantitative methods underlying their conclusions differ, Grollemund *et al.* (2015) and Koile *et al.* (2022) are not only both phylogenetic studies, but they are also based on exactly the same datasets of basic vocabulary, the same cognacy judgments and the same underrepresented and unbalanced sample of rainforest Bantu languages. What is more, recent scholarship has seriously challenged the idea that phylogenies based on modern Bantu languages may directly reflect the initial migration of Bantu speech communities (Bostoen 2018; Gunnink *et al.* 2022; Bostoen *et al.* forthcoming). Seidensticker *et al.* (2021) argue that a population collapse hit the entire Congo rainforest ~1,500 BP, which probably led to the extinction of many ancestral lineages of Bantu languages before the area was recolonized by Bantu speakers from ~1,000 BP onwards. Hence, many of the Bantu languages currently spoken in the Congo rainforest may have an ancestry there that is more than a millennium younger than previously assumed.

In order to shed new light on the initial expansion of Bantu languages through the Congo rainforest and how its signal in lexicon-based phylogenies got possibly blurred by language death and spread-over-spread-events, we present in this talk the preliminary results of a new comprehensive lexicon-based phylogeny. This on-going study focuses on two clades of rainforest Bantu languages, i.e., Central-Western (CWB) and West-Western or West-Coastal (WCB) Bantu. These two groups display different topologies in the phylogenies of Grollemund *et al.* (2015) and Koile *et al.* (2022). While in the first, WCB branches off after CWB, in the second CWB and WCB are parallel branches. Moreover, portions of what is CWB in Grollemund *et al.* (2015) cluster more closely with WCB in Koile *et al.* (2022).

While Grollemund *et al.* (2015), the most comprehensive phylogeny of the Bantu languages to date, includes 424 doculects for the entire family, our new phylogeny includes more than 350 varieties for two branches, CWB and WCB. Featuring many varieties spoken in the DRC that were never documented before (Kouarata *et al.* forthcoming), it closely reflects modern-day Bantu language diversity within the Congo rainforest. The vast majority of our data come from first-hand fieldwork and second-hand specialized sources other than Bastin *et al.* (1999). Cognacy judgments are performed by relying on a profound knowledge of the historical phonology of the languages in question (Rottland 1977; Koni Muluwa & Bostoen 2012; Bostoen & Koni Muluwa 2014; Donzo 2015; Pacchiarotti & Bostoen 2020; Pacchiarotti & Bostoen 2021; Pacchiarotti & Bostoen 2022). We use Lexedata (Kaiping *et al.* 2022) as a toolbox to edit and annotate our lexical dataset and MrBayes 3.2 (Huelsenbeck & Ronquist 2001; Ronquist *et al.* 2012) to produce a bayesian phylogeny, which is an important first step to refine our understanding of the layered history of rainforest Bantu languages. Additionally, our preliminary results will serve as the basis to develop and subsequently test new hypotheses regarding the colonization and recolonization of the Congo rainforest by Bantu-speaking peoples.

References

- Bastin, Y., A. Coupez and M. Mann. 1999. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study*. Tervuren: Royal Museum for Central Africa.
- Bostoen, K. 2018. The Bantu expansion. In T. Spear (ed.), *Oxford Research Encyclopedia of African History*. Oxford: Oxford University Press, Oxford Research Encyclopedias Online <https://oxfordre.com/africanhistory/view/10.1093/acrefore/9780190277734.001.0001/acrefore-9780190277734-e-191>.
- Bostoen, K., B. Clist, C. Doumenge, R. Grollemund, J.-M. Hombert, J. Koni Muluwa and J. Maley. 2015. Middle to late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of West Central-Africa. *Current Anthropology* 56: 354-384.
- Bostoen, K. and J. Koni Muluwa. 2014. Umlaut in the Bantu B70/80 languages of the Kwilu (DRC). *Transactions of the Philological Society* 112: 209-230.
- Bostoen, K., S. Pacchiarotti and W. Hubau. forthcoming. A la recherche des sentiers perdus du passé en forêt: Reflexions sur l'effondrement de la population en Afrique centrale il y a 1500 ans. *Bulletin des Séances de l'Académie Royale des Sciences d'Outre-Mer*.
- Donzo, J.-P. 2015. *Langues bantoues de l'entre Congo-Ubangi (RD Congo): documentation, reconstruction, classification et contacts avec les langues oubanguiennes*. Gand: Université de Gand (UGent), thèse de doctorat.
- Grollemund, R., S. Branford, K. Bostoen, A. Meade, C. Venditti and M. Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the United States of America* 112: 13296-13301.
- Gunnink, H., N. Chousou-Polydouri and K. Bostoen. 2022. Divergence and contact in Southern Bantu language and population history: A new phylogeny in cross-disciplinary perspective [07 Nov. 2022]. *Language Dynamics and Change*: 1-58.
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754–755.
- Kaiping, G., M.S. Steiger and N. Chousou-Polydouri. 2022. Lexedata: A toolbox to edit CLDF lexical datasets. *Journal of Open Source Software* 7: 4140.
- Koile, E., S.J. Greenhill, D.E. Blasi, R. Bouckaert and R.D. Gray. 2022. Phylogeographic analysis of the Bantu language expansion supports a rainforest route. *Proceedings of the National Academy of Sciences* 119: e2112853119.
- Koni Muluwa, J. and K. Bostoen. 2012. La diphtongaison dans les langues bantu B70-80 (Bandundu, RDC): typologie et classification historique. *Africana Linguistica* 18: 355-386.
- Kouarata, G.N., S. Pacchiarotti and K. Bostoen. forthcoming. Nouvelle contribution à l'inventaire et la documentation des parlers teke (bantou, B70) en République Démocratique du Congo. *Linguistique et Langues Africaines*.
- Pacchiarotti, S. and K. Bostoen. 2020. The Proto-West-Coastal Bantu velar merger. *Africana Linguistica* 26: 139-195.
- Pacchiarotti, S. and K. Bostoen. 2021. Final vowel loss in Lower Kasai Bantu (DRC) as a contact-induced change. *Journal of Language Contact* 14: 437-474.
- Pacchiarotti, S. and K. Bostoen. 2022. Erratic velars in West-Coastal Bantu: Explaining irregular sound change in Central Africa. *Journal of Historical Linguistics* 12: 1-65.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3), 539–542.
- Rottland, F. 1977. Reflexes of Proto-Bantu Phonemes in Yanzi (B85). *Africana Linguistica* 7: 375-396.
- Seidensticker, D., W. Hubau, D. Verschuren, C. Fortes-Lima, P. de Maret, C.M. Schlebusch and K. Bostoen. 2021. Population collapse in Congo rainforest from 400 CE urges reassessment of the Bantu Expansion. *Science Advances* 7, eabd8352: 1-13.

Bartholomae's law revisited and remodelled (conference abstract)

Bartholomae's Law affects the development of the clusters of $*D^h+T$ and $*D^h+S^1$ in old Indo-Iranian languages.² The progressive principle of the law goes contrary to the prevailing tendency of Aryan clusters to be assimilated in a regressive manner (note that the outcomes of the Law are often levelled on "regular" regressive later).

Since Bartholomae (1882), there have appeared numerous models trying to reconstruct the possible trajectories of the development (cf. for general overview Collinge 1985: 7–11; Mayrhofer 1986: 115–118; Szemerényi 1990: 106–109; Mayrhofer 2004: 46), especially remarkable are models of Anderson (1970), Sag (1974: 593), Mey (1972), Schindler (1976), D. G. Miller (1977), Ejerhed (1981), Lombardi (1991: 140) and Kobayashi (2004: 1117–125), none of them got general acceptance since all models require either atypical shifts of aspiration; either deaspiration of the left member or biphonemic nature of Aryan voiced aspirates.

Our proposed model for the development of D^hT/D^hS clusters follows the trajectory of spirantization and subsequent fortition, that reconstructed IE $*D^h$ had in Aryan value of the voiced spirant Δ was brought first by Walde (1887: 466), though not for the Bartholomae's clusters.

We assume the following trajectories for $*D^hT$ clusters:

- a. the left plosive (= IE $*D^h$) becomes a voiced spirant and the right voiceless plosive ($*T$) also becomes a voiced spirant ($D^h + T > \Delta\Delta$);
- b. in the second phase, both spirants became a subject of fortition to plosives; the left spirant became a voiced plosive, the right spirant changed into a voiced aspirate ($\Delta\Delta > DD^h$)³ in Indic,⁴ the right plosive is non-aspirated plosive in Iranian:

- i. $D^h + T > \Delta\Delta > DD^h$ (Indic)
- ii. $Dh + T > \Delta\Delta > DD/\text{Ð}\text{Ð}$ (Iranian)

NB: The process and its outcomes are similar for the clusters $*TD^h$ and $*D^hD^h$ left aside at this moment.

Similo modo, the trajectories for the development of the $*D^hS$ clusters are modelled:

- a. a voiced aspirate becomes a voiced spirant;
- b. a sibilant becomes voiced;
- c. a ΔZ cluster is despirantized in the left part of it in Iranian; T_s replaces the expected $\dagger DZ$ due to analogy in Indic, the Iranian state is assumed to be archaic, hence:

- i. $D^h + S > \Delta Z (\rightarrow TS)$ (Indic)
- ii. $D^h + S > \Delta Z > DZ$ (Iranian)

NB: The spirantization model of Bartholomae's law has one prominent advantage concerning the development of D^hS clusters: within the spirantization model, there is no need to introduce the "exotic" voiced aspirated sibilants (Z^h) at all.

The Indic outcomes with voiced aspirated plosives are hence not inherited but innovations; the existence of voiced spirants in Iranian hence reflects, in some range, the older Indo-Iranian situation.

Ondřej Šefčík
Masaryk University
Faculty of Arts
Brno
Czechia

¹ IE $ts > OIA ts$, Av. os ; IE $d^h s > OIA ts$, Av. oz ; IE $ss > OIA ts$, Av. os ; IE $k_s > OIA k_s$, Av. $o\check{s}$; IE $g^h s > OIA k_s$, Av. $o\check{z}$; II $ss > OIA k_s$, Av. $o\check{s}$.

² Some authors who consider the law being operating already in IE (Kuryłowicz 1935: 50–51; Lubotsky 2018: 1879), but we support the idea that the Law is exclusively Indo-Iranian (e.g., Szemerényi 1990: 107; Hoffmann/Forsmann 1996: 95–96). Noteworthy is that the possible validity of the Bartholomae's Law for Germanic was examined in recent years, especially by Görtzen (1998: 444–448) and Hill (2003: 218–220).

³ We assume, similarly to D. G. Miller (1977), that voice was a primary quality, not aspiration.

⁴ Either directly due to the same process or later, Walde (1897) assumes aspiration as a later feature both of OIA and Gr. aspirates.

Selected references:

- Anderson, Stephen R. 1970. On Grassmann's Law in Sanskrit. *Linguistic Inquiry* 1(4). 387–396.
- Bartholomae, Christian. 1882. Die arische Vertretung von med. Asp. + t und med. Asp. +s. *Arische Forschungen* I. 3–24. Halle: Niemeyer.
- Beekes, Robert S. P. 1988. *A Grammar of Gatha-Avestan*. Leiden: Brill
- Brugmann, Karl. 1897. *Vergleichende Laut-, Stammbildungs- und Flexionslehre der indogermanischen Sprachen. Zweite Bearbeitung. Erster Band: Einleitung und Lautlehre. Zweite Hälfte*. Strassburg: Karl. J. Trübner.
- Burrow, Thomas. 1955. *The Sanskrit Language*. London: Faber and Faber.
- Cheung, Johnny. 2007. *Etymological Dictionary of the Iranian Verb*. Leiden: Brill.
- Collinge, Neville E. 1985. *The Laws of Indo-European*. Amsterdam/Philadelphia: John Benjamins.
- Curtius, Georg. 1853. Die aspiraten der indogermanischen sprachen. *KZ* 2. 321–336.
- Ejerhed, Eva I. 1981. The Analysis of Aspiration in Sanskrit Phonology. *Nordic Journal of Linguistics* 4(2). 139–159.
- Görtzen, Jens. 1998. *Die Entwicklung der indogermanischen Verbindungen von dentalen Okklusi-ven mit besonderer Berücksichtigung des Germanischen* (Volume 94 of Innsbrucker Beiträge zur Sprachwissenschaft). Innsbruck: Institut für Sprach-wissenschaft der Universität Innsbruck.
- Hoffmann, Karl/Forssman, Bernard. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Kellens, Jean. 1984. *Le verbe avestique*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kellens, Jean. 1995. *Liste du verbe avestique*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kobayashi, Masato. 2004. *Historical Phonology of Old Indo-Aryan Consonants*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa. Tokyo University of Foreign Studies
- Kümmel, Martin J. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.
- Kuryłowicz, Jerzy. 1935. *Études indoeuropéennes I*. Kraków: Skład główny w księgarni Gebethnera i Wolffa.
- Lehmann, Winfred P. 1952. *Proto-Indo-European Phonology*. Austin: The University of Texas Press/Linguistic Society of America.
- Lipp, Reiner. 2009a. *Die indogermanischen und einzelsprachlichen Palatale im Indo-iranischen*. Band I: *Neurekonstruktion, Nuristan-Sprachen, Genese der indoarischen Retroflexe, Indoarisch von Mitanni*. Heidelberg: Universitätsverlag Winter.
- Lipp, Reiner. 2009b. *Die indogermanischen und einzelsprachlichen Palatale im Indo-iranischen*. Band II: *Thorn-Problem, indoiranische Laryngalvokalisation*. Heidelberg: Universitätsverlag Winter.
- LIV² = Rix, Helmut/Kümmel, Martin J./Zehnder, Thomas/Lipp, Reiner/Schirmer, Brigitte. 2001. *Lexikon der indogermanischen Verben*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Lombardi, Linda. 1991. *Laryngeal Features and Laryngeal Neutralization*. Dissertation, University of Massachusetts MacDonell, Arthur A. 1910: *Vedic Grammar*. Strassburg: Karl J. Trübner.
- MacDonell, Arthur A. 1916: *A Vedic Grammar for Students*. Oxford: Clarendon Press.
- Mayrhofer, Manfred. 1986. *Indogermanische Grammatik I.2. Lautlehre*. Heidelberg: Winter.
- Mayrhofer, Manfred. 2004. *Die Hauptprobleme der indogermanischen Lautlehre seit Bechtel*. Wien: Verlag der österreichischen Akademie der Wissenschaften.
- Mey, Jacob L. 1972. Was Bartholomae really a Grassmann? *Norwegian journal of linguistics* 1. 81–89
- Miller, D. Garry. 1977. Bartholomae's Law and an IE Root Structure Constraint. In: Hopper, Paul J. (ed.): *Studies in Descriptive and Historical Linguistics. Festschrift for W. P. Lehmann*. Amsterdam: John Benjamins. 365–392.
- Renou, Louis. 1996³. *Grammaire Sanscrite*. Paris: Librairie d'Amérique et d'Orient. Adrien Maisonneuve
- Sag, Ivan A. 1974: The Grassmann's Law Ordering Pseudoparadox. *Linguistic Inquiry* 5(4). 591–607.
- Sag, Ivan A. 1976: Pseudosolutions to the Pseudoparadox: Sanskrit Diaspirates Revisited. *Linguistic Inquiry* 7(4). 609–622.
- Schindler, Jochem. 1976: Diachronic and Synchronic Remarks on Bartholomae's Law and Grassmann's Law. *Linguistic Inquiry* 7(4). 622–637.
- Szemerényi, Oswald. 1990. *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Wackernagel, Jakob. 1896. *Altindische Grammatik. Band I. Lautlehre*. Göttingen: Vandenhoeck & Ruprecht.
- Walde, Alois. 1897. Die Verbindungen zweier Dentale und tönendes z im Indo-germanischen. *KZ* 34: 461–536.
- Whitney, William D. 1885. *The roots, verb-forms and primary derivatives of the Sanskrit language: a supplement to his Sanskrit grammar*. Leipzig: Breitkopf and Härtel.

Dialectal variation and the Second Sophistic: evidence from the Atticist lexica.

The Atticist lexica contain what their authors believed to be Classical Attic forms next to their alleged Koiné Greek equivalents, and provide the historical linguist with a useful insight into second century CE Greek speakers' perceptions on the Ancient Greek dialects. The lexica showcase the attempt by the educated Greek-speaking elite under the Second Sophistic to provide some sort of Greek standard by rejecting forms from every Greek dialect other than Classical Attic. While their authors often make errors in deciding what constitutes, and does not constitute, Classical Attic, these lexica are nevertheless of great use for the sociolinguist and dialectologist alike, as they demonstrate how Greek speakers of the Postclassical Period perceived, explained and categorised Greek dialectal variation.

The lexicographers show awareness of the Ancient Greek dialects, as they reject the use, in formal written Greek, of all features that they believe are not Attic. This is a topic of considerable concern to these scholars: rejected Ionic forms account for approximately 10% of glosses in Phrynichus' *Ecloga*, and they are all keen to display their knowledge of the dialects, rejecting forms used by 'the Ionic speaker' (e.g. Phrynichus *Ecl.* 156: "ὁ Ἴων"), explaining that certain forms are typical of the Aeolic dialect (e.g. Antiatticist ε79: "Αἰολικῶς") or talking of the common language of Doric, Ionic and Attic speakers (e.g. Moeris δ6: "κοινὸν Δωριέων Ἴωνων Ἀττικῶν"). However, their understanding and interpretation differs to our own modern delineation of the Ancient Greek dialects. Most significantly, they occasionally accept Homer, in addition to the canonised Attic orators and tragedians, as a model for the budding Atticising writer to follow (e.g. Moeris η9; Antiatticist β14).¹ It appears that the concept of dialect in this period is not geographic, but cultural, more akin to register.² For example, Ionic forms are often described as 'poetic',³ which suggests that the forms were linked, in the minds of the lexicographers, to the types of texts in which they were used, rather than to the language of a geographical region of Greece. When they discuss the dialects, they discuss the literary dialects, not how their contemporaries in Ionia and other parts of Greece spoke. Literary and linguistic preferences are subjective, and can often trigger strong feelings, and for this reason the lexicographers, and Phrynichus in particular, writes in a very unsubstantiated way about words he considers to be ἀνάττικον, claiming to be 'unable to endure' ("δυσχεραίνω" (*Ecl.* 32.)) and 'disgusted' ("ἐναυτίασα" (*Ecl.* 172)) by certain 'greatly distasteful' ("ἀηδέες πάνυ" (*Ecl.* 339; *Ecl.* 332.)) forms, some of which are 'so wrong that not even Menander uses [them]' ("οὕτως ἀδόκιμον ὡς μηδὲ Μένανδρον αὐτῶ χρήσασθαι" (*Ecl.* 307)). The latter statement shows that Phrynichus was explicitly aware that certain authors used certain dialects: the presupposition here is that Menander does *not* write in Classical Attic.⁴ Despite these differences, the lexicographers also hold themselves to certain criteria and rules that are familiar to historical linguists when it comes to evaluating dialect usages. For example, they consistently reject uncontracted vowels in favour of their contracted Attic equivalents (e.g. Moeris χ28), geminate -σσ- in favour of Attic -ττ- (e.g. Moeris β25), and cluster -ρσ- in favour of Attic -ρρ- (e.g. Moeris θ20). This raises the question of what similarities and differences we can find between modern day understanding of the dialects, and the perspective of a second century CE grammarian.

This paper therefore proposes to examine the Ancient Greek dialects by investigating evidence of dialectal variation from second century CE users of the language, and exploring what aspects of variation were meaningful to them. The paper will discuss the evidence for attitudes about dialectal variation in Phrynichus' *Ecloga*, Moeris' Lexicon, and the Lexicon of the 'Antiatticist', three lexica which survive to us in a more or less complete form. It will examine how their authors discussed the relationship between the literary dialects, with which they would have been familiar from school, and their own Koiné Greek, and show that the dialect that one used could and did invoke significant para-linguistic associations, notably that of social status, education, and background, a fact that is evident from the very practice and prescription of grammatical Atticism under the Second Sophistic.

¹ This contradicts Swain's (1996: 53) observation that 'all other dialects, including Homeric Greek, are firmly rejected.'

² This was also the case in the Classical Period: cf. Aristophanes (fr. 706), who talks of the διάλεκτος of the πόλις, using this term to refer to what scholars today would call a register, not a dialect.

³ This idea is also found in Strabo 1.2.6 and Hermogenes *On Style* 2. 319f.

⁴ In a similar vein, he (correctly) accuses Herodotus of writing in Ionic ("ἰάζων") in *Ecloga* 101, and Hecataeus of using a particular verb 'because he is an Ionian' ("Ἴων ὄν") in *Ecloga* 198.

Bibliography

- Hajdú, K., & Hansen, D. (1998). *Ps.-Herodian, De figuris / Das attizistische Lexikon des Moeris Überlieferungsgeschichte und kritische Ausgabe / Quellenkritische Untersuchung und Edition Aelius Herodianus*, Vol. 8/9, Berlin/New York: De Gruyter.
- Fischer, E. (1974). *Die Ekloge des Phrynichos*, Berlin/ New York: De Gruyter.
- Swain, S. (1996). *Hellenism and Empire: Language, Classicism and Power in the Greek World AD 50-250*, Oxford: Clarendon Press.
- Valente, S. (2015). *The Antiatticist: Introduction and Critical edition*, Berlin/ Boston: De Gruyter.

January 20th, 2023

Automating Comparative Reconstructions: Case Study in Austronesian and Ongan

Although comparative reconstruction has always been one of the key endeavors of linguistics, there exists no widely accepted method for evaluating its applications (Michalove, 1998). Instead, evaluation is conducted through debate, often spanning decades, as in the case of Altaic, Nostratic, and, more recently, Dene-Yeniseian. Previous attempts to introduce quantitative measures for genetic relatedness are heuristics for estimating similarity, usually either by calculating the average phonetic distance for each putative word-pair (Downey et al., 2008; Kondrak, 2003) or by computing the proportion of cognates between the two wordlists (Chang et al., 2015; Atkinson & Gray, 2003). Since none of these previous attempts engage with diachronic change directly, most researchers agree that, while they are useful when manual reconstruction is not feasible, traditional methods are still the gold standard (Kiparsky, 2015).

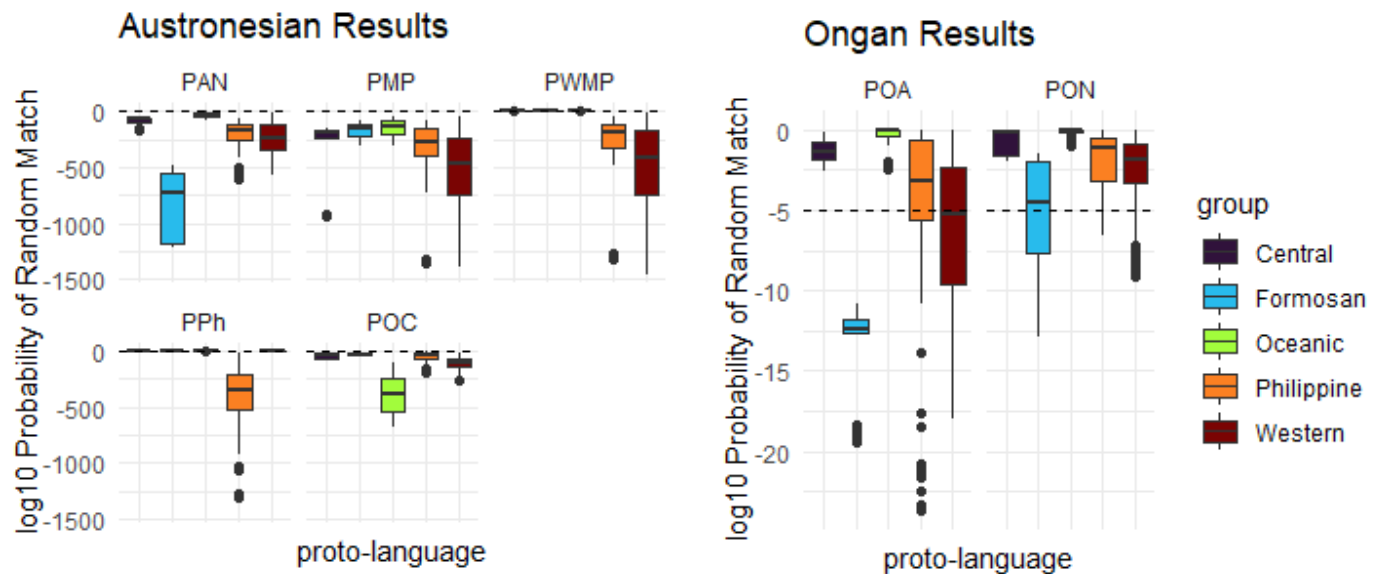
I present a probabilistic framework for evaluating comparative reconstruction attempts. The series of transformations – sound changes, borrowing, semantic change, etc – serves as the input to the framework's evaluation function. The output is the estimated probability that a randomly generated wordlist merits a reconstruction from the mother language using the same number of transformations or fewer than required by the daughter language. Thus, the framework evaluates reconstruction attempts themselves rather than the original dataset, setting it apart from previous quantitative measures.

The framework was incorporated into a simulated annealing learning algorithm, where reconstructions from a mother wordlist to a daughter wordlist were suggested stochastically with a bias toward decreasing the probability of a random match. The algorithm was tested on a genetically diverse sample of Austronesian languages and 5 Austronesian proto-languages. Figure 1 presents the probability of a random match in automated reconstructions from the proto-languages to the 5 Austronesian groups tested, as defined in the Comparative Austronesian Dictionary (Trussel & Blust, 2010). The results are in line with general knowledge in the Austronesian field. For the 237 comparisons between an Austronesian proto-language and a direct descendant, the algorithm always found a reconstruction with a probability of a random match below the chosen cut-off of reliability at .0001. The probability of a random match appears to be strongly correlated with the time depth of the reconstruction.

The case study was further extended to evaluate the putative Ongan-Austronesian connection (Blevins, 2007), a hypothesis not generally accepted in the field (Blust, 2014). Figure 2 presents the probability of a random match in reconstructions from proto-Ongan and proto-Ongan-Austronesian to the 5 Austronesian groups. In reconstructions from proto-Ongan-Austronesian to the Austronesian languages, the results are mixed with the algorithm finding probabilistically non-arbitrary reconstructions to 26 of the 74 of the Austronesian languages tested. Reconstructions from proto-Ongan-Austronesian to the Ongan languages are similarly mixed, with some extremely convincing and others not at all. In general, the results with respect to the Ongan-Austronesian hypothesis appear promising, but not conclusive.

This research is meant to introduce a framework for objective debate surrounding comparative reconstructions and controversial language groupings. The framework can also be used to reason about the comparative method more broadly. For example, the results of the case study reveal that the probability of a random match is mostly determined by the number of borrowings posited, as well the phonotactic complexity of the daughter language. The effect of individual sound changes on reconstruction arbitrariness is measurable but comparatively minor. Future implementations of the framework can be extended to other types of diachronic transformation, e.g. semantic change, morphological change, etc.

January 20th, 2023



Figures 1 & 2 : The log probability that a randomly generated wordlist merits a reconstruction of the same size or smaller than the one generated automatically by a simulated annealing algorithm for 74 Austronesian languages and 5 widely accepted Austronesian proto-languages (Figure 1) and 2 putative Ongan proto-languages (Figure 2). PAN = proto-Austronesian; PMP = proto-Malayo-Polynesian; PWMP = proto-West-Malayo-Polynesian; PPh = proto-Philippine; POC = proto-Oceanic; POA = proto-Ongan-Austronesian; PON = proto-Ongan.

References:

- Atkinson, Q., & Gray, R. (2005). Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology*, 54(4), 513-526.
- Blevins, Juliette (2007). A long lost sister of proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands. *Oceanic Linguistics*, 46(1), 155-198.
- Blust, Robert (2014). Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics*, 53(2), 300-391.
- Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1), 194-244.
- Downey, S., Hallmark, B., Cox, M., Norquest, P., & Lansing, J. (2008). Computational Feature-Sensitive Reconstruction of Language Relationships: Developing the ALINE Distance for Comparative Historical Linguistic Reconstruction. *Journal of Quantitative Linguistics*, 15(4), 340-369.
- Kiparsky, P. (2015). New perspectives in historical linguistics. *The Routledge Handbook of Historical Linguistics*.
- Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities*, 37(3), 273-291.
- Michalove, P., Georg, S., & Ramer, A. (1998). Current Issues In Linguistic Taxonomy. *Annual Review of Anthropology*, 27(1), 451-472.
- Trussel, Stephen & Robert Blust (2010). *Austronesian Comparative Dictionary*.
<https://www.trussel2.com/ACD/introduction.htm>

Sam Mirelman

Research Fellow, Department for the Study of Antiquity, University of Geneva
Sam.Mirelmann@unige.ch

Translation as Royal Legitimation: The Concepts of “Source” and “Target” Language in Sumerian-Akkadian Royal Inscriptions from the Old Babylonian Period (2000–1600 BC)

Bilingual Sumerian-Akkadian royal inscriptions from the Old Babylonian period (2000–1600 BC) are amongst the earliest examples of literary translation in world history. The concept and purpose of translation in this early corpus remains largely unexplored, despite its importance to historical linguistics. For example, in most instances translation involves the transference of meaning from an original “source” language to a “target” language. It remains an open debate whether in Sumerian-Akkadian royal inscriptions from the Old Babylonian period, one language is the original “source” and the other language is the “target”.

Sumerian is perhaps the earliest language recorded in writing. The date of its earliest attestation is disputed, due to the fact that early cuneiform writing is almost entirely logographic. Assuming that such early logographic writing does indeed represent Sumerian, the Sumerian language is attested from the late fourth millennium BC. The period during which Sumerian died out as an everyday language is also disputed, although this development almost certainly took place around the late third millennium BC. From the Old Babylonian period (2000–1600 BC) until the late first millennium BC, translation between Sumerian and Akkadian became commonplace. It is in this period that Sumerian became a clear marker of antiquity and prestige, a function which has been compared to the role of Latin in Medieval Europe. By contrast, Akkadian was the vernacular language of Mesopotamia. However, various dialects and registers of Akkadian may be identified, including a literary dialect.

In the 18th century BC the king of Babylon, Hammurabi, began to compose (or have composed for him) bilingual Sumerian-Akkadian inscriptions (these inscriptions are “virtual” bilinguals, because the Sumerian and Akkadian versions are written on different physical objects, typically clay nails). This is an important phenomenon, as it is certain in this instance that the Sumerian versions of such inscriptions were newly composed for the king, in a period during which Sumerian was certainly a language known exclusively amongst scholars and priests. The Sumerian in such inscriptions features several markers of language contact with Akkadian. Indeed, one may argue that in such inscriptions the Akkadian is the “source” language and the Sumerian is the “target”. However, if this is indeed the case it would go against the expected direction of translation, from the prestige language of Sumerian to the vernacular Akkadian. Indeed, it would suggest that Sumerian functioned as a means of conveying the impression of being a source language, due to its symbolic role as a marker of antiquity and authority. Thus, in the case of this very early example of literary translation, translation does not seem to function as a means of conveying meaning from one language to another. Instead, Sumerian-Akkadian translation functions as a means of conveying prestige and authority, and as a means of asserting royal legitimation.

In the case of Sumerian-Akkadian translation described above, one may arguably use the term “pseudo-translation” (Touy 1995: 40–52), or the concept of language contact through translation (e.g. Kranich 2014). The specific problems involved in the study of written, as opposed to oral, language contact must also be considered (Adams, Janse and Swain eds. 2002; Lavidas 2022).

References

- Adams, J. N., M. Janse and S. Swain (eds.). 2002. *Bilingualism in Ancient Society: Language Contact and the Written Text*. Oxford: Oxford University Press.
- Kranich, S. 2014. “Translations as a Locus of Language Contact.” In J. House (ed.), *Translation: A Multidisciplinary Approach*. London: Palgrave Macmillan, 96–115.
- Lavidas, N. 2022. *The Diachrony of Written Language Contact*. Brill’s Studies in Historical Linguistics 15. Boston/Leiden: Brill.
- Toury, G. 1995. *Descriptive Translation Studies and Beyond* (2nd Edition 2012). Amsterdam: John Benjamins Publishing.

Changes in the Polish address practices after the Second World War

After the Second World War, as a result of the Yalta and Potsdam conferences, Poland fell into the Soviet sphere of influence. The new political reality had a great impact on social life. The sociocultural changes involved radical changes in discourse practices, and in the national discourse system in general (Duszak, 2006; cf. Fairclough, 1992). They also affected Polish politeness: a transition from 'traditional Polish' politeness to 'more egalitarian' politeness (cf. Huszcza, 2005). There were changes in address practices (cf. Betsch, 2019). The aim of the study is to conduct an analysis of the changes in the Polish address practices which occurred within the four decades after the Second World War.

The theoretical framework used in the analysis is a combination of historical sociopragmatics and the Discourse-Historical Approach. Historical sociopragmatics, focusing on the interaction between specific aspects of social context and particular historical language use (Culpeper, 2011: 4), can account for the role of address forms in relation to the discourse in which they appear and the context of situation. The Discourse-Historical Approach (Wodak, 2002; Wodak et al., 2009), which is one of the main approaches to Critical Discourse Studies, is an interdisciplinary, problem-oriented approach which goes beyond the linguistic dimension and includes the historical, political and sociological dimensions in the analysis and interpretation of a specific discursive event (Wodak and Reisigl, 2015: 583; Reisigl, 2018). As such, the DHA could greatly contribute to historical sociopragmatics (Culpeper, 2010: 87).

The data used in the analysis come from Polish films depicting the then reality in Poland, produced in the years 1945-1989.

References

- Betsch, M. 2019. Address forms in grammars and textbooks of West and South Slavic languages in the eighteenth and nineteenth century. In: Paternoster, A., Fitzmaurice, S. (Eds.), *Politeness in Nineteenth-Century Europe*. John Benjamins Publishing Company, Amsterdam, pp. 39-74.
- Culpeper, J. 2010. Historical sociopragmatics. In: Jucker, A. H., Taavitsainen, I. (Eds.), *Historical Pragmatics*. Walter de Gruyter, Berlin/New York, pp. 69-94.
- Culpeper, J. 2011. Historical sociopragmatics. In: Culpeper, J. (Ed.), *Historical Sociopragmatics*. John Benjamins, Amsterdam/Philadelphia, pp. 1-8.
- Duszak, A. 2006. Why 'New'Newspeak?: axiological insights into language ideologies and practices in Poland. In: Mar-Molinero, C., Stevenson, P. (Eds.), *Language Ideologies, Policies and Practices*. Palgrave Macmillan, London, pp. 91-103.

Fairclough, N. 1992. *Discourse and Social Change*. Polity Press, Cambridge.

Huszcza, R. 2005. Politeness in Poland: from 'titlemania' to grammaticalised honorifics. In: Hickey, L., Steward, M. (Eds.), *Politeness in Europe. Multilingual Matters*, Clevedon/Buffalo, pp. 218-233.

Reisigl, M. 2018. The discourse-historical approach. In: Flowerdew, J., Richardson, J. E. (Eds.), *The Routledge Handbook of Critical Discourse Studies*. Routledge, London, pp. 44-59.

Wodak, R. 2002. The discourse-historical approach. In: Wodak, R., Meyer, M. (Eds.), *Methods of Critical Discourse Analysis*. Sage Publications, London, pp. 63-94.

Wodak, R., R. De Cillia, M. Reisigl, K. Liebhart, 2009. *The Discursive Construction of National Identity*. Edinburgh University Press, Edinburgh.

Wodak, R., M. Reisigl, 2015. Discourse and racism. In: Tannen, D., Hamilton, H. E., Schiffrin, D. (Eds.), *The Handbook of Discourse Analysis*. Wiley Blackwell, Chichester, pp. 576-596.

A case of *Verschärfung* in the Swedish dialect from Stora Rågö in Estonia

Ida Västerdal, PhD student, Gothenburg University, Sweden

From the 13th century until the 1940's, the northwestern coast of Estonia was inhabited by a Swedish population, speaking diverse Swedish dialects with many peculiar linguistic traits. There, on the small island of Stora Rågö, the local dialect was characterized by the insertion of consonant clusters *-ggv-* or *-ggj-* between vowels: e.g. *fruggver* 'wives' (singular *fru*), *liggjer* 'scythes' (singular *li*). Clearly, these stop insertions are almost identical in form to the insertions caused by the sound law known as *Verschärfung*, appearing in Old Norse and Gothic, as well as in modern Faroese. A study presented in Västerdal (2021) show that there are also close similarities regarding phonetic context and conditions for insertion to take place. It is therefore claimed that the stop insertions in the dialect of Stora Rågö (SR) is a fourth case of Germanic *Verschärfung*.

The term *Verschärfung*, or *Holtzmann's law*, describes a Proto-Germanic sound change found in Old Norse (ON) and Gothic (goth.). The sound change involved glides that developed into *-ggv-/-ggj-* in ON and *-ggw-/-ddj-* in goth., e.g. ON *tryggr*, goth. *triggws* 'trustworthy'; ON *tveggja*, goth. *twaddje* 'two' (Petersen 2002).

A later case of *Verschärfung* occurred in modern Faroese, where it is called *skerping*. In Faroese, stop insertions *-gv-* and *-ggj-* first appeared between two vowels in disyllabic words, e.g. *rógva* 'row', *nýggjur* 'new', but subsequently spread to monosyllabic words (Thráinsson et al. 2004).

Besides the similarities in the form of the stop insertions, there are also significant similarities regarding phonetic context between the stop insertions in SR and Faroese. In the latter, *-gv-* follows after the diphthongs /uu, ou/ and *-ggj-* after /ei, ui, ai, oi/, which are all monophthongized in this context (Thráinsson et al. 2004). In SR, *-ggv-* follows after /u, ʊ/ or a diphthong ending in /ʊ/. Equivalently, *-ggj-* follow after /i/ or a diphthong ending in /i/. All diphthongs preceding insertions are monophthongized before *-ggv-* and *-ggj-* (Västerdal 2021). At times, the similarities between SR and Faroese (Far.) are striking: e.g. SR *sjoggvín*, Far. *sjógvin* 'the sea'; SR *biggjar*, Far. *bíggjar* 'villages'. The results of the *Verschärfung* development in SR and Faroese seem closely related, yet there are no recordings of any contact between inhabitants of Stora Rågö and the Faroe Islands. Therefore it must be presumed that *Verschärfung* has developed spontaneously in both varieties, just as it has already been suggested that *skerping* in modern Faroese is a similar but separate development from the Old Norse *Verschärfung*.

In this presentation, the SR stop insertions and their phonetic context will be described and compared to the Old Norse, Gothic and Faroese equivalents. There will also be a brief discussion on the possible origins as well as the social and historical context that tentatively contributed to the development of *Verschärfung* in SR.

References

- Petersen, Hjalmar P. 2002. *Verschärfung* in Old Norse and Gothic. In: *Arkiv för nordisk filologi* 117, pp. 5–26. Lund: Arkiv för nordisk filologi.
- Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen & Zakaris Svabo Hansen. 2004. *Faroese. An overview and reference grammar*. Tórshavn: Føroya Fróðskaparfelag.
- Västerdal, Ida. 2021. *Verschärfung* in the Estonian-Swedish dialect of Stora Rågö. In: *Arkiv för nordisk filologi* 136, pp. 209–230. Lund: Arkiv för nordisk filologi.

Parallel Phases in the History of French

This talk will present new evidence from the history of French to evaluate the Parallel Phases Hypothesis. A fundamental issue in historical syntax concerns whether certain changes cluster and why. One recent proposal, the Parallel Phases Hypothesis (Poletto 2006; 2014; 2015), suggests that phases – semi-autonomous units of syntactic structure – change in a uniform fashion. Based on evidence from Old Italian, Poletto shows that head and phrasal movement are lost concomitantly at the CP, ν P, and DP edges, providing an apparently uniform account for the loss of Verb Second, scrambling, and a number of word-order operations licensed in the Old Italian nominal expression.

This talk will assess whether the predictions of the hypothesis extend to a closely related language, French; this is a necessary first step in assessing the validity of the hypothesis as, if it holds up crosslinguistically, the hypothesis has the potential to shed light on why, in a number of languages, several major word-order patterns change quite radically in a short period of time, yielding modern languages with significantly distinct typological footprints to their early counterparts (see, for example, Walkden 2014 on early Germanic, Wolfe 2018 on early Romance, and Willis 2007 on early Celtic).

Examination of Old and Middle French data from the *Base de Français Médiéval*, supplemented with a hand-annotated corpus of Renaissance and Classical French texts reveals that – as per the predictions of the hypothesis – Verb Second-related movement to the CP, object scrambling to the ν P edge, as well as widespread Adjective-Noun orderings and N-to-D movement to the DP edge are all lost in tandem as fully productive operations between 1400 and 1525. Moreover, the corpus data suggest that Poletto's hypothesis extends further to the discourse-pragmatic and syntactic features associated with the phrasal elements undergoing movement: while a wide range of objects can be focalised or topicalised to both the CP and ν P edge in Early Old French, by 1225 such objects show a strong tendency to either be focal or discourse 'old'. Moreover, in Middle French texts, focal objects at both the CP and ν P edges are heavily restricted, but QPs remain stable in both contexts. More generally, we suggest that at the CP, ν P, and DP edge, reanalysis of a maximally general movement operation proceeds first as a movement operation restricted by discourse-pragmatic status, which in turn is reanalysed as an operation where only categories bearing particular categorial features may move (e.g. [+Focus] > [+Q]). This progression is observable in the loss of Verb Second, the loss of scrambling, and the restrictions on adjective movement to the left periphery of the extended nominal expression.

The talk will conclude with the proposal that the Parallel Phases Hypothesis can be subsumed within the Input Generalization Principle proposed by Roberts (2007). Under this approach, the acquirer – unless receiving strong evidence to the contrary – assigns similar featural makeup to syntactic heads perceived as forming a uniform class. In the case under examination the class in question is that of phase-heads.

Giusti, G. 2015. *Nominal syntax at the interfaces: a comparative analysis of languages with articles*. Newcastle: Cambridge Scholars • Poletto, C. 2006. Old Italian Scrambling: The Low Left Periphery of the Clause. In P. T. Brandt & E. Fuss (eds), *Form, Structure, and Grammar: A Festschrift Presented to Günther Grewendorf on Occasion of His 60th Birthday*, 209–229. Berlin: Akademie • Poletto, C. 2014. *Word Order in Old Italian*. Oxford: OUP • Roberts, I. 2007. *Diachronic syntax*. Oxford: OUP • Walkden, G. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: OUP • Willis, D. 2007. Historical syntax. In *The syntax of Welsh*, 286–337. Cambridge: CUP • Wolfe, S. 2018. *Verb Second in Medieval Romance*. Oxford: OUP.

Abstract for ICHL 26 in Heidelberg, 4–8 Sept. 2023

Kerstin Roth (TU Dresden)

Rhetoric, stylistic and argumentative strategies of German language female authors in the 17th century

In the history of the German language, the 17th century is known for its so called *Sprachgesellschaften* and their aim to evolve and practice a well-formulated and written German language. This can be seen as an unofficial start for the standardization of the language, followed by the production of dictionaries and other literature that had the aim of showing how one could and should write in “good” German.

One question, however, is if and how women participated in this evolution? Who were they and why did they use the German language as they did?

This presentation focusses on different female authors from the 17th century who explain in their texts why they wrote in German, whom they addressed and which impact on the evolution of the German language this could have had.

Rhetoric, stylistic and argumentative strategies will build the center of this presentation. This is why the focus will lie on so called *Paratexts*, yet poems or letters in which the vernacular German is discussed will also be treated. *Paratexts* are texts that introduce a work to the addressees. In these texts (introductions, poems, etc.) the authors explain their aims. We like to show that the choice of German is a choice in the direction of an interested, but not (always) academically educated audience. The use of the vernacular German can also be seen as an approach to evoke intimacy between the authors and their public. The choice of the language is a choice between close and distant (vernacular vs. f. ex. Latin) or between a non-academic and an academic world.

Examples for this approach are texts from the scientists Maria Cunitz and Maria Sibylla Merian along with the poets Sophia Elisabeth Brunner and Johanne Charlotte Unzer. This study is based on results of cooperative research with researchers from the field of early modern literature. We intend to show that interdisciplinary work is needed to fully understand the language and literature of the 17th century.

References

Corinna Dziudzia & Sonja Klimek (2022) (Hgg.): *Gelehrte Frauen der Frühaufklärung. Einsame ‚Wunderthiere‘ oder vernetzte Akteurinnen?* Wiesbaden.

Claudia Opitz & Ulrike Weckel (1998) (Hgg.): *Ordnung, Politik und Geselligkeit der Geschlechter im 18. Jahrhundert.* Göttingen.

Kerstin Roth & Katharina Worms (t. b. in print): Selbstkonzeptionen von Autorinnen der frühen Neuzeit in ihren Paratexten. Sophia Elisabeth Brenner, Maria Cunitz, Maria Sibylla Merian und Johanne Charlotte Unzer. In: Peter Heßelmann (Hg.): *Simpliciana 2023.*

Katharina Worms (t. b. in print): „die Freyheit, von Liebe und Weine zu scherzen“: Rhetorik der Emanzipation in Johanne Charlotte Unzers anakreontischer Lyrik. In: Annina Klappert (Hg.): *Textfiguren der Emanzipation. Autorinnen des europäischen 17. und 18. Jahrhunderts wieder (und wieder) lesen.* Hannover.

Early New High German preposed adverbial clauses: integration and discourse functions

Barthe Bloom, Universität Potsdam, bloom@uni-potsdam.de

The study investigates the use of preposed adverbial clauses in Early New High German (henceforth: ENHG) narratives and evaluates whether the choice between integration, resumption and juxtaposition is determined by local and global discourse factors.

In ENHG, preposed adverbial clauses are positioned either i) before an argument, as in (1), ii) adjacent to a resumptive element, like in (2) and (3); or iii) directly adjacent to the finite verb of their host sentence, exemplified in (4) (König and van der Auwera, 1988).

- (1) Da der swartz ritter das gewar wart, **er** greiff yne mit dem helm
lit. 'And when the black knight noticed that, HE GRABBED him by the helmet.' (Pontus)
- (2) ob es in die warheit nit gesaht het, **so** muß es sterben
lit. 'If it hadn't told him the truth, SO MUST IT die.' (Tristrant)
- (3) vnd da alle ding bereyt waren **da** gieng sie zû dem Peter
lit. 'And when all things were ready, THEN WENT SHE to Peter.' (Magelone)
- (4) Do der künig das verstûnde sprach er zum graffen
lit. 'When the king understood that, SPOKE HE to the count.' (Fortunatus)

The four constructions illustrated above represent different degrees of integration, reaching from simple juxtaposition (1) to resumption (2) and (3) to full integration (4). This continuum is thought to be reflective of the diachronic development of the position of adverbial clauses (König and van der Auwera, 1988; Axel, 2004; Lötscher, 2006), i.e., from juxtaposed via resumption to complete integration.

This diachronic scenario predicts that at least one of the resumptive constructions was paradigmatically related to the integrating construction (Zehentner, 2019) at the time the use of the integrating construction was taking over, i.e., in ENHG (Axel, 2004). Moreover, it is expected that the juxtaposition construction has been ousted by resumption at this time, and found its own niche.

To test these predictions, a data set of 1500 preposed adverbial clauses from seven ENHG narratives – Pontus, Melusine, Wigalois, Wilhelm, Tristrant, Fortunatus, and Magelone – is investigated. The data indicate that the two resumptive constructions are associated with distinct discourse functions, and that resumption with *so* does not show systematic similarity to the integration construction.

First, the constructions are found to differ with regard to local discourse functions. For example, concessives and conditionals are normally resumed by *so*, whereas the temporal adverbials occur primarily in the integrating construction or with *da*. Juxtaposition has not developed its own niche here.

Second, the constructions' global discourse functions are compared. Specifically, it is hypothesized that the resumptives are associated with different narrative speed (Genette, 1983; Packard, 2008). The *so*-construction is relatively frequent in direct speech segments compared to the other three patterns. As such, it is associated with scenes, which are characterized by isochrony. In contrast, *da*-resumption likely condenses more story time into less narration, since the adverbial clauses that occur in this construction often summarize earlier events.

How this relates to the use of integration and juxtaposition presents the final part of the puzzle. In order to support the diachronic development, the juxtaposition construction should have a specialized function, while the integrating construction is expected to show a tight similarity to one of the resumptive constructions.

Primary sources

Fortunatus (1509). Schmitt, L. E. and R. Noll-Wiemann (eds.) 1974. *Fortunatus: Von Fortunato und seynem Seckel auch Wünschhütlein*. Hildesheim/New York: Georg Olms Verlag.

Magelone (1535). Müller, J.-D. 1990. *Romane des 15. und 16. Jahrhunderts. Nach den Erstdrucken mit sämtlichen Holzschnitten*. Frankfurt am Main: Deutscher Klassiker Verlag.

Melusine (1474). Müller, J.-D. 1990. *Romane des 15. und 16. Jahrhunderts. Nach den Erstdrucken mit sämtlichen Holzschnitten*. Frankfurt am Main: Deutscher Klassiker Verlag.

Pontus (1450-1500). Schneider, K. 1961. *Pontus und Sidonia: in der Verdeutschung eines Ungenannten aus dem 15. Jahrhundert / kritisch hrsg. von Karin Schneider*. Berlin: E. Schmidt.

Tristrant (1484). Elsner, H. (ed.). 1989. *Tristan und Isolde (Augsburg bei Antonius Sorg 1484) mit einem Nachw. von Helga Elsner*. Hildesheim: Georg Olms Verlag.

Wigalois (1519). Melzer, H. 1973. *Wigalois / mit eine Vorw. von Helmut Melzer*. Hildesheim: Georg Olms Verlag.

Wilhelm (1481 & 1491). Podleiszek, F. 1936. *Volksbücher von Weltweite und Abenteuerlust / bearbeitet von Dr. Franz Podleiszek*. Leipzig: Verlag von Philipp Reclam jun.

References

Axel, K. (2004). The syntactic integration of preposed adverbial clauses on the German left periphery: A diachronic perspective. In H. Lohnstein and S. Trissler (Eds.), *The Syntax and Semantics of the Left Periphery*, pp. 23–58. Berlin: Mouton de Gruyter.

Genette, G. (1983). *Narrative discourse: An essay in method*, Volume 25. Ithaca: Cornell University Press.

König, E. and J. van der Auwera (1988). Clause integration in German and Dutch conditionals, concessive conditionals, and concessives. In J. Haiman and S. Thompson (Eds.), *Clause combining in grammar and discourse*, pp. 101–133. Amsterdam: John Benjamins.

Lötscher, A. (2006). Linksperiphere Adverbialsätze in der Geschichte des Deutschen. Pragmatische Aspekte eines grammatischen Wandels. *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 127(3), 347–376.

Packard, S. (2008). Two-dimensional narrative velocity. *Journal of Literary Semantics* 37(1), 55–67.

Zehentner, E. (2019). *Competition in Language Change: The Rise of the English Dative Alternation*. Berlin: De Gruyter Mouton

Adverbs ending in *-(l)ig* ‘-ly’ and *-(l)igt* ‘-ly’ in Danish

In Modern Standard Danish, adverbs based on adjectives typically end in a *-t*, cf. *gør det omhyggeligt!* ‘do it carefully’, where the adverb *omhyggeligt* ‘carefully’ is based on the adjective *omhyggelig* ‘careful’. In the 19th century, this was not so; back then, adverbs based on adjectives would typically occur without such a *-t*, e.g. *hun indøvede stykket omhyggelig* ‘she practiced the play carefully’.

On the expression side, the adverbs without a final *-t* as well as the ones ending in a *-t* coincide with adjectival gender forms (e.g. *omhyggelig-Ø* ‘careful-C’ vs. *omhyggelig-t* ‘careful-N’). Therefore, they have often been analyzed as adjectives used as adverbs (e.g. Mikkelsen 1911; Diderichsen 1946). However, following Skautrup (1947, 1953) and Brink (2018), we consider them adverbs derived from adjectives (Schack & Jensen, *subm.*; Jensen & Schack, *subm.*). Historically, adjective-based adverbs with and without final *-t* are preceded by other derivational forms, most notably ones with the suffix *-e* (e.g. *listige* ‘cunningly’, *grumme* ‘badly, ugly’) and – if the adjective itself ends with the derivational suffix *-(l)ig* – by forms with the suffix *-(l)igen* (e.g. *retteligen* ‘properly’, *lystigen* ‘merrily’).

During the 17th and 18th centuries, adverbs ending in *-e* and in *-(l)igen* recede and give way to adverbial forms with no particular expression marker (e.g. *omhyggelig*) and later still to the adverbial forms ending in *-t* (e.g. *omhyggeligt*).

This diachronic outline concerns all adverbs based on adjectives, e.g. derived adjectives with the suffixes *-bar*, *-som*, (*mærkbar* ‘tangible’, *langsom* ‘slow’), compositional adjectives with the highly productive elements *-fuld*, *-løs* (*sorgfuld* ‘mournful’, *tankeløs* ‘thoughtless’), as well as simplex adjectives without a suffix (*smuk* ‘beautiful’, *streng* ‘strict’). However, in the literature on Danish, most attention has been devoted to adverbs ending in *-(l)ig*.

The adjectival suffix *-lig* is a cognate to the English adverbial suffix *-ly*. In the course of history, a number of adjectives historically with other endings have joined the adverbs ending in *-lig* (cf. Falk & Torp 1900; Skautrup 1947, 1953). In grammars of Modern Danish, adverbs ending in *-lig(t)* and *-igt(t)* are traditionally treated as one issue, and we follow this tradition.

In this paper, we present a study of the change from *-(l)ig* to *-(l)igt* in recent history. The study is based on two corpora, both representing written language. One corpus consists of literature and letters from the 19th century; the other of texts from the year 2017.

The study documents how forms without *-t* are replaced by forms with a final *-t*. The study also reveals that the replacement does not take place at an even pace in all adverbial functions. Manner adverbs without *-t* are significantly more frequent in the 19th century than forms without *-t*; in 2017, manner adverbs exclusively occur with *-t*. Temporal adverbs change from forms with *-t* being much more frequent than forms without *-t* in the 19th century to the opposite situation in 2017. As regards adverbs of degree, in the 19th century, they almost exclusively occur without *-t*; in 2017, the ratio of forms with and without *-t* is fifty-fifty.

This pattern supports ideas of adverbs being a morpho-syntactic more diverse lexical class than traditionally assumed.

References

- Brink, L. (2018): Adjektiv eller adverbium? *Nordiske Studier I Leksikografi* 14, 56–62.
- Diderichsen, P. (1946): *Elementær Dansk Grammatik*. København: Gyldendal.
- Falk, H. & A. Torp (1900): *Dansk-norskens syntax i historisk fremstilling*. Kristiania: Aschehoug & Co.
- Jensen, E.S. & J. Schack (subm.): Ubøjede adverbialer uden og med *-t* og historien bag. In: *Ny forskning i grammatik* 30.
- Mikkelsen, K. (1911): *Dansk ordføjninglære med sproghistoriske tillæg*. København: Lehmann & Stages Forlag.
- Schack, J. & E.S. Jensen (subm.): Uden og med *-t*. Adverbialer dannet af adjektiver på *-ig* og *-lig*. In: Schoonderbeek mfl. (red.). *19. Møde om Udforskningen af Dansk Sprog (MUDS)*.
- Skautrup, P. (1947): *Det danske sprogs historie. Bind 2*. København: Gyldendal.
- Skautrup, P. (1953): *Det danske sprogs historie. Bind 3*. København: Gyldendal.

Adjectival typology in four ancient Indo-European languages

Scholars usually agree that the part of speech (PoS) system of Proto-Indo-European (PIE) is similar to that in Latin, Greek, Sanskrit and Hittite and, in each case, it is based on three major classes: nouns, verbs and adjectives or schematically [N, A, V] (implicitly, Nussbaum 2022). Still, adjectives show the same endings as nouns, comparative morphology is lacking in Hittite, Tocharian, Armenian and Albanian, and most PIE adjectives show the “recent” inflection in **-e/o-*. Thus, many scholars argued that the adjective is a recent category (Wackernagel 2009: 466, Lehmann 1974: 208, Comrie 1997: 101): that is, PIE or, at least, pre-PIE could be a language “without adjectives” or “with only two major lexical classes”, nouns and verbs (i.e. [N, V]). However, this idea was further developed into two opposite directions: some scholars claimed that quality concepts were merged with nouns in PIE and, thus, PIE was a language with “noun-like adjectives” or a language of type [(NA) V] (Balles 2006, 2009); others claimed that quality concepts were merged with verbs in PIE and, thus, PIE was a language with “verb-like adjectives” or a language of type [N (AV)] (Alfieri 2016, 2021; Bozzone 2016). In the former scenario, the adjective class arose from appositional nouns added to the feminine motion: i.e. **[noun]-ieh₂/ih₂-Agr*, where Agr means “agreement” (Brugmann 1888: 420; Fritz & Meier-Brügger 2020: 225); in the latter, it arose from derived nouns built on verbal roots of (nearly) quality meaning added to the feminine motion: i.e. **[verb-NM]-ieh₂/ih₂-Agr*, where NM means “nominalizer”. The talk aims to show that, if a functional-typological definition of the adjective is accepted, the latter view is more plausible than to the former.

The starting point for this research is Croft’s definition of PoS (2001: 67ff.). In his view, the “adjective” is not a language-specific, formal class, but a zone of cognitive space defined in terms of semantics and syntax. In practice, the “adjective” is defined as the most typical Quality Modifier construction that is found in any language. This definition is applied to 4 ancient IE languages, namely Rig-Vedic Sanskrit, Homeric Greek, Classical Latin and Hittite. Thus, a sample of texts is gathered for each language – 51 hymns of the *Rig-Veda*, the first book of Homer’s *Iliad* and *Odyssey*, Sallust’s *De coniuratione Catilinae*, and an anthology of Hittite texts –, all the Quality Modifiers in each text are collected (between 800 and 1000 in each language) and their internal structure is analysed. The results are the following (the data on RV Sanskrit and Homeric Greek come from Alfieri 2016, 2021, Alfieri & Gasbarra 2021, while the data on Latin and Hittite are presented here for the first time).

Basically, the same six construction types code the ADJECTIVE slot in each language, namely (RV Sanskrit is quoted as an example for all IE languages): 1) the simple adjective or [adjective]-Agr, i.e. Skt. *kr̥ṣṇá-* ‘black’; 2) the deverbal adjective or [verb-NM]-Agr, i.e. Skt. *tap-ú-* ‘hot’, *mah-ánt-* ‘big’ < *tap-* ‘become/make hot’, *mah-* ‘be/make big’; 3) the denominative adjective or [noun-ADJ]-Agr, i.e. Skt. *pitr-īya-* ‘paternal’ < *pitár-* ‘father’, where ADJ means “adjectivalizer”; 4) the prepositional adjective or [preposition-ADJ]-Agr, i.e. Skt. *paramá-* ‘most distant’ < *parā* ‘away’; 5) the prefixed adjective that is, a nominal stem attached to a prefix or PRE-[...]-Agr, i.e. Skt. *su-vīra-* ‘having good heroes’ < *vīra-* ‘hero’; 6) the compound adjective or [...]_N-[...]_N-Agr, i.e. Skt. *hiraṇya-pāṇi-* ‘golden-palmed’. However, the frequency of each construction type is far different from a language to another (Tab. 1):

| | RV Skt. | Hom. Gk. | Hittite | Latin |
|---|---------|----------|---------|-------|
| [adjective]-Agr | 7.6% | 48.1% | 65.7% | 80.4% |
| [verb-NM]-Agr | 45.9% | 12.9% | 25.7% | 6.0% |
| [noun-ADJ]-Agr | 10.4% | 10.5% | 0.8% | 6.5% |
| [preposition]-Agr | 2.1% | 0.3% | 7.9% | 0.5% |
| Pre-[...]-Agr | 14.4% | 12.1% | 0% | 6.5% |
| [...] _N -[...] _N -Agr | 19.9% | 15.6% | 0% | 0% |

Tab. 1 shows that the most frequent Quality Modifier construction is [adjective]-Agr in Latin, Homeric Greek and Hittite, which are *specialized* languages of type [N, A, V] in Hengeveld's terms (1992), while it is [verb-NM]-Agr in RV Sanskrit, which is a *rigid* language with verb-like adjectives in Hengeveld's terms and falls into type [N (AV)]. The easiest way to interpret the difference between RV Sanskrit and the remaining IE languages is to claim that PIE was a language "without" adjectives or a language of type [N (AV)], in which quality concepts were coded "verbally" and the most typical "adjective" was *[verb-NM]-Agr; this PoS system is preserved in RV Sanskrit bar minor changes, while a previously neglected typological change of type [N (AV)] → [N, A, V] came about in the prehistory of Greek, Latin and Hittite, although the change occurred independently and with a different timing in each branch of the IE family.

- Alfieri, Luca. 2021. Parts of speech comparative concepts and Indo-European linguistics. Alfieri L., Arcodia G. & Ramat P. (eds.), *Linguistic categories, language description and linguistic typology*, 313-366. Amsterdam-Philadelphia: John Benjamins.
- , 2016. The typological definition of the (apparently historical) notion of root. *Archivio Glottologico Italiano* 102(1): 129-169.
- & Gasbarra, Valentina. 2021. The adjective class in Homeric Greek and the parts of speech change in the Indo-European languages. L. Biondi, F. Dedé & A. Scala (eds.), *Change in Grammar. Triggers, Paths and Outcomes*, 7-26. Alessandria: Ed. dell'Orso.
- Balles, Irene. 2009. Zu den i-stämmigen Adjektiven des Lateinischen. Lühr R. & Ziegler S. (eds.): *Protolanguage and Prehistory*. Akten der XII. Fachtagung der Indogermanischen Gesellschaft, Krakau, 11.-15. Oktober 2004. Wiesbaden 2009.
- , 2006. *Die altindische Cvi-Konstruktion: Form, Funktion, Ursprung*. Bremen: Hempen.
- Bozzone, Chiara. 2016. The Origin of the Caland System and the Typology of Adjectives. *Indo-European Linguistics* 4: 15-52.
- Brugmann, Karl & Delbrück, Bertold. 1887-1916. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. Vol. II.1: Brugmann K. 1888 [1906²]. *Lehre von den Wortformen und ihrem Gebrauch. Das Nomen*. Strassburg: Trübner.
- Comrie, Bernard. 1997. La famiglia linguistica indoeuropea: prospettive genetiche e tipologiche. A. Giacalone Ramat & P. Ramat (eds.), *Le lingue indoeuropee*, 95-122. Bologna: Il Mulino.
- Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Fritz, Matthias & Meier-Brügger, Michael. 2021¹⁰. *Indogermanische Sprachwissenschaft*. Berlin-Boston: Mouton De Gruyter.
- Hengeveld, Kees. 1992. *Non-verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton.
- Lehmann, Winfried P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Nussbaum, Alan J. 2022. Derivational properties of "adjectival roots" (expanded handout). Malzahn M., Fellner H.A. & Illés T. (eds.), *Zurück zur Wurzeln. Struktur, Semantik und Funktion der Wurzel im Indogermanisch*. Akten der 15. Fachtagung der Indogermanischen Gesellschaft vom 13. bis 16. September 2016 in Wien, 205-224. Wiesbaden: Reichert.
- Wackernagel, Jakob. 2009. *Lectures on Syntax: with Special Reference to Greek, Latin, and Germanic*. Translated and edited with notes and bibliography by David Langslow. Oxford: Oxford University Press [Id. 1920-1924¹. *Vorlesungen über Syntax mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch*. 2 vols. Basel: Birkhäuser].

The periodization of the Pre-Classical French through the study of *nennil* and *non* in grammars, remarks and treatises (15th-17th centuries)

Key words

Pre-Classical French, Periodization, Historical Pragmatics, Negation markers, Remarqueurs

Abstract

The periodization of Pre-classical French is relatively recent compared to the other generally accepted periods for French (Medieval, Classical and Modern French). Based on morphosyntactic criteria, Combettes (2003) isolates the period 1500-1650 (Preclassical French), where the language does not function as in the 15th century (Middle French), nor as in the late 17th century (Classical French). Combettes and Marchello-Nizia (2010) revisit the periodization of Pre-Classical French, taking 1550 as the initial boundary and 1650 as the final boundary. More recent studies (Ayes-Bennett & Caron 2016; Amatuzzi *et al.* 2020) have analysed the accuracy of the commonly accepted final boundary (1650), identifying a break around 1620-1630 which implies the culmination or the beginning of a number of morphosyntactic changes. Only in a decade, the French language underwent an upheaval that can be identified in oral corpora and which was also perceived in French language treatises, remarks and grammars of the time, as demonstrated by Ayres-Bennett and Caron (2016).

The aim of this presentation is to contribute to the reflection on the periodization of Pre-Classical French by analysing the evolution of the use of the disagreement markers *nennil* and *non* between 1450 and 1700. *Nennil*, the disagreement marker complementary to the agreement marker *oïl/ouy* of Medieval French, disappears during this period in favour of *non*, which will be paired with *oui* from Classical French onwards. After a brief presentation of the distribution and the pragmatic and interactional functioning of the markers *nennil*, *non* and *non* + verb substitute (*non feray*, *non est*), I will analyse the comments made about these markers in the 33 publications collected in the *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue (XIVe-XVIIe s.)* (Colombat & Fournier 2011).

Remarks such as those made by Vaugelas (1647) and Mauger (1659) will help me to date the transition from the use of medieval *nennil* to modern *non*, especially with regard to the complementarity with *oui* and the language level.

« [*&c* eft bien placé] apres *oïy*, & *non*, comme *Oïy Madame, Non Madame, il ne fe voit rien &c.* »

[*&c* is well placed after *ouy* and *non*, as in *Oïy Madame, Non Madame, il ne fe voit rien &c*]
(Claude Favre de Vaugelas, 1647, *Remarques sur la langue française utiles à ceux qui veulent bien parler et bien écrire*)

« *Nenny*, eft vn mot vulgaire; & jamais on ne s'en fert en écriuant; tant il eft de bas alloy. »
[*nenny* no is a vulgar word and is only used in writing because it has little value]

(Claude Mauger, 1659, *Grammaire française avec des augmentations*)

My results will support the conclusions of other authors on the date of transition from Pre-Classical to Classical French between 1620 and 1630. The study of pragmatic phenomena is innovative in the sense that periodization research generally analyses phonetic and/or morphosyntactic changes, at least for French. Also, this study of a corpus of grammars, remarks and treatises of the French language of the period complements other quantitative works.

References

- Amatuzzi, Antonella, *et al.* 2020. « Changement linguistique et périodisation du français (pré)classique: deux études de cas à partir des corpus du RCFC », *Journal of French Language Studies* 30, 301-326.
- Ayres-Bennett, Wendy & Philippe Caron, 2016. « Periodization, Translation, Prescription, and the Emergence of Classical French », *Transactions of the Philological Society* 114(1), 339-390.
- Colombat, Bernard & Fournier, Jean-Marie (dir.). 2011. *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue française*, Paris, Classiques Garnier Numérique.
- Combettes, Bernard (ed.). 2003. *Évolution et variation en français préclassique : études de syntaxe*, Paris, Champion.
- Combettes, Bernard & Christiane Marchello-Nizia, 2010. « La périodisation en linguistique historique : le cas du français préclassique », in : Combettes, Bernard. *et al.* (eds.), *Le Changement en français*, Berne, Peter Lang, 129-41.

Just a bystander? Semantic change in the English simple tenses

Juliette Kayenbergh, Hendrik De Smet

As the English progressive construction *BE Ving* grammaticalized from a stylistically employed device without clear grammatical meaning into an obligatory grammatical aspect marker expressing ongoingness, incompleteness and dynamicity (e.g. Kranich 2010), it entered into a paradigmatic relation with the simple form. Accordingly, Bybee (1994) argues that when the progressive grammaticalized, the zero-marked simple form became obligatorily associated with the expression of non-progressive meaning – a process she refers to as “the grammaticization of zero”. However, the timing of obligatorification of *BE Ving* for the expression of progressive meaning remains unclear. Indeed, Kranich (2010: 165) points out that the history of the simple tense forms has been largely ignored in the literature, leaving a major blind-spot in our understanding of the development of aspectual marking in English. The aim of our research is to investigate the impact of the rise of *BE Ving* on the meanings of the simple forms in order to test Bybee’s (1994) hypothesis.

The hypothesis is tested against the history of the simple forms and their alternates in the *BE Ving* construction for three verbs: *sleep* (mostly expressing a ‘undirected activity’ (Croft 2012)), *drink* (mostly expressing an ‘activity’ or ‘accomplishment’) and *stand* (mostly expressing a ‘state’) (Vendler 1957). Data are collected from the EEBO, CLMET and BNC corpora, covering Early Modern, Late Modern and Present-day English. The data consist of random samples capped at 250 instances per verb per period.

As a first step, we establish how the interpretations of simple forms, which range from imperfective (including habitual and progressive) to perfective, arise pragmatically from other elements in the context, such as adverbial modifiers and relations established to other profiled events. For example, both the adverbial *now* in (1) and the framing subordinate clause in (2) encourage a progressive reading, while in (3) the sequence of events and the explicit endpoint to the *sleep* event jointly trigger a perfective interpretation.

- (1) Beyond the fear of ling’ring punishment, Aspasia now, within her lover’s arms, Securely **sleeps**[.] (CLMET, 1726-49)
- (2) Vnto this Pertynax therefore wente Letus, and Electus, with a few of theyr complices aboute mydnyghte, whyles all men **slepte**. (EEBO, 1556)
- (3) And so [I] wente to my bedde full ryght Where I **slepte** styll and merely tyll foure of the clocke after mydnyght Than vp I rose by the candell light (EEBO, 1506)

As a next step, we exploit the presence of these pragmatic interpretative cues as proxies to the likely intended aspectual reading of any given instance. This way, we can gain insight into any semantic change in the simple tense and its timing. If Bybee’s (1994) hypothesis holds, the simple tense is expected to shift away from contexts favouring or imposing a progressive interpretation as *BE Ving* gains in frequency.

At the time of writing, the analysis has been completed for *sleep*. Preliminary results show striking differences between present and past tense simple forms. In the present tense, the simple form over time increasingly refers to habitual situations, whereas the share of episodic progressive meaning declines. Although this is in line with expectations, we still find PDE examples that express ongoingness in the present-tense simple form, suggesting that contrary to received wisdom the simple tense does not obligatorily express non-progressive meaning even in PDE. The past tense casts even more doubt on the initial hypothesis, revealing an essentially stable coexistence of the simple form and *BE Ving* in the expression of progressive meaning. Results thus paint a far more complex picture than the literature suggests. If anything, obligatorification of the non-progressive interpretation in simple forms is a slow and contextually sensitive process.

References

- Bybee, Joan L. 1994. Asymmetries in tense and aspect systems. In William Pagliuca (ed.), *Current Issues in Linguistic Theory*, vol. 109, 235-254. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.109.02byb>.
- Croft, William. 2012. *Verbs: aspect and causal structure* (Oxford Linguistics). Oxford [England] ; New York: Oxford University Press.
- Kranich, Svenja. 2010. *The Progressive in Modern English: a Corpus-Based Study of Grammaticalization and Related Changes* (Language and Computers 72). Amsterdam New York, NY: Rodopi.
- Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66(2). 143-160. <https://doi.org/10.2307/2182371>.

On semantic change in grammaticalization: Why it is never metaphoric

We understand metaphor as a “conscious or voluntary shift in a word’s meaning” (cp. Matisoff 1991: 384).

Researchers like Heine et al. (1991) and Matisoff (1991) argued that metaphor is a fundamental mechanism of semantic change in grammaticalization. This idea has however been contested from the very start. For instance, Bybee et al. (1994: 24-25) argued that if metaphor would be a relevant mechanism under grammaticalization, we would expect to see semantic leaps, but these are rarely if ever found. Semantic change in grammaticalization tends to progress gradually involving small-scale reanalyses (in the sense of Hansen 2021).

The marginal position of metaphor in grammaticalization is in sharp contrast to the fact that metaphor is often considered to be one of the most important mechanisms in lexical change (e.g., Geeraerts 2015: 422; Blank 1997: 157). This remarkable difference between semantic change of lexical and of grammatical elements has to our knowledge received little attention, and it calls for an explanation.

Juge (2007) argued that metaphor plays a marginal role in grammaticalization because (as opposed to for instance pragmatic inferencing) it presupposes a high degree of speaker awareness of the metaphorically employed unit. Awareness is incompatible with the mechanisms of grammaticalization, since these are subconscious, he argues. We do not disagree, but Juge’s account begs the question why the mechanisms of grammaticalization are necessarily subconscious – or, at least, evade awareness – and why this is not the case with lexical semantic change. This question represents a fundamental challenge to linguistic change as both lexical semantic change and grammaticalization of lexical items have the same point of departure, namely lexical units.

In this paper, we offer an answer to the question. Our account takes its point of departure in a revised version of Boye & Harder’s (2012: 21) definition of grammaticalization. According to the revised definition, grammaticalization consists in the conventionalization of attentionally backgrounded status. This crucially entails that the input to grammaticalization is attentionally backgrounded, and we argue that this restriction on the input is what makes grammaticalization incompatible with metaphorically employed expressions. From a speaker as well as a hearer perspective, metaphor demands attention or awareness (cp. Juge 2007: 45): for a source concept to structure a target concept, speaker and hearer must pay attention to the internal structure of the concepts at play. Furthermore, novel metaphors will typically attract attention in that they are atypical ways of expressing oneself for which there arguably must be a reason based on the relevance maxim.

Our argument thus looks as follows: 1. A precondition for being grammaticalized is contextually being attentionally backgrounded. 2. Metaphor draws attention to the metaphorically used expressions. 3. Therefore, the metaphorical unit cannot lose the competition for attention and be conventionalized as having ancillary status.

We furthermore argue that based on our account, we can explain why also metonymy in a narrow sense (that is, as a conscious and voluntary process) seems to be marginal in grammaticalization. One of the consequences of our argument is thus that we should re-evaluate lumping pragmatic inferencing and metonymy narrowly defined together under the heading metonymy.

We argue that other approaches to grammaticalization are incapable to account for the same facts.

References

Blank A. (1997). *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: M. Niemeyer Verlag.

- Boye, K. & P. Harder (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, 88(1), 1-44.
- Bybee, J., R. Perkins & W. Pagliuca (1994). *The Evolution of Grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Geeraerts, D. (2015). How words and vocabularies change. In Taylor, J. (ed.) *The Oxford Handbook of the Word*. Oxford: Oxford University Press, 416-430.
- Hansen, M-B. M. (2021). In defense of a pragmatic view of reanalysis. *Journal of Historical Syntax*, 5, 1-34.
- Juge, M. 2007. Metaphor and teleology do not drive grammaticalization. In J. Salmons & Sh. Dubenion-Smith (eds.), *Historical Linguistics 2005: Selected papers from the 17th International Conference on Historical Linguistics, Madison, Wisconsin, 31 July - 5 August 2005*. Amsterdam: John Benjamins, 33-48.
- Matisoff, J. A. 1991. "Areal and Universal Dimensions of Grammatization in Lahu". In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization II*. Amsterdam: John Benjamins, 383-453.

Predicative possession in the languages of the Ethiopian area

Abstract

In Ethiopian languages, *have*-possession and existential constructions of possession are the predominant predicative possession types represented by the two dominant languages namely Oromo and Amharic, respectively. Previous research has shown that the *have*-possession is a Cushitic feature (cf. Thomason 1983:216) whereas the existential construction of possession has no clear historical records for its origin in Semitic (Rubin 2005:57). However, the Amharic-type of existential constructions of possession is common in the whole Ethiosemitic. The goal of this paper is two-fold: (1) showing further predominance of the existential constructions of possession over the *have*-possession, and (2) giving an overview of the notional characteristics of predicative possession in the languages of the Ethiopian area. It depends on the data from informants, online corpora, native-speaker intuition being bilingual in Oromo and Amharic, and published grammars. While the *have*-possession is employed by a few Cushitic and Omotic languages, the existential constructions of possession have been widespread in the area.

On the other hand, the Ethiopian languages show more commonality with the notions coded by the predicative possession constructions. Not only the possessive notions that include the prototypical ownership and the non-prototypical kinship, part-whole, physical, abstract, etc. relations, but also the non-possessive domains such as attribution, experience, and location are expressed by the predicative possession. The paper shows that the Ethiosemitic influence tends to be accounted for predominance of the existential constructions of possession based on the synchronic analysis because such construction is observable in all Ethiosemitic that might have spread to others through Amharic, the lingua franca of the country. The distribution of these predicative possession constructions implies that the predicate structures and the associated notional domains probably add important information to the Ethiopian Language Area.

Key words: *have-possession, Ethiopian Language Area, notions, predicative possession, existential constructions*

Reference

- Thomason, S. G. (1983). Genetic Relationship and the Case of Ma'a (Mbugu). *Studies in African Linguistics*, 14(2), 195–231.
- Rubin, A. D. (2005). *Studies in Semitic Grammatication*. Winona Lake, Indiana: Eisenbrauns.

The long and winding road of the Danish evidential *vel* - from epistemic modality via concessivity to evidentiality

The meaning of the Modern Danish modal particle *vel* (cognate *well*) has been analysed as expressing an appeal to the addressee to verify the truth of the proposition (cp. Davidsen-Nielsen 1996: 286). For its Modern Swedish cognate *väl*, the same meaning has recently been rendered in terms of *engagement* (Bergqvist 2020: 471).

However, these analyses only address the contextualized meaning. In utterances like (1) and (2), it would be absurd to hold that the speaker appeals to the addressee for verification:

- (1) Jeg fik ca. 30 sting og det gjorde **vel** ondt i en måneds tid.
'I got approx. 30 stitches and it hurt VEL for a month approximately'
- (2) Jeg har **vel** sikkert gjort forsøg på at falde ind i dialekten.
'I have VEL certainly tried to adapt to the dialect'

Vel does not necessarily seek confirmation but marks the proposition as put forth without having evidence for it, i.e., as a guess or conjecture. Thus, *vel* is so to speak grammaticalized 'gut feeling' (distinct from epistemic possibility since it readily combines with epistemic necessity markers, cp. (2)). Based on this meaning the contextual function to seek confirmation emerges: If the addressee can (dis-)confirm the proposition, a cooperative move would be to do so. In my paper, I will present an account of the development of this meaning.

Based on corpus studies, starting its development a millennium ago, the semantic path of evidential *vel* appears to be reconstructable as follows:

- (3) 'good, in satisfactory manner > 'easily' > epistemic necessity > concessive > evidential conjecture

With this semantic path, the development of *vel* can be accounted for based on small-scale hearer-driven reanalyses in terms of Hansen (2021).

The development of concessives out of epistemic modality markers is fairly uncontroversial (cp. e.g., Sweetser 1990: 70-72; Bybee et al. 1994: 226-227, Squartini 2012). The central part of the development is the transition from concessive to evidential meaning (cp. Thurgood 1986: 217-218). In line with Aikhenvald (2004: 276; 2011: 610) who argues that one source of evidential meanings are so-called evidential strategies whereby non-evidential markers are used evidentially in particular contexts, I argue that the conjecture meaning emerges based on contextual meanings in concessive sequences as rendered in Couper-Kuhlen & Thompson (2000: 382): The second, conceding, move in these sequences is a reiteration and an acknowledgment of the validity of others' statements. As pointed out by Squartini (2012: 2123), the conceded proposition is thus reported. Therefore, such sequences provide latent, contextually given evidential meanings where the addressee puts forth a proposition as true relying on others' assessment and hence, crucially for *vel*, without having supportive evidence for the assessment herself. In a pragmatically driven reanalysis (Hansen 2021), this second aspect ('lack of evidence') is then conventionalized as the coded meaning of *vel*.

The exact nature of the evidential meaning can only be understood in taking into account the paradigmatic oppositions of *vel* diachronically. Being grammaticalized, *vel* enters into opposition with the other evidential modal particles *nok* and *vist* which respectively express that the evidence is only subjectively or intersubjectively available (cp. Hansen & Heltoft 2011: 1058-60). Thus, only through paradigmaticization (Lehmann [1982]2015: 174; Nørgård-Sørensen et al. 2011; Diewald & Smirnova 2012), *vel* semantically finds its current position in the paradigm of evidential modal particles.

Reference

- Aikhenvald, A. Y. (2004) *Evidentiality*. Oxford: Oxford University Press, Incorporated.
- Aikhenvald, A. Y. (2011). The grammaticalization of evidentiality. In: Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 605–13. Oxford: Oxford University Press, 605-613.
- Bergqvist, H. (2020). Swedish modal particles as markers of engagement: Evidence from distribution and frequency. *Folia Linguistica*, 54(2), 469-496.
- Bybee, J. L., R. Perkins, W. Pagliuca (1994). *The Evolution of Grammar. Tense, Aspect and Modality in the Languages of the World*. Chicago/London: The University of Chicago Press.
- Couper-Kuhlen, E. & S. Thompson (2000) Concessive patterns in conversation. In: Couper-Kuhlen, E., Kortmann, B. (Eds.), *Cause, Condition, Concession, Contrast*. Berlin and New York: Mouton de Gruyter, 381-410.
- Davidson-Nielsen, N. (1996). Discourse particles in Danish. In: Engberg-Pedersen E., Fortescue M., Harder, P., Jakobsen, L.F., Heltoft L. (eds) 1996 *Content, Expression and Structure. Studies in Danish Functional Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 283-314.
- Gabriele D. & E. Smirnova. 2012. "Paradigmatic integration: the fourth stage in an expanded grammaticalization scenario." In Davidse, K., T. Breban, L. Brems and T. Mortelmans (eds.) [SLCS 130] *Grammaticalization and Language Change. New reflections*. Amsterdam: John Benjamins Publishing Company, 111-133.
- Hansen, E., & Heltoft, L. (2011). *Grammatik over det danske sprog*. Odense: Syddansk Universitetsforlag.
- Hansen, M-B. M. (2021). In defense of a pragmatic view of reanalysis. *Journal of Historical Syntax*, 5, 1-34.
- Lehmann, C. ([1982]2015). *Thoughts on grammaticalization. 3rd edition 2015* (Classics in Linguistics 1). Berlin: Language Science Press.
- Nørgård-Sørensen, J., L. Heltoft, & L. Schøsler (2011). *Connecting grammaticalisation*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Squartini, M. (2012). Evidentiality in interaction: The concessive use of the Italian Future between grammar and discourse. *Journal of Pragmatics*, 44(15), 2116-2128.
- Sweetser, E. (1990). *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Thurgood, G. (1986). The nature and origins of the Akha evidentials system. In W. Chafe and J. Nichols (eds.) *Evidentiality: the linguistic coding of epistemology*. Norwood, NJ: Ablex Publishing Corporation, 214-222.

Grammaticalization as conventionalization of discursively secondary status: Isolating what is unique to grammaticalization, and deconstructing the lexical-grammatical continuum

While there is wide agreement about approximately what grammaticalization is, a more profound understanding faces two major challenges. One is that the presupposed distinction between grammatical and lexical is itself hard to get a grip on (e.g. Boye & Harder 2012: 1-6). The other challenge is that even with a theoretically anchored and well-defined distinction between grammatical and lexical, it is not clear that grammaticalization is a distinct type of language change rather than an epiphenomenon (e.g. Campbell 2001: 151; see also Section 4).

Boye & Harder (2012) offered a solution to the first challenge, arguing for an understanding of grammatical elements as defined by conventionalized discursively secondary status (roughly, attentional background status). However, they circumvented the second challenge. Rather than attempting to define grammaticalization as a diachronic phenomenon, they defined it in terms of its result, namely as “the diachronic change that gives rise to linguistic expressions that are by convention ancillary and as such discursively secondary” (Boye & Harder 2012: 22). This definition is problematic as it includes all changes under grammaticalization as long as the output is a grammatical (i.e. secondary by convention) element.

This paper has two aims: Firstly, it proposes a definition of grammaticalization which is still based on the understanding of grammatical elements in Boye & Harder (2012), but which targets the nature of grammaticalization as a diachronic phenomenon: Grammaticalization is the conventionalization of discursively secondary status.

Secondly, it discusses important implications of the proposed definition: 1) Grammaticalization basically applies to meaning. What is conventionalized as discursively secondary is basically a meaning; a sign is only ‘discursively secondary by convention’ by virtue of its meaning. 2) Grammaticalization is a special case of a well-known type of change, viz. conventionalization. 3) Grammaticalization covers a restricted range of phenomena. For instance, changes like phonological reduction and semantic bleaching that are often associated with grammaticalization are external to grammaticalization. 4) Grammaticalization is not a type of overall development like, for instance *go* > *gonna*, but a small and well-defined part of such larger changes. 5) Grammaticalization is a gradual process to the extent that conventionalization is a gradual process. However, 6) the lexical-grammatical continuum must be deconstructed into other continua, including the conventionalization continuum, a splitting continuum (e.g. Heine & Reh 1984: 57), and a discourse prominence continuum. 7) Degrammaticalization (understood as a process in which grammatical elements give rise to lexical ones) is infrequent because it requires attention to discursively secondary (i.e. attentionally backgrounded) elements.

The paper is theoretical in intent, but invokes both standard examples of grammaticalization (e.g. the grammaticalization of future tense auxiliaries) and non-standard examples (e.g. what Jespersen 1922 called ‘secretion’) to illustrate its points.

References

- Boye, K. & P. Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88.1. 1–44.
- Campbell, L. 2001. What’s wrong with grammaticalization? *Language Sciences* 23.2–3. 113–161.
- Jespersen, O. 1922. *Language: Its nature, development and origin*. London: Allen & Unwin.

R Deletion in Brazilian Portuguese: Diachronic and Synchronic Evidence for Lexical Diffusion

This paper provides diachronic and synchronic evidence that R deletion in coda position in Brazilian Portuguese (BP) is being implemented through Lexical Diffusion (Wang; Chen, 1977; Oliveira, 1997). R deletion in BP can occur in both nouns and verbs, either in middle or final position: *amor* > *amô*-- 'love', *cantar* > *cantá*-- 'to sing', *perfume* > *pe—fume* 'perfume', *perseguir* > *pe—seguir* 'to chase'. Final R in verbs is almost categorically deleted (95,4%, according to Oliveira (1983)). The first document to provide evidence for this deletion process is the Appendix Probi, a document from the 3rd century that provides a list of words that people were pronouncing or writing incorrectly according to Classic Latin. Item number 149 from that list reads “*persica non pessica*” ‘peach’. There is then evidence that in the 3rd century there were already words being written or pronounced without the R. Also, this document shows there was social stigma against that. Gil Vicente, a Portuguese playwright from the 15th century, also provides examples of the stigma that R deletion would carry. In his plays, when lower class people are speaking, they do not pronounce final Rs in verbs. Leite de Vasconcelos (1970) and Chaves de Mello (1971) also show examples of that. Oliveira (1997, p. 33), using data from interviews with native speakers, analyzes this process through the lenses of Lexical Diffusion and proposes that “all phonological changes are guided through Lexical Diffusion.” In addition to that, he proposes that social class is a strong factor determining R deletion. Other factors such as stress, word class and informal speech also contribute to that. Our study of this process investigated final R deletion in nouns only. Our goal was to understand which social or internal factors were favoring the R deletion or retention. We also aimed at understanding if the change was proceeding through Lexical Diffusion. Interviews with 30 native speakers were conducted, and a total of 2,606 occurrences of final R in nouns was gathered. Overall, our study showed deletion rate of 12%, and factors such as stress, word class, age group, and social class were relevant to the process. The data also showed evidence for Lexical Diffusion, as the change is spreading through the lexicon according to token frequency, with frequent words changing first. This claim is supported by Bybee (1995). In sum, our paper makes a connection with Historical Linguistics as it analyzes R deletion diachronically, but it also presents results of a synchronic experiment.

Diachronic pathways of definite articles distribution

Definite articles are commonly suggested as an example of an unstable grammatical feature (Croft 1996, Wichmann & Holman 2009, Dediu and Levinson 2012, Greenhill et al. 2017). Frajzyngier (2008: 18) describes grammatical instability as a system “[i]n which given functional distinctions are neutralized in significantly extensive environments”, a situation which may cause articles to emerge and subsequently erode from a system. However, the properties promoting article emergence, loss, and persistence on a global scale remain unclear. Different linguistic features have been posited as the predictors of their distribution – gender marking and nominal classifiers, verb typology, topic and focus marking, verb-final word order, flexible word order, switch reference (Givón 1978, 1983), ergative alignment (Du Bois 1987, 2017), case marking (Comrie 1989), and perfective aspect (Leiss 2000). Empirical studies identified case, verb-final word order, flexible subject order, and ergative alignment as potentially robust predictors of the absence of definite articles (Evers 2020). These relationships have been explained in the light of 1) efficiency (case and flexible argument order can serve to mark definiteness and languages with these features might be less likely to gain another grammatical means with the similar function) and 2) the information status in languages with ergative alignment and verb-final languages (A arguments are prototypically definite, and clause-final arguments are prototypically indefinite). However, the relationships between the four predictors are complex: case marking is typically found in verb-final and flexible word order languages (e.g. Levshina 2021), and many ergative languages have case marking. An explicit causal approach is thus required to reveal the diachronic processes behind the cross-linguistic distribution of definite articles.

Here we disentangle the causal relationships underlying the synchronic distribution of definite articles on a global sample of 1232 languages matched for typological information from Grambank (The Grambank Consortium 2022) and the global EDGE tree (Bouckaert et al. 2022). We use logistic regression with phylogenetic correction within phylogenetic path analysis in *phylopath* package (van der Bijl 2018) to first establish the existing causal relationships between the predictors themselves and then to evaluate how and whether case, word order, and ergativity explain the distribution of definite articles.

We find that verb-final languages are more likely to maintain and gain case marking, and languages with case marking are likely to develop flexible ordering of core arguments. We use this causal model as the basis for several competing models predicting definite articles distribution. Out of the tested predictors of definite articles, only verb-final word order proves robust on a global sample: articles are more likely to be absent or disappear in verb-final word order languages. This suggests that word order dependent preferences for negotiating information status primarily shape the evolution of definite articles. The previously observed patterns of complementary distribution between 1) definite articles and 2) case, ergative alignment, and flexible word order are not supported by causal inference. Instead, these suggested relationships might be products of complex interactions between case, word order, and ergativity.

References

- Bijl, Wouter van der. 2018. *phylopath*: Easy phylogenetic path analysis in R. *PeerJ*. PeerJ Inc. 6. e4718.
- Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate E Jones & Quentin Atkinson. 2022. Global language diversification is linked to socio-ecology and threat status [Preprint]. <https://osf.io/preprints/socarxiv/f8tr6/>.
- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80(1). 1–28.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- Croft, William. 2002. *Typology and universals*. Cambridge University Press.
- Dediu, Dan & Stephen C Levinson. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLoS ONE* 7. e45198.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
- Du Bois, John W. 2017. Ergativity in discourse and grammar. In Jessica Coon, Diane Massam & Lisa D. Travis (eds.), *The Oxford Handbook of Ergativity*, 23–58. Oxford: Oxford University Press.
- Evers, Stephanie. 2020. *To 'the' or not to 'the': Cross-linguistic correlations between existing morphosyntax and the emergence of definite articles*. Buffalo, NY: The State University of New York at Buffalo PhD Thesis.
- Frajzyngier, Zygmunt. 2008. Grammaticalization, typology and semantics. In María José López-Couso & Elena Seoane (ed.), 61–102. Amsterdam: John Benjamins.
- Givón, Talmy. 1978. Definiteness and referentiality. *Universals of Human Language* 4. 291–330.
- Givón, Talmy. 1983. Topic continuity in discourse: the functional domain of switch reference. In John Haiman & Pamela Munro (eds.), *Switch Reference and Universal Grammar*, vol. 2, 51–82. Amsterdam: Benjamins.
- Greenhill, Simon J, Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C Levinson & Russell D Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*. National Acad Sciences 114(42). E8822–E8829.
- Leiss, Elisabeth. 2000. Artikel und Aspekt. Die grammatischen Muster von Definitheit. In *Studia Linguistica Germanica* 55. Berlin: Walter de Gruyter.
- Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 2791.
- The Grambank Consortium (eds.). 2022. Grambank 1.0. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://grambank.clld.org>.
- Wichmann, Søren & Eric W Holman. 2009. *Assessing temporal stability for linguistic typological features*. Munich: LINCOM Europa.

As Syntax Interfaces with Information Structure: Old Icelandic Non-Canonical Scrambled Orders

This talk reports some findings from my research on word order types, derived by Scrambling and found with double object constructions with one non-finite verb in the corpus of *Íslendinga Sögur* (1998). The account of Scrambling is launched in an integrated Minimalist syntactic framework (Eythórsson 1995; Haugan 2001; Thráinsson 2001; Epstein & Seely 2006; Wallenberg 2009; Cheng & Corver 2013; Engels & Vikner 2014) and provides convergent support to the claim that Scrambling is an optional displacement operation raising internal Arguments and Adjuncts out of their source positions into phrasally-adjoined targets in the left periphery of vP. Remarkably, while Conservation of C-Command (Wallenberg 2009: 132) bans movement of constituents across c-commanding functional heads, information-structural and semantic factors step in to instigate movement out of the VP domain. The given-before-new-information packaging strategy is the default option with canonical scrambled orders (Lambrecht 1994; Choi 1999; Hinterhölzl & Petrova 2009; Meurman-Solin, Lopez-Couso & Los 2012; Bech & Eide 2014; Biberauer & Walkden 2015), whereby the main findings can be boiled down to the following: pronouns scramble almost obligatorily; definite objects scramble quite freely; indefinite objects scramble only rarely; heavier objects usually remain in the unscrambled position.

Scrambled orders attested in non-canonical V_{fin} -IO(Dat)- $V_{non-fin}$ -DO(Acc), V_{fin} -DO(Acc)- $V_{non-fin}$ -IO(Dat), V_{fin} -IO(Dat)-DO(Acc)- $V_{non-fin}$ and V_{fin} -DO(Acc)-IO(Dat)- $V_{non-fin}$ constructions are in focus here, with “non-canonical” being reserved for scrambled orders that exhibit some syntactic structure deviations or information structure ambiguities, are of low frequency, and seem to be ‘non-optimal’ in terms of violating a syntactic constraint or overriding a principle of information structure, cf:

1) ... að hann mun þeim manni gefa dóttur sína, hvat manna sem hann er, ef að bana yrði orminum, (Ragn 819)

‘... that he will give his daughter in marriage to that man, whatever kind of man that be, if he would slay the dragon,’

2) ... ef þeir hefðu hlaupið frá mannum en þeir mættu grið gefa honum, (Heið 1387)

‘... if they had run away from the man so they might give him mercy,’

3) Nú skal veita svör þínu máli, að eg vil öllum yður grið gefa skipverjum. (Laxd 1564)

‘Now I shall give answers to your request that I will give mercy to all of you, shipmen.’

4) Vil eg það ráð þér gefa sem hverjum öðrum að hann leiti sér þess ráðuneytis ... (Fljót 723)

‘I’ll give that counsel to you as to anyone else that he should seek for himself this solution ...’

In 1) the in-situ DO *dóttur sína* escapes the definiteness effect while the scrambled IO *þeim manni* obeys definiteness but is arguably focussed. In 2) the ex-situ bare nominal DO *grið* is indefinite and the in-situ IO *honum* is an anaphoric pronoun that fails to obey the anti-focality effect. In 3) the non-constituent string *öllum yður* is scrambled along with the indefinite DO *grið*, as the nominal element of the IO *skipverjum* stays in-situ. Left-dislocation of the DO *það ráð* and the non-constituent pronominal element of the IO *þér* in 4) is sensitive to definiteness and anti-focality but the second element of the coordinate IO *sem hverjum öðrum* stays in situ.

I will argue that Scrambling in O(ld) Ice(landic) occurs on the Syntax-Information Structure Interface, and, by corollary that it can be thought of as a type of information packaging syntactic device. The studied interactions of word order and information structure can be explained as interface interactions that license scrambled orders on the basis of their syntactic, information-structural, and semantic properties. Base-generated word order and Scrambling of pronouns favour unmarked interpretation, Scrambling definite phrases is a less marked option than Scrambling indefinite phrases, Scrambling focussed phrases is more marked than Scrambling unfocussed phrases.

Among the conclusions are the following: Non-constituents may also be targeted by Scrambling in OIce; OIce Scrambling may evoke non-canonical information-structural effects: ex-situ XPs (and non-constituents) may be construed in terms of contrastive, non-presupposed, emphatic, focussed interpretations; A possible mismatch between the locus of default sentence focus and the position of the allegedly focalized expression can trigger Scrambling in OIce.

References:

- Bech, Kristin, and Kristine Gunn Eide, eds. 2014. *Information Structure and Syntactic Change in Germanic and Romance languages*. Amsterdam, Philadelphia: John Benjamins.
- Biberauer, Theresa, and George Walkden, eds. 2015. *Syntax over Time: Lexical, Morphological and Information-Structural Interactions*. Oxford: Oxford University Press.
- Cheng, Lisa Lai-Shen, and Norbert Corver, eds. 2013. *Diagnosing Syntax*. Oxford: Oxford University Press.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, California: CSLI Publications.
- Engels, Eva and Sten Vikner. 2014. *Scandinavian Object Shift and Optimality Theory*. Houndmills, Basingstoke, Hampshire and New York: Palgrave Macmillan Ltd.
- Epstein, Samuel D. and T. Daniel Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.
- Eythórsson, Thórhallur. *Verbal Syntax in the Early Germanic Languages*. PhD diss., Cornell University, 1995.
- Haugan, Jens. *Old Norse Word Order and Information Structure*. PhD diss., Norwegian University of Science and Technology, 2001.
- Hinterhölzl, Roland, and Svetlana Petrova, eds. 2009. *Information Structure and Language Change. New Approaches to Word Order Variation in Germanic*. Berlin: Mouton de Gruyter.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Meurman-Solin, Anneli, Maria Jose Lopez-Couso, and Bettelou Los, eds. 2012. *Information Structure and Syntactic Change in the History of English*. Oxford: Oxford University Press.
- Thráinsson, Höskuldur. 2001. Object Shift and Scrambling. In *The Handbook of Contemporary Syntactic Theory*, edited by Marc Baltin and Chris Collins, 148-212. Oxford: Blackwell Publishing.
- Wallenberg, Joel. *Antisymmetry and the Conservation of C-command: Scrambling and Phrase Structure in Synchronic and Diachronic Perspective*. PhD diss., University of Pennsylvania, 2009.

Obscenity as a Window into Slavic Linguistic History

Slavic obscenity has received some diachronic attention (e.g. Hamp 1968, Isačenko 1964, Uspenskij 1996), but its study has been marginalized in Slavic countries. Trubačev's (1964-73) translation and annotation of Vasmer (1950-58) omitted the core obscenities *ebat'* (older *eti*) 'fuck', *pizda* 'cunt' and *xuj* 'cock, prick', and *bljad'* 'whore'. Stavyc'ka (2008) provides some comparative Slavic material, but there has been no attempt to map out the patterns of retention and innovation in Slavic obscenity, and these display significant phenomena.

Slavic obscenity is remarkably conservative from an Indo-European standpoint, arguably more so than any other branch. The root (j)eb- is cognate with forms indicating copulation in Greek, Sanskrit, and Sogdian, but evidence from Luwian and Tokharian show that the original etymon began as a euphemism that was obscenified by contamination in dialectal Indo-European. While universally preserved in Slavic, the verbal root is in retreat as the *vox propria* for 'fuck' in most of West Slavic, especially Sorbian, Czech, and Kashubian, less so in Polish, and not in Slovak. Similarly, Common Slavic **peizdā* began as a euphemism on the Indo-European dialectal level, with cognates in Albanian and Nuristani (Hamp 1968, Mallory and Adams 1997). Bulgarian and Kashubian have specific developments, and Sorbian shows competition. An old isogloss separates South Slavic *kur* 'cock' from North Slavic *xuj* 'prick', with Bulgarian being transitional.

Since obscenities are subject to euphemization, which euphemisms become contaminated and displace earlier obscenities, leaving the older obscenities to either become obsolete or shift meaning, it would appear that in Slavic, obscenities underwent a process of decontamination and recontamination. Evidence for this is suggested in old South Slavic and East Slavic texts (Vinodolski Zakon, Old East Slavic Birchbark Letters). Alternatively, evidence from the Birchbark Letters could suggest that in the Middle Ages the situation, at least for *pizda* and *eb-* was like that in, e.g., modern Romani, where the single lexical item, *mindž*, can be translated 'vagina/vulva' or 'cunt' depending on the context, e.g. medical or invective.

By contrast it appears that, unlike terms for sexual intercourse and female genitalia, the male member was subject to the usual processes of obscenified euphemism replacement (with items that correspond lexically to English 'prick' and 'cock' in the North and South, respectively) at some time during the break-up of Common Slavic, i.e. the early Middle Ages. Based on the evidence of the Vinodolski Zakon and the Birchbark Letters, it can thus be argued that Medieval Slavic *eti* and *pizda* were obscene only contextually and did not become restricted to obscenity until the early modern period (evidence argues for the same treatment of Russian *bljad'*). Still, the Common Slavic for the male member may already have been restricted to obscenity, whence its replacement by the obscenification of new euphemisms, *xuj* and *kur*, after the break-up of Common Slavic.

The history of core obscenities in Slavic thus illustrates the importance of studying obscenities in general. In the case of Slavic: 1) The remarkable conservatism of two out of three Slavic core obscenities suggests either decontamination and re-obscenification or a period characteristic of languages for which context determines obscenity; 2) Those languages with the strongest Germanic contact are most likely to innovate obscenities, consistent with facts of German (Stavyc'ka 2008); 3) The male member was treated differently from both female genitalia and sexual intercourse and in this regard Bulgarian shows connections to East Slavic, pointing to Macedonian's closer connection to the rest of South Slavic vis-à-vis Bulgarian.

References

- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European Languages*. Chicago: University of Chicago Press.
- Hamp, Eric P. 1968. Albanian *pidh*: Slavic *peizd'ǎ*. *International Journal of Slavic Linguistics and Poetics* 11.25-26.
- Isačenko A. V. 1964. Un juron russe du XVIe siècle. *Lingua Viget: Commentationes slavicae in honorem V. Kiparsky*, ed. by I. Vahros & K. Kahla, 68-70. Helsinki: Suomalaisen Kirjallisuuden Kirjapaino.
- Mallory, James P. and Douglas Q. Adams, 1997. *Encyclopedia of Indo-European Culture*. London/Chicago: Fitzroy Dearborn.
- Stavyc'ka, Lesja. 2008. *Ukraïnska mova bez tabu*. Kyiv: Krytyka.
- Trubačev, O.N., ed. and transl. 1964-1973. Maks Fasmer, *Ėtimologičeskij slovar's russkogo jazyka*. Moscow: Progress.
- Uspenskij, Boris. 1996. Mifologičeskij aspekt russkoj èkspressivnoj frazeologii. *Anti-mir russkoj kul'tury*, ed. by N. Bogomolov, 9-107. Moscow: Ladomir.
- Vasmer, Max. 1950-1958. *Russisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Anchoring patterns in emerging complement clauses in Slavic

Complement structures are considered to consist of a complement-taking predicate (CTP) and an element indicating the subordinated/complement status of the clause it introduces (Schmidtke-Bode 2014, Kehayov/Boye 2016). However, in actual discourse we regularly come across cases that do not allow an easy categorization of these elements (and their relation). This holds in particular for diachronic data. The challenge arises from the fact that the categorization of complementizer and CTP in a potential complementation structure is subject to a certain circularity. In standard cases this does not present a problem because we have strong paradigmatic and preference-based evidence for the functional loads of the respective units. In less obvious cases, however, it becomes more difficult. This is particularly true for the Slavic languages, which possess very few clear-cut indicators of subordination (for preliminary findings cf. Sonnenhauser 2021 on Sln. *naj*, Wiemer 2021; forthcoming).

In our presentation we aim to identify the impact of various types of predicates on the inference of cross-clausal discourse relations and the eventual emergence of structures with clausal complements in Slavic. Special attention is given to predicates which usually do not join the list of ‘classic’ CTPs. We focus on the relation of clausal structures containing optative / directive expressions such as Po. *niech*, Ru. *pust’*, Sln. *naj* to clauses preceding them. The categorization of a clause-initial element as an untypical complementizer might be based on the presence of a typical CTP in the preceding clause, whereas the identification of an untypical CTP might be triggered by a subsequent typical complementizer. Cf. example (1) with a directive-optative marker (DIR) like Russ. *pust’* and the regular CTP *poželat’* ‘wish’ and (2) with the atypical CTP *nie dość* ‘not enough’ and a standard complementizer (COMP) like Pol. *że*:

- (1) Russian (RNC; T.N. Tkačenko, 1995)
Ja poželala [ej skorej opraviti’sja] i [pust’ segodnjašnee DTP budet ej v nazidanie] [...]
‘I wished.CTP [her a fast recovery] and [may.DIR today’s accident be a lesson for her]’
- (2) Polish (PNC)
Nie dość, [że jest to bardzo niewygodne dla samych obywateli], to jeszcze urąga samej powadze sądu okręgowego, [...]
‘Not only.QUASI-CTP [that.COMP this is very inconvenient for the citizens themselves], it also hurts the very authority of the district court’

To avoid the pitfalls of circular argumentation we refrain from assumptions about the categorial affiliation of the elements in question. We explore the relation between predicates that have the potential of serving as anchors for cross-clausal discourse continuation and clause-initial elements that introduce non-first clauses in a chain of clauses, regardless of their mainstream classification. This includes an account of combinations of COMP elements (Po. *że*, Ru. *čto*, Sln. *da*) and DIR elements (Po. *niech*, Ru. *pust’*, Sln. *naj*), adjacent as well as non-adjacent.

Using sets of random samples from two periods (17th-19th c. vs. contemporary stage; see References) we aim to identify possible correlations between the (type of) predicate of a preceding and a following clause containing a DIR-element, with respect to the following criteria concerning the anchoring element: (i) lemma, (ii) semantic class, (iii) PoS, (iv) inflectional form. On this basis, we explore whether the identified anchoring patterns bear on structural features within the subsequent clause, such as the (non)-initial appearance of the DIR-element and the distance between this element and the finite verb. Initial position of DIR-elements has been considered favorable for them to function as complementizers, non-initial as favorable for AUX status.

References

- Kehayov, Petar & Kasper Boye. 2016. Complementizer semantics – an introduction. In: Kasper Boye & Petar Kehayov (eds.), *Complementizer Semantics in European Languages*, 1–11. Berlin & Boston: De Gruyter Mouton.
- Schmidtke-Bode, Karsten. 2014. *Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization*. Jena (unpubl. PhD thesis).
- Sonnenhauser, Barbara. 2021. Slovene *naj*: an (emerging) clausal complementizer? In: Wiemer, Björn & Barbara Sonnenhauser (eds.), *Clausal complementation in South Slavic*. Berlin, Boston: De Gruyter Mouton, 443–475.
- Wiemer, Björn. 2021. A general template of clausal complementation and its application to South Slavic: theoretical premises, typological background, empirical issues. In: Wiemer, Björn & Barbara Sonnenhauser (eds.), *Clausal complementation in South Slavic*. Berlin, Boston: De Gruyter Mouton, 29–159.
- Wiemer, Björn (forthcoming). Between analytical mood and clause-initial particles – on the diagnostics of subordination for (emergent) complementizers. *Zeitschrift für Slawistik* 68–2 (2023).

Corpora

Polish

<http://nkjp.pl/>

https://korba.edu.pl/query_corpus/

Russian

<https://ruscorpora.ru/new/>

Slovene

<http://nl.ijs.si/imp/>

<http://www.gigafida.net/>

The anticausative alternation in Italian and Spanish: a historical corpus-based perspective

The causative alternation is the grammatical alternation whereby languages express externally caused (CAUSAL) vs. spontaneously occurring (NONCAUSAL) events. Typological studies have shown that the alternation is encoded via different patterns across languages, based on whether the causal event, the noncausal one, both, or neither are overtly marked (see Tubino-Blanco 2020). A cross-linguistically peculiarity of several languages of Europe is that, in this area, one finds an exceptional preference towards explicit marking of the noncausal event (Nichols et al. 2004), which goes under the name of anticausativization (Haspelmath 1987; Zúñiga & Kittilä 2019: 41–53). In many of these languages, as is common cross-linguistically (Bahrt 2021), anticausativization is formally expressed by the same marker of reflexivity. Examples from Italian and Spanish are given in (1) and (2). Notably, alongside the anticausative pattern, in these languages one also finds labile verb pairs, as in (1b-c).

Regarding the alternation between anticausative and labile patterns, it has been shown that individual verbs may trigger anticausativization more frequently than others, to the effect that verb meanings can be ranked based on their likelihood to receive anticausative marking (Haspelmath 1993, 2016). To explain these preferences, scholars have resorted to either verb semantics or frequency effects. Semantics-based accounts appeal to notions such as spontaneity (Haspelmath 1987) and claim that verbs lexicalizing events less likely to occur spontaneously more frequently trigger anticausative marking (Haspelmath 1993: 106). In frequency-based approaches, marking asymmetries mirror frequency asymmetries, based on the assumption that higher usage frequency items are more predictable and favor shorter coding (Haspelmath 2021). This means that verbs that more routinely occur in noncausal contexts are less likely to occur in the anticausative pattern (Haspelmath et al. 2014). These findings have also been confirmed by corpus data from French and Spanish (Heidinger 2019).

A yet unresolved issue with frequency-based approaches is that “we cannot measure the earlier frequencies” (Haspelmath 2016: 601). To overcome this issue, this work aims at testing whether claims advanced by the frequency-based approach are borne out by historical data and what the interplay is between semantic and frequency effects.

To do so, we undertake a contrastive exploration of the distribution of anticausative marking in historical corpora of Spanish and Italian. The choice of Italian and Spanish is also due to the fact that systematic diachronic studies on anticausatives in these languages are relatively few (e.g., Portilla 2007; Cennamo 2012, 2021). Data for Italian come from the MIDIA corpus whereas data for Spanish come from the CDH corpus, both including texts ranging from the 13th to the 20th century. In particular, based on the 20 verb meaning pairs list in Haspelmath et al. (2014), reported in (3), we have sampled (max) 500 tokens of each of the corresponding Italian and Spanish verbs. For each verb, we extract data regarding their token frequency in causal vs. noncausal contexts and on the encoding of the alternation (anticausative vs. labile). This will allow us to explore whether the observed frequency of the verbs under investigation in causal and noncausal contexts changes over time and what this reveals about asymmetries in the encoding of the anticausative alternation in Italian and Spanish. In addition, we also consider a number of additional factors that have been claimed, on synchronic grounds, to play a role in the choice of the anticausative vs. labile pattern, including semantic features of the subject (animacy, control) and aspectual properties of the verbs (Cennamo 2012, 2021; de Benito Moreno 2022: Chap. 4). By resorting to regression modelling techniques, we assess whether the role of these factors is stable across time or differences can be pinpointed at specific language stages and across languages.

The results of this work will contribute to showing how anticausativization comes about, and will offer the first in-depth empirical assessment of how anticausative markers spread through the verbal lexicon across time.

Examples

- (1) a. *Il ragazzo bruciò il cibo* CAUSAL
 the boy burn.PST.3SG the food
 ‘The boy burnt the food.’
- b. *Il cibo si bruciò* NONCAUSAL-ANTICAUSATIVE
 the food REFL burn.PST.3SG
- c. *Il cibo bruciò* NONCAUSAL-LABILE
 the food burn.PST.3SG
 ‘The food burnt.’
- (2) a. *Juan rompió la mesa*
 J. break.PST.3SG the table
 ‘Juan broke the table.’
- b. *La mesa se rompió*
 the table REFL break.PST.3SG
 ‘The table broke.’
- (3) *boil, freeze, dry, wake up, go out/put out (fire), sink, melt, stop, turn, burn, fill, rise/raise, improve, rock, connect, gather, open, break, close, split*

References

- Bahrt, Nicklas N. 2021. *Voice syncretism*. Berlin: Language Science Press.
- Benito Moreno, Carlota de. 2022. *The Middle Voice and Connected Constructions in Ibero-Romance*. Amsterdam & Philadelphia: John Benjamins.
- Cennamo, Michela. 2012. Aspectual Constraints on the (Anti)Causative Alternation in Old Italian. *Transactions of the Philological Society* 110(3). 394–421.
- Cennamo, Michela. 2021. Anticausatives and lability in Italian and French. In Silvia Luraghi & Elisa Roma (eds.), *Valency over time*, 265–304. Berlin & New York: de Gruyter.
- Haspelmath, Martin. 1987. *Transitivity alternations of the anticausative type*. Köln: Institut für Sprachwissenschaft.
- Haspelmath, Martin. 1993. More on the typology of the inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (eds.), *Causatives and Transitivity*, 87–120. Amsterdam & Philadelphia: John Benjamins.
- Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2). 33–63.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57(3).
- Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyacı. 2014. Coding causal–noncausal verb alternations. *Journal of Linguistics* 50(3). 587–625.
- Heidinger, Steffen. 2019. Reflexive and unmarked anticausatives in French and Spanish: Frequency of transitive use and undergoer overlap. *Langages* 216(4). 53–69.
- Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8(2). 149–211.
- Portilla, Mario Chaves. 2007. Diacronía de las construcciones con pronombres reflexivos en español. *Filología y Lingüística* 33(1). 131–149.
- Tubino-Blanco, Mercedes. 2020. Causative-Inchoative in Morphology. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Zúñiga, Fernando & Seppo Kittilä. 2019. *Grammatical Voice*. Cambridge: Cambridge University Press.

Corpora

- Morfologia dell’Italiano in DIACRONIA* (MIDIA): www.corpusmidia.unito.it [Accessed: 28/12/2022]
- Real Academia Española (2013): Corpus del Diccionario histórico de la lengua española (CDH): <https://apps.rae.es/CNDHE> [Accessed: 28/12/2022]

The Chronicle of Lingbe, an Extinct Bantu Language of East Congo

Lingbe, aka Ngbele, was a Bantu language spoken in the Northeast of the Democratic Republic of Congo, near the Bomokandi river. UNESCO (Moseley 2010) and the Summer Institute of Linguistics (see Eberhard et al. 2022) consider the language to be extinct. Indeed, a recent questionnaire on the language, recorded in 2020, resulted in a 100-word list of a Nilo-Saharan language, possibly Mangbetu, rather than the Bantu language we were actually looking for.

We are presenting the chronicle of a death foretold: the extinction of Lingbe was already announced in the 1930s. Moeller (1936) writes that the Mangbele were fishermen, one of the “Bakango” peoples living at the Uélé and Bomokandi rivers. At Rungu and Wamba the Mangbele had already exchanged their language for Mayogo, an Ubangi language. Other Mangbele had been conquered by the Mangbetu and incorporated in Mangbetu and “Madjaga” communities. As mentioned, Mangbetu is a Nilo-Saharan, more specifically Central-Sudanic, language. The present paper explores historical sources from the early 20th century that document the language and its speakers as well as the on-going language shift. This allows us to study the context in which the language shift and ultimate extinction of Lingbe occurred.

Still, the language lived at least until the 1990s when the second author recorded a wordlist in Lingbe. As short as it may be, merely 67 words, it has already allowed the authors and colleagues to identify Lingbe as a Boan Bantu language, more specifically affiliated to the western Boan subgroup which includes Leboale, Kango and Ngelima. Interestingly, this is in line with the hypothesis forwarded by Moeller (who cites De Calonne-Beaufaict 1921) that they are closely related to the Ababua. The present paper compares the surviving linguistic material to the surrounding Bantu, Central-Sudanic and Ubangi languages, as such reconstructing as much as is possible on Lingbe’s early history and studying the process of language shift.

The present paper assembles all available documentation on Lingbe, be it historical sources, linguistic or other, and retrieves the utmost from the linguistic material. The research results offer unique insights into the history of a lost Bantu language.

References:

- De Calonne Beaufaict, Adolphe. 1921. *Azande : introduction à une ethnographie générale des bassins de l’Ubangi-Uele et de l’Aruwimi*. Bruxelles: M. Lamertin.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2022. *Ethnologue: Languages of the World. Twenty-fifth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> .
- Moeller, A. 1936. *Les Grandes Lignes des migrations des Bantous de la Province-Orientale du Congo belge*. Brussels : IRCB, Section des sciences morales et politiques, Mémoires, T. VI.
- Moseley, Christopher (ed.). 2010. *Atlas of the World’s Languages in Danger, 3rd edn*. Paris, UNESCO Publishing.

Dramatic texts as a source of stigmatization from below

Linguistic stigmatization has mostly been described as a "from above" process. However, stigmatizing grammarians must also have their structures from somewhere: either from their own observation and evaluation of language, or from stereotypes and stigmas that are already circulating within the speech community. Moreover, stigmatizations from the top must also descend to the bottom, where they must be incorporated into a broader discourse of linguistic registers and evaluative techniques.

Because fictional, literary language is typically a standard (written) variety, it frequently exploits the contrast with orality. This frequently involves orality features that are considered characteristic of certain groups of speakers, as we see in classical Greek drama, which clearly influenced modern European literatures.

In German literature we already find passages with fictionalizations of oral varieties in Middle High German texts. As a literary strategy, however, the opposition between written and oral language(s) becomes popular only from the early modern period. Especially within passages of direct speech and dramatic texts, oral varieties are adapted. Such oral variety implementations convey linguistic (and non-linguistic) stereotypes as for example the excessive use of diminution (by means of *-le*) and the apocope in the text of a Swabian peddler in the play "Der Eheteufel auf Reisen" (1822) by the Viennese author Josef Alois Gleich:

Ach du lieber Herrgott, was ist das für ein Lebe-ø , ja, es wäre schon recht, ich wollte auch gern tanze-ø und gute Bißle esse, wenn wir nur auch das Geldle dazu hätte-ø. Aber mein Mann, das ist ein Lump ohnegleiche-ø – den ganzen Tag sitzt er in Branntweinhäusle, und ich kann nicht begreife-ø, wo er auf den Abend das Geldle hernimmt. [Emphasis added]

'Oh, dear God, what kind of life is that, yes, it would be all right, I would also like to dance and eat good bits, if only we had the money for it. But my husband is an incomparable rascal - all day long he sits in a liquor house, and I can't understand where he gets the money for the evening.'

The lecture examines the extent to which such structures of intended orality serve stereotypes (in the definition according to Labov 1972 and Silverstein 2003) and can be used as a source of stigmatization from below using 200 German-language plays from the 16th to the 19th centuries. Selected phenomena (diminution, pronominal adverbs, variation within the VP, and verbal periphrasis) will be investigated and compared to previous studies on stigmatization processes in the history of German (esp. Davies & Langer 2006).

Davies, Winifred V. / Langer, Nils (2006): *The Making of Bad Language*. Frankfurt am Main: Peter Lang.

Labov, William (1972): *Sociolinguistic Patterns*. Conduct and Communication. Philadelphia: University of Pennsylvania Press.

Silverstein, Michael (2003): „Indexical Order and the Dialectics of Sociolinguistic Life.“ *Language and Communication* 23 (3): 193–229.

Divergence and contact in Cappadocian concessive conditionals

In this talk we present preliminary results from an ongoing investigation into concessive conditionals (CCs) in Cappadocian, a near-extinct variety of Greek that was heavily influenced by Turkish when it was spoken in Asia Minor from Byzantine times until the Greek-Turkish population exchange in 1923–1924. We investigate whether Cappadocian CCs deviate from their Standard Modern Greek counterparts and, if so, whether the deviations can be attributed to language contact with Turkish.

CCs are a special type of conditionals which express not one antecedent p , but a set of antecedents that all lead to the same consequent q : ‘if $\{p_1, p_2, p_3, \dots\}$, then q ’ (König 1986; Leuschner 2006, 2020). Three quantificational strategies to evoke this set are distinguished (ibid.): scalar concessive conditionals (SCCs) mention an extreme value p_n and imply that q also holds for other, less extreme values (cf. English *even if p_n*); alternative concessive conditionals (ACCs) express a disjunction which exhausts the scale at hand (cf. English *whether p_1 or p_2*); universal concessive conditionals (UCCs) express free-choice quantification over instantiations of a variable, often realized as an interrogative-like pronoun (cf. English *WH-ever*).

Under Haspelmath & König’s (1998) typology of CCs in European languages, Turkish qualifies as uniformly-coding, i.e. as a language that encodes all CCs as conditionals, while Standard Modern Greek is differentially-coding, i.e. a language in which only SCCs have overt conditional coding, while ACCs and UCCs have primarily quantificational, e.g. interrogative-like, coding. This typological divide makes Cappadocian an interesting case study.

We investigate the coding strategies of Cappadocian CCs in a corpus of 58 folktales from 11 villages (ca. 50,000 words, the largest corpus of Cappadocian to date). While Cappadocian CCs are differentially coded like their Standard Modern Greek counterparts, the actual coding is distinct between both varieties. In part, these differences are due to Turkish influence, as Turkish loan words are found in ACCs, e.g. *jáxot ... jákot ...* ‘whether ... or ...’ < Tr. *yahut* ‘or, else’ and sporadically in UCCs, e.g. *-dak* in *ótia-dak* ‘whatever’ < Tr. *dek* ‘until, as far as’. Mostly, however, Cappadocian CCs differ from their Standard Modern Greek equivalents in ways that cannot be attributed to Turkish. In Cappadocian SCCs, the focus particle *ke* ‘even’ invariably follows the conditional conjunction *an* ‘if’, whereas *ke* precedes *an* in SCCs in Standard Modern Greek (where *an ke* is purely concessive). And whereas Haspelmath & König (1998) suggest that Standard Modern Greek UCCs usually contain focus particles like *-dipote* ‘ever’ or *ke* ‘even’ and/or conditional *an*, Cappadocian UCCs lack any overt coding other than the WH-word in 68% of all instances.

Future studies should investigate whether these differences are a consequence of changes in Modern Greek, with Cappadocian preserving coding strategies from earlier stages of Greek due to its relative isolation from mainstream Greek since the defeat of the Byzantine Empire by the Seljuk Empire at the Battle of Manzikert in 1071. Future research should also try to account for SCCs in the Flořta dialect, which are introduced by *an ki* like exceptive conditionals. To our knowledge, concessive and exceptive conditionals are not coded identically in any other varieties of Greek nor, indeed, any other languages. We suggest this overlap can be explained either in terms of accidental homonymy or as scale/polarity reversal.

References

- Haspelmath, Martin & Ekkehard König (1998): Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.): *Adverbial constructions in the languages of Europe*. Berlin: de Gruyter, 563–640. <https://doi.org/10.1515/9783110802610.563>
- König, Ekkehard (1986): Conditionals, concessive conditionals and concessives: Areas of contrast, overlap and neutralization. In Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson (eds.): *On conditionals*. Cambridge: Cambridge University Press, 229–246. <https://doi.org/10.1017/CBO9780511753466.013>
- Leuschner, Torsten (2006): *Hypotaxis as building-site: The emergence and grammaticalization of concessive conditionals in English, German and Dutch*. Munich: Lincom Europa.
- Leuschner, Torsten (2020): Concessive conditionals as a family of constructions. *Belgian Journal of Linguistics* 34, 234–246. <https://doi.org/10.1075/bjl.00049.leu>

Gascon u-perfects and the analogical foregrounding of inflectional class

This paper combines historical corpus data and comparative reconstruction to elucidate the origin and spread of the thematic vowel /u/ found today throughout preterite and imperfect subjunctive forms (i.e. reflexes of Latin *perfectum* forms; morphomic distribution ‘PYTA’, Maiden 2018) of third-conjugation verb lexemes in Gascon (Gallo-Romance) varieties.

Modern Gascon verb inflection (Grosclaude & Nariò 1998, Romieu & Bianchi 2005, Massourre 2012) is typified by a three-way conjugational class distinction in most tense-aspect-mood categories, expressed principally via theme vowels. *Perfectum* reflexes are characterised by thematic /ɛ/ in the first conjugation, e.g. *cantè* [kan'tɛ] ‘sing.PRET.3SG’, *cantèsse* [kan'tese] ‘sing.IPF.SBJV.3SG’, continuing Latin -DEDĪ perfects; thematic /i/ in the fourth conjugation, e.g. *bastí* [bas'ti] ‘build.PRET.3SG’, *bastísse* [bas'tise] ‘sing.IPF.SBJV.3SG’, continuing Latin -IUĪ perfects; and thematic /u/ in the third conjugation, e.g. *venó* [be'nu] ‘sell.PRET.3SG’, *venósse* [be'nuse] ‘sell.IPF.SBJV.3SG’. While the early historical development of the first- and fourth-conjugation forms shows continuity with other Occitan varieties (Ronjat 1937, Skårup 1997, Wheeler 2012), thematic /u/ is unique to Gascon within southern Gallo-Romance.

Traditional accounts (Zauner 1896:444-446, Allières 1988:177, Massourre 2012:214) ascribe Gascon u-perfects to the analogical influence of the single item ‘be’, and historical reconstruction supports this view. Sound change only yields theme vowel /u/ in *perfectum* reflexes of ‘be’ (e.g. FUIT > *fo* /fo/ > modern *hó* /hu/ ‘be.PRET.3SG’), and there is no potential source of /u/ elsewhere in the paradigm: in Gascon, the past participle in -/yt/ < -ŪTUM remains distinct from the *perfectum* reflexes, whereas in northern Gallo-Romance, reflexes of Latin past participles in -ŪTUM exert analogical influence on *perfectum* reflexes (Pope 1952:370, 381).

While the initial emergence of u-perfects occurs at the pre-literary period and is not directly visible in textual evidence (Bourciez 1927), the chronology and direction of their subsequent geographical and lexical spread can be traced in data from the Linguistic Corpus of Old Gascon (LCOG; Field 2012, 2013) and the *Atlas linguistique de la Gascogne* (ALG; Allières 1971). The progressive lexical extension of u-perfects is associated with the levelling of root and stress alternations in former ‘strong’ perfects (those with root-stressed forms in a subset of persons), e.g. analogical *metó* [me'tu] ‘put.PRET.3SG’ replaces etymological *mes* [mes] < MĪSIT, a process largely complete by the late thirteenth century; and later, with reduction in distinctive root allomorphy characterising morphomic distribution patterns. In some varieties, u-perfects spread beyond the third conjugation and into all fourth conjugation lexemes; this development is visible in northern dialects from the thirteenth century onwards, and is associated with the introduction of a thematic ‘augment’ (Maiden 2004, Esher 2016), e.g. analogical *bastissó* [basti'su] ‘build.PRET.3SG’ replaces etymological *bastí* [bas'ti].

The Gascon developments are significant for theories of inflectional analogy since they provide evidence for analogical remodelling of multiple items on the model of a single, idiosyncratic item of high token frequency, a phenomenon rarely reported in existing literature (see e.g. Cowgill 1959:11, Fulk 2018:189, 209, 307–308 for possible examples in Germanic). A further noteworthy aspect is the historical tendency of Gascon inflectional systems to retain and reinforce conjugational class distinctions and morphomic stem contrasts based on thematic vowels, contrary to the tendency observed across other Occitan varieties, in which contrasts based on distinctive root allomorphy take precedence (compare Esher 2021a,b,2022).

References

- Allières, Jacques. 1971. *Atlas linguistique de la Gascogne. Volume V: Le verbe*. Paris: CNRS.
- Allières, Jacques. 1988. Quelques énigmes de la conjugaison gasconne ou le verdict de l'aire. In Christian Abry (ed.), *Espaces romans*. Grenoble: Ellug. 175–192.
- Bourciez, Jean. 1927. *Recherches historiques et géographiques sur le parfait en Gascogne*. Bordeaux: Feret.
- Cowgill, Warren. 1959. The inflection of the Germanic *ō*-presents. *Language* 35(1). 1–15.
- Esher, Louise. 2016. Morphomic distribution of augments in varieties of Occitan. *Revue Romane* 51(2). 271–306.
- Esher, Louise. 2021a. Hypercorrection and velar-to-labial change in Occitan preterites, *Zeitschrift für romanische Philologie* 137(1). 61–98.
- Esher, Louise. 2021b. La diachronie du prétérit occitan à la lumière des sources textuelles. *Linguistica Pragensia* 31(2). 115–136.
- Esher, Louise. 2022. The historical development of imperfect indicative and conditional inflection in Pyrenean Romance. *Transactions of the Philological Society*. <https://doi.org/10.1111/1467-968X.12234>
- Field, Thomas T. 2012. Variation et diachronie: le témoignage du corpus électronique gascon. In Mario Barra-Jover, Guylaine Brun-Trigaud, Jean-Philippe Dalbera, Patrick Sauzet & Tobias Scheer (eds.), *Études de linguistique gallo-romane*. Paris: Presses universitaires de Vincennes. 21–32.
- Field, Thomas T. 2013. *The Linguistic Corpus of Old Gascon*. Research database. <<http://mllidev.umbc.edu/gascon/>>.
- Fulk, R. D. 2018. *A comparative grammar of the early Germanic languages*. Amsterdam: Benjamins.
- Grosclaude, Michel & Gilabèrt Nariò. 1998. *Répertoire des conjugaisons occitanes de Gascogne*. Orthez: Per Nòste.
- Maiden, Martin. 2004. Verb augments and meaninglessness in Romance morphology. *Studi di grammatica italiana* 22. 1–61.
- Maiden, Martin. 2018. *The Romance verb. Morphomic structure and diachrony*. Oxford: OUP.
- Massourre, Jean-Louis. 2012. *Le Gascon: Les mots et le système*. Paris: Champion.
- Pope, Mildred K. 1952. *From Latin to modern French*. Manchester: Manchester University Press.
- Romieu, Maurice & André Bianchi. 2005. *Gramàtica de l'occitan gascon contemporanèu*. Bordeaux: Presses Universitaires de Bordeaux.
- Ronjat, Jules. 1937. *Grammaire istorique des parlers provençaux modernes*, vol. III. Montpellier: Société des Langues Romanes.
- Skårup, Povl. 1997. *Morphologie élémentaire de l'ancien occitan*. Copenhagen: Museum Tusulanum Press.
- Wheeler, Max. 2012. Vies d'analogia i d'explicació en l'evolució del pretèrit feble de la conjugació-e romànica. *Estudis Romànics* 34. 7–36.
- Zauner, Adolf. 1896. Die Konjugation im Béarnischen. *Zeitschrift für Romanische Philologie* 20. 433–470.

The role of French in the Johnsons' correspondence

Among the studies flourished in the field of historical sociolinguistics, mainly in the strand devoted to English history, during its Medieval and early modern phases, multilingual texts have been analysed using theories and models coming from contact linguistics, thus applying synchronic models and approaches to the study of the past (Eckert 2012, Nurmi & Phata 2004, among others).

This contribution aims at investigating the role and the usage of the French language in the correspondence between Otwell Johnson and his brother, John Johnson, merchants of the Staple in Calais. Their letters, along with those of their other brother and their partners, were preserved only because of legal reasons (Oldroyd 1998), and the entire collection was subsequently transcribed for biographical studies by Winchester (1955).

The letters written by Otwell Johnson to his brother display interesting usages of the French language, making them multilingual; although the writer makes predominant use of English, we can find French elements in 17 letters (out of 161). These non-English items can be broadly grouped into three different contact phenomena: language choice, inter-sentential code switching, and intra-sentential code switching (defined as in Ciccolone & Dal Negro 2021).

Applying an atomistic approach (Hernández-Campoy 2003) to the study of the contact phenomena, I will investigate the above-mentioned documents, trying to explore the nature of the switching forms they contain, from an intra-writer variation perspective (Auer 2015, Nevalainen & Raumolin-Brunberg 2003). After analysing the letters and the type of multilingualism in them, I will also try to evaluate the functions given to the French language in this specific context.

References:

- Auer, A. (2015). Stylistic variation. In Auer, A., Schreier, D., & Watts, R.J. (eds.), *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 133-155.
- Ciccolone, S., & Dal Negro, S. (2021). *Comunità bilingui e lingue in contatto. Uno studio sul parlato bilingue in Alto Adige*. Caissa.
- Eckert, P. (2012). Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87-100.
- Hernández-Campoy, J.M. (2003). Exposure to contact and the geographical adoption of standard features: two complementary approaches. *Language in Society*, 32(2), 227-255.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical Sociolinguistics. Language Change in Tudor and Stuart England*. London: Longman Pearson Education.

Nurmi, A., & Phata, P. (2004). Social stratification and patterns of code-switching in Early English letters. *Multilingua*, 23(4), 417-456.

Oldroyd, D. (1998). John Johnson's letters: the accounting role of Tudor merchants' correspondence. *The Accounting Historians Journal*, 25(1), 57-72.

Winchester, B. (1955). *Tudor Family Portrait*. London: Jonathan Cape.

Tracing the origins of resumption in Swedish

Modern Swedish is a strict Verb-Second language. However, the language regularly displays Verb-Third strings when an adverbial particle is inserted between a fronted adverbial and the finite verb (*adverbial resumption*) (1).

- (1) *Ändå så finns det där hatet där ute*
 still SÅ exist that there hatred.DEF there out
 ‘Still, there is such a hatred out there.’ (Bloggmix 2005)

This paper will provide a detailed account of the evolution of adverbial resumption through the history of the Swedish language, focusing on the resumptive particles *så* ‘so’, and *då* (2). The main difference between the two is that *då* must follow an element that conveys time, while there are no restrictions on the use of *så*. In Meklenborg’s terms, *så* is a generalized resumptive, while *då* is specialized (Meklenborg, 2020). *Så* is by far the more common (see *inter alia* Elmquist (1945); Ekerot (1998); Nordström (2010)).

Using the Korp corpus (Borin et al., 2012), we find that in Early Old Swedish, resumption is very common. It becomes less frequent in the following centuries, before surging in the 18th century.

- (2) *hauir bondæn alt synir . tha takær han ey mer æn*
 have farmer.DEF already sons . tha take.PRS he NEG more than
en sun
 a son
 ‘If the farmer already has sons, then he will not take more than one son’

(SL, 1203–1212)

Whereas resumption with *så* is the most prominent structure in Modern Swedish, the structure was marginal in Old Swedish. The dominating resumptive element in the earliest period of the Swedish language was *þa* ‘then’ (= *då*). Out of 3976 cases where a fronted CP is followed by a resumptive in Early Old Swedish, the resumptive is *þa* in 3969 cases, while only 7 cases contain the resumptive *swa* (= *så*). Resumption with *swa* is restricted to contexts where the initial CP conveys a condition or a comparison. *Swa* is therefore a specialized resumptive during this period.

In the 16th century, the picture changes completely. Out of 190 cases of fronted CP + resumptive, the resumptive is *så* in 162 cases. In the next centuries, the ratio of resumption with *då* drops further. Looking at the semantics of the fronted adverbial constituent, we find that resumption with *då* gets more and more restricted, while these resumption with *så* becomes more widespread. We can, in other words trace the evolution into specialized and generalized resumptives.

References

- Lars Borin, Markus Forsberg, and Johan Roxendal. Korp – the corpus infrastructure of språkbanken. In *Proceedings of LREC 2012. Istanbul: ELRA*, pages 474–478, 2012. URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf.
- Lars-Johan Ekerot. *Så-konstruktionen i svenskan. Konstruktionstypen "Om värdet tillåter, så genomförs övningen i funktionellt grammatisk perspektiv"*. PhD thesis, University of Lund, Lund, 1998.
- Axel Louis Elmquist. The use of resumptive så after non-adverbial expressions in swedish. *Scandinavian Studies*, 18:284–288, 1945.
- Christine Meklenborg. Resumptive Particles and Verb Second. In Rebecca Woods and Sam Wolfe, editors, *Rethinking V2*, pages 90–125. Oxford University Press, Oxford, 2020.
- Jackie Nordström. The swedish så-construction, a new point of departure. *Working Papers in Scandinavian Syntax*, 85:37–63, 2010.

Old and new approaches to vernacular glosses in textual culture of medieval Poland: The case of embedded glosses

Most of Old Polish texts are in fact fifteenth century copies with accompanying inscribed or reinscribed glosses (interlinear and marginal) introduced into the text. The aim of the present article is to show how to recognize these types of glosses in the text using elements of semantics, syntax, rhetoric and source studies. The majority of examples come from the most extensive monument of Polish and Slavic medieval prose and apocryphal literature in the Polish language: *Rozmyślanie przemyskie* [The Przemysl Meditation]. About several hundred glosses introduced into the text of *Rozmyślanie* have been recognized. The glosses have had an enormous impact on the form of the preserved copy.

In the only preserved copy of this apocryphal text all glosses are incorporated into the main body of the text, often haphazardly and in inappropriate places. Some of them were turned into headings of chapters by the last copyist.

During our paper we will show how to work with a text with a multi-layered structure (a "text within a text within a text"), what tools are most suited for proper recognition of glosses and how to analyse the degree of their integration with the main body of the text.

The ordering of matrix clauses and subordinate causal clauses in the Old Bailey Corpus 1720–1913

The paper investigates changes in the order of finite adverbial causal clauses and matrix clauses in Late Modern English. Such clauses most frequently follow their main clause, as in (1), but they can also precede it, as in (2).

(1) *I did not much wonder at it, **because she had attempted to cut her Throat before** [...]* (OBC-1726)

(2) [...] ***because it was to be my place to take them**, I kept company with them all four for two hours* (OBC-1755)

Diessel (2005:465) argues that main clauses containing final adverbial clauses are cognitively privileged because they are easier to parse (Diessel 2005: 465–9); initial positioning is only possible for short subordinate clauses. Likewise, discourse pragmatic function plays a role, with initial position more likely if the reason presented is given rather than new. Therefore, *because* clauses, usually providing new causes, are less likely to be initial than clauses introduced by *as* and *since*, which tend to introduce known causes (Diessel 2005: 465–6; Chafe 1984: 442–4). Studies of adverbial clauses in earlier stages of English have broadly confirmed many of these results: Given-/newness plays a role for the positioning of causal clauses in Early Modern English (cf. Claridge & Walker 2001: 46–8; Pentrel 2017: 278 for temporal clauses in the 17th century), and Eitelmann (2016: 409), e.g., stresses the general validity of end-weight in Late Modern English.

The present paper analyses 1055 causal subordinate clauses from the 24-million word Old Bailey Corpus 2.0, which contains transcripts of trials from London's Central Criminal Court from 1720–1913 (Huber, Nissel and Puga 2016). Using multiple logistic regression, it explores the historical development of the influence of clause length and the given-/newness of the cause on the ordering of causal and matrix clause in a formative phase of English syntax, which saw the demise of *for* and the establishment of both *as* and *because* as causal conjunctions (Rissanen 1998). It is hypothesised that, under the uniformitarian hypothesis, these well-known factors had similar effects on clause position as in present-day English. The results feed into a diachronic constructional analysis which models the historical changes in the network of English causal clauses (Kanetani 2019).

References

Chafe, Wallace. 1984. How People Use Adverbial Clauses. In Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society, February 17-20, 1984, 437–49. Berkeley Linguistic Society.

Claridge, Claudia & Terry Walker. 2001. Causal clauses in written and speech-related genres in Early Modern English. ICAME Journal (25). 31–64.

Eitelmann, Matthias. 2016. Support for end-weight as a determinant of linguistic variation and change. English Language and Linguistics. Cambridge University Press 20(3). 395–420.

Diessel, Holger. 2005. Competing Motivations for the Ordering of Main and Adverbial Clauses. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 43(3). 449–470.

Huber, Magnus and Magnus Nissel and Karin Puga. 2016. *Old Bailey Corpus 2.0*. hdl:11858/00-246C-0000-0023-8CFB-2.
http://fedora.clarin-d.uni-saarland.de/oldbailey/downloads/OBC_2.0_Manual%202016-07-13.pdf. (7 January 2023).

Kanetani, Masaru. 2019. *Causation and Reasoning Constructions*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Rissanen, Matti. 1998. Towards an integrated view of the development of English: Notes on causal linking. In Jacek Fisiak & Marcin Krygier (eds.), *Advances in English historical linguistics*, 389–406. Berlin & Boston: Mouton De Gruyter.

The shape of grammaticalization: matching the bridging context scenario with patterns of frequency use

Heine's theory of contexts (2002) has been recognized as a satisfying account of a grammaticalization process, detailing the diachronic steps leading to and stemming from the drastic semantic shift or reanalysis that lies at the core of a grammaticalization process (Detges & Waltereit 2002). Other grammaticalization scholars have offered similar scenarios for the key steps of a grammaticalization process (Diewald 2002), introducing a critical context in which a new semantic inference can take place, while this inferential mechanism as a trigger for grammaticalization has been scrutinized on its own (Traugott 1988, Nicolle 1998, Nicolle 2011). From then, the form spreads over new contexts and becomes conventionalized through further constructional changes (Smirnova 2015).

On the other hand, it is known that language change is reflected, in terms of frequency of use, by a common template, that of the S-curve (Kroch 1989, Blythe & Croft 2012). Furthermore, this S-curve has been specifically associated with semantic expansion (Feltgen et al. 2017). Therefore, it is to be expected that the context-based scenario outlined by Heine and others in the one hand, and the S-curve pattern of frequency change on the other hand, should match on the level of their diachronic development; however, no robust parallel between the two accounts has been offered so far. Notably, frequency has been considered as a driving force for semantic change in grammaticalization processes (Bybee & Thompson 1997, Bybee 2006); nevertheless, whether the frequency needs to reach a 'critical threshold' for the shift to occur (or in which stage of the S-curve does the form start transitioning from the bridging context to the switch context) remains largely unclear. The goal of this contribution is precisely to remedy to this state of things.

To do so, I provide three empirical case studies of grammaticalizations in French, based on corpus data from the Frantext database (ATILF 1998-2023). These three case studies have been chosen to test my findings over three widely different grammaticalization scenarios. I thus explore a grammaticalization leading to the entrenchment of a schematic construction (Trousedale 2014), *une espèce de N* ('a kind of N'); a grammaticalization highlighted by a syntactic shift of the form (Fischer 2010), *mis à part*, which starts being fronted by the nineteenth century; and a grammaticalization driven by a calque over an already existing form (Cornillie 2019), *d'une façon* ADJ ('in a(n) ADJ way'), calqued from *d'une manière* ADJ. For each of these studies, I follow qualitatively the emergence of the grammatical meaning by manually parsing the occurrences, and I compare it with the S-curve of frequency rise extracted from the quantitative data.

In all of these three cases, the S-curve appears in the wake of the grammaticalization 'trigger', that is, as a consequence of the semantic shift. Several hints support this hypothesis; e.g., in the case of *une espèce de N*, the frequency over the types (i.e. how many fillers can enter the construction) also follows an S-curve, which would indicate a lexical diffusion (Ogura 2007); but then the token frequency of each individual fillers follows the same, identical S-curve, suggesting that they all 'register' the same semantic shift. The competition between *d'une façon* and *d'une manière* also follows two mirror S-curves, indicative of a transfer of a 'semantic load' between the two, similar to other competition processes (*en vs dans* in Fagard & Combettes 2013). The S-curve is thus mostly associated with stages 3 and 4 in Diewald's scenario, or with a diffusion over switching contexts in Heine's account. As Heine emphasizes, this also shows that the trigger of a grammaticalization does not guarantee its entrenchment in language use. A further actuation step is necessary. Incidentally, it also shows that frequency plays an ambivalent role in a grammaticalization scenario: it follows the semantic expansion and derives from it, seems necessary to entrench it, but does not seem to foster or facilitate it.

References

- ATILF. (1998-2023). *Base textuelle Frantext* (En ligne). ATILF-CNRS & Université de Lorraine. <https://www.frantext.fr/>
- Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88(2), 269-304.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733.
- Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. *Annual Meeting of the Berkeley Linguistics Society*, 23(1), 378-388.
- Cornillie, B. (2019). On the pace of syntactic elaboration from Latin calques: Evidence from Spanish, Dutch and English. *Belgian Journal of Linguistics*, 33(1), 82-107.
- Detges, U., & Waltereit, R. (2002). Grammaticalization vs. reanalysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft*, 21(2), 151-195.
- Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. In Wischer, I. & Diewald, G. (ed.), *New reflections on grammaticalization*, John Benjamins Publishing, 103-120.
- Fagard, B., & Combettes, B. (2013). De en à dans, un simple remplacement?. *Langue française*, 178, 93-115.
- Fischer, S. (2010). *Word-order Change as a Source of Grammaticalisation*. John Benjamins Publishing.
- Heine, B. (2002). On the role of context in grammaticalization. In Wischer, I. & Diewald, G. (ed.), *New reflections on grammaticalization*, John Benjamins Publishing, 83-102.
- Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3), 199-244.
- Nicolle, S. (1998). A relevance theory perspective on grammaticalization, *Cognitive Linguistics*, 9(1), 1-35.
- Nicolle, S. (2011). Pragmatic aspects of grammaticalization. In Heine, B. & Narrog, H. (ed.), *The Oxford handbook of grammaticalization*, Oxford University Press, 401-412.
- Ogura, M. (2007). Lexical Diffusion and Complex Adaptive System. *Studies in Modern English*, 2007(23), 23-51.
- Smirnova, E. (2015). When secondary grammaticalization starts: A look from the constructional perspective. *Language Sciences*, 47, 215-228.
- Traugott, E. C. (1988). Pragmatic strengthening and grammaticalization. *Annual Meeting of the Berkeley Linguistics Society*, 14, 406-416.
- Trousdale, G. (2014). On the relationship between grammaticalization and constructionalization. *Folia Linguistica*, 48(2), 557-578.

WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose.

In Latin and ancient Greek linguistics, onomasiological studies on the lexicalizations of concepts have led to a productive line of research based on in-depth philological analysis of extensive textual evidence (e.g. Moussy 1999; Kaster 2001; Craig 2005; Moussy 2005; 2007; Zamboni 2008; Lecaudé 2010; Kutscher and Werning 2014). The main obstacle to large-scale onomasiological analyses is that they require specialized language resources organized by concepts (Black 2001) rather than by lemmas. The Historical Thesaurus of English (Kay 1984) has enabled studies on diachronic patterns in English lexicalization of various concepts (Wild 2010; Alexander 2018; Allan 2020). For ancient Greek and Latin, we can now use WordNet (Biagetti et al. 2021).

Building on the work done within the MultiWordNet project to automatically generate Latin synsets from dictionaries (Minozzi 2017), the Latin (<https://latinwordnet.exeter.ac.uk>) and Ancient Greek WNs (<https://greekwordnet.chs.harvard.edu>), directed by Dr William Short (University of Exeter), aim to create accurate machine-readable and machine-actionable representations of the semantic structures of Ancient Greek and Latin using the WN framework. Thanks to the same data structures and semantic descriptors (*synsets*, *semfields*) provided by the Princeton WN, they intend to model the idiosyncratic and unique features of the semantics of ancient languages, in terms that are compatible and comparable with models constructed for other languages, as well as integratable into larger computational systems. WNs for the ancient languages try to capture diachronic and generic variations in word meaning, and figurative structures that impact semantic development. Based on conceptual metaphor theory in cognitive linguistics (Lakoff and Johnson 1980), they capture both the metaphorical or metonymical relationships between word senses at the level of word meaning and large-scale figurative relations that may operate supra-lexically. In addition, they contain etymological information.

Linking onomasiological resources such as WNs to corpus data can help us answer quantitative questions on the lexicalization of concepts. In our study, we combine WNs for ancient languages and corpus data, with specific attention to Ancient Greek and Latin. For our analysis, we choose the lexical field of SEA, given its importance in the two cultures. For the Greeks and the Romans, the sea was a source of wealth and trade (e.g. Reed 2003; Wilkinson 2020; Boardman et al. 2021), and a key part of their military strategy (e.g. Harris 2017; Nash 2018), but also a place of mystery and danger (e.g. Berens 1979; Lindenlauf 2004; Nikoloska 2012; Beaulieu 2016).

We focus on Ancient Greek and Latin pairs of nouns (e.g. AGr. *thálassa*, *póntos*, *pélagos*, *háls* : Lat. *mare*, *pontus*, *pelagus*, *aequor* ‘sea’) and verbs (e.g. AGr. *pléō* : Lat. *navigo* ‘sail’) related to SEA. We enrich the Ancient Greek and Latin Dependency Treebanks (Celano 2019; v2.1 http://perseusdl.github.io/treebank_data/) with semantic information from the Ancient Greek and Latin WNs. Our corpus includes the treebanked passages from Herodotus’s *Histories* and Caesar’s *De bello Gallico*. We then compare the quantitative data from the two languages to understand in what way Ancient Greek and Latin lexicalize concepts related to SEA, investigating the best way to add semantic information to the Ancient Greek and Latin Dependency Treebanks. This is a challenging task as WNs often provide a large amount of potentially relevant semantic information, and, even in a specific textual context, words can possess multiple senses, or their sense(s) can be ambiguous. Our study will cast new light on the use of WNs for crosslinguistic comparison for ancient Indo-European languages. It will also represent an attempt to digitally link semantic and syntactic information for classical languages, integrating WNs and treebanks (for modern languages see e.g. Kingsbury et al. 2002; Hajnicz 2014).

REFERENCES

- Alexander, M. 2018. Lexicalization Pressure. In *International Conference of English Historical Linguistics XX*, August 2018, Edinburgh, plenary paper.
- Allan, K. 2022. Degrees of lexicalization in the history of English: antonymy and asymmetry. In *International Conference on Historical Linguistics*, Oxford, 1-5 August 2022. Book of Abstracts.
- Biagetti, E., Zanchi, C. & Short, W. M. 2021. Toward the creation of WordNets for ancient Indo-European languages. In *Proceedings of the 11th Global Wordnet Conference*, University of South Africa (UNISA). Global Wordnet Association, 258–266.
- Beaulieu, M-C. 2016. *The Sea in the Greek Imagination*. Philadelphia: University of Pennsylvania Press.
- Berens, E. M. 1979. *The Myths and Legends of Ancient Greece and Rome*. Boston: Longwood Press.
- Blank, A. 2001. Words and concepts in time: Towards diachronic cognitive onomasiology. In *Words in Time: Diachronic Semantics from Different Points of View*, 37–65.
- Boardman, J., G. R. Tsetschladze, A. Avram, and J. Hargrave. 2021. *The Greeks and Romans in the Black Sea and the Importance of the Pontic Region for the Graeco-Roman World (7th Century BC-5th Century AD): 20 Years on (1997-2017)*. *Proceedings of the Sixth International Congress on Black Sea Antiquities (Constanta - 18-22 September 2017)*. Oxford: Archaeopress.
- Celano G. G. 2019. The Dependency Treebank for Ancient Greek and Latin. In *Digital Classical Philology*, 279–298.
- Craig, W. 2015. The language of gender: lexical semantics and the Latin vocabulary of unmanly men. In M. Masterson, N. S. Rabinowitz, and J. Robson, (eds.), *Sex in antiquity: exploring gender and sexuality in the ancient world*, 461-481. London: Routledge.
- Hajnicz, E. 2014. Lexico-Semantic Annotation of Składnica Treebank by Means of PLWN Lexical Units. In *Proceedings of the Seventh Global Wordnet Conference*, 23–31.
- Harris, W. V. 2017. Rome at Sea: The Beginnings of Roman Naval Power. In *Greece and Rome* 64 (1), 14–26.
- Kaster, R. A. 2001. The Dynamics of ‘Fastidium’ and the Ideology of Disgust. In *Transactions of the American Philological Association*, 131, 143–89.
- Kay, C. 1984. The Historical Thesaurus of English. In R. R. K. Hartmann (ed.), *Lexeter '83 Proceedings*, 87-90. Tübingen: Niemeyer.
- Kingsbury, P., M. Palmer, and M. Marcus. 2002. Adding Semantic Annotation to the Penn Treebank. In *HLT '02: Proceedings of the second international conference on Human Language Technology Research*, 262–256.
- Kutscher, S. and D. A. Werning. 2014. *On ancient grammars of space: linguistic research on the expression of spatial relations and motion in ancient languages*. Berlin/Boston (Mass.): De Gruyter.
- Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. London: The University of Chicago Press.
- Lecaudé, P. 2010. *La notion de puissance: les équivalents latins du grec δύναμις*. Thèse dactylographiée soutenue le 11 décembre 2010 à l'Université de Paris-Sorbonne.
- Lindenlauf, A. 2004. The Sea as a Place of No Return in Ancient Greece. In *World Archaeology* 35 (3), 416–433.

- Minozzi, Stefano. 2017. Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval. In P. Mastrandrea (ed.), *Strumenti digitali e collaborativi per le scienze dell'antichità*. Venezia: Edizioni Ca' Foscari.
- Moussy, C. 1999. Les vocables latins servant à désigner le sens et la signification. In M. Baratin and C. Moussy (eds.), *Conceptions latines du sens et de la signification*, 13- 27. Paris: PUPS.
- Moussy, C. 2005. Champ lexical et polysémie: l'expression du pardon en latin. In *Journal of Latin Linguistics*, 9(1), 359–370.
- Moussy, C. 2007. Ambiguus, ambiguitas, anceps, utroqueuersus dans le vocabulaire de l'ambigüité. In C. Moussy and A. Orlandini (eds.), *L'ambigüité en Grèce et à Rome. Approche linguistique*, 57–64. Paris: PUPS.
- Nash, J. M. 2018. *Rulers of the Sea – Maritime Strategy and Sea Power in Ancient Greece 550-321 BC*. The Australian National University: PhD Thesis.
- Nikoloska, A. 2012. The Sea Voyage of Magna Mater to Rome. In *Histria Antiqua* 21, 365–371.
- Reed, C. M. 2003. *Maritime Traders in the Ancient Greek World*. Cambridge/New York: Cambridge University Press.
- Wild, K. 2010. Angelets, trudgeons and bratlings: the lexicalization of childhood in the Historical Thesaurus of the Oxford English Dictionary. In M. Adams (ed.), *Cunning passages, contrived corridors: Unexpected essays in the history of lexicography*, 289–308. Monza: Polimetrica.
- Wilkinson, Troy. 2020. *The Indian Ocean Trade and the Roman State*. University of Wales Trinity Saint David: MA Thesis.
- Zamboni, A. 2008. Sul lessico latino, tardo e neolatino dell'alienazione mentale. In R. Wright (ed.), *Latin vulgaire–latin tardif VIII: actes du VIIIe Colloque international sur le latin vulgaire et tardif, Oxford, 6-9 septembre 2006*. Hildesheim/New York: Olms-Weidmann.

Where did *wer* go?

Searching for s-curves in lexical change from Old English to Middle English

To refer to a male adult, speakers of Present Day English have several lexical items to choose from (e.g., *man*, *guy*, *dude*, *fella*, *bloke*, *gentleman*, and *geezer*). However, variation within this semantic field is not new. According to *The Thesaurus of Old English*, there were at least 25 lexical items which denoted ‘male adult’ in Old English (e.g., *ceorl*, *guma*, *man*, *wer*) which could occur in referentially comparable contexts, as in (1). To examine the evolution of this onomasiological set from Old English and Middle English, the present study uses variationist quantitative methods, addressing two research questions. First, what was the distribution of third-person male adult nouns referents in Old English and Middle English? In other words, which variants were most frequent, in which contexts did they occur, and how did their ranking change over time? Second, is there any evidence to suggest that these variants were conditioned, constrained, or influenced by any attested intra- or extralinguistic factors of variation?

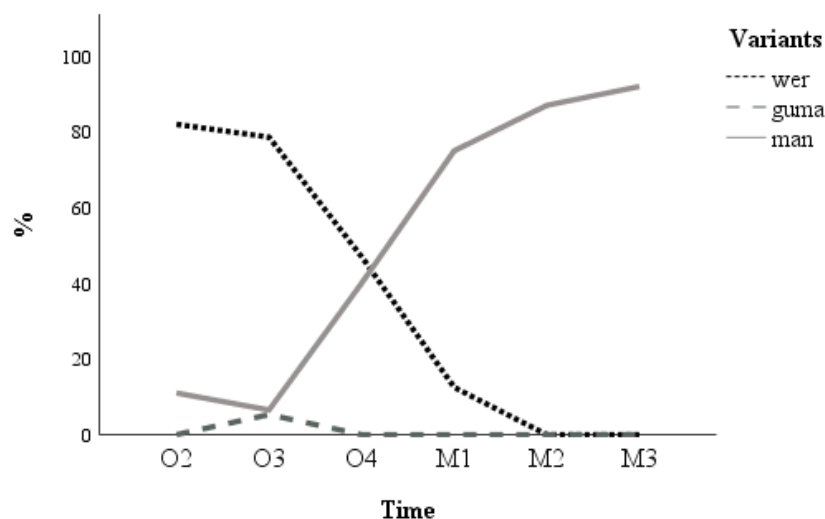
The *Helsinki Corpus of English Texts* (Rissanen et al., 1991) was used as the principal source of linguistic data. A list of third-person male adult noun referents was compiled using previous literature (e.g., Kleparski, 2005), dictionaries (e.g., Bosworth & Toller, 1882), and thesauruses (*The Historical Thesaurus of English*, *A Thesaurus of Old English*). Search queries were run to find instances of these variants in the corpus, which were subsequently downloaded and manually inspected for the removal of any functionally non-equivalent tokens (e.g., vocatives of address). Socio-historical context, as well as comparisons with Latin-based originals for translated texts, when available, were used to ensure only tokens with overt male referents were included in the envelope of variation. Each token was coded for both intra- (e.g., alliteration) and extra-linguistic factors (e.g., text type, provenance, time, text ID).

Results from the analysis demonstrate a significant shift in frequency from the favored variant *wer* in Old English to *man* in Middle English, a type of lexical replacement which coincides with collocational and frequency changes. As *wer* decreases, *man* takes on the former function of *wer*, with the diachronic shift in frequency following a prototypical s-curve distribution (Figure 1). Multivariate analyses using Rbrul (Johnson, 2009) found text type and text provenance to significantly constrain lexical choices, with certain variants (e.g., *rinc*, *scealc*, *knizt*) occurring more frequently in verse texts than prose texts on account of alliterative and metrical requirements.

Findings from the study are interpreted in the context of ongoing scholarship on lexical variation and change. First, the shift from *wer* to *man* illustrates a clear example of lexical replacement. As *wer* is replaced, it temporarily retreats to use as part of the semantic field ‘married man’ which later too is usurped by a competing lexical item *husband*. Second, although linguistic change does not have to follow an s-curve pattern, the shift from *wer* to *man* follows an s-curve trajectory, which is often used as a diagnostic for lexical replacement. Third, while work using apparent time data (Chambers, 1995) or short periods of time (Grieve et al., 2017) point to the applicability of s-shaped trajectories for lexical change, the present analysis of the semantic field of third-person male adult noun referents over approximately six hundred years adds a diachronic dimension to this discussion. Finally, in line with synchronic work on lexical variation, the present study shows how intra- and extra-linguistic factors harmoniously affect lexical choices and subsequently change diachronically.

- (1) (a) Ond on ðone ylcan dæg Crist gereorde fif ðusenda **wera** of fif hlafum ond of twam
 and on that same day Christ fed five thousand men of five bread and of two
 fixum, eac wifum ond cildum þara wæs ungerim
 fish also women and children which was uncountable
 ‘And on that same day, Christ fed five thousand men, with five loaves of bread and two
 fish. In addition, he also fed women and children, of which there were many’
 [Old English Martyrology, 950-1050]
- (b) on þære fyrde wæron þe ferdon fram Egipte, sixhund þusend **manna** butan wifum
 in the army were which traveled from Egypt six-hundred thousand men except women
 7 cildum
 and children
 ‘In that army, there were 600,000 men who travelled from Egypt, and that number
 does not include women and children’ [Ælfric’s Letter to Sigeward, 1050-1150]
- (c) ðonne onwæcneð eft wineleas **guma**
 then awakens again friendless man
 ‘Then the man without any friends woke up’ [The Wanderer, 950-1050]

Figure 1. Frequency of *wer*, *guma*, and *man* from Old English to Middle English



Selected References

- Bosworth, Joseph & Toller, Northcote. (1882). *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*. Oxford: Clarendon Press.
- Chambers, Jack. (1995). The Canada-US border as a vanishing isogloss: the evidence of chesterfield. *Journal of English Linguistics*, 23(1-2), 155-166.
- Grieve, Jack., Nini, Andrea & Guo, Diansheng. (2017). Analyzing lexical emergence in Modern American English online. *English Language & Linguistics* 21(1), 99–127.
- Kleparski, Grzegorz. (2005). Towards the semantics of Middle English synonyms of MAN. *Studia Anglica Resoviensia* 3, 88-95.
- Rissanen, Matti, Kytö, Merja., Kahlas-Tarkka, Leena, Kilpiö, Matti., Nevanlinna, Saara, Taavitsainen, Irma, Nevalainen, Terttu & Raumolin-Brunberg, Helena. (1991). *Helsinki Corpus of English Texts: Diachronic and Dialectal* (Helsinki).

The PRESENTATIVE > DEMONSTRATIVE Grammaticalization Pathway in Arabic

In their recent survey of lexical sources for the grammaticalization of demonstrative forms, Heine et al. (2020) identify the development of imperative verbs of perception to demonstratives as a cross-linguistically attested pathway. As a component of that discussion, they refer to the proposed origin of the French demonstrative *ce* in an earlier presentative construction *ecce ille* ‘Behold that!’ (cf. van Gelderen 2011), noting that, “while not a verb, *ecce* has a meaning similar to that of the imperative form of a perception verb” (Heine et al. 2020: 420) and proceeding with their analysis on that basis. In this presentation, I move to expand on this latter observation by adducing additional corroborating data from varieties of Arabic, and to discern in finer detail the relation between PRESENTATIVE > DEMONSTRATIVE and PERCEPTION VERB > DEMONSTRATIVE grammaticalization pathways. In so doing, I identify conceptual characteristics of the proposed diachronic sources that are especially consonant with established precepts of synchronic demonstrative function, which speak to the broader viability of a PRESENTATIVE > DEMONSTRATIVE grammaticalization pathway beyond the specific contexts of the languages here considered.

The material for this discussion comes primarily from the review and synthesis of three Arabic demonstrative series previously proposed in to originate in presentative predications, but not till now considered under a unified, concentrated lens. These include the Classical Arabic *hāḍā* proximals (Magidow 2013), the Egyptian Arabic *dawwa* proximals/*dukha* distals (Leddy-Cecere 2021), and Libyan Arabic *āhwa* proximals (Pereira 2008). Extending beyond earlier accounts, I will show that each of the above – while distinct from the others in time, place and etymology – may be plausibly analyzed as originating in a topic-dislocated presentative structure of the type ‘Behold it, X!’/‘X, behold it!’, through processes of grammaticalization (desemanticization, extension and decategorialization – cf. Heine 2007) and rebracketing. Classical Arabic *hāḍā* and Egyptian *dawwa/dukha* represent “reinforcing” developments that incorporate and modify an inherited demonstrative element in the grammaticalizing source construction, while Libyan *āhwa* does not.

Building on the observation preliminarily voiced by Heine et al., I assert that the linkage between PRESENTATIVE > DEMONSTRATIVE developments like these and those deriving demonstratives from imperative verbs of perception consists explicitly not in the *semantic* dimension of their source meanings, but rather in the shared *pragmatic* status of those sources as directive speech acts that implore an addressee to attend to a given demonstratum. This directive nature thus contributes to the deictic component of the demonstrative function – the identification/demonstration to the addressee of an index – while the inherited pronominals involved in all examples considered here contribute toward the demonstrative function’s classificatory component – assisting the addressee’s ultimate attachment of that index to a referent (cf. Nunberg 1993).

As a direction for future research, I will briefly consider the implications of these findings for a novel interpretation of two further Arabic demonstrative series – the Classical Arabic *ḍālika* and North Fertile Crescent Arabic *hāk* distals – as grammaticalizing from dative presentative/offertive constructions (‘Here you go, X!’). While more data remains to be gathered (particularly in the Classical Arabic case, for which a rich textual record survives), the entailment of a directive to attend in a directive to take/receive supports the viability of such a pathway in light of the PRESENTATIVE > DEMONSTRATIVE cases already considered, and further work in this area may serve to illuminate aspects of the interflow between person- and distance-oriented distinctions in demonstrative development.

References

- van Gelderen, E. (2011). “The grammaticalization of agreement.” In H. Narrog & B. Heine (eds.), *The Oxford Handbook of Grammaticalization*, 491-501. Oxford: Oxford University Press.
- Heine, B. (2007). “Grammaticalization.” In B. Joseph & R. Janda (eds.), *The Handbook of Historical Linguistics*, 575-601. Oxford: John Wiley & Sons, Ltd.
- Heine, B., T. Kuteva, H. Long, H. Narrog & F. Wu. (2020). “Where do demonstratives come from?”. *STUF* 73(3), 403-434.
- Julia, M. (2020). “About Latin *Ecce* ‘Behold! Lo! See! There!’ and Some Ancient Languages Presentatives. *International Journal of Language and Linguistics* 8(1), 17-23.
- Leddy-Cecere, T. (2021). “Interrogating the Egypto-Sudanic Arabic connection.” *Languages* 6, 123.
- Magidow, A. (2013). *Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity*. Austin: The University of Texas at Austin doctoral dissertation.
- Nunberg, G. (1993). “Indexicality and Deixis.” *Linguistics and Philosophy* 16(1), 1-43.
- Pereira, C. (2008). *Le parler arabe de Tripoli (Libye): Phonologie, morphosyntaxe et catégories grammaticales*. Paris: INALCO Doctoral dissertation.

Syntactic change and DLM in German: a corpus study

This study argues that the variation in the placement of relative clauses (RC) in German can be explained by the principle of dependency length minimization (DLM), which states that languages tend to place syntactically and semantically related words close together (e.g., Gibson, 2000, Futrell et al., 2015). Although Gildea & Temperley (2010) ascribe only a weak effect of DLM to German grammar, my diachronic corpus study will show that we can see a strong effect of DLM by looking at the RC position in the history of written German. I will testify to a two-stage change: The syntactic complexity reflected by a high number of intraposed RC first increases in the 17th and 18th centuries due to the growing standardization of the written language, which favors verb-final structures and reaches its peak at the end of the 18th century (Admoni, 1967), before it decreases continuously so that extraposition becomes the most frequent word order.

According to Hawkins (1994, 2004), word order variation in sentence production can be explained by processing economy constraints. Specifically, he states that, given alternative word orders within a construction, the one that allows faster recognition of the immediate constituents is expected to be preferred. Gibson (1998, 2000) takes a similar approach concerning sentence comprehension. He establishes that the memory cost is higher the more incomplete syntactic dependencies one has to keep track of during sentence processing. In German word order, auxiliary and main verb are placed discontinuously in a sentence. According to DLM, there should be a tendency to reduce the distance between the verbs by outsourcing heavy NPs behind the finite verb to the end of the sentence to avoid an overload of the working memory capacity.

Addressing the question of whether DLM plays a role in explaining the changing RC position in German, I investigate the placement of heavy NPs and PPs with relative clause modifiers from newspaper texts from 1600 until Present-day German. The relevant structures are intraposition where the RC is placed adjacent to its head noun (1) and extraposition where the nominal head appears preverbally and the RC occurs at the right edge of the sentence (2).

- (1) Von deß Orators Leuten **sollen** 2. [_{PP} vnter einem tumult/ [so die Türcken deß Nachts in jhrem Losament angefangen]] **vmbkommen sein** (1609: Relation)
‘Two people of the orator are said to have been killed during a tumult that the Turks started in their accommodation at night.’
- (2) sonst **weren** in Spannia auch [_{NP} Ampassatores vom König Matthiasen vnd Hertzogen von Savoia] **angelangt** / [so beim König schon Audienz gehabt] (1609: Relation)
‘Furthermore, ambassadors of King Matthias and of the Duke of Savoy had also arrived in Spain, who had already had an audience with the king.’

An intraposed word order with a long dependency (i.e., a high amount of language material between the verbs) leads to high working memory costs and therefore runs into the risk of processing difficulties. Differently, extraposition can facilitate sentence processing because the dependency length (DL) between the verbs is minimized. Complicating matters, however, is that the DL between RC and its antecedent is increased with extraposition, thus, presenting potentially competing motivations.

To measure syntactic complexity, I consider the factors (i) RC length, (ii) distance between the discontinuous verbs, and (iii) distance between the RC and its antecedent. My results show a strong effect of DLM on the development of German word order preferences: both DLs (between RC and its antecedent and between the separated verbs) decrease over time significantly. On the one hand, extraposition becomes the most frequent word order which leads to reduced distances between the discontinuous verbs over time. On the other hand, the DL between the extraposed RC and the antecedent decreases because it becomes the norm that no more than two words can occur here. This can be explained by the processing pressure in spoken modalities that influence the written standard in newspapers over time increasingly.

References

- Admoni, Wladimir G. 1967. Der Umfang und die Gestaltungsmittel des Satzes in der deutschen Literatursprache bis zum Ende des 18. Jahrhunderts. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 89, 144–199.
- Futrell, Richard, Kyle Mahowald, & Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33), 10336-10341.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95–126.
- Gildea, Daniel, and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science* 34, 286–310.
- Hawkins, John A. 1994. A Performance Theory of Order and Constituency. *Journal of Linguistics*, 33(1), 171–212.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Spread the German new(s): third-person reflexive *zich* in 17th-century Dutch newspapers

In Middle Dutch, reflexivity was commonly expressed by using personal pronouns (third person singular: *haar* ‘her’ or *hem* ‘him’; plural: *haar* ‘them’ or *hen/hun* ‘them’), sometimes followed by *zelf* ‘self’ (e.g. Mooijaart & Van der Wal, 2011: 46). Although attested from the 10th century onwards, especially in the eastern parts of the Low Countries, the specific reflexive form *zich*(*zelf*), currently used in Standard Dutch in both singular and plural third person, only became the standard third-person reflexive pronoun during the seventeenth century (e.g. Van der Sijs, 2021: 429).

This reflexive pronoun is generally accepted as being a grammatical borrowing from the neighbouring High German *sich* (e.g. Postma, 2011: 139; Van der Sijs, 2021: 429-430), yet the factors determining its expansion in the Dutch speaking world still remain subject of discussion (as summarised in Bennis, 2005). On the one hand, research suggests diffusion from above, through formal written texts, in particular religious texts from Germany (e.g. Hermodsson, 1952; Van der Wal & Van Bree, 2008: 214-215; Nobels, 2013: 115-121). On the other hand, Boyce-Hendriks’s study (1998: 209-224) suggests dissemination from below, through the informal spoken language of the many German immigrants settling in the Low Countries in the sixteenth and seventeenth centuries.

In this presentation, we intend to contribute to this ongoing debate by presenting the results of a study testing a third hypothesis: diffusion via newspapers. Newspapers as a genre came into existence in Germany in the early seventeenth century and quickly found their way to the Low Countries (Van Oostendorp & Van der Sijs, 2019: 23-76). Although we do not have exact circulation figures, circumstantial evidence indicates that Dutch newspapers, which were set in Dutch and not in Latin, were meant for and read by many people, from all social classes (Van Groesen, 2016). Newspapers can therefore be seen as the first mass medium (Van Oostendorp & Van der Sijs, 2019: 61), and could thus have served as the ideal vehicle for the dissemination of *zich*, especially given their close connection to German newspapers (e.g. Der Weduwen, 2017).

We test this hypothesis using the new Couranten Corpus (2022), a corpus containing Northern Dutch newspapers from all years between 1618 to 1700 (ca. 19 million words). Within these newspapers, we examine what third-person singular and plural reflexive constructions were used: the previously mentioned uses of personal pronouns and *zich* (both with and without *zelf*), Low German *sick* (Van der Wal, 2018), and possessive pronouns followed by *eigen* ‘own’ or *zelf* ‘self’ (see Weijnen, 1965: 49). We track the distribution of these constructions both over time as well as regarding the place of origin of the article that contains it, focussing in particular on German places of origin. Not only did newspapers originate in Germany, Dutch newspapers also contained many news reports from German speaking areas, and, at the start of the century, news reports were sometimes directly translated from German (Van Oostendorp & Van der Sijs, 2019: 29-31). Although we do not know the exact extent of this German ‘borrowing’, it is safe to assume that early news reports from German speaking areas were either translated from German, or came from a correspondent who lived in Germany and had heard the news from a German source, which both could have significantly influenced the use of *zich*, especially given the presumed lack of editing of incoming news reports (Demske, 2022). Comparing those news reports to domestic news and news reports from other language areas, as well as changes in their distributions over time, will thus provide valuable insight into (1) the influence of German on the use of *zich*, and (2) the possible role of newspapers in spreading this new form.

References

- Bennis, H. (2005). Een Duitse expansie. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 121(3), 201-210.
- Boyce Hendriks, J. (1998). *Immigration and linguistic change: a socio-historical linguistic study of the effect of German and southern Dutch immigration on the development of the northern Dutch vernacular in 16th/18th century Holland*. Madison. (unpublished dissertation).
- Couranten Corpus (version 1.0) (May 2022) [Online Service]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-u9>
- Demske, U. (2022). Variation Across Newspapers in Early Modern German: Degrees of Syntactic Complexity. *Journal of Historical Syntax*, 6(13-18). <https://doi.org/doi:10.18148/hs/2022.v6i13-18.136>
- Der Weduwen, A. (2017). *Dutch and Flemish Newspapers of the Seventeenth Century, 1618-1700 (2 Vols.)*: Brill. <https://doi.org/https://doi.org/10.1163/9789004341890>
- Hermodsson, L. H. (1952). *Reflexive und intransitive Verba im älteren Westgermanischen*. Uppsala: Almqvist und Wiksell.
- Mooijaart, M., & Van der Wal, M. J. (2011). *Nederlands van Middeleeuwen tot Gouden Eeuw* (2nd rev. ed.). Nijmegen: Uitgeverij Vantilt.
- Nobels, J. (2013). *(Extra)Ordinary letters: A view from below on seventeenth-century Dutch*. University of Leiden. Utrecht: LOT.
- Postma, G. (2011). Language contact and linguistic complexity—the rise of the reflexive pronoun *zich* in a fifteenth-century Netherlands border dialect. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical Change: Origins, Nature, Outcomes* (pp. 139-159): Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199582624.003.0007>
- Van der Sijs, N. (2021). *Taalwetten maken en vinden: het ontstaan van het Standaardnederlands*. Gorredijk: Sterck & De Vreese.
- Van der Wal, M. J. (2018). Early Modern migrants in a language contact setting: Characteristics of the Dutch Heusch correspondence (1664–1665). *Journal of Historical Sociolinguistics*, 4(2), 253-280. <https://doi.org/doi:10.1515/jhsl-2017-0029>
- Van der Wal, M. J., & Van Bree, C. (2008). *Geschiedenis van het Nederlands*. Houten: het Spectrum.
- Van Groesen, M. (2016). Reading Newspapers in the Dutch Golden Age. *Media History*, 22(3-4), 334-352. <https://doi.org/10.1080/13688804.2016.1229121>
- Van Oostendorp, M., & Van der Sijs, N. (2019). *Een mooie mengelmoes: meertaligheid in de Gouden Eeuw*. Amsterdam: Amsterdam University Press.
- Weijnen, A. A. (1965). *Zeventiende-eeuwse taal* (5th ed.). Zutphen: N.V. W.J. Thieme & Cie.

The use of “thanks” and “to thank” in Old Saxon and Old High German

Modern German counts on different nouns, verbs, adjectives, and prepositions to express gratitude. Many of these elements, such as the noun *Danke* (thanks) and the verb *danken* (to thank) were already common as early as in the 8th century (Pfeifer 1993). However, the use of these nouns and verbs in expressions of thankfulness have been studied only synchronically and in comparative studies (Zborowski 2005; Siebold 2021). The goal here is to provide a qualitative analysis of the expression of thanking with *Danke* and *danken* at the earlier stage of the history of German, filling a long-standing gap in the literature. More specifically, this investigation seeks to answer the following research questions:

1. How were *Danke* and *danken* used in Old Saxon and Old High German?
2. What can the use of *Danke* and *danken* tell us about historical expressions of thanking in Old Saxon and Old High German?

To answer these questions, I examined all the instances of *Danke* and *danken* in a corpus of religious and secular Old Saxon and Old High German texts taken from the *Referenzkorpus Altdeutsch* (Reference corpus Old German). The targeted structures of this investigations are as showed in (1), (2), (3), and (4):

- (1) *ôlat sagde themu the these uuerold giscôp*
 thank said to the one that this world created
 ‘He said thank to the one who created this world’
 (Hêliand, XLIX, 4091)
- (2) *endi gode thancode, sagde them ôlat*
 and god thanking, said (he) to him thank
 ‘He was thanking god and said thank to him’
 (Hêliand, LVI, 4633)
- (3) *er nú ana wánc hábet fora góte thanc*
 he now without doubt has before god thanc
 ‘He has to say thanks to god now without doubts’
 (Evangelienbuch, 20, 6)
- (4) *thir thánkon mit wórton joh mit wérkon*
 you I thank with words and with deeds
 ‘I thank you with words and deeds’
 (Evangelienbuch, 24, 91)

The data were extracted using the lemma search function offered by the online corpus. The frequency of the instances and the addressers and addressees involved in the expressions of thanking with *Danke* and *danken* were analyzed.

Regarding the first question, a total of 76 instances of *Danke* and *danken* were found, and most of them were in religious texts (71 in total). The data show that the words *Danke* and *danken* were mostly used to express gratitude, but there were instances in which they were found with the meanings of “reward/to reward” and “praise/to praise.” These results show how, Old Saxon and Old High German

speakers relied on the same linguistic elements to carry out different speech acts and how, probably, these speech acts may have been perceived similar to each other.

Turning now to the second research question, the analysis of the addressers and addressees involved in these instance shows how, in the texts selected, and independently of their genre (religious vs. secular), the expression of thanking (but also praising, and reward) with *Danke* and *danken* could only be found when the addresser was in lower social position than the addressee. These results seem to suggest that, in Old Saxon and Old High German, these types of expression could be carried out only when addressees and addressers had an asymmetrical relationship.

In conclusion, these results suggest that use of *Danke* and *danken* in expressions of thanking (but also in praising and rewarding) were restricted to specific communicative settings in which the social status of the speakers played a major role.

Selected references

Pfeifer, Wolfgang. 1993. *Etymologisches Wörterbuch des Deutschen. Digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache.*

<https://www.dwds.de/d/wb-etymwb>.

Siebold, Kathrin. "Implicit and explicit thanking in Spanish and German." *Linguistic Insights-Studies in Language and Communication* 132 (2012): 155-172.

Zborowski, Piotr. 2005. *Dankbarkeit vs. Höflichkeit und sprachliche Routine. der Dankakt im Schwedischen verglichen mit Polnisch und Deutsch.* Wydawnictwo Naukowe UAM.

What is *ke* and if so how many?

The Persian modal particle *ke* and its diachronic development

Keywords: Modal particles, discourse markers, common ground, uncontroversality, grammaticalisation, pragmaticisation, Indo-Iranian, Turkic

In this paper, I will argue that the polysemous Persian conjunct and relative pronoun *ke* (PIE *k^wís/^w*k^wós) has acquired the additional function of a modal particle in colloquial Persian. This stands in direct contrast to previous accounts which have rendered the particle a focus particle or an emphatic marker (Oroji and Rezaei 2013; Sadat-Tehrani 2002), claims that can be refuted based on the examples and restrictions of *ke* presented here. Due to their impalpable nature, modal particles (henceforth MPs) and discourse markers (DMs) were long neglected in linguistic research but have increasingly gained traction in recent years. While both MPs and DMs express the attitudes of the speaker towards a proposition - for which they have also been called "Würzwörter" ('words that add spice') in German - and do not change its truth value, MPs differ from DMs in that they are often well integrated in a phrase and have scope over only the proposition they appear in. The MP at hand *ke* can both appear after the topic as well as at the right periphery. Akin to the German MP *doch*, the basic properties of *ke* can be summarized as follows: ADVERSITY, (also sometimes referred to as CORRECTION (Döring 2016)), UNCONTROVERSIALITY (also described as COMMON GROUND/BACKGROUND) (Grosz 2016; Diewald 2006), and what I refer to as RELEVANCE/SALIENCE, such that:

$[[ke]](p) = p$ and the current question q stands in conflict with p which the speaker renders uncontroversial/part of the common ground but is retrieved for purposes of salience (cf. (Grosz 2016))

Consider the following examples: Person A: Shall I make lamb curry for Ali? Person B:

- | | |
|--|---|
| <p>(1) Ali <i>ke</i> gusht nemikhore. Ali <i>ke</i> meat eat:NEG.3SG.PRS (But) Ali doesn't eat meat.</p> | <p>(2) Ali isst doch gar kein Fleisch. Ali eat.3SG <i>doch</i> no meat. (But) Ali doesn't eat meat.</p> |
|--|---|

In the example given, *ke* highlights the adversity of p toward the current question q (eating a dish containing meat) and that person B thinks p should be known to A (common ground) but that p was apparently not salient enough or momentarily forgotten (relevance). In a quest to answer the hitherto unanswered question of how *ke* might have acquired the function of a MP, I will propose a diachronic development from a conjunction through a process of grammaticalisation along the path of (referential function) → (text-connective function) → (discourse function) as suggested by Diewald (2006) based on (Traugott 1995, 1999). I will argue that this development was especially facilitated by the deictic function of conjuncts as described by Hentschel (1986) and Diewald (2006) and the coordination of Persian subjunctive sentences in the form [A co][B] (Haspelmath 2004).

As shall become clear, the existence of an Old Turkish emphatic particle *är-ki* seems to seriously challenge this theory at first as it has been argued that the Turkish modal particle *ki* is derived from this OldkTurkish particle (Karakoc 2010) thereby insinuating that the Persian MP is in fact borrowed from Turkish and not vice versa. However, this theory can be dismissed on the basis of further, comparative evidence from the North-Afghan Badākhshāni dialect of Persian as well as due to the strong anchoring of second position MPs in other Indo-Iranian languages such as Pashto *kho*, Urdu *to* and Marathi *tər* (Bayer 2018; Deo 2022). Not only are these particles almost identical in function to *ke* but, coincidentally, are all also used as conjuncts denoting "but" and "then"/"so" respectively. This strongly suggests similar paths of grammaticalisation triggered perhaps by an Indo-Iranian predisposition for this kind of development. Even if one is to dismiss a development of MPs *out of* conjuncts, one cannot deny the intricate relationship that exists between the two, a matter worthy of further investigation.

References

- Bayer, Josef (2018): 'A Comparative Note on the Bangla Particle "to" and the German Particle "doch"', *Personal Essays, Poems, and Scholarly Papers* .
- Deo, Ashwini (2022): 'Could be stronger: Raising and resolving questions with Hindi =to', *Language* **98**(4), 716–748.
- Diewald, Gabriele (2006): 'Discourse particles and modal particles as grammatical elements', *Approaches to Discourse Particles* pp. 403–425.
- Döring, Sophia (2016): *Modal Particles, Discourse Structure and Common Ground Management. Theoretical and Empirical Aspects..* Phd, Humboldt-Universität zu Berlin.
- Grosz, Patrick G (2016): Discourse particles. In: D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann and T. E. Zimmerman, eds, *The Wiley Blackwell Companion to Semantics*. Wiley, Oxford.
- Haspelmath, Martin (2004): Coordinating constructions. In: M. Haspelmath, ed., *Typological Studies in Language*. Vol. 58, John Benjamins, Amsterdam/Philadelphia, pp. 3–39.
- Hentschel, Elke (1986): *Funktion und Geschichte deutscher Partikeln: Ja, doch, halt und eben*. Max Niemeyer Verlag, Tübingen.
- Karakoc, Birsal (2010): Mutmaßungen über die Etymologie des türkischen Suffixes KI. In: H. Boetschoten and J. Rentzsch, eds, *Turcology in Mainz*. 82 edn, Wiesbaden, pp. 101–112.
- Oroji, Mohammad Reza and Amir Rezaei (2013): 'Exploring 'ke' as a Focus Particle in Persian from both Form and Function Points of View', *Australian Journal of Linguistics* **33**(1), 76–84.
- Sadat-Tehrani, Nima (2002): 'The indifference-ke construction in modern conversational Persian', *Linguistica atlantica* **24**, 43–69.
- Traugott, Elizabeth Closs (1995): Subjectification in grammaticalization. In: S. Wright and D. Stein, eds, *Subjectivity and Subjectivisation*. Cambridge University Press, Cambridge, pp. 31–54.
- Traugott, Elizabeth Closs (1999): The Rhetoric of Counter-Expectation in Semantic Change: A Study in Subjectification. In: A. Blank and P. Koch, eds, *Historical Linguistics and Cognition*. Mouton de Gruyter, Berlin, New York, pp. 61–89.

The development of future-referring constructions (in Indo-European languages)

Abstract

The investigation of the evolution of grams encoding futurity both in individual languages or language families (cf., e.g., Fleischman 1982, Botne 1998, Whaley 2000) and in typological comparison (e.g., Ultan 1978, Dahl 1985, Bybee & Dahl 1989, Heine & Kuteva 2002: 331) has focused on their grammaticalization from various source constructions with etyma such as ‘will’, ‘have’, ‘become’ and motion verbs (‘go’, ‘come’). According to the “source determination” hypothesis, the origin of a gram determines its syntactic and semantic restrictions in its further development. Another frequent source for future-referring verb forms is hypoanalysis, i.e., a process by which peripheral and/or contextually conditioned functions of a construction, or a categorial opposition, become its inherent property due to a reduction in the inherited functional range (Croft 2000: 126f.); typical cases are North Slavic perfective present (PFV.PRS) > pfv. future (Wiemer 2020: 275f.) and the Latin futures in *-ē-* (e.g., *legēs* ‘you will say’), and *-b-* (continuing the SBJV of the root **b^huH-* ‘be’, e.g., 3SG *amā-bit* ‘will love’) and the future of the copula (*erō, eris, erit* etc.), which are based on a preceding subjunctive (Meiser 2006: 199, Weiss 2020: 441f.). This raises the question whether these two types of futures show systematic differences in their grammatical behaviour, e.g., regarding the frequent function of futures to also encode deontic, epistemic and other kinds of modality.

Using a dataset from Indo-European languages comprehensive over time and space, the paper will discuss the following hypotheses: (a) Contrary to the assumption of source determination, morpho-syntactic and semantic/pragmatic restrictions related to the source construction only pertain at an intermediate stage of the development of futures, whereas fully developed futures overcome these and display a characteristic range of functions (first of all, modal and illocutionary functions) independent of their source; (b) Futures arisen from hypoanalysis differ from futures based on grammaticalization in that they are stable with respect to their initial restrictions (e.g., regarding aspect); (c) The diachronic relation between future meaning and modal, especially epistemic, function is not unidirectional, i.e., either may precede the other. This applies to both types of futures; (d) Futures either stay what they are, or they disappear, but they do not develop into anything else (a “post-future” stage). In particular, if epistemic readings develop as a “sub-function” of futures, they do not oust the future reading as the default meaning.

Assumption (a) is largely confirmed by our dataset. Hypothesis (b) is motivated from parallel observations concerning, e.g., English futures (*going to* contracts to *gonna*, the latter is no longer available as reference to physical movement; *will* no longer codes volition, apart from archaic formula) and PFV.PRS > pfv. future in North Slavic: while the future reading is salient (as a default), non-deontic modal readings are still widely available. The latter, in turn, are among the dominant readings of PFV.PRS in South Slavic, while South Slavic futures based on WANT no longer code volition (see Engl. *will*) and are now morphologically or lexically distinct from WANT (see (1)). Concerning (c), languages differ as for whether their well-established futures are used for epistemic judgment referring to the moment of speech or not, regardless of the diachronic pathway of the future: for instance, Germ. *werden* and Span. *tener* do allow for such usage (see (2)), while Engl. *will*, futures in North Slavic and in Lithuanian do not. Moreover, Balkan Slavic has epistemic uses of future markers (see above), also in combination with *da*-clauses (= irrealis marking; see (3)), but the chronological relation to future meaning requires clarification, also in comparison to Engl. *will* (cf. Ziegeler 2006) and High Alemannic (vs Standard German) *werden* (*Schweiz. Idiotikon* 16, 1344, 1346-7, cf. <https://www.idiotikon.ch/woerterbuch/idiotikon-digital>). As for (d), no IE language shows a “post-future” stage for a “surviving” future, unless as a suppletive form in another paradigm (e.g., Span. 2SG.PRS *eres* ‘you are’ continuing Latin *eris* ‘you will be’; Lausberg 1972: 3.251, Penny 1993: 181; 2014: 191). We also discuss whether the lack of post-future stages might be an IE. feature.

Examples

- (1) South Slavic futures based on WANT: Bulg. *šte*, Mac. *ke*, Srb.-Cr. inflected *ć-u*, *ć-eš*, *ć-e...* – distinct from contemporary WANT: Bulg. *iskam*, Mac. *sakam*, Srb.-Cr. *hoć-u*, *hoć-eš*, *hoć-e...*
- (2) Germ. *Er wird gerade seinen Vortrag halten* ‘He must be having his lecture right now’
(personal knowledge)
- (3) Bulg. *Šte da ima poveče ot pedeset*. ‘S/He must be older than fifty.’
(Tomić 2006: 476)

References

- Botne, Robert. 1998. The evolution of future tenses from serial ‘say’ constructions in Central Eastern Bantu. *Diachronica* 15-2, 207-230.
- Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13, 51-103.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford, New York: Basil Blackwell.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language. Diachronic evidence from Romance*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 1998. The semantic development of old presents: new futures and subjunctives without grammaticalization. *Diachronica* 15-1, 29-62.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Lausberg, Heinrich. 1972. *Romanische Sprachwissenschaft. III. Formenlehre*. 2nd ed. Berlin, New York: Walter de Gruyter.
- Meiser, Gerhard. 2006. *Historische Laut- und Formenlehre der lateinischen Sprache*. 2. Aufl. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Penny, Ralph. 1993. *Gramática histórica del Español*. Barcelona: Editorial Ariel.
- Penny, Ralph. 2002[2014]. *A History of the Spanish Language*. 2nd ed. Cambridge: Cambridge University Press.
- Tomić, Olga Mišeska. 2006. *Balkan Sprachbund Morpho-syntactic Features*. Dordrecht: Springer.
- Ultan, Russell. 1978. The nature of future tenses. In: Greenberg, Joseph H., Charles A. Ferguson & Edith Moravcsik (eds.). *Universals of Human Language. Vol. 3. Word Structure*. Stanford: Stanford University Press, 83-123.
- Weiss, Michael L. 2009. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor: Beech Stave Press.
- Whaley, Maryka Lynn. 2000. *The evolution of the Slavic ‘be(come)’-type compound future*. Ohio: The Ohio State University (PhD thesis).
- Wiemer, Björn. Grammaticalization in Slavic. In: Bisang, Walter & Andrej Malchukov (eds.): *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies. Vol. 1: Grammaticalization scenarios from Europe and Asia*. Berlin, Boston: De Gruyter Mouton, 249-307.
- Ziegeler, Debra. 2006. Omnitemporal will. *Language Sciences* 28, 76-119.

A re-assessment of Early Runic Metrics

This paper on early runic metrics systematically excludes inscriptions of the transitional period and the Viking Age. Thus, the focus is on the metricity of the unsyncopated, linguistically archaic runic inscriptions of the older period AD 150-500/550. This has the advantage of systematically excluding syncopated or partially syncopated inscriptions such as the Eggja stone and the Blekinge inscriptions. Due to their linguistic status, these inscriptions will have to be subjected to a separate study.

The author takes a minimalist standpoint by first taking the position of the *Advocatus Diaboli*. This means that metrical criteria of the older runic inscriptions should speak for themselves without being directly derived from or equated with later language stages with their elaborated metrical systems. In other words, the present approach attempts to avoid constrictions and direct comparisons with the Old Germanic literary languages, especially Old Norse and Old High German. This is so because the language typology of Ancient Nordic cannot directly be equated with Old Norse in terms of syllabic metrics (cf. Pascual 2016 on Old English). Sievers' typology (Sievers 1893) is also rejected here as a straightforward tool for analysis, since it is oriented towards the Old Germanic literary languages. This raises the basic question of scientifically verifiable regularities of early runic metrics that meet criteria of validity and reliability.

The leitmotif of this analysis is the fusion and interaction of language, sentence rhythm and alliterative metrics. Not entirely unexpectedly, the Germanic long line emerges as the basic unit of Ancient Nordic metrics, since it can be relatively reliably verified in a small group of older runic inscriptions. Finally, this unit is identified as the proto-long line of North-West Germanic metrics.

Keywords

Early Runic metrics, Germanic long-line, proto-long-line, Germanic verse, older runic inscriptions, Sievers' s metrical types, formulaicity, resolution, heaviness requirement.

Micro-level conflict in the productivity of anticausativization strategies Evidence from the history of Icelandic

The objective of this paper is, first, to argue that different morphosyntactic strategies used to form anticausatives are productive in different periods of Icelandic, a language characterized by rich inflection. Second, we show how shifts in anticausativization strategies coincide with factors in other domains of grammar (cf. Cennamo 2022). Our analysis is carried out on the basis of the diachronic variation exhibited by selected predicates in historical corpora (e.g. ONP, RMH). Expanding on Ottosson (2013), we identify five different causative-anticausative patterns in Old and Modern Icelandic (1a–e); the case frames for the relevant predicates are given within square brackets.

- (1) a. Weak-strong alternation: caus. *sökkti* ‘sank’ [Nom–Acc] : anticaus. *sökk* ‘sank’ [Nom]
 b. *na*-verbs: caus. *braut* ‘broke’ [Nom–Acc] : anticaus. *brotnaði* ‘broke’ [Nom]
 c. Case-Preserving Anticausativization (CPA): caus. *hvessti* ‘sharpened’ [Nom–Acc] :
 anticaus. *hvessti* ‘got windier’ [Acc]
 d. Lability: caus. *lokaði* ‘closed’ [Nom–Dat] : anticaus. *lokaði* ‘closed’ [Nom]
 e. *st*-predicates: caus. *opnaði* ‘opened’ [Nom–Acc] : anticaus. *opnaðist* ‘opened’ [Nom]

Two of these strategies, (1a) and (1b), ceased being productive already in pre-Old Icelandic (before 1150 AD). Their unproductivity is reflected in the fact that occasionally some predicates may form an anticausative in more than one way (2a), or occur with double marking simultaneously (2b):

- (2) a. OldIcel *sökk-ti-st* ‘sunk’, a weak form with *-st* for an older strong form *sökk* in (1a)
 b. ModIcel *brot-na-ði-st* ‘broke’, with both *-na-* and *-st* instead of *brot-na-ði* in (1b)

The remaining three strategies (1c–e) show productivity to a varying degree in Modern Icelandic, engaging in a “micro-level conflict” against each other. We posit a hierarchy of strategies for the modern language such that the *st*-strategy (1e) is selected provided the relevant *st*-form does not already have another function (reflexive, reciprocal and denominals). The productivity of this strategy is clearly boosted by the ubiquity of the *st*-suffix with all kinds of verb formations. In cases where the *st*-form is dispreferred or unavailable, a different strategy (CPA or lability) is selected. The strategy of labile verbs (1d) seems to be gaining ground, having previously been rather limited (witness new verbs in the semantic domain of technology like *hlaða* ‘charge’ and *starta* ‘start’). In some cases its rise may be “accidental”, due to a common change called Nominative Substitution, whereby an oblique subject is replaced by a nominative (e.g. Svavarsdóttir 1982). However, even in Modern Icelandic, the emergence of new oblique subjects can still be triggered by CPA (1c).

In conclusion, the diachrony of anticausativization strategies in Icelandic suggests that in this highly inflected language, morphological marking with anticausatives ((1a–c), (1e)) is preferred over an unmarked option (1d). Nevertheless, various conflicting factors create a micro-level tension in the grammar, affecting the productivity of the different anticausativization strategies. Thus, Nominative Substitution replaces oblique subjects and the otherwise very productive *st*-suffix reaches a point of saturation due to its multifunctionality, leaving the labile option as the most viable one in Modern Icelandic.

References

Cennamo, Michela. 2022. Paths of change in the diachrony of anticausatives in Romance: the rise and fall of labiality. Paper presented at ICHL25, Oxford, 1–5 August 2022.

ONP = *Ordbog over de danske sprogsprog*. Copenhagen: Den Arnamagnæanske Kommission (eds). <<https://onp.ku.dk/onp/onp.php>>

Ottosson, Kjartan. 2013. The anticausative and related categories in the Old Germanic languages. In Folke Josephson & Ingmar Söhrman (eds.), *Diachronic and Typological Perspectives on Verbs*, 329–381. Amsterdam: John Benjamins.

RMH = Risamálheild [Gigaword Corpus]: Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. *Proceedings of LREC 2018*, 4361–4366, Japan. <<https://malheildir.arnastofnun.is/?mode=rmh2022>>

Svavarsdóttir, Ásta. 1982. „Þágufallssýki“: Breytingar á fallnotkun í frumlagssæti ópersónulegra setninga [“Dative Sickness”: Changes in the use of subject case in impersonal sentences]. *Íslenskt mál og almenn málfræði* 4:19–62.

The totalizing function of the Vedic particle *cid*

The interpretation of particles is often exceedingly difficult, especially in extinct languages (cf. Goldstein 2019:269–271). These problems become even more grave in an attempt to reconstruct the functions of a certain particle in Proto-Indo-European.

The enclitic particle $*=k^h id$ can be safely reconstructed for Proto-Indo-European. Its most productive reflexes are attested in Indo-Iranian but other Indo-European branches exhibit reflexes as well (cf. Dunkel 2014:448–451). In order to reconstruct not only the form but also the functions of this particle properly, thorough synchronic examinations of its reflexes in the oldest stages of the languages in which it is reflected are necessary, in particular those in which it is used productively.

This paper is concerned with a detailed analysis of one reflex, namely the particle *cid* in the Rigveda, the oldest Indo-Iranian text. It is attested there 691 times. In the Rigveda, *cid* can fulfill a number of functions, among others that of an additive focus particle ‘even, also’ (e.g. Grassmann 1873:454f., Lühr 2017:283–285). This paper will concentrate on one function of this particle which until now has not received proper treatment in the literature, namely its totalizing function.

It is a well-known fact that the particle *cid* is cliticized to interrogative proforms in order to form indefinites, e.g. *kás cid* ‘some, any, every’. However, some scholars assume that also *cid* alone possesses such a function (e.g. Gonda 1954–1955:281). Thus, *cid* itself has also been translated as ‘all’ (e.g. Grassmann 1873:455). I will argue that although such a translation of *cid* is adequate in certain contexts such occurrences are to be differentiated from those of the particle in indefinite proforms. For even though indefinites formed by interrogatives and additive particles like *cid* are typologically widespread, additives themselves usually do not function as indefinites or universal quantifiers (König 2017:40; Ying 2017:218–226). I assume that instead of being a quantifier *cid* possesses a function which according to Forker (2016:84–86) is also attested for additive particles in other languages and may be called ‘totalizing’. This means that it emphasizes that all elements of a set are referred to, but I will argue that in spite of the resemblance to universal quantification, *cid* is not a genuine quantifier. I will show that this function can be identified after numerals, universal quantifiers, demonstratives, the pronominal adjective *anyá-* ‘other’ and possibly after multiplicative adverbs.

I will also discuss another context where I assume this function of *cid*, namely after *purá* ‘before, of old’. When *purá* occurs with a verb in the perfect or present it expresses a norm or habit in the past (Mumm 2004:55–61). There, I consider it to emphasize that the event or state denoted by the predicate is true for the entire time span that is loosely defined by *purá*. In this case, *cid* can be translated as ‘always’ but again, unlike the English adverb, it does not quantify over a set of time points because the habituality is already expressed without the particle.

The results of the synchronic investigation of Vedic *cid* have consequences for the reconstructed semantics of Proto-Indo-European $*=k^h id$. Dunkel (2014:451) assigns it an additive and a generalizing function. However, both the Vedic and typological data suggest that the generalizing function is not part of the semantics of the particle itself but that it only occurs in combination with interrogative proforms. In order to corroborate this assumption and to determine whether $*=k^h id$ also possesses a totalizing function further synchronic analyses of its reflexes are necessary.

References

- Dunkel, George E. 2014. Lexikon der indogermanischen Partikeln und Pronominalstämme: Band 2. Lexikon. Universitätsverlag Winter, Heidelberg.
- Forker, Diana. 2016. Toward a Typology for Additive Markers. *Lingua* 180, 69–100.
- Goldstein, David, 2019. Discourse Particles in LSJ: A Fresh Look at $\gamma\epsilon$, In: Stray, Christopher, Clarke, Michael, Katz, Joshua T. (Eds.), Liddell and Scott. *The History, Methodology, and Languages of the World's Leading Lexicon of Ancient Greek*. Oxford University Press, Oxford, pp. 268–287.
- Gonda, Jan. 1954–1955. Notes on the Indo-European k^u i- and k^u o-pronouns. *Lingua* 4, 241–285.
- Grassmann, Hermann. 1873. *Wörterbuch zum Rig-Veda*. F.A. Brockhaus, Leipzig.
- König, Ekkehard. 2017. Syntax and Semantics of Additive Focus Markers from a Cross-linguistic Perspective: A Tentative Assessment of the State of the Art, In: De Cesare, Anna-Maria, Andorno, Cecilia (Eds.), *Focus on Additivity. Adverbial Modifiers in Romance, Germanic and Slavic Languages*. John Benjamins Publishing Company, Amsterdam, Philadelphia, pp. 23–43.
- Lühr, Rosemarie. 2017. Stressed and Unstressed Particles in Old Indic, In: Bayer, Josef, Struckmeier, Volker (Eds.), *Discourse Particles. Formal Approaches to their Syntax and Semantics*. Walter de Gruyter GmbH, Berlin, Boston.
- Mumm, Peter-Arnold. 2004. Altindisch *sma*. *International Journal of Diachronic Linguistics* 1, 19–86.
- Ying, Zhang, 2017. The Categorization of *Dou* (都) in Chinese: A Study from a Cross-linguistic Perspective. *Bulletin of Chinese Linguistics* 10, 214–234.

The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective

Two central questions regarding language evolution are a) how representations of linguistic structures emerge in individuals and b) how emergent linguistic structures spread throughout communities and became conventionalised (Author 1a). The concept of “protolanguage” (Tallerman 2012) or “early languages” (Heine & Kuteva 2007) is often evoked to bridge the gap between fully complex linguistic structures and an earlier stage of structured communication. However, questions a) and b) hold for the evolution of protolanguage just as much as they do for fully complex modern language. This paper discusses theoretical models of the interaction of two important dimensions and their potential for shedding light on these two questions from a usage-based perspective: 1) the individual cognitive level, in which processes of automation and entrenchment lead to the emergence of structured linguistic representations, and 2) the community level, in which processes of ritualisation and conventionalisation lead to the diffusion and stabilisation of usage patterns within communities of practice (Schmid 2020; Author 2).

First, regarding the level of individuals, domain-general cognitive processes such as automation and analogy are central to how structures emerge, are represented, and are stored in the minds/brains of individuals. The parallelism of dialogic interaction (Du Bois 2014) invites speakers to share (ad-hoc) structures and repeat them in future usage events. Frequent repetition promotes structures to be entrenched and automated as complex constructions that, over time, emancipate from their original concrete referential uses. That is, assisted by further domain-general capacities like chunking and metaphor, more schematic representations with procedural (i.e. grammatical) functions may emerge. Entrenchment and its associated effects have been implicated in the emergence of structure in language learning (Schmid 2016), language change (Hilpert 2017) and language evolution (Author 1b, Author 1a).

Secondly, community-wide processes such as conventionalisation and ritualisation refer to the establishment, negotiation and diffusion of community-wide, regularised practices (Schmid 2020; Author 2). Processes of conventionalisation and ritualisation depend on processes of social transmission interacting with individual processes of entrenchment, and as such represent an important part of explaining how communicative structures came to be adopted both in proto(linguistic) communities. Conventionalisation also represents the foundation for cumulative culture (Tomasello 1999), which enables the cumulative accretion of changes and increasing structuration of communicative systems. This in turn enabled the gradual change from protolanguage and early languages to modern human language (Heine & Kuteva 2007).

The channel of interaction between entrenchment processes and conventionalisation processes is usage, whereby rough alignments between mental form–meaning pairings and communal linguistic norms arise. In summary then, the present paper will present the interaction of the individual cognitive dimension on the one hand, and the community level on the other, as well as their underlying mechanisms, as crucial building blocks of a usage-based approach to the evolution of language.

References:

- Du Bois, J.. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3). 359–410.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction*. Oxford University Press.
- Hilpert, M. (2017). *Frequencies in diachronic corpora and knowledge of language*. In M. Hundt, S. Mollin, & S. E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives* (pp. 49-68). Cambridge University Press.
- Schmid, H.-J.. (Ed.). (2016). *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge*. De Gruyter.
- Schmid, H.-J.. (2020). *The dynamics of the linguistic system: usage, conventionalization, and entrenchment*. Oxford University Press.
- Tallerman, M. (2012). Protolanguage. In M. Tallerman & K. R. Gibson (eds.), *The Oxford handbook of language evolution* (pp. 479-491). Oxford University Press.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

The development of number strengthening in German declensional classes. A diachronic-dialectal corpus study

Keywords: Historical morphology, diachronic dialectology, plural marking, corpus linguistics, German

The present study addresses the implementation of overt marking of number in the plural forms of German nouns belonging to declensional classes which originally lacked number distinctivity. Initially, the phonological reduction of final syllables (cf. Braune 2018: 248-249, 265) had led to a formal overlapping of singular and plural forms within several of the subclasses of the original vocalic stems, most importantly within the paradigms of the neuter *a*-stems, see OHG/MHG nom./acc.sg. *wort* – nom./acc.pl. *wort* ‘word – words’, as well as within the feminine *ō*-stems, see OHG nom./gen./acc.sg. *gēba* – nom./acc.pl. *gēba*, MHG nom.-acc.sg. *gebe* – nom./acc.pl. *gebe* ‘gift – gifts’ (cf. Braune 2018: 250, 265; Klein et al. 2018: 73; Ronneberger-Sibold 2013: 19). In addition, there was massive overlapping within the paradigms of the consonantal *n*-stems in Old and Middle High German, where the oblique cases in the singular coincided with the plural forms, all ending in *-(e)n*, see OHG gen./dat./acc.sg. *zungūn* – nom.-acc.pl. *zungūn* (cf. Braune 2018: 282), MHG gen./dat./acc.sg. *zungen* – nom.-acc.pl. *zungen* ‘tongue – tongues’ (cf. Klein et al. 2018: 73). In the following periods, an expansion of overt plural markers took place reaching its peak in Early New High German. This reinforcement of number distinctivity is generally referred to as *Numerusprofilierung*, i.e. the strengthening of number distinctions, and is one of the most significant processes in the history of German nominal morphology. It consists of various strategies aiming to create number distinctivity. Two basic kinds of strategies can be distinguished:

- 1) changes in the plural paradigm to create number distinctivity, e.g. various overt plural markers are assigned to originally indistinctive forms, sometimes accompanied by the process of *Umlaut*, as in NHG sg. *Wort* – pl. *Worte/Wörter* ‘word – words’
- 2) changes in the singular paradigm as in the case of the *n*-stems where the oblique cases of the singular were identical with the plural ones as in MHG *herze* ‘heart’, gen./dat.sg. *herzen* – nom.-acc.pl. *herzen* (cf. Klein et al. 2018: 73) and the *n*-suffix is reanalyzed as a marker of plurality and removed from the singular paradigm cf. NHG sg. *Herz* – pl. *Herzen*

On the basis of the reference corpora of the historical periods of the German language, the reference corpus for Middle High German, *Referenzkorpus Mittelhochdeutsch* (REM), covering the period from 1050 to 1350 and the reference corpus for Early New High German, *Referenzkorpus Frühneuhochdeutsch* (REF), covering the period from 1350 to 1650 (cf. Klein/Dipper 2016; Dipper/Kwekkeboom 2018), the present study explores the rich meta-annotation provided in the corpora to investigate the diachronic and dialectal spread of plural morphology within members of declensional classes that were originally number indistinct. This investigation is part of a larger study that aims to account for the precise diachronic and dialectal representation of the individual processes pertaining to the overall phenomenon of *Numerusprofilierung* in German. It also serves to explore the methods of retrieving and compiling the relevant data from the different reference corpora to generate a multidimensional map of the spread of these processes over time and space.

References

- Braune, Wilhelm (2018): Althochdeutsche Grammatik. I. Laut- und Formenlehre. 16. Aufl. Neu bearbeitet von Frank Heidermanns. Berlin/Boston: De Gruyter.
- Dipper, Stefanie/Kwekkeboom, Sarah (2018): Historische Linguistik 2.0. Aufbau und Nutzungsmöglichkeiten der historischen Referenzkorpora des Deutschen. In: Kupietz, Marc/Schmidt, Thomas (ed.): Korpuslinguistik. Berlin/Boston: De Gruyter, p. 95-124.
- Enger, Hans-Olav (2014): Reinforcement in inflection classes. Two cues may be better than one. *Word Structure* 7.2, p. 153–181.
- Klein, Thomas/Dipper, Stefanie (2016): Handbuch zum Referenzkorpus Mittelhochdeutsch. Bochumer Linguistische Arbeitsberichte 19. Bochum: Ruhr-Universität Bochum.
- Klein, Thomas/Solms, Hans-Joachim/Wegera, Klaus-Peter (2018): Mittelhochdeutsche Grammatik. Pt. 2. Flexionsmorphologie. Vol. 1. Substantive, Adjektive, Pronomina. Berlin/Boston: De Gruyter.
- Ronneberger-Sibold, Elke (2013): Einführung in die historische Morphologie des Deutschen. In: *Jahrbuch für ungarische Germanistik*, S. 13-34.

Corpora

- Klein, Thomas/Wegera, Klaus-Peter/Dipper, Stefanie/Wich-Reif, Claudia (2016): Referenzkorpus Mittelhochdeutsch (1050–1350), Version 1.0, <https://www.linguistics.ruhr-uni-bochum.de/rem/>. ISLRN 332-536-136-099-5.
- Wegera, Klaus-Peter/Solms, Hans-Joachim/Demske, Ulrike/Dipper, Stefanie (2021). Referenzkorpus Frühneuhochdeutsch (1350–1650), Version 1.0, <https://www.linguistics.ruhr-uni-bochum.de/ref/>. ISLRN 918-968-828-554-7.

The Lost Cause: Inflection Class in Amarasi

1. Overview While the origin and development of inflection classes has long been an area of interest for historical linguistics, the majority of research in this area has focused solely on Indo-European languages (Maiden 2005, Collier 2013, Kaye 2015). With this gap in mind, this paper presents a novel perspective on the source of these classes by building on the observation that many Timoric languages appear to have undergone low-level innovations which have given rise to morphological inflection classes (which cannot be reconstructed for Proto-Austronesian).

In particular, this paper investigates an incipient inflection class system in Amarasi (Central Malayo-Polynesian: West Timor), which has two distinct paradigms of prefixal subject agreement that take the shapes CV (1a) and C (1b). The distribution of these prefix sets is partly phonologically predictable as in (1c) on the basis of regular phonotactic constraints (e.g. a ban on CCC clusters and cross-morpheme hiatus, and a dispreference for quadrisyllabic words).

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|----------------------------|-----------------------------|------------|--|---|---------------|------------|-------|--|--|------------|-------|---|------------|------------|--|---|------------|------------|--|--|--|----|----|--|---|-------------|-----------|-------|--|--|-----------|-------|---|-----------|-----------|--|---|-----------|-----------|--|---|-------|--------|-----|---------|-----|----|----------------|-----------|----|---|--------------|----------|------|---|-----------------|------------|
| (1) a. Syllabic Agreement (CV) | b. Asyllabic Agreement (C) | c. Phonotactic Distribution | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <table border="0"> <tr> <td></td> <td>SG</td> <td>PL</td> <td></td> </tr> <tr> <td>1</td> <td><i>ku/?u-</i></td> <td><i>ta-</i></td> <td>(IN.)</td> </tr> <tr> <td></td> <td></td> <td><i>mi-</i></td> <td>(EX.)</td> </tr> <tr> <td>2</td> <td><i>mu-</i></td> <td><i>mi-</i></td> <td></td> </tr> <tr> <td>3</td> <td><i>na-</i></td> <td><i>na-</i></td> <td></td> </tr> </table> | | SG | PL | | 1 | <i>ku/?u-</i> | <i>ta-</i> | (IN.) | | | <i>mi-</i> | (EX.) | 2 | <i>mu-</i> | <i>mi-</i> | | 3 | <i>na-</i> | <i>na-</i> | | <table border="0"> <tr> <td></td> <td>SG</td> <td>PL</td> <td></td> </tr> <tr> <td>1</td> <td><i>k/?-</i></td> <td><i>t-</i></td> <td>(IN.)</td> </tr> <tr> <td></td> <td></td> <td><i>m-</i></td> <td>(EX.)</td> </tr> <tr> <td>2</td> <td><i>m-</i></td> <td><i>m-</i></td> <td></td> </tr> <tr> <td>3</td> <td><i>n-</i></td> <td><i>n-</i></td> <td></td> </tr> </table> | | SG | PL | | 1 | <i>k/?-</i> | <i>t-</i> | (IN.) | | | <i>m-</i> | (EX.) | 2 | <i>m-</i> | <i>m-</i> | | 3 | <i>n-</i> | <i>n-</i> | | <table border="0"> <tr> <td>SHAPE</td> <td>PREFIX</td> <td>3SG</td> <td>EXAMPLE</td> </tr> <tr> <td>#CC</td> <td>CV</td> <td><i>na-skau</i></td> <td>‘invites’</td> </tr> <tr> <td>#V</td> <td>C</td> <td><i>n-inu</i></td> <td>‘drinks’</td> </tr> <tr> <td>σσσ+</td> <td>C</td> <td><i>n-marine</i></td> <td>‘is happy’</td> </tr> </table> | SHAPE | PREFIX | 3SG | EXAMPLE | #CC | CV | <i>na-skau</i> | ‘invites’ | #V | C | <i>n-inu</i> | ‘drinks’ | σσσ+ | C | <i>n-marine</i> | ‘is happy’ |
| | SG | PL | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | <i>ku/?u-</i> | <i>ta-</i> | (IN.) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | <i>mi-</i> | (EX.) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | <i>mu-</i> | <i>mi-</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | <i>na-</i> | <i>na-</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | SG | PL | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | <i>k/?-</i> | <i>t-</i> | (IN.) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | <i>m-</i> | (EX.) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | <i>m-</i> | <i>m-</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 3 | <i>n-</i> | <i>n-</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| SHAPE | PREFIX | 3SG | EXAMPLE | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| #CC | CV | <i>na-skau</i> | ‘invites’ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| #V | C | <i>n-inu</i> | ‘drinks’ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| σσσ+ | C | <i>n-marine</i> | ‘is happy’ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

2. Two classes However, **disyllabic #C** verbs appear to be idiosyncratically and lexically-specified to take one of the two prefix sets. Edwards (2020) observes that 75% of these verbs take C prefixes (2a, 3a) while 25% take CV prefixes (2b, 3b). Semantically-unrelated (near-)minimal pairs as in (2, 3) show that prefix set selection for verbs of this shape is not outwardly phonologically-conditioned, suggesting the existence of two conjugation classes.

| | | | |
|---------------------------|---------------------------|---------------------------|------------------------|
| (2) a. In n -reku. | b. In na -reku. | (3) a. Ho m -hani. | b. Ho mu -hana. |
| 3SG 3SG.C-strike | 3SG 3SG.CV-ruin | 2SG 2SG.C-dig | 2SG 2SG.CV-cook |
| ‘It strikes (x o’clock).’ | ‘S/he ruins, besmirches.’ | ‘You dig.’ | ‘You cook.’ |

3. Analysis I propose that the distribution of CV prefixes originates in verbs which took the PMP causative **pa-* prefix, and that the current system is the result of an interaction between historical antepenultimate vowel syncope and phonotactic constraints on cluster formation. Comparative data from neighbouring Rote languages (Tamelan 2021, Edwards 2021) shows that stem-initial #CC clusters in Amarasi originate from the application of antepenultimate vowel syncope to known derivational prefixes (4). This syncope clearly also derived the C prefix set from the original CV forms (e.g. 3SG **na-CVCV* > *n-CVCV*). Crucially missing from our comparanda is the highly productive and well-attested PMP causative prefix **pa-*, which should have produced Amarasi **h-*. Despite allowing many typologically unusual sonority-falling clusters like /fk/, /ft/, /mt/, Amarasi shows a complete absence of #hC. I argue that the descendants of these missing **pa-C* > **hC* clusters are verbs which synchronically take CV prefixes.

| | | | | | | | |
|-----|-------------|--------------------|----------------|--------------------|-------------------|-----------------|-------------------|
| | PMP | | Rote | | Amarasi | | |
| | Prefix | Function | Root | Verb | Meaning | Verb | Meaning |
| (4) | <i>*ka-</i> | achieved state | <i>*bəntəŋ</i> | <i>na-ka-bete</i> | ‘is tense, tight’ | <i>na-kbeet</i> | ‘is stiff, tight’ |
| | <i>*ma-</i> | stative | <i>*lapaR</i> | <i>na-ma-laʔa</i> | ‘is hungry’ | <i>na-mnaha</i> | ‘is hungry’ |
| | <i>*ta-</i> | spontaneous action | <i>?*belaj</i> | <i>na-ta-mbele</i> | ‘flies’ | <i>na-tpene</i> | ‘flies’ |

I propose that the phonotactically illicit **hC* was repaired by deleting the overt featural content of **h* but retaining its representation in phonological structure as a ghost consonant \emptyset_C (Piggott 1991; Kiparsky 2003). Recalling how #CC verbs require CV prefixes (1c), I posit that this silent etymological \emptyset_C on #C verbs created a covert #CC cluster that similarly blocks C prefixes. Thus, a verb like *na-hana* ‘cooks’ (3b) goes back to PMP **pa-panas* ‘make hot’ > **na-ha-hana* > *na- \emptyset_C hana*. Strong evidence for this analysis comes from i) the existence of verbs which show an (anti)-causative alternation as expressed only by a change in prefix set (5), captured straightforwardly by a derivational history as in (6); and ii) the retention of a small set of fossilised forms where exceptionally unsyncopeated causative *ha-* alters with stative *ma-* (Edwards 2020: 445), confirming the existence of **pa-* in Proto-Amarasi (7).

| | | | | | | | | |
|-----|---------------------|-----------------|----------------------|---------------|---------------------|----------|-----------------------|-------------|
| | C Prefix Set | | CV Prefix Set | | Stative Noun | | Causative Verb | |
| | Verb | Meaning | Verb | Meaning | Verb | Meaning | Verb | Meaning |
| (5) | <i>n-ʔate</i> | ‘serves (ITR.)’ | <i>na-ʔate</i> | ‘enslaves’ | <i>mainuan</i> | ‘open’ | <i>n-hainua-b</i> | ‘opens’ |
| | <i>n-mae</i> | ‘is ashamed’ | <i>na-mae</i> | ‘shames s.o.’ | <i>maʔekiʔ</i> | ‘smooth’ | <i>n-haʔeki</i> | ‘smoothens’ |
| | <i>n-peaʔ</i> | ‘breaks (ITR.)’ | <i>na-peaʔ</i> | ‘breaks s.t.’ | <i>maʔkafaʔ</i> | ‘light’ | <i>n-haʔkafa</i> | ‘lightens’ |
| | <i>n-punu</i> | ‘rots, decays’ | <i>na-punu</i> | ‘makes rot’ | <i>maputuʔ</i> | ‘hot’ | <i>n-haputu</i> | ‘heats up’ |

- (6) a. PMP **buRuk* ‘rotten’ > **na-punu* →_{syncope} *n-punu* ‘rots’
- b. PMP **pa-buRuk* ‘makes rotten’ > **na-ha-punu* →_{syncope} **na-h-punu* →_{CC red.} *na-punu* ‘makes rot’

References

- Collier, Scott James. 2013. *The evolution of complexity in Greek noun inflection*. Guildford: University of Surrey dissertation.
- Edwards, Owen. 2020. *Metathesis and Unmetathesis: Parallelism and Complementarity in Amarasi, Timor*. Berlin: Language Science Press.
- Edwards, Owen. 2021. *Rote-Meto Comparative Dictionary*. Canberra: The Australian National University Press.
- Kaye, Steven. 2015. *Conjugation class from Latin to Romance: Heterocclisis in diachrony and synchrony*. Oxford: University of Oxford dissertation.
- Kiparsky, Paul. 2003. Finnish noun inflection. In Diane Nelson & Satu Manninen (eds.), *Generative approaches to Finnic and Saami Linguistics*, 109–161. Stanford, CA: CSLI Publications.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert Booij & Jaap Marle (eds.), *Yearbook of Morphology 2004*, 137–175. Dordrecht: Springer.
- Piggott, Glyne L. 1991. Empty onsets: evidence for the skeleton in prosodic phonology. *McGill Working Papers in Linguistics/Cahiers linguistiques de McGill* 7. 41–71.
- Tamelan, Thersia. 2021. *A grammar of Dela; an Austronesian language of Rote in Eastern Indonesia*. Canberra: The Australian National University dissertation.

West Germanic 2.sg. *-st* Revisited: The Role of Supervescence

The Uniformitarian Principle restrains historical linguists who are reconstructing a proto-language not only from positing typologically unattested linguistic configurations (e.g., a verb system with 2 tenses in the indicative vs. 22 in the subjunctive, or a phonological system with no low vowels at all), but also from positing typologically unattested sociolinguistic situations: e.g., a system that lacks variation, or one without any stigmatized forms. When a proto-language with stigmatized forms (due, e.g., to hypercorrection) divides into several descendant languages such that each inherits those stigma-bearing items, all the ingredients are in place for subsequent loss of the stigma to allow the forms at issue to appear late (at a time subsequent to the initial stage of attestation) in the written texts of many or even all of the descendants. This delayed appearance of once-stigmatized but later-accepted forms can be called *supervescence* [a haplogogized blend of *super*+(*fer*)*vescence*], because it indeed involves, as it were, the bubbling up to the surface of formerly submerged forms. This phenomenon is already well-established for the attested stages of languages' histories: e.g., certain taboo-words have taken centuries to appear in print.

Joseph 2006, 2012, 2013 adduces evidence for plausible cases of such “bubbling up” in the histories of Germanic and Indo-Iranian languages, emphasizing that supervescence obviates an appeal to convergent Sapirian “drift” as an explanation for independent parallel developments in related languages. Janda & Joseph 2023 point out that supervescence (as one innovation, later multiply inherited) not only is more economical than “co-drift” (as multiple innovations, often across all descendant languages), but also is nearer to Sapir's own 1921 notion of drift — which includes several types, one of them involving the persistence and spread of patterns from a proto-language into its descendants (a parallel to pattern-persistence and -spread within one language). It must be stressed that distinct distributions and chronologies of eventually supervescent forms in different descendants of a common ancestral language are expected, because destigmatization itself is governed by sociolinguistic conditions that are unique to each descendant language.

Here, we argue that the appearance of 2.sg. *-st* (< earlier *-s*) in multiple West Germanic [WGmc.] languages is more comprehensible when analyzed as an instance of the supervescence of a once-stigmatized Proto-WGmc. [PWGmc.] form that was originally due to hypercorrection, not as an instance of co-drift whereby *-s* independently and repeatedly became *-st*. Crucially, WGmc. *-st* in OE and OHG has frequently been treated as a case of independently convergent development (cf., e.g., Greenberg 1957, Sihler 1986, Ringe 2002, Ringe & Taylor 2014), but *-st* actually appears in at least one stage of all five major WGmc. languages, hence also in both Old and Middle Frisian, Old (but not Middle) Dutch, and Middle Low German [though not (earlier) Old Saxon]. We agree fully with Greenberg 1957 (and other specialist works, such as Wilmanns 1906, Campbell 1962, Brunner 1965, Paraschkeow 2003, Braune & Reiffenstein 2004) that the pivotal hypercorrection was the addition of a /t/ after a verb-final *-s* that preceded a 2.sg. pronoun starting with *þ...* (or the like, depending on the language) — cf. OHG *gilaubist thū* < *gilaubistū* < *gilaubis thū* ‘believest thou’, e.g. — except that, unlike those authors, we locate this reanalysis in PWGmc. Lausberg 1972 cites a Romansch parallel: *chantast (tu)* ‘singingest thou’ < *cantas tu*.

Sihler 1986, Ringe 2002, and Ringe & Taylor 2014 treat the inversion context *...-s + þū* as a red herring in the rise of 2.sg. *-st* in OE: they view that configuration as syntactically minor, and analyze the analogical model of OE (plus OHG) preterite-present verbs already ending in *-st* (like OE *wāst* ‘thou knowest’) as rendering inversion structures irrelevant. However, we show, first, that the syntactic environments where the specifically relevant inversion occurs are some of the most basic in OE, and, second, that the preterite-presents (and similar verbs) in *-st* can have been extremely relevant to the reanalysis of *-s* as *-st* without having been the SOLE reason for it. Further, given the predominant view that PWGmc. was already a V2 language, we can project the hypercorrection-related inversion back into the proto-language, along with the various verbs already marked with 2.sg. *-st* — thereby further strengthening the case for supervescence.

References

- Braune, [Theodor] Wilhelm & Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik, Band I: Laut- und Formenlehre*. Tübingen: Max Niemeyer Verlag. (Sammlung kurzer Grammatiken germanischer Dialekte, A: Hauptreihe, 5. [Fifteenth edition; first edition (by Braune alone) 1886.]
- Brunner, Karl. 1965. *Altenglische Grammatik, nach der Angelsächsischen Grammatik von E[duard] Sievers*. Tübingen: Max Niemeyer Verlag. (Sammlung kurzer Grammatiken germanischer Dialekte, A: Hauptreihe, 3.) [Third edition; first edition 1942 (based on Sievers' grammar: third edition 1898, first edition 1882).]
- Campbell, Alistair. 1962. *Old English Grammar*. Oxford, UK: Clarendon Press. [Second, corrected hardback edition (first hardback edition 1959); paperback edition 1983, reprinted 1991.]
- Greenberg, Joseph H[arold]. 1957. "The Problem of Linguistic Subgroupings". Chapter 4, pp. 46-55, in his *Essays in Linguistics*. Chicago, IL: University of Chicago Press.
- Janda, Richard D. & Brian D. Joseph. 2023. "Drift as a Centenarian Vagabond: Initially Floated in 1921 but Still an Unmoored Concept". Paper presented at the 35th Annual Meeting of the North American Association for the History of the Language Sciences (NAAHoLS); Denver, CO; January 6-7.
- Joseph, Brian D. 2006. "On Projecting Variation Back into a Proto-Language, with Particular Attention to Germanic Evidence". On pp. 103-118 in Thomas D. Cravens (ed.), *Variation and Reconstruction*. Amsterdam: John Benjamins. (Current Issues In Linguistic Theory, 268.)
- Joseph, Brian D. 2012. "A Variationist Solution to Apparent Copying Across Related Languages". On pp. 151-164 in Lars Johanson & Martine Robbeets (eds.), *Copies vs. Cognates in Bound Morphology*. Leiden: Brill.
- Joseph, Brian D. 2013. "Demystifying Drift — A Variationist Account". On pp. 43-66 in Martine Robbeets & Hubert Cuyckens (eds.), *Shared Grammaticalization, with Special Focus on the Transeurasian Languages*. Amsterdam: John Benjamins.
- Lausberg, Heinrich. 1972. *Romanische Sprachwissenschaft III: Formenlehre*. Berlin: Walter de Gruyter & Co. (Sammlung Göschen, 7199.) [Second, revised edition; first edition 1962 (Sammlung Göschen, 1199).]
- Lühr, Rosemarie. 1984. "Reste der athematischen Konjugation in den germanischen Sprachen". On pp. 25-90 in Jürgen Untermann & Béla Brogyányi (eds.), *Das Germanische und die Rekonstruktion der indogermanischen Grundsprache: Akten des Freiburger Kolloquiums der Indogermanischen Gesellschaft; Freiburg [im Breisgau], 26.-27. Februar 1981*. Amsterdam: John Benjamins. (Current Issues in Linguistic Theory, 22.)
- Paraschkewow, Boris. 2003. "Zur Polygenese des -t in der Verbalendung -st". *Zeitschrift für germanistische Linguistik* 31.3:382-385.
- Ringe, Donald A., Jr. 2002. "Syncopated Present Indicative Forms in Old English". On pp. 125-156 in Alfred R. Wedel & Hans-Jörg Busch (eds.), *Verba et Litterae — Explorations in Germanic Languages and German Literature: Essays in Honor of Alfred L. Lloyd*. Newark, DE: Linguatext, Ltd.
- Ringe, Donald A., Jr. & Ann A. Taylor. 2014. *A Linguistic History of English, Volume II: The Development of English*. Oxford, UK: Oxford University Press. (Oxford Linguistics.)
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York, NY: Harcourt, Brace, & World. [Hardback edition; paperback edition 1955.]
- Sihler, Andrew. 1986. "Germanic 2nd-Person Endings in -st". *Münchener Studien zur Sprachwissenschaft* 47:193-222.
- Wilmanns, [Franz] Wilhelm. 1906. *Deutsche Grammatik: Gotisch, Alt-, Mittel- und Neuhochdeutsch, Band 3: 3. Abteilung, Flexion; 1. Hälfte, Verbum*. Strasbourg: Verlag Karl J. Trübner. (First and second edition [sic].)

From *de* to *ke*: functional transfer of a topic shift marker from Turkish to Cappadocian Greek

This presentation discusses how the particle *ke* is used as a topic shift marker in Cappadocian Greek (CG), which I argue is the result of contact-induced language change. CG is a cluster of closely related, critically endangered dialects of the Greek language that were spoken in the Turkish region of Cappadocia until the Greek-Turkish Population Exchange of 1924 (Janse 2020). Due to intensive influence of the Turkish language, CG became a textbook example of language contact cited by many scholars as a case of heavy borrowing (Thomason & Kaufman 1988: 215-22; Thomason 2001: 74; Winford 2003: 83-4; 2005: 402-9; Matras 2020: 231-4).

A salient example of this strong Turkish influence is the extension of the functional range of the CG proclitic particle *ke*. In CG, *ke* functions as a coordinating (‘and’), additive (‘also’) and scalar-additive (‘even’) particle, as was already the case in Ancient and Medieval Greek (Beekes & Van Beek 2010: 615). However, a new function appears in the following examples:

- (1) *k’ ekino ke lex: “irta na vró ta tría güzélja.”*
 and DEM.NOM **and** say.PRS.3SG
k’ ekín ke lex: “aderé éxo éks peðjá.”
 and DEM.NOM **and** say.PRS.3SG
 ‘And he says, “I came to find the three Fair Ones.”
 And she says, “Now I have six sons.”’ (Dawkins 1916: 306-7)

Alongside a first *k(e)*, which functions here as a coordinating particle, a second, seemingly redundant *ke* is introduced, usually positioned between the topic and the verb of the sentence. I argue that this specific use of *ke* marks a topic shift, which is the result of ‘functional transfer’ (FT) from the Turkish enclitic particle *dA*. FT is defined by Siegel (2012: 189) as “applying the grammatical functions of a morpheme from one language (the Source Language) to a morpheme in another language (the Recipient Language)”. More specifically, the transfer from *dA* to *ke* is a case of ‘type II’ (2012: 194-8), meaning that “[t]he functional range of a grammatical item or construction is extended and/or reduced” (Jennings & Pfänder 2018: 91).

In this case, the functional range of the Greek (RL) proclitic particle *ke* is extended and incorporates the function of topic shift marker associated with the Turkish (SL) enclitic particle *dA*, based on already existing similarities with Greek *ke*: both are used not only as a coordinating particle, but also as a (scalar-)additive particle (Göksel & Kerslake 2005: 110; Dawkins 1916: 605). Additionally, Turkish *dA* also functions as a topic shift marker, unlike Ancient and Medieval Greek *ke*:

- (2) *Ben sinema-ya git-ti-m. Ahmet de tiyatro-ya git-ti-Ø.*
 I cinema-DAT go-PFV-1SG Ahmet **TOP** theatre-DAT go-PFV-3SG
 I went to the cinema. **As for** Ahmet, he went to the theatre.

It is this function of topic shift marker that is included into the functional range of the CG particle *ke*. The analysis of this specific use of the particle was carried out in a corpus of 58 CG folktales (ca. 50,000 words, the largest CG text collection to date). It should be noted that topic shift marker *ke* is only found in combination with the verb *lé(γ)ο* ‘say’. This can be explained by the distribution of the Turkish model *dA*, which introduces a new topic “without changing the direction of the discourse” (Göksel & Kerslake 2005: 513). Without this continuity or connection between two events or situations, other topic shift markers, like the clitic *ise* would be used (Göksel & Kerslake 2005: 401). Subsequently, since a reported dialogue (introduced by the verb *lé(γ)ο* ‘to say’) between constantly shifting subjects is a prototypical example of switching topics without changing the action in the predicate, this could explain why the use of CG *ke* is seemingly restricted to this specific verb.

This paper contributes to the research on language contact by investigating the mechanisms by which the Greek dialect in Cappadocia – which was isolated from the rest of the Greek-speaking world from the 11th century onwards – underwent significant changes under the strong influence of the surrounding Turkish *superstratum*. Additionally, it contributes to the overall knowledge of this underdocumented Greek dialect at the brink of extinction.

References

- Beekes, R.S.P. & Van Beek, L. (2010). *Etymological dictionary of Greek*. Brill.
- Dawkins, R.M. (1916). *Modern Greek in Asia Minor*. Cambridge University Press.
- Göksel, A. & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. Routledge.
- Janse, M. (2020). Ἑλληνιστὶ γινώσκεις; (Acts 21:37): The survival of Cappadocian Greek. *Cursor: Zeitschrift für Freunde der Lateinischen Sprache und europäischen Kultur*, 16, 49-57.
- Jennings, W. & Pfänder, S. (2018). *Inheritance and Innovation in a Colonial Language: Towards a Usage-Based Account of French Guianese Creole*. Palgrave-McMillan.
- Matras, Y. (2020). *Language contact* (2nd ed.). Cambridge University Press.
- Siegel, J. (2012). Two Types of Functional Transfer in Language Contact. *Journal of Language Contact* 5: 187-215.
- Thomason, S.G. & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. University of California Press.
- Thomason, S.G. (2001). *Language contact. An introduction*. Edinburgh University Press.
- Winford, D. (2003). *An introduction to contact linguistics*. Blackwell.
- Winford, D. (2005). Contact-induced change: classification and process. *Diachronica*, 22(2), 373-427.

A computational approach to detect discourse traditions and register differences: a case study on historical French

Historical sociolinguists have demonstrated the crucial role of register/genre in mediating the spread of innovations throughout language communities (Nevalainen and Raumolin-Brunberg 2017). However, the traditional conceptualization of genre has been challenged by the concept of Discourse Traditions (Kabatek 2005), henceforth DTs.

The core idea in the DTs framework is that language is not a monolithic object and one cannot dispense with the impact of textual traditionality to study the evolution of individual phenomena (Kabatek 2005). Additionally, the detection of DTs represents a challenge for quantitative corpus linguistics, as each texts can allow for global or internal classifications (Kabatek 2013: 19). Although previous research has discussed distinctive classification features for textual genre, descriptions might be biased by a researcher's particular interest or object language. It is therefore worthwhile to explore whether thorough philological analysis can be complemented by bottom-up generated classifications.

The first goal of our paper is to leverage on the popularity of computational models for the semantic representation of words and texts, so-called 'vector space models' (Boleda 2020), for the unsupervised, bottom-up identification of DTs. Document-based vector space models represent a document's content by means of a frequency profile (i.e., vector) of the terms occurring therein. Afterwards documents can be compared by calculating a similarity value based on those frequency vectors. The rationale is that the co-occurrence of certain terms in a document will be correlated with certain DTs. For this endeavor we explore a corpus of 1400+ historical French theater plays dated between 1600 and 1930 (Author 2023). This corpus is annotated in terms of sub-genres (e.g., comedy, tragedy, pastoral, etc.), which might correlate with different registers.

The second goal is to verify how the automatic classification of documents improves or complements a traditional genre classification provided by the corpus metadata. Building on previous work on Spanish (Author *in press*), we evaluate this comparison by checking the impact of both classifications on a case study of syntactic alternation in French, namely the distribution and change in inverted (1) and clefted (2) interrogatives.

- (1) *Aimez-vous voyager?*
'Do you like to travel?'
- (2) *Est-ce que vous aimez voyager?*
'Is it that you like to travel?'

By including these two different operationalizations of DTs in a logistic regression, we show how the bottom-up classification (a) improves the overall fit of the regression model, (b) reveals unattested differentiation within theater texts, and (c) functions as a principled approach to distinguishing 'change from above' and 'change from below'. Overall, the proposed approach evidences the relevance of computational-semantic methods for historical (socio)linguistic research.

References

Author 2023.

Authors in press.

Boleda, Gemma. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*. 6(1). 213–234.

GITHE. 2015. Codea+ 2015. *Corpus de documentos españoles anteriores a 1800*.

- Kabatek, Johannes. 2005. Tradiciones discursivas y cambio lingüístico. *Lexis: Revista de lingüística y literatura* 29(2). 151–177.
- Kabatek, Johannes. 2013. ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania* 77(1).
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 2017. *Historical Sociolinguistics : Language Change in Tudor and Stuart England*. London: Routledge.

Where do all the NPs go? – A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900

Although in modern German, it is highly marked to place an NP in the postfield,¹ the phenomenon is not as rare as expected in early New High German (1650-1900) data (ex. A).

- A. ...weil er [...] von den meisten Medicis [genennet wird]_{RSB} **ein Schmid aller Kranckheiten**.
 ... as he ... by the most doctors called is a forger of all sicknesses.
 "...as he is called a forger of sicknesses by most physicians." (Abel 1699, 225)

However, studies concerned with extraposition in diachrony treat the placement of NP as a marginal phenomenon that can nearly exclusively be explained by the length of the NP (Ebert 1980, Sapp 2014) or pragmatic factors like givenness (Light 2011).

Although it is not mentioned as such in these studies, both explanations can be linked to processing difficulties which are resolved by extraposition. Processing difficulties can be rather objectively investigated using Information Density, namely Surprisal (ID; Shannon 1948). Levy and Jaeger (2007; 1) define ID as the "amount of information per unit comprising the utterance". It is calculated as the likelihood with which a word occurs in a context ($P(\text{word}) = -\log_2(\text{word}|\text{context})$). More expected combinations of words result in lower surprisal values and, thus, in lower perceiving difficulties (Hale 2001), as low surprisal values reduce the impact of the working memory (Levy & Jaeger 2007, Hale 2001, Levy 2008). We claim that the surprisal value of NPs is also relevant for their placement in the postfield. Therefore, we propose that NPs with high surprisal values are more likely to be extraposed.

To investigate this claim, we built a corpus of medical and theological texts from 1650 to 1900 taken from the Deutsches Textarchiv (DTA, BBAW 2019). We manually extracted extraposed and embedded NP and the sentence brackets using WebAnno (Eckart de Castilho et al. 2016). Then, we calculated a 2-Skip-Bigram-Language Model (Guthrie et al. 2016) to gain surprisal values for every word in the context. These surprisal values were used to calculate the mean Skipgram surprisal on lemmata for every annotated NP. Furthermore, we determined the length of the NP, the text genre (medical vs. theological), and the Orality Score (COAST, Ortmann & Dipper 2022) since extraposition is claimed to be more likely in conceptionally oral texts (Koch & Oesterreicher 2007) and the time of publication, the period. To determine the most influential factor for extraposition, logistic regression was performed with R (The R Core Team 2022).

As a result, we find that extraposition is indeed linked to high surprisal values ($z=-2.44$, $p<.05$ *) and that length is not significant ($z=-0.48$, $p<.63$), in contrast to the aforementioned literature. However, both the genre ($z=-2.58$, $p<.001$ **) and the interaction between Orality Score and the period ($z=-2.68$, $p<.001$ **) are more significant. That suggests an influence of genre and a change over time. The latter is furthermore supported by a slightly significant result for the interaction between length and period ($z=-1.75$, $p<.1$).

Following Speyer (2015: 499), we suggest that there are more processing capacities available behind the right sentence bracket because the main verb is eventually processed at this point. Thus, there is no uncertainty about the constituent function of the extraposed phrases, which causes further strain on the working memory. This leaves more capacities to process lexical difficulties, represented by the surprisal values. In our corpus, the effect is more pronounced than the influence of length. Furthermore, we detect indications of language change in the interactions and an influence of the genre, suggesting a difference in writing style that could yield further investigations.

¹ The postfield is the position behind the right sentence bracket (RSB) and the RSB is the position late in the clause where verbal material, which is distributed over two positions in the clause in German, occurs (Wöllstein 2014).

Example taken from:

Abel, H. (1699). *Wohlerfahrner Leib-Medicus Der Studenten*. Leipzig: Groschuff.

References:

- BBAW (2019). Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften; <http://www.deutschestextarchiv.de/>. [last accessed: 2023-01-19]
- Ebert, R. P. (1980). Social and stylistic variation in early new high german word order: The sentence frame (>satzrahmen<). 102. Jahresband, 357–398.
- Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A. and Biemann, C. (2016): A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In *Proceedings of the LT4DH workshop at COLING 2016*, Osaka, Japan.
- Guthrie, D., B. Allison, W. Liu, L. Guthrie, and Y. Wilks (2006). A closer look at skip-gram modelling. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.
- Hale, J. 2001. A probabilistic Early parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics*.
- Koch, P. & Oesterreicher, W. 2007. Schriftlichkeit und kommunikative Distanz. *ZGL* 35, 346-375.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106(3), 1126–1177.
- Levy, R. & Jaeger, F. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19. 849-856.
- Light, C. (2011). The information structure of subject extraposition in early new high german. In S. Müller (Ed.), *Proceedings of the HPSG 2011 Conference*
- Ortmann, K. and S. Dipper (2022a). Coast (conceptual orality analysis and scoring tool). <https://github.com/rubcompling/COASTcoast-conceptual-orality-analysis-and-scoring-tool> [last accessed: 2023-01-18]
- R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Sapp, C. D. (2014). Extraposition in middle and new high german. *The Journal of Comparative Germanic Linguistics* 17(2), 129–156.
- Speyer, A. (2015a). Auch früher wollte man informieren – Zum Einfluss der Informationsstruktur auf die Syntax in der Geschichte des Deutschen. *Zeitschrift für germanistische Linguistik* 43(3), 485–515.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* 27(3), 379 – 423
- Wöllstein, A. (2014). *Topologisches Satzmodell* (2 ed.). Heidelberg: Winter.

Pronoun history and information structure in 18th century non-religious Russian texts

The registers - situational and context-dependent language varieties (Biber & Conrad, 2019) - that dominated Russian in the diachronic point of view are religious original and translated texts, legends, and chronicles, which are also often bound to religious canons¹ (Azimova & Johnston, 2012; Comrie, 2018; Vakulenko, 1989). An apparent change came with the age of Enlightenment, which began in Central and Western Europe around the middle of the 17th century. The movement came to the Russian-speaking area in the late 17th-early 18th century and started the process of register development in the written language. The range of different registers in written language became broader and included, among other things, travelogues, letters from ordinary people, diaries, and others (Anciferova, 2012; Kotkov & Pankratova, 1964; Mayorov, 2006). This development period of Russian could be understood as the transition phase to the modern Russian language. The changes that take place at this stage should not be understood as selective changes that happened at a specific moment in the history of the language. Before a specific phenomenon becomes noticeable or written down, there must first be a tendency in the development of language history that leads to this change (Lieb, 1970, p. 43).

In transitional stages of language development, pronouns are sensitive and can indicate more profound linguistic changes. Languages can create new pronouns and pronoun functions over time. Among others, there may be denotational semantic reasons, such as the need to convey refined gender and sex distinctions (Lakoff, 1975). Overall, however, changes in pronouns can also affect the syntax, morphology, and information structure of the sentence or phrase. The historical corpus linguistics and digital humanities methods demonstrate the flexible nature of pronouns. For the history of Russian, Meyer (2011) demonstrated by statistical corpus analysis a shift from a consistent null-subject language to a partial null-subject language over a period of about 700 years (cf. Roberts & Holmberg, 2010 for the typology of null-subject languages). Among others, this change weakened overt pronouns & function from demonstrative, referent-introducing, and contrastive to coreferent, which apparently affected 1st and 2nd person before 3rd person. An appropriate comparative empirical analysis must be based on diachronic corpora in which pronouns are annotated in a unified manner across languages for their syntactic roles, information structural categories (focus, givenness, and topic), as well as coreferential properties and links to their discourse antecedents.

The talk aims at providing the first comprehensive, corpus-based investigation of the historical developments of the grammar and use of pronouns in 18th-century Russian non-religious texts. Focusing mainly on the history of subject pronouns, it uses a combination of grammar- and discourse-oriented methods to investigate how pronominal forms and their functions change over time. The analysis discussed during the talk involves compiling, annotating, and evaluating a corpus for Russian. Developing a general workflow that includes all phases of text processing for corpus linguistic investigations and would be projectable to other languages and language development stages is central to our approach to historical corpus data. Furthermore, explicit text selection principles are developed, allowing for a more objective look at language history and intended to redefine language diachronic definitions of past centuries.

¹ An exception are the birch bark letters from the 11th - 15th century, which represent the spoken register in written form and are now available in a large amount. However, they usually only contain a few lines, phrases, or words (Schaeken, 2019; Zaliznjak, 2004; Zaliznjak, 2006). Texts in other registers also exist, but they are not frequent and only in rare cases originate from a non-religious context.

References:

- Anciferova, O. «Журнал Путешествия...» Демидова Как Памятник Русского Языка Второй Половины XVIII Века (Doctoral Dissertation). Vol. Dissertation. Ekaterinburg: Ural Federal University, 2012.
- Azimova, N., and B. Johnston. 'Invisibility and Ownership of Language: Problems of Representation in Russian Language Textbooks'. *The Modern Language Journal (Boulder, Colo.)* 96, no. 3 (2012): 337–49.
- Biber, Douglas, and Susan Conrad. *Register, Genre, and Style*. Cambridge: University Press, 2019.
<https://doi.org/10.1017/9781108686136>.
- Comrie, Bernard. 'Russian'. In *The World's Major Languages / Editor, Bernard Comrie*, Third edition. London: Taylor and Francis, 2018.
- Kotkov, S., and N. Pankratova. *Istočniki Po Istorii Russkogo Narodno-Razgovornogo Yazyka XVII-Načala XVIII Veka*. Moskva: Nauka, 1964.
- Lakoff, Robin Tolmach. *Language and Woman's Place Text and Commentaries*. New York, 1975.
- Lieb, Hans-Heinrich. 'Sprachstadium und Sprachsystem: Umriss einer Sprachtheorie'. Stuttgart u.a. : Kohlhammer, 1970.
- Mayorov, A. *Региональный Узус Деловой Письменности XVIII Века (По Памятникам Забайкалья)*. Diss. Диссертация... Доктора Филологических Наук/АП Майоров, 2006. Moskva: V.V. Vinogradov Russian Language Institute of the Russian Academy of Sciences, 2006.
- Meyer, Roland. *The History of Null Subjects in North Slavonic. A Corpus-Based Diachronic Investigation*. Regensburg, 2011.
- Roberts, I., and A. Holmberg. 'Introduction: Parameters in Minimalist Theory'. Edited by T. Biberauer and et al. *Parametric Variation: Null Subjects in Minimalist Theory.*, no. Cambridge University Press (2010): 1–57.
- Schaeken, Jos. *Voices on Birchbark: Everyday Communication in Medieval Russia / Jos Schaeken*. Studies in Slavic and General Linguistics ; v. 43. Leiden, Boston: BRILL, 2019.
- Vakulenko, T. *Формирование и Функционирование Административноканцелярской Лексики в Русском Языке XVIII Века*. Kyiv: Kyiv Pedagogical Institute A.M.Gorky, 1989.
- Zaliznjak, Andrej Anatol'evič. 'Berestjanye Gramoty — Bescennyj Istochnik Svedenij o Drevnej Rusi i Ee Jazyke'. *Lekcii Laureatov Demidovskoj Premii (1993-2004)*, 2006, 229–32.
- . *Drevnenovgorodskij dialekt*. 2. izd., Pererabot. s učetom materiala nachodok 1995-2003 gg. Jazyk, semiotika, kul'tura. Moskva: Jazyki Slavjanskij Kul'tury, 2004.

Isoglosses and distributions of features – Analyses of the *Dialectological Atlas of the Russian Language*

The Dialectological Atlas of the Russian Language (abbreviated DARJa based on its Russian title) represents more than 4 decades of data collection, from 1938 onwards, and was published in Moscow during 1986-2005. It contains 313 maps, each corresponding to a linguistic feature, and covers 4196 locations. In 2015-16, researchers at Kazan Federal University extracted linguistic features and their values directly from the physical maps and created Excel files giving the values of features across locations (Isaev et al. 2016). We have processed these materials further, georeferencing the map of locations covered and manually extracting the latitude and longitude location of every location.

Thus, the DARJa data are now amenable to systematic, quantitative analyses. For instance, it is possible to define dialect areas in more principled ways than was hitherto possible. For instance, Zaxarova and Orlova (2004: 166) present a map of 28 dialect zones. A rather similar map can be generated from the DARJa data by computing Hamming distances based on the features present in different locations, classifying the locations in a UPGMA tree and cutting this tree into $k = 28$ clusters. This approach is principled and also versatile, since k can be any number.

The focus in this talk is on two issues of dialectological method, namely how to draw isoglosses computationally and how to measure the similarity of two distributions of feature values. After having binarized all features by taking each feature value as present/ absent we extract isoglosses, as follows. First, we fit a thin plate spline (Franke 1982) to each binarized feature, and produce a spatial interpolation on the region in question (following Wieling et al. 2011 and Guzmán Naranjo and Becker 2021). An example of this can be seen in Figure 1. This figure shows the spatial distribution of so-called *Akan'e* (weakening of unstressed *o*). If the probability of the interpolation is rounded to 0 and 1 we are left with clearly delimited regions that allow for extracting the isoglosses by applying an edge detection procedure to the map. Finding features which have a similar spatial distribution is also performed using the interpolated values. Here we use correlation distances between values across locations. Using these methods for drawing isoglosses and comparing distributions we go on to analyze the interplay between the many phonological, morphological, syntactical, and lexical features in DARJa.

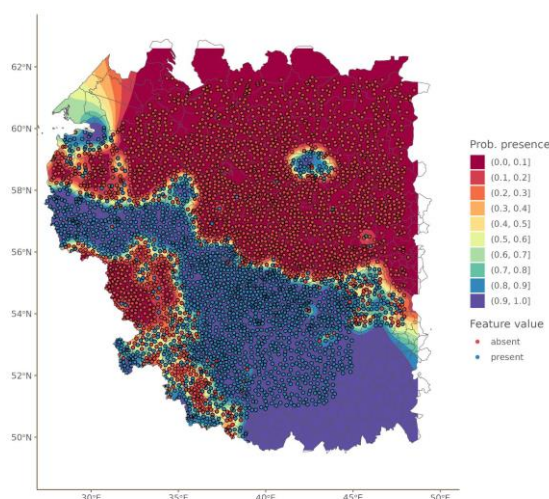


Figure 1: Map of interpolated probabilities for *Akan'e* (weakening of unstressed *o*). The color scheme shows a probability P of presence = 1 as blue vs. $P = 0$ as red, with the intermediate color range representing interpolated values.

References

- Franke, Richard. 1982. Smooth interpolation of scattered data by local thin plate splines. *Computers & Mathematics with Applications* 8.4, 273–281.
- Guzmán Naranjo, Matías & Laura Becker. 2021. Statistical bias control in typology. *Linguistic Typology* 26, 605–670. <https://doi.org/10.1515/lingty-2021-0002>
- Isaev, Igor', Valerii Solov'ev, Farid Salimov, Aleksandr Piljugin & Venera Bairaševa. 2016. Creating a database of Russian dialects and prospects for dialectometric studies. *Herald of the Russian Academy of Sciences* 86(6): 448-453.
- Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PloS ONE* 6, e23613.
- Zaxarova, Kapitolina F. & Varvara G. Orlova. 2004. *Dialektnoe členenie russkogo jazyka*. Moskva: URSS.

Hearsay in Historical German Newspapers (1740–1840)

The century ‘around’ 1800 was a central period in the history of European society. Between the late Enlightenment and the revolutions of 1848, a series of changes fueled by rapidly growing industrialization and urbanization took place, as well as advances in mass education, widespread politicization of the masses, and increasingly pressing requests for democratic policies.

This period was pivotal for the evolution of newspapers. Because of booming alphabetization and the blossoming of the bourgeoisie, newspapers garnered a growing readership who, in time, became increasingly exigent. This led to the professionalization of journalism and journalists. However, this important period and its impact on the language of newspapers have not been fully investigated in the field of the history of the German language.

In the paper, I will delineate the first results of a research project on the connections among changes in society, changes in the text genre ‘report,’ and changes in the formulation of reported speech in German newspapers in the century ‘around’ 1800.

The focus of the paper is the lexical and grammatical marking of information source of hearsay in reported speech. Following Wiemer (2010: 59), hearsay is understood as synonymous with reportive evidentiality, that is, “reported information with no reference to who it was reported by” (Aikhenvald 2018: 12). Hearsay markers in German comprise — but are not limited to — the reportive modal verb *sollen* (a) (Wiemer 2010: 81; Smirnova & Diewald 2013) and indirect speech with the generic 3SG subject pronoun *man* (b) (Jäger 2010: 179), as well as constructions with the verb *hören* (*hear*; Whitt 2009: 1088–1889) and with the reportive particle *angeblich* (*allegedly*; Wiemer 2010: 92).

- a. Sonst **sollen** Se. Königl. Majestät in Ruppin am See, 3 der größten Häuser erkaufte haben (BN 4.10.1740: 2)
[Furthermore, His Royal Majesty **is said to** have purchased 3 of the largest houses in Ruppin on the lake]
- b. **Man sagt**, daß der General Brocon [...] das Commando über die Truppen [...] führen werde; (BN 2.7.1740: 1)
[**It is said** that General Brocon will command the troops]

1. Research hypothesis:

Due to the changes delineated above, newspaper reports became more transparent in handling information and information source (Schröder 2017: 169–172); it is expected that this led to a decrease in the frequency of hearsay and an increase in quotative strategies, that is, “reported information with an overt reference to the authorship of the quoted source” (Aikhenvald 2018: 12).

2. Research object:

Reported speech, understood as any form of direct or indirect reporting of or reference to linguistically encoded content produced in another communicative situation in the *Wiener Zeitung* (WZ) and the *Berlinische Nachrichten von Staats- und gelehrten Sachen* (BN) (1740–1840), with a focus on hearsay.

3. Corpus:

Factual reports (*Kurzbericht*, *faktizierende Meldung*, *Erlebnisbericht*, Wille 2020: 150–155) in the WZ and BN from the years 1740, 1745 – 1770, 1775 – 1800, 1805 – 1830, 1835 (two issues per year).

4. Methodology:

Production of a taxonomy of hearsay markers present in the WZ and BN (1740–1840);

Quantitative investigation of the diachronic changes in frequency of different hearsay markers in comparison to quotative strategies in reported speech.

5. Results:

As expected, in the analyzed time span, hearsay markers became progressively rare. In the WZ issues from 1740 to 1775, between 38% and 47% of all instances of reported speech are hearsay; that is, they do not contain any indication of the information source. In the 19th century, a steady decrease in the frequency of hearsay is observable: In the 1835 issues of the WZ, only 7.5% of all reported speech is hearsay. This diachronic development can be explained fully in light of the changes in media and society in the century ‘around’ 1800.

However, it remains to be discussed why not all markers of hearsay decline at the same rate. Whereas constructions with the generic 3SG subject pronoun *man* decline from being one of the most frequent formulations of reported speech in 1740–1775 (16–20%) to disappearing almost entirely (from 3.2% in 1830 to 0% in 1835), the frequency of the reportive *sollen* remains constant, at ca. 6%.

References:

- Aikhenvald, A.Y., 2007. Information source and evidentiality: what can we conclude? *Italian Journal of Linguistics* 19, 209–227.
- Aikhenvald, A.Y., 2018. Evidentiality: The framework. In: Aikhenvald, A.Y. (Ed.) *The Oxford handbook of evidentiality*. Oxford University Press, Oxford, 1–43.
- Cornillie, B., Marín Arrese, J., Wiemer, B., 2015. Evidentiality and the semantics-pragmatics interface: An introduction. *Belgian Journal of Linguistics* 29, 1–17.
- Fabricius-Hansen, C., 2002. Nicht-direktes Referat im Deutschen - Typologie und Abgrenzungsprobleme. In: Fabricius-Hansen, C., Leirbukt, O., Letnes, O. (Eds.) *Modus, Modalverben, Modalpartikeln*. WVT Wiss. Verl. Trier, Trier, pp. 7–29.
- Habermann, M., 2019. Praktiken des Zitierens in der vormodernen Wissensliteratur. *Zeitschrift für germanistische Linguistik* 47, 102–136.
- Haß-Zumkehr, U., 1998. "Wie glaubwürdige Nachrichten versichert haben": Formulierungstraditionen in Zeitungsnachrichten des 17. bis 20. Jahrhunderts. Narr, Tübingen.
- Jäger, A., 2010. Reported speech constructions and grammaticalization of hearsay as evidentiality: a cross-linguistic survey. *STUF Sprachtypologie und Universalienforschung* 63, 177–195.
- Letnes, O., 2011. Über werden und sollen als modale und/oder evidentielle Marker. In: Diewald, G., Smirnova, E. (Eds.) *Modalität und Evidentialität. = Modality and evidentiality*. Wiss. Verl. Trier, Trier, 109–123.
- Meyer, H.F., 1972. Berlinische Nachrichten von Staats- und gelehrten Sachen (1740-1874). In: Fischer, H.-D. (Ed.) *Deutsche Zeitungen des 17. bis 20. Jahrhunderts*. Verlag Dokumentation, München, 103–114.
- Oggolder, C., 2016. Politik, Gesellschaft, Medien. Österreich zwischen Reformation und Erstem Weltkrieg. In: Karmasin, M., Oggolder, C. (Eds.) *Österreichische Mediengeschichte. Band 1: Von den frühen Drucken zur Ausdifferenzierung des Mediensystems (1500 bis 1918)*. Springer, Wiesbaden, 9–27.
- Reisner, A., Schiemer, A., 2016. Das Wien(n)erische Diarium und die Entstehung der periodischen Presse,. In: Karmasin, M., Oggolder, C. (Eds.) *Österreichische Mediengeschichte. Band 1: Von den frühen Drucken zur Ausdifferenzierung des Mediensystems (1500 bis 1918)*. Springer, Wiesbaden, 87–112.
- Schröder, T., 2017. Information und Meinung. Presstextsorten vor der Trennungsnorm. In: Pfefferkorn, O., Riecke, J., Schuster, B.-M. (Eds.) *Die Zeitung als Medium in der neueren Sprachgeschichte. Korpora - Analyse - Wirkung*. De Gruyter, Berlin/Boston, 165-175.
- Smirnova, E., Diewald, G., 2013. Kategorien der Redewiedergabe im Deutschen: Konjunktiv I versus *sollen*. *Zeitschrift für germanistische Linguistik* 41, 443-471.
- Squartini, M., 2018. Extragrammatical expression of information source. In: Aikhenvald, A.Y. (Ed.) *The Oxford handbook of evidentiality*. Oxford University Press, Oxford, pp. 273–285.
- Whitt, R.J., 2009. Auditory evidentiality in English and German: The case of perception verbs. *Lingua* 119, 1083-1095.
- Wiemer, B., 2010. Hearsay in European languages: towards an integrative account of grammatical and lexical marking. In: Diewald, G., Smirnova, E. (Eds.) *Linguistic realization of evidentiality in European languages*. De Gruyter, Berlin, 59-129.
- Wille, M., 2020. Die Tageszeitung des 18. Jahrhunderts auf dem Weg zum Massenmedium: Eine textlinguistische Untersuchung des „Hamburgischen Correspondenten“. Universitätsbibliothek, Gießen.

Towards Quantifying Social Behavior in Language Contact

This talk investigates social dynamics in contact-induced change using Exemplar Theory (ET, Pierrehumbert 2001). Language contact theory attributes the outcomes of contact to social factors and structural factors but theories differ in the primacy attributed to either. While Van Coetsem (1988) mostly attributes the outcomes to structural considerations (e.g. presence or absence of contrasts), Thomason & Kaufman (1988) mostly attribute outcomes to social ones (e.g. intensity of language contact). Situationally, however, it is difficult to define the role of social factors as scholarship often conflates different types of social factors (e.g. “intensity”, social evaluation, etc.).

In this talk, I examine the influence of intensity in the implementation of sub-phonemic shift. I define intensity as the amount of interactions that one community has with another. I create an ET model based on a case-study of Mexican Mennonites (Plautdietsch-speakers) described in Burns (2022). In the Plautdietsch community, Spanish-Plautdietsch bilingualism is traditionally characteristic of males. Males are community liaisons to the world beyond the religious community and learn Spanish at a young age by accompanying older males. While Plautdietsch males become proficient bilinguals, Plautdietsch females remain functionally monolingual.

Northern dialects of Mexican Spanish are undergoing a sub-phonemic shift involving deaffrication of $/tʃ/ > [ʃ]$. Deaffrication is believed to have been active in the early 20th century (Brown 1989), which is around the time that Canadian Mennonites developed their settlements in Mexico (Burns 2016). While such a shift lacks structural consequences in Mexican Spanish, in Plautdietsch, it can potentially lead to the merger of $/tʃ/$ and $/ʃ/$. Burns (2022) finds that a sample of Plautdietsch-speaking males born in Mexico around 1950 deaffricate, but females born around this time do not. Younger females, who have no knowledge of Spanish, exhibit some deaffrication (Burns 2022). This suggests that deaffrication entered the community through bilingual males, despite a dearth of lexical borrowings from Spanish, and recently expanded beyond this group to other members of the community.

I created an ET model (based on Pierrehumbert 2001) with static phonetic and phonological input in order to test how varying degrees of intensity, as defined above, impact the outcomes of contact in a community where only half of the members are expected to engage in bilingual interactions as outlined above for Plautdietsch. The script that ran 10,000 production–perception loops for gradiently defined tokens of $[tʃ]$ and $[ʃ]$. Spanish-speakers categorized both as types of $/tʃ/$, whereas Plautdietsch speakers categorized tokens as either members of $/tʃ/$ or $/ʃ/$. Conversational interactions were assigned between either Plautdietsch males and Plautdietsch females, Plautdietsch males and Spanish-speakers, or any community with itself (i.e. Plautdietsch males, Plautdietsch females, Spanish). This script was run 10 times, each time with random conversational assignment. Across all trials, Plautdietsch males and females started with the same grammatical input.

The outcomes of the model suggest that even when only half of a community is bilingual, sub-phonemic shift can occur, albeit delayed in the non-bilingual group. Initially, Plautdietsch males and females diverged as Plautdietsch males converge with the Spanish community, but Plautdietsch females eventually begin to shift. The ratio of conversational pairings indicates how quickly the community shifts (i.e. if Plautdietsch males have more pairings with Spanish speakers, the shift occurs faster). While the ratio of interactions impacts how quickly subsets of a community adopt the innovation, non-bilingual participants do not preclude the diffusion of innovation as Plautdietsch females and Spanish-speakers also shift. These results indicate when sup-phonemic shift does not occur in communities with bilingualism, other social and structural factors may be necessary explanatory factors.

References

- Brown, Dolores. 1989. El habla juvenil de Sonora, México: la fonética de 32 jóvenes. *Nueva Revista de Filología Hispánica* 37(1). 43–82.
- Burns, Roslyn. 2022. Mechanisms of sub-phonemic change in prescriptive bilingualism: The case of Mexican Plautdietsch. *Glossa: a journal of general linguistics* 7(1). DOI: 10.16995/glossa.5744
- Burns, Roslyn. 2016. The Plautdietsch vowel shift across space and time. *Journal of Linguistic Geography* 3(2). 79–94.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–158. Amsterdam: John Benjamins.
- Thomason, Sarah & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Van Coetsem, Frans. 1988. *Loan Phonology and the two Types of Transfer in Language Contact*. Dordrecht: Foris Publications.

Rehabilitating ‘non-proportional’ analogy

The proportional model of analogical change attributes morphological productivity to the solution of a proportion like (a) below, involving ad-hoc generalisation of relationships between small sets of inflected words. This captures the fact that similarity begets similarity in language change: items that look or behave alike, sometimes in quite superficial ways, are liable to become even more alike. Yet many examples of morphological innovation do not appear to be based on any proportional model: contrast the Greek subjunctive *rhégnūtai* in (a) with the variant form *rhégnuētai* in (b), which seems to be formed simply by combining a stem and suffix. Such changes fit better within a rule-based approach to productivity, in which speakers perform a global analysis of linguistic data and distill it into a mental grammar capable of reconstituting surface forms as needed.

a. *phéretai* ‘it is carried’ (ind.) is to *phérētai* (subj.) as *rhégnutai* ‘it shatters’ (ind.) is to ***rhégnūtai***

b. ***rhégnu-ētai*** SHATTER-mediopassive.3sg.subj (cf. *phér-ētai* CARRY-mediopassive.3sg.subj)

Such phenomena led the Neogrammarians to concede that not all analogical processes are proportional (e.g. Osthoff 1897, Paul 1886:95). Attempts to reconcile ‘non-proportional’ and proportional analogy have since fallen into two camps: either all morphological change is treated as non-proportional within a rule-based approach (e.g. Kiparsky 1968), or all analogical change is treated as proportional (see e.g. Hill 2020, Garrett 2008, Fertig 2016). The former approach fails to capture the role of surface similarity in morphological change, while the latter relies on explaining away apparently non-proportional examples by finding either a proportional model or an alternative explanation such as dialect borrowing. Nonetheless both proportional and non-proportional analogy are still widely invoked in historical linguistics, and the theoretical conflict between them remains unresolved.

I will outline a new way of reconciling proportional and ‘non-proportional’ analogy by expanding analogical proportions to include any number of known surface forms, rather than being limited to the three forms of classical proportions (*‘a is to b as c is to x’*). Using computational procedures a local analysis of these forms is performed on the fly and used to predict an unknown inflectional form. In this framework, rules and analogies are seen as notational variants at opposite ends of a spectrum: while traditional rules are maximally general in scope, traditional analogical proportions are based on a single exemplar. This accommodates both types under a single umbrella, while retaining the insights that speakers look for similar models when producing morphological forms, and that the generalisations revealed by morphological innovations can be quite local and idiosyncratic (Joseph 2011). I will show how by integrating both types of analogy with a common formalism, they can be subjected to the same measures (e.g. of phonological/morphosyntactic similarity, type and token frequency) and subjected to a unified statistical analysis, to reveal factors that affect the probability of a putative analogical change actually taking place.

Fertig, D. (2016). Mechanisms of paradigm leveling and the role of universal preferences in morphophonological change. *Diachronica*, 33(4), 423-460.

Garrett, A. (2008). Paradigmatic uniformity and markedness. In J. Good (ed.), *Linguistic universals and language change*, 125-143. New York: OUP.

Hill, E. (2020). Analogy in inflectional change: Modification or whole-word replacement?. *Language* 96(1), 38-58.

Joseph, B. (2011). A localistic approach to universals and variation. In Peter Siemund (ed.), *Linguistic universals and language variation*, 404-425. Berlin: Mouton de Gruyter.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E. and Harms, R.T. (Eds.), *Universals in linguistic theory* (171-202). New York: Holt, Rinehart and Winston.

Osthoff, H. (1879). Kleine Beiträge zur Declinationslehre der Indogermanischen Sprachen. *Morphologische Untersuchungen*, 2, 1-147.

Paul, H. (1886). *Principien der Sprachgeschichte*. 2nd ed. Halle: Niemeyer.

Aspectual uses of *saber* + infinitive in South American Spanish varieties: a corpus-based study

Silvina Espíndola Moschner, Malte Rosemeyer (Freie Universität Berlin)

Keywords: modality, aspect, grammaticalization

In addition to its well-attested modal meaning (ability), the construction *saber* ‘to know’ + infinitive can exhibit aspectual meanings in certain American Spanish varieties (cf. Kany 1945, Di Tullio 2006, Pfänder 2009, Duque Enríquez 2021). Thus, it can be used to express present or past habituality (1) or, in the preterite form, encode completed events (2).

- (1) a. *(El perro) En la calle sabe estar todos los días, salta a las personas, molesta.*
‘(The dog) is usually in the street every day, it jumps on people, it annoys them.’
b. *A veces sabía entrar a las 9:00 de la mañana y salía a las 3:00 de la mañana.*
‘Sometimes I used to come at 9:00 in the morning and leave at 3:00 in the morning.’
- (2) *el barrio albaicín que supo ser asiento de las cortes de los monarcas ziríes en el siglo xi*
‘the Albaicín neighborhood that was once the seat of the courts of the Zirid monarchs in the 11th century’

It is not well understood to what extent the grammaticalization processes leading to (1) and (2) can be considered independent or whether they are interrelated. Likewise, some scholars have assumed the meaning change illustrated in (1) to be induced by contact with Quichua, where the verb *yachay* has both the meanings of ‘to know’ and ‘to be used to’ (cf. Vázquez 1991, Duque Enríquez 2021). This study approaches these questions by comparing four American Spanish varieties (Argentinian, Uruguayan, Bolivian and Ecuadorian Spanish), all of which possess aspectual uses of *saber*.

We conduct a quantitative analysis of $n = 6,000$ occurrences of *saber* + infinitive in the eSTenTen, a synchronic corpus of blogs and newspapers (Kilgarriff & Renau 2013). Using logistic regression analysis, we demonstrate that the functions of *saber* + infinitive can be predicted from a set of contextual properties, namely semantic features of the subject, predicate type and type of adverbial modification. While these proxies predict both habituality and completion readings, their effect is moderated by whether *saber* is inflected for imperfective or perfective aspect. Crucially, many of these examples, such as (3), can also be taken to express participant-external possibility, i.e. a possibility that does not depend on the agent, but on external circumstances.

- (3) *en la estancia la ema se saben ver (estos animales)*
‘at the farm la ema one usually sees/can see [= it is possible to see] (these animals)’

Our analysis shows that there are differences in the degree to which the grammaticalization processes leading to (1) and (2) have been implemented in the varieties under study. Thus, while in some dialects (Argentinian, Uruguayan) both habituality and completion readings are attested, in others (Ecuadorian, Bolivian) habituality readings predominate. These findings allow us to postulate the grammaticalization cline in (a), i.e. that the semantic extension of *saber* from the expression of participant-internal to participant-external modality constituted a decisive step in the development of aspectual values, and (b) that the development of habituality and completion readings then followed different grammaticalization paths. The results also suggest that the development of habitual *saber* in Ecuador and Bolivia has been facilitated by contact with Quichua.

- (a) participant-internal possibility (ability) > participant-external possibility > aspectual values

Corpus:

SkE. Sketch Engine: Corpus query system. Corpus: esTenTen18. <https://www.sketchengine.eu/estenten-spanish-corpus/>

References:

- Aijmer, K. (2004). The semantic path from modality to aspect. *Corpus approaches to grammaticalization in English*, 13, 57-78. Amsterdam/Philadelphia: Benjamins.
- Bosque, I. (2000). ¿Qué sabe el que sabe hacer algo? Saber entre los predicados modales. *Palabras: Víctor Sánchez de Zavala in memoriam*, 303-324. Universidad del País Vasco/Euskal Herriko Unibertsitatea.
- Bybee, J., Perkins, B. & Pagluica, W. (1994). *The evolution of grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Di Tullio, A. (2006). Auxiliares y operadores aspectuales en el español rioplatense. *Signo y seña*, 15, 267-285.
- Duque Enríquez, P. (2021). El ‘saber habitual’: Una gramaticalización inducida por factores internos y externos en el español andino ecuatoriano. *Spanish in Context*, 18(3), 362-386.
- Heine, B. (2002). On the role of context in grammaticalization. *Typological studies in language* 49, 83-102. Amsterdam/Philadelphia: Benjamins.
- Hengeveld, K., & Olbertz, H. (2021). Habituals in Spanish. *Linguistics in Amsterdam*, 14(1), 81-104.
- Kany, C. (1945). *Sintaxis hispanoamericana*. Madrid: Gredos.
- Narrog, H. (2012). *Modality, subjectivity, and semantic change: a cross-linguistic perspective*. Oxford University Press.
- Linell, P. (2009). *Rethinking language, mind, and world dialogically*. IAP.
- Pfänder, S. (2009). *Gramática mestiza. Con referencia al castellano de Cochabamba*. Instituto boliviano de lexicografía y otros estudios. La Paz, Bolivia.
- Vatrican, A. (2015). La modalidad en la gramática: las construcciones saber/poder + infinitivo. *Revista Española de Lingüística*, 45(2), 115-141.
- Vázquez, H. (1991). *Reparos sobre nuestro lenguaje usual*. Núcleo del Azuay de la Casa de la Cultura Ecuatoriana “Benjamín Carrión”.
- Vendler, Z. (1957). Verbs and times. *The philosophical review*, 66(2), 143-160. Duke University Press.

Maria Flaksman
LMU Munich

LOST IN TRANSLATION: ONOMATOPOEIC WORDS IN OLD ENGLISH GLOSSES

Onomatopoeic (or imitative) words are words with iconic correlation between form and meaning. Iconicity – as opposed to arbitrariness – is a relationship of resemblance (Peirce, 1940). Onomatopoeic words (in some languages also termed ‘ideophones’), thus, are rough ‘copies’ or imitations of sounds they denote (e.g., English *meow*, *bang*, *buzz*, *plop*). But while onomatopoeia enjoys increasing popularity in research on modern languages (see Akita, 2013; Antilla, 1975; Dingemanse, 2012; Hinton et al., 1994; Moreno-Cabrera, 2020, Voelz et al., 2001, etc.) it is rarely in focus of historical-comparative research (Carling et al., 2020; Georgescu, 2018; Koleva-Zlateva, 2008; Liberman, 2010). However, diachronic approach towards lexical iconicity yields significant results: exceptions from regular sound changes and otherwise inexplicable phonetic and/or semantic development are successfully explained through onomatopoeia (Campbell, 2013; Durkin, 2009; Hock, 1991; Lühr, 1988; Malkiel, 1990; Sadowski, 2001). It has also been suggested (Flaksman, 2017) that language change causes general *de-iconization* of imitative lexicons, which triggers new onomatopoeic coinage. Research on onomatopoeic words in ancient and reconstructed languages is rare (e.g., see Kozlova, 2013; Anderson, 1998) although evidence for their presence does appear in etymological dictionaries (e.g., Kroonen, 2013; Lehmann, 1986).

The *aim* of this talk is twofold: (1) to provide evidence for the existence of onomatopoeic words in Old English and (2) to discuss the specific problems related to translation of onomatopoeia.

This *comparative* research is based on the *material* of the two reference lists of onomatopoeic words from both Old English and Latin. The lists were comprised by the *method* of continuous sampling from the etymological dictionaries: de Vaan (2008) and Holthausen (1974) respectively. Words marked ‘onomatopoeic’, ‘expressive’, ‘(sound) imitative’ were selected (for example, L. *cuculus* ‘cuckoo’, *murmurare* ‘to rumble’, *crocire* ‘to croak’, *stridere* ‘to make a shrill sound’; OE *dynnan* ‘to make a noise, din’, *fneósun* ‘a sneezing’, *giellan* ‘to yell’). Words which originated as denotations of sounds were also added.

On the second stage of the research, the corpus of Old English glosses from Meritt (1945) was analysed for the presence of regular translation patterns of imitative words (e.g., *balare/bláetan* ‘to bleat’).

The talk will focus on the following *research questions*: (1) whether Latin onomatopoeic words were translated with Old English onomatopoeic words; (2) whether there were any cognate Latin/Old English pairs of words and (3) whether the Latin/Old English onomatopoeic word pairs are typologically similar (on the typology of onomatopoeic words – see Voronin, 2005).

References

- Akita, K. 2013. Constraints on the semantic extension of onomatopoeia. In *Public Journal of Semiotics*, 5(1), 21–37.
- Anderson, E. 1998. *A Grammar of Iconism*. London: Associated University Press.
- Antilla, R. 1975. Affective vocabulary in Finnish: an(other) invitation. In *Ural-Altische Jahrbucher*, 47, 10–19.
- Bosworth, J., et al. *An Anglo-Saxon Dictionary Online*. ed. by T. N. Toller, Faculty of Arts, Charles University in Prague. URL: <https://bosworthtoller.com/> (accessed 2 Jan. 2023).
- Campbell, L. 2013. *Historical Linguistics. An Introduction*. Edinburgh: Edinburgh UP.
- Carling, G., E. Johansson, N., & Frid, J. 2020. *The evolutionary mechanisms of vocal iconicity. A study on basic vocabulary of the Indo-European family*. Abstract from SLE, 2020.
- de Vaan, M. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden: Brill.
- Dingemanse, M. 2012. advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10), 654–672.
- Durkin, Ph. 2009. *The Oxford Guide to Etymology*. Oxford: Oxford University Press
- Flaksman, M. 2017. Iconic treadmill hypothesis. the reasons behind continuous onomatopoeic coinage. In M. Bauer, A. Zirker, O. Fischer & Ch. Ljungberg (Eds.), *Dimensions of Iconicity* (pp. 15–38). Amsterdam: John Benjamins.
- Georgescu, S. 2018. Las ‘creaciones expresivas’ del *Diccionario crítico etimológico castellano e hispánico en contexto románico e indoeuropeo*. Un nuevo (y a la vez antiguo) enfoque. *Revista de Filología Románica*, 35, 119–139.
- Hinton, L., Nichols, J., & Ohala J. J. (Eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Hock, H.H. 1991. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Holthausen, F. 1974. *Altenglisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.
- Koleva-Zlateva, Zh. 2008. *Slavianskaya leksika zvukosimvolocheskogo proiskhozhdeniya* [Slavonic words of phonoiconic origin]. *Studia Slavica Hungarica*, 53/2, 381–395.
- Kozlova, T. O. 2013. Iconically expressible meanings in proto-Indo-European roots and their reflexes in daughter branches. In L. Elleström, O. Fischer & Ch. Ljungberg (Eds.), *Iconic Investigations* (pp. 311–330). Amsterdam: Benjamins.
- Kroonen, G. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden-Boston: Brill.
- Lehmann, W. P. 1986. *A Gothic Etymological Dictionary*. Leiden: Brill.
- Lieberman, A. 2010. Iconicity and Etymology. In C. J. Conradie, R. J.M. Beukes, O. Fischer, & Ch. Ljungberg (Eds.), *Synergy* (pp. 243–258). Amsterdam: John Benjamins.
- Lühr, R. 1988. *Expressivität und Lautgesetz im Germanischen*. Heidelberg: Carl Winter Universitätsverlag.
- Malkiel, Ya. 1990. *Diachronic Problems in Phonosymbolism*. Amsterdam-Philadelphia: John Benjamins
- Meritt, H.D. 1945. *Old English glosses (a collection)*. Oxford: Oxford University Press.
- Moreno-Cabrera, J. C. 2020. *Iconicity in Language: An Encyclopaedic Dictionary*. Cambridge: Cambridge Scholars Publishing.
- Peirce, Ch. S. 1940. *The Philosophy of Peirce: Selected Writings* (ed. by J. Buchler). New York: Harcourt, Brace and Company.
- Sadowski, P. 2001. The sound as an echo to the sense. the iconicity of English gl- words’. In O. Fischer & M. Nänny (Eds.). *The Motivated Sign* (pp. 69–88). Amsterdam: Benjamins.
- Voeltz, E. F. K. & Kilian-Hatz, Ch. (Eds.). 2001. *Ideophones* [Typological Studies in Language 44]. Amsterdam-Philadelphia: John Benjamins.
- Voronin, S. V. 2005. *Iconicity. Glottogenesis. Semiosis: Sundry Papers*. St. Petersburg: St. Petersburg University Press.

A new perspective on the evolution of mood and negation markers in Proto-Basque

Borja Ariztimuño & Iker Salaberri
(University of the Basque Country)

Keywords: <reconstruction, morphosyntactic change, tense-aspect-mood, negation, Basque>

Recent decades have witnessed an increasing interest in historical Basque linguistics in general, and in reconstructing Proto-Basque (PB) in particular. This has led to meaningful advances in the reconstruction of PB phonology and morphology and, to a more limited extent, morphosyntax. Against that background, this paper addresses the evolution of PB markers of mood, modality and negation, which have not received much attention despite being well documented in textual records.

Early recorded stages of Basque present a complex picture concerning the expression of mood, modality and negation: some verbs are both marked with the suffix *-ke/-te* —which has various temporal, aspectual and modal uses— and accompanied by the modal auxiliary *ahal* (negative counterpart *ezin*), as in *Ayn fuerte eta galanto labradu-ric nola ecin ayn-beste mundu-an al ç-a-te-an* so strongly and beautifully build-PTCP that cannot so-much world-LOC can 3SG.PST-be-*te*-PST ‘Built so strongly and beautifully that it could not have been possible on Earth’ (Lazarraga 1567–1602, *apud* Bilbao et al. 2020: 76). By contrast, the same suffix seems to suffice as a modal marker when accompanied by the standard negator *ez*, cf. *Bat-a ez=pa-d-a nahi, ez=k-ita-ke-k gudu-ka ni eta hi* one-DEF NEG=SUB-3SG.PRS-PRS want NEG=1PL-AUX-*ke*-ALLC ‘If one of us doesn’t want to, we cannot fight each other, you and I’ (Oihenart 2003 [1657]: 256). In yet other cases the same functions are indicated only by means of *ahal/ezin*, compare *Ehor-c hura gayxteri-a-z ecin l-eça-n inbia* no.one-ERG 3SG evilness-DEF-INS cannot HYP-AUX-SUB envy ‘So that no one could envy her out of evilness’ (Etxepare 1980 [1545]: 94). All three kinds of verbal forms seem to be largely synonymous.

This state of affairs raises several questions: first of all, the aforementioned variation is suggestive of ongoing change, but it is unclear which marker, if any, originally served to indicate mood in the proto-language. Moreover, the numerous uses of the suffix *-ke/-te* —ranging from potential mood, through future reference, to doubt (conjecture) on the speaker’s part towards the truth of the proposition, among others (Lafon 1970)— provide few clues as to which use is original and what path of change subsequently ensued. According to Rebuschi (1984: 275–276) modal, conditional and conjectural uses are all derived from an initial predicative meaning, whereas Mounole (2014: 340–341), on the basis of cross-linguistically common paths of change (Bybee 1994: 265–266), argues that the shift must have been from modal to temporal. Inherent to these proposals is the view that the use of modal markers *ahal/ezin* is an innovation, i.e., they came to reinforce verbal forms with *-ke/-te* in order to avoid ambiguous modal-temporal readings.

In turn, it has been recently suggested that the suffix *-ke/-te* must have had an original negative meaning, which over time yielded temporal (future) reference (Ariztimuño & Salaberri 2022, from an earlier proposal by Ariztimuño 2016). Following this line of thought, here we put forward a three-stage path of change for the uses of *-ke/-te*:

Stage #1: *-ke/-te* is used as a marker of future tense, whereas the modal auxiliary *ahal/ezin* is the only means of indicating potential mood;

Stage #2: *-ke/-te* undergoes semantic bleaching and develops new meanings, including epistemic modality (conjecture). Consequently, verbal forms with *-ke/-te* become, in some contexts, ambiguous between a temporal and an epistemic modal meaning;

Stage #3: in order to avoid ambiguity, epistemic modal uses of *-ke/-te* are reinforced by the modal auxiliary *ahal/ezin*. As a consequence, *-ke/-te* is increasingly identified with and ultimately takes on a potential modal meaning it previously did not have.

This chain of stages (#1-3) is meant to account for two facts: (a) modal uses of *-ke/-te* seem to stem from an initial temporal meaning, judging by the fact that all other uses of this suffix (predicative, apodosis, conjecture i.e. epistemic modality, etc.) can also be derived by the same path; (b) the existence of rare potential verb forms with *ahall/ezin* but without *-ke/-te*—such as *Eta nehor-c ecin ihardets c-i-eço-yo-n hitz-ic* and no one-ERG cannot answer 3SG.PST-AUX-SBJ-3SG-PST word-PART ‘And no one could answer him a word’ (Leizarraga 1990 [1571]: 367)—in 16th-century Basque and the widespread prevalence of *ahall/ezin* suggest that the modal auxiliary must have been the original means of expressing mood (and negation) in the proto-language. Accordingly, this paper reconstructs a typologically uncommon path of change (temporal > modal) which is, however, supported by a careful interpretation of the data.

Abbreviations

1 = 1st person; 2 = 2nd person; 3 = 3rd person; ALLC = allocutive; AUX = auxiliary; DEF = definite; ERG = ergative; HYP = hypothetical; INS = instrumental; ITER = iterative; LOC = locative; NEG = negator; PART = partitive; PL = plural; PRS = present tense; PST = past tense; PTCP = participle; SBJ = subjunctive; SG = singular; SUB = subordinator.

References

- Ariztimuño, Borja. 2016. Euskal ezeztapena(k) ‘Jespersen Zikloen’ argitan [Basque negation(s) in light of Jespersen’s Cycles]. Talk held at the *Monumenta Linguae Vasconum mintegi irekiak* conference, University of the Basque Country, Vitoria-Gasteiz, 05/12/2016.
- Ariztimuño, Borja & Iker Salaberri. 2022. Reconstructing negation in a language isolate: Jespersen’s Cycle(s) and grammaticalization in Basque. Talk held at the *25th International conference in historical linguistics*, University of Oxford, Oxford, 03/08/2022.
- Bilbao, Gidor, Ricardo Gómez, Joseba A. Lakarra, Julen Manterola, Céline Mounole & Blanca Urgell. 2020. *Lazarraga eskuizkribuaren edizioa eta azterketa* [Edition and analysis of the Lazarraga manuscript], vol. 2: *Testua* [Text]. Bilbo/Bilbao: University of the Basque Country.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. London/Chicago: University of Chicago Press.
- Ettxepare, Bernard. 1980 [1545]. *Linguae vasconum primitiae* (ed. by Patxi Altuna). Bilbo/Bilbao: Euskaltzaindia, Ediciones Mensajero.
- Lafon, René. 1970. Le suffixe *-ke*, *-te* dans la conjugaison basque. *Bulletin de la Société Linguistique de Paris* 65(1). 184–212.
- Leizarraga, Joanes. 1990 [1571]. *Jesus Christ gure iaunaren testamentu berria, Othoizta ecclesiasticoen forma, Catechisme, Kalendrera, ABC edo christinoen instructionea* [New Testament of our lord Jesus Christ, Form of ecclesiastic prayers, Catechism, Calendar, ABC or christians’ teachings]. Bilbo/Bilbao: Euskaltzaindia.
- Mounole, Céline. 2014. Le verbe basque ancien: étude philologique et diachronique. *International Journal of Basque Linguistics and Philology* [ASJU] 48(1/2). 1–506.
- Oihenart, Arnaud. 2003 [1657]. *Euskal atsotitzak eta neurtitzak* [Basque proverbs and poems] (ed. by Patxi Altuna and Jose A. Mujika). Bilbo/Bilbao: Euskaltzaindia.
- Rebuschi, Georges. 1984. *Structure de l’énoncé en basque*. Paris: Société d’Études Linguistiques et Anthropologiques de France.

Verified Computational Rule-based Historical Phonology in Standard ML and Isabelle/HOL

This paper introduces an implementation of rule-based phonology in Standard ML and a formal definition and verification of the core components of such phonology in Isabelle/HOL, an interactive theorem prover. This phonology is used to automatically derive modern reflexes of Spanish, Portuguese, Chinese and Sino-Korean from their ancestral etyma using one underlying model. The architecture of this program is as such: We first implement a featureful segmental inventory, which means that each segment is not merely a character literal but data types with feature information. Then we define the constituents of syllables and syllables themselves by gluing the segments into nested lists. Once syllables are defined, phonological words come naturally, as they can be implemented as lists of syllables. Finishing the definition of phonological words means that we can represent all of the etyma and reflexes in the languages that we are investigating in this project. The rest of the program deals with the definition of the operations on those data, namely sound changes, and convenient utilities to help us define all of the sound changes happened in the history of these four languages.

A sound change in our system is represented as function mapping a phonological word to another, which matches the intuition of a working linguist. One may stop here and start implementing all of the sound changes as recursive ML functions on lists, as the phonological words are so represented in our system, but this approach is tedious, certainly may result in a lot of boilerplate, for that many sound changes in world languages only differ in some details and are structurally similar; it is also error-prone, as the reduplication of similar routines in the code base usually is. Thus instead of writing those sound changes entirely by hand, most of the sound changes in our system are created through schema that decouples the problem into manageable modular pieces. Here is the concrete explanation: just like we have roughly three tiers in the representation of a phonological word: the segments (which themselves are products of features), syllables (that have constituents like onset, nucleus, and coda), and phonological words. Our strategy is to define utilities that would rewrite one tier at a time and compose them into an actual sound change.

These two components, the component that represents data and the component that rewrites data, constitutes the trusted kernel of the program; this kernel is what we are going to verify in Isabelle/HOL. Isabelle is a member of HOL family of theorem provers. It is based on Higher-Order Logic, which a battle-tested logical system that is more than enough to verify our system to secure the desired behaviors. This verified kernel is shared among all the languages whose history we implemented; the only two language dependent parts of our system are the etyma that are to be rewritten and the respective sound changes in those languages. It should be noted that although there are four languages in our project, we only need two sets of etyma: Latin for Spanish and Portuguese, Chinese for both Chinese and Sino-Korean. Even more, Spanish and Portuguese share all the sound changes in our system until their split in Medieval times. Spanish and Portuguese are chosen precisely because of their similarities, so that we can demonstrate another feature of our system: our program is able to output and parse reflexes coupled with their history represented a list of sound changes, which enabled us to: define sound changes that are shared among Spanish and Portuguese, apply them to the etyma, and store the results so that the now separate Spanish and Portuguese modules can deal with the Romance etyma respectively.

The entirety of this program is written in Standard ML '97 and is able to be compiled both by the SML/NJ and MLton compilers. The proof scripts are written in Isabelle/HOL 2022. The source code of this project will be released under the BSD-3-Clause license.

[Wil38] Edwin B. Williams. *From Latin to Portuguese: Historical Phonology and Morphology of the Portuguese Language*. University of Pennsylvania Press, 1938.

- [Llo87] Paul M. Lloyd. *From Latin to Spanish*. American Philosophical Society, 1987.
- [Pau90] Lawrence C. Paulson. “Isabelle: The Next 700 Theorem Provers”. In: *Logic and Computer Science* (1990), pp. 361–386.
- [Pau96] Lawrence C. Paulson. *ML for the Working Programmer*. Cambridge University Press, 1996.
- [Pen02] Ralph Penny. *A History of the Spanish Language*. Cambridge University Press, 2002.
- [Har03] Steven Lee Hartman. *Phono (Version 4.0): Software for Modeling Regular Historical Sound Change*. 2003. URL: <https://langhist.weebly.com/files/theme/ver40.pdf> (visited on 10/18/2022).
- [Hua05] José Ignacio Hualde. *The Sounds of Spanish*. Cambridge University Press, 2005.
- [Sch09] Axel Schuessler. *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensa*. University of Hawai’i Press, 2009.
- [BS14] William H. Baxter and Laurent Sagart. *Old Chinese: A New Reconstruction*. Oxford University Press, 2014.

Where and How?
Request verb constructions in Ancient Greek
 Abstract

Ancient Greek shows different ways of expressing speech acts of request (cf. Dickey 1996 and 2016; Denizot 2011). One of the most common strategies consists of using a request verb followed by a structure that expresses the content of the request, in other words, what the speaker would like the addressee to perform. Various factors influence how to make a request and where to insert it, and a wide range of variation can be observed across languages. However, two elements always play a significant role in shaping the construction of the request: the relationship between speaker and addressee and the context in which the utterance is performed (i.e., communication setting, discourse structure, and topic).

This paper explores speech acts of request in Ancient Greek that indicate the usage of a request verb. Specifically, it aims at addressing the variation within the usage of request verbs in relation to where they occur (i.e., type of text and their position within the discourse) and how they are constructed (i.e., their corresponding syntactic constructions, their pragmatic functions, and the co-occurrence of discourse particles).

The main contribution of the present paper is a combination of diachronic and synchronic analyses focusing on different language stages (i.e., Classical and Postclassical Greek) and various types of texts. In particular, it integrates data from documentary papyri into the analysis of literary sources, which has typically been the focus of scholarly research. Documentary papyri provide evidence of the Greek language with a continuity of more than a millennium (4th cent. BCE – 8th cent. CE) and reflect language usages that are more oriented to everyday communication purposes instead of adhering to literary conventions and genre constraints (cf. Dickey 2011).

In addition to verbs such as αἰτέω / *aitéō* ('ask, demand'), ἐρωτάω / *erōtáō* ('ask about'), and παρακαλέω / *parakaléō* ('exhort/beg'), two verbs are particularly worth considering: δέομαι / *déomai* and ἀξιόω / *axióō*. They occur in literary sources of the Classical period with the meaning of 'to be in want or need' and 'think worthy of / esteem', respectively; later on, they are used in documentary papyri for introducing a request (cf. Di Bartolo 2021). Specifically, petitions of the Ptolemaic period are the first type of documentary texts that attest to the new meaning of these two verbs. Accordingly, δέομαι / *déomai* occurs in petitions addressed to kings (i.e., ἐντεύξεις / *enteúxeis*) and ἀξιόω / *axióō* in petitions addressed to officers (i.e., ὑπομνήματα / *hupomnēmata*) in order to express a request (cf. Baetens 2020; Di Bitonto 1967 and 1968). Later on, papyrus private letters also show the occurrence of both verbs in request speech acts.

The investigation also addresses questions related to language change, synchronic variation, and information structure, discussing different occurrences of the above-mentioned request verbs. Instances will be collected by means of secondary literature, TLG, the two papyrus databanks *DDbDP* and *Trismegistos*, and the linguistically annotated treebank corpus *PapyGreek*. The paper illustrates the changes within the range of the syntactic patterns of these verbs between the Classical and the Postclassical periods and considers the pragmatic function of discourse particles occurring with them. A comparison with analogous request constructions in other ancient Indo-European languages, such as Latin, will be drawn. Furthermore, the paper addresses the semantic shifts of δέομαι / *déomai* and ἀξιόω / *axióō* in the context of Greek petitions, explaining them in terms of Traugott's subjectification theory (cf. Traugott 1995).

References:

- Baetens, Gert. 2020. *I am wronged. Petitions and Related Documents from Ptolemaic Egypt (332 - 30 BC)*. Leuven: Trismegistos online publications.
- Di Bitonto, Anna. 1967. “Le petizioni al re”. *Aegyptus* 47. 5-57
- Di Bitonto, Anna. 1968. “Le petizioni ai funzionari nel periodo tolemaico”. *Aegyptus* 48. 53-107.
- Denizot, Camille. 2011. *Donner des ordres en grec ancien: étude linguistique des formes de l'injonction*. Mont-Saint-Aignan: Presses des l'Universités de Rouen et du Havre.
- Di Bartolo, Giuseppina. 2021. *Studien zur griechischen Syntax dokumentarischer Papyri der römischen Zeit*. (Papyrologica Coloniensia Vol. XLIV). Paderborn: Ferdinand Schöningh.
- Dickey, Eleanor. 1996. *Greek Forms of Address. From Herodotus to Lucian*. Oxford: Oxford University Press.
- Dickey, Eleanor. 2011. The Greek and Latin Languages in the Papyri. In Roger S. Bagnall (ed.), *The Oxford Handbook of Papyrology*, 149–169. New York / Oxford: Oxford University Press.
- Dickey, Eleanor. 2016. Emotional language and formulae of persuasion in Greek papyrus letters. In Ed Sanders and Matthew Johncock (eds.), *Emotion and Persuasion in Classical Antiquity.*, 237-262. Stuttgart: Franz Steiner.
- Traugott, Elisabeth C. 1995. “Subjectification in Grammaticalization”. In Dieter Stein and Susan Wright (eds.), *Subjectivity and Subjectivisation*, 31–54. Cambridge: Cambridge University Press.

Evolving rhythms:

A quantitative assessment of rhythmic alternation in the history of English

Stress-based languages such as English favor alternating rhythms made up of stressed and unstressed syllables (Selkirk 1984; Kelly & Bock 1988). At a basic level of rhythmic structure, this means that consecutive stressed syllables, i.e. ‘clashes’ (e.g. *búild bák bétter*), and consecutive unstressed syllables, i.e. ‘lapses’ (e.g. *survival of the fittest*), are dispreferred. Non-optimal patterns typically trigger prosodic repair mechanisms such as pauses or prominence adjustments. However, Breiss and Hayes (2020) have demonstrated statistically that rhythmic optimization reaches beyond phonology, showing that clashes are not only repaired with prosodic means but outright avoided through syntactic (or lexical) choices (see also Schlüter 2005; Shih et al. 2015; Anttila 2016).

The proposed study extends this line of inquiry into diachrony by asking (a) whether a global preference for rhythmically optimized patterns can be detected throughout the history of English, and, if so, (b) whether it is possible to identify lexical or morpho-syntactic developments that have helped to stabilize or even improve rhythmic well-formedness.

To this end, the study investigates rhythmicity in the history of English, measured in terms of the occurrence probabilities of clashes and lapses in word bigrams sampled from Middle, Early Modern and Modern English texts. The data for the analysis come from the Penn-Helsinki Parsed corpora of English (Kroch & Taylor 2000; Kroch, Santorini & Delfs 2004; Kroch, Santorini & Diertani 2016). Quantitative analysis is carried out with R (R Core Team 2023), using linear and generalized additive models (GAMs) (Wood 2017). Apart from time period, various other predictor variables encoding prosodic and morpho-syntactic constituency will be taken into account.

Identifying clashes and lapses in the historical texts crucially depends on the correct interpretation of (the reflexes of) unstressed inflections and monosyllabic function words. The former are generally not pronounced as syllabic in Present-Day English, but their status in Middle English is not always obvious in the individual case (e.g. ME *makede* ‘made’). Monosyllabic function words generally exhibit low stress probabilities in Present-Day English, but might have been more prominent in earlier stages of the language (e.g. ME *thou schalt haue* ‘you will have’). To account for these complications in a systematic manner, evidence from contemporaneous metrical verse (e.g. Chaucer, Lydgate, Shakespeare, Spenser) will be used to calculate probability scores for inflectional syllabicity and monosyllable stress with the help of machine learning techniques such as conditional inference trees and hierarchical clustering (Levshina 2015).

Preliminary results suggest that overall rhythmicity has not changed much since Middle English. However, it can also be shown that the diachronic process of schwa loss (Minkova 1991) must have posed a major challenge to rhythmic well-formedness, as it would have significantly increased the occurrence of clashes, had it not been offset by various structural adjustments, including analyticization (e.g. ME *Gódes sóne* vs. PDE *son of Gód*). The results are interpreted in terms of language evolution (Croft 2000; Baumann & Ritt 2017): prosodic preferences act as a selective pressure tipping the balance in favor of rhythmically more optimal syntactic (or lexical) variants.

References

- Anttila, Arto. 2016. Phonological effects on syntactic variation. *Annual Review of Linguistics* 2(1), 115–137. doi:10.1146/annurev-linguistics-011415-040845.
- Baumann, Andreas & Nikolaus Ritt. 2017. On the replicator dynamics of lexical stress: accounting for stress-pattern diversity in terms of evolutionary game theory. *Phonology* 34(3), 439–471.
- Breiss, Canaan & Bruce Hayes. 2020. Phonological markedness effects in sentence formation. *Language* 96(2), 338–370. doi:10.1353/lan.2020.0023.
- Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow, England; New York: Longman.
- Kelly, Michael H. & J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology* 14(3), 389–403. doi:10.1037/0096-1523.14.3.389.
- Kroch, Anthony, Santorini, Beatrice, & Delfs, Lauren. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony, Santorini, Beatrice, & Diertani, Ariel. 2016. The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony & Taylor, Ann. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. <http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4>.
- Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam: John Benjamins. doi:10.1075/z.195.
- Minkova, Donka. 1991. *The history of final vowels in English: the sound of muting*. Berlin, New York: Mouton de Gruyter.
- R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schlüter, Julia. 2005. *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. Berlin; New York: Mouton de Gruyter.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (eds.), *Rhythm in cognition and grammar: a Germanic perspective*. Berlin: De Gruyter Mouton, 207–233.
- Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC. doi:10.1201/9781315370279.

The History of /pf/ in New Braunfels German: Another Case of Rule Inversion?

The status of rule inversion, i.e. “reversal of the input and output of a rule and complementation of the environment” (McCarthy 1991: 194), as a mechanism of language change remains controversial. While scholars like Vennemann (1972) have argued that it is a relatively common process, others, like McCarthy (1991), have contended that rule inversion is at best very rare. Additional examples of rule inversion would help resolve this controversy. This paper therefore addresses another potential example of rule inversion, involving the history of the affricate /pf/ in New Braunfels German (NBG), a critically endangered New World variety of German. According to Eikel (1954), which is based on data collected in the 1940s and 1950s, /pf/ did not appear word-initially in NBG, meaning that words beginning with [pf] in Standard German, e.g. *Pferd* ‘horse’, *Pfeffer* ‘pepper’, and *Pfirsich* ‘peach’, began with [f] in NBG. It did, however, appear word-medially and word-finally, e.g. in *Topf* ‘pot’. In light of the numerous similarities between Standard German phonology and NBG phonology discussed by Eikel, this situation indicates that a sound change from /pf/ to /f/ (deaffrication) had taken place.

The situation soon changed: Gilbert (1972: Map 103), which is based on data collected in the 1960s, notes that his informants pronounced words like *Pferd* with an initial [pf] (e.g. 100% of his informants produced an initial [pf] in *Pferd*). In other positions within the word, Gilbert’s informants used both [pf] and [f]. This indicates that the earlier sound change had been undone, as /f/ had been affricated in word-initial position to [pf].

In the data collected by members of the Texas German Dialect Project (TGDP; www.tgdp.org) since 2001, the situation has changed again. According to Boas (2009), only 8% of his informants pronounced words like *Pferd* with an initial [pf]. This shows that the affrication process indicated by the Gilbert data has largely been undone, i.e. that /pf/ has again been deaffricated to /f/.

I argue that the best account of the NBG facts is a relatively straightforward sequence of sound changes, /pf/ > /f/ > /pf/. This account outperforms other possible analyses of the Texas German data, like the model of new dialect formation proposed by Trudgill (2004), which can account for the differences between the Eikel data and the Gilbert data, but not for the differences between the Gilbert data and the TGDP data (Boas 2009). These changes admittedly do not correspond precisely to the classical definition of rule inversion, since they take place in the same environment, and do not involve the “complementation of the environment,” as true rule inversion does, but they do involve the “reversal of the input and output of a rule.” While this particular development may therefore not be a clear-cut example of rule inversion, the reasoning here could be extended to cover other phonological phenomena in Texas German (e.g. the unexpected occurrence of front rounded vowels in some contexts), which may yet reveal such examples of rule inversion in Texas German.

References

- Boas, Hans C. 2009. *The life and death of Texas German*. Durham: Duke Univ. Press.
 Eikel, Fred. 1954. *The New Braunfels German dialect*. Dissertation, Johns Hopkins.
 Gilbert, Glenn. 1972. *The linguistic atlas of Texas German*. Austin: UT Press.
 Trudgill, Peter. 2004. *New-dialect formation*. Edinburgh: Edinburgh Univ. Press.
 Vennemann, Theo. 1972. Rule inversion. *Lingua* 29: 209-242.

**Ditransitive GIVE-construction in three Hainan Min-Chinese:
Interaction between inherited structures and contact-induced changes**

Xiuwei Zeng (The Chinese University of Hong Kong)

Ditransitive *GIVE*-constructions in Sinitic languages can be classified into two types: (1) the “canonical” [V-IO-DO] construction (i.e., “give me a book”), which is found in Mandarin, Northern Chinese and Southern Min-Chinese; and (2) [V-DO-IO] construction (i.e., “give a book me”), which is common in Southern Chinese (Hashimoto 1976). Hainan Min-Chinese is a variety of Southern Min-Chinese consisting of various dialects. It has undergone intense language contact with the indigenous Kra-Dai languages (Hlai and Be) and other Chinese varieties on the Hainan Island for more than one millennium (Liu 2006). Cao (2008) claims that Hainan Min differs from other Southern Min varieties in employing the [GIVE-DO-IO] construction, as well as using the verb /ʔio/ ‘take’ as a *GIVE* verb. Zhang (2011) further argues that this is the result from the omission of dative markers in prepositional dative constructions (i.e., “take a book (tə) me”) under the pressure of contact with [V-DO-IO]-type Cantonese.

This paper presents evidence from three Hainan Min dialects (Haikou, Qionghai, Gangmen) showing that [GIVE-IO-DO] (Southern Min) construction is strongly preferred, and that the [GIVE-DO-IO] (non-Southern Min Southern Chinese) construction is still in its infancy of development. By comparing ditransitive *GIVE*-constructions in the three Hainan Min dialects, with reference to other Southern Min varieties, early modern vernacular Southern Min texts, other Southern Chinese varieties and the indigenous Be language, I have identified the origin and the historical strata of *GIVE* verbs. In present-day Hainan Min, *bun* ‘distribute’, *ʔio* ‘take’ and *khi* ‘beg/give’ all coexist as *GIVE* verbs. I argue that Hainan Min inherited *khi* and [GIVE-IO-DO] construction from Southern Min, while the sememe {GIVE} of *bun* and *ʔio* were introduced through language contact with Hakka (Sinitic) and Be (Kra-Dai) respectively at different stages of historical development.

Previous studies have also proposed an implicational universal about ditransitive constructions in Sinitic languages: “**absence of [GIVE-IO-DO] constructions** \supset **absence of R-type GIVE verbs**” (Zhang 2011; Phua 2015; Phua and Xiang 2020). R-type *GIVE* verbs refer to *GIVE* verbs that introduce recipient argument (Li and Wu 2015; cf. Margetts and Austin 2010). It is further postulated that for a verb meaning *TAKE/HOLD/DISTRIBUTE* to be used as a ditransitive *GIVE* verb in the [GIVE-IO-DO] construction, it has to undergo DO-fronting and preposition incorporation to be a R-type *GIVE* verb (Li and Wu 2015; Xia 2017). These proposals are challenged by the Haikou and Gangmen dialects of Hainan Min, which use *bun* ‘distribute’ and *ʔio* ‘take’ as ditransitive *GIVE* verbs in the [GIVE-IO-DO] construction, as they do not allow *bun* and *ʔio* to introduce recipient argument without preposition in between. Preposition incorporation is either in progress (Haikou dialect), or has not begun (Gangmen dialect). I argue that these unusual syntactic behaviors of *bun* and *ʔio* are due to the fact that they did not undergo the development from *DISTRIBUTE/TAKE* verbs into *GIVE* verbs through DO-fronting and preposition incorporation. Instead, the polysemous patterns of *bun* (‘distribute’/ ‘give’) and *ʔio* (‘take’/ ‘give’) were copied into Hainan Min at the time of contact, despite that the corresponding morpheme of *ʔio* in Be only appears in [GIVE-DO-IO] construction. *bun* and *ʔio* replaced the native *GIVE* verb *khi* and occupied its position in the inherent [GIVE-IO-DO] constructions, making Hainan Min a rare exception to established patterns. The findings illustrate examples of polysemous pattern being transferred in contact situation regardless to the mismatch in syntactic structures, and it is mapped onto the inherent structures in the recipient language without causing change in word order.

keywords: ditransitive construction, language contact, Hainan Min-Chinese

References

- Cao, Zhiyun. ed. 2008. *Hanyu fangyan dituji* (Linguistic Atlas of Chinese Dialects). Beijing: Commercial Press.
- Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: problems of typological diversity in Sino-Tibetan. *Computational Analyses of Asian and African Languages* 3:49–65.
- Li, Xuping and Yicheng Wu. 2015. Ditransitives in three Gan dialects: valence increasing and preposition incorporation. *Language Sciences* 50:66-77.
- Liu, Xinzhong. 2006. *Hainan minyu de yuyin yanjiu* (Phonetics and phonology of the Min dialect in Hainan). Beijing: China Social Sciences Press.
- Margetts, Anna and Peter K. Austin. 2007. Three Participant Events in the Languages of the World: Towards a Crosslinguistic Typology. *Linguistics* 45 (3), 393–451.
- Phua, Chiew Pheng and Mengbing Xiang. 2020. Liancheng xinquan kejiahua de jiyu dongci he jiyulei shuangjiwu jiegou (The GIVE verbs and ditransitive give-type construction in Liancheng Xinquan Hakka dialect). *Yuyanxue Luncong* 61:50-91.
- Phua, Chiew Pheng. 2015. *Shanggu hanyu yuge jushi yanjiu* (Study in dative constructions in archaic Chinese). Beijing: Commercial Press.
- Xia, Liping. 2017. Jufa kucang liebian: cong liandongshi dao jiyulei shuangjiwu jiegou (Syntactic inventory fission: from serial verb construction to ditransitive GIVE-construction). *Bulletin of Linguistic Studies* 18:25-46.
- Zhang, Min. 2011. Hanyu fangyan shuangjiwu jiegou nanbei chayi de chengyin: leixingxue jianjiu yinfa de xinwenti (Revisiting the alignment typology of ditransitive constructions in Chinese dialects). *Bulletin of Chinese Linguistics* 4 (2), 87–270.

A Phylogenetic Study of the Cariban Family: Combining Linguistic and Archaeological Data

This study presents a preliminary linguistic phylogenetic analysis of the Cariban language family, a group of indigenous languages in Northern South America and Central Brazil. The family comprises approximately three dozen languages, most of which are spoken by small communities of a few hundred speakers (Hammarström et al., 2022). The languages are closely related in phonology and lexicon. However, except for a few shallow clades, most of the family's history is unknown and subject to many theories on its homeland and expansion routes (Meira & Franchetto, 2005; Gildea, 2012).

We follow the best practices in current computational historical linguistics (Hoffman et al., 2021; Greenhill et al., 2020; Jäger, 2019; Tresoldi et al., 2022), including initial analyses with neighbournets (Huson & Bryant, 2006) and Bayesian MCMC inference (Bouckaert et al., 2019) using different evolutionary models. We collected the data from reliable sources (Matter, 2020; de Tauste, 1680; Courtz, 2008; Largo, 2011; Ruiz Blanco 1888 [1690], von den Steinen 1892; Koehn & Koehn, 1986). It is organised into an independent, normalised, and open-access database in CLDF format (Forkel et al., 2018), carrying cognate assignments made by experts (Carvalho et al., *forth.*). We will discuss the results of our classification in the context of earlier classifications of the Cariban language family, including those by Derbyshire (1999), Meira (2006), Gildea (2012), and Meira et al. (2015), such as the statistical support of our findings for consensual and nearly consensual clades (e.g., Parukotoan, Pekodian) and for other clades proposed in the literature (e.g., the “Venezuelan” branch, Wayana-Apalai, Panare-Pemongan). Besides our maximum clade credibility (MCC) tree informed by linguistic and archaeological data, we will present our first phylogeographic inference models (Lemey et al., 2009).

This study is part of a larger initiative to analyse the linguistic history of South America. It will contribute analyses and insights into the evolution and relationships within the Cariban language family, including the location of its homeland, the date of its first expansions, and its migration movements. For general phylogenetic inference, it will contribute to the discussion on the solution and strategies for incorporating archaeological data, a necessary step when classifying language families without written records extending over multiple centuries. The classification based on lexical evidence will motivate the search for shared innovations in phonology and morphology, paving the way for the reconstruction of intermediate-level proto-languages using traditional and computer-assisted methods. It will likewise expand the foundations for research on language contact among South American native languages.

References

- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019) “BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis”. *PLoS computational biology*, 15(4), e1006650.
<https://doi.org/10.1371/journal.pcbi.1006650>
- Courtz, H. (2008). *A Carib grammar and dictionary*. Magoria books.
- de Tauste, F. (1680). *Arte y vocabulario de la lengua de los indios chaymas, cumanagotos, cores, parias, y otros diversos de la provincia de Cumana, o Nueva Andalucia*. Madrid: Bernardo de Villa.
- Derbyshire, D. (1999). “Carib”. In Dixon, R. M. W. & Aikhenvald, A. (eds.), *The Amazonian Languages*, 23-64. Cambridge: Cambridge University Press.
- Carvalho, F. O., Ferraz Gerardi, F., Aragon, C. & Tresoldi, T. (forth.). *Cariban Lexical Database*.
- Forkel, R., List, J.-M., Greenhill, S. J., Bank, S., Rzymiski, C., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A. & Gray, R. D. (2018). “Cross-linguistic Data Formats, advancing data sharing and reuse in comparative linguistics”. *Scientific Data*, 5:180205.
<https://doi.org/10.1038/sdata.2018.205>
- Gildea, S. (2012). “Linguistic studies in the Cariban family”. In Campbell, L. & Grondona, V. (eds.), *The Indigenous Languages of South America: A Comprehensive Guide*, 441-494. Berlin, Boston: Berlin: Mouton.
- Hammarström, H., Forkel, R., Haspelmath, Martin & Bank, S. (2022). *Glottolog 4.7*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
<https://doi.org/10.5281/zenodo.7398962> . Available online at <http://glottolog.org>, accessed on January 17th, 2023.
- Hoffmann, K., Bouckaert, R., Greenhill, S. J. & Kühnert, D. (2021). “Bayesian phylogenetic analysis of linguistic data using BEAST”. *Journal of Language Evolution*, 6: 119–135.
<https://doi.org/10.1093/jole/lzab005>
- Huson, D. H. & Bryant, D. (2006). “Application of Phylogenetic Networks in Evolutionary Studies”. *Molecular Biology and Evolution*, 23(2): 254-267.
<https://doi.org/10.1093/molbev/msj030>
- Greenhill, S. J., Heggarty, P., & Gray, R. D. (2020). “Bayesian Phylolinguistics”. In Janda, R. D., Joseph, B. D. & Vance, B. S. (eds) *The Handbook of Historical Linguistics*, Volume II, pp. 226–253. Wiley-Blackwell: New Jersey.

- Jäger, G. (2019). “Computational historical linguistics”. *Theoretical Linguistics*, 45(3-4): 151-182. <https://doi.org/10.1515/tl-2019-0011>
- Koehn, E., & Koehn, S. (1986). “Apalai”. In D. S. Derbyshire & G. K. V. I. Pullum (Eds.), *Handbook of Amazonian languages*, Vol. 17: 33-127. Berlin: de Gruyter.
- Largo, W. (2011). *Una Gramática del Yukpa (Colombia)*. Privately published.
- Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. (2009). “Bayesian Phylogeography Finds Its Roots”. *PLoS Computational Biology* 5, e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>
- Matter, F. (2020). *Comparative Cariban Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <https://cariban.clld.org/>, accessed on January 17th, 2023.
- Meira, S. (2006). “Cariban Languages”. In Brown, K. (ed.), *Encyclopedia of Language and Linguistics*, 199-203. 2nd edn. Amsterdam: Elsevier.
- Meira, S. and B. Franchetto (2005). The Southern Cariban languages and the Cariban family. *International Journal of American Linguistics* 71: 127-192.
- Meira, S., Birchall, J., Chousou-Polydouri, N. (2015). “A character-based internal classification of the Cariban family”. Talk. *48th Annual Meeting of the Societas Linguisticae Europaea*. Leiden, Netherlands. Available online at https://www.academia.edu/15980095/A_character-based_internal_classification_of_the_Cariban_language_family, accessed on January 17th, 2023.
- Ruiz Blanco, M. (1888 [1690]). *Arte y tesoro de la lengua cumanagota (Vol. 3)*. Leipzig, Germany: B.G. Teubner.
- von den Steinen, K. (1892). *Die Bakairi-Sprache: Wörterverzeichnis, Sätze, Sagen, Grammatik mit beiträgen zu einer Lautlehre der Karaibischen Grundsprache*. Leipzig: K. F. Koehler's Antiquarium.
- Tresoldi, T., Rzymiski, C., Forkel, R., Greenhill, S. J., List, J.-M. & Gray R. (2022). “Managing historical linguistic data for computational phylogenetics and computer-assisted language comparison”. In Berez-Kroeker, A. L., McDonnell, B., Koller, E. & Collister, L. B. (eds). *Open Handbook of Linguistic Data Management*. MIT Press. <https://doi.org/10.7551/mitpress/12200.003.0033>

Evidence for a Chibcha-Jê connection

The genealogical composition of the South American continent is a major puzzle of historical linguistics. This talk discusses lexical and grammatical evidence for a genealogical relationship between Chibchan languages of Central America and northern South America and Macro-Jê languages, predominantly spoken in the Brazilian lowlands, south of the Amazon. The correspondences include seventeen grammatical morphemes and twenty-four lexical items (Pache 2023), such as corresponding suppletive forms – Proto-Chibchan *ⁿdaʔ ~ *taʔ ‘to go₁’, *^mbã ‘to go₂’, Proto-Northern Jê *tẽ ‘to go₁’, *mõ ‘to go₂’ – and (near-)homophones such as Proto-Chibchan *siʔ ‘thorn, tooth (incisor)’, *sih ‘meat, flesh’, Proto-Northern Jê *-jĩ ‘thorn’, *-jĩ ‘meat’ (Pache 2018, Nikulin 2020). Among the further evidence is an unusual parallel grammaticalization path in Chibchan and Macro-Jê languages, which may reflect variation that existed in the shared ancestor language (see Joseph 2012). For instance, Boruca, a Chibchan language of southeastern Costa Rica, has an inessive postposition *kabá* (Quesada Pacheco 2019: 103); its cognate counterpart in Rama (eastern Nicaragua) encodes benefactive and purposive meanings (Craig 1989: 207), as shown in (1).

Rama (Chibchan)

- (1) *jaŋ-kama*
 what-PURPOSE
 ‘why’ (Craig 1989: 206)

Both Boruca inessive *kabá* and Rama purposive *kama* derive from a Proto-Chibchan form *ka^mba by regular sound change (see Pache 2018). The reflex of its Proto-Northern Jê counterpart, locative *kãm ~ *kãm (Nikulin 2020: 511) likewise indicates the inessive in Apinajé (Northern Jê, central Brazil) (Oliveira 2005: 147), and, in a particular construction illustrated in (2), purpose.

Apinajé (Northern Jê)

- (2) *ja kamã*
 this PURPOSE
 ‘for that, for this reason, that’s why’ (Oliveira 2005: 147)

Together, the correspondences in question suggest a genealogical connection between Chibchan and Macro-Jê languages. This also has implications for our understanding of migration and the populating of the Americas.

References

- CRAIG, COLETTE G. 1989. A Grammar of Rama. Université de Lyon. Report to National Science Foundation, BNS 8511156. Unpublished manuscript.
- JOSEPH, BRIAN D. 2012. A variationist solution to apparent copying across related languages. Copies versus Cognates in Bound Morphology, ed. Lars Johanson and Martine Robbeets, pp. 151–64. Studies in Language, Cognition and Culture, 2. Leiden: Brill.
- NIKULIN, ANDREY. 2020. Proto-Macro-Jê: um estudo reconstrutivo. Ph.D. dissertation, Universidade de Brasília.
- OLIVEIRA, CHRISTIANE CUNHA DE. 2005. The Language of the Apinajé People of Central Brazil. Ph.D. dissertation, University of Oregon.
- PACHE, MATTHIAS. 2018. Contributions to Chibchan Historical Linguistics. Ph.D. dissertation, Universiteit Leiden.
- _____. 2023. Evidence for a Chibcha-Jê connection. International Journal of American Linguistics 89 (2).
- QUESADA PACHECO, MIGUEL ÁNGEL. 2008. Gramática de la lengua guaymí (ngäbe). Languages of the World/Materials, 474. Munich: LINCOM Europa.

Epistemic modality out of ‘playfulness’: Modern Greek *pezi*

The emergence of epistemic constructions is a well-known and much-discussed issue, with various claims regarding the possible sources, the directionality of changes and the morphosyntactic properties of such modals and constructions (cf. e.g. Traugott & Dasher, 2002, Narrog, 2012, Hilpert, Cappelle & Depraetere 2021, among many others). The article aims to contribute to this trend of research by presenting the emergence of a novel epistemic construction in Modern Greek, which had escaped notice till now and is based -rather surprisingly- on the verb *pezo* (=play), as illustrated below:

- (1) O Mike Flanagan ... ***pezi na ine*** to mono atomo pu bori.
 The Mike Flanagan ... plays that be the only person that can
 “Mike Flanagan... may be the only person that is able to”.
 (*watchandchill.gr*, retrieved 19/1/23)

Based on an examination of available corpora of Modern Greek (TextCorpora, Corpus of Modern Greek), as well as an online search, the article shows that:

- a) The epistemic construction ‘*pezi na*’ emerged recently (last 10-15 years), most probably in spoken varieties of Modern Greek, and has gained considerable token frequency in everyday registers
- b) The morphosyntactic properties of the construction follow closely well-known typological (and Modern Greek) trends, for instance the verb is attested only in 3rd person singular, it exhibits only two tense forms (past / non-past) etc.

As far as the diachronic origin of the construction is concerned, the article argues for a multiple-source causation, illustrating the inter-relationship between the epistemic ‘*pezi na*’ and specific constructions involving the very same verb with different but related semantics (‘*dhen pezete*’ = you cannot beat this! / ‘*pezete*’= something is at stake / ‘*pezete*’= something is not finalized or is not certain yet) that were already at place.

Drawing on the observations mentioned above, the article:

- a) Broadens our perspective on the possible sources of epistemic constructions cross-linguistically
- b) Discusses the complex inter-relationship between constructions and its possible outcomes, drawing on recent Construction Grammar insights (cf. e.g. Sommerer & Smirnova, 2020)
- c) Attempts to draw on its usage-based approach to answer the elusive actuation question: Why ‘*pezi na*’ develops into an epistemic construction in 21st c. Greece?

References

- Hilpert M., B. Cappelle & I. Depraetere (eds.), 2021. *Modality and Diachronic Construction Grammar*. Amsterdam: Benjamins.
- Narrog, H. 2012. *Modality, Subjectivity, and Semantic Change. A cross-linguistic perspective*. Oxford: Oxford University Press.
- Sommerer L. & E. Smirnova (eds.), 2020. *Nodes and Networks in Diachronic Construction Grammar*. Amsterdam: Benjamins.
- Traugott, E. & R. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Explaining the speed of lexical change in historical Dutch

Why are different words replaced by new synonyms at different rates? In some domains, new variants replace each other rapidly (e.g. ‘awesome’ or ‘lit’ to refer to something COOL), while other domains display more stability (e.g. ear). Research has shown that the speed of lexical change is influenced by word-related features, like frequency, word class, length or age of acquisition (Bochkarev, Solovyev & Wichmann 2014; Monaghan 2014; Pagel, Atkinson & Meade 2007; Wichmann & Holman 2013). In this paper, we analyze whether characteristics of concepts play a role as well. Taking our lead from Franco et al. (2019) who showed that concept characteristics such as familiarity, vagueness and affect-sensitivity influence the amount of synchronic lexical variation in the base dialects of Dutch, we test whether these characteristics affect the speed of diachronic change in Dutch as well.

The data we use come from the ‘Middelnerlands Woordenboek’ (Middle Dutch Dictionary: 1250-1550) and the ‘Woordenboek der Nederlandsche Taal’ (Dictionary of the Dutch Language: 1500-1976), two large dictionaries of historical Dutch. We extract data from the digitized versions of these dictionaries with the DiaMaNT tool (Depuydt & de Does 2018), a semantic historical computational lexicon for Dutch, zooming in on 279 concepts from two semantic fields: body parts and clothing terms. In particular, for each body part or clothing concept we record all the variants that are available as (historical) synonyms to express the concept, as well as the times at which they were used by relying on the citations available in the dictionary. For example, for the body part JAW, we record that it occurs with 5 variants between 1500 and 1550, including ‘kaak’ (the current Standard Dutch lexeme), ‘kinnebak’, ‘pellorijn’ and ‘kieuw’.

Next, we divide the dataset into 50-year periods. For each period, we calculate two types of information: (1) the number of variants in use at each time point, and (2) the proportion of variants in use at a given period that were also used during the previous period. Using this information, we can answer two research questions: (1) is the number of synonyms for the concept diachronically stable, or are there fluctuations (diachronic stability)?; (2) how quickly do variants disappear from the data, how quickly are they replaced with new synonyms (the rate of lexical replacement)?

Our hypotheses are that the factors that play a role in synchronic data, affect diachronic change as well, viz. familiarity, vagueness and affect-sensitivity. Moreover, we may find differences between the body part concepts and the clothing concepts as the former concepts have a higher degree of universality and may therefore only rarely be referred to with novel lexical items.

Preliminary results on the body part concepts indicate that there are some trends in the data that confirm the correlation between familiarity and diachronic stability on the one hand, and affect-sensitivity and diachronic stability on the other. For vagueness, the picture is less clear. Further data collection and analyses will take place in the coming months.

References

- Bochkarev, V., V. Solovyev & S. Wichmann. 2014. Universals versus historical contingencies in lexical evolution. *Journal of The Royal Society Interface* 11(101). 20140841. <https://doi.org/10.1098/rsif.2014.0841>.
- Depuydt, Katrien & Jesse de Does. 2018. The Diachronic Semantic Lexicon of Dutch as Linked Open Data. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France*.
- Franco, Karlien, Dirk Geeraerts, Dirk Speelman & Roeland Van Hout. 2019. Concept characteristics and variation in lexical diversity in two Dutch dialect areas. *Cognitive Linguistics*. De Gruyter Mouton 30(1). 205–242. <https://doi.org/10.1515/cog-2017-0136>.

- Monaghan, Padraic. 2014. Age of acquisition predicts rate of lexical evolution. *Cognition* 133(3). 530–534. <https://doi.org/10.1016/j.cognition.2014.08.007>.
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449(7163). 717–720. <https://doi.org/10.1038/nature06176>.
- Wichmann, Søren & Eric W. Holman. 2013. Languages with longer words have more lexical change. In *Languages with longer words have more lexical change*, 249–282. De Gruyter Mouton. <https://doi.org/10.1515/9783110305258.249>.

The diachronic study of Bangla case marking system

Syed Shahrier Rahman ^{PhD}

Professor, Department of Linguistics, University of Dhaka

ss.rahman@du.ac.bd

Mithun Banerjee

Associate Professor, Department of Linguistics, University of Dhaka

mithun.b@du.ac.bd

Abstract

Bangla case marking is one of the least focused areas of grammatical exploration. Since this language allows an extensive set of word order variations, case marking elements of Bangla grammar are used to license their distributions in the syntactic domain. However, a few case markers have different functionality than others create complexity and ambiguity when attempting to identify case features. It raises an important query: Is there any historical data that we have access to that might explain why the Bangla case markers correlate one-to-many correspondences? From a diachronic point of view, this article examines the missing linkages among the case marking components of the Bangla language and attempts to characterise their historical evolution.

The eight-case system of Old Indo-Aryan (OIA) has been mostly lost throughout the centuries by the Indo-Aryan languages (Butt & Ahmed, 2011). Direct (the result of the merger of nominative, accusative and dative) and Oblique (mostly deriving from the Old Indo-Aryan genitive) are the only two cases that have survived however, in some languages, including Bangla, isolated traces of some other oblique cases, such as instrumental, locative, or ablative, can still be found, sometimes even within the declension paradigm (Beekes, 1995). Those bound or free morphemes of varied provenance, including many that exist in this usage already in OIA, take over for the lost cases and replenish the inventory of case markers. It's worth noting that, towards the conclusion of the Middle Indo-Aryan (MIA) era, some (oblique) case forms may become indistinguishable from bare stems because of the degradation of the nominal inflexion. Old Bangla possessed an ergative construction in the perfect aspect (Chatterji, 1970, pp. 947-8), similar to the MIA ergative clause. In contrast, modern Bangla, however, has lost this pattern by allowing the same subject case-marking for its non-perfect and perfect counterparts. New postpositions, often added to the oblique case form, have seen a dramatic growth in usage and grammaticalisation in Bangla.

- a) ami toma=ke golpo-ti bolte chai
 I.NOM you=ACC story-the.DET.DAT tell.INF want.PRS
 "I want to tell you the story."
- b) ama=r toma=ke golpo-ti bolte hobe
 I.GEN you=ACC story-the.DET.DAT tell.INF have.FUT
 "I will have to tell you the story."
- c) ami toma=r shonge golpo-ti niye alap korlam
 I.NOM you.GEN with.PostP story-the.DET.DAT about.PostP discuss.VN do.PST
 "I discussed the story with you."

Because of this grammaticalisation, a new case can emerge from the combination of a postposition and the nominal stem or oblique case (Blake, 2001). A bulk of markers of genitive and dative containing *k-* and/or *r-*, can be found in Bangla. A detailed chronological analysis of the major turning points in the corresponding grammaticalisation scenarios, therefore, is necessary to understand the facts behind the case marking system of Bangla language.

Keywords: Case, Bangla language, Case marking elements, Grammaticalisation

References

- Beekes, R. S. P. (1995). *Comparative Indo-European Linguistics. An Introduction*, Benjamins.
 Blake, B. J. (2001). *Case*. Cambridge University Press.

- Butt, M. & Ahment, T. (2011). The redevelopment of Indo-Aryan case systems from a lexical semantic perspective, *Morphology* ; 21, 3-4, 545-572 <https://dx.doi.org/10.1007/s11525-010-9175-0>
- Chatterji, S. K. (1970[originally published in 1926]). *The Origin and Development of the Bengali Language*, George Allen and Unwin Ltd.

===

German V2-Argument Clauses from a Diachronic Perspective

This study investigates argument realizing verb-second clauses (arg.V2) in the Early New High German period (ENHG) on the basis of a corpus of narrative texts from the 15th and 16th c. (Pontus und Sidonia (1450), Melusine (1456), Wigalois (1472), Wilhelm von Österreich (1481), Tristrant und Isalde (1484), Huce Scheppel (1500), Fortunatus (1509), Schöne Magelone (1527), and Goldener Esel (1538)). I will argue that arg.V2 are licensed by two different pragmatical factors independently: *at-issueness* (Simons et al. 2010) and *mediated assertivity* (Reis 1997).

In ENHG (as well as in Present Day German (PDG)), argument clauses of verbs that denote an act of assertion can either be realized as asyndetic V2 clauses (a.) or as syndetic verb-end clauses (VE) (b.), b. being the canonical structure of subordinate clauses.

- a. sy sagten all [sy **wißten** es nit] (Fortunatus)
 they said all they KNEW it NEG
 ‘They all said they didn’t know it’
- b. der schray (...) [das Fortunatus nit umb die ding **wißt**] (Fortunatus)
 he screamed (...) that Fortunatus NEG about the things KNEW
 ‘He screamed (...) that Fortunatus didn’t know about these things’

The V2-structure is assumed to be a main clause phenomenon in German and cross-Germanic (Holmberg 2015), hence arg.V2 are often argued to come close to main clauses (Reis 1997, Gärtner 2002, Truckenbrodt 2006). Although the PDG counterparts of arg.V2 have been thoroughly investigated, their licensing conditions are still a matter of debate (Jacobs 2020, Djärv 2022). There are at least two theoretical approaches to the pragmatical licensing conditions of arg.V2 in German. Firstly, arg.V2 are claimed to be *mediated assertions* (Reis 1997, Gärtner 2002), as the truth value of the proposition is usually asserted by a matrix subject. Secondly, arg.V2 are assumed to mark *at-issue*-content (in the sense of Simons et al. 2010), that is, assertions that are relevant for the *Question under Discussion* (in the sense of Klein & von Steutterheim 1992) of a discourse (Antomo 2015).

The diachronic perspective on the licensing conditions of arg.V2 is often unconsidered. Since the formal distinction between dependent and independent clauses had already developed in the Old High German period (Axel 2007), former studies have primarily focused on arg.V2 in the Old and Middle High German period. Petrova (2020) has shown that there are noticeable parallels in the typology of matrix verbs of arg.V2 between these periods and PDG. However, diachronic frequency changes indicate that there is a stronger association between the discourse pragmatical status of a clause and the V2-structure in ENHG: In our corpus of ENHG narratives, argument clauses of verbs of saying have V2 in 48,8 % of cases, whereas narratives from the 18th and 19th c. (Deutsches Textarchiv) show arg.V2 in 34,78 % of cases.

In fact, the corpus data provide evidence that the different structures of argument clauses followed rather strict discourse-pragmatical principles after verbs of saying in ENHG – this is especially prominent in the *Fortunatus*, which is the only non-translated text in the corpus. I will show that VE is very much restricted to propositions that have a discourse antecedent and that are therefore presupposed (which is the case in b.). The function of V2 is twofold: On the one hand, V2 is accessible for *mediated assertions*, whereby it is not relevant if the narrator is committed towards the truth of the proposition or not (for example, in a., the proposition of the argument clause can immediately be identified as a lie). On the other hand, V2 marks argument clauses with *at-issue*-content that independently adds to the discourse. The proposition may even be presupposed in such contexts. These two factors can, but do not necessarily have to interact. The pragmatical two-sidedness suggests that the V2-structure can hardly be ascribed to one particular pragmatic function and the existing theories do not necessarily oppose one another.

References

- Antomo, M. 2015. Abhängige Sätze in einem fragebasierten Diskursmodell. Göttingen: Georg-August-Universität Göttingen dissertation.
- Axel, K. 2007. Studies on Old High German Syntax - Left sentence periphery, verb placement and verb-second. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Djärv, K. 2022. On the interpretation and distribution of embedded main clause syntax: new perspectives on complex discourse moves. In *Glossa: a journal of general linguistics* 7(1). doi: <https://doi.org/10.16995/glossa.5752>
- Freywald, U. 2013. Uneingeleiteter V1-und V2-Satz. In J. Meibauer, M. Steinbach & H. Altmann (eds.), *Satztypen des Deutschen*, 317-337. Berlin: De Gruyter.
- Gärtner, H.-M. 2002. On the force of V2 declaratives. In *Theoretical Linguistics* 28, 33-42.
- Holmberg, A. 2015. Verb Second. In T. Kiss & A. Alexiadou (eds.), *Syntax – Theory and Analysis: An International Handbook of Contemporary Syntactic Research*, 342-383. Berlin, New York: Mouton de Gruyter.
- Jacobs, J. 2020. What kind of main clause phenomenon is V2 in German? In H. Lohnstein and A. Tsiknakis (eds.), *Verb Second*, 169-206. Boston/Berlin: De Gruyter.
- Klein, W. & C. von Steutterheim. 1992. Textstruktur und referentielle Bewegung. In *Zeitschrift für Literaturwissenschaft und Linguistik* 86, 67-92.
- Petrova, S. 2020. Variation and change in the licensing of dependent V2 in German. In H. Lohnstein & A. Tsiknakis (eds.), *Verb Second*, 251–276. Boston/Berlin: De Gruyter.
- Reis, M. 1997. Zum syntaktischen Status unselbstständiger Verbzweit-Sätze. In C. Dürscheid, K.-H. Ramers und M. Schwarz (eds.), *Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag*, 121-144. Tübingen: Niemeyer.
- Simons, M., D. Beaver, J. Tonhauser & C. Roberts. 2010. What projects and why. *Proceedings of SALT 20*, 309–327.
- Truckenbrodt, H. 2006. On the semantic motivation of syntactic verb movement to C in German. In *Theoretical Linguistics* 32, 257–306.

Semantic factors influencing the change in position of German adnominal genitives in the 17th to 19th centuries

Lisa Dücker (CAU Kiel)

In contemporary German, adnominal genitives usually follow the noun they modify: *das Haus der Frau* (the house the-GEN woman-GEN ‘the woman’s house’). This is especially true for common nouns; proper nouns have at the same time retained the ability of appearing before the head noun: *Marias Haus* (Maria-GEN house ‘Maria’s house’) but may also appear postnominally. This current word order in noun phrases with adnominal genitives is the result of a diachronic change that has been going on for several hundred years. In Old High German, genitive attributes usually appeared in the prenominal position but from the later Old High German period onwards, more and more adnominal genitives have changed to the postnominal position, starting with abstract common nouns. Since the 18th century, it has been mostly proper nouns that still appear prenominally (c.f. Demske 2001).

This syntactic change is well-studied and a number of morphological, syntactic, semantic and pragmatic factors have been found to influence this phenomenon (for an overview, see e.g. Ackermann 2021). With the exception of Ackermann’s (2021) study on onymic genitives, those factors have usually been studied in isolation, although it has, of course, been noted that properties such as the animacy of the attribute and the semantic relation between the attribute and the noun it modifies (e.g. possessive, partitive) interact with one another. With regards to the semantic relation between the nouns, it has additionally been proposed that the two word orders can correlate with distinct readings, especially with regards to subjective and objective genitives, meaning that *Lenas Entdeckung* (Lena-GEN discovery) will likely be read as a discovery made by Lena, while *die Entdeckung Lenas* (the discovery Lena-GEN) suggests the reading that Lena is being discovered (c.f. Eisenberg & Smith 2002).

In my talk I will analyse noun phrases containing adnominal genitives from three corpora: the Deutsches Textarchiv corpus, which contains written texts from the 17th to 19th centuries, the RIDGES corpus consisting of texts about herbology from the 15th to 19th centuries, and the SiGS corpus, which is made up of handwritten protocols of witch trials from the 16th and 17th centuries. I investigate three semantic properties of noun phrases containing a genitive attribute, namely the animacy of the attribute, the semantic relation between attribute and the head noun, and whether the adnominal genitive is a proper noun or not. These three factors will be analysed in isolation and in interaction as independent variables in a *random forest* model (Tagliamonte & Baayen 2012).

This analysis will show which adnominal genitives are the last to change to the postnominal position at a time when most genitives have already changed to the postnominal position. It will also show the importance of differences in text genre (c.f. Peter 2015). I propose that proper nouns and highly animate nouns (i.e. nouns denoting humans, superhumans, animals and plants), as well as nouns found to be in a subjective, objective, possessive or auctorial relation with the noun they modify will be found to occur in the prenominal position much later than common nouns in general. It is especially inanimate attributes and those that occur in a different kind of semantic relation with the head of the phrase that will appear mostly postnominally early on.

References

- Ackermann, Tanja (2021): Pre- and postnominal onymic genitives in (Early) New High German. *Journal of Historical Linguistics* 11 (3), 499–533.
- Demske, Ulrike (2001): *Merkmale und Relationen. Diachrone Studien zur Nominalphrase des Deutschen*. Berlin: De Gruyter.

Eisenberg, Peter, and George Smith (2002): Der einfache Genitiv. Eigennamen als Attribute. In Corinna Peschel (ed.): *Grammatik und Grammatikvermittlung*. Frankfurt a.M.: Peter Lang, 113-126.

Peter, Klaus (2015): Zur Semantik pränominaler onymischer Genitivattribute. *Zeitschrift für germanistische Linguistik* 43 (2), 199–232.

Tagliamonte, Sali A. & R. H. Baayen (2012): Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24 (2), 135–178.

Existential HAVE in Late Latin: insights on its diachrony in the passage to Romance

This paper investigates the rise of existential HAVE in the transition from Latin to Romance and its paths of development. Existential HAVE constructions in Romance are usually viewed in the literature as stemming from the possessive scheme associated with the transitive verb of possession HAVE (Gaeta 2013; Ciconte 2015: 231; Cruschina 2015: 58, among others).

Early and Late Latin data, however, provide evidence for a different origin of these constructions. More specifically, it is shown that existential HAVE in Romance continues and further develops patterns which became available in Late Latin to ‘introduce a new entity or situation into the world of discourse’ within a spatio-temporal frame, starting from:

- (i) *the stative-locative meaning* of the verb HAVE (‘being in a (physical/abstract) state/place, location’) (... *ille geminus, qui Syracusis habet* ... (Plt. *Men.* 68-69) that.NOM twin.NOM who Syracuse.loc has ‘...that (other) twin, who is in Syracuse ...’),
- (ii) its *non-lexical uses*, witnessed by existential-like ‘impersonal’ constructions consisting of an adverb+the 3rd singular of the verb alternating with the verb BE, *esse* (e.g., *bene habet/est, recte habet/est* ‘it is good’, attested in Early (e.g., Plautus) and Classical authors (e.g., Cicero),
- (iii) its *functional equivalence with the copula esse* ‘be’, occurring in Early and Classical Latin (Baldi & Nuti 2010: 273, note 34; 278, 376; Pinkster 2015: 97; Ciconte 2015), attested also in equative clauses in Late Latin (*ubi omnia aequalia habent* (Orib. Syn. VII, 49, 10) where all.N.PL alike.N.PL have.PRS.IND.3PL ‘Where all these are alike’ (Luque Moreno 1998: 140)

Rare examples of existential-like HAVE are reported for Early Latin (Cato, III-II BC) (2), with the verb in the active impersonal form and the nominal in preverbal position in the accusative case (Baldi & Nuti 2010: 275): ... *nisi calicem pertusum cauum habeat* unless cup hole.ACC hollow have.SBJV.PRS.3S ‘... except that **there is a bowl with a pierced hole**’ (Cato, agr. 80,1).

Existential(-like) constructions with HAVE are well attested in Late Latin, in 4th-6th c. texts (e.g., *Itinerarium Egeriae, Mulomedicina Chironis, Palladius, Oribasius, Anthimus*), occurring with [–human], most typically inanimate, indefinite/non-specific pivots (exs. (1) – (2)) and are found also in spatial (*inde ad sanctam Teclam habebat de civitate forsitan mille quingentos passus* from-there to saint Tecla have.IMP.3SG from city one thousand five-hundred.M.PL.ACC steps.M.PL.ACC ‘From that place to the mountain of God it was perhaps four miles’ (*Itin. Eger.* 23,2) and temporal constructions (*Pater eius ... ex quo hinc profectus est habet annos XIII* father.NOM his.GEN from which.ABLfrom-here leave.PST.PTCP.M.SG.NOM be.PRS.IND.3SG **have.PRS.IND.3SG years.ACC** 14 ‘It has been fourteen years since his father left (from) here’ (Hist. Apoll. RA 31) (Svennung 1935: 475-477, 572-573; Leumann, Hofmann & Szantyr 1965: §221, c; Cennamo 2011: 177-179; Pinkster 2015: 97; Panayotakis 2016 and further examples and further references therein).

HAVE is in the third person singular ‘impersonal’ active (1) and passive (2) forms, with variability of the construction involving three syntactic domains:

- (i) **word order: post and pre-copular NP/pivot** (1a) *In Hebraeo ... non habet hunc numerum* (Hier. Ezech. 11. 297B) in Hebrew not have.PRS.IND.3SG this number.ACC ‘In Hebrew this number does not exist’ vs (1b)... *ibi ... altarium ... habet* (*Itin. Eger.* 4,4) there ... altar.ACC have.PRS.IND.3SG ‘... there ... there is an altar’; (2a) *Ibi habetur capella* (Pard. 369, y. 673) there have.MP. PRS.IND.3SG goat.NOM ‘There is a goat there’ vs (2b) ... *non alter habetur* (Comm. Ap. 374) not other.NOM have.MP.PRS.IND.3SG ‘there is nobody else’
- (ii) **±agr of the copula habere with the post-/pre-copular NP/pivot**: [+AGR]: (2a) *habetur capella*, (2b) *alter habetur*; [-AGR]: (2c): *habetur ... reliquias* (Diehl, ILCV 2013) have.MP.PRS.IND.3SG remnants ‘... there are remnants of ...’; (1c) ... *unde ergo habet zizania* (Vulg. *Matth.* 13, 26-7; Ciconte 2015: 231) whence therefore have.PRS.IND.3SG darnel.weeds.ACC ‘From where, therefore, are the darnel weeds?’

- (iii) **case-marking of the pre/post-copular NP/pivot: NOM vs ACC: (2a) *habetur capella* vs (2d) *habetur ... tumulum*** (Greg. Tur. glor. conf. 35) have. PRS.IND.3SG.MP tomb.ACC (Mikilová 2016: 158)

It will be shown that the data investigated point to the presence of ‘impersonal’ HAVE and lack of pivot agreement (in number /case) as the overt markers of a change in progress, leading to the subsequent reanalysis of the locative argument as a non-referential, unspecified argument, the abstract spatio-temporal argument of predication (as witnessed in early Italo-Romance), and ultimately to the new Romance existential construction (step III on the scheme in (3)), where the theme argument (y), the pivot, ‘takes the predicating function and is thus predicated of the unspecified (spatio-temporal) argument’ (x) (Bentley 2015: 152 and references therein)

Considering the diachronic steps in the changes in the logical structure of HAVE from verb of location/locative copula to an existential copula, illustrated in (3) (adapted from Bentley 2015: 151-152, Ciconte 2015: 231)

(3) I. Be/Have-Loc’ (location, theme) > II. Be/Have-Loc’ (x, theme) > III Be’/Have’ (x,y/pivot)

the Late Latin data investigated appear to witness an intermediate stage between steps I and II. *Habere* still occurs in its locative and copular functions. It does not consistently occur in the default third person singular, preceded/followed by a non-agreeing nominal, the theme, that is still an argument in the construction, and the locative phrase/adverb in the pattern is fully referential (unlike at stage II, where it has lost its locative meaning, as witnessed in some early Northern Italian texts) (Parry 2013; Bentley 2015: 152; Ciconte 2015).

References

- Baldi, Ph. & Andrea N. 2010. Possession. In Ph. Baldi & A. Nuti (eds), *New Perspectives on Historical Latin Syntax*, vol. 3. Berlin: Mouton De Gruyter, 239-387.
- Bentley, D. 2015. Predication and argument realization. In D. Bentley, F. M. Ciconte, S. Cruschina. 2015, 99–160.
- Cennamo, M. 2011. Impersonal constructions and accusative subjects in Late Latin. In A. Malchukov & A. Siewierska (eds), *Impersonal Constructions: a Crosslinguistic Perspective*. Amsterdam: J. Benjamins, 169-188.
- Ciconte, F. M. 2015. Historical context. In D. Bentley, F. M. Ciconte, S. Cruschina. 2015, 217–267.
- Cruschina, Silvio. 2015. Patterns of variation in existential constructions. *Isogloss 1/1*: 33–65.
- Gaeta, L. 2013. Existential constructions. In E. van Gelderen, M. Cennamo & J. Barðdal (eds.), *Argument structure in flux: the Naples/Capri Papers*. Amsterdam & Philadelphia: J. Benjamins, 477-509.
- Leumann, M., J. B. Hofmann, A. Szantyr. 1965. *Lateinische Grammatik*. vol. 1. München: Beck.
- Löfstedt, E.. 1933. *Syntactica*, vol 2. Lund: Gleerup.
- Mikulová, J. 2016. *Haberi* + passive perfect participle. *Glotta* 92: 152-174.
- Moreno, J.L. 1978. En torno al sintagma “*Haber* impersonal + sustantivo y sus orígenes latinos. *Revista Española de lingüística* 8.1: 125-147.
- Panayotakis, S. 2016. Expressions of time in early and late Latin: the case of temporal *habet*. In J. N. Adams & N. Vincent (eds), 202: 216.
- Parry, M. 2013. Variation and change in the presentational constructions of north-western Italo-Romance varieties. In E. van Gelderen, M. Cennamo & J. Barðdal (eds.), *Argument structure in flux: the Naples/Capri Papers*. Amsterdam & Philadelphia: J. Benjamins,
- Pinkster, H. 2015. *Oxford Latin Syntax I. The Simple Clause*. Oxford, OUP.
- Raimos Guerreira, A. 1998. Consideraciones sobre la expresión de la posesión externa en latín. In B. García Hernández (ed), *Estudios de Lingüística Latina*, vol 2, Madrid: Ediciones Clásicas, 673-688.
- Reichenkron, G. 1933. *Passivum, Medium und Reflexivum in den Romanischen Sprachen*. Berliner Beiträge zur Romanischen Philologie. III, 1. Jena & Leipzig: Gronau.
- Thielmann, Ph. 1885. *Habere* mit dem Part. Perf. Pass. *Archiv für Lateinische Lexicographie* 2: 372-423.
- Svennung, J. 1935. *Untersuchungen zu Palladius und zur Lateinischen Fachund Volkssprache*. Uppsala: Almqvist & Wiksell.

Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics

Maltese is a prime example of a language that emerged through extensive language contact, joining the two linguistic worlds of Semitic and Italo-Romance languages. Previous studies have demonstrated this on the basis of comparative methods (Comrie, 2009; Comrie & Spagnol, 2016; Lucas & Čéplö, 2020), mostly focusing on the non-concantenative and concantenative morphology of Maltese, with broken plurals such as *kelb-klieb* ‘dogs’ belonging to the first and sound plurals such as *fjura-fjuri* ‘flowers’ belonging to the second language family. The present study aims to extend earlier comparative studies by applying a computational method to the classification of a word’s origin.

To do so, we trained a simple two-layer neural network (NDL, Baayen et al. (2011)) to classify 2-phones from 2347 Tunisian Arabic nouns from Gugliotta and Dinarelli (2020) as Semitic and to classify 2-phones from 2347 Italian nouns from the **subtlex-it** corpus¹ as Non-Semitic. Subsequently, the trained network was tasked with the classification of 6511 Maltese singular and plural nouns from Nieder et al. (accepted) as belonging to the categories **semitic** vs. **non-semitic**.

The network achieved an overall high classification accuracy of 97% in the training data. When the network was required to classify Maltese nouns, which were unknown to the network, as Semitic vs. Non-Semitic, the overall classification accuracy was at 70.76%. When inspecting the classification in more detail, we find that nouns with a Non-Semitic origin show a smaller classification probability as Semitic than those with a Semitic origin (39.0% vs. 84.6%, $\Delta = 45.6\%$, $z\text{-value} = 34.34$, $p < 0.001$). Moreover, we were interested how classification depended on the nouns’ morphological class (sound plural, broken plural, singular sound, singular broken).

While all nouns with a Semitic origin across all morphological classes were classified as Semitic, plurals with a Non-Semitic origin showed a higher classification probability as Semitic when they show a broken pattern. This indicates that Semitic nouns in Maltese are more similar to their Semitic relatives than Non-Semitic nouns to their Non-Semitic relatives in terms of their phonotactic characteristics.

Our results show that it is possible to classify Maltese nouns using a simple two-layer network with a training based on Tunisian and Italian nouns only. The network is sensitive to the phonotactics of individual languages and can use language-specific phonotactic knowledge to classify a language that is the result of extensive language contact.

References

- Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. [Publisher: American Psychological Association]. *Psychological review*, 118(3), 438–481.
- Comrie, B. (2009). Maltese and the world atlas of language structures [Publisher: John Benjamins]. *Introducing Maltese Linguistics*, 3–12.
- Comrie, B., & Spagnol, M. (2016). Maltese loanword typology. In G. Puech & B. Saade (Eds.), *Shifts and patterns in Maltese* (pp. 315–330). De Gruyter Mouton. <https://doi.org/10.1515/9783110496376-012>
- Gugliotta, E., & Dinarelli, M. (2020). TArC: Incrementally and Semi-Automatically Collecting a Tunisian Arabish Corpus. *LREC*.
- Lucas, C., & Čéplö, S. (2020). Maltese. <https://doi.org/10.5281/zenodo.3744525>
- Nieder, J., Chuang, Y.-Y., van de Vijver, R., & Baayen, R. (accepted). A Discriminative Lexicon Approach to Word Comprehension, Production and Processing: Maltese Plurals. *to appear in Language*.

¹downloaded from <http://crr.ugent.be/programs-data/subtitle-frequencies>

From direct quotation to a chain of extended quotations: the history of Hungarian *úgymond* 'so to speak'

It is a well-known phenomenon of grammaticalization that the verb or verb construction of a quotative main clause serves as a source for new elements (e.g. Harris–Campbell 1995: 171, see also 168–172). The topic is quite well-researched for many languages (comprehensive analyses: Kuteva et al. 2019: 357–358, 375–388; Spronck–Casartelli 2021).

In present-day Hungarian there are two discourse markers which started from the quotation and arose from the verb *mond* 'say': *mondván* ('giving the reason') and *úgymond* ('so to speak'). Their history goes back approximately 600 years. At the 25th ICHL conference I discussed the element *mondván* (From quotative to causation – the history of Hungarian *mondván*). The present research focuses on the history of the marker *úgymond*. For this expression, the source used to be a collocation: the demonstrative pronoun *úgy* ('so') and the verb *mond* ('he/she says') fused and lexicalized as a quotative marker, which later acquired an extended discourse marker function (Dömötör 2015).

In the 15th century, and at the beginning of the 16th century (Old Hungarian period) *úgy mond* ~ *úgymond* occurred in two kinds of grammatical roles. On the one hand, as a collocation it was a part of a prototypical quotative main sentence (1–2). On the other hand, the fused form was also present as a grammaticalized quotative marker (3).

- (1) ő **úgy** **mond** **vala** ön+benne: „Ha en csak ő
 he so say-PRS3SG AUX.PST self+he.INE if I only he
 ruhá-já-t illet-end-em, megvigasz-om”
 dress-POS-ACC touch-FUT-3SG.DEF comfort-PRS3SG
'He says to himself: If only I can touch his dress, I would be comforted' (MunC.15rb, 1466)
- (2) az szoror-ok [...] **úgy** **mond-anak** **vala** öneki: „Mi teneked benne?”
 the sister-PL so say-PRS3PL AUX.PST she-DAT what you-DAT it.INE
'The sisters told her: What do you care?' (MargL. 84, 1510)
- (3) ez gonosz füge+fá-ról **mond-á** ur-unk Jézus az igé-k-et:
 this ill fig+tree-DEL say-PST3SG lord-POSS3PL Jesus those word-PL-ACC
 „Íme **úgymond** három esztende-je vagy on, hogy gyümölcs-öt keres-ek [...]”
 lo QUOTDM three year-POSS3SG is that fruit-ACC seek-PRS1SG
'Our Lord, Jesus told us these words about this ill fig tree: Lo, QUOTDM for tree years I have been looking for fruits [...]' (DignAp.11, 1521)

The quotative role of *úgymond* was attested during the Middle Hungarian period (4) and also in subsequent periods, and it also appears today. Meanwhile, the collocation did not lose its original function for a long time (until the 19th century).

- (4) **mond-ja** a fatens báty-ja: „No, megad-á **úgymond** nékem Jóczikné
 say-3SG.DEF the witness brother-POS3SG well give-PST3SG.DEF QUOTDM I.DAT Jóczik's wife”
'The brother of the witness says: Well, Jóczik's wife QUOTDM clobbered me' (TMK, 1716)

Later, from the middle of the 20th century, the function of the quotative marker has extended and shifted towards attitude marking. It is used in present Hungarian primarily in this role, with the meaning 'so to speak'.

On the one hand, the discourse marker *úgymond* signals the speaker's attitude that they use a word or collocation which was taken from an unnamed outside source (possibly from the public discourse), but they distance themselves from that usage (5).

- (5) Minden **úgymond** rossz anyá-nak megvan a maga történet-e
 each DM bad mother-DAT has-PRS3SG the own story-POSS3SG
'Each, so to speak, bad mother has got her own story' (hvg.hu/360/202239 2022.10.11.)

On the other hand, the element can signal that the speaker reflects on their own language use. This usage of *úgymond* is the result of a further extension, as it goes beyond linking to the outside source. Among other things, the speaker can signal they use the following element unusually or inaccurately (6).

- (6) Amikor ezek-et a tartalm-ak-at **úgymond** leszerződte-t-ék, a jogtulajdonos-ok [...]
 when this-ACC the content-PL-ACC DM contract-PST-3PL.DEF the rightsholder-PL
'When they, so to speak, contracted this contents, the rightsholders [...]' (napidroid.hu/netflix 2023.01.14.)

In my presentation, I investigate the steps of the functional extension of *úgymond*. I analyse the first, determining step (grammaticalization) and also focus on the relation between the different discourse marker functions.

References

Dömötör Adrienne 2015. A *mondván* és az *úgymond*: diskurzusjelölő elemek keletkezése idéző szerkezetekből. [*Mondván* 'giving the cause' and *úgymond* 'so to speak': the emergence of discourse markers from quoting constructions.] *Magyar Nyelv* 111: 21–37.

Harris, Alice C. and Campbell, Lyle 1995: *Historical syntax in crosslinguistic perspective*. Cambridge, Cambridge University Press.

Kuteva, Tanja, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog and Seongha Rhee 2019. *World Lexicon of Grammaticalization*. Cambridge, Cambridge University Press.

Spronck, Stef and Casartelli, Daniela 2021. In a manner of speaking: how reported speech may have shaped grammar. *Frontiers in Communication* 6:624486. www.frontiersin.org 1–22.

Cited sources

MunC. = Munnich codex (1466)

MargL. = Margaret legend (1510)

DignAp. = Booklet on the dignity of the saint apostles (1521)

TMK = Történeti Magánéleti Korpusz [Old and Middle Hungarian corpus of informal language use]

Synthetic or analytical: factors which explain the formal variation of future and conditional in Old Catalan

Aina Torres-Latorre

In Romance languages, future and conditional are newly created verbal tenses. They are the result of grammaticalization processes of Latin periphrasis, mainly of the periphrasis CANTĀRE HABEŌ, which has been the successful one in most languages. During the Middle Ages, the grammaticalization process of the two verbal tenses was not yet complete in many Romance varieties, such as Portuguese, Castilian, Aragonese, Catalan, Occitan, or some Italian dialects. Two types of forms could be found: the so-called synthetic forms (*cantaré* ‘I will sing’) and the so-called analytical forms (*cantar-lo he* ‘I will sing it’). Analytical forms differed from synthetic forms because of the need of a weak pronoun between the infinitive and the auxiliary. The presence of a clitic within the verbal tense shows the absence of univerbation, one of the characteristics of grammaticalization (see Lehmann 1985, 2020). Hence, synthetic forms are more grammaticalized than analytical forms.

The aim of this work is to study the variation between synthetic and analytical forms in Old Catalan. The distribution between the two types of forms is not arbitrary. First, it follows syntactic-pragmatic restrictions. In medieval Romance, clitic placement was due to different syntactic environments: ones which entailed preverbal clitics and others in which pronouns were postverbal (for Old Catalan, Batllori *et al.* 2005, Francalanci *et al.* in press). Clitic placement followed the same principles with future and conditional (Sentí & Bouzouita 2022). In this case, preverbal clitics only appeared with synthetic forms (*lo cantaré* ‘I will sing it’), but postverbal pronouns could be used with analytical forms (*cantar-lo he* ‘I will sing it’) or with synthetic forms (*cantaré-lo* ‘I will sing it’). To study the variation between synthetic and analytical forms only those who appear in the same environments can be compared, that is, analytical forms and synthetic forms with postverbal clitics.

Then, in second place, the distribution of synthetic forms with postverbal clitics and analytical forms is not arbitrary either. Some studies based on Old Castilian (Bouzouita 2016a, 2016b) and Old Navarro-Aragonese (Primerano & Bouzouita submitted) suggest some morphological and syntactic factors which could explain a preference for synthetic forms instead of analytical forms, the most common ones in these languages. The proposed factors are: (i) verbal tense (if it is a future or a conditional), (ii) verbal conjugation, (iii) presence of a verbal syncope, (iv) presence of a non-finite verbal form after the future or conditional. This study intends to analyse the behaviour of these factors in Old Catalan, and to suggest differences between Catalan and the other languages studied. The mentioned factors will be reviewed based on data from the 11th century to the 16th century extracted from the *Corpus Informatizat del Català Antic* (CICA).

Corpus

CICA = *Corpus Informatitzat del Català Antic*, Torruella, Joan (dir.), with Manuel Pérez Saldanya and Josep Martines: <http://www.cica.cat>.

Bibliography

BATLLORI, Montserrat / IGLÉSIAS, Narcís / MARTINS, Ana Maria (2005): «Sintaxi dels clítics pronominals en català medieval», *Caplletra*, 38: 137-177.

BOUZOUITA, Miriam (2016a): «La posposición pronominal con futuros y condicionales en el código escurialense I.i.6: un examen de varias hipótesis morfosintácticas», in: Kabatek, Johannes: *Lingüística de corpus y lingüística iberorrománica*, Berlín: De Gruyter.

— (2016b): «La posposición pronominal con futuros y condicionales en el castellano medieval: ¿un caso de contacto de lenguas?», presented in the seminar *El orden de palabras en las lenguas iberorrománicas medievales*, 29th of July of 2016, University of Girona.

FRANCALANCI, Leonardo, Ana Maria MARTINS, Francisco ORDÓÑEZ & Josep E. RIBERA (in press): «Els pronoms febles (I): forma, posició i ordre», in Martines, Josep; Pérez Saldanya, Manuel; Rigau, Gemma (ed.): *Gramàtica del català antic*, Amsterdam: John Benjamins.

LEHMANN, Christian (1985): «Grammaticalization: Synchronic variation and diachronic change», *Lingua e stile*, 20: 303-318.

— (2020): «Univerbation», *Folia Linguistica Historica*, 41: 205-252.

PRIMERANO, Antoine & Miriam BOUZOUITA (submitted): «La gramaticalización de los futuros y condicionales en el navarroaragonés de los siglos XIII-XIV: una comparación con el castellano medieval», *Revue Romane*.

SENTÍ, Andreu & Miriam BOUZOUITA (2022): «The syntax of clitics in Old Catalan: Llibre dels Fets», *Catalan Journal of Linguistics*, 21: 1-31.

Computational Anatolian phylogeny using maximum parsimony

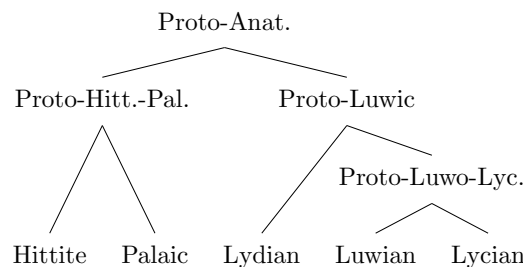
The Anatolian languages constitute an extinct branch of the Indo-European language family, attested across modern day Turkey from ca. the 19th cent. BCE to the 2nd cent. CE (Zinko 2017). Prominent members include Hittite, Luwian, Lydian, and Lycian.

Previous traditionally oriented work on the internal phylogeny of Anatolian has not reached a consensus. For example, in some studies (Oettinger 1979; Kloekhorst 2022), Palaic, Luwian, and Lycian form a clade, whereas other treatments (Melchert 2003; Rieken 2017) tend to assume a closer relationship between Lydian, Luwian, and Lycian to the exclusion of Palaic. It is consequently warranted to explore alternative methods for determining the topology of the Anatolian tree.

Recent studies applying computational methods to linguistic phylogenies have mostly operated with lexical cognate data as the sole input (e.g. Bouckaert et al. 2012; Chang et al. 2015; Ringe et al. 2002; Ringe et al. use some morphological characters but still principally base their analysis on lexical data). This MO is not viable for Anatolian, as the languages involved are too scarcely attested to allow for the compilation of a reliable cognate data set. Indeed, the material available for Carian, Sidetic, and Pisidian is too scanty for any analysis. Rather, Anatolian phylogeny must operate primarily with phonological and morphological data.

A solid candidate method for conducting computational phylogeny using non-cognate based data sets is *Maximum Parsimony*. This study employs PAUP* (Swofford 2003) to infer an Anatolian tree on the basis of a data set consisting of 27 characters (12 phonological and 15 morphological) gathered from existing literature and additional research. The taxa involved are Hittite, Palaic, Lydian, Luwian, and Lycian. Considering that the characters used here are predominantly the result of the historical-comparative method, a root state is often possible to assign confidently. Consequently, our tree is rooted. Characters are assigned a weight from 1–4 on the basis of pre-established parameters (e.g. *unconditioned sound changes* are weighted 1 and *sporadic sound changes* are weighted 4). It should be noted that these parameters are unavoidably to some extent subjective, but we do not expect any strong objections from specialists.

Our analysis gives the following most parsimonious tree:



A bootstrap analysis (Felsenstein 1985) indicates that our tree is highly robust. However, considering that the innovations assumed for Proto-Hittite-Palaic under this topology are rather trivial, it may be most prudent to assume a polytomy between Hittite, Palaic, and Proto-Luwic, pending further evidence.

The method used in this study could be exploited for other scarcely attested extinct language families (e.g. Sabellic). An advantage over alternative methods is furnished by the transparency in the grounds on which trees are evaluated (cf. Hammarström et al. 2019: 236). Accordingly, a classically trained historical linguist and/or specialist on the language family at hand is given the opportunity to qualitatively assess the validity of the developments postulated by the analysis.

References

- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., . . . Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, *337*(6097), 957–960.
- Chang, W., Hall, D., Cathcart, C., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, *91*, 194–244.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.
- Hammarström, H., Rönchen, P., Elgh, E., & Wiklund, T. (2019). On computational historical linguistics in the 21st century. *Theoretical Linguistics*, *45*(3-4), 233–245.
- Kloekhorst, A. (2022). Anatolian. In T. Olander (Ed.), *The Indo-European Language Family. A Phylogenetic Perspective* (pp. 63–82). Cambridge: Cambridge University Press.
- Melchert, H. C. (2003). The dialectal position of Lycian and Lydian within Anatolian. In M. Giorgieri, M. Salvini, M.-C. Trémouille, & P. Vannicelli (Eds.), *Licia e Lidia prima dell'Ellenizzazione. Atti del Convegno internazionale. Roma, 11-12 ottobre 1999* (pp. 265–272). Roma: Consiglio Nazionale delle Ricerche.
- Oettinger, N. (1979). Die Gliederung des anatolischen Sprachgebietes. *Zeitschrift für vergleichende Sprachforschung*, *92*, 74–92.
- Rieken, E. (2017). The dialectology of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 298–308). Berlin – Boston: De Gruyter Mouton.
- Ringe, D., Warnow, T., & Taylor, A. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society*, *100*(1), 59–129.
- Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. <http://paup.phylosolutions.com> /. Sunderland, Massachusetts: Sinauer Associates.
- Zinko, C. (2017). The documentation of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 239–249). Berlin – Boston: De Gruyter Mouton.

The Diachrony of Person-Number Marking of Subjects in Celtic

Mark Darling (University of Oxford), Marieke Meelen (University of Cambridge), David Willis (University of Oxford)

The system of marking person and number of verbal subjects in the Celtic languages has undergone considerable changes in the history of the language family. The earliest examples of the family – the continental Celtic languages of antiquity (Jordán Cólera 2019; Lambert 2003) and the early stages of Irish (Thurneysen 1998), Welsh (Evans 1964), Breton (Lewis and Piette 1990), and Cornish (Lewis 1990) from the medieval period – attest richly inflected verbal systems, with distinct verbal morphology for almost all persons and numbers. Alongside this, they attest null subjects: a non-NP subject of a verb is unexpressed in unmarked contexts.

This contrasts markedly with the situation found in later stages of the Celtic languages. Through the Middle Irish and into the Early Modern Irish period, both the verbal and the nominal morphology of the language becomes increasingly impoverished, and subject pronouns, which were already obligatory with the copula and the defective verb *ol* ‘says’ in later Old Irish, become required in ever more contexts. This development has traditionally been described as occurring first with verbs in the passive voice, with subject pronouns then spreading to be required in most contexts that had previously required a null subject. In modern standard Irish, most persons are expressed by the combination of an uninflected verbal form and a personal pronoun, while there remain inflected forms for some persons in some tenses, which require a null subject, e.g., *léim* ‘I read’ vs. *léann sé* ‘he reads’. Inflected forms are distributed to different persons of the verb in different tenses, and the amount of inflection preserved varies between dialects of the modern language. There are, however, questions to be answered as to the precise details of how the development takes place, and how the various types of pronouns and pronominal particles in the language participate in and are affected by the change from null subjects to overt pronominal subjects.

The developments undergone by Welsh are somewhat more complex than those found in Irish. While very little verbal morphology was lost over the history of the language – modern spoken Welsh retains distinctive personal endings for almost all persons of the verb, with syncretism being limited to the 1st and 3rd persons plural in the preterite and conditional – there has been considerable change regarding the null subject parameter in the language. In Middle Welsh an alternation can be observed between pronominal subjects in preverbal position, which must always be overt, and those in postverbal position, which may be either overt or null. Over the course of the development from Middle Welsh to modern spoken Welsh, this flexibility has been lost: subjects can only be post-verbal, and must always be overt, e.g., *gweles i* ‘I saw’, *gweloch chi* ‘you (pl.) saw’. Moreover, Welsh exhibits other interesting agreement properties, such as plural NPs triggering singular agreement on verbs, e.g., *gwelodd y plant* ‘the children saw (sg.)’ vs. *gwelon nhw* ‘they saw (pl.)’.

In this paper, we will use two newly developed historical corpora to examine the factors involved in the loss of null subjects in both Irish and Welsh. We will examine the extent to which Information Structure (IS) factors influence this development in both languages, as opposed to the development being purely morphosyntactic. For instance, it seems plausible that overt subject pronouns in Middle Welsh are more frequent in new-subject and/or topic-switch contexts. Do overt subjects spread from these environments to less IS-marked contexts? It is not clear that there is any correlation with verbal morphology. On the other hand, in Irish, it is far less clear that IS-factors play any role, and a correlation with morphological richness seems to be more evident. We will test these hypotheses against data from a selection of medieval and early modern texts from the two languages.

- Evans, D. Simon. 1964. *A Grammar of Middle Welsh*. Dublin: Dublin Institute for Advanced Studies.
- Jordán Cólera, Carlos. 2019. *Lengua y Epigrafía Celtibéricas*. 2 vols. Vol. 1. *Monografías de filología griega*. Vol. 29. Zaragoza: Universidad de Zaragoza.
- Lambert, Pierre-Yves. 2003. *La langue gauloise: description linguistique, commentaire d'inscriptions choisies*. 2 ed. Paris: Éditions Errance.
- Lewis, Henry. 1990. *Handbuch des Mittelkornischen*. Translated by Stefan Zimmer. Edited by Wolfgang Meid. *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck.
- Lewis, Henry, and J. R. F. Piette. 1990. *Handbuch des Mittelbretonischen*. Translated by Wolfgang Meid. Edited by Wolfgang Meid. *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck.
- Thurneysen, Rudolf. 1998. *A Grammar of Old Irish*. Translated by Daniel A. Binchy and Osborn Bergin. Dublin: Dublin Institute for Advanced Studies. 1946.

The loss of word-initial consonants in Kera'a – A challenge for phonological theory
 Uta Reinöhl, Kirsten Culhane, Naomi Peck (all: University of Freiburg)

Kera'a (or Idu Mishmi), spoken in the districts Lower Dibang Valley and Dibang Valley of Arunachal Pradesh, shows an intriguing loss of most initial consonants in multi-syllabic words. This loss affects the progressive dialect Midu, while the consonants in question – non-aspirated stops, *h*-, nasals, glides – are retained in the more conservative dialect Mithu. For example, *aphe* 'bat' corresponds to *kaphe*, *ili* 'pig' to *bili*, *iku* 'dog' to *miku*, and *am^we* 'wild boar' to *yam^we*. While initial consonant loss has been noticed elsewhere in and outside of Trans-Himalayan (e.g., in Kiranti languages (Guillaume Jacques, personal communication), in at least 50 Australian languages (Hale 1964, Blevins 2001), in Sogeram languages of New Guinea (Don Daniels, personal communication)), it remains a challenge for the basic and common assumption in phonological theory that CV is the optimal syllable type given its wide-spread distribution in languages around the world, early acquisition in childhood, and the impression that it strikes an optimal balance between ease of articulation and signal clarity. Furthermore, consonants in word-initial position are generally considered to be 'prominent' cross-linguistically (Beckman 1998, Barnes 2002); for example, word-initial consonants undergo strengthening in a number of languages (Keating et al 2003).

A reduction of the consonant onset in word-initial position is thus unexpected for these and other theoretical and empirical reasons, and also because there is no obvious pathway of phonological change that would lead to it. For Australian languages, it has been argued that a stress shift from the first to the second syllable invited the consonant loss, but this account does not hold up for some of the Australian languages showing the change, including Ogunyjan and Oykangand–Olgol (e.g. Blevins 2001). For Kera'a, Reinöhl (2022) claims that a stress shift also does not play a role, but this claim still awaits a detailed empirical investigation. Since the change is ongoing rather than historical in Kera'a, we are able to draw on synchronic, acoustic data, comparing Midu and Mithu variants of the same words. Having explored the role of stress, we will consider alternative hypotheses for how to motivate this rare and theoretically important change. One of the alternative hypotheses that we will focus on is inter-vocalic lenition in connected speech, resulting from the prior loss of consonantal codas in Kera'a. This talk thus builds on Reinöhl (2022) studying initial consonant loss in Kera'a in greater phonological and phonetic detail, and examines the implications of the Kera'a data for one of the most wide-spread assumptions in phonological and specifically phonotactic theory.

Barnes, Jonathan. 2002. Domain-initial strengthening and the phonetics and phonology of positional neutralization. *Proceedings of the North East Linguistic Society* 32, 1-20.

Beckman, Jill N. 1998. *Positional faithfulness*. Doctoral Dissertation, University of Massachusetts Amherst.

Blevins, J. 2001. Where have all the onsets gone? Initial Consonant Loss in Australian Aboriginal Languages. In Jane Simpson, David Nash, Mary Laughren, Peter Austin, and Barry Alpher (eds.) *Forty years on: Ken Hale and Australian languages*. Canberra, The Australian National University: 481-492.

Hale, Kenneth. 1964. Classification of Northern Paman Languages, Cape York Peninsula, Australia: a research report, *Oceanic Linguistics* 3, 248-265.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In *Phonetic interpretation: Papers in Laboratory Phonology VI*, ed. John Local, Richard Ogden, and Rosalind Temple, 145–163. Cambridge: Cambridge University Press.

Reinöhl, Uta. 2022. Locating Kera'a (Idu Mishmi) in Its Linguistic Neighbourhood: Evidence from Dialectology. In Mark W Post, Stephen Morey & Toni Huber (eds.), *Ethnolinguistic Prehistory of the Eastern Himalaya*, 232–263. Leiden: Brill.

Anglo-Scandinavian Contact Influence on Verbs Entering the Causative Alternation

Keywords: Anglo-Scandinavian contact, labilization, causative alternation, corpus linguistics

The English causative alternation and which verbs participate in it in true syntactically labile fashion is a puzzle attempted from many angles and widely discussed (Levin & Rappaport Hovav 1994, 2012; cf. Schäfer 2009, a.o.). Diachronically, several derivational morphemes have contributed to the English set of causative-anticausative verb pairs (van Gelderen 2011, 2018; García García 2012, 2020; Ottósson 2013). One of them is the Germanic causativizer *-j* that is reflected in Old English (OE) as in *meltan* and *miltan* ‘(cause to) melt’ or *byrnan* and *bærnan* ‘(cause to) burn’ (cf. van Gelderen 2018, p. 85-96). However, while pairs like *fall* and *fell* or *sit* and *set* are still distinct in Present Day English (PDE), a number of these verb pairs became formally indistinct when the derivation had become both intransparent and unproductive (cf. Ottósson 2013, García García 2020, van Gelderen 2011). The resulting merged lexemes like *melt* and *burn* have thus become syntactically labile and entered the causative alternation (cf. Levin 1993), some of them in early Middle English already. In contrast to Visser’s (1963) account of 55 labile verbs in OE, McMillion (2006) documents a stark increase to over 800, of varying sources and derivations, until PDE. However, one possible factor in the labilization of English *-j* derived verbs entering the causative alternation however is scarcely discussed: Anglo-Scandinavian contact. Linguistic contact between these somewhat mutually intelligible and very closely related languages has been shown to have led to changes in the lexicon (Durkin 2014; Dance, Pons-Sanz & Schorn 2019). A number of verbs in the alternating verb classes of PDE (Levin 1993) show cognate influence of varying degrees, like *melt*, *burn* and *run*.

This work investigates whether the set of Germanic derived causative verbs in English has been influenced by contact with Old Norse cognate verb pairs. Specifically concerning the labilization of Germanic derived causative verbs in English, this work explores a possible correlation of the level of Old Norse contact influence on these lexical items and their beginning participation in the causative alternation during late Old English and Middle English. To this end a corpus analysis of OE derived causative verbs, their anticausative base verbs as well as their descendant lexemes in Middle English illustrates if and when these lexemes enter into the causative alternation in Middle English. Data are extracted from the *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor, Warner, Pintzuk & Beths 2003) and *The Penn-Helsinki Parsed Corpus of Middle English (2nd edition)* (Kroch & Taylor 2000). Work on Norse cognate verb pairs influences on these lexemes (Dance, Pons-Sanz & Schorn 2019) is considered in relation to the timing and extent of their labilization. If the analysis shows that verbs with significant Old Norse cognate influence show differences in timing or pace of labilization to lexemes without significant Old Norse influence, one might consider contact between Anglo-Scandinavian cognate verb sets as a factor in the labilization of these originally derived causative verbs. Formal and functional overlap between cognate verb pairs not only across their derivational relation, but also cross-linguistically due to the close genealogical relationship of the languages in contact might have led to increased ambiguity in a set of verbs already affected by derivational intransparency due to the loss of *-j*-causativization (cf. García García 2012, 2020; van Gelderen 2011; Ottósson 2013).

References

- Dance, R., S. Pons-Sanz, and B. Schorn. (2019). *The Gersum Project: The Scandinavian Influence on English Vocabulary* (Cambridge, Cardiff, and Sheffield, 2019)
- Durkin, Philip. 2014. *Borrowed Words: A History of Loanwords in English*. Oxford/New York: Oxford University Press.
- García García, Luisa. 2012. Morphological Causatives in Old English: The Quest for a Vanishing Formation1. *Transactions of the Philological Society*, 110(1), 122–148. <https://doi.org/10.1111/j.1467-968X.2012.01287.x>
- García García, Luisa. 2020. The Basic Valency Orientation of Old English and the Causative Ja-Formation: A Synchronic and Diachronic Approach. *English Language & Linguistics* 24 (1). 153–77. <https://doi.org/10.1017/S1360674318000345>.
- Gelderen, Elly van. 2011. Valency changes in the history of English. *Journal of Historical Linguistics*, 1(1), 106–143. <https://doi.org/10.1075/jhl.1.1.05van>
- Gelderen, Elly van. 2018. *The diachrony of verb meaning: Aspect and argument structure*. New York: Routledge.
- Kroch, Anthony & Ann Taylor. 2000. *The Penn-Helsinki Parsed Corpus of Middle English, second edition (PPCME2)*. Philadelphia: University of Pennsylvania <https://www.ling.upenn.edu/ppche/ppche-release-2010/PPCME2-RELEASE-3/>
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, Beth & Malka Rappaport Hovav. 1994. A Preliminary Analysis of Causative Verbs in English. *Lingua* 92. 35–77.
- Levin, Beth & Malka Rappaport-Hovav. 2012. Lexicon Uniformity and the Causative Alternation. In M. Everaert, M. Marelj, & T. Siloni (Eds.), *The Theta System: Argument Structure at the Interface*. 150–176. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199602513.001.0001>
- McMillion, Alan. 2006. *Labile Verbs in English: Their Meaning, Behavior and Structure*. Dissertation. Stockholm: Department of English, Stockholm University. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-994>
- Ottósson, Kjartan. 2013. The anticausative and related categories in the Old Germanic languages. In F. Josephson & I. Söhrman (Eds.), *Diachronic and Typological Perspectives on Verbs* (Vol. 134). 329–382. Amsterdam: Benjamins <https://doi.org/10.1075/slcs.134.12ott>
- Schäfer, Florian. 2009. The Causative Alternation. *Language and Linguistics Compass*, 3(2), 641–681. <https://doi.org/10.1111/j.1749-818X.2009.00127.x>
- Taylor, Ann., Warner, A., Pintzuk, S. and Beths, F. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Electronic texts and manuals available from the Oxford Text Archive.
- Visser, Frederik Theodor. 1963. *An historical syntax of the English language*, vol. 1. Leiden: Brill.

The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database

One of the less discussed limits of Comparative Method is the database requirements for establishing regularity of correspondences. We may consider a hypothetical example assessing relatedness among three languages – A, B, C. For Correspondence 1, there are 20 cognate sets where Langs A and B have /t/ and Lang C has /s/. For Correspondence 2, there are 2 sets where Lang A has /n/ and Lang B has /ŋ/, 2 different sets where Langs A and C have /n/, and no other [anterior] nasal sets. Correspondence 1, A /t/, B /t/ C /s/, is more securely based than Correspondence 2, A /n/, B /ŋ/, C /n/ and it is not clear that Correspondence 2 is sufficiently supported (Mailhammer 2015; Mailhammer & Harvey 2018).

The basic aim of the Comparative Method is to determine whether inheritance is better supported than chance or contact as explanations for similar forms (Harrison 2003; Weiss 2014). We propose that it is possible to statistically evaluate chance vs inheritance in cases where the database is limited, such as Correspondence 2. There are different mathematical approaches have been applied to assess correspondences in hypothesised distant genetic relationships (Ringe 1993; Oswald 1993).

We propose a further development of Ringe (1992), assessing how likely it is that sound correspondences are accidental using a binominal distribution formula. The key information needed for this method is how frequent on average each relevant phoneme is in each relevant position across all the languages in the sample. Frequencies are multiplied for each phoneme considered and the formula determines whether a match in Y languages across a sample of X languages is accidental. We propose two innovations. The first is assessing the frequency of phonemes across entire lexicons to avoid false positives (Baxter 1993). The second is to compare only identical matches, i.e. identical phonological forms that share a common meaning.

We exemplify this method using a database from Australian languages, which provide a good testing ground, as there are proposals for extensive genetic relationships among Australian languages (Harvey & Mailhammer 2017; Koch 2014), but there are limited numbers of potential cognates and establishing correspondences is problematic (Miceli & Round 2022). We assembled phonologically identical forms expressing 25 lexical concepts across Australian languages, and we examined full lexicons from 35 Australian languages to determine average frequencies of phonemes in all phonotactic positions.

Results show that widespread identities are very unlikely to be due to chance. For example, there are 27 languages where the form /pu/ conveys the meaning 'hit'. It can be ruled out that this match is accidental on a 0.000 level, irrespective of whether average or the highest attested frequencies for /p/ and /u/ are used in the calculation. This even holds if the frequencies of /p/ in initial and /u/ in second position were 50%, which is of course unrealistic across the lexicon of any language. Given that chance is not supported, the remaining hypotheses on shared forms, such as /pu/ 'hit', are contact or inheritance. We do not consider the contact vs inheritance choice here, but note that there are well-established criteria bearing on the choice: e.g. continuity vs discontinuity of attestation (Harvey & Mailhammer 2017); variation in probability of borrowing by part-of-speech class and semantic domain (Tadmor, Haspelmath & Taylor 2010; Tadmor 2009).

References:

- Baxter, William. 1993. Response to Oswalt and Ringe. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 217–236. Amsterdam & Philadelphia: John Benjamins.
- Harrison, S. P. 2003. On the Limits of the Comparative Method. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*. Maldon, MA: Blackwell.
- Harvey, Mark & Robert Mailhammer. 2017. Reconstructing remote relationships: Proto-Australian noun class prefixation. *Diachronica* 34(4). 470–515.
- Koch, Harold. 2014. Historical relations among the Australian languages. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia: a comprehensive guide* (The World of Linguistics 3), 23–89. Berlin & Boston: Walter de Gruyter.
- Mailhammer, Robert. 2015. Etymology. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 423–441. London: Routledge.
- Mailhammer, Robert & Mark Harvey. 2018. A reconstruction of the Proto-Iwaidjan phoneme system. *Australian Journal of Linguistics* 38. 329–359.
- Miceli, Luisa & Erich Round. 2022. Where have all the sound changes gone? Examining the scarcity of evidence for regular sound change in Australian languages. *Linguistics Vanguard* 8(s5). 509–518. <https://doi.org/10.1515/lingvan-2021-0094>.
- Oswalt, Robert. 1993. A probabilistic evaluation of North Eurasian Nostratic. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 199–216. Amsterdam & Philadelphia: John Benjamins.
- Ringe, Don. 1993. Probabilistic evidence for Indo-Uralic. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 153–197. Amsterdam & Philadelphia: John Benjamins.
- Ringe, Donald. 1992. On calculating the factor of chance in language comparison. *Transactions of the American Philosophical Society* 82(1). 1–110.
- Tadmor, Uri. 2009. Loanwords in the World's Languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the World's Languages*, 55–75. Berlin/New York: De Gruyter Mouton.
- Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.
- Weiss, Michael. 2014. The Comparative Method. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 127–145. London: Routledge.

‘Old presents’ and the layered history of the Andi verb

In a seminal article on the typology of verbal inflection, Haspelmath (1998) points out that many anomalous features in the shape and behaviour of imperfective verb forms cross-linguistically can be explained as a side effect of grammaticalization. Once an innovative present tense grammaticalizes into the TAM system of a given language, the pre-existing formation whose central function it takes over – labelled an ‘old present’ – may become restricted to more peripheral roles (whether in terms of semantics or of lexical distribution) that have only their diachronic heritage in common. This observation predicts that at a particular moment in a language’s history, traces of numerous diachronic layers of present formation may be preserved side by side in its synchronic morphology and morphosyntax. The formal complexity of such a system thus provides clues to its development.

In this paper we present a verb system of just this kind in Andi, an understudied minority language of the East Caucasian family, and show that the unusual functional distribution of its morphological material makes sense as the result of a particularly multi-layered history, in which each successive imperfective formation has encroached upon the domain of the one preceding it.

Andi, belonging to the Avar-Andic branch of East Caucasian and spoken by approximately 20,000 people in a handful of villages in mountain Daghestan, is a largely unwritten language attested only since the late 19th century: our material is drawn from the two printed works comprising the Andi corpus (Magomedova & Alisultanova 2010 and Luke 2015), complemented by dialect descriptions and the results of recent fieldwork. A striking feature of Andi morphology is the division of its verb system into two formal zones based on distinct, lexically listed inflectional stems – neither of which, however, has an identifiable function in its own right, e.g. they do not straightforwardly encode tense or aspect. Instead, the longer of these two stems serves as the basis for a disparate range of somewhat peripheral verb forms, including certain specialized converbs, the negative (but not positive) imperative, the typologically notable ‘counterexpectation present’ (Maisak & Verhees 2020), the future, and the present habitual – but not the basic present itself, which uses Stem 1.

| ‘comb’ | Stem 1 <i>roxo-</i> | Stem 2 <i>roxud-</i> |
|--------|--|--|
| | <i>roxo-∅</i> AOR ‘combed’ | <i>roxud-ja</i> FUT ‘will comb’ |
| | <i>roxo-rado</i> PRS ‘combs, is combing’ | <i>roxud-o</i> HAB ‘(generally) combs’ |
| | <i>roxo-ddu</i> PF ‘has combed’ | <i>roxud-abiddu</i> COUNTEREXP.PRS ‘still isn’t combing (!)’ |
| | <i>roxo-r</i> MSD ‘(action of) combing’ | <i>roxud-obʔ:ij</i> ANT.CVB ‘before combing’ |
| | <i>rox-o!</i> IMP ‘comb!’ | <i>roxud-os:ub!</i> NEG.IMP ‘do not comb!’ |

Table 1. Some examples of Gagatli Andi finite and non-finite verb forms based on Stems 1 and 2

We show that this complex synchronic situation can be explained by reconstructing a series of developments whereby each ‘new present’ takes over the central functions of the preceding one, while small sets of lexical items may resist the change. Thus the current PRS *roxorado* ousted what is now FUT *roxudja*, which once had a more general non-past value, as attested by the existence of an identical participial form referring to inherent characteristics, e.g. [*hinuk:u*] *dašdja* ‘openable [from inside]’ (Salimov 2010: 222); this value also survives on finite modal verbs, e.g. FUT *ʔidja* ‘may’.

However, *roxudja* was itself an innovation, marginalizing earlier non-past *roxudo*; the latter formation in fact still serves as the basic present for precisely one verb, meaning ‘go’. We give cross-dialectal evidence that *-dja* and the other Stem 2 forms are based on the formation underlying *-do* – some of them via an imperfective participle in *-dob*, which survives only as the suffix deriving ordinals.

Meanwhile, internal reconstruction and Avar-Andic parallels allow us to identify *-do* as an innovation itself: it is the grammaticalized present in *-o* of an iterative in **-id-*. And remarkably, a few exceptional verbs retain a present signalled by this *-o* directly. This means we can identify formations from at least *four* diachronic layers coexisting with basic present value for different verbs in Andi.

References

- Haspelmath, Martin (1998). The semantic development of old presents: new futures and subjunctives without grammaticalization. *Diachronica* 15.1: 29-62.
- Luke (2015) = *Lukašdi bosam rerhanol'li' xabar* [The good news according to Luke]. Moscow: Institut perevoda Biblii.
- Maisak, Timur, and Samira Verhees (2020). *Counterexpectation present in Andi*. MS., HSE University, Moscow.
- Salimov, Khangerej (2010). *Gagatlinskij govor andijskogo jazyka* [The Gagatli dialect of Andi]. Makhachkala: IJaLI.

Quasi-Suffixaufnahme in Classical Armenian

The phenomenon of *Suffixaufnahme*, or case stacking, refers to a type of morphosyntactic agreement whereby a dependent noun or phrase shows case agreement with its head noun in addition to its regular, functional case marking. Prototypically, the dependent noun occurs in the genitive case, signifying appurtenance, but other cases may also be involved (PLANK 1995).

The phenomenon is most well known from its occurrence in Old Georgian (BOEDER 1995), Hurrian (WEGNER 2007:69–75), and Urartian (SALVINI AND WEGNER 2014:29–31), but also occurs in other languages of the Caucasus and ancient Middle East as well as in some languages of Australia (e.g. Lardil, cf. RICHARDS 2013); the vast majority of the languages concerned exhibit agglutinative morphology. Examples (1) and (2) illustrate this structure in Old Georgian and Hurrian, respectively.

- | | | | |
|-----|---|-----|---|
| (1) | <i>šəcevn-ita čmid-isa sameb-isa-jta</i> help-INS holy-GEN trinity-GEN-INS “with the help of the Holy Trinity” (Sos 1980 no. 2) | (2) | <i>sen(a)=iffu=ue=ne=z asti=i=z</i> brother-1SG.POSS-GEN-CON-ERG wife-3SG.POSS-ERG “my brother’s wife” (Mil. III 7) |
|-----|---|-----|---|

Agreement by *Suffixaufnahme* is not obligatory in all languages which possess the pattern, but may be used to disambiguate or in a limited subset of case combinations.

Indo-European languages are not commonly known to exhibit this particular agreement pattern, although limited parallels exist (e.g. in Slavonic, cf. CORBETT 1995). In Classical Armenian, however, a very similar type of agreement does occur: dependents of heads in the accusative marked by the direct object proclitic *z=* may optionally also receive the same proclitic marking regardless of their functional case. This type of agreement is most common with genitive dependents (3), but also extends to other cases (4) and even subordinate clauses (5).

- | | |
|-----|--|
| (3) | <i>covac’uc’anēr z=vardapetut’ean=n z=xorut’iwn</i> plunge-into.3SG.PST OBJ=teaching.GEN.SG=DET OBJ=depth.ACC.SG “he plunged into the depth of the teaching” (Koriwn §111) |
| (4) | <i>xoselov z=noc’anē z=amenayn č’arut’iwn</i> tell.INF.INS OBJ=3PL.ABL OBJ=all wickedness.ACC.SG “relating all their (lit. from them) wickedness” (Elišē III.234) |
| (5) | <i>varesc’ē z=tiezerakan išxanut’iwn=d z=or awandeał ē dma</i> use.3SG.AOR.SBJV OBJ=universal power.ACC.SG=DET OBJ=REL.NOM.SG give.PTCP be.3SG 3SG.DAT <i>y=Astucoy</i> from=God.ABL.SG “he will use the universal power which was given to him by God” (Elišē II.130) |

This paper aims to explore two dimensions of this agreement phenomenon in Classical Armenian: firstly, a classification of its usage in 5th-century texts by types of cases and nominal phrases involved; and secondly, an attempt at an explanation of the development of this pattern.

Historically, the direct object proclitic is a preposition; other prepositions in Classical Armenian do not, however, show similar agreement by repetition in non-translated texts. Since Armenian nominal morphology rarely distinguishes NOM and ACC, an internal explanation based on the need or wish to disambiguate cannot be excluded. Yet, given the occurrence of *Suffixaufnahme* in the linguistic area, potential interference from Old Georgian or Urartian cannot be excluded *a priori* despite the limited evidence of other contact phenomena (cf. YAKUBOVICH 2010 on morphological interaction between Armenian and Urartian).

References

- BOEDER, W. (1995) "Suffixaufnahme in Kartvelian," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 151–215.
- CORBETT, G.G. (1995) "Slavonic's Closest Approach to Suffixaufnahme: The Possessive Adjective," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 265–282.
- PLANK, F. (1995) "(Re-)Introducing Suffixaufnahme," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 3–112.
- RICHARDS, N. (2013) "Lardil "Case Stacking" and the Timing of Case Assignment," *Syntax* 16 (1), 42–76.
- SALVINI, M. and WEGNER, I. (2014) *Einführung in die urartäische Sprache*, Wiesbaden: Harrassowitz.
- WEGNER, I. (2007) *Hurritisch. Eine Einführung*, second revised edition, Wiesbaden: Harrassowitz.
- YAKUBOVICH, I. (2010) "Morphological negation in Urartian," *Aramazd. Armenian Journal of Near Eastern Studies* 5 (1), 141–165.

Can fortis stops spirantise without aspiration?

We have known for over 200 years that fortis stops can spirantise to fortis fricatives in phonological change, but it is not clear that we know *why*. Grimm (1822) showed clearly (and others had noticed earlier), for example, that the fortis stops (in bold) in the Latin words in (1) correspond to the fortis fricatives in Gothic, and that Latin preserves Proto-Indo-European stops while Gothic illustrates a Germanic innovation of fricatives.

- (1) Latin Gothic
pes, frater, canis *fotus, broþar, hunds*

A change of this type (something along the lines of an unconditioned: p, t, k > f, θ, x) has also been recognised in other languages, including: Greek, Proto-Iranian, Proto-Italic, High German and Liverpool English. The latter two cases preserve evidence that a fortis *affricate* stage can (or must) intervene between the fortis stop and fortis fricative stages, which would mean that the change should be understood as: p, t, k > pf, tθ, kx > f, θ, x (ignoring the precise place of articulation of the fricatives). This paper is intended as an exploration of what it might mean to say that we *understand* this type of change.

One crucial facet of ‘understanding a type of change’ is to be certain about the nature of the pre-change phonological state into which it can be innovated – any notion that some aspect of a pre-change state might *cause* a change clearly requires this. A major claim along these lines is that: aspiration is required for fortis stops to spirantise in this way. For example, Salmons (2021, 138) writes that “aspiration is often taken for granted as a, or the, motivation for” changes like this, echoing a long tradition, including Whitney (1884, 92), who wrote that “the spirants (f, th, and so on) are almost universally derived from the full mutes ... and they come especially from such mutes as were originally aspirated”. If this claim can be shown to be true, we could reasonably see it as a firm step in the direction of understanding the fortis-stop-to-fortis-fricative change. The claim has never been rigorously tested, however. I test it in this paper.

In order to work out if this claim is true, we need two things:

- (i) a phonetic and/or phonological rationale to link aspiration and affrication/spirantisation to allow us to argue that the claim is plausible
- (ii) a consideration of all (or, rather, many) cases of changes of the fortis-stop-to-fortis-fricative type, to check if the pre-change fortis stops were aspirated in every case

I first show that there is reason to think that condition (i) can be met. A number of such rationales have been proposed: e.g., Davis & Iverson (1995) consider how fission and spreading of place features can account for affrication, which lays the ground for deaffrication to fricatives; Scheer (1999) argues for the inherent incompatibility of the elements representing aspiration and occlusion in a single segment if the former is incorporated into a unitary segment, leading directly to a fricative; and Honeybone (2002) considers a misperception analysis (whereby postaspiration could be reanalysed as affrication, which, again, could allow for deaffrication).

I focus for the main part on (ii). This calls for an engagement with diachronic phonological typology. Honeybone (2016) argues that diachronic phonological typology is possible but complex, requiring both wide-ranging typological surveys *and* detailed analysis of instances of a change. Consonant with this, I argue that previous relevant typological surveys (Kümmel 2007, Cser 2003, Kirchner 1998) largely fit with the claim that fortis stops must be aspirated in order to be able to spirantise, but not completely. I then show that a detailed analysis of potential counterexamples, informed by an understanding of how laryngeal phonology (‘voicing’ and ‘aspiration’) works (following such work as Iverson & Salmons 1995), promises to remove these counterexamples on a principled basis, allowing us argue with some degree of certainty that fortis stops can only spirantise if they are aspirated.

References

- Cser, András. 2003. *The typology and modelling of obstruent lenition and fortition processes*. Budapest: Akadémiai Kiadó.
- Davis, Garry & Iverson, Gregory. 1995. Segment organization in the High German Consonant Shift. *American Journal of Germanic Linguistics and Literatures* 7, 111-127.
- Grimm, Jacob. 1822. *Deutsche Grammatik*. 4 volumes. Second edition. Göttingen: Dieterichsche Buchhandlung.
- Honeybone, Patrick. 2002. *Germanic obstruent lenition: some mutual implications of theoretical and historical phonology*. PhD thesis, University of Newcastle upon Tyne.
- Honeybone, Patrick. 2016. Are there impossible changes? $\theta > f$ but $f \not> \theta$. *Papers in Historical Phonology* 1, 316-358.
- Iverson, Gregory & Salmons, Joseph. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369-396.
- Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles.
- Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden: Reichert.
- Salmons, Joseph. 2022. *Sound change*. Edinburgh: Edinburgh University Press.
- Scheer, Tobias. 1998. A theory of consonantal interaction. *Folia Linguistica* 32, 201-237.
- Whitney, William. 1884. *Language and the study of language: twelve lectures on the principles of linguistic science*. London: Trübner.

Tonogenesis in Baltic and Slavic languages

Tijmen Pronk, Leiden University

Some Slavic and Baltic languages have contrastive tones. Lithuanian, Slovene and Serbian/Croatian/Bosnian/Montenegrin combine contrastive lexical stress with two or more tones. Latvian combines fixed word-initial stress with two or three tones, depending on the dialect. There are three hypotheses about the origin of these tones:

1. The tonal contrast was inherited from Proto-Indo-European
2. The tonal contrast is a shared innovation of Baltic and Slavic languages
3. The tonal contrast arose independently in Baltic and Slavic

There is no consensus among proponents of the last two hypotheses about how the Baltic and Slavic tones would have arisen. Three scenarios have been proposed in the literature:

- a. Tonogenesis through reanalysis of an earlier vowel length contrast
- b. Tonogenesis through reanalysis of an earlier phonation contrast
- c. Tonogenesis through reanalysis of phonetic cues for stress

There are typological parallels for each of these scenarios and all scholars agree that at least two of these mechanisms played some role in the evolution of tone in Baltic and Slavic.

The evaluation of the proposed scenarios is hampered by the fact that the presence of stress has caused changes to the realization of the tones, often depending on whether they occur in stressed or unstressed syllables, as well as the rise of additional tones. This complicates a direct comparison of most of the attested tones across different languages. Instead, it is necessary to peel off layers of innovations in the individual languages first. This is illustrated in the table below, which shows the reflexes of what is generally considered to be a single original tone in 1. medial syllables, 2. initial syllables that are always stressed and 3. initial syllables that sometimes stressed and sometimes unstressed.

| | Aukštaitian Lithuanian | Žemaitian Lithuanian | Latvian | Common Slavic | |
|----|---------------------------|-------------------------|---------|------------------|------------------|
| 1. | -i:- | -j:- | -j:- | -í- | denominal suffix |
| 2. | sè:ti | sje:te | sé:t | sǎeti | 'to sow' |
| 3. | gì:v- | gì:v- | dzj:v- | zì:v- | 'alive' |

In this paper it will be argued that scenario 3b explains the attested distributions best: a contrast between modal and laryngealized syllables can be reconstructed for the common ancestor of the Baltic and Slavic and oppositions that were exclusively tonal arose only after Baltic and Slavic had become separate entities.

Literature

Dybo, V.A., S.L. Nikolayev & S.A. Starostin. 1978. A tonological hypothesis on the origin of paradigmatic accent systems. *Estonian Papers in Phonetics* 1978, 16-20.

Jasanoff, J. 2017. *The Prehistory of the Balto-Slavic Accent*. Leiden/Boston.

Kortlandt, F. 2010. Glottalization and tonogenesis in Athabaskan, Balto-Slavic and Germanic. *Suvremena Lingvistika* 69, 37-44.

Olander, T. 2009. *Balto-Slavic Accentual Mobility*. Berlin.

- Pronk, T. 2013. On Indo-European Tones, Accentuation, and Ablaut. In: *Indo-European Accent and Ablaut* (G. Keydana, P. Widmer & T. Olander, eds.). Copenhagen, 149-176.
- Ratliff, M. 2015. Tonoexodus, Tonogenesis, and Tone Change. In: *Handbook of Historical Phonology* (P. Honeybone, J. Salmons, eds.). Oxford, 245-261.

Metaphor, Overtness and Word Order Routinization

This talk argues for a central role of metaphor in constraining and changing clausal syntax in two important, interconnected ways. First, functor-argument metaphors require the overt expression of arguments (Reinöhl 2016, [anonymized reference]). For example, it is possible to say *Everyone was waiting at the hotel. Finally, Kim arrived*. By contrast, people do not use the functor *arrived* metaphorically without a goal argument: *Everything had been pointing to that conclusion all along. *Finally, Kim arrived*. What they say is *Finally, Kim arrived at it*. Second, overt arguments are required for word order routinization ([anonymized reference]): covert constituents do not need to be ordered. This means that metaphor plays a special role in the conventionalization of word order, because it ensures that in each use of the functor, a decision about the ordering of the functor and its necessarily overt argument(s) must be made.

Our talk builds on cross-linguistic as well as experimental research into metaphor-driven argument overtness, and expands this research with evidence that functor-argument metaphors also feed into word order routinization. This research is supported by Indo-Aryan historical data and framed in accounts of linguistic routinization and automatization (Bybee 2003, Haiman 1994). Hitherto discussion of linguistic routinization has focussed on processes associated with grammaticalization, and we see the effect of metaphor most clearly there.

Reinöhl (2016) shows how the New Indo-Aryan postpositional phrases grammaticalized from various parts of speech, including spatial nouns and adverbs. While these could optionally take arguments in Old and Middle Indo-Aryan, those arguments had become obligatory by New Indo-Aryan, a syntactic shift accompanying semantic reanalyses of functor-argument metaphors. Reinöhl argues that these semantic reanalyses could only occur when the functor and argument were overt, were adjacent, and stood in a particular order. This overtness condition and ordering pattern persists to this day.

Work on grammaticalization has noted the connection between the grammaticalization of lexical items and the rigidification of word order. Lehmann (2002 [1982]) views these as parallel developments. Others (Hopper 1996, Bybee 2003) describe string routinization as a facilitating mechanism for the fixing of word order. However, a crucial missing piece in this story is an explanation for why string routinization should necessarily go hand-in-hand with grammaticalization. We offer, as this missing piece, the requirement that metaphor-creating arguments need to be overt. The role of functor-argument metaphors in establishing bridging cases for grammaticalization is well-known (Svorou 1988, Meillet 1975[1912]). But in precisely these cases, the argument to the functor must be overt. When overt and adjacent, and thus a potential formal whole, functor and argument can then semantically develop greater gestalt properties (Himmelman 1997).

Metaphor has already been shown as fundamental in how we use language. More recently it has been shown to drive argument overtness. Here we argue that it also has a central role in the routinization of word order.

References:

- Bybee, Joan. 2003. "Mechanisms of Change in Grammaticization: The Role of Frequency", in: B. D. Joseph & J. Janda (eds.) *The Handbook of Historical Linguistics*. Oxford: Blackwell, 602-623.
- Haiman, John. 1994. "Ritualization and the Development of Language", in: William Pagliuca (ed.). *Perspectives on Grammaticalization*. Amsterdam/Philadelphia: Benjamins, 3-28.
- Himmelmann, Nikolaus. 1997. *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur*. Tübingen: Max Niemeyer.
- Hopper, Paul J. 1996. "Some recent trends in Grammaticalization", *Annual Review of Anthropology* 25, 217-236.
- Lehmann, Christian. 2002 [1982]. *Thoughts on Grammaticalization* (2nd, revised edition). (Arbeitspapiere Seminar für Sprachwissenschaft 9). Erfurt: Seminars für Sprachwissenschaft der Universität Erfurt.
- Meillet, Antoine. 1975 [1912]. "L'évolution des formes grammaticales", in: Antoine Meillet. *Linguistique Historique et Linguistique Générale*. Paris: Champion, 130-148.
- Reinöhl, Uta. Mai 2016. *Grammaticalization and the Rise of Configurationality in Indo-Aryan*. Oxford: Oxford University Press.
- Svorou, Soteria. 1988. *The experiential basis of the grammar of space. Evidence from the languages of the world*. Buffalo: SUNY dissertation.

**Complexity in counting systems:
early systems vs. modern numerical ones**

Brigitte L.M. Bauer
The University of Texas at Austin
Max Planck Institute for Psycholinguistics

Numerical numerals are at the cross-roads of linguistics, mathematics, cognitive sciences, archaeology, and anthropology. Yet counting or quantification is not necessarily numerical nor is it necessarily language-bound. Infants, for example, recognize (change in) quantity and speakers of languages with few numerals easily identify large quantities.

A strong trend today is the spreading of a counting system that is numerical, decimal, and based on arithmetical operations, especially addition and multiplication. Numerical counting systems with high upper limits—such as decimal systems—are easily qualified in the literature as “complex”. Yet the criteria to identify “complexity” often remain implicit or may be open for discussion.

While the decimal numerical system continues to spread globally, it reduces numeral variation even if many languages in the past and today feature residues of earlier counting or measure systems. The Latin measures of length, for example, were based on body parts, with La. *pes* ‘foot’ equaling sixteen *digiti* (‘fingers’) or four *palmi* (‘palms’), and the *gradus* ‘step’ equaling 2.5 *pedes*, and five *pedes* equaling one *passus* ‘pace’. Similarly in today’s Indo-European languages, residues of earlier quantification systems relate quantity to commodity, each with their own (base) units for example, cf. Engl. *stone* (weight), *pint* (volume, liquid), or UK’s pre-1971 monetary *pound* system; Fr. *pouce* (length), *muid* (volume, dry/liquid), and so forth.

These residues allow to identify the main features of earlier systems, which ultimately may have their origins in non-numerical systems, such as the one based on tokens in the early stages of the agricultural revolution in the Near East.

In this talk, I discuss the spread of the decimal system identifying its various manifestations and evaluating its main characteristics against early systems of quantification as we know them from residues of the type mentioned above. This comparison will provide data and insights to assess the concept of “complexity” in counting systems.

The classification of South Cushitic.

In his seminal work on the classification of African languages, Greenberg (1963) has South Cushitic as one of the primary branches of Cushitic. This proposal has been immensely influential in the interpretation of the history of East Africa. On the basis of this classification, the assumption is that the presence of South Cushitic in Tanzania is ancient (Ehret 1980), and pre-dates the entry of Bantu and Nilotic peoples (Ambrose 1998). As a consequence, (South) Cushitic has been linked to the Savannah Pastoral Neolithic cultural complex that is recognised in archaeology, and it has been proposed that the introduction of agriculture and cattle-keeping in Tanzania can be attributed to the South Cushitic speakers. Ehret's (1980) reconstruction of South Cushitic lexicon and phonology has been extremely influential in recognising linguistic contact in many of East Africa's Bantu and Nilotic languages despite the fact that this reconstruction has been severely criticised (Philippson 2013). Serious doubts on the classification of South Cushitic as primary branch have been raised though. Hetzron (1980) has pointed to many grammatical resemblances between South Cushitic and East Cushitic languages that argue for inclusion of South Cushitic within East Cushitic. In Tosco's (2000) Cushitic overview this uncertainty is represented by the fact that South Cushitic figures differently in his genetic trees: as primary branch of Cushitic for the classical view and as primary branch within East Cushitic reflecting Hetzron's suggestions; the issues are discussed in detail in Kießling (2001). Kießling and Mous (2003) provide an extensive lexical and phonological reconstruction of the four South Cushitic Tanzanian languages that are still spoken; while Kießling (2002) is a detailed morphological reconstruction of these languages. Kießling and Mous (2003) pointed out wider Cushitic parallels where they could but this did not enable them to suggest a position of Tanzanian Cushitic in the Cushitic tree. The challenges are the lexical innovations that Tanzanian Cushitic must have undergone and for many of these no other languages could be suggested as sources. In addition, the other languages that were classified as South Cushitic do not offer much for reconstruction: the languages Aasáx and Qwadza are obsolete and the data on these are unreliable because they were collected from rememberers rather than speakers (Kruijsdijk 2023); Ma'á, often characterised as a mixed language, is Bantu, and not Cushitic, and some of the original Cushitic lexicon after language shift survives in a parallel register which also contains words from a variety of other sources including manipulated words from the basis Bantu vocabulary (Mous 2003). The last suggested member, Dahalo, is now considered to be (marginally) East Cushitic rather than South Cushitic (Tosco 1989, Tosco and Blazek 1994).

I propose that Tanzanian Cushitic is a primary branch of Cushitic after all. I also argue that the earliest South Cushitic expansion into Tanzania was followed by two others that have left their (lexical) impact on Tanzanian Cushitic. The latest is the pre-Oromo influence on Tanzanian Cushitic. The suggestion for such language contact showing transfer from pre-Oromo lexical and morphological material featured already in Kießling and Mous (2003), and was recently substantiated by Rapold (2023). There is plenty of reconstructed Tanzanian Cushitic material that is clearly Cushitic but did not undergo the Oromoid innovations. An earlier expansion is formed by speakers of the Dullay-Yaaku subgroup (see Hayward 1978 that this is a subgroup). Recently Sands and Tosco (2022) have argued that early Dullay-Yaaku speakers must have been in contact with Hadza (a language isolate and in the area of Tanzanian Cushitic). I provide further evidence for this intrusion by showing Dullay-Yaaku influence on proto-Tanzanian Cushitic while the Tanzanian Cushitic proto lexicon also contains Cushitic lexical evidence that pre-dates Dullay-Yaaku. The consequences for the interpretation of East Africa's history are far-reaching: There was not one migration of Cushitic speakers into Tanzania but at least three. For all Cushitic lexical transfer into Bantu and Nilotic languages of Tanzania and Kenya, the source needs to be considered.

References

- Ambrose, Stanley H. 1998. Chronology of the Later Stone Age and Food Production in East Africa. *Journal of Archaeological Science* 25:377–392.
- Appleyard, David 2006. *A comparative dictionary of the Agaw languages*. Cologne: Rüdiger Köppe.
- Ehret, Christopher 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin: Dietrich Reimer.
- Greenberg, Joseph 1963. *The Languages of Africa*. Bloomington: Indiana University
- Hayward, Richard J. 1978. The Qawko Dialects and Yaaku. *Abbay: Documents pour Servir à la histoire de la Civilisation Éthiopienne* 9: 59-70.
- Hetzron, Robert. 1980. The Limits of Cushitic. *Sprache und Geschichte in Afrika* 2:7-126.
- Kießling, Roland. 2001. South Cushitic links to East Cushitic. In *New data and New Methods in Afroasiatic Linguistics*, Andrzej Zaborski (ed.), pp. 95-102. Wiesbaden: Harrasowitz.
- Kießling, Roland. 2002. *Die Rekonstruktion der südkuschitischen Sprachen (West-Rift): von den systemlinguistischen Manifestationen zum gesellschaftlichen Rahmen des Sprachwandels*. (KuS 19). Cologne: Rüdiger Köppe.
- Kießling, Roland and Maarten Mous. 2003. *The lexical reconstruction of West Rift (Southern Cushitic)*. Cologne: Rüdiger Köppe.
- Kruisdijs, Iris 2023. The position of Asá and Qwadza within Cushitic. To appear in *Selected papers from the 53rd Annual Conference on African Linguistics*.
- Mous, Maarten. 2003. *The making of a mixed language: the case of Ma'a/Mbugu*. Amsterdam: John Benjamins.
- Philippon, Gérard 2013. Southern Cushitic loans in North-Eastern Bantu – a reconsideration of the evidence. In *Proceedings of the 5th International Conference on Cushitic and Omotic Languages, Paris, 16-18 avril 2008*, edited by Marie-Claude Simeone-Senelle and Martine Vanhove, pp. 81-93. Cologne: Rüdiger Köppe.
- Rapold, Christian 2023. Pre-Oromo links to Tanzanian Cushitic. Ms.
- Sands, Bonny and Mauro Tosco 2022. Early East African and Peripheral East Cushitic: Foragers and pastoralists in early East Africa. Paper presented on December 14 2022 in the Rift-Valley Network Seminar Series. <https://zenodo.org/record/7440446#.Y8IKq0HMJmM>
- Tosco, Mauro. 1989. The Classification of Dahalo: Another Perspective. Paper, read at the “2nd Symposium on Cushitic and Omotic Languages”, Torino.
- Tosco, Mauro. 2000. Cushitic Overview. *Journal of Ethiopian Studies* 33: 87-121.
- Tosco, Mauro & Václav Blažek, 1994. Between South and East Cushitic: Reconsidering the Position of Dahalo. Paper, read at the 9th Afrikanistentag, 20.9.94, Köln.

Demonstrative modifiers in Middle Hungarian: a complex picture of renewal

BACKGROUND: The definite article grammaticalized in Old Hungarian to systematically encode the definiteness of the noun phrase, and the source of the article was the distal demonstrative *az* 'that'. The demonstrative system, quite atypically (although attested elsewhere, cf. van Gelderen 2011: 207–208), renewed via two different strategies: one involving reinforcement and one involving doubling. In Hungarian, even two pairs of reinforced demonstratives started to be used along with the old demonstrative modifier, namely *ezen/azon* 'this/that' and *imez/amaz* 'this/that', while a determiner doubling construction also emerged in the 16th century. In the latter case, the old demonstrative pronoun, agreeing in number and case, joined to the noun phrase already determined by the definite article.

EMPIRICAL RESEARCH: The reinforcement strategy and the determiner doubling strategy did not appear simultaneously, but following one another, which raises a couple of questions. Queries in historical corpora allowed me to specify numerically the change of ratio between the various strategies. Furthermore, corpus study will clearly show how the use of doubling constructions increased through centuries to arrive at its present day dominance in frequency. Data have been drawn from the Old Hungarian corpus, on the one hand, which is mainly composed of religious texts and from two Middle Hungarian corpora, on the other hand. Middle Hungarian sources include both Bible translations and private documents, thus the empirical research could consider diachronic processes, and variation between registers and individual sub-corpora as well.

In order to approach the most intriguing question as to why two different strategies emerged to renew the old system, the distributional properties of the modifiers as well as the semantic and pragmatic aspects of their uses have also been examined.

ANALYSIS: In the talk, I propose a syntactic structure for each of the patterns and also aim to model the syntactic change from one construction to the other. Reinforced demonstratives take over the construction type of the old demonstrative strategy, while the definite article emerges as the result of a reanalysis, one that corresponds to the so called Head Preference Principle (van Gelderen 2008 and 2011). At the same time, the determiner doubling strategy can be analyzed as an adjunction, which is supported by word order peculiarities (Author 2014) and by remnant constructions featuring two copies of the demonstrative modifier in apposition, rather than the combination of a demonstrative and an article. However, demonstratives in doubling constructions have been further reanalyzed and integrated into the DP domain, in accordance with the so-called 'Specifier Incorporation', (cf. the universal economy principles in: van Gelderen's 2008)

As a result of the closer examination of distributional phenomena, reinforcement strategies turn out not to have covered all the possible functions (see Himmelmann 1996 and Diessel 1999 for the possible contexts of uses). The reinforced pronouns *ezen/azon* 'this/that' were originally identity markers (meaning 'the same'). Their use only gradually expanded into the general anaphoric use, and never bleached enough to also express an exophoric (extra-linguistic deictic) function. The other pair of reinforced pronouns (*imez/amaz*) remained restricted to special contexts, first of all to the so called recognitional use. That is to say, neither of them encoded exophoric deixis. It was only the determiner doubling construction, emerged somewhat later, but completely neutral with respect to all the pragmatic contexts, that filled the gap. Thus the newest strategy of all could naturally replace the original construction, and at the same time spread into the pragmatic contexts in which the reinforced demonstrative were used.

REFERENCES

- Author 2014.** ‘The DP-cycle in Hungarian and the functional extension of the noun phrase.’ In: É. Kiss, K. (ed.): *The Evolution of Functional Left Peripheries in Hungarian Syntax*. Oxford. 56-82.
- Diessel H. 1999.** *Demonstratives. Form, Function and Grammaticalization*. Amsterdam.
- Gelderen, E. van 2008.** Linguistic cycles and Economy Principles. The role of Universal Grammar in language change. In: Eythórsson, Thórhallur (ed.): *Grammatical Change and Linguistic Theory. The Rosendal Papers*. Amsterdam, 245-264.
- Gelderen, E. van 2011.** *The Linguistic Cycle. Language Change and the Language Faculty*. Oxford.
- Himmelmann, N. 1996.** ‘Demonstratives in narrative discourse: a taxonomy of universal uses.’ In: Fox, B. (ed.): *Studies in anaphora*. Amsterdam, 205-254.

The natural stability of ‘unnatural’ morphology

Models can serve as powerful tools for uncovering how a simple change process may lead to striking emergent outcomes [1,2], and likewise, how small revisions in assumptions can lead to dramatic shifts in how a system is predicted to behave [3]. Here we use simple modelling to demonstrate that analogical changes in inflectional systems can be expected to have a particularly counterintuitive, yet empirically well-supported, long-term effect: namely, inflectional classes’ and stem alternation patterns’ resilience against levelling, even as they undergo constant analogical change [4]. We first underscore why this resilience is so surprising. Next, we explain why analogical reasoning in inflectional systems is expected to trigger changes based not only on similarity but also on dissimilarity. Finally, we implement the predicted change processes in a simple iterated evolutionary model and show that their long-term consequence is to support not only the ongoing coalescence and coherence of morphological classes but also their resistance to complete collapse and levelling.

In inflectional systems, idiosyncratic morphological class systems such as inflectional classes and stem alternation patterns are empirically ubiquitous [5, 6] yet they present a stark theoretical challenge. Relative to their absence, these systems (i) incur a learning cost [7], yet (ii) offer no clear functional benefit [8]. Given that they are constantly undergoing analogical changes [4], standard evolutionary logic predicts that they *ought to disappear*—and indeed, early modelling work has implied this conclusion [9,10]. Yet in real languages and families, stem alternations and inflection class systems routinely persist across millennia, even as other, more functionally motivated inflectional phenomena collapse and fade. So, what makes these systems so resilient?

Recent work in experimental and computational psychology [11] has emphasised that inferential reasoning occurs within an *inductive context*, which is appropriate to the situation at hand, and which shapes the inferences that are more or less likely. For instance, people reason differently when items are related by physical distances, versus via a taxonomic tree. Carefully applying the notion of inductive context to inflectional systems reveals a significant implication: analogy is expected to licence both similarity-enhancing and dissimilarity-enhancing inferences. When these two types of inference are implemented in an iterative model, they give rise to two dynamic forces: one of attraction and one of repulsion. As in many attraction-repulsion systems in nature [12,13], this dynamic leads the system to self-organise into areas of internal coherence (i.e., morphological classes) while maintaining difference across them (i.e., avoidance of total levelling).

In conclusion, inflectional classes and other ‘morphomic’ categories [14, 15] have long been held in suspicion within certain theoretical circles [16], and have even earned the label ‘unnatural’ [17]. On the contrary, however, here we demonstrate that they are natural phenomena *par excellence*: they are natural (emergent) outcomes of natural (rational) inference, and they just so happen to be naturally tenacious survivors of aeons of unceasing analogical change.

References

1. Boyd, R. & P. J. Richerson . 1988. *Culture and the evolutionary process*. Chicago: University of Chicago Press.
2. Smaldino, P. E. 2017. Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read & A. Nowak (eds), *Computational social psychology*, 311-331. New York: Routledge.
3. Guest, O. & A. E. Martin. 2021. How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789-802.
4. Maiden, M. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
5. Bach, X. 2018. *The origins of inflectional classes*. D.Phil dissertation, University of Oxford.
6. Herce, B. 2020. *A typological approach to the morphome*. Doctoral dissertation, University of Surrey.
7. Clark, E. V. 1998. Morphology in language acquisition. In A. Spencer & A. Zwicky (eds), *The handbook of morphology*, 374-389. Oxford: Blackwell.
8. Anderson, S. R. 2015. Dimensions of morphological complexity. In M. Baerman, D. Brown & G. G. Corbett (eds), *Understanding and measuring morphological complexity*, 11-26. Oxford: Oxford University Press.
9. Ackerman, F. & R. Malouf. 2015. The No Blur Principle effects as an emergent property of language systems. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA.
10. Esher, L. 2015. Approaches to the function and emergence of metamorphomic structure. Presentation given at Décembrettes 9. Toulouse, December 2015.
11. Kemp, C. & J. B. Tenenbaum. 2009. Structured statistical models of inductive reasoning. *Psychological review*, 116(1), 20.
12. Lennard-Jones, J. E. 1931. Cohesion. *Proceedings of the Physical Society*, 43(5), 461–81.
13. D’Orsogna, M. R., Chuang, Y. L., Bertozzi, A. L., & Chayes, L. S. 2006. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical review letters*, 96(10), 104302.
14. Aronoff, M. 1994. *Morphology by itself*. Cambridge: MIT Press.
15. Round, E. R. 2015. Rhizomorphomes, meromorphomes and metamorphomes. In M. Baerman, D. Brown & G. G. Corbett (eds), *Understanding and measuring morphological complexity*, 29-52. Oxford: Oxford University Press.
16. Bermúdez-Otero, R. & A. R. Luís. 2016. A view of the morphome debate. In A. R. Luís & R. Bermúdez-Otero (eds), *The morphome debate*, 309–340. Oxford: Oxford University Press.
17. Aronoff, M. 2016. Unnatural kinds. In A. R. Luís & R. Bermúdez-Otero (eds), *The morphome debate*, 11-32. Oxford: Oxford University Press.

Diachronic stability of case functions: oblique in Romani dialects

It is well known that case marking is susceptible to change as a result of internal development and language contact (Johanson 2009). In this study I show how different functions of a case form are lost and preserved in Romani, an Indo-Aryan language that has been spoken in Europe since the Middle Ages (Matras 2002).

Structurally, the Romani case system is similar to those found in other new Indo-Aryan languages (NIA) and consists of several layers (Masica 1991). The first layer distinguishes between direct and oblique forms, and the second layer of case markers is attached to the oblique form, cf. Table 1.

Table 1. Case marking in Romani (Kalderash dialect)

| Case | <i>manúš</i> ‘person’ | | <i>bakró</i> ‘sheep’ | |
|------|-----------------------|---------------------|----------------------|--------------------|
| | SG | PL | SG | PL |
| DIR | <i>manúš</i> | <i>manúš</i> | <i>bakró</i> | <i>bakr-é</i> |
| OBL | <i>manuš-és</i> | <i>manuš-én</i> | <i>bakr-és</i> | <i>bakr-én</i> |
| ABL | <i>manuš-és-tar</i> | <i>manuš-én-dar</i> | <i>bakr-és-tar</i> | <i>bakr-én-dar</i> |
| DAT | <i>manuš-és-ke</i> | <i>manuš-én-ge</i> | <i>bakr-és-ke</i> | <i>bakr-én-ge</i> |
| ... | | | | |

In most NIA languages, the oblique serves exclusively as a base form for secondary cases and cannot be used independently. In Romani, however, the oblique has several distinguished functions, as it marks i) animate direct objects, ii) possessor, iii) the recipient of the verb ‘give’, and iv) the experiencer of certain verbs (‘feel pain’, ‘like’) (Matras 2002: 85–87). This range of functions is arguably inherited from Middle Indo-Aryan (MIA) stage, as Romani oblique affixes are continuation of the MIA oblique forms (Beníšek 2009).

The goal of the study is to establish how the different functions of the oblique are preserved across Romani dialects. The data come from the Romani Morpho-Syntax database (RMS; <https://romani.humanities.manchester.ac.uk/rms/>) which contains questionnaire-based elicited data on various Romani dialects from 120 locations in Europe. Table 2 summarizes the distribution of case marking among the aforementioned functions of the oblique in the dataset.

Table 2. Case marking in the contexts typical for the oblique in RMS.

| Function | OBL | DAT | LOC | Other |
|--------------------------|-------|-----|-----|---------------------------|
| Experiencer, ‘like’ | 12,5% | 25% | — | 64% = DIR |
| Experiencer, ‘feel pain’ | 45% | 4% | 24% | 23% = possessive pronouns |
| Recipient, ‘give’ | 50% | 50% | 28% | — |
| Possessor | 58% | 2% | 30% | — |
| Animate direct objects | 88% | — | — | not marked |

I suggest that the three main factors which determine the stability of the oblique marking are i) type- and token frequency of the functions in speech, ii) the availability of other marking for similar semantic roles, and iii) case marking in contact languages. The lexically determined oblique marking (‘like’, ‘feel’, ‘give’) is less stable than the construction related oblique marking (possessor). This is especially clear with the verb ‘like’ which is often borrowed together with the argument structure of the source language. As usually the semantic roles of ‘recipient’ and ‘experiencer’ are marked in Romani with the dative, there is no wonder that the same marking is found alongside the oblique in these contexts. When the dative marking is additionally supported via language contact, the oblique marking can become obsolete (as it happens in the Romani dialects of Eastern Europe). Finally, the function of the oblique related to the differential object marking is very frequent in speech and cannot be replaced by other cases. That is why the oblique is usually preserved unless the differential object marking is lost altogether (sometimes together with the inflectional case systems as it happens in the Romani dialects of Italy).

References

- Beníšek, Michael. 2009. Middle Indo-Aryan Ablative and Locative Markers in Romani. *Indo-Iranian Journal* 52, 335–362.
- Johanson, Lars. 2009. Case and contact linguistics // *The Oxford handbook of case*. Oxford University Press. 494–501.
- Masica, C. 1991. *Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2002. *Romani. A linguistic introduction*. Cambridge: Cambridge University Press.

The Charition Mime: Decoding the “Indian Language” through Typology and Entropy

The 2nd century CE papyrus P.Oxy III.413 records a Greek play titled “Charition” (Χαρίτιον) written by an anonymous author. The play, a parody of Euripides’ “Iphigenia in Tauris,” features a Greek maiden held captive in India and rescued by her brother by getting the local king and people drunk with wine (Page, 1992; Tsitsiridis, 2005; Webb, 2008; Whitmarsh & Thompson, 2013; Crevatin, 2009). Linguistically, the most exciting aspect of this work is the inclusion of large sections of dialogue in an “Indian” language. This language was initially identified as Dravidian, usually as proto-Kannada, by European and Indian scholars (Hultzs, 1904; Rice, 1926; Sastri, 1926). However, this interpretation was later dismissed by most Western philologists following Barnett (1926).

There is documented contact between Greek-speaking Egypt and India at the time (Salomon, 1991; Seldeslachts, 1998), and the “barbaric” language does not appear to be gibberish. Hultzs (1904) and Rice (1926) noted that there would not be much profit in accurately transcribing a gibberish language (including at least one apparent correction of a misspelling), and some phonological and morphological patterns seem compatible with natural languages. They even identified some words as related to modern Kannada, such as $\kappa\omicron\tau\tau\omega\varsigma$ (ultimately a causative related to ಕುಡಿದಿ [kuḍi], “to drink”) and $\kappa\omicron\nu\zeta\epsilon\iota$ (ultimately related to ಕೊಂಚ [koñca], “a little”). Recent proposals have attempted to connect the language to either Kannada or Tulu (Varadpande, 1981; Shivaprasad Rai, 1985; Upadhyaya, 1996), with scholarly work not necessarily impressive (cf. Hall, 2010). No modern study has provided a solid linguistic assessment of whether the language is natural or related to known languages by using expected practices, such as identifying regular sound correspondences.

Our work combines typological examination, linguistic knowledge of Dravidian languages, and quantitative analysis to uncover the nature of the unknown language. The typological assessment considers known features of Dravidian languages (Krishnamurti, 2003) and proposed translations (e.g., Varadpande, 1981) to review the alleged word correspondences in light of the comparative method. The statistical assessment experiments with two approaches, involving a review of the proposed English translations and the production of translations in other languages using automatic methods (Vaswani et al., 2017). In the first approach, the resulting corpus is employed for comparing information measures, such as perplexity (Brown et al. 1992), cross-entropy (Murphy, 2012), and bits-per-character (Shannon, 1951), comparing the unknown language to the translations aggregated by language families. In the second approach, we experiment with classifier models (Pedregosa et al., 2012) operating over skip n-grams (Manning & Schütze, 1999).

With such a multidisciplinary approach, we aim to provide more concrete answers on whether the unknown language in the “Charition” play displays characteristics of natural languages, specifically those of the Dravidian family, and evaluate the reliability of proposed translations. Our methods apply to other undeciphered or partially deciphered ancient texts. At the same time, for this play, they could provide valuable linguistic data for the study of the diachronic development of the Dravidian language family.

References

- Barnett, L. D. (1926). "The Alleged Kanarese Speeches in P. Oxy. 413." *The Journal of Egyptian Archaeology*, 12(1/2), 13–15. <https://doi.org/10.2307/3854174>
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., Lai, J. C., & Mercer, R. L. (1992). "An estimate of an upper bound for the entropy of English." *Computational Linguistics*, 18(1), 31-40.
- Crevatin, F. (2009). "Nota a P.Oxy. III, 413." *Zeitschrift Für Papyrologie Und Epigraphik*, 171, 199–200.
- Hall, E. (2010). "Iphigenia in Oxyrhynchus and India: Greek tragedy for everyone". In Tsitsirides, S. (ed.), *Parachoregema*. University of Crete Press.
- Hultsch, E. (1904), "Remarks on a papyrus from Oxyrhynchus." *Journal of the Royal Asiatic Society*, 1904: 399-405.
- Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge Language Surveys. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486876>
- Manning, C. D. & Schütze, H., *Foundations of Statistical Natural Language Processing*, MIT Press: 1999.
- Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT.
- Page, D. L. (1992). *Select Papyri*. Loeb classical library. Greek authors. Volume 3. Harvard University Press. ISBN 0674993977, 9780674993976.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., et al. (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*, 12: 2825-2830.
- Shivaprasad Rai, P. (1985). "Sariti: A 2000 Year Old Bilingual Tulu-Greek Play", *International Journal of Dravidian Linguistics*, 14/2: 320-30,
- Rice, E. P. (1929). "On the proposed identification of Kanarese in Oxyrh. Pap. No 413". In Powell, J. U. and Barber, E. A. (eds.), *New Chapters in the History of Greek Literature*, 2: 215-22.
- Salomon, R. (1991). "Epigraphic Remains of Indian Traders in Egypt". *Journal of the American Oriental Society*, 111(4), 731–736. <https://doi.org/10.2307/603404>
- Sastri, R. Shama (1926). "A Greek farce with Old Kannada passages", *Annual Report of the Mysore Archaeological Department*. Bangalore: Mysore Archaeological Department.

- Seldeslachts, E. (1998). “Translated Loans and Loan Translations as Evidence of Graeco-Indian Bilingualism in Antiquity”. *L’Antiquité Classique*, 67: 273–299.
- Shannon, C. E. (1951). “Prediction and entropy of printed English”. *Bell System Technical Journal*, 30(1): 50-64.
- Tsitsiridis, S. (2005). “Greek Mime in the Roman Empire (P.Oxy.413: Charition and Moicheutria)”. *The Classical Quarterly* 55 (2): 435-446, 2005.
- Varadpande, M. L. (1981). *Ancient Indian And Indo-Greek Theatre*. New Delhi: Abhinav Publications.
- Upadhyaya, U. P. (1996). *Coastal Karnataka*. Rashtrakavi Govind Pai Samshodhana Kendra.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Webb, R. (2008). *Demons and Dancers: Performance in Late Antiquity*. Harvard University Press.
- Whitmarsh, T., & Thomson, S. (2013). *The Romance between Greece and the East*. Cambridge University Press.

Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure

Although some stem alternation patterns (aka. 'morphemes') in Romance (N, L, PYTA, see Maiden 2018) have been quite substantially described and analysed, others remain underexplored. Here we focus on a pattern of alternation that involves irregular stem shortening (S) in parts of the paradigm (e.g. It. *d-ire* 'say.INF' [vs *dic-iamo*], or *f-a* 'do.3SG.PRS.IND' [vs *fac-iamo*]). Because these short stems have never been subject to a systematic pan-Romance investigation (but see Malkiel [1977], Maiden [2004:237], Mariño Paz [2019], and Dubert García [2021]), we explore their paradigmatic and lexical domain across Romance, and their likely historical origin and motivation.

An initial qualitative inspection (in Maiden et al.'s 2010 ODRVM database) identified the reflexes of *faciō*, *dīcō*, *habeō*, *sapiō*, *possum*, and *volō* as the verbs that show these alternations most frequently. For these we coded, across the 70+ varieties in the database, the paradigmatic distribution of short stems in the paradigm, which yielded 2773 short-stem forms (18.67%), and 12082 long-stem forms (81.33%). Short stems were found to be most frequent in the cells 3SG.PRS.IND, 2SG.PRS.IND, 3PL.PRS.IND, and 2SG.IMP. A quantitative phylogenetic reconstruction of the ancestral states (i.e. presence of a short or long stem in a particular cell in a particular verb) at different points in time along the family tree found that the likelihood of short stems increases through time quite early in the history of the family. Results suggests that various short stems are likely ancestral to Proto-Western-Romance (i.e. Romance minus Sardinian and Balkan).

We propose an explanation for the timing of the emergence of short stems and their lexical and paradigmatic domain. The first part of the explanation relies on the well-known relation between length of expression and frequency of use (e.g. Zipf 1935, Bybee 2006, Gahl et al. 2012). The cells and verbs in which short stems are most common are all extremely frequent: among the 10 most frequent cells and among the 20 most frequent verbs respectively in Latin (Delatte et al. 1981). Having shorter forms for the expression of very common lexical and morphosyntactic meanings is an adaptive property for the efficient transfer of information, which would provide a motivation to prefer these in situations of competition (note that short stems would have appeared accidentally in the paradigm from regular sound change in some forms like *fa*<*fac* 'do.2SG.IMP' and *di*<*dīc* 'say.2SG.IMP').

The exact domain for the spread of short stems, however, must have been influenced, in addition, by the paradigmatic domains of extant stem alternations. The domain of short stems corresponds closely to those cells that partake in N alternations (i.e. stem-vowel differences related to stress), but not in L alternations (i.e. stem-final consonant alterations resulting from palatalization). N-L is an area of the paradigm, hence, within which a single stem would have been expected, whereas stem differences could easily exist with other parts of the paradigm. This links with the timeline of the better-known Romance morphemes. While PYTA is ancestral to all of Romance (i.e. it was present already in Classical Latin), those known as L and N emerged later. The sound changes that generated them all took place in Western-Romance but not always in the varieties that split before. The morphomic niche for the short-stem allomorphy, thus, is the direct result of the cross-classification of the domains of L and N, which means that it must have followed (and not preceded) the emergence in Romance of the L and N morphemes.

References

- Bybee, Joan. 2006. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. L.A.S.L.A, Liege.
- Dubert García, Francisco. 2021. Las formas reducidas de «facēre» en el noroeste de la península ibérica. Convergencia y divergencia geolingüísticas." *Studia linguistica romanica* 6: 163-191.
- Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language* 66, 4: 789-806.
- Herce, Borja. 2019. Morpheme interactions. *Morphology* 29, 1: 109-132.
- Maiden, Martin. 2004. When lexemes become allomorphs: on the genesis of suppletion. *Folia Linguistica* 38: 227-256.
- Maiden, Martin. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
- Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin & Maria Goldbach. 2010. Oxford online database of Romance verb morphology. Available at: <http://romverbmorph.clp.ox.ac.uk/>.
- Malkiel, Yakov. 1977. Editorial Post-Script: Old Spanish far, fer, fazer. *Romance Philology* 31, no. 2: 257-262.
- Mariño Paz, Ramón 2019. O cambio *dizes* > *dis*, *diz* > *dí*, *dizen* > *din* nas fontes do galego medio. Xosé A. Fernández Salgado, Aquilino S. Alonso Núñez (eds.). *Non haberá illa, pro hai o nome. Homenaxe a Antón Palacio*. Vigo: Universidade de Vigo, 97-113. <https://secretaria.u.vigo.gal/uv/web/publicaciones/public/show/342>.
- Zipf, George Kingsley. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston, MA: Houghton Mifflin Company.

Contact-induced change of Negative Indefinites – the case of Meadow Mari

Since the remarks about the borrowing of negative indefinite pronouns and adverbs (in short: negative indefinites (NI)) in Haspelmath's seminal 1997 book, research on contact-induced change of NI has greatly advanced. Newer research pertains not only to the question how and to what extent markers of NI, i.e., their morphology, can be borrowed, but also how language contact can shape the functional distribution of series of NI (e.g. Elšík & Matras 2006; Karjalainen 2019; Lucas 2013, 2020). In a recent monograph, Breitbarth et al. (2020: Ch. 7) argue, based on van Coetsem's (1988, 2000) model of language contact, that the outcome of contact-induced change of NI depends on the psycholinguistic dominance relations of bilingual speakers in contact situations: if speakers transfer material from an L2 into their dominant language, then transfer of form is usually what is found, whereas in situations of long-lasting bilingualism, the functions of NI may also structurally converge. In this talk I discuss an intriguing case of contact-induced change of NI in Meadow Mari which was likely caused both by direct borrowing of form and structural convergence, combining the two possible developments.

Meadow Mari is a Uralic language spoken by roughly 320,000 people (according to the 2021 All-Russian population census), largely in the Mari El Republic and adjacent areas, as well as in the Ural Mountains. It is situated in the Volga-Kama area, an area of intense historical language contact between Uralic (Mari, Udmurt) and Turkic (Tatar, Chuvash) languages, as well as later intense contact of those languages with Russian. The Uralic and Turkic languages of the area show a large number of lexical, phonological, and morphological convergences which are regarded as outcomes of this contact (e.g. Berezcki 1984; Hesselbäck 2005; Wintschalek 1993). The research of syntactic convergence in the area is still at a nascent stage, however.

NI in Mari, Chuvash, and Russian on the one hand show a morphological parallel; in all three languages, NI are prefixed with *ni-*, a borrowing from Russian in Chuvash and Mari (Egorov 1964; Sibatrova 2021). However, the NI of Mari and Chuvash show additional, structural convergences to the exclusion of Russian which have so far remained unnoticed. For example, in both languages they are licit as standard of comparison (1–2) whereas this is not possible in Russian (3). The data suggests that this structural convergence between Mari and Chuvash arose independently of the borrowing of form from Russian.

- (1) Chuvash
Väl axal'-ten mar [nikam-ran ästa letčik] šutlan-nă.
 3SG simple-ABL NEG nobody-ABL skilled pilot consider-prt.perf
 'Not for no reason was he considered an unsurpassed pilot.' (Chuv.-Rus.-Corpus)
- (2) Meadow Mari
Kö tide saska-m kočk-eš, tudo [nigö deč vijan da patər] lij-eš.
 who this fruit-ACC eat-3SG 3SG nobodyfrom strong and powerful be-3SG
 'He who eats this fruit will be stronger and more powerful than anyone else.' (Mari Corpus)
- (3) Russian
*Zdes' prijatn-ee ži-t' čem *nigde / gde-libo v mire.*
 here comfortable-CMPR live-INF than nowhere where-ever in world.PREP
 'Here it is better to live than anywhere in the world.' (elicited)

In this talk I present ongoing research into the mechanisms of contact-induced change of NI. Based on corpus and elicited language data from Mari, Chuvash, and Russian, showing both convergent and divergent structural patterns of NI, I argue that the present-day distribution of the Mari NI can be attributed to multi-layered language contact with Chuvash and Russian, involving both syntactic convergence and morphological borrowing. Besides sketching the development of the Mari NI series, the talk will also discuss possible challenges that systems of negative indefinites as found in Mari and Chuvash can pose for typologies of negative indefinites.

References

- Bereczki, Gábor. 1984. Die Beziehungen zwischen den finnougriischen und türkischen Sprachen im Wolga-Kama-Gebiet. *Nyelvtudományi Közlemények* 86, 307–314. [The relations between the Finno-Ugric and Turkic languages of the Volga-Kama Area]
- Breitbarth, Anne, Christopher Lucas & David Willis. 2020. *The History of Negation in the Languages of Europe and the Mediterranean. Volume II: Patterns and Processes*. Oxford: Oxford University Press.
- Chuv.-Rus.-Corpus = Dvujazyčnyj korpus čuvašskogo jazyka. <https://ru.corpus.chv.su/> [Bilingual Russian-Chuvash Corpus]
- Egorov, V. G. 1964. *Etimologičeskij slovar' čuvašskogo jazyka*. Čeboksary: Čuvaškoe knižnoe izdatel'stvo. [Etymological dictionary of Chuvash]
- Elšik, Viktor & Yaron Matras. 2006. *Markedness and Language Change. The Romani Sample*. Empirical Approaches to Language Typology 32. Berlin & New York: Mouton de Gruyter.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Hesselbäck, André. 2005. *Tatar and Chuvash Code-copies in Mari*. Uppsala: Uppsala Universitetet.
- Karjalainen, Heini. 2019. Borrowing morphology: The influence of Russian on the Veps system of indefinite pronouns. In: Sofia Björklöf & Santra Jantunen (eds.) *Multilingual Finnic – Language contact and change*. Helsinki: University of Helsinki, 55–87.
- Lucas, Christopher. 2013. Negation in the history of Arabic and Afro-Asiatic. In: Dave Willis, Christopher Lucas & Anne Breitbarth (eds.) *The History of Negation in the Languages of Europe and the Mediterranean. Volume I: Case Studies*. Oxford: Oxford University Press, 399–452.
- Lucas, Christopher. 2020. Contact and the expression of negation. In: Christopher Lucas & Stefano Manfredi (eds.) *Arabic and contact-induced change*. Berlin: Language Science Press, 643–667.
- Mari Corpus = Meadow Mari Corpus of the Giellatekno/Divvun Korp infrastructure: https://gtweb.uit.no/u_korp/?mode=mhr
- Sibatrova, Serafima. 2021. *Marijsko-russkie jazykovye kontakty: morfologija marijskogo jazyka*. Joškar-Ola: MarNIIJaLI. [Mari-Russian language contact: Morphology of Mari]
- Van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.
- Wintschalek, Walter. 1993. *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal*. Wiesbaden: Harrassowitz. [Areal linguistics exemplified by syntactic correspondences in the Volga-Kama Area]

The Rise of Raising in Early Modern English

Modern English distinguishes between two structures, raising and control, that are surface-identical yet, nonetheless, differ structurally in crucial ways. These two structures are exemplified in (1):

- (1) a. David_i wants [PRO_i to like syntax]. [Control]
 b. David_i seems [_{t_i} to like syntax]. [Raising]

In (1a), *want* is a control predicate, which means that the DP *David* does not raise from within the infinitival clause; instead there is a PRO subject. In contrast, *seem* is a raising verb that requires *David* to raise to the specifier of the highest TP. The difference between the two structures is that raising verbs, unlike control verbs, do not assign a theta-role to their external argument. Raising and control represent two subclasses of verbs that select non-finite complements in Modern English. The question is what makes a predicate either raise or control? Many attempts have been made to capture the semantic distinction between the two predicate classes (e.g. Stiebels, 2007). However, there are several confounds, including that some predicates in Modern English vacillate between raising and control complements. Furthermore, the conditions that license either raising or control may vary cross-linguistically and some languages do not even encode the distinction (Polinsky, 2013). Therefore, the question is what conditions license the emergence of such a distinction between raising and control predicates in the first place?

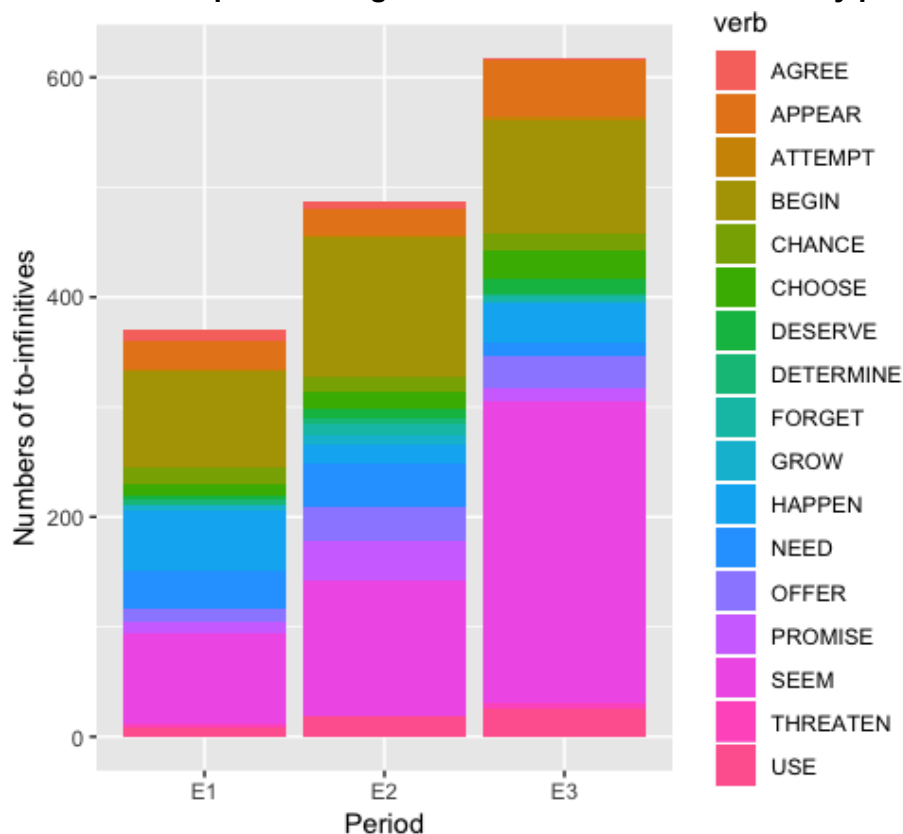
The dramatic change in the history of English complementation has been extensively studied (De Smet, 2013; Los, 2005; Rohdenburg, 2006). However, prior accounts have mostly left out the raising-control distinction. The only existing prior diachronic study is Higgins's (1990) investigation of the development of *promise* and *happen* which led him to conclude that raising verbs emerged from control verbs. Hitherto, however, it has been unclear what triggered the conditions for this change. In this paper, we analyze the distribution of high-frequency raising and control verbs using the Penn-Helsinki Parsed Corpus of Early Modern English (Kroch, Santorini & Delfs, 2016). This period is of particular interest for this distinction since at the time non-finite complements were well established in the grammar and it seems like there were more subtle argument structure distinctions emerging with inevitable variation (e.g. Fanego, 2004).

Our pilot study focuses on the ten most frequent verbs in Early Modern English which (in present-day English) take raising and control complements respectively. The complement-taking predicate was coded by hand based on PPCEME data. Figure 1 (overleaf) shows that there is a massive rise in frequency of to-infinitive complements overall during this period, but that this rise is driven by a very small number of verbs – particularly *seem* and *appear* – which are prototypical raising verbs. This indicates that raising as a structure emerges around or before 1500 and spreads dramatically during the Early Modern period on a lexically specific basis. In our paper we will also investigate the diagnostic structural properties of control and raising verbs (e.g. expletive subjects) and test models of lexically-specific syntactic change based on the Tolerance Principle (Yang, 2016; Irani, 2019).

References

- De Smet, Hendrik. 2013. *Spreading patterns: diffusional change in the English system of complementation*. Oxford: Oxford University Press
- Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change. *Diachronica* 21(1), 5–25.
- Higgins, F. Roger. 1990. By mishap out of control: on the meaningful descent of raising in English. Ms., University of Massachusetts at Amherst.
- Irani, Ava. 2019. *Learning From Positive Evidence: The Acquisition Of Verb Argument Structure*. PhD dissertation, University of Pennsylvania.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2016. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, release 3 (<http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3>).
- Los, Bettelou. 2005. *The rise of the to-infinitive*. Oxford: Oxford University Press.
- Polinsky, Maria. 2013. Raising and control. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 577–606. Cambridge: Cambridge University Press.
- Rohdenburg, Günter. 2006. The role of functional constraints in the evolution of the English complementation system. In Christiane Dalton-Puffer, Nikolaus Ritt, Herbert Schendl & Dieter Kastovsky (eds.), *Syntax, style and grammatical norms: English from 1500–2000*, 143–166. Frankfurt: Lang.
- Stiebels, Barbara. 2007. Towards a typology of complement control. In Barbara Stiebels (ed.), *Studies in Complement Control*, 1–80. ZAS Working Papers in Linguistics 47. Berlin: ZA.
- Yang, Charles. 2016. *The Price of Linguistic Productivity*. MIT Press.

Figure 1: the 10 most frequent raising and control verbs in PPCEME by period



Conservative pressure on the progressive: the passival

Keywords: passive construction, progressive construction, grammaticalization, Early Modern English, functional load, idiolect

This talk discusses the internal reorganization of the progressive construction in Early Modern English (ca. 1500-1700), with special attention for the so-called ‘passival’, which was active in form but passive in meaning (*the house was building* ‘the house was being built’). I will argue that, while the passival was on the increase in Early Modern English, it ‘needed’ to disappear in order to make further progress of the progressive possible. As such, it may be seen as a ‘false turn’ in a grammaticalization process that was eventually resolved by resourceful language users.

While the progressive construction started to increase its scope in the late Middle English period (e.g., Kranich 2010), its grammaticalization was completed only with the introduction of the progressive passive (*the house was being build*) in the late eighteenth century (Hundt 2004: 101). This introduction has been characterized as coming ‘out of the blue’ rather late (Anderwald 2016: 201–202). I will argue that its lateness is due to the passival being the more natural development out of the more nominal origins of the progressive, which originated in part as a prepositional gerundial construction of the type *he was on hunting*. While the preposition gradually eroded, the original nominal characteristics still shimmer through when in Early Modern English the verb’s object was occasionally expressed by *of* NP (e.g. *He was killing of this man* instead of *he was killing this man*) instead of as a direct object. More generally, nominalizations typically do not formally differentiate agent and patient (cf. the famous Latin example *amor matris* where the mother could be both subject or object of the love). In that respect, the passival is a natural functional expansion of the gerundial origin of the progressive. However, as the progressive became more and more verbal, the inclusion of ever more animate and inanimate contexts created competition within the progressive construction, with the passival blocking progress in two different domains. Passivals with human agents, which were occasionally attested, as in (1), had difficulty gaining ground because of obvious confusion with active progressives.

- (1) *The Place where they say the Virgin Mary Prayed for St. Stephen, while he was Stoning.*
 ‘The place where they say the Virgin Mary prayed for St. Stephen, while he was being stoned.’ (Nathaniel Crouch 1672. EMMA)

Second, active progressives with inanimate subjects remained virtually restricted to intransitives and passivals.

Data for the analysis come from the analysis of 15 prolific authors from the EMMA corpus (Petré et al. 2019), collectively constituting a corpus of more than 20 million words. While the unstable nature of the passival has been hinted at in earlier work (e.g., Hundt 2004), the EMMA corpus allows for a more fine-grained and also individual-level analysis. Such an analysis shows that the functional block of the passival is a consistent cognitive reality across most (but not all) authors. However, where this cognitive dissonance within the progressive construction remained inconspicuous in early generations, it came more and more to the fore in later generations as the passival became more prevalent, eventually leading to its demise in favour of the progressive passive, in spite of the latter’s semi-artificial introduction by professional writers.

References

Anderwald, Lieselotte 2016. *Language between Description and Prescription: Verb Categories in Nineteenth-century Grammars of English*. Oxford: Oxford University Press.

- Kranich, Svenja. 2010. The progressive in Modern English: A corpus-based study of grammaticalization and related changes. *The Progressive in Modern English*. Leiden: Brill.
- Hundt, Marianne 2004. The passival and the progressive passive – a case study of layering in the English aspect and voice systems. In Hans Lindquist and Christian Mair (eds.), *Corpus Approaches to Grammaticalisation in English*. Amsterdam: Benjamins, pp. 79–117.
- Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing and Odile A.O. Strik 2019. Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME journal* 43: 83–122.

Another look at Noun-Genitive vs. Genitive-Noun in Early New High German

It has long been recognized that historical varieties of German exhibit variation between prenominal vs. postnominal genitives (GN vs. NG orders), with an overall development away from GN and toward NG under various conditions (Behagel 1923). In broad strokes, retention of GN order is associated with possessive genitives (as opposed to subjective, objective, partitive, or explicative genitives), genitives denoting persons, and single-words, proper names, or pronouns. By 1700, near the beginning of the Modern German era, GN order accounts for only 10% of adnominal genitives, and GN becomes restricted to proper nouns by 1750 (Niehaus 2016). In Early New High German (ENHG; 1350-1650) there is often variation between GN and NG within these conditions even within individual texts. Despite several examinations of this variation and change (e.g., Ebert 1988; Lunt Lanouette 1990, 1998; Pickl 2020), this topic remains surprisingly under-researched.

In this paper, we explore the hypothesis that in at least some varieties of ENHG, the grammar of adnominal genitives includes two properties not found in Modern German:

- (1) Adnominal genitives are generated in the Specifier position of the NP.
- (2) NG order is derived by an optional rule that extraposes the genitive DP or a subpart of it and which is sensitive to the length/weight of the genitive phrase.

Property (1) accounts for examples like (3), in which pre-nominal genitives appear to the right of attributive adjectives, and property (2) accounts for instances in which the pre-nominal genitive is modified by a post-nominal phrase.

- (3) *eyn besunder* [_{DP} *Rulands*] *streitgesel* (Fierrabras 196 [1533])
a certain Ruland-GEN battle-companion ‘a certain combatant of Ruland’
- (4) [_{DP} *Josephs* _{TPP}] *sun* [_{PP} *von aramathea*] (Karrenritter 472 [1430])
Joseph-GEN son of Aramathea ‘Joseph of Aramathea’s son’

Taken together, these properties yield the tendency for NG order with longer genitive DPs, but GN with one-word genitives, as well as the split construction illustrated in (4).

We present data from an ongoing corpus-based study of GN vs. NG variation in ENHG. The corpus will ultimately consist of 60 texts, with one text per 50-year bin from 10 dialects, representing a variety of genres. Texts in the corpus are constituency parsed according to the Penn annotation system (e.g., Kroch 2020). Preliminary data have been extracted from the first four completely parsed texts, yielding the following results thus far:

- By text: GN varies widely from 84% in *Karrenritter* (1430) to 48% in *Fierrabras* (1533) to just 7% in *Geistlicher Mai* (1529). More texts need to be analyzed to tease apart the effects of time, dialect, and genre.
- Proper vs. common noun genitive: In texts other than *Geistlicher Mai*, proper nouns appear in GN order more frequently (67-83%) than common nouns.
- Genitive type: for both GN and NG, the vast majority of adnominal genitives are possessives, thus the effect of this factor is inconclusive so far.
- Length: One-word genitives most strongly favor GN (67% excluding *Geistlicher Mai*) and longer genitives increasingly disfavor it (down to 19% for five or more words, again excluding *Geistlicher Mai*.)

We suggest that the Modern Standard German pattern in which only personal possessives in *-s* (which are probably no longer genitives, see Fuß 2011) and possessive adjectives occur prenominally and full genitive DPs occur only in the order NG, results from a reanalysis triggered by the increased application of the extraposition rule in the late ENHG period.

References

- Behaghel, Otto. 1923. *Deutsch Syntax: Eine geschichtliche Darstellung*. Volume 4. Heidelberg: Carl Winter.
- Ebert, Robert Peter. 1988. Variation in the position of the attributive genitive in sixteenth century German. *Monatshefte* 80: 32-49.
- Fuß, Eric. 2011. Eigennamen und adnominaler Genitiv im Deutschen. *Linguistische Berichte* 225: 19-42.
- Kroch, Anthony. 2020. *The Penn Parsed Corpora of Historical English*. University of Pennsylvania. <https://www.ling.upenn.edu/hist-corpora/>
- Lunt Lanouette, Ruth. 1990. *Prenominal specifiers in the Early New High German Noun Phrase*. Unpublished dissertation, Princeton University.
- Lund Lanouette, Ruth. 1998. The attributive genitive in the history of German. *Germanic linguistics: Syntactic and diachronic*, ed. by Rosina L. Lippi-Green and Joseph Salmons, 85-102. Philadelphia: John Benjamins.
- Niehaus, Konstantin. 2016. *Wortstellungsvarianten im Schriftdeutschen: Über Kontinuitäten und Diskontinuitäten in neuhochdeutscher Syntax*. Heidelberg: Winter.
- Pickl, Simon. 2020. Polarization and the emergence of a written marker: A diachronic corpus study of the adnominal genitive in German. *Journal of Germanic Linguistics* 32: 145-182.

Latin *placēre* as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis

Traditionally, the Latin verb *placēre* ‘like, please’, which is a two-place predicate licensing a nominative and a dative argument, is analyzed as a Nom-Dat verb, which means that it is the nominative that is considered to be the subject, whereas the dative is assumed to be the object. However, examples like the ones in (1) below show that either order, Dat-Nom as in (1a) and Nom-Dat as in (1b), is acceptable in Latin. The fact that both word order patterns are equally fine has been observed in the field, but it has generally been attributed to what is termed ‘free word order’ (Devine & Stephens 2006, Spevak 2010, inter alia).

- (1) a. (Ov. *Tr.* 4,10,19)
 At **mihi iam puero caelestia sacra** placebant
 and I.DAT even boy.DAT mystic.NOM service.NOM like.IMPF.3PL
 “And I, even as a boy, liked the mystic services”
- b. (Cic. *Orat.* 2, 42, 179)
Qui ordo tibi placeat inquit Catulus
 what.NOM arrangement.NOM you.DAT like.SBJV.PRS.3SG say.PRS.3SG Catulus.NOM
 “What arrangement would please you, said Catulus [...]”

The present paper advocates a more radical approach, namely in terms of *alternating predicates*. Such structures have also been shown to exist in Germanic (Barnes 1986 for Faroese, Allen 1995 for Old English, Barðdal 1998 for the history of the Mainland Scandinavian languages, Barðdal 2001 for Modern Icelandic, Barðdal, Eythórsson & Dewey 2019 for Modern German) and in Romance (Illoaia 2022 for Romanian), and they may also exist in Baltic, Slavic, Hittite and Sanskrit (cf. Barðdal 2023: Ch. 3). Alternating predicates systematically occur with two diametrically opposed argument structures: a Dat-Nom argument structure and a Nom-Dat argument structure. As a consequence, (1a) would contain a dative subject and a nominative object, whereas the opposite is true for (1b), which would contain a nominative subject and a dative object.

Here we focus on one specific verb in Latin, *placēre* ‘like, please’, confining our analysis to occurrences of this verb as a two-place predicate, which licenses a nominative and a dative argument. The data are drawn from the *LatinISE* corpus, which stretches a period of approximately 500 years, from the Archaic Period (3rd–2nd century BCE) up to the Late Latin period (7th century). Our dataset comprises 350 occurrences of *placēre*, annotated for (pro)nominativity, person, definiteness, length, and animacy.

In line with Eythórsson & Barðdal (2005), Barðdal & Eythórsson (2012, 2018), and Barðdal (2023), we define subject as the leftmost argument of the argument structure. This definition is based on a generalization across a range of diagnostics which have been successfully applied to various Germanic languages and the behavior of the arguments relative to these. Unfortunately, research on the modern linguistic concept of subject is still in its early stages within the Latin scholarship. The issue was first dealt with by Michaelis (1992), later to be taken up by Baños Baños (2003) and Fedriani (2009, 2014).

More recently, Barðdal et al. (2023) have shown that several subject tests identified for the Germanic languages may be successfully applied to Latin and Ancient Greek. On this basis, we discuss the behavior of the two arguments of *placēre* with regard to these subject tests and show that either argument, the nominative or the dative, passes the subject tests in Latin. We focus in particular on data involving word order, raising-to-object, raising-to-subject, and control infinitives.

Barðdal et al. (2023) further document that ordinary nominative subjects in Latin precede the object in ca. 70% of the cases, thereby establishing a baseline against which to compare the statistics obtained for alternating predicates. We compare our Latin word order statistics with corresponding statistics from Old English (Allen 1995) and Old Norse-Icelandic (Elens, Somers & Barðdal 2023), arguing that Latin *placēre* ‘like, please’, shows the same distributional properties as alternating predicates in the Early Germanic languages.

References

- Allen, Cynthia L. 1995. *Case Marking and Reanalysis: Grammatical Relations from Old to Early Modern English*. Oxford: Oxford University Press.
- Baños Baños, José Miguel, 2003. *Paenitet* y los verbos impersonales de sentimiento en latín: sintaxis y pragmática del acusativo personal [*Paenitet* and the impersonal verbs of feeling in Latin: syntax and pragmatics of the personal accusative]. *Praedicatoria: Estudios sobre complementación verbal en griego y latín*, ed. by José Miguel Baños Baños, Concepción Cabrillana, María Esperanza Torrego & Jesús de la Villa, 51–77. Santiago de Compostela: Universidad de Santiago de Compostela.
- Barðdal, Jóhanna. 1998. Argument structure, syntactic structure and morphological case of the impersonal construction in the history of Scandinavian. *Scripta Islandica* 49: 21–33.
- Barðdal, Jóhanna. 2001. The perplexity of Dat-Nom verbs in Icelandic. *Nordic Journal of Linguistics* 24: 47–70.
- Barðdal, Jóhanna. 2023. *Oblique Subjects in Germanic: Their Status, History and Reconstruction*. To appear with Mouton de Gruyter.
- Barðdal, Jóhanna, Eleonora Cattafi, Leonardo Biondo, Laura Bruno & Serena Danesi. 2023. Non-nominative subjects in Latin and Ancient Greek: Applying the subject tests on Early Indo-European languages. *Indogermanische Forschungen* 128.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. “Hungering and lusting for women and fleshly delicacies”: Reconstructing grammatical relations for Proto-Germanic. *Transactions of the Philological Society* 110(3): 363–393.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2018. What is a subject? The nature and validity of subject tests. *Non-Canonically Case-Marked Subjects: The Reykjavík–Eyjafjallajökull Papers*, ed. by Jóhanna Barðdal, Na’ama Pat-El & Stephen Mark Carey, 257–273. Amsterdam: John Benjamins.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1): 107–170.
- Barnes, Michael. 1986. Subject, nominative and oblique case in Faroese. *Scripta Islandica* 37: 13–46.
- Devine, A. M., & Laurence D. Stephens. 2006. *Latin Word Order: Structured Meaning and Information*. Oxford: Oxford University Press.
- Elens, Wannes, Joren Somers & Jóhanna Barðdal. 2023. The alternating behaviour of ‘like’ in Old Norse-Icelandic: Facts or Fiction. Ms. Ghent University.
- Eythórsson, Thórhallur & Jóhanna Barðdal. 2005. Oblique subjects: A common Germanic Inheritance. *Language* 81(4): 824–881.
- Fedriani, Chiara. 2009. The “Behaviour-Before-Coding-Principle”: Further Evidence from Latin. *Archivio Glottologico Italiano* XCIV: 156–184
- Fedriani, Chiara. 2014. *Experiential Constructions in Latin*. Leiden: Brill.
- Ilioiaia, Mihaela. 2022. Toward canonical marking of core arguments in Romanian? The case of *plăcea*. A presentation delivered at ICHL 25, Oxford.
- Michaelis, Laura A. 1992. On Deviant Case-Marking in Latin. *Advances in Role and Reference Grammar*, ed. by Robert D. Van Valin, 311–373. Amsterdam: John Benjamins.
- Spevak, Olga. 2010. *Constituent Order in Classical Latin Prose*. Amsterdam: John Benjamins.

The diachronic development of future markers in Chinese

Linguistic analyses of future marking distinguish two different viewpoints: 1) an objectivistic view, i.e., the *branching future* concept, according to which the future is a kind of modality (Portner 2009: 236), and 2) a subjective perspective, in which future tense simply refers to a time following speech time (Reichenbach 1947, Bochnak 2019). Within the cartographic approach, future tense is hosted in a functional projection separate from the projections hosting different kinds of modality.

- (1) ModP_{epist}>TP(Past)>TP(fut)>AspP_{habit}>ModP_{volition}>AspP...>ModP_{obligation}>ModP_{permission/ability} (modified from Cinque 2004)

Chinese does not have morphological tense marking; temporal and aspectual distinctions are expressed analytically. However, future is the most regularly expressed temporal (and/or aspectual) concept in Archaic Chinese. Future markers by default appear in complementary distribution with other aspecto-temporal markers, and in Archaic Chinese they permit a purely temporal reading. This is shown in example (2) with the future marker *jiāng*, which is semantically similar to the Pre- and Early Archaic future marker *qí* 其, frequently appearing in oracular predictions (Djamouri 2009).

- (2) 公 將 以 某 日 薨 (Lüshi chunqiu LAC/EMC)
 gōng jiāng yǐ mǒu rì hōng
 duke FUT YI such.and.such day pass.away
 ‘The duke will pass away on such-and-such day.’

In Early Middle Chinese, new future markers grammaticalize from the lexical verbs *dāng* 當 ‘correspond to’, which first develops into a deontic auxiliary ‘should’, and from the volitional verb *yù* 欲. Both markers include modal, besides their temporal readings. Similar to *jiāng*, they appear in complementary distribution with other aspecto-temporal markers in the TP layer. When they appear in combination with future *jiāng*, DANG and YU either have to be analyzed as pre-modal verbs, or the combination functions as a disyllabic future marker.

- (3) a. 若干 百年 當 至 于 闐 國。
 Ruògān bǎi nián dāng zhì yútiān guó
 Several hundred year DANG arrive Khotanstate
 ‘After several hundred years it will/should arrive in Khotan.’
 b. 『欲云何作?』
 yù yúhézuò
 YU how do
 ‘How will you (do you want to) do it?’

The only syntactic difference between EMC future DANG and YU, and LAC *jiāng* is the position of negation. Since DANG and YU are verbal heads, NEG has to precede them, but it has to follow the aspecto-temporal adverb *jiāng*. Meisterernst (2020) proposed two different functional projections hosting future tense and deontic modality for LAC and EMC. Contrastingly, we propose one unified zone within TP (following Ramchand and Svenonius 2014), which can be targeted by either a root modal necessity marker or a by future marker; epistemic necessity is hosted in a higher projection. In LAC, this zone could be occupied by either a modal negator or a future marker; modal verbs were confined to the lexical, i.e., the *vP* layer. The situation changes in EMC, when true deontic modals emerged, which were hosted in the TP layer in the same zone as modal negators and future markers in LAC.

- (4) [CP ModP_{epist} [TP TPzone FUT/ASP/NEG_{Mod}/MOD_{deont} [_{vP} VPzone (MOD_{circum}) vP]]]

Similar to what Ramchand and Svenonius propose, the markers in the *vP* external zone are characterized by a relative independence of the aktionsart feature of the *vP*.

References

Bochnak, Ryan M. 2019. Future reference with and without future marking. *Language and Linguistics Compass* 13,1. <https://doi.org/10.1111/lnc3.12307>

- Djamouri, Redouane. 2009b. Jiaguwen zhong 'qi' de yongfa 甲骨文中“其”的用法 [Use of *qi* in the Oracle Bone Inscriptions]. In Feng Li & Yang Yonglong (eds.), *Hanyu shiti wenti lishi yanjiu* 漢語時体問題歷史研究 [Diachronic Studies on Tense and Aspect in Chinese], Beijing: Yuwen chubanshe.
- Meisterernst, Barbara. 2020. A syntactic approach to the grammaticalization of the modal marker dāng 當 in Middle Chinese, *Journal of Historical Syntax*, 4,5. 2020-05-29.
- Portner, Paul (2009). *Modality*. Oxford: OUP
- Ramchand, Gillian and Svenonius, Peter. 2014. Deriving the Functional Hierarchy. *Language Sciences* 46: 152-174.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Dover Publication.

Properties of Complex Compounds in Old Japanese

This project investigates complex compounds in Old Japanese (OJ), the language of 8th century Japan. Properties of compounds are well studied for modern Japanese, but less so for earlier stages of the language, and this project will add to our understanding of how the Japanese language developed over time.

Mithun (1984) describes noun incorporation as a syntactic morphological process. In her framework there are 4 types of noun incorporation. This framework is hierarchical:

- 1) lexical compounding where the noun is assigned the role of patient, location, or instrument by the incorporating verb (IV)
- 2) the manipulation of case, where the incorporated noun (IN) loses its argument status and another noun receives it
- 3) the manipulation of discourse structure, where the IN is part of information structure representing known (and not new) information
- 4) classificatory noun incorporation, where the IN narrows the scope of the IV but can be accompanied by an NP which classifies the IN

OJ shows evidence of types 1 and 4.

In addition, Rosen (1989) claims there are 2 separate word formation processes.

- 1) argument structure of the IV changes, so that the verb loses an argument
- 2) argument structure is unaltered

Rosen claims that a language may have one or the other process, but not both. OJ, however, shows both processes.

The data for the initial study were extracted from the Oxford Corpus of Old Japanese (OCOJ), a linguistically annotated corpus of the language of 8th century Japan, the earliest attested stage of the Japanese language. The data were then further annotated with information about orthography, verbal inflection, syntactic structure, whether the noun adjacent to a verb was bound or free, whether *rendaku* (sequential voicing) occurred, and whether the incorporated noun was modified.

Criteria for determining whether or not nouns are incorporated in OJ are 1) positive evidence for incorporation; 2) positive evidence against incorporation; and 3) absence of evidence against incorporation.

Unlike Modern Japanese, the data show that OJ has both noun and phrasal incorporation; incorporated nouns in OJ can be modified by adjectives, possessives, and relative clauses. A variety of semantic roles can be assigned to the incorporated noun, including both arguments and adjuncts. Subjects of intransitive verbs can be incorporated, and, finally, incorporated structures are not always nominalized and do not always occur in light verb constructions.

- Frellesvig, Bjarke. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press.
- Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, & Peter Sells. n.d. The Oxford Corpus of Old Japanese. <http://vsarpj.orinst.ox.ac.uk/corpus/index.html>
- Kinuhata, Tomohide. 2010. 上代語の名詞抱合について [On incorporated nouns in Old Japanese]. 語文, 92/93, 34-44.
- Kageyama, Taro. 1982. "Word formation in Japanese". *Lingua* 57, 212-258.
- Martin, Samuel E. 1987. *The Japanese language through time*. New Haven: Yale University Press.
- Mithun, Marianne. 1984. "The evolution of noun incorporation". *Language* 60, 847-894.
- Omodaka Hisataka, editor. 1967. 時代別国語大辞典：上代編 [A dictionary of the Japanese language by periods: Old Japanese volume]. Tokyo: Sanseidō.
- Rosen, Sara Thomas. 1989. "Two type of noun incorporation: A lexical analysis." *Language* 65, 2, 294-317.
- Russell, Kerri L. 2012. "Noun incorporation in Old Japanese (上代日本語における名詞抱合について)" presented at 通時コーパスプロジェクト・オックスフォード大VSARPJプロジェクト合同シンポジウム(NINJAL Diachronic Corpus Project – Oxford VSARPJ Project Joint Symposium Corpus Based Studies of Japanese Language History).
- Russell, Kerri L., John Whitman, and Yuko Yanagida. 2014. "Was Old Japanese a polysynthetic language?" presented at International Symposium on Polysynthesis in the World's Languages 国際シンポジウム「世界の言語における複統合性」National Institute for Japanese Language and Linguistics, 20-21 February, 2014
- Shibatani, M. and T. Kageyama. 1988. Word formation in a modular theory of grammar: Post-syntactic compounds in Japanese. *Language*, 64(3):451-484.
- Stekauer, Pavol, Salvador Valera, and Lívia Kortvélyess. 2012. *Word-formation in the World's languages: A typological survey*. Cambridge: Cambridge University Press.

The Alternating Behavior of ‘Like’ in Old Norse-Icelandic: Facts or Fiction

In a recent article, Sigurðsson & Viðarsson (2020) claim that the verb *líka* ‘like’ in Old Norse-Icelandic is an alternating Dat-Nom/Nom-Dat verb, as opposed to Modern Icelandic where this same verb is uncontroversially a non-alternating Dat-Nom verb. The difference between the two is that alternating verbs may instantiate two diametrically opposite argument structures, i.e. Dat-Nom and Nom-Dat, while non-alternating Dat-Nom verbs only instantiate one of these, the Dat-Nom argument structure (Bernóusson 1982, Barnes 1986, Jónsson 1997–98, Barðdal 1999, 2001, Barðdal, Eythórsson & Dewey 2014, 2019, Platzack 1999, Wood & Sigurðsson 2014, Somers & Barðdal 2022, inter alia).

Sigurðsson & Viðarsson (2020) base their claims on both language internal Old Norse-Icelandic evidence and on the comparative evidence from the other early Germanic languages (Fischer & van der Leek 1983, Allen 1986, 1995, Barðdal 1998, Eythórsson & Barðdal 2005). Starting with the comparative evidence, it has been argued for Old English that *lician* is an alternating verb in that language (Allen 1995: 141) and the same has been argued for *galeikan* in Gothic (Eythórsson & Barðdal 2005: 833).

Turning to the language internal evidence for an alternating analysis of *líka* in Old Norse-Icelandic which Sigurðsson & Viðarsson introduce, this consists of data involving two subject tests, i) control infinitives and ii) word order. Sigurðsson & Viðarsson present several examples of control infinitives with the verb *líka* where it is indeed the nominative and not the dative that is left unexpressed in such structures. This they take as conclusive evidence that the nominative behaves syntactically as a subject and the dative as an object. Likewise, Sigurðsson & Viðarsson also present a handful of examples involving word order distribution, which are incompatible with a Dat-Nom analysis of the argument structure of *líka*, and call instead for a Nom-Dat analysis of the relevant structures.

While we agree with Sigurðsson & Viðarsson on their analysis of the relevant control infinitives, in that there is no doubt that these examples show that it is indeed the nominative that is left unexpressed and not the dative, we still call into question the relevance of their data set. As they acknowledge themselves, all their examples of control infinitives are from translated texts, and the same applies to their word order examples. In general, translated examples may well be taken to speak for authenticity, but for these particular examples, we argue that the relevant translations are word-for-word glosses of the Latin verb *placere* which sometimes means ‘like’ and sometimes ‘please’, depending on its argument structure (cf. Cluyse, Somers & Barðdal 2023).

Therefore, in order to shed light on this issue, we present word order statistics for *líka* in Old Norse-Icelandic. The data have been extracted from three different sources: i) the Saga Corpus, ii) the Icelandic Text Archive, and iii) the Dictionary of Old Norse Prose (ONP), resulting in a dataset containing approximately 200 occurrences of *líka*. It turns out that there is a major divide between native and translated texts. This means that in texts originally written in the Old Norse-Icelandic vernacular, *líka* consistently occurs with a Dat-Nom order, unless the nominative contains a demonstrative pronoun, then the Nom-Dat word order is preferred. This suggests that *líka* could only instantiate the Dat-Nom argument structure construction in texts originally written in Old Norse-Icelandic, with the Nom-Dat word order representing topicalizations. In contrast, in the translated texts, the proportions between Dat-Nom and Nom-Dat with *líka* are more even and are not tied to word class. Therefore, since any “alternating” behavior of *líka* is confined to translated texts, we conclude that this seeming behavior is a translation effect.

Our alternative analysis of the data involving Old Norse-Icelandic *líka* above makes a certain prediction, namely that the existing alternating analysis of Old English *lician* and Gothic *galeikan* may be equally faulty as the analysis provided by Sigurðsson & Viðarsson, as most if not all instances in Old English and Gothic are also translations. In other words, our analysis predicts that the apparent alternating behavior of ‘like’ in Old English and Gothic may also be a translation effect.

References:

- Allen, Cynthia L. 1986. Reconsidering the history of *like*. *Journal of Linguistics* 22. 375–409.
- Allen, Cynthia L. 1995. *Case Marking and Reanalysis: Grammatical Relations from Old to Early Modern English*. Oxford: Oxford University Press.
- Barðdal, Jóhanna. 1998. Argument structure, syntactic structure and morphological case of the impersonal construction in the history of Scandinavian. *Scripta Islandica* 49. 21–33.
- Barðdal, Jóhanna. 1999. The dual nature of Icelandic psych-verbs. *Working Papers in Scandinavian Syntax* 64. 78–101.
- Barðdal, Jóhanna. 2001. The perplexity of Dat-Nom verbs in Icelandic. *Nordic Journal of Linguistics* 24. 47–70.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2014. Alternating predicates in Icelandic and German: A sign-based construction grammar account. *Working Papers in Scandinavian Syntax* 93. 50–101.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1). 107–170.
- Barnes, Michael. 1986. Subject, nominative and oblique case in Faroese. *Scripta Islandica* 37. 13–46.
- Bernódusson, Helgi. 1982. Ópersónulegar setningar [Impersonal sentences]. University of Iceland Master's Thesis.
- Cluyse, Brian, Joren Somers & Jóhanna Barðdal. 2023. Latin *placere* as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis. Ms., Ghent University.
- Eythórsson, Thórhallur & Jóhanna Barðdal. 2005. Oblique subjects: A common Germanic Inheritance. *Language* 81(4). 824–881.
- Fischer, Olga C. M. & Frederike C. van der Leek. 1983. The demise of the Old English impersonal construction. *Journal of Linguistics* 19. 337–368.
- Icelandic Text Archive = Stofnun Árna Magnússonar. (n.d.) Leit. In *Íslenskt Textasafn*. <http://corpus.arnastofnun.is/leit.pl>.
- Jónsson, Jóhannes Gísli. 1997–98. Sagnir með aukafallsfrumlagi [Verbs selecting for oblique subjects]. *Íslenskt mál* 19–20. 11–43.
- ONP = University of Copenhagen. (n.d.). Líka vb. In M. Arvidsson, S. Battista, H. Degnbol, P. Ellyton, Þ. Helgadóttir, B. C. Jacobsen, E. Þ. Jóhannsson, A. S. Kjeldsen, J. E. Knirk, J. Lindholm, E. Rode, C. Sanders, A. Sigurðardóttir & T. Wills (Eds.), *ONP: Dictionary of Old Norse Prose*. <https://onp.ku.dk/onp/onp.php?o98918>.
- Platzack, Christer. 1999. The subject of Icelandic psych-verbs: A minimalistic account. *Working Papers in Scandinavian Syntax* 64. 103–116.
- Wood, Jim & Halldór Ármann Sigurðsson. 2014. *Let*-causatives and (a)symmetric DAT-NOM constructions. *Syntax* 17(3). 269–298.
- Saga Corpus = Stofnun Árna Magnússonar. (n.d.). Líka. In *Saga Corpus*. <https://malheildir.arnastofnun.is/?mode=fornrit#?lang=en>.
- Sigurðsson, Einar Freyr & Heimir van der Feest Viðarsson. 2020. Frá skiptisögn til ósamhverfrar aukafallssagnar: Um líka í fornu máli [From an alternating to a non-alternating oblique subject verb: About *líka* 'like' in Old Icelandic]. *Orð og tunga* 22. 39–68.
- Somers, Joren & Jóhanna Barðdal. 2022. Alternating Dat-Nom/Nom-Dat verbs in Icelandic: An exploratory corpus-based analysis. *Working Papers in Scandinavian Syntax* 107. 83–110.

The rise of *do*-support during Scots anglicisation: Insights from the *Parsed Corpus of Scottish Correspondence*

Lisa Gotthard, University of Edinburgh

This paper explores the rise of *do*-support in Scots, and investigates whether it exhibited similar functions to English ‘intermediate’ *do* (as analysed by Ecay (2015)) before regularising into its current function. In doing so, this study provides more insight into syntactic change in Scots during the period of *anglicisation*, starting in the 16th century, when English became favoured over Scots in writing. While *anglicisation* is usually discussed in terms of lexis and orthography (e.g. Devitt (1989); Meurman-Solin (1993b); Kniezsa (1997)), the 16th to 18th century has been obscure within Scots syntax research; aided by the new *Parsed Corpus of Scottish Correspondence* (PCSC; Gotthard 2022), this gap now begins to be filled.

Do-support is the mandatory insertion of the auxiliary *do*, which has historically been bleached of its semantic meaning and today has a strictly morpho-syntactic function: carrying tense and agreement features when the transfer of these features between the subject and main verb is interrupted and no other auxiliary is present (E.g., *I do not eat cake*). English *do*-support is extensively researched (e.g., Ellegård 1953; Denison 1985; Ecay 2015; Garrett 1998; Kroch 1989; Nurmi 2011; Poussa 1990; Tieken-Boon van Ostade 1990; van der Auwera and Genee 2002; Warner 2002), but the feature in Scots has received less attention. It has been suggested that Scots *do*-support is a transfer from English, supported by that the feature emerged in Scots during the height of anglicisation (Gotthard 2019, 2022, 2023; Meurman-Solin 1993a), and that it remains variable in more traditional dialects (Jamieson 2015; Jonas 2002; Smith 2000), but it could also have developed independently, from the causative *do* auxiliary inherited from Old English. Based on observations already made by Ellegård (1953), Ecay (2015) identifies an ‘intermediate’ *do* auxiliary in English pre-1575; an agentive marker which merges in a lower syntactic position than post-1575 *do*. This leads to the question of whether Scots *do* follows the same path, or is adopted with the same function as post-1575 English *do* – if a candidate feature emerges in the receiving language in its fully grammaticalised form, then it is more likely to be a transferred feature (e.g. Pa-Tel 2013).

In order to investigate (i) the emergence and trajectory of *do* in Scots, and (ii) whether Scots *do*-support is a plausible outcome of anglicisation, proportions of affirmative and negative declarative *do* in the PCSC are calculated across different syntactic contexts. The likelihood of the feature being an anglicisation outcome is evaluated by assessing the results against criteria for contact-induced change (e.g., Thomason and Kaufman 1988; Thomason 2001; Pa-Tel 2013; Robbeets and Cuyckens 2013; Poplack and Levey 2010). It is found that Scots *do* emerges towards the end of the 16th century, and remains at low proportions (around 20%) until ca. 1700 when it increases more dramatically, and initially behaves largely consistent with what Ecay (2015) observed for ‘intermediate’ *do* in English. The social context and timing of the rise of Scots *do* suggest that it is a contact-induced change, but the fact that the auxiliary shows ‘intermediate’ *do* qualities warrants further discussion; if this is truly an intermediate stage in the grammaticalisation of *do* then the analysis is compromised, but this *do* auxiliary may also be as different type of *do*, spreading northward from English into Scots.

References

- Denison, D. (1985). The origins of periphrastic do: Ellegård and visser reconsidered. In Eaton, R., Fischer, O., Koopman, W. F., and van der Leek, F., editors, *Papers from the 4th international conference on English historical linguistics* 4th International Conference on English Historical Linguistics, pages 45–60. Amsterdam/Philadelphia: John Benjamins Publishing.
- Devitt, A. J. (1989). *Standardizing written English: diffusion in the case of Scotland, 1520-1659*. Cambridge: Cambridge University Press.
- Ecay, A. (2015). *A multi-step analysis of the evolution of English do-support*. PhD thesis, University of Pennsylvania.
- Ellegård, A. (1953). *The auxiliary do: The establishment and regulation of its use in English*. Stockholm: Almqvist and Wiksell.
- Garrett, A. (1998). On the origin of auxiliary do. *English Language and Linguistics*, 2:283–330.
- Gotthard, L. (2019). Why do-support in Scots is different. *English Studies*, 100(3):1–25.
- Gotthard, L. (2022). *Syntactic change during the anglicisation of Scots: Insights from the Parsed Corpus of Scottish Correspondence*. PhD thesis, University of Edinburgh.
- Gotthard, L. (2023). Subject-verb agreement and the rise of do-support during the period of anglicisation of Scots. To appear in proceedings of ICEHL21, Leiden 2021.
- Jamieson, E. (2015). An investigation of verb raising in the Shetland dialect of Scots. Master's thesis, University of Edinburgh.
- Jonas, D. (2002). Residual V-to-I. In Lighfoot, D., editor, *Syntactic effects of morphological change*, pages 251–270. Oxford: Oxford University Press.
- Knieszsa, V. (1997). The origins of scots orthography. In Jones, C., editor, *The Edinburgh History of the Scots Language*, pages 24–46. Edinburgh: Edinburgh University Press.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1:199–244.
- Meurman-Solin, A. (1993a). Periphrastic do in sixteenth- and seventeenth century Scots. In Meurman-Solin, A., editor, *Variation and change in early Scottish prose*, pages 259–276. Helsinki: Suomalainen Tiedeakatemia.
- Meurman-Solin, A. (1993b). *Variation and change in early Scottish prose*. Helsinki: Suomalainen Tiedeakatemia.

- Nurmi, A. (2011). The rise and regulation of periphrastic do in negative declarative sentences: a sociolinguistic study. In Kastovsky, D. and Mettinger, A., editors, *The History of English in a Social Context: A Contribution to Historical Sociolinguistics*, pages 339–362. New York: de Gruyter Mouton.
- Pa-Tel, N. (2013). Contact or inheritance? criteria for distinguishing internal and external change in genetically related language. *Journal of Language Contact*, 6(2):313–28.
- Poplack, S. and Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In *Language and Space – An international handbook of linguistic variation: Volume 1 – Theories and methods*. Berlin: Mouton de Gruyter.
- Poussa, P. (1990). A contact-universals origin for periphrastic do with special consideration of old english-celtic contact. In Adamson, S., Law, V., Vincent, N., and Wright, S., editors, *Papers from the 5th international conference on English historical linguistics*, volume 65. Amsterdam/Philadelphia: John Benjamins Publishing.
- Robbeets, M. and Cuyckens, H. (2013). *Shared Grammaticalization: With Special Focus on Transeurasian Languages*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Smith, J. (2000). 'you ø na hear o' that kind o' things': Negative do in Buckie Scots. *English World Wide*, 21(2):231–259.
- Thomason, S. G. (2001). *Language Contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, S. G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tieken-Boon van Ostade, I. (1990). The origin and development of periphrastic auxiliary do: A case of destigmatisation. *NOWELE. North-Western European Language Evolution*, 16(1).
- van der Auwera, J. and Genee, I. (2002). English do: on the convergence of languages and linguists. *English Language and Linguistics*, 6:283–307.
- Warner, A. (2002). What drove do? In Kay, C., Horobin, S., and Smith, J., editors, *New perspectives on English Historical Linguistics*, volume I, pages 229–242. Amsterdam/Philadelphia: John Benjamins Publishing.

Exploring language variation and change in the distant past.

The modelling of ‘prehistorical’ language features through comparative analyses and reconstructions offers a unique insight into language dynamics in the distant past (e.g., Dunn et al. 2011; Carling & Cathcart 2021). However, to analyse the particulars of the process of language change in its socio-historical embeddings, we are constrained to analyse written documentation. But not all records are suitable for this endeavour; historical sociolinguistic research has mostly focused on well-attested post-medieval European languages (e.g., Nevalainen & Raumolin-Brunberg 2003), and could be expected to find major limitations for ancient languages (Winter 1999). Then, how far back in time can we go to study variation and change in languages? What are the temporal limits for assessing the *uniformitarian principle* (Bergs 2012)?

While sociolinguistic variation has indeed been accounted for in the study of languages going back at least to classical antiquity (e.g., Adams 2013), this presentation explores the study of language variation and change ca. 4 millennia before present. Using corpora of Old Babylonian epistolary texts (cf. Hernáiz 2020), significant observations can be made regarding phonological and morphological variables in Akkadian, one of the earliest recorded languages. Two case studies will be discussed: the deaffrication of sibilants and gender syncretism in the demonstrative/pronominal paradigm.

Despite clear limitations, the characteristics of the continuously growing record of Akkadian provides documented data (i.e., not reconstructed) that illustrate the co-existence of variants of linguistic variables in communities of the distant past. The analysis of these variables offers a higher resolution image of language change in the context of socio-political fluctuations, population movements and contact between lectal varieties.

Key words

Language variation and change, syncretism, deaffrication, ancient languages, Akkadian

Bibliography

Adams, J. N. (2013). *Social variation and the Latin language*. Cambridge University Press.

Bergs, A. (2012). The Uniformitarian Principle and the risk of anachronisms in language and social history. In Hernández-Campoy, J. M. & Conde-Silvestre, J.C (eds). *The handbook of historical sociolinguistics*, 83-101.

Carling, G., & Cathcart, C. (2021). Reconstructing the Evolution of Indo-European Grammar. *Language*, 97(3), 561-598.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345), 79-82.

Hernáiz, R. (2020). *Studies on linguistic and orthographic variation in Old Babylonian letters*. Ugarit-Verlag.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical sociolinguistics: language change in Tudor and Stuart England*. Longman.

Winter, W. (1999). Sociolinguistics and dead languages. In Ernst Håkon Jahr (ed). *Language Change. Advances in Historical Sociolinguistics*, Mouton de Gruyter. 67-84.

When change fails: evidence from French

INTRODUCTION: In this paper I focus on a series of pronouns found in French between the 12th and the 16th centuries. These pronouns pose a theoretical challenge as they bear a strong morphology (*moy, toy, soy*, etc.), which differentiates them from clitics, yet they appear in a derived position, like clitics. This construction only concerns infinitival complements as exemplified in (1) and (2).

- (1) *Vous auriez tort de moy faire desplaisir.*
 you would.have wrong to me make.INF displeasure
 “You would be wrong to make me unhappy.”
- (2) *et pour cuider qu’il deust soy appaisier.*
 and for believe.INF that-he must REFL appease.INF
 “and to believe that he had to appease himself.”

This construction has been described (Moignet 1970, Pearce 1990, de Kok 1993, Roberts 1997) but it is yet to receive a formal analysis. The sporadic occurrences of clitics in the immediate periphery of the infinitive are traditionally attributed to a preference for the use of pre-infinitival strong pronouns when clitic climbing is not possible (Moignet 1970, Robert 1997). This hypothesis is supported by the obligatoriness of clitic climbing until ca. 1600 (Martineau 1990). Nevertheless, recent research has shown that clitics could cliticise on infinitives from the earliest periods on (Olivier 2022), which challenges this view. The objectives of this paper are (a) to present a quantitative report of the evolution of such pronouns, (b) to characterise their nature, (c) to provide a formal analysis of the construction and (d) to account for their loss.

METHODOLOGY: I have created and manually tagged a corpus of legal texts from 1150 to 1856. The construction under focus is attested between the 12th and the 16th centuries, therefore I will limit the discussion to this timeframe. I counted 66 occurrences. This is extremely low compared to the number of enclisis and proclisis (Table 1); and it dismisses the traditional hypothesis that clitics are banned in this context.

| ENCLISIS | PROCLISIS | CLITIC CLIMBING | PRE-INFINITIVAL PRONOUN |
|----------|-----------|-----------------|-------------------------|
| 190 | 486 | 1,414 | 66 |

Table 1: Distribution of pronouns in the corpus (12th-16th c.)

In restructuring clauses, the pronoun is almost always a clitic and climbs. I have counted two occurrences where the pronoun does not climb and remains strong, see example (2). This construction is extremely rare. In non-restructuring clauses, there is no particular environment that appears to favour the presence of a pre-infinitival pronoun instead of a clitic.

ANALYSIS: DP objects can precede and follow infinitives in Old French (de Kok 1993: 261).

The pronouns under focus are strictly pre-infinitival, which evidences that they appear in a derived position. I propose that their hybridity (they are not clitics but they appear in a derived position) can be explained alongside Cardinaletti & Starke’s (1999) tripartition, namely that they are Weak Pronouns (WP). Being phrases, WPs target a specifier, which I take to be that of *vP*. This construction is present for a limited amount of time (4 centuries) and in low quantities (Table 1). Put informally, it exposes a change that is not *successful* in the diachrony. Failed changes are reported cross-linguistically (Postma 2010, Meyer 2020, Ringe & Yang 2022), and they are cases where a new construction is either rapidly reanalysed, or where it is challenged by another construction. I will propose that the latter is true for French, with the rise of proclisis. Infinitival proclisis appears only a century after we first observe the construction under focus here, which rapidly overshadowed the innovative use of WPs in the 12th century. The two constructions are linearly identical and clitics were favoured during acquisition. Further, clitic climbing becomes optional in restructuring clauses towards the end of the 16th century (Olivier 2022) which substantially increases the acquirers’ exposure to proclisis. This pivotal moment connects to the loss of WPs in the corpus.

SELECTED REFERENCES:

- Cardinaletti, A., & Starke, M. (1999)** The typology of structural deficiency. *Clitics in the languages of Europe*, 145-233.
- Chomsky, N. (2004)** Beyond Explanatory Adequacy.
- Déchaine, R.M. & Wiltschko, M. (2002)** Decomposing pronouns. *Linguistic Inquiry* (33)3, 409-442.
- De Kok, A. (1985)** *La place du pronom personnel régime conjoint en français: une étude diachronique* (No. 23). Rodopi.
- de Kok, A. (1993)** Sur l'origine de l'emploi des formes toniques du pronom. In *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, New Brunswick, 14-18 August 1989* (Vol. 106, p. 249). John Benjamins Publishing.
- Hirschbühler, P., & Labelle, M. (1994)** L'évolution des propositions infinitives négatives en français. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (5), 59-90.
- Hirschbühler, P., & Labelle, M. (2000)** Evolving Tobler-Mussafia effects in the placement of French clitics. *Amsterdam studies in the theory and history of linguistic science series 4*, 165-182.
- Labelle & Hirschbühler (2005)** Changes in clausal organization and the position of clitics in Old French. *Grammaticalization and parametric variation*, 60-71.
- Moignet, G. (1965). *Le Pronom personnel français: essai de psycho-systématique historique..* (Vol. 9). C. Klincksieck.
- Martineau, F. (1990)** *La montée du clitique en moyen français: une étude de la syntaxe des constructions infinitives*. University of Ottawa (Canada).
- Olivier, M. (2022)** *A Corpus Study of Clitic Placement with Infinitives in the Diachrony of French* (Doctoral dissertation, Ulster University).
- Olivier, M. (2022)** Diachronie de la proclise et de l'enclise avec l'infinitif en français médiéval (12e-15e siècles). *Studia linguistica romanica*, (8), 10-35.
- Pearce, E. H. (1990)** *Parameters in Old French Syntax: Infinitival Complements: Infinitival Complements* (Vol. 18). Springer Science & Business Media.
- Postma, G. (2010)** The impact of failed changes. *Continuity and change in grammar*, 269-302.
- Roberts, I. (1997)** Restructuring, head movement, and locality. *Linguistic inquiry*, 423-460.
- Roberts, I. (2010)** *Agreement and head movement: Clitics, incorporation, and defective goals* (Vol. 59). MIT Press.
- Yang, C. (2016)** *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

Usage-based evolutionary models reveal context-specific word order change in Indo-European

In the early days of Indo-European studies, Schleicher (1868) published his famous reconstruction of a fable in Proto-Indo-European to demonstrate that is possible to reconstruct all aspects of a language, from lexicon to morphology, domains where the comparative method has been applied widely and successfully, but also syntax. After the foundational work of Delbrück (1893-1900) and others (Wackernagel 1892; Brugmann 1925), the interest in syntactic reconstruction ceased for a couple of decades. In the 1970s, both Lehmann (1974) and Friedrich (1975) proposed reconstructions of basic word order patterns in Proto-Indo-European following the Greenbergian framework of conditional word order universals. Recently, computational phylogenetic methods have been applied to model the diachronic dynamics and reconstruct syntactic traits and other grammatical features (Greenhill et al. 2010, 2017; Dunn et al. 2011; Carling and Cathcart 2021, and more).

Proto-Indo-European word order was very likely flexible to some extent and allowed non-basic word order for emphasis and to mark information-structural properties (Viti 2014; Lühr 2015). Studies of documented word order changes highlight the importance of synchronic variation as a precondition for change (England 1991; Harris and Campbell 1995; Ross 2007; Heine 2008). Therefore, we propose a new approach to infer the evolutionary dynamics of word order under different pragmatic conditions. Instead of coding word order as an abstract type, we take observed instances in specific pragmatic contexts as a starting point.

To control for pragmatics and information-structure, we extracted a set of 46 sentences in 36 modern Indo-European languages from a parallel corpus (Levshina 2016, with additional data collected by the authors to enhance the coverage of Indo-European languages). Our sentence sample includes different types of subjects and objects (pronouns, nouns, and object clauses) to cover a wide range of constructions that are common in naturalistic speech. We use Bayesian phylogenetic comparative methods to infer transition rates between the states of a binary feature that encodes the order of object and verb. The follow-up analysis investigates whether some of these contexts are more prone to change than others.

The long-term probability of being in one state or the other varies between sentences, with some sentences having a higher probability for OV, while others have a higher probability for VO. This suggests the co-existence of different word order patterns in Proto-Indo-European.

By applying k-means clustering on the mean posterior rates, we identified sets of sentences that evolve in a similar way: one cluster contains verbs of speech and mental verbs with complement clauses which are almost exclusively VO in all modern Indo-European languages, even in those with basic OV order. A second cluster encompasses verbs with object pronouns which tend to precede the verb in many languages of the Romance and Slavic branch that otherwise prefer post-verbal nominal objects. The last cluster contains mostly nominal objects.

Our study did not provide evidence for the initial hypothesis that pragmatic factors lead to more variation and therefore faster rates of change. It is likely that our sentence sample was too small to have sufficient data for word order variation conditioned by pragmatic factors. Specific constructions that share semantic and structural properties can still be identified based on their distinct rates of change. This suggest that these factors play a major role in the evolution of word order.

References

- Brugmann, Karl. 1925. *Die Syntax Des Einfachen Satzes Im Indogermanischen*. Berlin; Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111586304>.
- Carling, Gerd, and Chundra Cathcart. 2021. “Reconstructing the Evolution of Indo-European Grammar.” *Language* 97 (3): lan.0.0253. <https://doi.org/10.1353/lan.0.0253>.
- Delbrück, Berthold. 1893-1900. *Vergleichende Syntax Der Indogermanischen Sprachen*. Grundriss Der Vergleichenden Grammatik Der Indogermanischen Sprachen 3–5. Strassburg: Karl J. Trübner.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russel D. Gray. 2011. “Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals.” *Nature* 473: 79–82.
- England, Nora C. 1991. “Changes in Basic Word Order in Mayan Languages.” *International Journal of American Linguistics* 57 (4): 446–86.
- Friedrich, Paul. 1975. *Proto-Indo-European Syntax: The Order of Meaningful Elements*. Journal of Indo-European Studies. Butte: Montana College of Mineral Sciences.
- Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray. 2010. “The Shape and Tempo of Language Evolution.” *Proceedings of the Royal Society B: Biological Sciences* 277 (1693): 2443–50.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. “Evolutionary Dynamics of Language Systems.” *Proceedings of the National Academy of Sciences of the United States of America* 114: E8822–29.
- Harris, Alice C., and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge Studies in Linguistics 74. Cambridge: Cambridge University Press.
- Heine, Bernd. 2008. “Contact-Induced Word Order Change Without Word Order Change.” In *Language Contact and Contact Languages*, edited by Peter Siemund and Noemi Kintana, 33–60. Amsterdam; Philadelphia: Benjamins.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin; London: University of Texas Press.
- Levshina, Nathalia. 2016. “Verbs of Letting in Germanic and Romance: A Quantitative Investigation Based on a Parallel Corpus of Film Subtitles.” *Languages in Contrast* 16 (1): 84–117. <https://doi.org/10.1075/lic.16.1.04lev>.
- Lühr, Rosemarie. 2015. “Traces of Discourse Configurability in Older Indo-European Languages?” In *Perspectives on Historical Syntax*, edited by Carlotta Viti, 203–32. Studies in Language Companion Series 169. Amsterdam & Philadelphia: Benjamins.
- Ross, Malcolm. 2007. “Calquing and Metatypy.” *Journal of Language Contact* 1: 116–43. <https://doi.org/10.1163/000000007792548341>.
- Schleicher, August. 1868. “Eine Fabel in Indogermanischer Ursprache.” *Beiträge Zur Vergleichenden Sprachforschung Auf Dem Gebiete Der Arischen, Celtischen Und Slawischen Sprachen* 5 (2): 206–8. <http://www.jstor.org/stable/23458804>.
- Viti, Carlotta. 2014. “Reconstructing Syntactic Variation in Proto-Indo-European.” *Indo-European Linguistics* 2 (1): 73–111. <https://doi.org/10.1163/22125892-00201004>.
- Wackernagel, Jakob. 1892. “Über Ein Gesetz Der Indogermanischen Wortstellung.” *Indogermanische Forschungen* 1: 333–436.

References

- Berthele, Raphael, Matthew Whelpton, Áshild Næss & Pieter Duijff. 2015. Static spatial descriptions in five Germanic languages. *Language Science* 49:82–101.
- Friðjónsson, Jón G. 2005. Kerfisbundnar breytingar á notkun nokkurra forsetninga í íslensku [Changes in the use of some prepositions in Icelandic]. *Íslenskt mál* 27:7–40.
- Grestenberger, Laura. 2015. Number marking in German measure phrases and the structure of pseudo-partitives. *The Journal of Comparative Germanic Linguistics* 18:93–138.
- Jónsson, Jóhannes Gísli. 2017. Avoiding genitive in Icelandic. In Höskuldur Thráinsson et al. *Syntactic variation in Insular Scandinavian*, 141–163. Amsterdam: John Benjamins.
- Kjartansson, Helgi Skúli. 1979. Eignarfallsflótti: Uppástunga um nýja málvillu [Genitive-Avoidance: A proposal on a new linguistic error]. *Íslenskt mál* 1:88–95.
- ONP = *Ordbog over ded norrøne prosasprog*. Copenhagen: Den Arnamagnæanske Kommission (eds). Access at: <https://onp.ku.dk/onp/onp.php>
- RMH = Risamálheild [Gigaword Corpus]: Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. *Proceedings of LREC 2018*, 4361–4366, Japan. Accessed at: <https://malheildir.arnastofnun.is/?mode=rmh2022>
- ROH = *Ritmalssafn Orðabókar Háskólans*. Stofnun Árna Magnússonar í íslenskum fræðum (eds.). Access at: <http://ritmalssafn.arnastofnun.is/>
- Rögnvaldsson, Eiríkur. 2021. *Hliðiná*. A blog post on topics in Icelandic language use. Published January 9th, 2021. Accessed at: <https://uni.hi.is/eirikur/2021/01/09/hlidina/>

Differential Place Marking and the reconstruction of the Proto-Nakh system of spatial cases

Within the East Caucasian language family, the Nakh branch (often assumed to be the first to have split off from the protolanguage) has often been described as typologically dissimilar to all other languages of the family (collectively known as Daghestanian) with respect to its marking of spatial relations. Where Daghestanian languages of various branches typically display two-slot systems, with one suffix marking location (in, on, under, through, at, near) and another marking direction (allative, ablative, essive) (Daniel & Ganenkov 2012), the Nakh languages have been analysed as having fewer, monomorphemic spatial cases and as making use of more postpositions (e.g. Nichols 2011).

Two recent advancements have shed new light on the Nakh data. Firstly, reanalysis of the Tsova-Tush data (the third Nakh language besides Chechen and Ingush) has allowed the recognition of a Daghestanian-style two-slot system of spatial cases, (Author, forthc.). See Table 1 for 12 of the 33 (combinations of) spatial cases in Tsova-Tush. Secondly, the Tsova-Tush data displays clear features of Differential Place Marking. The notion of Differential Place Marking identifies splits in the coding of locative, allative or ablative roles depending on subclasses of nouns, in particular place names (toponyms), inanimate common nouns and human nouns (Haspelmath 2019).

| | Goal | Location | Source | Compatible nouns |
|-------------|------|----------|---------|---|
| 'near, at' | -go | -go-ħ | -go-ren | animates |
| 'among, in' | -lo | -lo-ħ | -lo-ren | liquids, masses, collections |
| 'in' | -i | -i-ħ | -i-ren | rooms, buildings, containers, place names |
| Default | -∅ | -ħ | -ren | other |

Table 1: Tsova-Tush spatial cases (excerpt)

Based on these findings, two questions arise:

- Should a two-slot system be reconstructed for Proto-Nakh, and if so, are the morphemes cognate with those found in Daghestanian languages?
- Is Differential Place Marking an innovation in Tsova-Tush, or is it inherited from Proto-Nakh?

This paper has three goals. (1) It puts the Tsova-Tush data in typological and areal perspective, and concludes that it shows clear parallels with the Daghestanian-style system, and furthermore that it obeys the typological universals concerning Differential Place Marking (i.e. there is less phonetic material in spatial suffixes on place names compared to other nouns, there is more phonetic material in spatial suffixes on animate nouns, and more phonetic material in non-spatial cases on place names). (2) It re-evaluates the Chechen and Ingush data, concluding that they show clear traces of a former two-slot system with many cognate morphemes (see Chechen comparative case *-l*, allative case *-ie/-ga/-a*); (3) It reconstructs the spatial case system for Proto-Nakh, with clear cognates to Daghestanian languages (as reconstructed by Alekseev (1997)), but concludes that the Differential Place Marking features of Tsova-Tush are secondary.

By answering the above questions, this paper aims to be an important case-study in (1) the grammaticalisation of locative markers; (2) the history and internal branching of the East Caucasian family, where often cognate sets of morphemes are established (e.g. Desheriev 1963:436), but attempts at reconstruction are few and far between; and (3) the recent topic of Differential Place Marking in Caucasian languages, where related phenomena such as Differential Subject Marking have been observed previously (see Arkadiev 2017).

- Alekseev, Mixail. 1999. "Reconstruction of the Proto-East Caucasian locative morphemes". In: van den Berg, Helma (ed.) *Studies in Caucasian linguistics: selected papers of the eighth Caucasian colloquium*. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), p. 116-124.
- Arkadiev, Peter. 2017. Multiple ergatives. From allomorphy to differential agent marking. *Studies in Language* 41(3): 717-780. <https://doi.org/10.1075/sl.41.3.06ark>
- Daniel, Michael & Ganenkov, Dmitry. 2008. "Case Marking in Daghestanian: Limits of Elaboration", in: Malchukov, Andrej L. & Spencer, Andrew (eds) *The Oxford Handbook of Case*. Oxford Academic. p. 668-685. <https://doi.org/10.1093/oxfordhb/9780199206476.013.0047>
- Desheriev, Yunus. 1963. *Sravnitel'no-istoričeskaja grammatika naxskix jazykov [Historical comparative grammar of the Nakh languages]*. Grozny: Čečeno-ingušskoe knižnoe izdatel'stvo.
- Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF* 72(3):313-334. <https://doi.org/10.1515/stuf-2019-0013>
- Nichols, Johanna, 2011. *Ingush grammar*. University of California Press
- Author (forthcoming). *Intensive language contact in the Caucasus: the case of Tsova-Tush*.

Contact and the origins of headed *wh*-relatives in Hungarian

Headed *wh*-relatives are a feature of Standard Average European (Haspelmath, 1998, 2001). De Vries (2002) found that 40% of Indo-European (IE) languages had such relative clauses, but only 2.3% of the non-IE languages in his sample did. The non-IE languages that do include Hungarian, Finnish, and Georgian; (1) is an example from Hungarian, from (Comrie 1998: 60).

- (1) A fiú, akit láttam
 The boy who.ACC I saw ‘the boy I saw’

Comrie (1998) and Hendery (2012) explain the cross-linguistic distribution of headed *wh*-relatives in terms of contact. However, the precise mechanisms of this contact-induced distribution are obscure. In particular it is hard to learn functional vocabulary like *which*: there are fluid pairings between category and denotation which give rise to persistent ambiguity. This underspecified nature of function-word meaning makes change likely, but direct borrowing difficult. Instead, it seems probable that Hungarian developed a precursor of headed *wh*-relatives through contact. Headed *wh*-relatives then emerged as Hungarian followed a recurring pathway found extensively in IE languages.

In IE languages, the Proto-Indo-European indefinite/interrogative pronouns **kwi-/kwo-*, which did not head relative clauses, are the source of IE *wh*-relative forms. There is a pathway from conditional to correlative (Belyaev & Haug, 2014, 2020) and on to headed relative (Haudry, 1973) with multiple possible pathways through the semantic space (Gisborne & Truswell, 2018). Conditional-correlative constructions, formed around indefinite-interrogative pronouns (a class of words found in interrogatives, conditionals, and other dependent contexts, Haspelmath, 1997), are therefore the source of headed *wh*-relatives in the IE daughter languages, where they have developed through parallel evolution. Word-order conditions the change. Conditional protases are topics (Haiman, 1978) and the indefinite/interrogative pronoun is focused. Kiparsky (1995) argues that early Indo-European (Vedic, Hittite, Greek) had the structure (TOPIC)-(FOCUS)-Clause: the emergence of conditional-correlatives involves topicalization of the conditional clause and, typically, fronting of the indefinite/interrogative pronoun.

Examples of conditional-correlatives are found in modern Hungarian: (2) is from Lipták (2009: 27).

- (2) Amelyik kutya közel jön hozzám, azt elkergetem
 REL.which dog close comes to.me that.ACC chase.away

‘Which(ever) dog comes close to me, I’ll chase it away’=‘If a dog comes, close to me I’ll chase it away’

However, structures like (2) were not possible in proto-Hungarian, which Kiss (2013) reconstructs as strict SOV/head-final, with grammaticalized discourse roles for S and O. S is always the (primary) topic and O the focus or secondary topic. For patterns like (2) to emerge in Hungarian, a word-order change was necessary. Kiss (2013) argues that the SOV of proto-Hungarian developed into (TOPIC)-(FOCUS)-V-X* in Old Hungarian, giving Hungarian a word-order type consistent with the early IE languages that developed conditional correlatives.

Contact-driven change depends on both the complexities of the sociolinguistic context, and on what can plausibly be transferred from one language to another in bilinguals and bilingual use. Some contact-driven changes seem more plausible than others. As noted above, functional vocabulary is hard to learn. Similar abstract forms to *wh*-relatives appear resistant to borrowing: Sorbian has co-opted its demonstratives to certain functions of definite marking without having all the properties of definite articles (Heine & Kuteva 2005: 71-73). And yet, contact induces word-order change (Heine & Kuteva, 2005). We argue that Kiss’ word-order change provides the relevant context for the development of a headed *wh*-relative in Hungarian. It is also possible that Hungarian borrowed left-adjoined conditional-correlatives formed on indefinite-interrogative pronouns. In either case, given the necessary pre-conditions, the potential pathway of change exists, making it possible for *wh*-relatives to emerge in Hungarian without having to borrow abstract, underspecified forms. The account developed here allows us to develop a plausible understanding of the role of contact in the diffusion of areal phenomena by understanding the context in which ‘replica’ development (Heine and Kuteva, 2005) is enabled.

References

- Belyaev, O., & Haug, D. (2014). *The genesis of wh-Based correlatives: From indefiniteness to relativization*.
- Belyaev, O., & Haug, D. (2020). The genesis and typology of correlatives. *Language*, 96(4), 874–907. <https://doi.org/10.1353/lan.2020.0065>
- Comrie, B. (1998). Rethinking the typology of relative clauses. *Language Design*, 1, 59–86.
- De Vries, M. (2002). *The syntax of relativization*. Universiteit van Amsterdam.
- Gisborne, N., & Truswell, R. (2018). *Parallel evolution of relative clauses in Indo-European*.
- Haiman, J. (1978). Conditionals are topics. *Language*, 54(3), 564–589.
- Haspelmath, M. (1997). *Indefinite pronouns*. Oxford University Press.
- Haspelmath, M. (1998). How young is standard average European. *Language Sciences*, 20(3), 271–287.
- Haspelmath, M. (2001). The European linguistic area: Standard Average European. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 1492–1510). De Gruyter.
- Haudry, J. (1973). Parataxe, hypotaxe et corrélation dans la phrase latine. *Bulletin de La Société Linguistique de Paris*, 68, 147–186.
- Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge University Press.
- Hendery, R. (2012). *Relative clauses in time and space: A case study in the methods of diachronic typology*. John Benjamins.
- Kiparsky, P. (1995). Indo-European origins of Germanic syntax. In A. Battye & I. Roberts (Eds.), *Clause structure and language change* (pp. 140–169). Oxford University Press.
- Kiss, K. É. (2013). From Proto-Hungarian SOV to Old Hungarian Top Foc V X*. *Diachronica*, 30(2), 202–231.
- Lipták, A. (2009). The landscape of correlatives: An empirical and analytical survey. In *Correlatives cross-linguistically* (pp. 1–46). John Benjamins Publishing Company.

Areal alignment and the loss of ATR harmony in Riverine Bua languages (Chad)

Several linguistic areas have been identified in Africa, one of the most discussed in recent years being the Macro-Sudan Belt (MSB; Güldemann, 2008, 2010, 2018; Clements and Rialland 2008). Areas of intermediary size have also been identified (e.g. Central Africa within the MSB, cf. Dryer 2009, Idiatov 2018, Güldemann 2018b: 457; Rolle, Lionnet & Faytak 2020, a.o.). Such linguistic areas suggest a form of areal pressure, i.e., languages have a strong tendency to adapt their linguistic profile to the area where they are spoken, or where their speakers migrate into (cf. Güldemann 2020, Rolle et al. 2020 for examples of such changes).

The goal of this talk is to illustrate the inner workings of areal alignment, with a detailed case study of Bua languages (southern Chad), focusing on the changes that affect the vowel systems. Bua languages form a tight family of 9 languages, traditionally classified in the “Adamawa” grouping within Niger-Congo. The family is divided into two markedly different branches: Riverine languages (Lua, Ba, Tun) and Inland languages. Inland languages have a robust ATR contrast and harmony and no interior (i.e. central and other non-peripheral) vowels, while Riverine languages have no ATR contrast or harmony, and have interior (mostly central and front rounded) vowels. This difference is illustrated with the inventories of most Inland languages (and Proto-Bua, as I will show) in (1), and Lua (Riverine) in (3).

| | | |
|---------------------------------|-----------------|---|
| (1) Proto-Bua (and most Inland) | (2) Loss of ATR | (3) Central vowels > Lua (Boyeldieu 1985) |
| ATR: + *i *u | *i *u | i ị u <i>high</i> |
| – *ɪ *ʊ | | |
| + *e *o | *e *o | e ə̣ o <i>mid</i> |
| – *ɛ *ɔ | *ɛ *a *ɔ | ɛ a ɔ <i>low</i> |
| – *a | | |

Based on a comparative Bua database of about 600 terms (Author et al. 2023), I show that:

1) Proto-Bua had a 9-vowel system very close to that of Inland languages, with robust ATR contrast and harmony – a 2IU system in Casali’s (2008) typology (step (1) above).

2) Riverine languages are doubly innovative: firstly, they lost the ATR contrast by merging the [-ATR] high vowels *ɪ and *ʊ with the [+ATR] mid vowels /e/ and /o/ respectively, and reinterpreted ATR harmony as height harmony in a rectangular vowel system (step (2) above), and secondly, they independently innovated central vowels (step (3)). This change confirms the tendency for ATR attrition to be caused by the loss of [-ATR] high vowels [ɪ ʊ], which are known to be perceptually confusable with either [i u] or [e o] (Casali 2003: 342; Rose 2018, a.o.). The confusability with [e o] was the phonetic precursor to the merger that took place in Riverine languages – a confusability that is still noticeable in present-day Inland languages, as I will show.

3) These changes are the result of areal alignment: the resulting vowel systems of Riverine languages are almost identical to those of neighboring languages: Laal (isolate), East Chadic, and Sara-Bongo-Bagirmi (Central Sudanic) languages. Conversely, the Inland languages (and by extension proto-Bua) have a general phonological profile that is unexpected in the area.

I will conclude with preliminary remarks and questions about (i) the relationship between ATR and interior vowels in the MSB; (ii) the sociolinguistic underpinnings of areal alignment – notably the crucial role of small-scale multilingualism (Lüpke 2016; DiCarlo et al. 2019); and (iii) what these diachronic changes tell us about the linguistic history of the region, notably about the time depth of the Central African ATR-deficient / Interior Vowel zone (Rolle et al. 2020), which the speakers of the distant ancestor of Bua languages (likely to have had ATR harmony) must have crossed in a distant past on their way from Nigeria to their current location.

References:

- Author et al. 2023. [redacted for review purposes]
- Boyeldieu, Pascal. 1985. *La Langue lua ('niellim'), groupe boua, Moyen-Chari, Tchad: phonologie, morphologie, dérivation verbale*. (Descriptions de langues et monographies ethnolinguistiques, 1.) Paris: SELAF.
- Boyeldieu, Pascal, Raimund Kastenholz, Ulrich Kleinewillinghöfer, and Florian Lionnet. 2018. The Bua Group languages (Chad, Adamawa 13): A comparative perspective. In Rajja Kramer and Roland Kießling (eds.), *Current Approaches to Adamawa and Gur Languages*, Afrika und Übersee Supplements, Vol. 34. Cologne: Köppe. Pp. 53-126.
- Casali, Roderic F. 2003. [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7:307-382.
- Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2. 496-549. <http://onlinelibrary.wiley.com/doi/10.1111/j.1749-818X.2008.00064.x/abstract>.
- Di Carlo, Pierpaolo, Jeff Good, and Rachel A. Ojong Diba. 2019. Multilingualism in rural Africa. *Oxford Research Encyclopedia of Linguistics*. Retrieved 2 Aug. 2022, from <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-227>.
- Dryer, Matthew S. 2009. Verb-object-negative order in central Africa. In Norbert Cyffer, Erwin Ebermann & Georg Ziegelmeyer (eds.), *Negation patterns in West African languages and beyond*. *Typological Studies in Language* 87, 307–362. Amsterdam: John Benjamins.
- Güldemann, Tom. 2008. The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In Heine, Bernd and Derek Nurse (eds.), *A linguistic geography of Africa*. Cambridge: Cambridge University Press, 151-185
- Güldemann, Tom. 2010. Sprachraum and geography: linguistic macro-areas in Africa. In Lameli, Alfred, Roland Kehrein and Stefan Rabanus (eds.), *Language and space: an international handbook of linguistic variation, volume 2: language mapping*. *Handbooks of Linguistics and Communication Science* 30,2. Berlin: Mouton de Gruyter, 561-585, Maps 2901-2914.
- Güldemann, Tom. 2018. Language contact and areal linguistics in Africa: introduction. In Güldemann, Tom (ed.), *The languages and linguistics of Africa*. *The World of Linguistics* 11. Berlin: Mouton de Gruyter, 445-446.
- Güldemann, Tom. 2020. Changing profile when encroaching on hunter-gatherer territory: towards a history of the Khoe-Kwadi family in southern Africa. In Güldemann, Tom, Patrick McConvell and Richard Rhodes (eds.), *The languages of Hunter-Gatherers*, 114-146. Cambridge: Cambridge University Press.
- Idiatov, Dmitry 2018. An areal typology of clause-final negation in Africa: Language dynamics in space and time. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 115–163. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110607963-005.
- Lüpke, Friederike. 2016. Uncovering small-scale multilingualism. *Critical Multilingualism Studies* 4(2): 35-74.
- Rolle, Nicholas, Florian Lionnet, and Matthew Faytak, 2020. Areal patterns in the vowel systems of the Macro-Sudan Belt. *Linguistic Typology* 24(1): 113-179. DOI: 10.1515/lingty-2019-0028
- Rose, Sharon. 2018. ATR Vowel Harmony: new patterns and diagnostics. *Proceedings of the 2017 Annual Meeting on Phonology*. DOI: 10.3765/amp.v5i0.4254.

Morphosyntactic variation in Swahili: Tracing descriptions past and present

Swahili is a Bantu language spoken by some 100 million people across East Africa. In addition to its high speaker numbers and geographic spread, Swahili is one of the few African languages for which we have written materials from over a century ago. There are Swahili manuscripts dating from the 18th century – many of which record text and language which is likely to be older. Interest in this written tradition in part gave rise to descriptions of Swahili by European linguists, missionaries and grammarians towards the end of the 19th century. Another key feature of the descriptive accounts of Swahili from this period was the interest in (dialectal) variation in the language, with Swahili spoken along the coast from Southern Somalia to northern Mozambique, on a number of islands and into the ‘mainland interior’ of East Africa.

Many of these early accounts focused on lexical and phonological differences. However, there are also descriptions of grammatical differences on which the present talk focuses. We revisit a number of these earlier sources with a view to examining the variation described, compare these accounts to present-day variation, and to develop a more comprehensive picture of morphosyntactic variation and change in Swahili (cf. Nurse and Hinnebusch 1993). Morphosyntactic variation remains an under-examined aspect of Swahili, and the talk aims to show how the study of this variation in the early sources contributes to a better understanding of the interacting processes of language change, dialect contact and convergence, and effects of standardisation in the history of Swahili.

Early materials are taken from a number of key sources. We consider the data in Steere’s (1870) seminal *Handbook of the Swahili Language* which is based primarily on the Zanzibari dialect Kiunguja - the dialect on which ‘Standard Swahili’ was based and developed in the twentieth century. We also examine also Velten’s (1901) *Safari za Wasuahili* ‘Travels of the Swahili people’, which is a key reference for this time period, Sacleux (1909) which draws on descriptions of ten Swahili dialects (although focusing on Kiunguja), and Stigand’s (1915) comparative study which provides a more explicit account of dialectal variation in Swahili, primarily in phonology and the lexicon. Finally we consider a set of more recent sources in Lambert (1957, 1958a,b) which examine the Chijomvu, Kingare, Chifundi and Vumba dialects of the southern Kenyan coast. We examine the following domains of variation in these sources:

- Verbal morphology – e.g. subject and object agreement, plural marking; behaviour of monosyllabic verb stems such as the retention of infinitive *ku-* prefix
- Nominal domain and nominal dependents - e.g. variation in class 1/2 prefixes *m-* ~ *mu-*; variation in class 1/2 possessive agreement *y-ake* ~ *w-ake*, noun class membership, personal pronouns *mimi* ~ *mie*;
- Tense-aspect-mood distinctions – e.g. near and distant past tense *-li-/-liki-*;
- Negation strategies – e.g. the use of the negative word *hapana*
- Locative formation – e.g. noun class prefixes, *vahali* ~ *mahali*, class 16 concord *va-*

We use these sources and examples of variation to contribute to the discussion on the historical development of Swahili, and pathways of grammatical change more broadly, including in contexts of high linguistic diversity, language contact and multilingualism. The talk aims to provide a better understanding of morphosyntactic variation in Swahili, Swahili dialects, and how the dialects have changed over time. We show that many of the areas in which this earlier variation was described are similar to those found in present day Swahili, or in Bantu languages more widely. We also pay attention to issues of authorship and voice, as well as how processes of standardisation have impacted on the Swahili seen in the present day.

References

- Lambert, H. E. 1957. *Ki-Vumba, a dialect of the southern Kenya coast*. (Studies in Swahili Dialect, II.) Kampala: East African Swahili Committee.
- Lambert, H. E. 1958a. *Chi-Jomvu and ki-Ngare: Sub-dialects of the Mombasa Area*. (Studies in Swahili Dialect 3). Kampala: East African Swahili Committee.
- Lambert, H. E. 1958b. *Chi-Chifundi: a Dialect of the Southern Kenya coast*. (Studies in Swahili Dialect, V.) Kampala: East African Swahili Committee.
- Nurse, Derek and Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: A Linguistic History*. University of California Press.
- Sacleux, C. 1909. *Grammaire des dialectes swahilis*. Paris: R. P. du Saint Esprit
- Steere, Edward. 1919. *A handbook of the Swahili language as spoken at Zanzibar*. London: Society for promotion Christian knowledge.
- Stigand, Chauncy Hugh. 1915. *Dialect in Swahili: A Grammar of Dialectic Changes in the Kiswahili Language*. Cambridge University Press. Cambridge
- Velten, Carl. 1901. *Safari za Wasuaheli*. Gottingen: Vandenhoeck & Ruprecht.

Outliers in variation and change: atypical users of the variants of negation in Old and Middle Hungarian

According to Weinreich et al. (1968: 188)’s classic observation, „Not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity.” The present paper aims to discuss the history of sentence negation in Hungarian, with a focus on variability and heterogeneity that **precede** change. This variation can be investigated in negative sentences that contain Verb Modifiers (VM); the basic features of the variants are summarized in the table below. The coexistence of the variants can be broken down into three periods: a) stable variation with the predominance of the A-variant (from the first written documents until the 19th century); b) the radical frequency growth of the M-variant (during the 19th century); c) stable variation with the predominance of the M-variant (since the 20th century).

| word order | VM – NEG – V | NEG – V – VM |
|------------------------------------|--|--|
| example | <i>el nem megyek</i> away _[VM] not go.1sg 'I am not going away' | <i>nem megyek el</i> not go.1sg away _[VM] 'I am not going away' |
| structure (É. Kiss 2014) | Negative particle: adjoined to the verb in a lower structural position (PredP); verb modifier: precedes the negated verb | Negative particle: merged into a higher structural position (NegP), eliciting verb movement; verb modifier: follows the negated verb |
| shorthand | A (djunction-based)- variant | M (ovement-based)- variant |

The period that precedes change is investigated based on three groups of texts: a) Old Hungarian codices; b) Old and Middle Hungarian informal texts; c) 16th century (=early Middle Hungarian) translations of the New Testament. The overall distribution of the two variants seems to be fairly similar in the three groups, but each group features an outlier (see the table below). The paper addresses the questions 1. whether the outliers differ in the same way from the majority, or they are different from each other as well; 2. whether these can be seen as the forerunners of the 19th-century change.

| group | Old Hungarian codices | 16 th -century translations of the New Testament | Old and Middle Hungarian informal texts |
|------------------------------|---|---|---|
| rate of the M-variant | 22.8% (N=3472) | 24.3% (N=745) | 15.4% (N=3125) |
| outlier | the Hussite codices (15th century) | Károlyi's translation (1590) | Kolozsvár witch trials (1572-1592) |
| M-var. in the outlier | 71.7% | 43.3% | 55.2% |
| source of data | http://omagyarkorpusz.nytud.hu/en-intro.html | | https://tmk.nytud.hu/ |

The tentative answers to the questions can be summarized as follows:

- a) The Hussite subgroup may represent an innovative dialect, being on the way to generalizing the use of the M-variant. However, this dialect disappeared. b) The more frequent use of the M-variant in Károlyi's translation may be Károlyi's idiolectal trait. Still, it cannot be excluded that as a highly influential text, it had a (long-term, indirect) impact on the choice of patterns. c) Kolozsvár subgroup: although the unmarked word order is VM – V in neutral sentences, this group displays a persistent use of the V – VM pattern in this context. The reasons for this are not fully understood yet, but it might be the case that the choice between the A- and the M-variant is also influenced by a factor that is operative in neutral sentences as well, that is, independent of negation.
2. There is one important feature all these outliers share both with each other and with the rest of the sources: if clause types that are attested frequently enough are ordered on a scale based on the frequency of the M-pattern, these scales are almost the same: reason / main > complement > relative > conditional (> *until*-clauses). This corroborates the suggestion (Author 2017, 2022) that the variants differed in their pragmatic load in the first period of stable variation, the M variant being more emphatic; the outliers in Old and Middle Hungarian were early overusers of the emphatic variant, but this did not lead to a change in the entire population of speakers at that time.

References

É Kiss, Katalin 2014. The evolution of functional left peripheries in the Hungarian sentence. In *The evolution of functional left peripheries in Hungarian syntax*, ed. Katalin É. Kiss, 9–55. Oxford: OUP.

Weinreich, Uriel & Labov, William & Herzog, Marvin I. 1968. Empirical foundations for a theory of language change. In: *Directions for historical linguistics*, eds Winfred P. Lehmann & Yakov Malkiel 95–195. Austin: University of Texas Press.

Author 2017. A magyar tagadó mondatok szórendje és a konstansráta-hipotézis [Word order in Hungarian negative sentences and the Constant Rate Hypothesis]. In: *Nyelvelmélet és diakronia 3.*, eds Katalin É. Kiss & Attila Hegedűs & Lilla Pintér, 91–110. Budapest – Piliscsaba: PPKE BTK.

Author 2022. Contrast maintenance in syntax: a highly atypical case of the Negative Cycle. [Manuscript]

Towards a new classification of Western Bantu languages using non-lexical data

The present contribution aims at revisiting Nurse and Philippson's 2003 paper wherein they propose a new classification of Bantu languages based on the study of 30 phonological and morphological criteria. According to the authors, classifications based solely on the study of the lexicon can lead to errors because of the imperfect notion of similarity (cognate) used to classify languages, possible borrowings and the rapid evolution of vocabulary. Therefore, Nurse and Philippson proposed the first non-lexically based historical classification of 80 Bantu languages across the entire family. While Nurse and Philippson's approach and methodology were innovative, they encountered some problems: (i) the 80 Bantu languages selected for their study inadequately covered the North-Western area with only 8 languages spoken in zone A, 1 in zone B and 4 spoken in zone C, (ii) the available data lacked both in quantity and quality, and (iii) the study lacked the use of advanced classificatory techniques.

In order to address these problems, we propose here to work on a new sample of languages and on the selection of new non-lexical data. For this paper, we have decided to focus first on the study of Western Bantu languages that are spoken in Cameroon, Gabon, Equatorial Guinea (EG), Congo, DRC, Angola, Namibia, Zambia and Botswana by building a database containing approximately 100 languages from zones A, B, C, D, K, R, H, L. We selected morphological criteria from both the nominal and verbal domains as well as phonological criteria to be used to classify these languages. Such criteria include but are not limited to the presence/absence of gender categories, patterns of syncretism within nominal paradigms, singular/plural correspondences, verbal derivation, tense-aspect morphology, and sound changes.

Brown et al. (2023) tests the historical informativity of these criteria focusing solely on 32 languages in Zone A and B spoken in Cameroon, Equatorial Guinea and Gabon with promising results (i.e., Northwest Bantu). The classification produced therein consists of strong genetic groupings that correspond to what is found in the lexically based classifications in Grollemund 2012 and Grollemund et al. 2015—namely among A70-80-90 languages and the West Coastal languages (B20-50-60 in our sample). Even more revealing was the emergence of larger grouping containing languages from Zone A10-20-30 and B10-30. The majority of these languages are spoken along the coast of Cameroon, EG and Gabon. It is therefore possible that the non-lexical criteria considered for this study have uncovered a contact relationship among these languages that the lexical data failed to show.

The resulting phylogenetic classification of the Western Bantu languages based on non-lexical parameters has also proven to be enlightening. For example, a clear division between the Forest Bantu languages (Zones A, B, C, parts of H and parts of D10-20-30) and the rest of the sample emerges when looking at the expression of perfectivity. Forest languages almost exclusively exploit the suffix *-i* for perfectivity (and related temporal categories) while the others utilize the suffix *-ile*. Furthermore, phonological evidence further divides Forest Bantu in two groups. We find that North-western languages (Zones A, B20 and some C languages) have the reflexes \emptyset and *k* for the PB consonants **k* and **g* respectively. These results reveal the utility of considering non-lexical data in doing language classification.

References

Brown, Braden, Rebecca Grollemund, Gérard Philippson and Derek Nurse. Publication expected in January 2023. Towards a new classification of North-Western Bantu languages using non-lexical data. In: *Bantoid and Bantu in Cameroon: an historical re-assessment*, Grollemund, R., Nurse, D. and Watters, J. (eds). Berlin/Boston, De Gruyter Mouton.

Grollemund Rebecca. 2012. *Nouvelles approches en classification: Application aux langues bantu du Nord-Ouest*. Thèse en Sciences du Langage (Université Lumière Lyon 2).

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences (PNAS)* 112.13296-301.

Grollemund, Rebecca and Gérard Philippson. Publication expected in 2023. A new phylogenetic classification of North-Western Bantu languages based on consonant correspondences. In: *Bantoid and Bantu in Cameroon: an historical re-assessment*, Grollemund, R., Nurse, D. and Watters, J. (eds). Berlin/Bo ston, De Gruyter Mouton.

Nurse Derek & Gérard Philippson. 2003 Towards a historical classification of the Bantu languages. In: *The Bantu Languages*, eds Nurse D & Philippson G (Routledge, London-New-York), pp 164-181.

An evolutionary loner in Southern African Bantu: The classification of Yeyi

Bantu is Africa's largest language family and part of the world's largest phylum, i.e., Niger-Congo. The relatedness of Bantu languages has been recognized since nearly two centuries, and their internal classification is better understood today than ever before. Nonetheless, there are still several important uncertainties obscuring not only the diachronic linguistic processes that gave rise to Bantu diversification, but also the migrations of ancestral Bantu speakers and other population dynamics underlying them. One of the persisting mysteries of Bantu genealogy is the unclassifiable Yeyi language, a minority language spoken in northwestern Botswana and northeastern Namibia. While the Bantu origin of Yeyi is straightforward and undisputed, it is not known what its closest relatives are, even not to which major Bantu branch it belongs (Gowlett, 1997; Seidel, 2005, 2009). This lack of genealogical resolution severely limits our understanding of the deep history of the Yeyi speech community and leaves us with questions on their origin and timing of their migration into their current habitat, and the deeper connections they might once have had with other language communities.

One of the possible explanations why it is so complicated to identify Yeyi's closest Bantu relatives is its incorporation of a large degree of linguistic influence from neighboring Khoisan languages. Khoisan languages, previously analyzed as a single phylum but more recently and accurately as (at least) three separate language families (Güldemann, 2014), are characterized by their use of click phonemes. These typologically rare consonants not only occur in Khoisan languages, but also in certain Bantu languages spoken in Southern Africa. They are therefore seen as a clear indicator of Bantu-Khoisan contact (Pakendorf et al., 2017). Yeyi has the largest click inventory of all attested Bantu languages (Fulop et al., 2003; Sommer & Voßen, 1992), which suggests extensive Khoisan influence. This is also seen in the presence of Khoisan loanwords, and the borrowing of certain bound verbal affixes (Gunnink, 2022).

In this paper, we intend to shed new light on the classification of the Yeyi language, with an aim to better understand its genesis and historical development. Using a lexicon-based, Bayesian phylogenetic approach, we will provide a new Bantu classification with the explicit objective of clarifying the position of Yeyi. We therefore include all Bantu languages of the wider geographic region, including even more far-flung languages that have previously been hypothesized to bear some relationship to Yeyi. The resultant linguistic phylogeny will show the (potential) linguistic affiliations of Yeyi. These will subsequently be studied in more detail in order to identify specific lexical, phonological or morphological innovations that Yeyi may share with its putative sister languages. The role of non-Bantu contact, particularly in terms of Khoisan influence, will also be taken into account. Together, they will provide new insights into the history of the Yeyi language and its speakers.

Fulop, S. A., Ladefoged, P., Liu, F., & Vossen, R. (2003). Yeyi clicks: Acoustic description and analysis.

Phonetica, 60, 231–260.

Gowlett, D. F. (1997). Aspects of Yeyi diachronic phonology. In W. H. G. Haacke & E. E. Elderkin (Eds.), *Namibian languages: Reports and papers* (pp. 235–263). Rüdiger Köppe.

- Güldemann, T. (2014). 'Khoisan' linguistic classification today. In T. Güldemann & A.-M. Fehn (Eds.), *Beyond 'Khoisan': Historical relations in the Kalahari basin* (pp. 1–40). John Benjamins Publishing Company.
- Gunnink, H. (2022). *Morphological Khoisan influence in the Southern African Bantu language Yeyi*. 43(1), 3–35. <https://doi.org/10.1515/jall-2022-8892>
- Pakendorf, B., Gunnink, H., Sands, B., & Bostoen, K. (2017). Prehistoric Bantu-Khoisan language contact: A cross-disciplinary approach. *Language Dynamics and Change*, 7(1), Article 1.
- Seidel, F. (2005). The Bantu languages of the Eastern Caprivi: A dialectometrical analysis and its historical and sociolinguistic implications. In *South African Journal of African Languages* (Vol. 4, pp. 207–242).
- Seidel, F. (2009). Layered language genesis in the 'catch basin' of the Linyanti and Okavango swamps: The case of Yeyi. In *Sprache und Geschichte in Afrika* (Vol. 20, pp. 231–264).
- Sommer, G., & Voßen, R. (1992). Schnalzwörter im Yei (R.41). *Afrika Und Übersee*, 75, 1–42.

Diachronic Null Subject Use across Latin American Spanish: Comparing Corpora¹

It has been widely noted that varieties of Latin American Spanish such as Dominican Spanish (Toribio 2000) increasingly allow both overt and null subject pronouns in the same contexts (opposed to the use of the same term to describe 3rd person referential split in earlier literature), resulting in their labelling as partial null subject languages (PNSLs). Their existence has challenged the NSP as formulated by Rizzi (1982, 1986) and suggested the occurrence of null subject reduction. Given Latin American Spanish's rich history of language contact over the last five centuries, the observed decrease in use of null subjects is likely a result of contact-induced simplification. Following Trudgill's (2011) sociolinguistic typology, this contact would be characterized as short-term adult second-language acquisition which is consistent with the scenario for the African slaves brought over to the Americas by the Spanish from the 16th century onward. Indeed, there are numerous non-standard varieties of Spanish spoken by Afro-Hispanic communities throughout Latin America. These Afro-Hispanic Languages of the Americas (AHLAs) have been described as the result of "conventionalized advanced second languages," much in line with Trudgill's proposal (Sessarego 2013, 2017). The L2-difficulty of the rules surrounding null subject use would have led adult learners to overproduce overt subject pronouns. Their errors would have then become nativized in the speech of the next generation. Similar to the Jespersen cycle for negation, the partial status of null subjects in varieties such as Dominican Spanish could reflect language in change as the variety moves from a null subject language to a non-null subject language á la van Gelderen's (2011) Subject Cycle. What is needed then is a robust diachronic analysis of null subjects.

To achieve this, I constructed a corpus of 57 texts from eight countries over four centuries to evaluate the diachronic rate of null subjects in Latin American Spanish. The corpus is supplemented by transcriptions of previous fieldwork in current AHLA varieties as well as historical texts written in AHLA vernacular. The aim was/is to determine whether the corpus supports the hypothesis that the rate of overt subject pronouns has risen significantly since contact with African L2-speakers. Broadly, we expected to see the highest variation in the form of more overt subject pronouns in varieties with the highest Afro-Hispanic populations. However, despite an attempt to balance the corpus by genre (equal literary and non-literary texts), preliminary data from the Dominican Republic, Bolivia, Panama, and Spain (as a control) has shown much variation, leaving no clear diachronic patterns (Figure 1). In order to see if something similar to but independent of genre could be affecting the pronominal data, Rosemeyer's (2019) method of measuring orality levels in Portuguese plays was adapted for use in this corpus. This ORSCORE was found to have a significant ($p=.001$) relationship with overtness rates (Figure 2²). With a view to account for this obstruction, we set out to compare forms of the private verbs *creer* 'to believe/think' and *pensar* 'to think' which should lessen the influence of orality as its frequency in a text is one of the five criteria used to measure orality. Unfortunately, our corpus is too small to restrict the data to one lexical item, so we are now in the process of sifting through data from the *CORDIAM* and *CORDE/CREA* corpora. We hope to tease apart this orality effect to find underlying diachronic trends.

¹ For consideration as a talk or poster.

² There are two outlying texts with very high ORSCORES (1.77 and 2.35) that might seem to skew the data. However, the best-fit line remains the same when they are excluded.

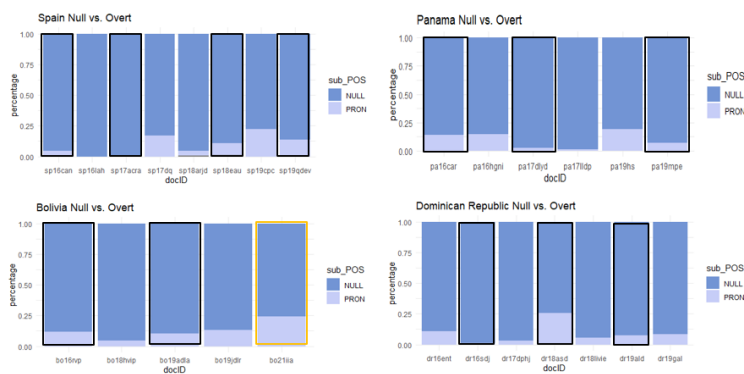
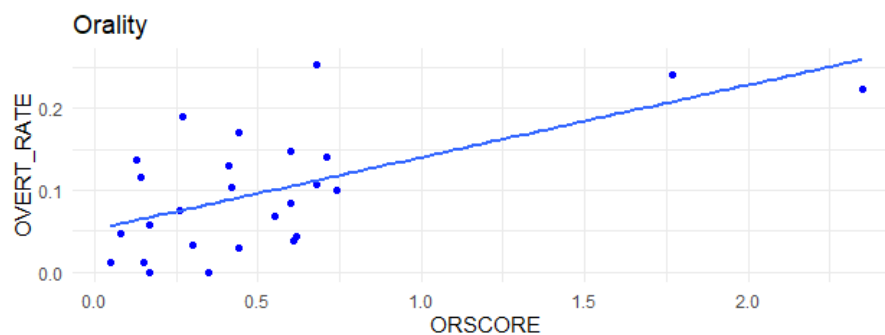
Figure 1:³ Null vs. Overt Subject Pronouns

Figure 2: Degree of orality plotted against proportion of overt subject pronouns per text



References:

- [CORDIAM] Academia Mexicana de la Lengua, *Corpus Diacrónico y Diatópico del Español de América*, www.cordiam.org
- Gelderen, Elly van. 2011. *The Linguistic Cycle*. Oxford: OUP.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>>
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <<http://www.rae.es>>
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17: 501–57.
- Rosemeyer, Malte. 2019. "Actual and apparent change in Brazilian Portuguese wh-interrogatives." *Language Variation and Change* 31(2): 165–191. CUP.
- Sessarego, Sandro. 2013. "Afro-Hispanic Contact Varieties as Conventionalized Advanced Second Languages". *IBERIA* 5 (1): 99-125.
- Sessarego, Sandro. 2017. "Revisiting the Null Subject Parameter: New Insights from Afro-Peruvian Spanish". *Isogloss* 3 (1): 43-68.
- Toribio, Almeida J. 2000. "Setting parametric limits on dialectal variation in Spanish". *Lingua: International Review of General Linguistics* 110 (5): 315–341.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: OUP.

³ The docID corresponds to the country, century, and title of a text (e.g. dr16ent = Dominican Republic, 16th century, Entremés). The document genre is bolded, and the golden bolded text is a transcript from Afro-Bolivian vernacular.

Persistence and Change of Colexifications in Indo-European

Languages differ in how they organize their lexicon: two given concepts can be expressed by two different word forms in one language, but by the same form in another language. This encoding of different concepts with the same word form is known as colexification (François 2008). Studies have shown that highly associated concepts are colexified more frequently across languages (Xu et al. 2020), and that colexification patterns are less genetically stable than the phonological form of vocabulary, at least in Europe (Gast & Koptjevskaja-Tamm 2022). Colexification patterns can also be borrowed, as various examples of areally common colexifications show (e.g., Schapper 2022 on the colexification of BONE and STRENGTH in Melanesia or Segerer & Vanhove 2022 on the colexification of color terms in Africa).

The present study is concerned with colexification in the Indo-European language family, and its relationship to lexical material. It aims to answer the question of how strongly a colexification is tied to the phonological material its concepts are expressed with, and if it is likely to persist when phonological material is replaced. To investigate this, we use vocabulary lists from the CLICS database (Rzymiski et al. 2021), and etymology data from IELex (Chang et al. 2015) in order to split the vocabulary items into cognate sets.

When relating colexifications to etymological data, there are two main options:

- (i) Colexifications are independent of etymology and are equally found among unrelated lexemes denoting the same concept.
- (ii) Colexifications are dependent on etymology and are more frequently found with lexemes of a certain etymology.

(i) suggests that colexifications arise, persist, and spread independently of phonological substance, and that, more generally, the development of patterns is unrelated to the development of matter. (ii) on the other hand, would suggest that colexifications arise rarely and are then bound to a certain phonological form. In this scenario colexifications mostly spread and persist together with phonological matter.

As it has been shown that certain colexifications are more common cross-linguistically even among unrelated languages in different areas, we expect that colexifications in the Indo-European family are largely (but not completely) independent from the etymology of their word forms.

In a further stage of this study, we plan to use phylogenetic methods in order to investigate the history of certain colexifications in the Indo-European language family. We also plan to conduct a similar study on Austronesian languages in order to compare language families that differ strongly in terms of culture and geographical environment.

Literature

Chang, Will, David Hall, Chundra Cathcart, and Andrew Garrett. "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis." *Language* (2015): 194-244.

François, Alexandre. "Semantic maps and the typology of colexification." *From polysemy to semantic change: Towards a typology of lexical semantic associations* 106 (2008): 163.

Gast, Volker, and Maria Koptjevskaja-Tamm. "Patterns of persistence and diffusibility in the European lexicon." *Linguistic Typology* 26, no. 2 (2022): 403-438.

Rzymiski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast et al. "The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies." *Scientific data* 7, no. 1 (2020): 13.

Schapper, Antoinette. "Baring the bones: The lexico-semantic association of bone with strength in Melanesia and the study of colexification." *Linguistic Typology* 26, no. 2 (2022): 313-347.

Segerer, Guillaume, and Martine Vanhove. "Areal patterns and colexifications of colour terms in the languages of Africa." *Linguistic Typology* 26, no. 2 (2022): 247-281.

Xu, Yang, Khang Duong, Barbara C. Malt, Serena Jiang, and Mahesh Srinivasan. "Conceptual relations predict colexification across languages." *Cognition* 201 (2020): 104280.

PARADIGMATIC REDUNDANCY IN THE COMPLEMENT SYSTEM OF BASQUE

While most languages do not seem to have more than four types of complement clauses (Noonan 2007: 147), in Basque at least seven have to be distinguished. There is, however, a large overlap in the distribution of finite and non-finite types, which raises the question if Basque acquired finite complement clauses as a result of contact with Romance languages.

Basque has been in increasingly intense contact with Romance languages for more than two millennia, but only the last 500 years are documented in texts. Thus most instances of Romance influence cannot be observed directly but have to be inferred.

The Basque inventory of complement types consists of finite indicative and subjunctive clauses, as well as verbal nouns (VN) inflected in different cases and showing different morphosyntactic behaviour, and the infinitive-like perfective participle. Table 1 is a very rough representation of the mapping between types of matrix predicates and types of complement clauses. It shows that finite clauses are not only restricted to two classes of matrix predicates but also generally replaceable with a non-finite alternative without or with only a subtle change in meaning. The only exception to this are some utterance and cognition verbs like *esan* ‘say’ or *pentsatu* ‘think’, for which it is cross-linguistically not uncommon to take only direct speech complements (Dixon 2006: 28). Finite complements in Basque, thus, strictly speaking add nothing to the system besides a syntactically more integrated alternative to direct speech.

| | | | | | | | |
|--|--|---|-----------------------|-------------|----------------------|------------|-------|
| utterance, cognition, propositional attitude | | manipulation | intention | implicative | immediate perception | phasal | modal |
| indicative | | subjunctive | | | | | |
| VN in structural case | | purposive VN, infinitival interrogative | VN in structural case | inessive VN | | infinitive | |

Table 1: Mapping between types of matrix predicates and finite and non-finite types of complement clauses.

This synchronic redundancy suggests that one class of complement clauses intruded into contexts occupied by the other. There are two possibilities:

Hypothesis A: Finite types spread into contexts occupied by non-finite ones. Having non-finite rather than finite complement clauses would fit the typological profile of Basque as a head-final language. In fact, OV languages tend to have preverbal complement clauses (Schmidtke-Bode/Diessel 2017: 10) and preverbal complement clauses are more commonly non-finite than finite (ibid.: 12-19). In the Romance contact languages, on the other hand, finite complement clauses are frequently used, and a subjunctive with similar functions as in Romance started to develop in Basque only at the beginning of the writing tradition (Mounole 2014: 328). Finite complementation in general seems, however, to be quite old since it was already well established at the time of the first texts, and the complementisers used are neither borrowed nor calqued from Romance.

Hypothesis B: Non-finite types spread into contexts occupied by finite ones. This would mean that the complex polypersonal forms of finite verbs are being replaced by simpler forms, a phenomenon that can also be observed in some other domains, like the potential mood being replaced by analytic constructions with *ahal izan* ‘be able’. While this is a plausible scenario for the subjunctive forms, which are very rarely found in non-embedded clauses and can thus almost completely be dispensed with when substituted by other subordinate forms, the indicative forms are less dispensable since they are still used in independent clauses.

Thus, a combination of both scenarios seems most plausible, namely a spread of finite complement clauses (whether they had developed independently or as a consequence of contact) replicating the Romance model and, probably more recently, a gradual substitution of subjunctive complements by non-finite forms.

REFERENCES

- Dixon, R. M. W. (2006): “Complement Clauses and Complementation Strategies in Typological Perspective.” In: Dixon, R. M. W. / Aikhenvald, Alexandra (eds.): *Complementation: A cross-linguistic typology*. Oxford: Oxford University Press. 1-48.
- Noonan, Michael (2007): “Complementation.” In Shopen, Timothy: *Language Typology and Syntactic Description*, vol. 2. Cambridge: Cambridge University Press. 52-150.
- Mounole, Céline (2014): *Le verbe basque ancien: étude philologique et diachronique*. ASJU 48. Bilbo: UPV/EHU.
- Schmidtke-Bode, Karsten / Diessel, Holger (2017): „Cross-linguistic patterns in the structure, function, and position of (object) complement clauses.” *Linguistics* 55(1): 1-38.

Doing Conversation Analysis in Latin: The Case of Hedging

Rules underlying conversation and linguistic phenomena specific to interaction, have received a lot of attention, especially in the areas of conversation analysis¹ and pragmatics of conversation²: we now know that, while superficially messy, conversation follows a set of underlying rules which all speakers know and are in principle expected to adhere to: rules about when to take the conversational floor; desirable and undesirable responses; strategies to convey one's undivided attention to one's co-interactant, etc. When these rules are intentionally or unintentionally breached, for instance when an undesirable response is given—such as declining an invitation or providing an uninformative response—the situation calls for various mitigating strategies to avoid giving offence. One of these strategies is *hedging*—modifying one's commitment to the truth of one's statement.

In this paper, I shall look at hedging behaviour in Latin. Hedging has received some attention in Latin, notably in terms of politeness and language characterization,³ but less from the point of view of Conversation Analysis. Using such sources as Plautus, Terence, Cicero and Petronius, and methodology developed within Conversation Analysis, I will address the following questions:

- Which expressions are used as hedges in Latin? Are they similar to hedges in other languages?
- In what contexts do they occur—i.e., which antecedents trigger them?
- Does the choice of hedging depend on the context of production (addressee, purpose of production, time) or genre (comedy, correspondence, novel)?
- What, if anything, does hedging tell us about spoken Latin? In other words, is Conversation Analysis applicable to Latin sources?
- Do the sources in Latin reflect hedging behaviour which has been shown to obtain on other languages?

Cross-cultural and cross-linguistic research has shown that hedging behaviour depends on cultural and cognitive factors. To gain a better understanding of hedges (linguistic devices) and hedging (communicative strategy), it is important that large bodies of data in different languages and cultures be studied. Latin, a large-corpus language, provides a wealth of material to study this phenomenon in detail and thus enrich our understanding of cognitive underpinnings of hedging, of cultural differences and commonalities, and of the interaction of

¹ See, for instance, Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language* 50, 696–735.

² See, for instance, Haugh, M. (2012). Conversational interaction. In K. Allan & K. Jaszczolt (Eds.), *The Cambridge Handbook of Pragmatics* (Cambridge Handbooks in Language and Linguistics, pp. 251-274). Cambridge: Cambridge University Press.

³ Barrios-Lech, Peter (2016). *Linguistic Interaction in Roman Comedy*. Cambridge, Cambridge: University Press: 142–165;

Mencacci, Francesca (2022). 'Text as Interaction: Vt Mihi (Quidem) Videtur as a Hedging Device in Latin Literary Texts.' In *Politeness in Ancient Greek and Latin*, edited by Luis Unceta Gómez and Łukasz Berger, 103–126. Cambridge: Cambridge University Press.

hedges and genre. This paper is intended as a contribution to the growing cross-cultural body of research on hedging.

It Ain't Over till It's Over. Bilingualism and language decay in Sicilian inscriptions

Inscriptions and ancient sources both indicate that Sicily has been characterized by multilingualism since before the classical period, when Greek, Punic, and indigenous languages were spoken in several areas of the island, and often coexisted in some areas. However, Sicilian indigenous languages disappeared with the end of the classical period and Punic survived only (and very scarcely) only until 1st c. BCE circa. Greek, on the other hand, is the only language attested in inscriptions throughout the entire Sicily, and it remained so until the Roman conquest, when Latin spreads across the island. During the Roman period, and until the late antiquity, Greek/Latin bilingualism was common in Sicily, although with important diatopic differences: e.g., in Syracuse (Korhonen 2012) and surrounding areas Greek clearly remained the most prestigious variety until at least the imperial period, while it is likely that rural areas were predominantly Latin-speaking (Korhonen 2016). Finally, and famously, Greek must have undergone language death in western Sicily during the early Middle Ages.

While it is relatively uncomplicating determining when a language no longer appears in inscriptions, establishing when it ceased to be spoken is a rather complicated task. Even if a language is not inscribed on a stone, it might still be present in some parts (e.g., strata) of a society, and not anymore used in inscriptions because of the increasing loss of prestige, or because of political reasons. In these cases, it is sometimes possible to determine the presence of these not explicitly represented language through the presence of language contact. For giving a famous Sicilian example, it has been convincingly suggested that behind the bilingual (Greek-Latin) inscription ISic000470 there is a Punic speaker (Susini 1968; Tribulato 2011). In other words, while a language is undergoing decay, and before it completely disappears as a spoken variety, it might still be detected through language contact. For what concerns ancient Sicily, I argue that we can access this information through what I call “implicitly bilingual” inscriptions, i.e., monolingual Greek inscriptions that show phenomena of language contact.

Despite the great importance of the topic of language decay and death, there are very few works on Greek language death (e.g., JANSE 2003), and even fewer on Punic and Sicilian indigenous languages (partially, POCETTI 2012 and MARCHESINI 2012), and no-one has yet exploited bilingual inscriptions to investigate the reasons for language death and language vitality. In this paper, I will use “implicitly bilingual” inscriptions to assess the dynamics of language decay and language death between Punic and Greek and between indigenous languages and Greek, and to assess the possibility that Punic has been spoken longer than it is attested. Furthermore, I will adopt this model to analyze Latin inscriptions from Western imperial Sicily that show phenomena of language contact with Greek, in order to assess the Greek language decay in areas that we know Greek was the minority language.

AMADASI GUZZO, M. G. 1990, *Iscrizioni fenicie e puniche d'Italia*, Roma

JANSE, M. 2003, Introduction: Language death and language maintenance. Problems and prospects. In: Janse, M. & Tol, S. (eds.), *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches*, Oxford, xi–xvii.

KORHONEN, K. 2011, Language and Identity in the Roman Colonies of Sicily. In: Sweetman, R. (ed.) *Roman Colonies in the First Century of Their Foundation*, Oxford, 7–31.

KORHONEN, K. 2016, *Questioni del bilinguismo siciliano antico in lunga durata*, «Linguarum Varietas» 6: 115–120.

MARCHESINI, S. 2012, The Elymian language. In: Tribulato, O. (ed.) *Language and Linguistic Contact in Sicily*, Cambridge, 95–114.

POCETTI, P. 2012, Language relations in Sicily: evidence for the speech of the Σικανοί Σικελοί and others. In: Tribulato, O. (ed.) *Language and linguistic contacts in ancient Sicily*, Cambridge, 49–94.

SUSINI, G. 1967. *Il lapicida romano*. Roma.

TRIBULATO, O. 2011, *The stone-cutter's bilingual inscription from Palermo (CIL X 7296 = IG XIV 297): a new interpretation*, «ZPE» 177: 131–40.

The emergence of word-initial voiced stops in Proto-Hungarian

This presentation aims to bring new solutions to the old problem of the emergence of word-initial voiced stops in the prehistory of Hungarian. Word-initial voiced stops are frequent in modern Hungarian, but it is well-known that these arose during the relatively late prehistoric period, during the existence of Hungarian as an independent branch (Abaffy 2003: 122). It has been assumed that they emerged through sporadic changes $*p > b$, $*t > d$ (but not $k > *g$) in word-initial position in some inherited words, and became more frequent with the influx of Turkic and Iranian loanwords (Abaffy 1994: 16–18; Bárczi 1958: 113). However, it is problematic that in some cases Hungarian shows voiced stops even though the attested Turkic data points to voiceless stops (for example, Hu *gödény* ← reconstructed Western Old Turkic **güdän* or **kütän*, cf. Middle Turkic *kutan* ‘a kind of bird’; Hu *dara* ← Turkic **dari* or **tari* cf. Old Turkic *tarig* ‘crops, millet, sowing’; Ligeti 1986: 193–194; Róna-Tas & Berta 2011: 1074–1075, 1078), and the more scarce Iranian loans also show a confusing picture, as corresponding voiced stops are found in some loans but not in others: Hu *bűz* ‘smell’ ← Iranian **bauda-* (> Ossetic *bud*, *bodæ* id.), Old Hungarian *kazdag* ‘rich’ ← Iranian (Alanic) **gæzdig* (> Ossetic *qæznyg*, *ğæzdug* id.) (Abaffy 1994).

It is also problematic that it is not known under which conditions the word-initial voiced stops arose in Hungarian in the first place, as these are not regular reflexes of any Proto-Uralic phonemes (contrary to word-internal voiced stops that developed regularly from the clusters of nasal and stop, Sammallahti 1988: 520). However, it has been claimed in presentations of Hungarian historical phonology that some inherited words have an irregular voiced reflex (Abaffy 2003: 118; Gerstner 2018: 107). The situation is somewhat similar as with the related Permic branch of the Uralic family, where similar voicing of has taken place (for example Proto-Permic **bur* < Proto-Uralic **para* ‘good’; Proto-Permic **dijn* < Proto-Uralic **tūji* ‘base of a tree’); however, in Permic this process seems to be a regular, conditioned change, voicing assimilation caused by word-internal voiced consonant (Csúcs 2005: 154–158), whereas in Hungarian the situation is often described as sporadic or no conditions are not given (Maticsák 2020: 360; Bereczki 2003: 69–70; Abaffy 2003: 118). It has been suggested that word-internal voiced consonants caused the voicing of word-initial stop (MSzFE: 105; UEW: 374–375; Honti 2017: 15), similarly as in Permic, but the situation is complicated as numerous counter examples can be found. The voicing of stops has been sometimes been considered a shared, areal innovation of Proto-Hungarian and Proto-Permic, but the evidence is inconclusive (Csúcs 2021: 44).

Most of the alleged examples of the sporadic voicing in inherited vocabulary are problematic etymologies, showing also irregular vocalism and semantic mismatch in many cases; for example, Hungarian *bal* ‘left’ allegedly reflects Proto-Uralic **palV* (MSzFE: 105; UEW: 351–352), but the only cognate is Udmurt Ud *pal’l’an* ‘left’ that shows irregular vocalism; Hu *bőr* ‘skin’ is allegedly a reflex of Proto-Uralic **perV* ‘bark’ (MSzFE: 110–111; UEW: 374), but *ő* is not the regular reflex of Proto-Uralic **e*. Some words have competing etymologies: for example, Hungarian *dug* ‘to squeeze’ has been considered both a loan from Turkic **dig-* ~ **tig-* ‘to squeeze’ (Róna-Tas & Berta 2011: 303–306) and a reflex of Proto-Uralic **tun̄ki-* (MSzFE: 135; UEW: 537–538). Because of this, the idea that Proto-Uralic stops can be reflected by Hungarian voiced stops in anlaut is very doubtful.

In this presentation, all the relevant Uralic etymologies and loanwords will be discussed, and the aim is to determine when and under what conditions the word-initial voiced stops became possible in Proto-Hungarian.

References

- Abaffy, Erzsébet 1994: Ósmagyar jövevényszavaink vállalomása egykorú fonémarendszerünkről. In: István Szathmári & Erzsébet E. Abaffy & Éva B. Lőrinczy (eds.), *Bárczi Géza emlékkönyv*. Budapest. 9–20.
- Abaffy, Erzsébet 2003: Ósmagyar kor. Hangtörténet. In: Jenő Kiss & Ferenc Pusztai (eds.), *Magyar nyelvtörténet*. Budapest. 106–128, 301–351, 596–609.
- Bárczi, Géza 1958: *Magyar hangtörténet*. Budapest.
- Bereczki, Gábor 2003: *A magyar nyelv finnugor alapjai*. Budapest.
- Csúcs, Sándor 2005: *Die Rekonstruktion der permischen Grundsprache*. Budapest.
- Csúcs, Sándor 2021: A másodlagos permi-magyar nyelvi érintkezések kérdése. In: László Klima & Attila Türk (eds.), *Párhuzamos történetek*. Budapest. 41–50.
- Gerstner, Károly 2018: Hangtörténet. In: Jenő Kiss & Ferenc Pusztai (eds.), *A magyar nyelvtörténet kézikönyve*. Budapest. 103–128.
- Honti, László 2017: *A magyar és a nyugati ótörök szókészleti kapcsolatairól*. Budapest.
- Ligeti, Lajos 1986: *A magyar nyelv török kapcsolatai a honfoglalás előtt és az Árpád-korban*. Budapest.
- Maticsák, Sándor 2020: *A magyar nyelv eredete és rokonsága*. Budapest.
- MSzFE = Gyögy Lakó 1971–1978: *A magyar szókészlet finnugor elemei*. Budapest.
- Róna-Tas, András & Árpád Berta 2011: *West Old Turkic. Turkic loanwords in Hungarian*. Wiesbaden.
- Sammallahti, Pekka 1988: Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In: Denis Sinor (ed.), *The Uralic languages*. Leiden. 478–554.
- UEW = Károly Rédei 1986–1991: *Uralisches Etymologisches Wörterbuch*. Wiesbaden.

The Evolution of Spatial Orientation Systems in Mayan and Nuristani

The diversity of spatial orientation systems in the world's languages has fascinated linguists, anthropologists and cognitive scientists alike (cf. Levinson 1998) and it has been taken as one of the few compelling arguments in favor of linguistic relativity (cf. Levinson 1996: 195-196).

Nevertheless, we still know little about how complex geomorphic orientation systems evolve diachronically or about the lexical sources that end up as elements of their paradigms. Palmer's (2015: 210) hypothesis that "a correlation will exist between a language's system of absolute spatial reference and the topography of the language locus" would suggest that certain environments favor the development of certain kinds of spatial systems. If this is the case, then it should be possible to compare the development of spatial orientation systems within language families that, e.g., have members both in mountainous and in flat environments.

In two case studies, we trace the evolution of spatial orientation paradigms from a variety of constructions involving motion verbs in Mayan languages and from adverbial formations in Nuristani languages. The fact that diverse sources lead to similar outcomes in unrelated languages spoken in similar environments, whereas related languages spoken in different environments do not develop the same amount of semantic distinctions, lends credence to the idea that the physical environment can under certain circumstances have a direct impact on linguistic structures and that linguistic coordinate systems "are constructed in response to the environment" (Palmer 2015: 210).

References

- Levinson, Stephen. 1996. Relativity in spatial conception and description. In: Gumperz, John & Stephen Levinson, *Rethinking Linguistic Relativity* (Studies in the Social and Cultural Foundations of Language 17), 177-202. Cambridge: Cambridge University Press.
- Levinson, Stephen. 1998. Studying spatial conceptualization across cultures: Anthropology and cognitive science. *Ethos* 26(1). 7-24.
- Palmer, Bill. 2015. Topography in language: Absolute frame of reference and the topographic correspondence hypothesis. In: De Busser, Rik & Randy LaPolla (eds.), *Language Structure and Environment. Social, Cultural and Natural Factors* (Cognitive Linguistic Studies in Cultural Contexts 6), 179-226. Amsterdam: John Benjamins.

Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture

A wide variety of work in both the Greenbergian typological tradition (Greenberg 1963, 1978; Hawkins 1979; Dryer 1989; Nichols 1992) and in the comparative-syntax paradigm (Baker 2001, 2008; Longobardi and Guardiano 2009; Roberts 2019) has established that different surface/typological features are frequently *correlated*; individual feature values can favour or disfavour others. Such correlations arise over historical time in a dynamical process in which both “vertical” genetic descent and “horizontal” contact processes play a role, and it has been suggested that the present-day global distribution of language types is reflective of the stationary state of this stochastic process (Greenberg 1978; Maslova 2000; Kauhanen et al. 2021; Jäger and Wahle 2021). This makes possible, among other things, the estimation of linguistic rates of change using a variety of techniques (Dediu and Cysouw 2013; Murawaki and Yamauchi 2018; Kauhanen et al. 2021).

What has received far less attention so far is the question of how *geographical* patterns of linguistic features relate to the purely linguistic properties of those features. Although Kauhanen et al. (2021) put forward a method for inferring rates of change of features from their present-day geospatial distributions, their method is restricted to statistically independent features. In this paper, using a set of word order features as our data, we show that feature correlations and geospatial distributions in fact stand in a systematic relationship. We show that typologically dispreferred types (such as the disharmonic combination OV & prepositions) tend to be surrounded by a greater variety of types than typologically preferred types (such as the harmonic combination OV & postpositions), which favour more uniform geographical environments. In particular, languages of dispreferred type are often found geographically “sandwiched” between languages of preferred type.

This *Sandwich Conjecture* is operationalized in terms of a notion of *neighbourhood entropy*, an information-theoretic measure of the extent of typological variability in a language’s immediate geographical neighbourhood. We apply this method to 28 word order feature pairs harvested from WALS (Dryer and Haspelmath 2013), and use an unrelated (and demonstrably uncorrelated) feature (presence/absence of voicing contrast) as a control feature. Measuring the typological correlation of two features with the usual ϕ coefficient (cf. Jäger and Wahle 2021), measuring “sandwichness” with neighbourhood entropy, and using permutation tests to factor out random noise, we then show that a statistically significant linear relationship exists between feature correlations and geospatial patterning: the more two features are correlated purely typologically (i.e. simply by virtue of the number of languages exhibiting the various possible feature combinations), the more sandwiched the corresponding geospatial distribution of types is. Hence, feature correlations are reflected in the geospatial patterning of said features.

This finding suggests a possible mechanism for the diachronic stability of dispreferred types, an otherwise surprising fact given that disharmonic feature combinations should be expected to resolve one way or the other into harmony over sufficient historical time. The Sandwich Conjecture explains this paradoxical stability by proposing that dispreferred types are preferentially found in typologically rich geographical environments, meaning that ample opportunities for horizontal transfer (contact) exist to sustain disharmony.

In addition to the above quantitative results, we present converging evidence from proof-of-concept computer simulations in which the same finding emerges.

References

- Baker, Mark. 2001. *The Atoms of Language*. New York: Basic Books.
- . 2008. “The Macroparameter in a Microparametric World.” In *The Limits of Linguistic Variation*, edited by Theresa Biberauer, 351–73. John Benjamins.
- Dediu, Dan, and Michael Cysouw. 2013. “Some Structural Aspects of Language Are More Stable Than Others: A Comparison of Seven Methods.” *PLOS ONE* 8: e55009.
- Dryer, Matthew S. 1989. “Large Linguistic Areas and Language Sampling.” *Studies in Language* 13: 257–92.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>.
- Greenberg, Joseph H. 1963. “Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- . 1978. “Diachrony, Synchrony, and Language Universals.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 1:61–91. Stanford, CA: Stanford University Press.
- Hawkins, John A. 1979. “Implicational Universals as Predictors of Word Order Change.” *Language* 55 (3): 618–48.
- Jäger, Gerhard, and Johannes Wahle. 2021. “Phylogenetic Typology.” *Frontiers in Psychology* 12: 682132.
- Kauhanen, Henri, Deepthi Gopal, Tobias Galla, and Ricardo Bermúdez-Otero. 2021. “Geospatial Distributions Reflect Temperatures of Linguistic Features.” *Science Advances* 7: eabe6540.
- Longobardi, Giuseppe, and Cristina Guardiano. 2009. “Evidence for Syntax as a Signal of Historical Relatedness.” *Lingua* 119 (11): 1679–1706.
- Maslova, Elena. 2000. “A Dynamic Approach to the Verification of Distributional Universals.” *Linguistic Typology* 4: 307–33.
- Murawaki, Yugo, and Kenji Yamauchi. 2018. “A Statistical Model for the Joint Inference of Vertical Stability and Horizontal Diffusibility of Typological Features.” *Journal of Language Evolution* 3: 13–25.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

Exploiting phylogenetic modeling to uncover directionality in the emergence of universals

Extensive investigations into word order and hierarchical universals have uncovered well-supported cross-linguistic tendencies for certain features of languages to go hand in hand (Greenberg 1963, 1966, Keenan & Comrie 1977, Kozinsky 1981, Hawkins 1983, Dryer 1992, Jäger and Wahle 2021, Authors). Many linguists hold that such universals are to be accounted for in terms of diachrony, i.e. in the words by Bickel et al. (2015: 29), "most – perhaps even all – statistical universals are not really synchronic in nature, but are rather the result of underlying diachronic mechanisms that cause languages to change in preferred or 'natural' ways". However, we still know very little about these 'diachronic mechanisms'.

Here, we initiate an investigation into possible mechanisms through which correlated change proceeds in a strict quantitative, phylogenetic context. We consider for a set of highly supported word order universals (including the correlation pairs identified by Dryer 1992) and highly supported hierarchical universals (including well-known dependencies in the (morphological) marking of person, number and gender) which of the two features changed first. For example, in the well-supported universal "In languages with prepositions, the genitive almost always follows the governing noun" (Greenberg's 1963 Universal no. 2), which word order feature changed first: the order of adposition and noun or the order of genitive and noun?

The data that we use comes from Grambank (Skirgård et al. 2022), a large typological database featuring morphosyntactic data on 2400+ languages. The global language tree we use has been created by Bouckaert et al. (2022). We use two methods to investigate directionality in correlated evolution. First, Nunn and Cooper's (2015) so-called "species-pairs evolutionary lag test" (SPELT), which explicitly tests whether change in one feature lags behind change in another feature. Second, ancestral state estimations at various time depths are contrasted to show which feature changed to the state predicted by the universal first. This is done using *BayesTraits* (Meade and Pagel 2022).

We envision a number of possible outcomes of this study. First of all, we may find that there are no global patterns at all: to go back to Greenberg's (1963) Universal no. 2, we may find that it is (1) sometimes the order of genitive and noun that changes first, and the order of adposition and noun that changes second, and (2) sometimes the other way around. Secondly, we may find that there is some evidence for directed change, but only for a subset of the universals that we investigate. Thirdly, we may identify strong patterns of directionality, possibly unveiling what are the central pivot features (potentially including the order of object and verb, Dryer 1992) that trigger change in related features. Confirming or rejecting this third hypothesis would advance our understanding of the causal factors in grammatical change and also point toward explanations for cross-linguistic tendencies that all syntactic theories should be able to account for.

References

Authors <>

Bickel, Balthasar, Alena Witzlack-Makarevich, Taras Zakharko, and Giorgio Iemmolo. 2015. 'Exploring Diachronic Universals of Agreement: Alignment Patterns and Zero Marking across Person Categories'. In *Agreement from a Diachronic Perspective*, edited by Jürg Fleischer, Elizabeth Rieken, and Paul Widmer, 29–51. Berlin: De Gruyter Mouton.

Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate E Jones, and Quentin Atkinson. 2022. 'Global Language Diversification Is Linked to Socio-Ecology and Threat Status'. Preprint. SocArXiv. <https://doi.org/10.31235/osf.io/f8tr6>.

Dryer, Matthew S. 1992. 'The Greenbergian Word Order Correlations'. *Language* 68 (1): 81–138.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of grammar*. Cambridge: The MIT Press.

Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. (Janua Linguarum, Series Minor, 59.) The Hague: Mouton.

Hawkins, John A. 1983. *Word Order Universals*. San Diego: Academic Press.

Jäger, Gerhard, and Johannes Wahle. 2021. 'Phylogenetic Typology'. *Frontiers in Psychology* 12: 682132.

Keenan, Edward L., and Bernard Comrie. 1977. 'Noun Phrase Accessibility and Universal Grammar'. *Linguistic Inquiry* 8 (1): 63–99.

Kozinsky, Isaak (1981). *Nekotorye grammaticheskie universalii v podsistemax vyrazenija subjektno-objektnyx otnosenij*. [Some grammatical universals in subsystems of expression of subject-object relations.] Doctoral dissertation, Moskovskij gosudarstvennyj universitet.

Meade, Andrew, and Mark Pagel. 2022. *BayesTraits V4.0.0*. Available at <http://www.evolution.reading.ac.uk/BayesTraitsV4.0.0/BayesTraitsV4.0.0.html>

Nunn, Charles L., and Natalie Cooper. 2015. 'Investigating Evolutionary Lag Using the Species-Pairs Evolutionary Lag Test (SPELT)'. *Evolution* 69 (1): 245–53.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bownen, Patience L Epps, Jane Hil, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu H. Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L. M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tânia R.A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya, Samamé Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O.C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabach, Frederick W.P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉婧婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin Atkinson & Russell D. Gray 2022. "Grambank Reveals the Importance of Genealogical Constraints on Linguistic Diversity and Highlights the Impact of Language Loss." SocArXiv. December 4. [doi:10.31235/osf.io/grh45](https://doi.org/10.31235/osf.io/grh45).

Periphrastic perfects reflect the lexical semantic distinctions of their auxiliaries

Compound perfect constructions commonly develop from resultative constructions, often with auxiliaries glossed as ‘be’ or ‘have’. Some formally similar constructions, however, exhibit important features not typically associated with perfects. In Portuguese, for instance, the ‘have’ + participle construction with a present tense auxiliary marks recent habitual action: *Tenho lido muitos artigos* ‘I have been reading many articles’ (not ‘I have read many articles’).

This paper contrasts the Portuguese periphrasis with *ter* (< Latin TENĒRE) + perfective participle with other Romance compound perfects formed with reflexes of Latin HABĒRE (cf. Harris 1982) and contextualizes it with regard to the effects lexical semantics on grammaticalization patterns involving semantically similar source material. I argue that previous analyses of auxiliaries have erroneously equated lexical verbs of similar meanings and consequently overlooked certain inference patterns that contribute to semantic change in grammaticalization (cf. Traugott & Dasher 2002). I also address the ongoing need for greater terminological care in this area.

While scholars such as Bybee et al. (1994) tend to treat sources of auxiliaries as equivalents if they can be glossed with a single English verb, Juge has argued (2002) that such an approach incorrectly predicts nearly identical outcomes of the grammaticalization of seemingly similar constructions that in fact result in significantly different structures. In the case of the Portuguese resultative construction, the key factor is the distinction between Latin HABĒRE and TENĒRE, both commonly glossed ‘have’.

These two verbs were not exact synonyms in Latin and did not yield exact synonyms in the Romance languages (cf. Harre 1991). While both indicate possession, TENĒRE and its reflexes also mean ‘hold’ and ‘keep’. The latter sense is a key factor in the development of Portuguese *ter* as an auxiliary. Certain predicates, when paired with a verb meaning ‘keep’, favor an iterative reading. If I assert that I keep my grass cut, for example, someone may conclude that I mow it periodically. For this reason, the label ‘iterative’ fits Portuguese *ter* + participle better than ‘perfect’.

Indeed, the term ‘perfect’ itself presents certain complications. First, it is often conflated with the aspectual term ‘perfective’, although some languages, such as Catalan, clearly show that these are orthogonal categories. Therefore in many cases the term ‘anterior’ is preferable to ‘perfect’ (cf. Bybee et al. (1994), among others; e.g., English future anterior *She will have arrived*).

Cruse (1986) suggests analyzing what he calls a lexeme’s sense-spectrum, or the collection of senses it encodes. For example, Spanish *mismo* has a range of senses that correspond to some of the senses of the English lexemes *same*, *very*, *right*, *oneself*, and *exactly*. This approach facilitates the identification of more fine-grained semantic relations that shape patterns of lexicalization and grammaticalization and discourages treating similar lexemes in different languages as being more similar than they are, which is an especially common problem in cases involving verbs of motion (cf. Juge 2007).

A widely recognized characteristic of verbs that become grammaticalized is that such lexemes are often highly polysemous before becoming grammaticalized and that they then show greater polysemy as a result of undergoing grammaticalization. The nature of the polysemy patterns shown by such lexemes, however, is still not well understood. Close analysis of the lexical semantics of grammaticalization—including interactions with factors like pragmatic inferencing—in familiar languages with well-attested histories allows typologists and historical linguists to more accurately apply insights gained from the examination of constructions in these languages to those in more poorly documented languages.

References

- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Dias da Costa, Albano. 1976. Periphrastic expressions in Portuguese. In Jürgen Schmidt-Radefeldt (ed.), *Readings in Portuguese linguistics*, 187-243. Amsterdam: North-Holland Publishing.
- Emanatian, Michele. 1992. Chagga ‘come’ and ‘go’: Metaphor and the development of tense-aspect.” *Studies in Language* 16.1–33.
- Harre, Catherine. 1991. *Tener plus past participle: A case-study in linguistic description*. London: Routledge.
- Harris, Martin. 1982. The ‘past simple’ and the ‘present perfect’ in Romance. In Nigel Vincent & Martin Harris (eds.), *Studies in the Romance verb*, 42-70. London: Croom Helm.
- Juge, Matthew L. 2002. Polysemy patterns in semi-lexical items in Spanish and Catalan. *Studies in Language* 26(2). 315-335.
- Juge, Matthew L. 2007. Metaphor and teleology do not drive grammaticalization. In Joseph C. Salmons & Shannon Dubenion-Smith (eds.), *Historical linguistics 2005: Selected papers from the 17th International Conference on Historical Linguistics, Madison, 31 July-5 August 2005*, 33-48. *Current Issues in Linguistic Theory*, 284. Amsterdam: John Benjamins.
- Parkinson, Stephen. 1988. Portuguese. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 131-169. New York: Oxford University Press.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Keywords

grammaticalization
 lexical semantics
 perfect
 anterior
 periphrasis

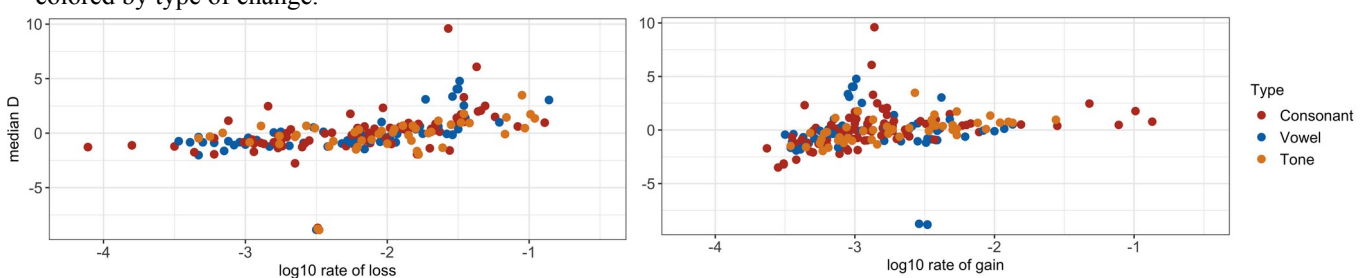
Is tone change more rapid and irregular than segmental change? - A Mixtec case study

Despite the abundance of tonal languages around the world (Yip 2002), the diachrony of tone is still poorly understood, especially when compared to segmental sound change (Campbell 2021). This is even more so regarding tone change *per se* as opposed to tonogenesis (Ferlus 2004, Dockum 2019, among others). This *lacuna* has contributed to the assumption that tones are inherently unstable and can change unpredictably (Ratliff 2015). In this talk, I address the questions of whether tones change faster than segments and whether they can be used to inform subgrouping in the Mixtec languages of southern Mexico (Otomanguean). All Mixtec languages exhibit complex systems of lexical and grammatical tone that have to be reconstructed to the proto-language (Dürr 1987, Swanton & Mendoza Ruíz 2021) and most probably all the way back to proto-Otomanguean (Rensch 1976, Campbell 2021). As such, these languages provide an ideal case study for testing assumptions about tone change.

I created a database of tonal and segmental sound changes across a sample of 42 Mixtec languages. The changes were identified based on cognate sets derived from a 209-item basic vocabulary list. All entries were converted to IPA and standardized with regard to the tone notation for consistent identification of sound changes. For each cognate set, I reconstructed a proto-form (both tones and segments) applying the comparative method and incorporating previous reconstructions where available (Josserand 1983, Dürr 1987, Swanton & Mendoza Ruíz 2021). I established tone correspondences and tone changes across the 42 languages of the sample applying the comparative methods as with segments. The results of this work are stored in multiple, interlinked databases that can be expanded and re-used for other research questions in the future. Based on a posterior distribution of phylogenetic trees from a previous study (AUTHOR et al. submitted), I calculated phylogenetic signal with the metric D (Fritz & Purvis 2010) and estimated rates of gain and loss with a Hidden Markov Model (Beaulieu et al. 2013) for each segmental and tonal change identified.

The results, summarized in Fig. 1, show that tone change in Mixtec does not behave differently from segmental change in any significant way. Many tone changes carry phylogenetic signal and can thus contribute to our understanding of the internal structure of this language family just like segmental changes. Tones also do not change faster or slower than segments overall, exhibiting similar transition rates as segments. These two measures suggest that tone change operates much the same way as segmental change and should be investigated on a par with segmental change.

Fig. 1: Median D (phylogenetic signal, y-axis) and rates of gain and loss (change rate, x-axis) per sound change colored by type of change.



References

- AUTHOR et al. submitted. [Title withheld to ensure blind review]
- Beaulieu, Jeremy, O'Meara, Brian & Donoghue, Michael. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic biology*, 62(5), p.725–737.
- Campbell, Eric W. 2021. Why Is Tone Change Still Poorly Understood, and How Might Documentation of Less-studied Tone Languages Help? In: Epps, Law, Pat-El (eds.). *Historical Linguistics and Endangered Languages*. Routledge
- Dockum, Ricker. 2019. *The tonal comparative method: Tai tone in historical perspective*. Doctoral dissertation: Yale University
- Dürr, Michael. 1987. A preliminary reconstruction of the Proto-Mixtec tonal system. *Indiana* 11. 19–61.
- Ferlus, Michel. 2004. The origin of tones in Viet-Muong. In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001* (pp. 297–313). Arizona State University Programme for Southeast Asian Studies Monograph Series Press.
- Fritz, Susanne & Andy Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4), p.1042–1051.
- Josserand, J. Kathryn. 1983. Mixtec dialect history. Tulane University PhD dissertation.
- Ratliff, Martha. 2015. Tonoexodus, tonogenesis, and tone change. In: Honeybone, Patrick & Joseph Salmons (eds.). *The Oxford Handbook of Historical Phonology*. Oxford University Press. pp. 245-261.
- Rensch, Calvin R. 1976. *Comparative Otomanguean Phonology*. Bloomington: Indiana University Press
- Swanton, Michael & Juana Mendoza Ruíz. Observaciones sobre la diacronía del tono en el Tu'un Savi (mixteco) de Alcozauca de Guerrero. In Arellanes & Guerrero (eds.), *Estudios lingüísticos y filológicos en lenguas indígenas mexicanas: Celebración de los 30 años del Seminario de Lenguas Indígenas*. México, D.F: Universidad Nacional Autónoma de México
- Yip, Moira. 2002. *Tone*. Cambridge University Press.

Beyond the paradigm: Change and expansion in Thai pronominal reference

This paper examines the historical development of personal pronouns in Thai. Building on work examining etymology and evolving dictionary definitions of Thai pronouns beginning in the 19th century (Author xxx), this study expands its scope to all available pedagogical and reference materials about Thai, ranging from the 17th century to the mid-20th century. We show how the set of pronouns has expanded, and how both the denotative meanings and the social context of their usage has changed over the centuries.

Multiple synchronic studies have explored the social dimensions of Thai pronouns (Cooke 1965, Palakornkul 1972, Saisuwan 2016). Saisuwan, for example, theorizes pronoun selection along dimensions of formality, power, and social distance. Thai is notable for its lack of a single pronominal paradigm, or even set of paradigms; many options are available for every cell in a conventional pronoun paradigm, each with distinct socioindexical meaning. Lexical items that can behave pronominally in Thai include proper names, occupational titles, social status terms, and kinship terms, in addition to a rich assortment of ‘true’ pronouns, including both ones inherited from Proto-Tai and ones borrowed from Khmer, Pali, Chinese, and English.

Here we focus on materials produced for second-language learners of Thai, because they often provide more direct explanation of social dimensions of pronoun usage. Some primary material from early written Thai texts (13th-15th centuries) is also surveyed in order to help establish which terms were in use before European contact. Works in our survey include grammatical descriptions (*e.g.*, de La Loubere 1693, Low 1828, Buell 1840, Jones 1842, Pallegoix 1850, Noss 1954), dictionaries (*e.g.*, Jones 1833, Caswell 1846, Pallegoix 1854, Bradley 1873, Michell 1892, Hays 1894, Pallegoix and Vey 1896, Cartwright 1907, Virajbhakaya 1924, McFarland 1939, Haas 1947, 1964), textbooks and primers (*e.g.*, McFarland 1900, Frankfurter 1900, Cartwright 1906, Haas & Subhanka 1945, Noss 1964).

Reasons the history of Thai pronouns makes an interesting historical question include:

(1) A pronoun paradigm is robustly reconstructible for Proto-Tai, with reflexes across the full geographic spread of the Tai branch of the Kra-Dai family (Matsuyama 1962, Benedict 1975, Strecker 1984). This reconstructed paradigm has a three-way number distinction and three-way person distinction, plus clusivity distinction in the 1st person, and animacy distinction in the 3rd person. Thai retains just three of these in unmarked usage, indicating that Thai pronouns reflect relatively recent changes in the social structure of Thai society.

(2) Some facets of Thai pronominal usage beyond the paradigm are found cross-linguistically in Southeast Asia (*e.g.* Cooke 1965, comparing Thai, Burmese and Vietnamese). This indicates language contact as a major factor in the process of pronominal expansion, but we must also distinguish more recent areal phenomena from older ones.

Tracing changing social factors in the history of Thai pronoun usage helps us better understand how this kind of large typological shift can take place – from a closed pronoun paradigm to a relatively open system that relies heavily on situational context and awareness of one’s social relationship to everyone they interact with.

Selected references

Author. 2021. xxxxx

Benedict, Paul K. 1975. Austro-Thai language and culture with a glossary of roots. HRAF Press.

Cooke, Joseph R. 1965. Pronominal reference in Thai, Burmese, and Vietnamese. Ph.D. dissertation, University of California, Berkeley.

de La Loubere, Simon. 1693. A new historical relation of the kingdom of Siam [translated by A.P. Gen. R.S.S].

Low, James. 1828. A grammar of the T'hai, or Siamese language. Baptism Mission Press.

Matsuyama, Osamu. 1962. Kyôtsû taigo no ninshô daimeishi no taikai ni tsuite [On the system of personal pronouns in Common Tai]. Tokyo Gaikokugo Daigaku Ronshu 9:1-8.

Palakornkul, A. 1972. A socio-linguistic study of pronominal strategy in spoken Bangkok Thai. PhD dissertation, University of Texas at Austin.

Saisuwan, Pavadee. 2016. Male femininity in Thai among men who identify. PhD dissertation, Queen Mary University of London.

Strecker, David. 1984. Proto-Tai personal pronouns. Ph.D. dissertation, University of Michigan.

**The Postil Time Machine: “God help those who have begun writing down these books in Lithuanian”
(BP1591 II 77,2–4)**

JOLANTA GELUMBECKAITĖ (Frankfurt/Main), ØYVIND EIDE (Köln), MORTIMER DRACH (Frankfurt/Main), CHRISTIAN CHIARCOS (Augsburg), MAX IONOV (Köln)

The *Postil Time Machine* (PosTiMe, 2021–2025) is a novel research project on knowledge transfer in 16th-century Europe; specifically, on the situation in Lithuania Minor, the Lithuanian-speaking part of the Duchy of Prussia. It aims to reveal the intratextual and intertextual relationships of the Old Lithuanian Lutheran postils – the *Wolfenbüttel Postil* (WP1573/1574) and the *Bretkūnas Postil* (BP1591) – and their relationship to their Latin and German translation sources and models. These postils are collections of sermons for the Sundays and feast days of the liturgical year (Gelumbeckaitė 2017). Early Modern postils as “the applied distillation of Christianity delivered on a regular basis by the clergy to the laity” (Frymire 2010: 443) were broadly influential texts offering references not only to Biblical texts and the Church Fathers, but also the classics, theologians and historians, bridging scholarly and popular culture. The main objective of PosTiMe is the genetic and historical linguistic study of these Lithuanian texts and their presentation in a digital *time machine*-type publication (cf. Kaplan 2013).

Previous work on Old Lithuanian in the Digital Humanities was focused mainly on digitizing Old Lithuanian texts and making them searchable (e.g. in TITUS and the *Senieji raštai* database). More recently, linguistically deeply annotated corpora have been created (SLIEKKAS, CorDon). The *Postil Time Machine* follows this corpus-linguistic tack while contributing new data, a systematic philological edition of its texts, and a visualization of inter-textual relations.

Specifically, this interdisciplinary linguistic-philological and technological project consists of the following tasks: (1) the identification of the construction principles and translation strategies of the Lithuanian texts, (2) the study of the textual relations of the postils both with the translation sources and among each other, (3) the detection of intra- and intertextual references as well as the alignment with the so-identified originals using machine translation methods (which necessitates the creation of a language technologies stack for Old Lithuanian), (4) the linguistic-historical interpretation of their contents in the form of a linguistically deeply annotated reference corpus, (5) the modeling as a graph (and the development of a Linked Open Data interface in RDFa based on and connected to the TEI documents, cf. Tittel et al. 2018 and Chiarcos & Ionov 2018) as well as the representation of these structures as interactive visualizations, and (6) the implementation of a platform to make the research results searchable, traversable and generally accessible. The scholarly digital edition of the postils is carried out in accordance with the principles formulated by Patrick Sahle, who stresses that “a digitised edition is not a digital edition” (Sahle 2016: 33).

Literature

- Chiarcos, Christian, Maxim Ionov. 2018. *Linking the TEI: Approaches, Limitations, Use Cases*. Paper presented at DH2019, Utrecht, Niederlande.
- Frymire, John M. 2010. *The Primacy of the Postils. Catholics, Protestants, and the Dissemination of Ideas in Early Modern Germany*. Leiden, Boston: Brill (Studies in Medieval and Reformation Traditions 147).
- Gelumbeckaitė, Jolanta. 2017. Predigtkultur in Litauen: Corpus der altlitauischen Postillen. In: Assel, Heinrich, Johann Anselm Steiger, Axel E. Walter (Hg.). *Reformatio Baltica: Kulturwirkungen der Reformation in den Metropolen des Ostseeraums*. Berlin, Boston: De Gruyter, 573–585.
- Kaplan, Frédéric 2013. “How to build an information time machine”. Presented at TEDxCaFoscariU, Venice, Italy.
- Sahle, Patrick. 2016. “What is a Scholarly Digital Edition?” In: Driscoll, Matthew James, Elena Pierazzo (eds). *Digital Scholarly Editing. Theories and Practices*. Open Book Publishers.
- Tittel, Sabine, Helena Bermúdez-Sabel, Christian Chiarcos. 2018. “Using RDFa to link text and dictionary data for Medieval French.” In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 7–12. Miyazaki, Japan.

Databases and Corpora

- CorDon, <https://titus.fkdg1.uni-frankfurt.de/cordon/menu/eng/start.html>
- Senieji raštai, <http://seniejirastai.lki.lt/home.php>
- SLIEKKAS, <https://titus.uni-frankfurt.de/sliekkas/index.html>
- TITUS, <https://titus.uni-frankfurt.de/indexd.htm>

**‘Chained to the rhythm’:
Using agent-based simulation to model the evolution of
stress pattern diversity in English**

English is characterized by a relative heterogeneity of stress patterns (e.g. *léntil* vs. *hotél*, *ábject* vs. *intáct*, *íncrease* [N] vs. *incréase* [V]). We lay out a usage-based explanation for the historical evolution of stress pattern diversity in languages such as English and present it in the form of an agent-based model. We find that the predictions derived from such a model are in line with diachronic corpus data.

In stress-based languages such as English, physiological and cognitive constraints (Lehiste 1970; Pitt & Samuel 1990; Peelle & Davis 2012) favor an alternating rhythm made up of sequences of stressed and unstressed syllables (Hayes 1984; Selkirk 1984; Schlüter 2005). These preferences can affect lexical stress diachronically by biasing words or entire word classes towards those patterns which most successfully produce alternating rhythmic sequences in combination with other words in language use (e.g. *the íncrease wórries us* vs. *híkes incréase the cóst of living*; cf. Kelly & Bock 1988, Kelly 1989). In evolutionary terms (Croft 2000), the rhythmic preferences operating at the level of phrasal phonology exert a selective pressure on lexical stress, constantly testing the viability of a pattern within its usage context.

We choose agent-based simulation (Wilensky & Rand 2015) as a method for probing this line of argumentation. The agent population in our model is made up of constituent types (i.e. a proxy for lexical items) defined by linguistic attributes, notably stress pattern, syllable weight and morpho-syntactic class. In each round of the simulation, a predetermined number of agents are probabilistically selected to occur and interact with one another within one of a range of possible syntagmatic contexts to form a rhythmic phrase. The phrase is evaluated with respect to prosodic criteria (rhythmic alternation and weight-to-stress) and the agents are rewarded or penalized accordingly. These payoffs continuously update the agents' fitness attribute, which in turn determines the agents' chances of successfully reproducing into the next generation.

The simulation suggests that stress pattern diversity will stably establish itself if the occurrence contexts of polysyllables also include monosyllabic material at a sufficiently high rate. In such a setting, diverse rather than uniform lexical stress patterns will reduce the probability of rhythmically suboptimal clashes and lapses. This prediction matches diachronic data from the Penn-Helsinki Parsed corpora of English (Kroch & Taylor 2000; Kroch, Santorini & Delfs 2004; Kroch, Santorini & Diertani 2016).

References

- Croft, William. 2000. *Explaining language change: an evolutionary approach* (Longman Linguistics Library). Harlow, England ; New York: Longman.
- Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15(1), 33–74.
- Kelly, Michael H. 1989. Rhythm and language change in English. *Journal of Memory and Language* 28(6), 690–710. doi:10.1016/0749-596x(89)90004-1.
- Kelly, Michael H. & Bock, J. Kathryn. 1988. Stress in time. *Journal of Experimental Psychology* 14(3), 389–403. doi:10.1037/0096-1523.14.3.389.
- Kroch, Anthony, Santorini, Beatrice, & Delfs, Lauren. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony, Santorini, Beatrice, & Dierani, Ariel. 2016. The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony & Taylor, Ann. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania.
<http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4>.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, Mass: MIT Press.
- Peelle, Jonathan E. & Davis, Matthew. 2012. Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3. doi:10.3389/fpsyg.2012.00320.
- Pitt, Mark A. & Samuel, Arthur G. 1990. The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance* 16(3), 564-573.
- Schlüter, Julia. 2005. *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. Berlin; New York: Mouton de Gruyter.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.
- Wilensky, Uri & Rand, William. 2015. *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, Mass.: MIT Press.

Fall of the jers: A multi-factorial analysis of the sound change progression in the Old Novgorodian birchbark texts

The graphics of the birch bark texts produced in Novgorod from the 11th to the 13th centuries show effects of the fall of the jers in the weak position in that the jers in word-initial and word-medial syllables of word forms are rendered inconsistently, or not at all (Janin / Zaliznjak 1993; Zaliznjak 2004). According to Eom et al. (2004), the progression of the underlying sound change process, as reflected in these texts, is associated with the process of optimization of word structure. However, it is still an open question whether and how the position of weak jers in the phonological word influenced the progression of their fall (Zaliznjak 2004: 63). This paper investigates the progression of this process in birch bark texts as a function of structural and usage-based factors.

The data were coded and analyzed with respect to the following variables: (estimated) time of text creation; jer position in the phonological and morphological word; length of the phonological and morphological word (in syllables); status of the morpheme containing the weak jer; alternation of the target form with inflectional forms having strong jers in the word paradigm; nature and frequency of the resulting consonant clusters (if present in the lexicon). A multifactorial analysis of the data was performed using logistic regression with the random variables “text or author” and “time” (cf., Baayen, 2008; Szmrecsanyi, 2013).

The analysis demonstrates that the fall first affects the jers in the initial syllables of morphological words, then the jers in the word-medial syllables of morphological words, and finally the jers in the proclitics. In addition to the random variable “time”, the writing or non-writing of the weak jers in the word-initial and the word-medial syllables is determined by (a) the status of the morpheme containing the jer (suffixes are affected earlier than prefixes and prepositions), (b) the resulting consonant cluster’s frequency and (c) the cluster’s phonological makeup. The progression of the fall of jers can thus be explained by an interaction of structural – both morphological and phonological – and usage-based factors.

References

- Baayen, Harald R. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.
- Eom, Jeehyeon / Werner Lehfeldt / Peter Meyer. 2004. Der Jerausfall und seine Folgen für die Wortstruktur der slavischen Sprachen. *Zeitschrift für Slavische Philologie* 63, 249-309.
- Janin, Valentin L. / Zaliznjak, Andrej A. 1993. *Novgorodskie gramoty na bereste (iz raskopok 1984-1989 gg.)*. Moskva.
- Szmrecsanyi, Benedikt. 2013. The great regression: Genitive variability in Late Modern English news texts. In: Borjars, Kersti, David Denison, and Alan Scott (eds.) *Morphosyntactic categories and the expression of possession*, 59–88. Amsterdam: Benjamins.
- Zaliznjak, Andrej A. 2004. *Drevnenovgorodskij dialekt*. Moskva: Jazyki slavjanskoj kultury.

Exploring the sources of animacy distinctions

Keywords: Diachronic morphology, Animacy Hierarchy, source- and result-oriented explanations, noun inflection

Animacy is a semantic category that influences the grammatical structure of languages in different ways. Its manifold effects have been linked to the functioning of a typological hierarchy, called either Nominal or Referential or Animacy Hierarchy, which has been claimed to be a true universal (Kiparsky 2008), not just a widely held typological generalization. Nevertheless, recent research on the sources of several animacy markers (Cristofaro 2013, 2019) has shown that at least some of them do not arise as morphemes specifically signaling animacy distinctions or that the function that they serve synchronically is preconditioned by the circumstances of their origin (in the Indic language Maithili, for instance, the pluralizer *lokain* is used only with human nouns presumably as a consequence of its original lexical meaning being ‘people’).

This source-oriented approach reveals that the association of certain morphemes with animacy can be regarded as secondary or contingent, a mere accident of their historical evolution. If this were the case of every morphological marker of animacy, the significance of the Animacy Hierarchy as a typological generalization (or even universal, be it absolute or statistical) would somehow be compromised, at least at a certain level of analysis (see Cristofaro & Zúñiga 2018). Some of the examples that we will examine in this paper are certainly suggestive of grammaticalization (or regrammaticalization) processes having led to the present animacy-based values (as occurred with the emergence of specialized inflectional markers such as the nominal dative singular case in *-ovi* in Slovak and other Slavic languages, formerly a common ending of a specific declension class).

However, the documented (or else presumed) existence of shared paths of change across languages as well as the convergent development of animacy-based innovations (materialized in similar asymmetries in the expression of plurality, differential object marking, animate first ordering in syntax, etc.) appear to demand a more comprehensive explanation, beyond the specific one referring to the history of each individual marker. Actually, this convergence in the outcomes, captured by such generalizations as the Animacy Hierarchy itself, suggests interpreting animacy as a fundamental cognitive (and linguistic) property, which is also evidenced by cognitive and psychological experiments (Jones et al. 1991, New et al. 2007, Trompenaars et al. 2021).

The diachronic coherence in the rise of animacy distinctions is a tendency identifiable both in morphological markers of the category (animacy as a feature) and in its behavior as a factor conditioning the selection of values of other features (animacy as a condition; cf. Santazilia 2020). Moreover, the grammatical consequences of the manifestation of animacy can ultimately be explained in terms of identical selective pressures and analogous evolutionary responses (Haspelmath 2019: 15), in much the same way as some biological traits (the wings of birds, bats, insects and old pterosaurs, for instance) are considered to have arisen for similar functional reasons, in spite of the fact that they have rather different evolutionary histories. Thus some general, overarching principles may turn out to be responsible for the attested structural similarities among languages (which would explain why universally preferred constructions may emerge even without any preconditions in the grammaticalization source, see Seržant & Rafiyenko 2021).

References

- Cristofaro, Sonia. 2013. The referential hierarchy: reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries. Studies in memory of Anna Siewierska*, 69–93. Berlin & New York: De Gruyter Mouton.
- Cristofaro, Sonia. 2019. Taking diachronic evidence seriously: result-oriented vs. source-oriented explanations of typological universals. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne M. Michaelis & Ilja A. Seržant (eds.), *Explanation in typology. Diachronic sources, functional motivations and the nature of the evidence*, 25–46. Berlin: Language Science Press.
- Cristofaro, Sonia & Fernando Zúñiga. 2018. Synchronic vs. diachronic approaches to typological hierarchies. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony*, 3–27. Amsterdam & Philadelphia: John Benjamins.
- Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne M. Michaelis & Ilja A. Seržant (eds.), *Explanation in typology. Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press.
- Jones, Susan S., Linda B. Smith & Barbara Landau. 1991. Object properties and knowledge in early lexical learning. *Child development* 62 (3). 499–516.
- Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalization. In Jeff Good (ed.), *Linguistic universals and language change*, 23–53. Oxford: Oxford University Press.
- New, Joshua, Leda Cosmides & John Tooby. 2007. Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 104 (42). 16598–16603.
- Santazilia, Ekaitz. 2020. The two faces of animacy. *Studies in Language* 44 (4). 812–830.
- Seržant, Ilja A. & Dariya Rafiyenko. 2021. Diachronic evidence against source-oriented explanation in typology: Evolution of prepositional phrases in Ancient Greek. *Language Dynamics and Change* 11. 167–210.
- Trompenaars, Thijs, Teresa A. Kaluge, Rezvan Arabi & Peter de Swart. 2021. Cognitive animacy and its relation to linguistic animacy: evidence from Japanese and Persian. *Language Sciences* 86 (101399): 1–17.

***Tone split and tone replacement:
toward the three-tone system of the ‘Western’ SBB Languages
(Central Sudanic, Central Africa)***

Pascal Boyeldieu (UMR 8135, CNRS, INaLCO, EPHE)

Abstract

Sara-Bongo-Bagirmi (SBB) languages represent a group of some 40 African languages that are scattered between Lake Chad in the North-West, and Lake Albert in the South-East, thus covering parts of Chad, Sudan, South-Sudan, the Central African Republic, and the Democratic Republic of Congo. The *SBB languages* form the largest sub-branch of Greenberg’s (1963) *Central Sudanic (Nilo-Saharan phylum)* and are geographically inserted among various languages or linguistic groups such as Chadic, Adamawa, Ubangian, other Central Sudanic, Eastern Sudanic, and Arabic.

While languages in the East display two-tone systems directly reflecting the historical *SBB configuration, a large subgroup of ‘western’ languages (*OCC for *occidental*) later innovated in developing a new tone in the high frequencies. More recently the western languages Yulu and Gula Koto independently underwent a splitting of their low tone, thus creating a now phonologised extra-low level. The main correspondences between the different types of systems are summarised in Table 1. In order to avoid possible ambiguities (e.g. H tone, but in which system?), tone levels will from now on be symbolised by numbers (0/1/2/3) as indicated below:

| *SBB & eastern languages 2 tones | > | *OCC & western languages 3 tones | > | Yulu & Gula Koto 4 tones |
|-------------------------------------|---|-------------------------------------|---|-----------------------------|
| High (2) | | High (3) | | High (3) |
| Low (1) | | Mid (2) | | Mid (2) |
| | | Low (1) | | Low (1) |
| | | | | X-Low (0) |

Table 1. Tone systems and tone level numbering

If it has been necessary for the following to mention the fourth tone of Yulu and Gula Koto, this paper is nevertheless specifically concerned with the emergence of a third tone level in the ‘western’ sub-branch, a change that affected a fair number of languages and played an important role in the history of this linguistic group. Moreover this transformation has followed different ways according to the grammatical category – verbs or nouns – it affected and therefore offers some particular interest for the general understanding of tone change.

Verbs

Most reconstructible *SBB verbs display an original VCV form (with current reflexes of a VCV, sometimes CV or CVCV shape) that is associated with a tone pattern defined by two tonal components (e.g. 12 for *àpá* or *àpā*, 31 for *ápà*, etc.). In all languages the most frequent verb form – usually labelled as *definite aspect*, *aorist* or *perfective* – undergoes tone alternations conditioned by the nature of the subject and determining several tone classes.

Let us illustrate this specific behaviour with the case of Modò, an eastern language that is clearly the best representative of the common *SBB situation under the respect of the verb morphology. Modò has four verb classes resulting from the different combinations of two tone highs in two-syllable patterns : A (12), B (22), C (11), and D (21). These tone patterns remain unchanged in the presence of a lexical subject. However a subject index of 3^d person lowers the first component of the tone pattern (1, 2 > 1) while a 1st-2^d person subject index raises this initial component (1, 2 > 2). Consequently the four-way contrast of the isolated verbs is reduced to two when they are preceded by a personal index:¹

¹ The same patterns appear with plural indexes, although with a different distribution.

| Class | A | B | C | D |
|---|----|----|----|----|
| Lexical subject | 12 | 22 | 11 | 21 |
| 3 ^d subj. index | | 12 | | 11 |
| 1 st -2 ^d subj. index | | 22 | | 21 |

Table 2. Tone classes of Modo

Also each class is identified by three alternating tone patterns (e.g. : class A = 12//12/22, class B = 22//12/22).

Now, *in all other languages*, the lexical subject for some reason was aligned with the 3^d subject index pattern so that, in the eastern languages at least, the verbs were, in all contexts, reduced to two classes, namely AB (12/22) and CD (11/21).

In the ‘western’ languages, the same happened for the reflexes of *SBB*A and *SBB*B that merged in *OCC*AB (*12/*22). Yet the contrast in the reflexes of *SBB*C and *SBB*D2 was not only preserved but even reinforced by enlisting a new *OCC***31** tone pattern and generating, by analogy with the *11/*21 alternation of *OCC*C, a specific *21/***31** alternation for *OCC*D verbs as shown in the correspondences displayed in Table 3:²

| | | | | | |
|------|-------------------|-------------------|-------------------|-----------------------|--------------------|
| *SBB | *A *12//12/*22 | *B *22//12/*22 | *C *11//11/*21 | *D1 *? | *D2 *21//11/*21 |
| *OCC | *AB *12/*22 | | *C *11/*21 | *D *21/* 31 | |

Table 3. Morphological function of tone level 3 at the *OCC level

Appendix A. displays some lexical illustrations of *SBB*C and *SBB*D2 reflexes.

Nouns

In nominal comparative series, ‘western’ languages may show, *in a quite irregular way*, one or the other of four patterns involving a level 3 tone and reconstructible, at an *OCC level, as ***31**, ***32**, ***13** or ***23**. Reflexes of the latter may occur while other ‘western’ languages preserve the tonal reflex of an original *SBB and/or *OCC formula.

Some lexical illustrations of such irregularities are given in Appendix B.

The best explanation one can offer for such a situation is the following: the ‘western’ languages having undergone an important lexical replacement, probably due to contact with (an) up to now unidentified three-tone language(s), numerous nouns defined in terms of three contrastive tones integrated as such the *OCC language system. With the possible contribution of the tone changes simultaneously affecting verbs, these new patterns contaminated the older nominal lexicon in form of free variants that were progressively eliminated, each ‘western’ language – or subgroup of ‘western’ languages – finally preserving one or the other of the two competitive tone patterns, inherited from either *SBB or *OCC.

Outcomes

In verbs, the *OCC class *D formula ***31** represents a genuine *reflex* of *SBB *21. Moreover, it has to be emphasised that the shift *SBB *21 > *OCC ***31** was constrained by the morphological alternation role played by the tones in the frame of class *D verbs. With class *C verbs the same original tone pattern remained unchanged: *SBB *21 > *OCC *21.

In nouns, the *OCC formulas ***31**, ***32**, ***13**, and ***23** are new patterns that do not result from any prior unit. In case they affect nouns dating to the original *SBB vocabulary, they strictly represent *replacements* of the previous tone patterns still attested in the eastern languages. In this latter case the ‘western’ form of a noun combines older segments with new tones, thus revealing a partial loss and discontinuity in its history since its *SBB origin.

Nevertheless, the two processes, probably reinforcing each other, jointly contributed to the emergence of a new *OCC system that led to the current ‘western’ languages.

² The *SBB*D1 formula calls for comments that I shall not mention in this abstract.

The presentation is directly inspired by a historical comparative study of the SBB tone systems (Boyeldieu 2000). It will provide comparative data (also available in Boyeldieu, Nougayrol & Palayer 2006: <https://llacan.cnrs.fr/SBB/>) and highlight the linguistic and geographical conditions of the tonal change.

Boyeldieu, Pascal. 1996. Mutation et réfection dans l'expansion des registres tonals : l'exemple des langues sara-bongo-baguirmiennes. *Revue de Phonétique Appliquée* (Université de Mons-Hainaut), 120, p.163-178.

———. 2000. *Identité tonale et filiation des langues sara-bongo-baguirmiennes (Afrique centrale)*. Köln: Rüdiger Köppe Verlag (SUGIA, Beiheft 10).

Boyeldieu, Pascal, Pierre Nougayrol & Pierre Palayer. 2006. *Lexique comparatif historique des langues sara-bongo-baguirmiennes*. Paris : CNRS-LLACAN. Electronic resource: <https://llacan.cnrs.fr/SBB/>

Greenberg, Joseph H. 1963. Languages of Africa. *International Journal of American Linguistics*, 29, 1, part II. [2^d ed.: 1966. *The Languages of Africa*. Indiana University, Bloomington, La Haye: Mouton]

| | Eastern languages | | | *SBB*C (*11//*11/*21) | | | | |
|-------|------------------------|---------------------------------------|---------------------|--------------------------|-------------------------------|-------------------|--------------------|---------------|
| | C Modo 11//11/21 | CD Bongo (1)1/(2)2 ¹ | CD Baka 11/21 | C Yulu 11=00/21=20 | ABC Gula Koto 1(1)/2(2) | CD Na 11/31 | CD Sar 11/31 | |
| V/069 | òlè//òlè/ólè | yà/yá ¹ | òyò/óyò | ùlè/ùlè | òy/òy | ò/ò | òy/òy | ‘die’ |
| V/071 | òpò//òpò/ópò | òp/óp ¹ | ànò/ánò | òpò/òpò | òp/òp | òpò/òpò | òp/òp | ‘eat’ |
| V/152 | àtò//àtò/átò | òtù/òtù ¹ ‘rot’ | | ààcè/ààcè | àt/àt | átù/átù | òtè/òtè | ‘smell (bad)’ |

| | Eastern languages | | | *SBB*D2 (*21//*11/*21) | | | | |
|-------|------------------------|---------------------------------------|---------------------|---------------------------|-------------------------|-------------------|--------------------|----------|
| | D Modo 21//11/21 | CD Bongo (1)1/(2)2 ¹ | CD Baka 11/21 | D Yulu 21/31 | D Gula Koto 10/31 | CD Na 11/31 | CD Sar 11/31 | |
| V/112 | úpò//úpò/úpò | | òfò/ófò | òfò/ófò | | | | ‘kill’ |
| V/115 | ádí//ádí/ádí | èdì/édì ¹ | (edj) | āādā/āādā | èd/èd ¹ | àri/àri | èdà/édà | ‘rain’ |
| V/111 | átì//átì/átì | èti/étì ¹ | èci/éci | | èt/èt ¹ | àti/àti | | ‘sneeze’ |

Appendix A. Lexical illustrations of *SBB*C and *SBB*D2 reflexes

| | Eastern languages | | | *SBB*11 | *OCC*11 (or *31) | | | |
|--------|----------------------------|-------------|------------|---------------|--------------------|--------------------|---------------------|----------------------|
| | Modo 11 | Bongo 11 | Baka 11 | Yulu 11=00 | Gula Koto 11 | Na 11 | Sar 11 | |
| N/0017 | kòmò | kòmò | kòmò | (kām(ə) !?) | kām | kām | kùm | ‘eye, face’ |
| N/0462 | kòtò | kòtò | (mokoto) | kòt(ə) | | | | ‘sterility, sterile’ |
| N/0536 | | kùdi | | kēj(ə) | kèɖ | kùjè | kàdè | ‘worm, grub, maggot’ |
| N/0029 | dòkòròsò (dò-kò ròsò ?) | | (kési) | | ³¹ kéc’ | kòkè | kòsə | ‘iron, hoe’ |
| N/0019 | (gáli !?) | gèl | | gääl(ə) | gèl | ³¹ gáli | gèl | ‘left (side)’ |
| N/0353 | tibò | | | vööv(ə) | | ³¹ júbù | ³¹ ndúbè | ‘bellows’ |

| | Eastern languages | | | *SBB*12 | *OCC*12 (or *OCC*31/*32/*13) | | | |
|--------|-------------------|-------------|------------|---------------|------------------------------|--------------------------|--------------------|------------------------------|
| | Modo 12 | Bongo 12 | Baka 12 | Yulu 12=02 | Gula Koto 11 | Na 12 | Sar 12 | |
| N/0083 | máá | máhá | (mbasa) | máàs(ə) | màs | màsā | màsə | ‘ <i>Tamarindus indica</i> ’ |
| N/0094 | bàďú | bòďú | (uəɖu) | bääď(ə) | váď | bārū | bòr | ‘warthog’ |
| N/0107 | ǫú | hiǫú | sùǫú | | kùǫ | ǫū | yibə | ‘oil, (grease)’ |
| N/0112 | kánó | | | | ³² kún | ³² kúnū | ³¹ kón | ‘nose’ |
| N/0058 | (tòkpè !?) | tíkí | | | tít | tihī, ¹³ tihí | tī | ‘bowels’ |
| N/0097 | (kòwé ?) | kùlǫhí | | | cic | kòkī | ¹³ kòsə | ‘cucumber’ |

| | *SBB*22 | | | *OCC*22 (or *OCC*31/*32/*23) | | | | |
|--------|-------------------|------------|-------------|------------------------------|--------------------|--------------------|--------------------|----------------------------------|
| | Eastern languages | Modo 22 | Bongo 22 | Baka 22 | Yulu 22 | Gula Koto 22 | Na 22 | |
| N/0196 | kénzè | kínjí | kénzè | kēēnj(ə) | kānz | kānjē | kānjə | ‘fish’ |
| N/0198 | népé | níhí | éfé | nēēp(ə) | lēhē | nōhē | nāā | ‘moon’ |
| N/0203 | kúpó | kúhú | kófó | kōōf(ə) | kōhō | kōhō | kōō | ‘seed’ |
| N/0227 | | hídó | (sida) | sūūj(ə) | ³² sód̄ | | yāda | ‘ <i>Anogeissus leiocarpus</i> ’ |
| N/0433 | yóri | | | ³¹ sóór(ə) | | ²³ kīró | ²³ yāró | ‘dirt’ |
| N/0223 | kólógbé | higé | (sige) | ³² síg(ə) | ³² kéd̄ | ³² kóbē | ³² yégā | ‘rat, mouse’ |

Appendix B. Irregular occurrences of level 3 tones in ‘western’ languages

Talk: Reconstructing the Kugama tone system

Author: Lora Litvinova

The paper presents the first results of the internal reconstruction of the tone system of Kugama, an Adamawa language of Nigeria. The Kugama tone system distinguishes three tone levels: High (H), Mid (M), and Low (L). I argue that the three-level tone system in Kugama originates from a two-level tone system. The current three levels have emerged primarily through the phonologization of an earlier phonetic raising of H and L in certain environments. Thus, the H level of the current three-level tone system goes back to the phonetic raising of H before a tautosyllabic L. For instance, this is suggested by the following synchronic tonal rule: M becomes H before a floating ^L, as in (1). The H tone of the two-level system that did not precede a tautosyllabic L was reinterpreted as M in the present-day Kugama tone system.

- (1)
- | | | | |
|---------|------------------|--|---------------------|
| /vélí | | | ʔᵛōᵛkī/ |
| vélí | ^L | | ᵛōᵛkī ¹ |
| husband | GEN ² | | daughter |
| ‘groom’ | | | |

The L tone of the two-level system was equally phonetically raised before a raised H and subsequently this phonetically raised L was also reinterpreted as M. For example, nouns with an MH^L tone pattern in their Free Form (main allomorph) have an LM tone pattern in their Construct Form 1 (a secondary allomorph): *kǎm* |kǎm^L| ‘body hair’ and *kǎm* ‘body hair.CF1’, as in (2).

- (2)
- | | | |
|-------------------|--|-----------|
| /kǎm | | púūpī/ |
| kǎm | | púūpī |
| body_hair.CF1 | | whiteness |
| ‘white body hair’ | | |

I will discuss other relevant phonetic and phonological properties of the current Kugama tone system as well as provide further evidence for the proposed internal reconstruction.

¹ Vertical bars |a| represent the morphophonological level.

² CF1 - Construct Form 1, GEN - Genitive

VSO orders in the *Egeriae* and *Antonini Placentini itineraria*; new evidence for the evolution towards Old Romance inversion systems

It is well documented that the Old Romance languages featured more subject-verb inversion than their modern descendants. Although many modern varieties of Italo-Romance and Ibero-Romance allow inversion with unaccusative verbs, inversion with transitive verbs is heavily constrained and not freely available outside particular constructions, such as for examples wh-questions (Rizzi 1990). In most documented varieties of Old Romance on the other hand, inversion is found with all types of verbs without distinction, a fact that has prompted some historical Romanists to consider these varieties 'V2 systems' because of the similarity with Modern Germanic V2 languages (Vance 1997; Poletto 2014; Wolfe 2018).

The origin of the Old Romance inversion systems is poorly understood, but their wide diatopic distribution suggests internal development within the Latin/Romance-family. Evidence has been sought in Late Latin, and one text in particular, the *Itinerarium Egeriae*, has attracted much attention. In this late 4th century text, VSO order is quite widespread, leading Ledgeway to the conclusion that the position of the verb is already the same as in Old Romance (Ledgeway 2017)

I will present data showing that this claim is too strong. A full quantitative and qualitative analysis of *Itinerarium Egeriae* reveals that the VSO orders are not the result of very high verb movement, but rather arise through a very low position of the subject. Furthermore, I will add data from a complete analysis of a text that has not featured prominently in the debate on the evolution of word order, namely the late 6th century *Itinerarium Antonini Placentini*. While VSO-order is quantitatively even more robust in this text than in *Egeria*, closer qualitative analysis shows that inversion is almost exclusively found with intransitive verbs and passives:

- (1) *Illic currit fluvius Asclepius*
there runs river.NOM Asclepius.NOM
'There the river Asclepius runs'
- (2) *Super his locis descendit ros sicut pluvia*
Over this places.ABL descends dew.NOM as rain.NOM
'Over these lands a dew falls like a rain'
- (3) *In qua etiam synagoga posita est trabis*
In which indeed synagogue placed is beam
'In this synagogue a beam is placed'

The combined evidence of these textual witnesses suggests that, while we need not abandon the hypothesis of a Romance-internal evolution of inversion, we should reconsider its diachrony. The new data from the youngest text reveals a new alignment system for the arguments of the verb which is sensitive to theta-roles, with thematic arguments liberally appearing in postverbal position, while agent arguments (in transitives) favour the preverbal position. If correct, this analysis entails that generalised inversion in Old Romance cannot have developed as early as the late 4th century, but must rather be postponed until at least the 7th century, thereby adding a new important piece to our understanding of the diachrony of word order in Romance.

References

- Ledgeway, Adam (2017). "Late Latin Verb Second: The Sentential Word Order of the *Itinerarium Egeriae*". In: *Catalan Journal of Linguistics*. Catalan Journal of Linguistics 16. Ed. by Jaume Mateau Fantanals and Renato Oniga, pp. 163–216.
- Poletto, Cecilia (2014). *Word Order in Old Italian*. Oxford: Oxford University Press.
- Rizzi, Luigi (1990). "Speculations on verb second". In: *Grammar in Progress*. Ed. by Joan Mascaró and Marina Nespore. Walter de Gruyter, pp. 375–386.
- Vance, Barbara S (1997). *Syntactic Change in Medieval French : Verb-Second and Null Subjects*. Dordrecht: Kluwer.
- Wolfe, Sam (2018). *Verb Second in Medieval Romance*. Oxford: Oxford University Press.

The Greek suffix -θ- and the Caland System

Andrea Santamaria (University of Bologna)

Via Zamboni, 32 – 40123 Bologna (BO), Italy

andrea.santamaria3@unibo.it

The present work addresses the distribution of the Greek verbal suffix -θ-, its connection to the Caland System and its development from the Homeric Greek up to the Byzantine period.

The presence of forms showing a Caland behavior in Greek was recognized as early as Wackernagel (1897), only few years after the related observations of Caland (1892; 1893) himself on Indo-Iranian languages. After more than a century of studies, Greek proved to match a high number of Caland suffixes attested elsewhere (cf. Ritsch 1974²: 66; Meißner 1998; Bichlmeier 2014). These suffixes commonly contribute to deriving adjectives (e.g. ἐρυθρός ‘red’ < **h₁rud^h-r^o-*), nouns (e.g. ἔρευθος ‘redness’ < **h₁reud^h-e/os-*) and verbs (e.g. ἐρέυθω ‘I make/become red’ < **h₁reud^h-eh₁-*). Among them, the suffix -θ- was tentatively associated with some Caland formations (e.g. Nussbaum 1976: 90 and Rau 2009: 152-153 fn. 80). Still, no definitive evidence was put forward in favor its attribution to the Caland System nor its behavior *vis-à-vis* the other Caland suffixes was clearly detected.

Since Benveniste (1935: 188-210), a suffix -θ- was recognized in a wide range of forms, such as adjective abstracts (e.g. πλῆθος ‘multitude, quantity’), *nomina agentis* (e.g. τένθης ‘gourmand’), verbs (mostly θω-presents such as βρίθω ‘I am heavy’) and adverbs (e.g. μίνυνθα ‘a short time’). The adjective abstract μέγεθος ‘greatness’ was analyzed as deriving a property concept adjective and would therefore speak in favor of a ‘marginal’ Caland *-*d^ho-* (cf. Nussbaum 1976: 90). On the other hand, few verbal forms such as λήθω ‘I hide’ were traced back to constructions employing the function verb **d^heh₁-* (cf. Hackstein 2002 and Schutzeichel 2014). However, as a complete investigation of forms in -θ- and their Indo-European relatives is still pending, both the typology of derived stems and their position within the Caland System remain opaque.

Thus, the present work focuses on (a) the typology, from both a morphological and a semantic perspective, of stems showing a derivative in -θ- and their diachronic developments; (b) the Indo-European cognates of -θ- and the role they played within the Caland System. Accordingly, I argue for the following classes:

a) Verbs deriving property concept roots showing both other Caland formations and, sometimes, root aorists. This group is the oldest one. Moreover, a PIE suffix *-*d^h-* (perhaps ultimately traceable back to a grammaticalized **d^heh₁-*) with such roots would find parallels outside Greek (e.g. Gr. πλήθω ‘I am full’, Av. *frāda-* ‘thrive’ and Lat. *plēbēs* ‘common people’). Semantically, these verbs are commonly the inchoative member of a causative/inchoative alternation based on the opposition of derived stems (e.g. βρίθω ‘I am laden with’ : βριάω ‘I make strong’). Although such a typology of alternation was largely neutralized in Homeric Greek (cf. πύθω ‘I cause to rot’), it is revealingly traceable back to the Caland System.

b) Deverbal and denominal verbs in -θ- without Caland cognates and analogical to Caland forms in -θ-. In fact, up to the Byzantine period, new (inchoative) presents in -θω were mostly built on thematic stems either from *s*-stem nouns (e.g. τελέθω ‘I come to end’ : τέλος ‘coming to pass’) or *e*-graded presents (e.g. φλεγέθω ‘I burn up’ : φλέγω ‘I burn sth.’). In Homeric Greek, few θ-aorists are opposed to thematic aorists (e.g. ἔσχεθον : ἔσχον ‘I had’) and could therefore be directly related to the emergence of θ-passives.

c) Re-derived nouns. The vast majority are abstracts in -θεσ- analogical to μέγεθος. They are variously shaped from non-Caland property concept roots (e.g. ὄχθος ‘eminence’), verbs (e.g. κέλευθος ‘road, path’) or θω-presents (e.g. λήθος ‘forgetfulness’). Nevertheless, these nouns often belong to new ‘paradigms’ derived *einzel sprachlich* through suffixes previously employed within the Caland System (e.g. ἔσθος ‘garment’, ἑάνός ‘fine robe’ and ἔννυμι ‘I put clothes on another’). The same behavior is to be found for *nomina agentis* in -θης and *nomina actionis* in -θη, which can partner with abstracts in -θεσ- too (e.g. ἐσθής ‘raiment’ and λήθη ‘forgetting’). Unsurprisingly, such suffixes were also employed to derive new adjectives from nouns in -θ- (e.g. λαθραῖος ‘secret’).

Literature

Benveniste, E. 1935. *Origines de la formation des noms en indo-européen*. Paris.

Bichlmeier, H. 2014. Caland System and Greek. In: G. Giannakis, ed., *Encyclopedia of Ancient Greek Language and Linguistics*. Pp. 256-260. Leiden-Boston.

- Bozzone, C. 2016. The Origin of the Caland System and the Typology of Adjectives. *IEL* 4: 15-52.
- Caland, W. 1892. Beiträge zur Kenntniss der Avesta, no. 19. *KZ* 31: 266-268
- Caland, W. 1893. Beiträge zur Kenntniss der Avesta, no. 26. *KZ* 32: 592
- Hackstein, O. 2002. Uridg. *CH.CC > *C.CC. *HS* 115/1: 1-22.
- Hackstein, O. 2012. When Words Coalesce: Chunking and Morphophonemic Extension. In H.C. Melchert, ed., *The Indo-European Verb*. Pp. 87-104. Wiesbaden.
- Kölligan, D. 2018. Funktionsverbgefüge und Sekundärwurzeln. In E. Rieken *et alii*, ed., *100 Jahre Entzifferung des Hethitischen. Morphosyntaktische Kategorien in Sprachgeschichte und Forschung*. Pp. 219-237. Wiesbaden.
- Lehmann, W.P. 1942. The Indo-European *dh*-determinative in Germanic. *Language* 18: 125-132.
- Magni, E. 2010. L'evoluzione semantico-funzionale dell'elemento -θ- nella morfologia verbale del greco. In I. Putzu *et alii*, ed., *La morfologia del greco tra tipologia e diacronia*. Pp. 266-28. Milano.
- Meißner, T. 1998. Das „Calandsche Gesetz“ und das Griechische - nach 100 Jahren. In W. Meid, ed., *Sprache und Kultur der Indogermanen : Akten der X. Fachtagung der Indogermanischen Gesellschaft, Innsbruck, 22.-28. September 1996*. Pp. 237-354
Innsbruck
- Meißner, T. 2006. *S-stem Nouns and Adjectives in Greek and Proto-Indo-European. A Diachronic Study in Word Formation*. Oxford.
- Nussbaum, A.J. 1976. *Caland's "Law" and the Caland System*. PhD Dissertation, Harvard University.
- Rau, J. 2009. *Indo-European Nominal Morphology. The Decades and the Caland System*. Innsbruck.
- Ritsch, E. 1974² [1937¹]. *Wortbildung der homerischen Sprache*. 2nd ed. Berlin-New York
- Schutzzeichel, M. 2014. *Indogermanische Funktionsverbgefüge*. Münster.
- Tucker, E. 1990. The Creation of Morphological Regularity: Early Greek Verbs in -έω, -άω, -όω, -ύω and -ίω. Göttingen.
- Wackernagel, J. 1897. *Vermischte Beiträge zur griechischen Sprachkunde*. Basel
- Willi, A. 2018. *Origins of the Greek Verb*. Cambridge.

Continuative relative clauses in Greek documentary papyri

Although relativization and its connected phenomena have received great scholarly attention in Ancient Greek, the most extensive examinations of Greek relative clauses concern the Archaic (Probert 2015), and the Classical period (Perna 2013, Faure 2021). Within the Post-Classical period, some consideration has been paid to New Testament Greek (Du Toit 2016, Hayes 2018): in particular, a recent contribution by Du Toit (2022) investigates the characteristics and function of continuative relative clauses in the New Testament.

However, this type of clauses has not yet been studied in the Post-Classical Greek papyri, where relative constructions are still largely unexplored (Kriki 2013; Bentein and Bağrıaçık 2018; Bentein and Cattafi forthcoming), despite the sociolinguistic potential of these documents in the investigation of morphosyntactic variation (Bentein 2019; Bentein, Cattafi and La Roi forthcoming).

The aim of this paper is therefore to analyse the behavior of continuative relative clauses in Greek documentary papyri from Egypt, by taking into account letters, petitions and contracts from the first to the eighth century AD.

First, I will discuss how the papyrological examples relate to the general concept of continuative relative clauses as pointed out in linguistic studies (cf. Lehmann 1984, *inter alia*) and in the other periods of Greek. For instance, in these texts, very different constructions such as (1) from a petition and (2) from a contract can be broadly ascribed to the “continuative” category.

- (1) Συρίων γενόμενος δεκάπρωτος [ἀπὸ τῆς αὐτῆς κώμης Θρασῶ ἀναπίσας μου τὸν ἄνδρα Καῆτ ὀνό]ματι ποιμένιν αὐτοῦ τὰ πρόβατα — ὅστις [ἀδίκως τὰς τοῦ] προκειμένου ἀνδρὸς αἴγας καὶ πρόβατα τὸν [ἀριθμὸν ἐξήκο]ντα συναπέσπασεν αὐτῷ

“Syrion having become decaprotus of the aforesaid village, persuaded my husband Ganis to pasture his flock—it was he who wrongfully removed into his own keeping my husband's goats and sheep 60 in number” (P. Sakaon 36, ll. 8-11 – III AD)

- (2) εἰς σπορὰν καὶ κατάθεσιν ὧν ἐὰν αἰρῶμαι φόρου ἀποτάκτου πυροῦ ἀρταβῶν τριάκοντα ἄνπερ φόρον μετρήσω τῷ Ἐπειφ μηνὶ τῆς (αὐτῆς) ζ᾽ ἰνδικ(τίωνος) ἄνπερθέτως

“to sow and plant with whatever I choose, at a total rent of thirty artabas of wheat, which rent I shall measure out in the month of Epeiph of the same 7th indiction without delay” (P. Charite 7, ll. 11-15 – IV AD)

Second, I will highlight some linguistic aspects of continuative relative clauses in papyri, in particular their syntactic type and their relationship with the antecedent (Bentein and Bağrıaçık 2018), and the relative markers that introduce the clauses.

Finally, I will explore their communicative and stylistic meanings, such as the role of linking different textual units (cf. Tabachovitz 1943: 11 on the possibility of using relative clauses as discourse connectives), by looking at the type of documents where continuative clauses tend to appear and at the function they perform within the text.

References

- Bentein, K. and M. Bağrıaçık. 2018. ‘On The Third Type of Headed Relative Clause in Post-Classical and Early Byzantine Greek’. *TPhS* 116 (3): 529–54.
- Bentein, K. 2019. ‘Dimensions of social meaning in post-classical Greek towards an integrated approach’. *JGL* 19 (2), 119-167.
- Bentein, K. and E. Cattafi (forthcoming). ‘Relative clauses’. In Giannakis, G., P. Filos, M. Janse, B. Joseph and I. Manolesou (eds.). *Encyclopedia of Greek language and linguistics*. Leiden: Brill.
- Bentein, K., E. Cattafi, and E. La Roi (eds.) (forthcoming). *(In)subordination in Post-Classical Greek: syntax, context and complexity*. Berlin: De Gruyter.
- Faure, R. 2021. *The Syntax and Semantics of Wh-Clauses in Classical Greek. Relatives, Interrogatives, Exclamatives*. Leiden: Brill.
- Hayes, M. E. 2018. *An Analysis of the Attributive Participle and the Relative Clause in the Greek New Testament*. New York: Peter Lang.
- Kriki, E. 2013. ‘Morfosintaktiki neoterismi sti glossa ton mi filologikon papiron: I anaforikes protasis’. PhD Dissertation, Thessaloniki: Aristotle University of Thessaloniki.
- Lehmann, C. 1984. *Der Relativsatz*. Tübingen: Gunter Narr Verlag.
- Perna, E. 2013. ‘La frase relativa in greco antico. Analisi sintattica basata sul dialetto attico di Platone’. PhD Dissertation, Padua: University of Padua.
- Probert, P. 2015. *Early Greek Relative Clauses*. Oxford: University Press.
- Tabachovitz, D. 1943. *Études sur le grec de la basse époque*. Uppsala: Skrifter.
- Toit, H. C. du. 2016. ‘Some Syntactic Features of Relative Constructions in the Greek New Testament’. *Stellenbosch Papers in Linguistics* 45: 49–75.
- Toit, H. C. du. 2022. ‘The Continuative Relative Clause in the Greek New Testament’. *Novum Testamentum* 64, 162-183.

From distal demonstrative to resultative marker (through definite article): evidence from Basque

1. Introduction

Demonstratives are at the origin of a variety of grammatical morphemes: definite articles, complementizers, conjunctions, etc. (see Diessel 1999: 39, Kuteva *et al.* 134-147). Some of these grammaticalization processes have also been described for Basque: for instance, it is beyond dispute that the definite nominal inflection emerged through the grammaticalization of demonstratives (Manterola 2015). This paper explores a grammaticalization path that to our knowledge is yet to be described in other languages, namely the one leading from demonstrative to resultative marker, through the intermediate stage of definite article.

2. Resultative constructions in Modern Basque

Resultative constructions in Basque consist of a predicative construction containing a past participle with the definite article and the auxiliary verb, as exemplified in (1). They are similar to adjectival predicates, in which the adjective appears also with the determiner (2):

- | | | | | | | | | | | | |
|-----|----|-------------------------------|-------------|----|--|----|--|-----|-----|----------------|------|
| (1) | a. | Mikel | iritsi-a | da | | b. | Mikel | eta | Ane | iritsi-ak | dira |
| | | Michael | arrived-the | is | | | Michael | and | Ann | arrived-the.PL | are |
| | | ‘Michael has already arrived’ | | | | | ‘Michael and Ann have already arrived’ | | | | |
| (2) | a. | Mikel | handi-a | da | | b. | Mikel | eta | Ane | handi-ak | dira |
| | | Michael | big-the | is | | | Michael | and | Ann | big-the.PL | are |
| | | ‘Michael is big’ | | | | | ‘Michael and Ann are big’ | | | | |

Resultative constructions contrast with perfect analytic construction in that the latter do not attach the definite article to the past participle, see examples in (3):

- | | | | | | | | | | | | |
|-----|----|-----------------------|---------|----|--|----|--------------------------------|-----|-----|---------|------|
| (3) | a. | Mikel | iritsi | da | | b. | Mikel | eta | Ane | etorri | dira |
| | | Michael | arrived | is | | | Michael | and | Ann | arrived | are |
| | | ‘Michael has arrived’ | | | | | ‘Michael and Ann have arrived’ | | | | |

That the morphological origin of *-a* (plural *-ak*), the definite article, is to be found in the distal demonstrative *ba* ‘that’ is uncontroversial (Azkue 1923-1925, Trask 1997). It is however obvious that its modern functions range well beyond a simple definiteness marker, see examples in (2).

3. Historical data

In Old Basque, adjectives and nouns involved in predicative constructions were usually not marked with a definite article (3).

- | | | | | | |
|-----|--|------|----|--------------------|--|
| (3) | Eihera | hon | da | dabileno | (Oihenart, 1657, <i>Proverbes</i> , 206) |
| | mill.the | good | is | as.long.as.it.runs | |
| | “Le moulin est bon tandis que la meule se remue” | | | | |

Likewise, the oldest Basque texts exhibit a language stage in which resultativeness may be expressed through sheer past participle forms, see (4):

- | | | | | | | |
|-----|---|------|----------|----------|-----|--|
| (4) | Habia | egin | deneko, | xori-a | hil | (Oihenart, 1657, <i>Proverbes</i> , 206) |
| | cage.the | done | AUX.TEMP | bird-the | die | |
| | “Pour lors que la cage a été faite, l’oiseau est venu à mourir” | | | | | |

The spread of the definite article to contexts where no definite interpretation is allowed is observable on written records. This spread is most readily noticeable in predicative constructions (2), and is present already in texts of the 16th century (Manterola 2015).

4. The historical evolution of resultative phrases

The definite articles *-a/-ak* spread to resultative contexts only from the 16th century onwards (Mounole 2014). This paper argues that this spread occurred on the model of adjective predicates; the need for number marking may have played a role, since number is only morphologically overt in definite phrases.

References

- Azkue, Resurrección M^a. 1923-1925. *Morfología vasca*. Euskaltzaindia: Bilbo.
- Diessel, Holger, 1999, “The morphosyntax of demonstratives in synchrony and diachrony”, *Linguistic Typology* 3, 1-49.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, Seongha Rhee, 2019, *World Lexicon of Grammaticalization*, 2nd edition. Cambridge: Cambridge University Press.
- Manterola, Julen. 2015. *Towards a history of Basque morphology: articles and demonstratives*. Doctoral dissertation, UPV/EHU.
- Mounole, Céline. 2011. *Le verbe basque ancien: étude philologique et diachronique*. Doctoral dissertation, UPV/EHU and Michel de Montaigne-Bordeaux III. [Published in *ASJU* 48, 2014 (2018)]
- Trask, Robert Lawrence. 1997. *The history of Basque*. Londres, New York: Routledge.

Areality through Migration: Investigating the Structure of Numeral Classifiers in the Eastern Himalayan Region Reveals Historic Contact Events

Numeral classifiers exhibit striking distribution in certain hotspots around the world, serving as strong areal markers due to their highly recessive nature genetically and areally (Nichols, 2003:299). By investigating one such known hotspot, the Eastern Himalayan Region (henceforth: EHR ; often called North-East India in the prior literature) – I show that the surface proliferation of a grammatical feature can actually be the result of migration and distant contact events, rather than an indication of intense, sprachbund-like convergence as was previously assumed for this feature. In doing so, I problematize the notion of an ‘areal feature,’ and the methods used to define them.

The EHR is a complex linguistic zone with four distinct language families - Trans-Himalayan, Indo-Aryan, Austroasiatic and Kra-Dai - all with long-standing similarities in history, culture and ecological environment (Konnerth et. al, 2020). Previous work identifying the EHR as a linguistic area sees classifiers as a defining areal feature (Moral, 1997), with recent work (Chelliah and Lester, 2017 ; Cathcart et. al, 2020) pointing to contact as the explanation for the prominence of classifiers in the region, or their emergence in Eastern Indo-Aryan, respectively. By surveying 22 of the languages in Assam (the most populous state of the region), I find evidence to the contrary, and present the first detailed investigation of the spread of the feature in this region that reveals clues about the linguistic prehistory of the EHR, and South Asia more broadly.

By analyzing the structure of the classifier systems in the area, I find that contact is an unlikely explanation for the high frequency of classifier languages in the area. The premise of my analysis is the fact that classifiers here are not lexical borrowings from other languages in the region, but instead have cognates in their own distant linguistic relatives. Hence, a situation of language contact must be similar to that elucidated by Matras and Sakel (2007) where structural, instead of lexical borrowing takes place, replicating a *pattern* from the donor language: a manifestation of metatypy (Ross, 1996, 2007). The key elements I identified for this analysis were Noun (N), Numeral/Quantifier (Q), and Classifier (CL) as established by previous scholars (Jones, 1970; Allan, 1977). When examining the order of these key elements, I find that the structure of the classifier phrase in *every* language of the area is predicted by its genetic affiliation, reflecting inheritance, rather than contact. This argument is further supported by the fact that the languages of the same family outside of the area (such as Standard Thai, Odia and Vietnamese) do not show a difference in the order of key elements as those in the EHR. All of this points towards contact being an insufficient explanation for the perceived areality of numeral classifiers. Some interesting oddities remain – Eastern Indo-Aryan and Khasian do share the order of key elements, and Tibeto-Burman is an outlier in all senses – the order of its classifier phrase is distinct from all languages in the area as well as its own genetic relatives outside of the area.

What then, explains the distribution of numeral classifiers in this region? I suggest that what has been considered an areal feature of the EHR can be better explained by two distant contact events and the subsequent migration of those communities into the region. This hypothesis is supported by previous work in historical linguistics – Peterson (2010, 2017) highlights classifiers as a potential contact feature between Eastern Indo-Aryan and Proto-Munda. Moreover, Dockum (2016) points towards metatypy between Southwest Tai and parts of Austroasiatic, which could have been an earlier contact event that feeds into this. I weigh the linguistic data against contributions from allied fields such as archaeogenetics, anthropology, and history, to build a timeline of when speaker groups could have migrated into the EHR, and elucidate the kind of contact that has occurred in the region. Understanding how patterns of grammatical features in the EHR can tell us about its history allows us to build a better picture of how contact and migration shape the linguistic diversity we see in other such complex linguistic zones with high genetic diversity, such as the Daly River region, Western South America, or Mainland South-East Asia.

References

- Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311. <https://doi.org/10.1353/lan.1977.0043>.
- Cathcart, Chundra A., Andreas Hölzl, Gerhard Jäger, Paul Widmer & Balthasar Bickel. 2020. Numeral classifiers and number marking in Indo-Iranian: A phylogenetic approach. *Language Dynamics and Change* 11(2). 273–325. <https://doi.org/10.1163/22105832-bja10013>.
- Chelliah, Shobhana & Nicholas Lester. 2016. Contact and convergence in the Northeast. In *The Languages and Linguistics of South Asia: A Comprehensive Guide*. Walter de Gruyter GmbH & Co KG.
- Dockum, Rikker 2016. Numeral classifiers in areal perspective: Khmer and Thai ‘syntactic borrowing’ revisited. *Austroasiatic Syntax in Areal and Diachronic Perspective. Workshop of the Austroasiatic Working Group*. September 5-7, 2016. [Slides]
- Jones, Robert B. 1970. Classifier Constructions in Southeast Asia. *Journal of the American Oriental Society*. American Oriental Society 90(1). 1–12. <https://doi.org/10.2307/598428>.
- Konnerth, Linda, Stephen Morey, Mijke Mulder, Mark W. Post & Kellen Parker van Dam. 2020. Languages and Peoples of the Eastern Himalayan Region and the North East Indian Linguistics Society: Taking stock. *Himalayan Linguistics* 19(2). <https://doi.org/10.5070/H91151002>.
- Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical Borrowing in Cross-Linguistic Perspective* (Empirical Approaches to Language Typology [EALT]). Berlin, New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110199192>.
- Ross, Malcolm. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford University Press, USA.
- Ross, Malcolm. 2007. Calquing and Metatypy. *Journal of Language Contact*. Brill 1(1). 116–143. <https://doi.org/10.1163/000000007792548341>.
- Nichols, Johanna. 2003. Diversity and Stability in Language. In *The Handbook of Historical Linguistics*, 283–310. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470756393.ch5>.
- Moral, Dipankar. 1997. North-East India as a linguistic area. *Monkmer Studies* 43-54.

Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2).

<https://doi.org/10.5070/H99223479>.

Peterson, John. 2017. Fitting the pieces together - Towards a linguistic prehistory of eastern-central South Asia (and beyond). *Journal of South Asian Languages and Linguistics*. Berlin, Germany: Walter de Gruyter GmbH 4(2). 211–257.

<https://doi.org/10.1515/jsall-2017-0008>.

Rumpled chicken come home to roost.

From [TO CARD – IMPURITY] to [TO PURIFY/HEAL (someone) – from DISEASE]. Evidence from Anatolian, Ancient Greek, and Old Indic.

The Modern Greek and Italian expression ο κόμπος φτάνει στο χτένι and *tutti i nodi vengono al pettine* (litt. *the knot(s) came to the comb*) clearly means *rumpled chicken come home to roost*. The main historical dictionaries of proverbs trace its origin back to the textile production in the antiquity, but the details are unknown (e.g. La Pucci 2007: 1023).

I advance the hypothesis that this proverb goes back to a phraseological collocation [TO COMB – IMPURITY (from wool)], that is, [TO CARD (wool)], which may reflect Proto-Indo-European (PIE) heritage, and which developed a metaphorical meaning already in antiquity. In particular, I argue for the development of a meaning [to COMB – a IMPURITY/DISEASE (from someone)], that is, [TO PURIFY/HEAL (someone or something) – from a DISEASE]. In other words, healing someone from a disease became like carding wool from impurities. This development is attested in the Anatolian and Ancient Greek branch of the Indo-European language family, although each branch has undergone a specific development starting from the same basic collocation. Anatolian shows a metaphorical narrowing into [TO COMB (down) – DISEASE – from a BODY PART] meaning [TO PURIFY/HEAL (someone) – from a DISEASE] (CTH 765.1, CTH 409.I). The very same collocation remains unattested in Ancient Greek. Instead, it shows a long simile between carding wool and the purification of the city of Athens from a mortal plague, together with a series of lemmata concerning processing wool (Aristoph. *Lys.* 574-86). In addition to that, in a controversial Old Indic passage, the collocation [TO CARD – X] is used metaphorically (1×, AVP 2.31.4, early 1st mill. BC) in the sense of ‘to prevent death/ill-minded event’. Although these differences, Anatolian as well as Indo-Iranian and Greek share the same Proto-Indo-European verbal root **kse(-n)-* ‘to card, to comb’ in the above-mentioned passages: Luw. *kiša-* ‘to comb’ (+ *katta* ‘down’), Ved. *kṣan-* ‘to card’, Gk. *ζαίνω* ‘to comb’ (+ *κατὰ* ‘down’). I will conclude that it is possible to venture the hypothesis that the basic parallel between surviving an unfavourable event and the carding process was already established at a PIE level, although it is methodologically unlikely to infer the very existence of a PIE metaphor even from the attestation of its basic elements attested in the daughter languages (as regards this methodological issue see recently Melchert 2020). Should this hypothesis be correct, it would enrich the numerous metaphors from the semantic field of textile production attested in the PIE languages – for a comparative perspective on this topic see, most recently, Olsen 2018. This study integrates the increasing interest on the identification of cognitive metaphors within the Indo-European languages, with the main scope to separate those elements that can be traced back to a PIE stage from those that are the result of later historical events. After the seminal work of Lakoff & Johnson 1980, who investigated the persistent use of metaphorical language in all areas of human experience, only some attempts have been made so far within the Indo-European languages (most recently, e.g. Kölligan 2020, 2022, van Beek 2017).

Bibliography

Beek, Lucien van (2017). Die Bildersprache des Rechts im Indogermanischen: Griechisch *ἰθεῖα δίκη* und *δίκην βλάπτειν*. In H. Bichlmeier and A. Opfermann (eds), *Das Menschenbild bei den Indogermanen. Akten der Arbeitstagung der Indogermanischen Gesellschaft in Halle* (31. März – 2. April 2011), 129–150. Hamburg: Baar.

Kölligan, Daniel (2020). Seething anger: Latin *furor*. In: L. Repanšek, H. Bichlmeier, and V. Sadovski (eds.), *vácāmsi miśrā kṛṇavāmahai*. Proceedings of the international conference of the Society for Indo-European Studies and IWoba XII, Ljubljana 4–7 June 2019, 397–412. Hamburg: Baar.

Kölligan, Daniel (2022). Conceptual metaphors and historical linguistics – some case studies. Talk at the Università Cattolica del Sacro Cuore, Milan. March 22, 2022.

Lakoff, George and Mark Johnson (1980). *Metaphors we Live by*. Chicago: University of Chicago Press.

Melchert, Craig. 2020. Hittite *handā(i)*- ‘to align, arrange, etc.’ and PIE Metaphors for ‘morally right’. In M. Serangeli and Th. Olander (eds.), *Dispersals and Diversification. Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, 166-178.

Olsen, B. A. 2018. Notes on the Indo-European terminology of sheep, wool and textile production. In R. Iversen & G. Kroonen (eds), *Digging for words. Archaeolinguistic case stories from the XV Nordic TAG Conference held at the University of Copenhagen 16–18 April 2015*, 66–77. Oxford.

Secondary lateral obstruents in South Cushitic and their significance for the linguistic history of East Africa

The precise position of South Cushitic within Cushitic (Afroasiatic) is a matter of controversy. Proposals include South Cushitic as (1) a primary branch of Cushitic (Greenberg 1963:48-49, Tosco 2020:292), (2) a coordinate branch of East Cushitic under an East–South Cushitic node (Ehret 1995:489-490, 2008:159), (3) a sub-branch of East Cushitic (Tosco 2000:109), (4) part of Lowland East Cushitic (Appleyard 2012:278), and (5) a sub-branch of (Southern) Lowland East Cushitic (Hetzron 1980:77, 101).

South Cushitic is also unique in Cushitic for having lateral obstruents, **ɟ** and **ɟʔ**. These phonemes have been reconstructed to Proto-Cushitic (Ehret 1995, 2008), from which South Cushitic is thought to have inherited them (cf. Bender (2020:138) for an alternative view). Mous (2012:347) identifies the presence of lateral obstruents in South Cushitic as problematic for establishing its position within East Cushitic: The further down in the tree it is posited, the more events of loss of lateral obstruents have to be assumed in the rest of Cushitic, and/or the more wide-ranging the waves obliterating these sounds must have been; hence the less likely such a classification is *a priori*.

In this paper I propose that the lateral obstruents in at least some South Cushitic lexical items are secondary, i.e. not inherited *as lateral obstruents* from Proto-Cushitic. In these instances, then, lateral obstruents are no obstacle for a lower-level classification of South Cushitic. I further argue that these items are historically connected to Pre-Oromo, one of the lowest branches in Cushitic. I trace the sound changes involved in their development and establish the relative chronology of some of these processes.

An example is Proto-South Cushitic ***ɟaʔ** ‘to love, like, want’ (Kießling & Mous 2003:254) which is connected to Oromo **ɟa:l-** ‘to love’ < Proto-East Cushitic ***geʔl-** ‘*id.*’ (Sasse 1979:36). The following diagram shows the development of this root from Proto-East Cushitic to Oromo and Proto-South Cushitic, based on sound changes taken from the literature (Black 1974, Sasse 1979, Kießling & Mous 2003) and the assumption of an additional, cross-linguistically supported sound change **ɟ > ɟ** that links the East and South Cushitic strands of development:

| | East Cushitic | | South Cushitic | |
|-----------------------------------|------------------|--|-------------------|-----------------------------------|
| | *geʔl- | | | |
| changes within Pre-Oromo | | | | changes within Pre-Proto-S. Cush. |
| ʔ > ɟ | geʔl- | | | |
| g > ɟ / _ {i, e} | jeʔl- | | | |
| e > a / [most contexts] | jaʔl- | | ɟaʔl- | |
| ʔ > V: / { _ C, C _ } | ja:l- | | ɟaʔl- | ɟ > ɟ |
| | | | ɟaʔ- | non-initial l > ∅ |
| Oromo | ɟa:l- | | *ɟaʔ- | Proto-South Cushitic |

I discuss whether the link between South Cushitic and Pre-Oromo thus established is one of common inheritance or language contact and conclude that lateral transfer from Pre-Oromo to South Cushitic is the most likely scenario. In the final part of the paper the implications of this late emergence of seemingly archaic phonemes are drawn for the history of Cushitic and its speakers in East Africa.

References

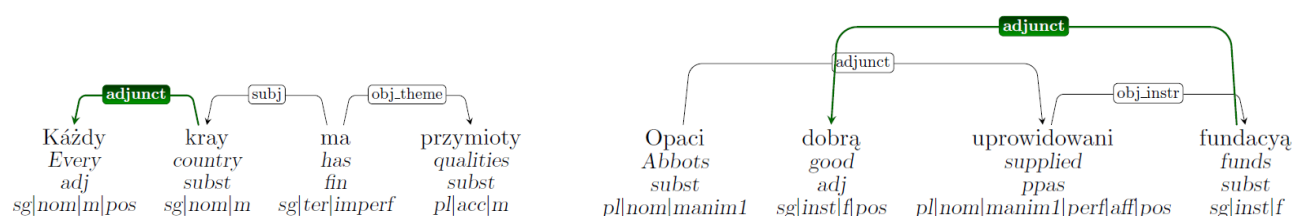
- Appleyard, David L. (2012). Cushitic. In: Lutz Edzard (ed.), *Semitic and Afroasiatic: Challenges and opportunities*. (Porta linguarum orientalium, Neue Serie, 24). pp. 199-295. Wiesbaden: Otto Harrassowitz.
- Bender, M. Lionel (Grover Hudson, ed.) (2020). *Cushitic lexicon and phonology*. (Research in African studies, 28). Berlin: Peter Lang.
- Black, Paul David (1974). Lowland East Cushitic: Subgrouping and reconstruction. PhD dissertation. New Haven: Yale University.
- Ehret, Christopher (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, tone, consonants, and vocabulary*. (University of California publications in linguistics, 126). Berkeley: University of California.
- Ehret, Christopher (2008). The primary branches of Cushitic: Seriating the diagnostic sound change rules. In: John D. Bengtson (ed.), *In hot pursuit of language in prehistory. Essays in the four fields of anthropology. In honor of Harold Crane Fleming*. pp. 149-160. Amsterdam: John Benjamins.
- Greenberg, Joseph H. (1963). *The languages of Africa*. (Publications, 25). Bloomington: Research center in anthropology, folklore, and linguistics, Indiana University.
- Hetzron, Robert (1980). The limits of Cushitic. *Sprache und Geschichte in Afrika* 2:7–126.
- Kießling, Roland & Maarten Mous (2003). *The lexical reconstruction of West-Rift Southern Cushitic*. (Cushitic language studies, 21). Cologne: Rüdiger Köppe.
- Mous, Maarten (2012). Cushitic. In: Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*. (Cambridge language surveys). pp. 342-675. Cambridge: Cambridge university.
- Sasse, Hans-Jürgen (1979). The consonant phonemes of Proto-East-Cushitic (PEC): A first approximation. *Afroasiatic linguistics* 7(1):1-67.
- Tosco, Mauro (2000). Cushitic overview. *Journal of Ethiopian studies* 33(2):87-121.
- Tosco, Mauro (2020). East Cushitic. In: Vossen, Rainer & Gerrit J. Dimmendaal (eds), *The Oxford handbook of African languages*. pp. 290-299. Oxford: Oxford University.

Discontinuous noun phrases containing adjective or adjective-like modifiers in Middle Polish texts. Preliminary research conducted on an experimental dependency treebank

In my presentation, I will show examples of unusual – compared to the modern state – order of words in noun phrases containing an adjective or adjective-like modifier in the Polish language of the 17th and 18th centuries (e.g. *dziwna trafiła się awantura* ‘there was a **strange** riot’, lit. ‘**strange** there was a **riot**’; *o pożytkach z xiąg tłumaczenia wynikających* ‘about the **benefits** resulting from translation of books’, lit. ‘about the **benefits** from books’ translation **resulting**’; *u mego widział mię stryja* ‘he saw me at **my** uncle’s’, lit. ‘at **my** he saw me **uncle**’s’). The uniqueness of the quoted and similar examples from Middle Polish texts lies in the fact that other sentence elements are relatively often placed between the components of nominal phrases, especially those elements that are not components of this nominal phrase. The elements of such discontinuous constructions can be so far apart as to make the sentence difficult to understand. Contrary to the current state of the language, such an order was very common in Middle Polish (cf. e.g. Ostaszewska 2002).

This phenomenon is known to historians of the Polish language, but the novelty of my research is that it is conducted on a dependency treebank (a syntactically annotated corpus) – the first such resource created for pre-modern Polish. Middle Polish Micro-Treebank contains 1,000 sentences selected from the Electronic Corpus of the 17th- and 18th-century Polish Texts (KorBa, www.korba.edu.pl). The treebank was created as an experimental resource; a further expansion is planned. One of the aims of the study presented here was to show what possibilities of studying historical syntax are offered by such a resource.

A dependency tree is a graph that unambiguously reflects the dependencies between the components of a sentence. It reflects e.g. the left or right position of the dependent relative to its governor (e.g. *dziwna awantura* vs. *awantura dziwna*) and the distance between them. It is worth mentioning that the graphical visualization of trees shows certain features of word order. In the case of a “typical”, linear word order, graph edges (visualized as arrows) do not intersect (as in the phrase *Każdy kraj ma przymioty* ‘**Each** **country** has qualities’). In the case of inversion and discontinuity of noun phrases with an adjective, we get trees with crossing edges (*Opaci dobrą uprowidowani fundacją* ‘**Abbots** supplied with **good** **funds**’, lit. ‘**Abbots** **good** supplied **funds**’).



Such a syntactically annotated corpus allows for various studies on the relative position of a noun and its modifier(s). In my speech, I will present some possible analyses and their results, carried out on a sample of Middle Polish texts.

References (selection)

Berdicevskis, A., and Eckhoff, H. (2020). *A Diachronic Treebank of Russian Spanning More Than a Thousand Years*. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 5251–5256, Marseille, France. European Language Resources Association.

<https://aclanthology.org/2020.lrec-1.646/>

Gruszczyński, W., Adamiec, D., Bronikowska, R., Kieraś, W., Modrzejewski, E., Wieczorek, A., and Woliński, M. (2022). *The Electronic Corpus of 17th- and 18th-century Polish Texts*. *Language Resources and Evaluation* 56, 309–332.

Klemensiewicz, Z. (1968). *Zarys składni polskiej [An outline of the Polish syntax]*. Warszawa: PWN.

Mel'čuk, I. A. (1988). *Dependency syntax: theory and practice*. Albany: SUNY Press.

Ostaszewska, D. (Ed.) (2002). *Polszczyzna XVII wieku. Stan i przeobrażenia [Polish language of the 17th century. State and transformations]*, Katowice.

Wróblewska, A., and Rybak, P. (2019). Dependency parsing of Polish. *Poznań Studies in Contemporary Linguistics*, 55(2):305–337.

The dominant-recessive hypothesis does not account for overlapping suppletion

Börjars & Vincent (2011) hypothesize that when two lexemes undergo suppletion, one is dominant and the other is recessive. The dominant lexeme is semantically more general and receives forms from the the recessive, which is less general semantically. For example, English *go* is more general than *wend*, and *go* is the lexeme that incorporated a form, *went*, of *wend*. This paper addresses several weaknesses of this proposal, especially as it relates to overlapping suppletion.

The first complication surrounding the dominant-recessive hypothesis (DRH) concerns the documentation of suppletion and its development. While many of the best-known cases of suppletion come from languages with ample historical documentation, not all examples of suppletion fit that model. Thus the DRH must be used with care so as not to become circular. A further issue with lower levels of documentation relates to situations in which some lexemes' forms survive only in suppletive paradigms. For example, the roots of English *am~be~was* are no longer represented elsewhere in the language, unlike the case of *go~went* vs *wend* mentioned above.

Cross-linguistically, separate survival of lexemes participating in suppletion is rare. Juge (1999, 2019) has identified a relatively small number of languages with overlapping suppletion, that is suppletion in which suppletive forms belong to two or more lexemes. For example, the Spanish copula *ser* 'be' and *ir* 'go' exhibit overlapping suppletion in the preterit and related paradigms (Table 1).

Such cases raise the question of how to gauge the semantic generality of the participating lexemes, in contrast with the relatively simple evaluation of verbs meaning 'go' and 'walk', for example, Juge (1999, 2019) has proposed that it is possible to correlate the semantic distance between lexemes with the non-overlapping, optionally overlapping, or non-optionally overlapping nature of the suppletive paradigms found in different cases. A key difference between these approaches is that the DRH is much more difficult to apply to situations in which the participating lexemes do not belong to the same semantic field (e.g, existence vs motion).

Furthermore, the DRH does not account for sound change as a source of suppletion. This reflects in part the traditional disregard of the various types of suppletion besides the best-known kind, incursion (cf. Juge 1999, 2013, 2019). In such cases, like the suppletive form of the present tense of the English copula, there is no evident role for semantics at all. Instances of analogically-created suppletion also challenge the DRH (along with ideas regarding the role of analogy as a regularizing process), as in the Galician (Fisterran dialect) form *iña* 'go (imperfect indicative)' created on analogy with *viña* 'come (imperfect indicative)' (Juge 2013).

The Galician cases raises a broader difficulty found not only in suppletion studies but also in grammaticalization research, namely faulty lexical semantic analysis, especially concerning what it means for one lexeme to be more general than another and how to apply such an evaluation to cross-linguistic analysis.

While the DRH may contribute to the analysis of some instances of suppletion, it must be combined with more detailed lexical semantic analysis, including measures of semantic distance, and contextualized among the multiple types of sources of suppletion.

| | <i>ser</i> † ‘to be’ | | | <i>ir</i> ‘to go’ | | |
|----|----------------------|-----------|-----------------|-------------------|---------------|-----------------|
| | present | imperfect | preterit | present | imperfect | preterit |
| 1s | soy | era | fui | voy | <i>iba</i> | fui |
| 2s | eres | eras | fuiste | vas | <i>ibas</i> | fuiste |
| 3s | es | era | fue | va | <i>iba</i> | fue |
| 1p | somos | éramos | fuimos | vamos | <i>ibamos</i> | fuimos |
| 2p | sois | erais | fuisteis | vais | <i>ibais</i> | fuisteis |
| 3p | son | eran | fueron | van | <i>iban</i> | fueron |

Table 1—Overlapping and non-overlapping suppletion (Juge 1999)

References

BÖRJARS, KERSTI & NIGEL VINCENT. 2011. ‘The pre-conditions for suppletion’, in Alexandra Galani, Glyn Hicks & George Tsoulas (eds.), *Morphology and its interfaces*. Amsterdam: John Benjamins, 239-265.

JUGE, MATTHEW L. 1999. ‘On the rise of suppletion in verbal paradigms’, in Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: BLS, 183-194.

JUGE, MATTHEW L. 2013b ‘Analogy as a source of suppletion’. In Ritsuko Kikusawa & Lawrence Reid (eds.), *Historical linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25-30 July, 2011*. Amsterdam: John Benjamins, 175-197.

JUGE, MATTHEW L. 2019. The sense that suppletion makes: Towards a semantic typology on diachronic principles. *Transactions of the Philological Society*. 117(3). 390-414. doi: 10.1111/1467-968X.12175.

Keywords

suppletion
 lexical semantics
 morphology
 dominant-recessive hypothesis

The case of Italian *segunte*: an European instance of current change from verb to demonstrative?

My proposal aims at discussing a potential instance of grammaticalization in current Italian, i.e. the change of the adjective *segunte* ‘following’ from an endophoric meaning to the function of proximal demonstrative.

There is no agreement in the literature regarding the development of demonstratives from lexical sources. Contra Diessel (2006), Heine et al. (2020) recently showed that demonstratives may originate at least from locative adverbs and from verbs, quoting many examples from studies on non-European languages. My analysis attempts to show that something very similar is occurring in current Italian, namely the potential origin of a new demonstrative from a verbal source.

Segunte is an adjective originating through transcategorization from the present participle of the verb *seguire* ‘follow’. It means “that comes immediately after in time, in space, in an ordering”. Given this lexical meaning, in its standard usage it has a mandatory cataphoric value, both as adjective and when it is nominalized, in the absence of a head noun. However, in my corpus – primarily consisting of a variety of Italian written by students (exam texts, chapters of theses, term papers, and e-mails), but also of some institutional communication texts – *segunte* also occurs either pointing to a previous referent (e.g. -Question: *Identifica il sintagma nominale nella frase seguente* (‘Identify the noun phrase in the following sentence’) -Answer (after the sentence): *Nella seguente frase...* (‘In the following sentence...’)) or without any endophoric values (e.g. *Nel seguente capitolo* (‘In the following chapter’), with reference to the chapter in which the PP occurs, not only at its very beginning).

Therefore, my hypothesis is that an innovation is taking place, and that it is in the direction of grammaticalization as a spatial deictic, specifically as proximal demonstrative, due to the presence of the feature [+ (immediately) PROXIMAL] in *segunte*.

Considering that in standard Italian the proximal demonstrative is *questo* ‘this’, my first reason is that in my data the equivalence [ART + *segunt-* (+ N)] ~ *quest-* (+ N) works everywhere. Moreover, the phenomenon fits diachronically into a process of desemantization of *segunte* that has already begun in its transcategorization from present participle to adjective (loss of causative value and intransitivization). Eventually, in my data there are bridge contexts, *segunte* appears frozen in prenominal position, there is extension of contexts (from immediately proximal to proximal deixis), and obligatory co-occurrence with the article, hence forming a phonological word.

A tentative explanation for the phenomenon is a restructuring of the demonstrative system, not unrelated to the weakening of the definite article.

References

- Croft, W. (2000), *Explaining Language Change: an Evolutionary Approach*, London, Longman.
- Da Milano, F. (2010), “Dimostrativi, aggettivi e pronomi”, *Enciclopedia dell’italiano*, Treccani. (https://www.treccani.it/enciclopedia/aggettivi-e-pronomidimostrativi_%28Enciclopedia-dell%27Italiano%29/).
- De Cesare, A. M. (2010), “Deittici”, in *Enciclopedia dell’italiano*, Treccani. (https://www.treccani.it/enciclopedia/deittici_%28Enciclopedia-dell%27Italiano%29/).
- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. (2006), “Demonstratives, joint attention, and the emergence of grammar”, *Cognitive Linguistics* 17(4). 463–489.
- Diessel, H. (2013a), “Distance Contrasts in Demonstratives”, in M. S. Dryer - M. Haspelmath (eds.), *The World Atlas of Language Structure Online* (<http://wals.info/chapter/41>).
- Diessel, H. (2013b), “Pronominal and Adnominal Demonstratives”, in M.S. Dryer - M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. (<http://wals.info/chapter/42>).
- Diessel, H. & Coventry K. R. (2020), “Demonstratives in Spatial Language and Social Interaction: An interdisciplinary Review”, *Frontiers in Psychology*, 11.
- Dixon, Robert M. W. (2003), “Demonstratives: A cross-linguistic typology”, *Studies in Language* 27, 61 – 112.
- Heine, B., T. Kuteva, H. Long, H. Narrog and Wu F. (2020), “Where do demonstratives come from?” *STUF* 73(3). 403-434.
- Ferrari, A. (2010a), “Anafora”, in *Enciclopedia dell’italiano*, Treccani, (https://www.treccani.it/enciclopedia/anafora_%28Enciclopedia-dell%27Italiano%29/).
- Ferrari, A. (2010b), “Catafora”, in *Enciclopedia dell’italiano*, Treccani, (https://www.treccani.it/enciclopedia/catafora_%28Enciclopedia-dell%27Italiano%29/).
- Giacalone Ramat, A. (1994), “Fonti di grammaticalizzazione. Sulla ricategorizzazione di verbi e nomi come preposizioni”, in Cipriano, P., Di Giovine, P. and Mancini, M. (eds.), *Miscellanea di studi linguistici in onore di Walter Belardi*, Roma, Il Calamo, vol. 2, 877-896.
- Giacalone Ramat, A. and Mauri, C. (2008), “From cause to contrast: a study in semantic change”, in Verhoeven, E., Skopeteas, S., Shin, Y.M., Nishina, Y. and Helmbrecht, J., *Studies on Grammaticalization* (eds.), Berlin, Mouton de Gruyter, 303-320.
- Heine, B., Claudi, U. and Hünemeyer, F. (1991), *Grammaticalization. A Conceptual Framework*, Chicago, University of Chicago Press.
- Hopper, P. J. e Traugott, E. C. (1993), *Grammaticalization*, Cambridge, Cambridge University Press.
- Kortmann, B (1992), “Reanalysis completed and in progress: participles as source of prepositions and conjunctions”, in Kellermann, G. and Morrissey, M. D. (eds.), *Diachrony within Synchrony. Language History and Cognition*, Frankfurt am Main, Lang, 429-453.
- Lehmann, Ch. (1995), *Thoughts on Grammaticalization*, II Ausg., München, Lincom Europa.
- Szantyka, I. A. (2017), “Lo studio morfo-sintattico dei costrutti determinanti con il pronome dimostrativo variabile *quello*”, *Lublin Studies in Modern Language and Literature* 47, 96 – 111.
- Meillet, A. (1948), “L’*évolution des formes grammaticales*”, in Meillet, A., *Linguistique historique et linguistique générale*, vol. 1°, 2 éd., Paris, Champion, 130-148.
- Ramat, P. (2005), “Su degrammaticalizzazione e transcategorizzazione”, in Ramat, P., *Pagine linguistiche. Scritti di linguistica storica e tipologica* (ed.), Roma-Bari, Laterza, 89-105.
- Vanelli, L. (1989), “Dimostrativi e articoli: deissi e definitezza”, in G. L. Borgato, A. Zamboni (eds.), *Dialettologia e varia linguistica per Manlio Cortelazzo*, Padova, Unipress, 369 - 381.

Prosody Reveals Syntactic Structure:
Secondary Predication in Metrical Finite Corpus Data

The mapping of syntax to prosody is regulated by correspondence requirements that hold between abstract syntactic structure and prosodic structure (Selkirk 2011; Elfner 2012; Ito and Mester 2013, among others). Preferentially, syntactic constituents map to prosodic constituents of the same level: morphosyntactic words (X^0) map to prosodic words (ω), syntactic phrases (XP) map to prosodic phrases (ϕ), and clauses (CP/TP) map to intonation phrases (ι). Given the crucial interaction between syntax and prosody, prosodic structure can be used to identify and differentiate syntactic structure.

Cross-linguistically, secondary predicates tend to be marked by special prosody: they are either in prominent positions and/or are subject to isolation from their respective VPs. Depictives are thought to be more prosodically independent than resultatives (cf. Irimia 2012: 208 and references therein). A further distinguishing feature of secondary predicates is the tendency to occur in STAGE-LEVEL predicates expressing a non-permanent state (Carlson 1977; Simpson 2005; Casaretto 2020). Following Kratzer (1995), STAGE-LEVEL predicates have an extra argument position for events. Likewise, secondary predicates correspond to additional syntactic structure (Kratzer 2005; Irimia 2012) which maps to a separate prosodic domain.

This study examines the distribution of secondary predication across finite metrical corpora, including the works of Homer (Greek) and the RigVeda (Vedic Sanskrit). The central goal of the present study is to identify the diagnostics for secondary predication in Greek and Vedic. Our survey suggests that secondary predicates in Greek and Vedic tend to exhibit uniform prosodic behavior. In particular, secondary predicates are separated from postverbal nominals by (i) caesura, (ii) line break, (iii) the process of *enjambment* whereby syntactic units are broken across multiple prosodic domains at the expense of Selkirk (2011)'s MATCH constraints, or a combination of these strategies. The Greek (1) and Vedic (2) data below illustrate this point.

- (1) **ton** d' ōs oun enoēse podarkēs dīos
 he.ACC but thus really see.AOR.ACT.3SG swift.NOM.SG.M divine.NOM.SG.M
 Achilles // **gymnon**
 Achilles.NOM.SG // naked.ACC.SG.M
 “now as brilliant swift-footed Achilles saw him **naked**” (II 21.49-50)
- (2) purutrā vṛtró aśayad **vyàstah**
 in.many.places Vṛtra.NOM.SG.M. lie.IMPF.3SG fling.apart.PTCP.NOM.SG.M.
 “Vṛtra lay (there), **flung apart** in many places” (RV 01.32.7d)

Prosodic isolation of secondary predicates in Greek is accomplished by enjambment: the depictive/resultative APs and postverbal nominals are parsed in different lines (1). The Vedic data in (2) demonstrate an additional isolation strategy: line-finality and post verbal position. Our findings therefore lend further support to the importance of caesurae and line-boundaries in syntactic analyses of ancient metrical corpora (Hale and Kissock 2021). The tendency to combine with STAGE-LEVEL (rather than INDIVIDUAL-LEVEL) predicates is also apparent—the secondary predicates *seeing him naked* (1) and *laying flung apart* (2) express transient properties and not permanent ones, as predicted. These facts set secondary predicates apart from attributive APs, which do not have complex syntactic structure corresponding to recursive ι domains preserved via isolation strategies in finite metrical corpora.

Selected references. Carlson, G.N. (1977). “Reference to kinds in English”. PhD thesis, UMass Amherst. ■ Casaretto, A. (2020). “Onsecondary predicates in Vedic Sanskrit–Syntax and Semantics” IJDL. ■ Hale, M. and M. Kissock (2021). “n the syntax of comparative clauses in Vedic Sanskrit. . . like someone eating the foam off the water”. *Trends in South Asian Linguistics* 367. ■ Kratzer, A. (1995). *Stage-level and individual-level predicates*. ■ Kratzer, A. (2005). “Building resultatives”. In Maienborn, Claudia and Angelika Wöllstein, *Event arguments: foundations and applications*. ■ Selkirk, E. (1996). “The prosodic structure of function words”. In: *Signal to syntax*. ■ Selkirk, E. (2011). “The syntax–Phonology Interface”. *The handbook of phonological theory*.

Quality vs. quantity: Contrast maintenance and tradeoff in Southwestern Tai vowels

Contrastive vowel length is frequently found in languages in the Tai branch of the Kra-Dai family, and has been reconstructed to Proto-Tai (Pittayaporn 2009b). In many Tai languages, however, only one length contrast remains, between /a/ and /a:/. In this study, we explore the loss of the last standing phonemic length contrast in the Upper Chindwin dialect of Tai Khamti (TK), a member of the Southwestern Tai subgroup, which is spoken in Khamti Township, Sagaing Region, Myanmar. Draw on evidence from both reconstruction and original fieldwork, we show that this quantity contrast is undergoing a compensatory change toward a quality contrast in this dialect.

Background. Prototypical Southwestern Tai (SWT) languages have nine phonemic vowel qualities forming a schematic 3x3 grid of height and backness. Pittayaporn (2009a, 2009b) reconstructs 7 quality contrasts and 7 length contrasts in Proto-Tai, which became 9 quality contrasts and 4 length contrasts in its descendant Proto-Southwestern Tai (see **Tables 1a-b**), which can be interpreted as a historical tradeoff whereby a reduction in the number of quantity contrasts went hand-in-hand with an increase in the number of quality contrasts. However, most modern SWT languages preserve at least one length contrast, regardless of how many of the 9 quality contrasts they preserve.

| | | |
|---------|---------|---------|
| *i, *i: | *u, *u: | *u, *u: |
| *e, *e: | *ɤ, *ɤ: | *o, *o: |
| | *a, *a: | |

Table 1a. PT vowels per Pittayaporn 2009b.

| | | |
|---------|---------|---------|
| *i, *i: | *u, *u: | *u, *u: |
| *e: | *ɤ: | *o: |
| *ɛ: | *a, *a: | *ɔ: |

Table 1b. PSWT vowels per Pittayaporn 2009a.

In the other direction, Bangkok Thai, the most populous and best studied Tai language, has accrued quantity contrasts apparently without any requisite tradeoff toward fewer vowel qualities. Sukhothai Thai, a closely related historical lect reconstructed from surviving texts, had 7 contrastive length contrasts circa the 13th-15th centuries (Maspong 2015), and Bangkok Thai expanded length contrasts to all 9 simple vowels by the late 19th or early 20th century (Author xxxx). However, even within the duration contrasts of Thai, vowel quality may act as a secondary or redundant cue.

Data. Recordings of 9 TK speakers ages 24 to 78 were made at 5 locations in 2014-2015. Each speaker recorded a 436-item Southeast Asia wordlist (SIL 2002), with 1-3 repetitions for each lexical item. The lexical material and token counts vary somewhat between speakers, but the total corpus used in this study has approximately 2,000 tokens each of /a/ and /a:/.

| All speakers | /a/ | /aa/ | all V | LP5 (24F) | /a/ | /aa/ | all V |
|-----------------|------------|------------|-------|-----------------|------------|------------|-------|
| all contexts | 94 | 175 | 156 | all contexts | 107 | 153 | 144 |
| sonorant codas | 107 | 147 | 130 | sonorant codas | 108 | 123 | 120 |
| obstruent codas | 96 | 105 | 125 | obstruent codas | 100 | 95 | 121 |

Table 2a-b. Vowel duration in milliseconds for all 9 TK speakers vs. speaker LP5.

Results. On the dimension of quality, **Figure 1** plots F1 and F2 for the 9 Tai Khamti speakers, and shows varying degrees of divergence between the two vowel means, but with several speakers showing large vowel quality differences. On the dimension of quantity, **Table 2a** shows that across the sample population that the length contrast has been fully neutralized before obstruent codas, but there is on average a 40ms difference before sonorant codas. Further, **Table 2b** shows that for certain speakers, such as LP5, the length contrast is becoming neutralized before sonorant codas, with an average difference of just 15ms between /a/ and /a:/. Both lines of evidence, from duration and formants, suggest that the total loss of contrastive vowel length is complete or nearly so among some members of the Tai Khamti speaker community, which we predict will continue to progress in the coming years. At the same time, this contrast is apparently being maintained via cues from vowel quality. This represents an interesting case of compensatory change, unusual among the Southwestern Tai languages.

References

Author. xxxx.

Jonsson, Nanna L. 1991. Proto Southwestern Tai. Ph.D. dissertation, State University of New York at Albany.

Li, Fang Kuei. 1977. A handbook of comparative Tai. Ocean Linguistics Special Publications 15.

Maspong, Sireemas. 2016. Vowel length contrasts in the language of Sukhothai period. M.A. thesis, Chulalongkorn University. (in Thai)

Pittayaporn, Pittayawat. 2009a. Proto-Southwestern Tai revisited: a new reconstruction. *Journal of the Southeast Asian Linguistics Society* 2:121-143.

Pittayaporn, Pittayawat. 2009b. The phonology of Proto-Tai. Ph.D. dissertation, Cornell University.

Sarawit, Mary Elizabeth Sautter. 1973. The Proto-Tai vowel system. Ph.D. dissertation, University of Michigan.

Strecker, David. 1983. Proto-Tai vowels revisited: a comparison and critique of the work of Sarawit and Li. *Linguistics of the Tibeto-Burman Area* 7(2): 33-74.

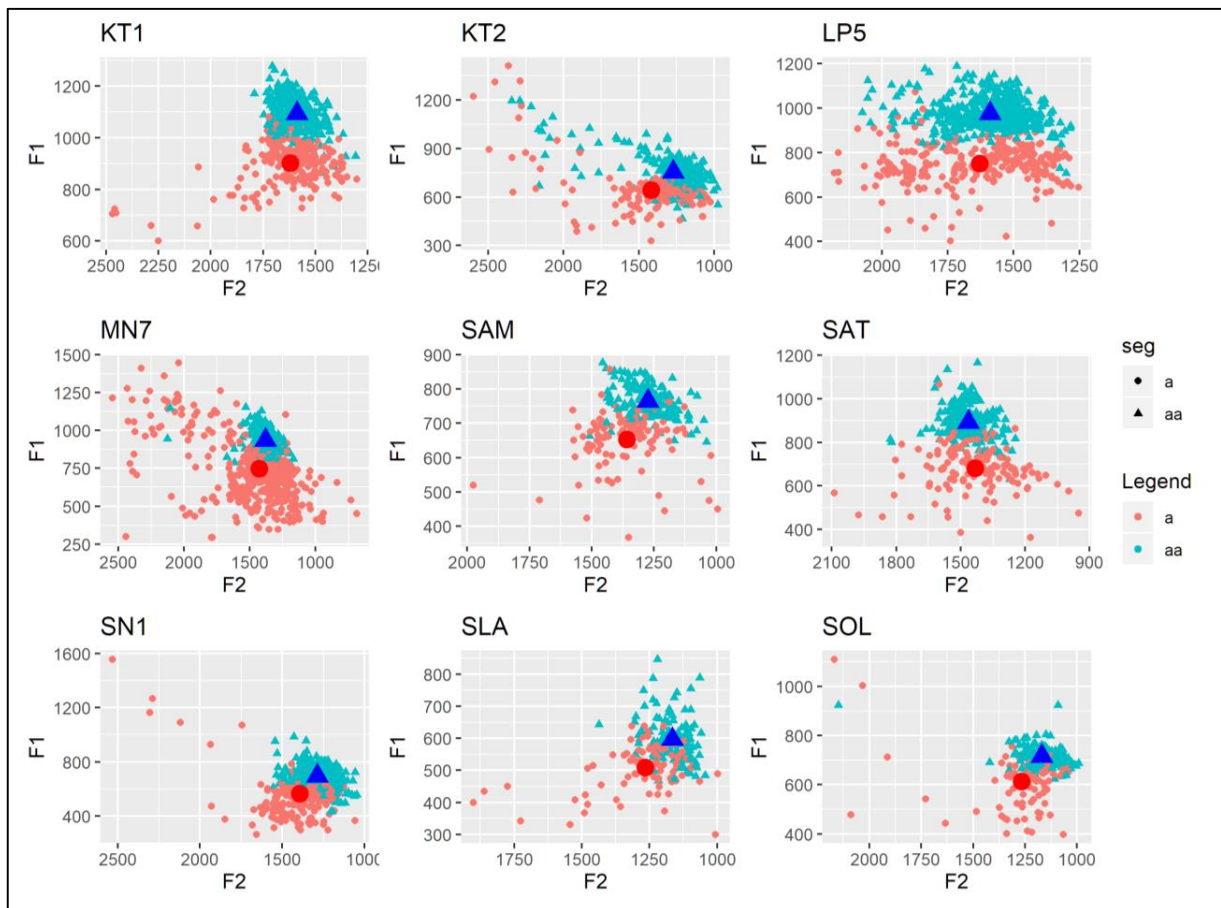


Figure 1. Comparison of F1 and F2 formants, with means, of /a/ and /a:/ in 9 Tai Khamti speakers.

Morphologization of Phonological Processes as Integration

The proposed paper presents a view of morphologization of phonological processes in which various degrees of integration of these processes into the morphology are possible, and follow a potential trajectory of increasing embeddedness. Debates on morphologization have focused on questions such as whether a phonological process needs to have become inactive, unproductive, or opaque as a precondition for morphologization (cf. e.g. Maiden 1991/2011). While this paper addresses the latter question likewise, the focus here is on how morphological systems integrate and absorb the phonological process and how they interact with the phonology thereafter. Different degrees of integration of phonological processes will be demonstrated with morphologization of vowel harmony in Turkish, consonant mutation in Welsh, fricative voicing in English, and German "umlaut".

In the following, we are assuming a modular grammar including a dynamic morphological component interfacing with lexicon, phonology, and syntax. Further, we assume that the phonology can only "see" phonological structures and only these can provide environments for phonological processes. If a process is restricted to specific morphological or lexical environments, the executive function over it belongs in the morphology and/or lexicon. The process itself will be applied by the phonology, but only under external orders.

As a first step towards morphologization of phonological processes, we can suggest that even prior to any loss of transparency/regularity or phonemicization of erstwhile allophones, there may develop an association between the phonological process and one or more morphological ones. Thus, for example, we would expect that an affinity between vowel-fronting and 'plural' would have been noted by speakers before the weakening of unstressed /i/ to schwa and thus loss of motivating environment in German, likewise between fricative voicing and 'plural' before fricative voicing became inactive as a phonological process in English. This "noting" could take the form of adding a redundant command to the phonology to the morphological operations creating plurals via suffixation etc. Such a scenario is preferable to one in which speakers are staring at the extra fronted vowels or voiced fricatives trying to find a use for them after the demise of productive phonological processes, i.e. wondering what to do with the resulting "junk" (Lass 1990).

While the cases of German umlaut and English fricative voicing could begin with "affinities" and be incorporated into already existing morphological processes such as affixation, ultimately potentially becoming the lone process corresponding to a particular morphological function, e.g. mapping onto 'plural', other pathways may involve rather more general associations with morphological functionality. In Turkish, for example, vowel harmony is observable in most words of two or more syllables, so that its value for associated morphological operations could not be much more concrete than 'I am a word(-form)'/ 'this is a word', i.e. a support for morphological structure in a general sense. Likewise with Welsh consonant mutation: the phonological process affected the entire obstruent series to begin with (Jackson 1953), so the syntactic contexts in which the initial consonant of some element of a construction was modified were particularly varied; here the contribution to the morphology would be something like 'I'm in construction' or 'this is a compound' etc. While Turkish vowel harmony may never progress beyond playing a supporting role in inflection and derivation, there is potential even for processes whose initial morphological function was very general to be "promoted" to the status of sole process in a morphological operation, as can be observed in Modern Welsh where in some dialects the nasal consonant alternation of a word-initial consonant has become the sole expression of 1 p.sg. possession ([kar] 'car', [vəŋar] 'my car' > [kar] 'car' ~ [ŋar] 'my car' (Jones 1998)). This is, of course, the highest degree of embedding into the morphology a once phonological process can undergo. In principle, then, a phonological process can be integrated first as a support to one or more morphological processes with only a very indirect relationship to morphological semantics, then in time become more closely associated with existing morphological semantic functions, eventually to figure as a morphological process with full status. However, long-term stability can be achieved at any of these stages.

References Cited

- Jackson, K. 1953. *Language and history in early Britain*. Cambridge: Harvard University Press.
- Jones, M. 1998. *Language obsolescence and revitalization: linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Clarendon Press.
- Lass, R. 1990. "How to do things with junk: Exaptation in language evolution." *Journal of Linguistics* 26:79-102.
- Maiden, M. 1991. *Interactive morphonology: Metaphony in Italy*. London: Routledge.

The expression of predicative possession in Avestan

While the linguistic expression of the predicative possession has been widely investigated in the languages belonging to the major Indo-European branches such as Ancient Greek (Benvenuto&Pompeo 2012, Kulneff-Eriksson 1999), Latin (Baldi&Nutti 2010, Bolkestein 1983, 2001), Old Indian (Bauer 2000, Danesi&Barðdal 2018), in the case of the Ancient Indo-Iranian group these kinds of constructions have been observed by some scholars in previous studies, but have not been discussed in great depth. Regarding such earlier research, it is worth mentioning Benveniste’s well-known 1960 study where, on the basis of examples from many Indo-European languages including Old Persian and Avestan, the predicative constructions with dative and with genitive are classified respectively as «*prédicat de possession*», expressing “possession”, and as «*prédicat d’appartenance*», expressing “belonging”. Èdel’man’s 1975 paper and Bauer 2010 also deserve mention, since they contain a brief overview over the expressions of possession in the Iranian languages.

The aim of this paper is to provide a detailed account of the functional distribution of predicative possessive constructions in the Avestan language,

The investigation of the expressions of possession will be based on the situation documented in the Old Avestan texts. According to Kellens and Pirart (1988), the texts which are definitely Old Avestan are: the five *Gāθās* (Y. 28-35, 43-51, 53), the *Yasna Haptañhāiti* (Y. 35.2-41.5), two fragments, that is to say, the *Ahuna Vairiia* (Y. 27.13) and the *Airiiaman Išiiia* (Y. 54.1); Y. 27.7 and Y. 56.1. The analysis will be focused on the two possessive constructions “verb to be plus dative” as in (1) and “verb to be plus genitive” as in (2), in order to identify the differences in their syntactic and pragmatic functions and in order to examine them from a semantic perspective.

(1) Av. Y. 62.1 *ušta buiiāṭ ahmāi naire*
 “Salvation be to this man”

(2) Av. Y. 43.7 *ciš ahī kahiiā ahī*
 “Who are you? Whose are you?”

In order to highlight the syntactic features of the two constructions we will take into account syntactic functions and the semantic roles of the Avestan genitive and dative, the relevant statistical tendencies regarding word order, focusing on the pragmatic role of the constituents.

References

- Baldi, Philip&Pierluigi Cuzzolin (2005) Considerazioni etimologiche, areali e tipologiche dei verbi di «avere» nelle lingue indoeuropee. In: Sándor Kiss, József Herman, Luca Mondin&Giampaolo Salvi eds., *Latin et langues romanes: études de linguistique offertes à József Herman à l'occasion de son 80ème anniversaire*. Tübingen: 27-36.
- Baldi, Philip&Andrea Nuti (2010) Possession. In: Philip Baldi&Pierluigi Cuzzolin eds., *New perspectives on historical latin syntax constituent syntax: quantification, numerals, possession, anaphora* Berlin: 239-388.
- Bauer, Brigitte (2000) *Archaic syntax in Indo-European: the spread of transitivity in Latin and French* (Trends in linguistics. Studies and monographs; 125). Berlin.
- Benveniste, Emile (1960) “Être” et “Avoir” dans leurs fonctions linguistiques *Review of Bulletin de la Société de Linguistique de Paris* 55/1. 113-134.
- Benvenuto, Maria Carmela&Flavia Pompeo (2012) Expressions of predicative possession in Ancient Greek: "εἶναι plus dative" and "εἶναι plus genitive" constructions. *Review of AION. Annali dell'Università degli Studi di Napoli «L'Orientale»*. Dipartimento di Studi Letterari, Linguistici e Comparati 1: 77-103.
- Bolkestein, A. Machtelt (1983) Genitive and Dative Possessors in Latin. In: S. C. Dik ed., *Advances in functional grammar*. Dordrecht: 55-91.
- Bolkestein, A. Machtelt (2001) Possessors and experiencers in Classical Latin. In: Irène Baron, Michael Herslund&Finn Sørensen eds., *Dimensions of possession*. Amsterdam: 269-283.
- Cantera, Alberto (2012) *The transmission of the Avesta*, (Iranica). Wiesbaden.
- Danesi, Serena&Jóhanna Barðdal (2018) Case marking of predicative possession in Vedic: The genitive, the dative, the locative. In: Barðdal, Jóhanna, Na'ama Pat-El&Stephen M. Carey eds., *Non-canonically case-marked subjects: the Reykjavík-Eyjafjallajökull papers*. Amsterdam, Philadelphia: 181–212.
- Èdel'man, Džoj I. (1975) Les verbes 'être' et 'avoir' dans les langues iraniennes. In: Françoise Bader ed., *Mélanges linguistiques offerts à Emile Benveniste*. Louvain: 151-158.
- Isačenko, Alexander (1974) On have and be Languages: A Typological Sketch. In: Michael S. Flier ed., *Slavic Forum: Essays in Linguistics and Literature*. The Hague-Paris: 43-77.
- Kellens, Jean&Eric Pirart (1988) *Les textes vieil-avestiques*. Wiesbaden.
- Kulneff-Eriksson, Karin (1999) *On 'have' in Ancient Greek. An investigation on εἶναι and the construction εἶναι with a dative as expressions for 'have'*, (Studia Graeca et Latina Lundensia 7). Lund.
- Skjærvø, P. Oktor (2009) Old Iranian: Avestan and Old Persian. In: Gernot Windfuhr ed., *The Iranian Languages*. London – New York.

Computational Approaches for Romance Related Words Discrimination

Abstract

Natural languages are living eco-systems, they are constantly in contact and, by consequence, they change continuously. Traditionally, the main problems in historical linguistics (“How are languages related?”, “How do languages change across space and time?”) have been investigated with comparative linguistics instruments. The main idea of the comparative method is to perform a property-based comparison of multiple sister languages in order to infer properties of their common ancestor. It is a time-consuming manual process that required a large amount of intensive work.

The identification of cognates is a fundamental process in historical linguistics, on which any further research is based. On the other hand, discriminating between lexical borrowings and inherited words is considered one of the most difficult and important tasks in HL (Jäger, 2019), for which “the computerised approach” is regarded as the appropriate solution even by classical linguists (Heggarty, 2012). We propose here computer-assisted methods for identifying cognates, for discriminating between cognates and borrowings, and for discriminating between inherited and borrowed Latin words.

Firstly, we introduce a method to automatically determine if a pair of words (u, v) are cognates or not, and we use it on a large database comprising the main Romance languages (Romanian, Italian, French, Spanish and Portuguese), applying it as well in subsequent tasks. Given an input pair of words, the initial task is to automatically determine if they are cognates or not. We developed a machine learning method for automatically producing the answer based on sequence alignment. To align pairs of words, we employed the Needleman Wunsch global alignment algorithm, which has been successfully used in natural language processing and computational biology. We used words as input sequences and a basic substitution matrix, which gives equal scores to all substitutions, disregarding diacritics (e.g., we ensure that e and \acute{e} were matched). For the machine learning part, we used an ensemble of methods. We applied our method to multiple data sets, showing that our approach improves on previous results, also having the advantage of requiring less input data, which is essential in historical linguistics, where resources are generally scarce. In the process of discriminating between cognate and borrowing, we tried to answer the following question: given a pair of words, are they cognates, borrowings, or neither? For the automatic discrimination between inherited and borrowed Latin words, the best results were obtained by a system based on SVM using features extracted from the word-etymon pairs. We apply our method on both graphic and phonetic forms of the words.

Keywords

Romance languages, cognates, borrowings, inherited words.

Acknowledgements

This research is supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI UEFISCDI, project number 108, CoToHiLi, within PNCDI III.

References

[1] Ciobanu, Alina Maria and Liviu P. Dinu. 2019. Automatic Identification and Production of Related Words for Historical Linguistics. *Computational Linguistics*, vol. 45, No. 4, pages 667–704.

- [2] Alina Maria Cristea, Liviu P. Dinu, Simona Georgescu, Mihnea-Lucian Mihai and Ana Sabina Uban, 2021. Automatic Discrimination between Inherited and Borrowed Latin Words in Romance Languages In: *Proc. EMNLP 2021(Findings)*, Punta Cana, 2021
- [3] Alina Maria Ciobanu and Liviu P Dinu, 2014. Automatic detection of cognates using orthographic alignment. In *Proc. ACL 2014*, June 22-27, 2014, Baltimore, MD, USA
- [4] Alina Ciobanu and Liviu P Dinu, 2015. Automatic discrimination between cognates and borrowings. In *Proc. ACL 2015*, July 26-31, 2015, Beijing, China
- [5] Gerhard Jäger. 2019. Computational Historical Linguistics. *Theoretical Linguistics* | Volume 45: Issue 3-4, 45: Issue 3–4.
- [5] Paul Heggarty. 2012. Beyond Lexicostatistics: How to Get More out of "Word List" Comparisons. In *Quantitative Approaches to Linguistic Diversity: Commemorating the Centenary of the Birth of Morris Swadesh*, pages 113–137. Benjamins

Emergence of alternate argument alignment patterns in Northwest Kainji

Northwest Kainji (NWK, Nigeria; Benue-Congo) languages display at least two argument alignment patterns: accusative and neutral. Head nouns in NWK occur with either prefixed or suffixed adnominal gender marking in most noun phrases (see Hoffmann 1967 and Dettwelier 2015 for Dakakari/C'Lela [dri]; Bendor-Samuel et al. 1973 for Duka/Ut-Hun [uth]; Author (2007, 2019) for Ut-Ma'in [gel], and D. Heath 2020 for Us-Saare [uss]). Labeled as ambifixes by Arkadiev (2022, mention of Ut-Ma'in [gel]) and typed as “clitic-like” by Güldemann & Fiedler (2022, mention of Dakakari/C'Lela [dri]), these powerhouse morphemes manifest grammatical categories of number (singular/plural/mass), class membership (including derivational functions adding designations of humanness, animacy, size, and shape), and, in certain morphosyntactic configurations, they flag grammatical role. However, affix placement, i.e., where, and whether or not, a particular form of adnominal marker occurs, is governed by the morphosyntactic configuration of the noun phrase. Further, the argument alignment pattern that occurs in any given clause also depends on the internal structure of a particular NP, see (i) and (ii), agnostic of grammatical role. That is, the argument flagging function on the adnominal markers is present, but potentially secondary to other referential functions.

- (i) Unexpanded NPs display **accusative alignment** by means of adnominal marking:
 - a. Unexpanded **subject** NPs occur with an adnominal gender **suffix** (nominative), (1–2);
 - b. Unexpanded **object** NPs occur with an adnominal gender **prefix** (accusative), as in (3).
- (ii) NPs that are expanded in any way, i.e., contain modifiers, display **neutral alignment** (4–6).

Two alignment patterns also exist to varying degrees within the pronominal system. In some NWK, noun class agreement pronouns manifest accusative alignment with distinct forms for subject and object. Most personal pronouns display neutral alignment; however, first person singular pronouns have distinct forms for subject.

- (1) *sē* [kó:t-jǝ] *rwōn* *ōr-vástè*
 then guinea.fowl-C7.NOM exit C5-last
 ‘Then a guinea fowl exited last.’ (Ut-Ma'in [gel], Author 2019: 104)
- (2) *kòná* [kó:t-jǝ] *zǝ-t.è...* *ōr-kjàt ...*
 there guinea.fowl-C7.NOM say-PRF C5-difficult
 ‘There a guinea fowl has said, “Difficult...”’ (Ut-Ma'in [gel], Author 2019: 104)
- (3) *á=b* *hján* [*ū-kó:t*]
 COND=2SG see C7.ACC-guinea.fowl
 ‘If you see a guinea fowl, ...’ (Ut-Ma'in [gel], Author 2019: 104)
- (4) [*jà=t-ǝ=s-té=tǝ*] *āzgōssè*
 fruit=C6-ASSOC=C4-tree=C6.DEF roll.out
 ‘Those fruit rolled out.’ (Ut-Ma'in [gel], Author 2019: 102)
- (5) [*jà=t-ǝ=s-té=tǝ*] *fámǝ* *t-móǝré*
 fruit=C6-ASSOC=C4-tree=C6.DEF resemble C6-mango.fruit
 ‘Those fruits resemble mangoes.’ (Ut-Ma'in [gel], Author 2019: 102)
- (6) *é=kár-g-ǝs:* [*jà=t-ǝ=s-té=tǝ*]
 C2.SUBJ=pick-PST-ITR fruit=C6-ASSOC=C4-tree=C6.DEF
 ‘They gathered those fruits.’ (Ut-Ma'in [gel], Author 2019: 102)

All NPs that demonstrate traditionally understood accusative alignment, may better be understood as demonstrating König's (2008: 8, 158) Type 2 Marked Nominative system, crucially because of the wide range of functions that the accusative (prefix) form fills. In contrast, the nominative form has a very restricted set of functions. In this paper I present the morphosyntax of these alignment systems across four NWK languages, survey the morphosyntactic conditions that determine when the alternate patterns arise in each language, and propose rationale for the development of the suffixed marked nominative pattern.

References:

Author. (2007). ...

Author. (2019). ...

Arkadiev, Peter, and Yuri Lander. (2022). *Ambifixes: Cross-Linguistic Variation and Possible Origins*.

Presented at the Princeton Phonology Forum 2022. Princeton University. December 2.

Bendor-Samuel, John, Donna Skitch and Esther Cressman. (1973). *Duka sentence, clause and phrase*. *Studies in Nigeria Languages* 3. Zaria, Nigeria: Institute of Linguistics.

Dettweiler, Stephen H. 2015. C'Lela Grammar Portrait. Electronic ms.

<https://www.sil.org/resources/archives/63135>.

Güldemann, Tom, and Ines Fiedler. (2022). *Cliticization cycles: Adnominal gender-number affixes in Niger-Congo*.

Presented at the 25th International Conference on Historical Linguistics, University of Oxford. August 2.

Heath, David M. (2020). U̅s-Saare Phonological and Grammar Sketch. Electronic ms.

<https://www.sil.org/resources/archives/87200>.

Hoffmann, Carl. (1967). An outline of the Dakarkari noun class system and the relation between prefix and suffix noun-class systems. *La classification nominale dans les langues négroafricaines*, ed. G. Manessy, 237-259. Paris: CNRS.

König, Christa. (2008). *Case in Africa*. Oxford: Oxford University Press.

Drivers of Diversity in the Construal of Quantity in the World's Languages

How have almost all the world's cultures developed diverse but conceptually related systems for exact reference to quantity? Observations of numeral and lexicogrammatical number systems for quantification influence the following conclusions about the origin and diversity of exact quantification resources. Numerals originate from counting practices and evolve through recursion (Wiese, 2007). Number, as a category of grammar, develops from lexical resources such as numerals and deictic expressions (Aikhenvald, 2018; Corbett, 2000). To account for the diversity in distribution of quantification resources across natural language, some studies invoke the complexity trade-off thesis: complexity in a system of quantification trades off complexity in another. For example, languages with numeral classifiers are said to have low-limit cardinal numeral systems and lack facultative plural number marking (Greenberg, 1987; Croft, 1994; Aikhenvald, 2000). Sociocultural and ecological factors such as the influence of climate and agricultural practices (Divale, 1999; Epps et al., 2012), as well as cognitive constraints like the economy of expression constraints (Haspelmath and Karjus, 2017) also drive diversity. The theories about diversity, however, seems to be based on a non-representative sample of the world's languages and often do not integrate influences of genealogy and contact, the key drivers of language diversity in Historical Linguistics framework. The association between diversity in numeral system and the full range of diverse lexicogrammatical number marking resources have not been explored. This study investigates the drivers of diversity in the distribution of resources for construing exact quantity in a representative sample of the world's languages, considering trade-offs and genealogy influences.

The study takes a diachronic approach to test whether (1) numeral system complexity trades-off lexicogrammatical number complexity over time. Numeral system complexity is coded as a continuous variable by measuring the restrictedness of the numeral system, presence of numeral classifiers and the robustness of the base system on a scale of 0-1. Data on numerals and grammatical number are respectively obtained from Numeralbank (Barlow et al., 2020) and World Atlas of Classifier Languages (WACL) (One-Soon, 2022). Grammatical number data from Grambank Consortium (2021) and World Atlas of Language Structures (WALS) are used to measure lexicogrammatical number complexity on a scale of 0-1 considering the presence and diversity of number features (e.g., singular, dual).

The study examines this trade-off in four geographically distributed language families: Sino Tibetan, Austronesian, Pama-Nyungan and Atlantic Congo. The evolution of quantification is modelled in two ways: An independent model of the evolution of numeral system complexity and grammatical number complexity and a dependent model of evolution. The study uses the phylogenetics Generalised Least Squares method and maps the data on Bayesian Phylogenetic tree data of the respective families. The talk presents preliminary results that hint a family specific co-evolution pattern.

References

- Aikhenvald, A. Y. (2018). Number systems in grammar - position paper. *Language and Culture Research Centre: 2018 Workshop*
- Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford & New York: Oxford University Press.
- Barlow, R. & Beller, S. & Bender, A. & Bibiko, H. & Forkel, R. & Greenhill, S. & Gray, R. & Hammarström, H & Jordan, F. & Rzymski, C. & Verkert, A. (eds.) (2020). *Numeralbank*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (available online at <http://numerals.clld.org>).
- Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.
- Croft, W. (1994). Semantic universals in classifier systems. *Word* 45.2: 145-171.
- Divale, W. (1999). Climatic instability, food storage, and the development of numerical counting: A cross-cultural study. *Cross-Cultural Research*, 33, 341-368.
- Epps, P., Bower, C., Hansen, C. A., Hill, J. H., & Zentz, J. (2012). On numeral complexity in hunter-gatherer languages. *Linguistic Typology*, 16, 41-109.
- Greenberg, J. (1974). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Proceedings of the 11th International Congress of Linguistics, Bologna / Florence, Aug-Sept 1972*. Bologna: 1974: 17-37.
- Haspelmath, M. & Karjus, A. (2017). Explaining asymmetries in number marking: singularatives, pluratives and usage frequency. *Linguistics*, 55(6): 1213-1235.
- One-Soon, Hammarström, H, Allasonnière-Tang, Marc, (2022). Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard*
- The Grambank Consortium, (ed.). 2021. *Grambank 1.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://grambank.clld.org> (accessed 4 July 2020).
- Wiese, H. (2007). The co-evolution of number concepts and counting words. *Lingua*, 117, 758-772.

Proto-Malayo-Polynesian: Some Phonetic Evidence for *l

Philippine languages are considered strong evidence for assertions about reconstructed Proto-Malayo-Polynesian (PMP). We address the phonetic articulation of Proto-Austronesian *l in PMP. We show reflexes via reported articulatory data from nine Philippine languages including both the north and south Philippine languages (Olson et al. 2010). Audio recordings support our work in three languages (Olson et al. 2009, 2008), one of which has also undergone an ultrasound study (Mielke et al. 2011). We suggest that a novel articulation (an interdental approximant) was in common use as far back as PMP, minimally in allophonic distributions, but possibly as the default pronunciation. Considering this evidence should cause us to carefully reconsider how we reconstruct words within PMP and higher reconstructions. The alternative hypothesis is to find pathways for independent innovations around the periphery of an archipelago in nine different Philippine languages. The proposed historical articulation, an interdental approximant, is phonetically rare in the world's languages, but not unattested (Everett 1982; Harley 2012). It is therefore likely overlooked for its historical significance in reconstructing PMP. We suggest that the diverse reflexes (ɭ, l, d, Ø, ʔ, n, y, ɔ̃) for *l in Austronesian languages are based on the most salient acoustic cues for the specific speech communities. Zorc (1975: 264-6) acknowledges the irregular correspondences involving liquids but leaves the diversity unexplained. One possibility, which we advocate for, is that these variations are all strategies to bring the tongue inside the mouth due to social pressures while maintaining an auditory or perceptual cue. By appealing to visual motivations for sound change we link sound change process to language as a multi-modal experience (Vigliocco 2014; Ambrazaitis & House 2017; Perniss 2018). Finally following Havenhill and Do (2018), we agree that linking historical change phenomena in primarily oral-languages to how they are visually experienced leads scholars to a deeper understanding of community communication practices through time.

Correspondences for the interdental and PAN *l

| English | Butbut | Lubuagen | Majukayong | Minangali | Kagayanen | Kalagan | Southern Catanduanes Bicolano | Blust (1999) |
|---------|-----------------------|----------|------------|-----------|-----------|---------|-------------------------------|-----------------|
| [eng] | [kyb] | [knb] | [knb] | [knb] | [knb] | [knb] | [knb] | [knb] |
| three | tu'ɔ̃u | ti'ɔ̃u | tu'ɔ̃u | tuɔ̃u | 'tallo | toɔ̃o | tuɔ̃u | *telu |
| moon | 'h ^w uɔ̃an | 'buɔ̃an | so'ɔ̃ag | soɔ̃ag | 'buɔ̃an | boɔ̃an | buɔ̃an | *bulaN / *qiNas |
| path | 'tʃaɔ̃an | 'keɔ̃sa | 'qaɔ̃sa | ʔaɔ̃sa | 'daɔ̃an | daɔ̃an | daɔ̃a | *zalan |

References

Ambrazaitis, Gilbert, and David House. 2017. "Multimodal Prominences: Exploring the Patterning and Usage of Focal Pitch Accents, Head Beats and Eyebrow Beats in

- Swedish Television News Readings.” *Speech Communication* 95: 100–113. doi:10.1016/j.specom.2017.08.008.
- Blust, Robert. 1999. “Subgrouping, Circularity and Extinction: Some Issues in Austronesian Comparative Linguistics.” In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, edited by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert, and Stephen Trussel. 2013. “The Austronesian Comparative Dictionary: A Work in Progress.” *Oceanic Linguistics* 52 (2). University of Hawai‘i Press: 493–523. <https://www.jstor.org/stable/43286361>.
- Everett, Daniel Leonard. 1982. “Phonetic Rarities in Pirahã.” *Journal of the International Phonetic Association* 12 (2): 94–96. doi:10.1017/S0025100300002498.
- Harley, Matthew. 2012. “Unusual Sounds in Nigerian Languages.” In *Advances in Minority Language Research in Nigeria*, edited by R. Blench and Stuart McGill, 38–65. Kay Williamson Educational Foundation: African Languages Monographs 5. Köln: R. Köppe.
- Havenhill, Jonathan, and Youngah Do. 2018. “Visual Speech Perception Cues Constrain Patterns of Articulatory Variation and Sound Change.” *Frontiers in Psychology* 9: Article 728. doi:10.3389/fpsyg.2018.00728.
- Mielke, Jeff, Kenneth S. Olson, Adam Baker, and Diana Archangeli. 2011. “Articulation of the Kagayanen Interdental Approximant: An Ultrasound Study.” *Journal of Phonetics*, *Speech Reduction*, 39 (3): 403–12. doi:10.1016/j.wocn.2011.02.008.
- Olson, Kenneth S., Emy T. Ballenas, and Nilo M. Borromeo. 2009. “Buhi’non (Bikol) Digital Wordlist: Presentation Form.” *Language Documentation & Conservation* 3 (2): 213–25. <http://hdl.handle.net/10125/4440>.
- Olson, Kenneth S, Glenn Machlan, and Nelson Amangao. 2008. “Minangali (Kalinga) Digital Wordlist: Presentation Form.” *Language Documentation & Conservation* 2 (1): 146–56. <http://hdl.handle.net/10125/1772>.
- Olson, Kenneth S., and Jeff Mielke. 2008. “Acoustic Properties of the Interdental Approximant in Kagayanen.” *The Journal of the Acoustical Society of America* 123 (5). Acoustical Society of America: 3460–3460. doi:10.1121/1.2934307.
- Olson, Kenneth S., Jeff Mielke, Josephine Sanicas-Daguman, Carol Jean Pebley, and Hugh J. Paterson III. 2010. “The Phonetic Status of the (Inter)Dental Approximant.” *Journal of the International Phonetic Association* 40 (2): 199–215.
- Perniss, Pamela. 2018. “Why We Should Study Multimodal Language.” *Frontiers in Psychology* 9: Article 1109. doi:10.3389/fpsyg.2018.01109.
- Vigliocco, Gabriella, Pamela Perniss, and David Vinson. 2014. “Language as a Multimodal Phenomenon: Implications for Language Learning, Processing and Evolution.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130292. doi:10.1098/rstb.2013.0292.
- Zorc, David R. 1975. “The Bisayan dialects of the Philippines: subgrouping and reconstruction.” Ph.D. dissertation, Cornell University

Hypotheses and scenarios in North Germanic Tonogenesis

The presence of a lexical tonal distinction in North Germanic (*accent 1* vs. *accent 2*) occasions the following two things to explain (the explananda):

- A) The origin of a lexical representation
- B) The origin of a lexical distinction

A lexical *representation* is here a lexical tone or a lexical foot (inducing a tonal effect). A lexical *distinction* is here a contrast that is expressed in the tonal structure, and which isn't predictable from non-lexical information. An example would be the tonal contrast between the monomorphemes 'ketchup' 'id.' (accent 1) and 'senap' 'mustard' (accent 2) in Swedish. One of the members in this pair contains some lexical property that causes the tonal melodic difference. That lexical property is part of the representation.

Lexical distinctions are tied to lexical representations and that gives us a logical order for the explananda. We should have an account for explanandum A in order to properly address explanandum B, while the reverse does not hold.

Nevertheless, research tradition seems to have given primacy to the explanation of B, the origin of a lexical distinction, over and above A, the origin of a lexical representation. This is apparent in the importance given to two historical changes that are taken to be crucial for the development of the contrast in simplex forms (ca 1000–1200 AD): cliticization of the definite article (*and hinn* > *and-en* 'the mallard'), and epenthesis before sonorants (*segl* > *segel* 'sail'), both resulting in disyllabic forms with accent 1, which come to contrast with accent 2, taken to previously dominate in polysyllables.

I discuss the consequences of both orderings among the explananda, i.e., B>A and A>B. I contend that the B>A stance at best provides a *scenario* for explanandum A, but not a proper hypothesis. A *hypothesis* for explanandum A should propose to explain 1) how the marked tonal contour comes into being, and 2) how it gets phonologized in the relevant forms. The B>A stance fails on the phonologization issue.

The A>B stance shifts attention away from minimal pairs in simplex forms to postlexical tonal patterns that exist and persist in all dialects. I argue that tonal patterns and changes that are apparent today admit a hypothesis for tonogenesis in North Germanic that comes out favourably by Ockham's razor.

Selected references

- Bye, Patrik. forthc. The Evolution of Pitch Accent in North Germanic: Pragmatics, Phonology, Geography, and History. Pavel Iosad & Björn Köhnlein (eds.).
- d'Alquen, K. and Brown, R. 1992. The origin of Scandinavian accents I and II. I. Rauch, G.F. Carr, & R.L. Kyes (eds.), *On Germanic Linguistics: Issues and Methods*. Berlin, New York: Walter de Gruyter, 61–80.
- Elstad, Kåre. 1980. Some remarks on Scandinavian tonogenesis. *Nordlyd*, 3, 62–77.
- Goldshstein, Yonatan. 2020. Stødets naturlige historie. Y. Goldshstein, I. Schoonderbeek Hansen & T. Thode Hougaard (eds.), *18. Møde om Udforskningen af Dansk Sprog*, Århus, 151–172.
- Gårding, Eva. 1977. *The Scandinavian Word Accents*. Lund: Gleerup.
- Iosad, Pavel. 2016. Tonal Stability and Tonogenesis in (North) Germanic. I. Giles, L. Chapot, C. Coijmans, R. Foster & B. Tesio (eds.), *Beyond Borealism: New Perspectives on the North*. Norvik Press, 78–95.
- Kingston, John. 2011. Tonogenesis. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–30.
- Kock, Axel. 1901. *Die alt- und neuschwedische Akzentuierung*. Strassburg: Trübner.
- Köhnlein, Björn. 2016. Contrastive foot structure in Franconian tone-accent dialects. *Phonology* 33, 87–123.
- Liberman, Anatoly. 1982. *The Scandinavian Languages. Germanic Accentology*. Vol. 1. Minneapolis: University of Minnesota Press.
- Lorentz, Ove. 2002. Delayed peak and tonal crowding in Scandinavian tonogenesis. Ms., University of Tromsø.
- Morén-Duolljá, Bruce. 2013. The Prosody of Swedish Underived Nouns: No Lexical Tones Required. *Nordlyd* 40.1, 196–248.
- Myrberg, Sara & Tomas Riad. 2015. The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics* 38, 115–147.
- Oftedal, Magne. 1952. On the Origin of the Scandinavian Tone Distinction. *Norsk tidsskrift for sprogvidenskap* 16, 201–225.
- Perridon, Harry. 2006. On the origin of the Scandinavian word accents. *Zeitschrift für Dialektologie und Linguistik*, Beiheft 131, 91–106.
- Riad, Tomas. 1998. The Origin of Scandinavian Tone Accents. *Diachronica* XV, 63–98.
- Stroh-Wollin, Ulla. 2016. The emergence of definiteness marking in Scandinavian – new answers to old questions. *Arkiv för nordisk filologi* 131, 129–169.
- Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46, 60–79.
- Wetterlin, Allison & Aditi Lahiri. 2015. The diachronic development of stød & tonal accent in North Germanic. *Historical Linguistics 2013: Selected Papers from the 21st International Conference on Historical Linguistics, Oslo, 5–9 August 2013*, 53–67. Amsterdam: John Benjamins.

A panchronic corpus of Old East Slavic and Russian : bringing together Slavic historical and modern corpus resources

A panchronic corpus is a resource representing texts of multiple different historical periods of a given language or a branch of a language group. Typically, most large-scale diachronic corpora capture a given language only within a single historical period (“Old”, “Middle” or “Modern” lect), which is the case, for example, with the COHA language of American English and the GRAC corpus of Ukrainian (both, roughly, encompassing the period of 1820s-2020s). Families of historical corpora may be further divided by centuries, which is the case with historical corpora of Polish (see eg. <https://spxvi.edu.pl/> for the 16th century, <https://sxvii.pl/> for the 17th century, <https://korba.edu.pl/> for 1600-1772).

On the other hand, panchronic corpora are essentially large-scale diachronic corpora encompassing the bulk of the known history of the lect in question. A good example of a panchronic corpus is the *Frantext* database (<https://www.frantext.fr/>) that includes the texts from the whole written history of French starting from the 9th to the 21st century. A useful tool, it allows for building queries for Old French, Middle French and Modern French alike, but its lemmatization and annotation heavily depends on the modern orthographic and grammatical standard and is far from being accurate even with high-frequency tokens. Another panchronic resource is *Corpus Corporum* of the Zurich University (Roelli 2014), representing different stages of Latin and built as merger of different pre-existing Latin databases.

Panchronic corpora can be used for statistical study of linguistic *phénomènes de longue durée* on different levels, including orthography, morphosyntax, grammaticalized constructions, and semantics. It is of particular use in studying the so-called submerged phenomena (see e. g. for Latin: Adams, Vincent eds. 2016) that are not reflected in written sources during a large timespan but are shared by earliest and latest attestations of the language.

The paper presents the experience in bringing together the existing corpus resources within the Russian National corpus for Old East Slavic (a common ancestor of Russian, Ukrainian and Belarusian), Middle Russian, and Modern Russian, as well as a separate corpus of Old East Slavic birchbark letters. This unified resource is now searchable as the Panchronic corpus within the Russian national corpus. The source corpora had been annotated using different morphological tagsets and lemma standards stemming to different historical dictionaries. Main issues in bringing together these resources are related to mapping correspondences between the Old East Slavic phonetic rendering of lemmas prior to the loss of the short vowels known as yers (ѣ and ѥ), and, further, between Middle and Modern Russian, using rule-based and neural network algorithms with manual post-correction. General phenomena of historical linguistics such as split and merger of different lemmas due to phonetic changes and semantic divergence are discussed within this context. The issue of unified annotation of changing and emerging grammar is also to be addressed, particularly within the context of grammaticalization of East Slavic aspect and animacy.

The panchronic corpus within the RNC is also annotated by semantic classes, using Modern Russian cognates; parallels between this solution and the approach in the Historical thesaurus of English (Kay et al. eds. 2009) are discussed in the talk. As many words changed their semantics drastically this approach has inherent setbacks and should be used with caution but they can be compensated by gains in research availability for the majority of lexicon. This is illustrated in the talk by an example of searching within the panchronic corpus of “lists of sins” (a literary tradition known both in literature and vernacular birchbark writing) using a simple semantic query of three abstract nouns with negative connotation in a row, that yields relevant results in Old East Slavic, Middle Russian, and Early Modern Russian tiers alike.

The technology can be further applied to both Ukrainian and Belarusian (building of a comprehensively annotated Old Ukrainian / Old Belarusian / Ruthenian corpus being a prerequisite) as well, and also to other Slavic languages.

References

- J. N. Adams, N. Vincent (eds.) 2016. *Early and Late Latin Continuity or Change?* Cambridge, Cambridge University press
- C. Kay et al. (eds.) 2009. *Historical Thesaurus of English*. Glasgow : Glasgow University.
- Ph. Roelli, 2014. The Corpus Corporum, a new open Latin text repository and tool. *Archivum Latinitatis Medii Aevi* 72, 2014

The Sardinian substrate lexicon and its Mediterranean comparanda

The Sardinian language has a large number of lexical items that cannot have been inherited from Latin, but likely go back to one or more languages spoken on Sardinia before the Roman conquest (Wagner 1932; 1997, 254–80). The rapid developments in ancient DNA studies, which have shown that there is a significant degree of genetic continuity of the Sardinian population from the Neolithic up to the present day (Chiang et al. 2018; Calò et al. 2021), have made conditions more favorable than ever to uncovering and contextualizing the island’s linguistic prehistory. Proposals for the affiliation of Sardinia’s hypothetical substrate language(s) have included Basque, Etruscan, Berber, and less well-defined concepts such as “Eurafrican” or “Hispano-Caucasian” (Hubschmid 1953; Blasco-Ferrer 1988; Pittau 1995; Argiolas 2020).

Here, I zoom in on one set of Sardinian words of presumed pre-Roman origin whose comparanda show a fascinating distribution, being found in the languages of the Iberian peninsula and Southern France, as well as in Greek and languages of the Balkans. Examples of this Mediterranean-wide distribution include: Sardinian *golóstju*, Basque *gorosti*, Lengadocian *agaloûssès*, Greek κήλαστρος ‘holly’; Sardinian *kèya*, Catalan *sitja*, Old Provençal *setgia*, Greek καίατα, καίαδᾶς ‘(storage) pit, silo’; Sardinian *kòsti*, Basque *gazitgar*, Old Provençal *agast*, Greek ἄκαστος ‘maple’ (Hubschmid 1953, 29, 38–39, 80–82).

Although individual Sardinian substrate words with a Mediterranean distribution have been discussed at various occasions in the past, they have not been treated as a defined group. By a detailed discussion of six words with a similar distribution, I aim to shed light on how they relate to each other and on their implications for the pre-Roman linguistic situation of Sardinia as well as that of the Mediterranean region in general. I will do this by addressing the following questions:

1. Do the comparisons withstand formal and semantic scrutiny?
2. If so, how can the observed distribution be accounted for in term of language distribution and/or contact?
3. What (pre-)historical events could have led to the attested situation, based on current insights from archeological and genetic studies?
4. What implications does the Sardinian substrate lexicon have for our understanding of the pre-Indo-European linguistic situation in the Mediterranean region in general, and in Sardinia in particular?

My results will not only shed more light on the linguistic prehistory of the Mediterranean as a whole, but also constitute an interesting case study on the methodology of substrate research.

- Argiolas, Valeria. 2020. “Un État Des Études Sur l’action Du Substrat/Adstrat ‘Libyque’ En Sarde: Problèmes et Méthodes.” *Asinag* 15: 31–61.
- Blasco-Ferrer, Eduardo. 1988. “Il Sostrato Paleosardo: Fine d’un Rebus.” *Zeitschrift Für Romanische Philologie* 127 (1): 63–110.
- Calò, Carla Maria, Giuseppe Vona, Renato Robledo, and Paolo Francalacci. 2021. “From Old Markers to next Generation: Reconstructing the History of the Peopling of Sardinia.” *Annals of Human Biology* 48 (3): 203–12.
- Chiang, Charleston W.K., Joseph H. Marcus, Carlo Sidore, Arjun Biddanda, Hussein Al-Asadi, Magdalena Zoledziewska, Maristella Pitzalis, et al. 2018. “Genomic History of the Sardinian Population.” *Nature Genetics* 50: 1426–34.
- Hubschmid, Johannes. 1953. *Sardische Studien*. Bern: A. Francke.
- Pittau, Massimo. 1995. *Origine e Parentela Dei Sardi e Degli Etruschi: Saggio Storico-Linguistico*. Sassari: Carlo Delfino.
- Wagner, Max Leopold. 1932. “Über Die Vorrömische Bestandteile Des Sardischen.” *Archivum Romanicum: Nuova Rivista Di Filologia Romanza* XV: 207–47.
- . 1997. *La Lingua Sarda. Storia, Spirito e Forma [1950]*. Edited by Giulio Paulis. Nuoro: Ilisso.

CONTRIBUTIONS BY PARTICIPANT

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------------|------------------------|-----|-------|---|
| Al-Laith, Ali | Al-Laith et al. | D4 | W03.1 | A Diachronic Analysis of Using Sentiment Words in Scandinavian Literary Texts from 1870–1900 |
| Alfieri, Luca | Alfieri & Pozza | D1 | 86 | Adjectival typology in four ancient Indo-European languages |
| Alfieri, Luca | Alfieri | D4 | W06.2 | On adjectivalizers in Rig-Vedic Sanskrit |
| Amaral, Patrícia | Amaral et al. | D4 | W03.6 | Model evaluation for diachronic semantics: A view from Portuguese and Spanish |
| Andersen, Henning | Andersen | D2 | W02.3 | Macro-changes at the dawn of history: The Slavic Expansion |
| Ariztimuño, Borja | Ariztimuño & Salaberri | D1 | 152 | A new perspective on the evolution of mood and negation markers in Proto-Basque |
| Arnold, Laura | Arnold | D5 | W11.4 | Tone splits from vowel height in the Austronesian language of Raja Ampat |
| Assenzi, Lucia | Assenzi | D5 | 145 | Hearsay in Historical German Newspapers (1740–1840) |
| Auderset, Sandra | Auderset | D2 | 244 | Is tone change more rapid and irregular than segmental change? – A Mixtec case study |
| Auderset, Sandra | Auderset et al. | D5 | W11 | The diachrony of tone: connecting the field |
| Authier, Gilles | Shamseddinov & Authier | D1 | 31 | Contact-driven grammaticalization and drift of new terminal tenses from go-periphrasis in Azeri and Kryz (East Caucasian) |
| Banerjee, Mithun | Rahman & Banerjee | D2 | 169 | The diachronic study of Bangla case marking system |
| Barðdal, Jóhanna | Cluyse et al. | D1 | 208 | Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis |
| Barðdal, Jóhanna | Elens et al. | D1 | 213 | The Alternating Behavior of ‘Like’ in Old Norse-Icelandic: Facts or Fiction |
| Bartolotta, Annamaria | Bartolotta | D1 | W05.2 | The right-left conceptual mapping in a comparative and diachronic perspective |
| Baudel, Étienne | Baudel et al. | D4 | W10 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Baudel, Étienne | Baudel et al. | D4 | W10.6 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Bauer, Brigitte L. M. | Bauer | D5 | 191 | Complexity in counting systems: early systems vs. modern numerical ones |
| Belelli, Sara | Belelli | D5 | W08.6 | A historical-comparative glimpse on Laki dialects |
| Beniamine, Sacha | Round et al. | D2 | 196 | The natural stability of ‘unnatural’ morphology |
| Benvenuto, Maria Carmela | Benvenuto & Bichlmeier | D4 | 278 | The expression of predicative possession in Avestan |
| Berge, Anne | Berge | D1 | W01.2 | Prehistoric climate changes and their effects on the development of the Eskaleut languages |
| Biagetti, Erica | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |
| Biagetti, Erica | Brigada Villa et al. | D2 | W14.2 | Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages |
| Bichlmeier, Harald | Benvenuto & Bichlmeier | D4 | 278 | The expression of predicative possession in Avestan |
| Billing, Oscar | Billing & Elgh | D5 | 178 | Computational Anatolian phylogeny using maximum parsimony |

| Participant | Authors' abbrev. | Day | No. | Title |
|-----------------------------|--------------------------------------|-----|-------|--|
| Billing, Oscar | Rönchen et al. | D4 | W03.7 | Using simulated data to evaluate models of Indo-European vocabulary evolution |
| Bjørn, Rasmus G. | Bjørn & Kilani | D1 | W07.0 | Interactions at the dawn of history: An introduction to the workshop |
| Bjørn, Rasmus G. | Bjørn & Kilani | D1 | W07 | Interactions at the dawn of history: Methods and results in prehistoric contact linguistics |
| Björnsdóttir, Sigríður | Björnsdóttir et al. | D1 | 203 | The rise of raising in Early Modern English |
| Bloom, Barthe | Bloom | D3 | 80 | Early New High German preposed adverbial clauses: integration and discourse functions |
| Blum, Frederic | Blum & List | D2 | W14.6 | A computational evaluation of regularly recurring sound correspondences |
| Bogdanowska-Jakubowska, Ewa | Bogdanowska-Jakubowska & Bogdanowska | D3 | 75 | Changes in the Polish address practices after the Second World War |
| Bogdanowska, Nika | Bogdanowska-Jakubowska & Bogdanowska | D3 | 75 | Changes in the Polish address practices after the Second World War |
| Bonmann, Svenja | Bonmann et al. | D2 | 32 | Towards a New Reconstruction of the Proto-Yeniseian Sound System |
| Börjars, Kersti | Börjars & Vincent | D1 | 52 | Auxiliary, light or lexical: the history of GO verbs |
| Bossuyt, Tom | Bossuyt & Daveloose | D2 | 104 | Divergence and contact in Cappadocian concessive conditionals |
| Bostoen, Koen | Gunnink et al. | D5 | 232 | An evolutionary loner in Southern African Bantu: The classification of Yeyi |
| Bostoen, Koen | Bostoen et al. | D1 | W07.2 | Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction |
| Bostoen, Koen | Pacchiarotti et al. | D5 | 69 | Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu |
| Boye, Kasper | Boye | D3 | 93 | Grammaticalization as conventionalization of discursively secondary status: Isolating what is unique to grammaticalization, and deconstructing the lexical-grammatical continuum |
| Boye, Kasper | Vincent et al. | D4 | PL6 | Linguistic models (with a focus on morphosyntactic change) |
| Boye, Kasper | Westergaard & Boye | D3 | 90 | On semantic change in grammaticalization: Why it is never metaphoric |
| Boyeldieu, Pascal | Boyeldieu | D2 | 255 | Tone split and tone replacement: toward the three-tone system of the 'Western' SBB Languages (Central Sudanic, Central Africa) |
| Božović, Đorđe | Božović | D5 | W11.1 | Tone, stress and length interactions in Central Neo-Štokavian |
| Bradley, David | Bradley | D1 | W01.5 | Climate change and the dispersal of Proto-Tibeto-Burman |
| Brigada Villa, Luca | Brigada Villa et al. | D2 | W14.2 | Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages |
| Bronikowska, Renata | Bronikowska | D5 | 40 | Middle Polish adverb-like predicates ending in -a compared to other adverbial and adjectival predicates – corpus-based approach |
| Brosig, Benjamin | Brosig & Dolgor | D2 | W12.2 | From spatial noun to medial demonstrative: the case of Khalkha Mongolian |
| Brown, Braden | Brown & Grollemund | D5 | 230 | Towards a new classification of Western Bantu languages using non-lexical data |
| Bru, Mathilde | Bru | D2 | 72 | 'So wrong that not even Menander uses it!': the Atticist lexicographers on the Ancient Greek dialects |
| Brunner, Thomas | Brunner | D4 | 112 | The ordering of matrix clauses and subordinate causal clauses in the Old Bailey Corpus 1720–1913 |

| Participant | Authors' abbrev. | Day | No. | Title |
|----------------------|---------------------|-----|-------|--|
| Bugaeva, Anna | Satō & Bugaeva | D4 | W10.1 | On stative/active intransitive split within tripartite alignment: A case of Kuril Ainu |
| Calabrese, Andrea | Calabrese | D4 | W06.1 | Inflectional vocalic pieces in Latin verbal morphology: a synchronic and diachronic analysis |
| Camilleri, Maris | Camilleri | D4 | W09.5 | Parallels in the development from locative and existential predications to possessive structures in Arabic and Hebrew |
| Capano, Marta | Capano | D5 | 237 | It Ain't Over till It's Over. Bilingualism and language decay in Sicilian inscriptions |
| Caso, Anabelle | Caso & Hale | D2 | 273 | Secondary predication in metrical texts: syntax-prosody mapping in ancient Indo-European languages |
| Cassarà, Alessia | Cassarà et al. | D4 | W13.3 | Marked vs. unmarked unaccusativity with alternating verbs: Linking diachronic and experimental data |
| Cassarà, Alessia | Cassarà et al. | D4 | W13 | New methods for old languages: the comparability of data |
| Cassarà, Alessia | Cassarà et al. | D4 | W13.0 | New methods for old languages: the comparability of data |
| Cathcart, Chundra | Herce & Cathcart | D5 | 199 | Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure |
| Cattafi, Eleonora | Cattafi | D5 | 264 | Continuative relative clauses in Greek documentary papyri |
| Céline, Mounole | Manterola et al. | D1 | 58 | The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments |
| Cennamo, Michela | Cennamo | D4 | 173 | Existential HAVE in Late Latin: insights on its diachrony in the passage to Romance |
| Chankova, Yana | Chankova | D2 | 97 | As Syntax Interfaces with Information Structure: Old Icelandic Non-Canonical Scrambled Orders |
| Cluyse, Brian | Cluyse et al. | D1 | 208 | Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis |
| Coenen, Pascal | Coenen | D1 | 127 | The totalizing function of the Vedic particle <i>cid</i> |
| Concu, Valentina | Concu | D2 | 122 | The use of "thank" and "to thank" in Old Saxon and Old High German |
| Conradie, Jac | Conradie | D1 | 59 | The Afrikaans auxiliary <i>het</i> 'have' from clitic to desinence |
| Cornillie, Bert | Inglese et al. | D4 | 100 | The anticausative alternation in Italian and Spanish: a historical corpus-based perspective |
| Creissels, Denis | Creissels | D4 | W09.4 | 'Be/have' verbs in historical perspective |
| Currie, Oliver | Currie | D3 | 63 | The emergence of a Welsh biblical literary standard and the evidence of early modern manuscript sermons |
| Cuyckens, Huybert | Nijs et al. | D2 | W02.4 | An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German |
| Däbritz, Chris Lasse | Däbritz | D4 | W09 | "Your birch-bark bag has something" – Grammaticalization and diachrony of locative, existential and possessive predication |
| Däbritz, Chris Lasse | Däbritz | D4 | W09.1 | "Your birch-bark bag has something" – Grammaticalization and diachrony of locative, existential and possessive predication |
| Darling, Mark | Darling et al. | D2 | 180 | The Diachrony of Person-Number Marking of Subjects in Celtic |
| Das, Patrick | Das | D5 | 266 | Areality through Migration: Investigating the Structure of Numeral Classifiers in the Eastern Himalayan Region Reveals Historic Contact Events |
| Daveloose, Eline | Bossuyt & Daveloose | D2 | 104 | Divergence and contact in Cappadocian concessive conditionals |
| Daveloose, Eline | Daveloose | D3 | 137 | From <i>de</i> to <i>ke</i> : functional transfer of a topic shift marker from Turkish to Cappadocian Greek |

| Participant | Authors' abbrev. | Day | No. | Title |
|-----------------------------|--------------------------|-----|-------|---|
| De Cesare, Ilaria | Zehentner & De Cesare | D2 | W04 | Ambiguity (avoidance) as a factor in language change |
| de Rossi, Nicolò | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |
| De Smet, Hendrik | Kayenbergh & De Smet | D5 | 89 | Just a bystander? Semantic change in the English simple tenses |
| De Smet, Hendrik | Felser | D2 | W04.3 | Structural ambiguity in language comprehension and production |
| de Vaan, Michiel | Neri & de Vaan | D2 | W12.6 | Origin and development of the Albanian demonstratives |
| de Vos, Machteld | de Vos | D4 | 121 | Spread the German new(s): third-person reflexive <i>zich</i> in 17th-century Dutch newspapers |
| Dedvukaj, Lindon | Dedvukaj | D1 | 68 | Reanalyzing the Historical Constructions of Albanian Prepositions |
| Degaetano-Ortlieb, Stefania | Jenset et al. | D4 | W03.2 | Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i> |
| Degaetano-Ortlieb, Stefania | Degaetano-Ortlieb et al. | D4 | W03 | Computational models of diachronic language change |
| Dellert, Johannes | Dellert & Blaschke | D2 | W14.5 | Configurable Language-Specific Tokenization for CLDF Databases |
| Demolin, Didier | Ricquier & Demolin | D3 | 102 | The Chronicle of Lingbe, an Extinct Bantu Language of East Congo |
| Deng, Bingcong | Deng | D1 | W01.6 | Climate change reflected in early Sino-Tibetan borrowings for crops and animals |
| Deo, Ashwini | Vincent et al. | D4 | PL6 | Linguistic models (with a focus on morphosyntactic change) |
| Dereza, Oksana | Dereza et al. | D4 | W03.8 | Evaluating historical word embeddings: strategies, challenges and pitfalls |
| di Bartolo, Giuseppina | di Bartolo | D3 | 154 | Where and How? Request verb constructions in Ancient Greek |
| Dinu, Liviu P. | Dinu et al. | D3 | 279 | Computational approaches for Romance related words discrimination |
| Dockum, Rikker | Dockum & Lu | D2 | 245 | Beyond the paradigm: Change and expansion in Thai pronominal reference |
| Dockum, Rikker | Dockum & Wang | D4 | 274 | Quality vs. quantity: Contrast maintenance and tradeoff in Southwestern Tai vowels |
| Dockum, Rikker | Auderset et al. | D5 | W11 | The diachrony of tone: connecting the field |
| Dolgor, Guntsetseg | Brosig & Dolgor | D2 | W12.2 | From spatial noun to medial demonstrative: the case of Khalkha Mongolian |
| Dömötör, Adrienne | Dömötör | D2 | 175 | From direct quotation to a chain of extended quotations: the history of Hungarian <i>úgymond</i> 'so to speak' |
| Drach, Mortimer | Gelumbeckaitė et al. | D3 | 247 | The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian" |
| Drinka, Bridget | Drinka et al. | D2 | W02 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Drinka, Bridget | Drinka et al. | D2 | W02.1 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Dücker, Lisa | Dücker | D4 | 171 | Semantic factors influencing the change in position of German adnominal genitives in the 17th to 19th centuries |
| Egedi, Barbara | Egedi | D2 | 194 | Demonstrative modifiers in Middle Hungarian: a complex picture of renewal |
| Elgh, Erik | Billing & Elgh | D5 | 178 | Computational Anatolian phylogeny using maximum parsimony |

| Participant | Authors' abbrev. | Day | No. | Title |
|-----------------------------|--------------------------------|-----|-------|--|
| Ellison, T. Mark | Reinöhl & Ellison | D5 | 190 | Metaphor, Overtness and Word Order Routinization |
| Elter, W. Juliane | Elter | D5 | 183 | Anglo-Scandinavian Contact Influence on Verbs Entering the Causative Alternation |
| Engelberg, Stefan | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |
| Enrique-Arias, Andrés | Enrique-Arias | D2 | W02.8 | Political influence as a factor in morphosyntactic variation: demonstratives <i>este</i> and <i>aqueste</i> in medieval Aragonese |
| Esher, Louise | Esher | D4 | 105 | Gascon <i>u</i> -perfects and the analogical foregrounding of inflectional class |
| Esher, Louise | Round et al. | D2 | 196 | The natural stability of 'unnatural' morphology |
| Espíndola Moschner, Silvina | Espíndola Moschner & Rosemeyer | D1 | 148 | Aspectual uses of <i>saber</i> + infinitive in South American Spanish varieties: a corpus-based study |
| Étienne, Baudel | Baudel | D4 | W10.4 | Reconsidering the classification of Hachijō: A glimpse from historical phonology |
| Eyþórsson, Þórhallur | Eyþórsson & Sigurðardóttir | D2 | 126 | Micro-level conflict in the productivity of anticausativization strategies: Evidence from the history of Icelandic |
| Farina, Andrea | Farina et al. | D2 | 114 | WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose |
| Feltgen, Quentin | Feltgen | D1 | 113 | The shape of grammaticalization: matching the bridging context scenario with patterns of frequency use |
| Figura, Lisa | Figura | D4 | W13.1 | Dative Experiencer Psych Verbs in (Old) French |
| Flaksman, Maria | Flaksman | D3 | 150 | Lost in translation. Onomatopoeic words in Old English glosses |
| Fonteyn, Lauren | Degaetano-Ortlieb et al. | D4 | W03 | Computational models of diachronic language change |
| Fonteyn, Lauren | Fonteyn et al. | D2 | 29 | From ecological to lexical diversity: measuring vocabulary richness in historical corpora |
| Forkel, Robert | Jäger et al. | D2 | W14 | Exploiting Standardized Cross-Linguistic Data in Historical Linguistics |
| Forkel, Robert | Forkel & Greenhill | D2 | W14.4 | Phlorest: A Database of Consistent and Reusable Language Phylogenies |
| Franco, Karlien | Franco | D5 | 168 | Explaining the speed of lexical change in historical Dutch |
| Fransen, Theodorus | Dereza et al. | D4 | W03.8 | Evaluating historical word embeddings: strategies, challenges and pitfalls |
| Fried, Mirjam | Vincent et al. | D4 | PL6 | Linguistic models (with a focus on morphosyntactic change) |
| Friedman, Victor | Friedman | D2 | 98 | Obscenity as a Window into Slavic Linguistic History |
| Fries, Simon | Bonmann et al. | D2 | 32 | Towards a New Reconstruction of the Proto-Yeniseian Sound System |
| Fromm, Nathalie | Fromm | D1 | 129 | The development of number strengthening in German declensional classes. A diachronic-dialectal corpus study |
| Funk, Ekaterina | Tikhonov et al. | D2 | 143 | Pronoun history and information structure in 18th century non-religious Russian texts |
| Gehrmann, Ryan | Auderset et al. | D5 | W11 | The diachrony of tone: connecting the field |
| Gelumbeckaitė, Jolanta | Gelumbeckaitė et al. | D3 | 247 | The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian" |
| Gfeller, Kim | Gfeller | D1 | 234 | Persistence and Change of Colexifications in Indo-European |
| Gholami, Saloumeh | Karim & Gholami | D5 | W08 | Filling in the diachronic gaps: the view of Old Iranian from the present |

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------------|-----------------------------|-----|-------|---|
| Gholami, Saloumeh | Karim & Gholami | D5 | W08.1 | Filling in the diachronic gaps: the view of Old Iranian from the present |
| Gholami, Saloumeh | Gholami & Naghshbandi | D5 | W08.5 | Polyptoton for the purpose of emphasizing within Iranian languages |
| Gibson, Hannah | Gibson et al. | D4 | 227 | Morphosyntactic variation in Swahili: Tracing descriptions past and present |
| Ginevra, Riccardo | Ginevra | D1 | W05.3 | Indo-European Poetics meets Cognitive Linguistics: an integrated approach to the comparative reconstruction of metaphoric and metonymic expressions |
| Gisborne, Nikolas | Gisborne & Truswell | D4 | 222 | Contact and the origins of headed <i>wh</i> -relatives in Hungarian |
| Giuliani, Martina | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |
| González Saavedra, Berta | Orqueda & González Saavedra | D2 | W12 | From and Towards Demonstratives: Grammaticalization Processes and Beyond |
| Gopal, Deepthi | Gopal et al. | D1 | 240 | Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture |
| Gosemann, Laura | Gosemann | D4 | 118 | Syntactic change and DLM in German: a corpus study |
| Gotthard, Lisa | Gotthard | D1 | 214 | The rise of <i>do</i> -support during Scots anglicisation: Insights from the <i>Parsed Corpus of Scottish Correspondence</i> |
| Gotthard, Lisa | Björnsdóttir et al. | D1 | 203 | The rise of raising in Early Modern English |
| Gray, Russell | Shcherbakova et al. | D2 | 96 | Diachronic pathways of definite articles distribution |
| Gray, Russell | Verkerk et al. | D3 | 242 | Exploiting phylogenetic modeling to uncover directionality in the emergence of universals |
| Grestenberger, Laura | Grestenberger et al. | D4 | W06 | Categorizers in diachrony |
| Grestenberger, Laura | Grestenberger et al. | D4 | W06.0 | Categorizers in diachrony |
| Grimm, Nadine | Grimm | D5 | W11.5 | A diachronic study of grammatical tone in northwestern Bantu |
| Grollemund, Rebecca | Brown & Grollemund | D5 | 230 | Towards a new classification of Western Bantu languages using non-lexical data |
| Grossman, Eitan | Sæbø & Grossman | D5 | W11.6 | A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis |
| Gugán, Katalin | Gugán | D1 | 228 | Outliers in variation and change: atypical users of the variants of negation in Old and Middle Hungarian |
| Gunnink, Hilde | Gunnink et al. | D5 | 232 | An evolutionary loner in Southern African Bantu: The classification of Yeyi |
| Günther, Laura | Bonmann et al. | D2 | 32 | Towards a New Reconstruction of the Proto-Yeniseian Sound System |
| Guzmán Naranjo, Matías | Mertner & Guzmán Naranjo | D2 | W14.7 | Exploring the Geographical Distribution of Missing Data Using Approximate Gaussian Processes |
| Guzmán Naranjo, Matías | Jonjić et al. | D2 | 144 | Isoglosses and distributions of features – Analyses of the <i>Dialectological Atlas of the Russian Language</i> |
| Gvozdanović, Jadranka | Gvozdanović | D2 | W02.6 | Ideology, language choice and language change |
| Hakimov, Nikolay | Hakimov | D4 | 251 | Fall of the jers: A multi-factorial analysis of the sound change progression in the Old Novgorodian birchbark texts |
| Hale, Mark | Caso & Hale | D2 | 273 | Secondary predication in metrical texts: syntax-prosody mapping in ancient Indo-European languages |

| Participant | Authors' abbrev. | Day | No. | Title |
|---------------------------|-------------------------|-----|--------|---|
| Halfmann, Jakob | Halfmann & Korobzow | D3 | 239 | The Evolution of Spatial Orientation Systems in Mayan and Nuristani |
| Hamans, Camiel | Hamans | D1 | 67 | A revolution in the history of affix-formation |
| Hansen, Magnus Pharao | Hansen & Davletshin | D1 | W07.1 | Tracing borrowings in and out of proto-Nahuatl |
| Hartmann, Stefan | Pleyer et al. | D1 | 128 | The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective |
| Harvey, Mark | Mailhammer & Harvey | D1 | 184 | The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database |
| Haspelmath, Martin | Haspelmath | D2 | W04.7 | Ambiguity avoidance vs. expectation sensitivity as functional factors in language change and language structures: Beyond argument marking |
| Hasselbach-Andee, Rebecca | Hasselbach-Andee | D4 | W06.3 | One or All: The Development of Singulatives to Collectives in Semitic |
| Heggarty, Paul | Heggarty | D1 | W01.8 | Languages, ecology and climate change: Worldwide perspectives and the test-case of the Andes |
| Hengeveld, Kees | Hengeveld | D4 | W09.3 | The development of locative, existential and possessive predication from a functional perspective |
| Herce, Borja | Herce & Cathcart | D5 | 199 | Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure |
| Hernáiz, Rodrigo | Hernáiz | D5 | 215 | Exploring language variation and change in the distant past |
| Hill, Eugen | Bonmann et al. | D2 | 32 | Towards a New Reconstruction of the Proto-Yeniseian Sound System |
| Hirvonen, Johannes | Hirvonen | D5 | 200 | Contact-induced change of Negative Indefinites – the case of Meadow Mari |
| Hofmann, Klaus | Ritt & Hofmann | D2 | 248 | 'Chained to the rhythm': Using agent-based simulation to model the evolution of stress pattern diversity in English |
| Hofmann, Klaus | Hofmann | D2 | 157 | Evolving rhythms: A quantitative assessment of rhythmic alternation in the history of English |
| Holopainen, Sampsa | Holopainen | D4 | 238 | The emergence of word-initial voiced stops in Proto-Hungarian |
| Honeybone, Patrick | Honeybone | D2 | 188 | Can fortis stops spirantise without aspiration? |
| Hu, Hai | Amaral et al. | D4 | W03.6 | Model evaluation for diachronic semantics: A view from Portuguese and Spanish |
| Hualde, José Ignacio | Hualde | D2 | 57 | The diachrony of Basque accentuation: comparative method and internal reconstruction |
| Hualde, José Ignacio | Manterola et al. | D1 | 58 | The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments |
| Huang, Yang | Huang | D2 | 56 | The Expression of Negation in Sabde Minyag |
| Huback, Ana Paula | Huback & Fontes Martins | D2 | 94 | R Deletion in Brazilian Portuguese: Diachronic and Synchronic Evidence for Lexical Diffusion |
| Hudson, Mark | Hudson | D1 | W01.10 | Risk, resilience and the ecology of farming/language dispersals |
| Idiatov, Dmitry | Idiatov | D2 | 65 | Vowel reduction to /i/ in functional morphemes in Northern Sub-Saharan Africa |
| Iezzi, Luca | Iezzi | D5 | 108 | The role of French in the Johnsons' correspondence |
| Igartua, Iván | Igartua | D4 | 254 | Exploring the sources of animacy distinctions |
| Inglese, Guglielmo | Inglese et al. | D4 | 100 | The anticausative alternation in Italian and Spanish: a historical corpus-based perspective |

| Participant | Authors' abbrev. | Day | No. | Title |
|---------------------------|----------------------|-----|-------|--|
| Ionov, Max | Gelumbeckaitė et al. | D3 | 247 | The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian" |
| Ishiyama, Osamu | Ishiyama | D2 | W12.3 | On the Development of Demonstratives into Personal Pronouns |
| Jäger, Gerhard | Jäger et al. | D2 | W14 | Exploiting Standardized Cross-Linguistic Data in Historical Linguistics |
| Janda, Richard | Janda & Joseph | D1 | 135 | West Germanic 2.sg. <i>-st</i> Revisited: The Role of Superscence |
| Jarosz, Aleksandra | Baudel et al. | D4 | W10 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Jarosz, Aleksandra | Baudel et al. | D4 | W10.6 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Jenkins, Chris | Maurer et al. | D4 | W03.3 | Quantifying Changes in English Noun Compound Productivity and Meaning |
| Jensen, Eva Skafte | Jensen & Schack | D4 | 84 | Adverbs ending in <i>-(l)ig '-ly'</i> and <i>-(l)igt '-ly'</i> in Danish |
| Joseph, Brian | Joseph | D1 | W01.9 | (Im)mobility, climate, and language: Towards a geoanthropology of the Balkans |
| Joseph, Brian | Janda & Joseph | D1 | 135 | West Germanic 2.sg. <i>-st</i> Revisited: The Role of Superscence |
| Juge, Matthew | Juge | D1 | 270 | The dominant-recessive hypothesis does not account for overlapping suppletion |
| Kaltenbach, Lena | Cassarà et al. | D4 | W13 | New methods for old languages: the comparability of data |
| Kaltenbach, Lena | Cassarà et al. | D4 | W13.0 | New methods for old languages: the comparability of data |
| Kamil, Iris | Grestenberger et al. | D4 | W06 | Categorizers in diachrony |
| Kamil, Iris | Grestenberger et al. | D4 | W06.0 | Categorizers in diachrony |
| Karim, Shuan Osman | Karim & Gholami | D5 | W08 | Filling in the diachronic gaps: the view of Old Iranian from the present |
| Karim, Shuan Osman | Karim & Gholami | D5 | W08.1 | Filling in the diachronic gaps: the view of Old Iranian from the present |
| Kauhanen, Henri | Gopal et al. | D1 | 240 | Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture |
| Kayenbergh, Juliette | Kayenbergh & De Smet | D5 | 89 | Just a bystander? Semantic change in the English simple tenses |
| Kilani, Marwan | Bjørn & Kilani | D1 | W07.0 | Interactions at the dawn of history: An introduction to the workshop |
| Kilani, Marwan | Bjørn & Kilani | D1 | W07 | Interactions at the dawn of history: Methods and results in prehistoric contact linguistics |
| Kim, Ronald I. | Kim | D5 | W08.3 | Steppe Iranian in the <i>longue durée</i> : contact, relative chronology, and internal reconstruction |
| Kinuhata, Tomohide | Kinuhata | D4 | W10.3 | Reconstructing the Proto-Japonic demonstrative system |
| Kiparsky, Paul | Kiparsky | D1 | PL1 | The word-order cycle |
| Kirby, James | Kirby & Pittayaporn | D5 | W11.3 | Tone and voicing in Cao Bằng Tai: implications for tonal evolution and change |
| Kisiel, Anna | Kisiel & Sobotka | D2 | 51 | The functional interpretation of semantic and syntactic shifts in the domain of North Slavic "conversive" preposition-pronominal constructions |
| Klævik-Pettersen, Espen | Klævik-Pettersen | D4 | 261 | VSO orders in the Egeriae and Antonini Placentini itineraria; new evidence for the evolution towards Old Romance inversion systems |
| Klosa-Kückelhaus, Annette | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |

| Participant | Authors' abbrev. | Day | No. | Title |
|------------------------|--------------------------|-----|-------|---|
| Knapen, Martijn | Knapen | D1 | W01.3 | Seals and sea ice: the (possible) climatic background of Amuric influence on Ainu |
| Kölligan, Daniel | Kölligan | D1 | W05.7 | Conceptual metaphors and etymology |
| Kölligan, Daniel | Kölligan & van Beek | D1 | W05 | Conceptual metaphors in a comparative and diachronic perspective |
| Kölligan, Daniel | Kümmel et al. | D1 | 124 | The development of future-referring constructions (in Indo-European languages) |
| Korobzow, Natalie | Halfmann & Korobzow | D3 | 239 | The Evolution of Spatial Orientation Systems in Mayan and Nuristani |
| Korobzow, Natalie | Bonmann et al. | D2 | 32 | Towards a New Reconstruction of the Proto-Yeniseian Sound System |
| Kozhanov, Kirill | Kozhanov | D2 | 197 | Diachronic stability of case functions: oblique in Romani dialects |
| Krasnoukhova, Olga | Krasnoukhova et al. | D4 | W09.2 | Negated but similar – Negation in the domains of locative, existential, and possessive predication: The case of Indo-European |
| Kreidl, Julian | Kreidl | D5 | W08.2 | Bactrian influence on local languages of Eastern Afghanistan |
| Krielke, Marie-Pauline | Degaetano-Ortlieb et al. | D4 | W03 | Computational models of diachronic language change |
| Kübler, Sandra | Amaral et al. | D4 | W03.6 | Model evaluation for diachronic semantics: A view from Portuguese and Spanish |
| Kümmel, Martin Joachim | Kümmel et al. | D1 | 124 | The development of future-referring constructions (in Indo-European languages) |
| Lahiri, Aditi | Lahiri | D2 | PL2 | Phonological grammars: Pertinacious constraints on change |
| Landwehr, Isabell | Jenset et al. | D4 | W03.2 | Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i> |
| Larrivèe, Pierre | Poletto et al. | D4 | 36 | Learning how to count – a treebank analysis of V2 word order in two Medieval Romance languages through time |
| Leddy-Cecere, Thomas | Leddy-Cecere | D1 | 116 | The PRESENTATIVE > DEMONSTRATIVE Grammaticalization Pathway in Arabic |
| Lindgren, Freja | Lindgren & Tresoldi | D3 | 198 | The Charition Mime: Decoding the “Indian Language” through Typology and Entropy |
| Lionnet, Florian | Lionnet | D5 | W11.2 | Accent and tone: the double origin of the Paicî tone system |
| Lionnet, Florian | Lionnet | D2 | 225 | Areal alignment and the loss of ATR harmony in Riverine Bua languages (Chad) |
| List, Johann-Mattis | Blum & List | D2 | W14.6 | A computational evaluation of regularly recurring sound correspondences |
| List, Johann-Mattis | Jäger et al. | D2 | W14 | Exploiting Standardized Cross-Linguistic Data in Historical Linguistics |
| Litvinova, Lora | Litvinova | D2 | 258 | Reconstructing the Kugama tone system |
| Luján, Eugenio | Luján & Ngomo Fernández | D2 | W12.7 | From demonstratives to articles in the Celtic languages |
| Luraghi, Silvia | Brigada Villa et al. | D2 | W14.2 | Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages |
| Luraghi, Silvia | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |
| Maiden, Martin | Maiden | D1 | 43 | A morphological freeloader: Ibero-Romance <i>cabrer</i> |
| Mailhammer, Robert | Mailhammer & Harvey | D1 | 184 | The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database |

| Participant | Authors' abbrev. | Day | No. | Title |
|-----------------------|--------------------------|-----|-------|---|
| Majtczak, Tomasz | Majtczak | D4 | W10.5 | Old, Middle and New: Periodisation as a back-burnered topic in the diachronic research of Japanese |
| Manterola, Julen | Mounole & Manterola | D1 | 265 | From distal demonstrative to resultative marker (through definite article): evidence from Basque |
| Manterola, Julen | Manterola et al. | D1 | 58 | The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments |
| Markopoulos, Theodore | Markopoulos | D3 | 167 | Epistemic modality out of 'playfulness': Modern Greek <i>pezi</i> |
| Marr, Clayton | Marr | D4 | W03.4 | A computerized investigation of Albanian diachronic phonology |
| Marten, Lutz | Marten | D3 | PL3 | Historical linguistics and <i>ubuntu</i> translanguaging: Towards a model of multilingualism, language change and linguistic convergence in the Bantu Linguistic Area |
| Marten, Lutz | Gibson et al. | D4 | 227 | Morphosyntactic variation in Swahili: Tracing descriptions past and present |
| Maselli, Lorenzo | Bostoen et al. | D1 | W07.2 | Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction |
| Maselli, Lorenzo | Pacchiarotti et al. | D5 | 69 | Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu |
| Maurer, Maximilian | Maurer et al. | D4 | W03.3 | Quantifying Changes in English Noun Compound Productivity and Meaning |
| Mazzola, Giulia | Rosemeyer et al. | D5 | 138 | A computational approach to detect discourse traditions and register differences: a case study on historical French |
| Mazzola, Giulia | Inglese et al. | D4 | 100 | The anticausative alternation in Italian and Spanish: a historical corpus-based perspective |
| McCarley, Gemma | McCarley | D4 | 233 | Diachronic Null Subject Use across Latin American Spanish: Comparing Corpora |
| McCrae, John P. | Dereza et al. | D4 | W03.8 | Evaluating historical word embeddings: strategies, challenges and pitfalls |
| McGillivray, Barbara | Jenset et al. | D4 | W03.2 | Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i> |
| McGillivray, Barbara | Farina et al. | D2 | 114 | WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose |
| Meisterernst, Barbara | Meisterernst | D1 | 211 | The diachronic development of future markers in Chinese |
| Mendoza, Imke | Mendoza et al. | D4 | 99 | Anchoring patterns in emerging complement clauses in Slavic |
| Mertner, Miri | Mertner & Guzmán Naranjo | D2 | W14.7 | Exploring the Geographical Distribution of Missing Data Using Approximate Gaussian Processes |
| Mesthrie, Rajend | Mesthrie | D2 | W02.9 | Macro sociohistorical forces, contact, convergence and the development of modern linguistic areas: insights from South Africa |
| Meyer, Peter | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |
| Meyer, Robin | Meyer | D2 | 187 | Quasi- <i>Suffixaufnahme</i> in Classical Armenian |
| Meyer, Roland | Tikhonov et al. | D2 | 143 | Pronoun history and information structure in 18th century non-religious Russian texts |
| Miletic, Filip | Maurer et al. | D4 | W03.3 | Quantifying Changes in English Noun Compound Productivity and Meaning |

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------------|-----------------------------|-----|-------|---|
| Mirelman, Sam | Mirelman | D3 | 74 | Translation as Royal Legitimation: The Concepts of "Source" and "Target" Language in Sumerian-Akkadian Royal Inscriptions from the Old Babylonian Period (2000–1600 BC) |
| Mithun, Marianne | Mithun | D2 | W12.1 | Further Pathways Towards Demonstratives |
| Mofidi, Roohollah | Mofidi | D1 | 55 | Competition in the aspect-mood domain: The standardization of a diachronic data set of New Persian |
| Mohammadirad, Masoud | Mohammadirad | D5 | W08.4 | Remarks on the category of copula in Gorani dialects |
| Mounole, Céline | Mounole & Manterola | D1 | 265 | From distal demonstrative to resultative marker (through definite article): evidence from Basque |
| Mous, Maarten | Mous | D5 | 192 | The classification of South Cushitic |
| Munteanu, Andrei | Munteanu | D3 | 73 | Automating Comparative Reconstructions: Case Study in Austronesian and Ongan |
| Næss, Åshild | Næss | D2 | W12.5 | Demonstratives taking over discourse: the grammaticalisation of deictic clitics in Äiwoo |
| Neels, Jakob | Pleyer et al. | D1 | 128 | The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective |
| Neri, Sergio | Neri & de Vaan | D2 | W12.6 | Origin and development of the Albanian demonstratives |
| Nevalainen, Terttu | Drinka et al. | D2 | W02 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Nevalainen, Terttu | Drinka et al. | D2 | W02.1 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Ngomo Fernández, Esteban | Luján & Ngomo Fernández | D2 | W12.7 | From demonstratives to articles in the Celtic languages |
| Nichols, Johanna | Nichols | D2 | W02.2 | Reconstructing prehistoric sociolinguistics from modern grammatical evidence |
| Nieder, Jessica | Nieder & Tomaschek | D5 | 174 | Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics |
| Nijs, Julie | Nijs et al. | D2 | W02.4 | An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German |
| Ochs, Samira | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |
| Olivier, Marc | Olivier | D1 | 217 | When change fails: evidence from French |
| Ongenaë, Tim | Ongenaë | D2 | 66 | Towards a Diachronic Account of P-lability in Latin: The Semantic Extension of the Active Intransitive as an Anticausative Strategy in Latin |
| Orlandi, Georg | Baudel et al. | D4 | W10 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Orlandi, Georg | Baudel et al. | D4 | W10.6 | The (Pre)History of the Languages of Japan – Current issues and prospects |
| Orqueda, Verónica | Orqueda & González Saavedra | D2 | W12 | From and Towards Demonstratives: Grammaticalization Processes and Beyond |
| Orqueda, Verónica | Orqueda & Pooth | D2 | W12.8 | Latin <i>ecce</i> : arguments in favor of its development from a PIE demonstrative |
| Pacchiarotti, Sara | Bostoen et al. | D1 | W07.2 | Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction |
| Pacchiarotti, Sara | Pacchiarotti et al. | D5 | 69 | Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu |

| Participant | Authors' abbrev. | Day | No. | Title |
|-------------------------|--------------------------|-----|-------|--|
| Pache, Matthias | Pache | D5 | 166 | Evidence for a Chibcha-Jê connection |
| Pan, Tao | Pan | D3 | 38 | <i>ille ego</i> and Recognitional Use of Demonstratives |
| Paterson, Hugh | Paterson, H. | D4 | 282 | Proto-Malayo-Polynesian: Some Phonetic Evidence for */ |
| Paterson, Rebecca | Paterson, R. | D1 | 280 | Emergence of alternate argument alignment patterns in Northwest Kainji |
| Peck, Naomi | Reinöhl et al. | D2 | 181 | The loss of word-initial consonants in Kera'a – A challenge for phonological theory |
| Perekhval'skaya, Elena | Perekhval'skaya & Vydrin | D5 | W11.7 | Tonal density and its correlation with the types of tonal systems: Diachronic aspects |
| Persohn, Bastian | Persohn | D1 | 37 | When 'still' comes to signal a near past |
| Petré, Peter | Petré | D1 | 206 | Conservative pressure on the progressive: the passival |
| Piccione, Mariapaola | Cassarà et al. | D4 | W13 | New methods for old languages: the comparability of data |
| Piccione, Mariapaola | Cassarà et al. | D4 | W13.0 | New methods for old languages: the comparability of data |
| Pierce, Marc | Pierce | D2 | 158 | The History of /pf/ in New Braunfels German: Another Case of Rule Inversion? |
| Pinzin, Francesco | Poletto et al. | D4 | 36 | Learning how to count – a treebank analysis of V2 word order in two Medieval Romance languages through time |
| Pittayaporn, Pittayawat | Kirby & Pittayaporn | D5 | W11.3 | Tone and voicing in Cao Bằng Tai: implications for tonal evolution and change |
| Pleyer, Michael | Pleyer et al. | D1 | 128 | The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective |
| Pompei, Anna | Pompei | D4 | 272 | The case of Italian <i>segunte</i> : an European instance of current change from verb to demonstrative? |
| Pompeo, Flavia | Pompeo | D1 | W05.4 | New meanings and old constructions: the conceptualization of 'fearing' and 'protecting' in Old Persian in comparison with other Indo-Iranian languages |
| Potochnik, Thomas | Potochnik | D3 | 236 | Doing Conversation Analysis in Latin: The Case of Hedging |
| Pounder, Amanda | Pounder | D4 | 275 | Morphologization of Phonological Processes as Integration |
| Pozza, Marianna | Alfieri & Pozza | D1 | 86 | Adjectival typology in four ancient Indo-European languages |
| Pronk, Tijmen | Pronk | D2 | 189 | Tonogenesis in Baltic and Slavic languages |
| Rainsford, Tom | Trips & Rainsford | D4 | W13.2 | How to use Yang's Principles to model acquisition in diachrony. The case of psych verbs |
| Rapold, Christian | Rapold | D2 | 268 | Secondary lateral obstruents in South Cushitic and their significance for the linguistic history of East Africa |
| Razguliaeva, Mariia | Tikhonov et al. | D2 | 143 | Pronoun history and information structure in 18th century non-religious Russian texts |
| Reetz, Malika | Reetz | D4 | 170 | German V2-Argument Clauses from a Diachronic Perspective |
| Reinöhl, Uta | Reinöhl & Ellison | D5 | 190 | Metaphor, Overtness and Word Order Routinization |
| Reinöhl, Uta | Reinöhl et al. | D2 | 181 | The loss of word-initial consonants in Kera'a – A challenge for phonological theory |
| Reiter, Viktoria | Grestenberger et al. | D4 | W06 | Categorizers in diachrony |
| Reiter, Viktoria | Grestenberger et al. | D4 | W06.0 | Categorizers in diachrony |
| Riad, Tomas | Riad | D2 | 283 | Hypotheses and scenarios in North Germanic tonogenesis |

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------------------|------------------------|-----|--------|--|
| Ricquier, Birgit | Ricquier & Demolin | D3 | 102 | The Chronicle of Lingbe, an Extinct Bantu Language of East Congo |
| Riegger, Chiara | Björnsdóttir et al. | D1 | 203 | The rise of raising in Early Modern English |
| Ritt, Nikolaus | Ritt & Hofmann | D2 | 248 | 'Chained to the rhythm': Using agent-based simulation to model the evolution of stress pattern diversity in English |
| Ritt, Nikolaus | Ritt & Böhm | D2 | W04.5 | Sound changes tend to reduce morphotactic ambiguity |
| Robbeets, Martine | Robbeets | D1 | W01 | From climate change to language change |
| Robbeets, Martine | Robbeets | D1 | W01.1 | From climate change to language change |
| Rodríguez-Somolinos, Amalia | Rodríguez-Somolinos | D5 | 48 | From inference to hearsay: the development of the French parentheticals <i>à ce qu'il paraît, comme il paraît, il paraît, paraît-il</i> |
| Roland, Pooth | Orqueda & Pooth | D2 | W12.8 | Latin <i>ecce</i> : arguments in favor of its development from a PIE demonstrative |
| Rönchen, Philipp | Rönchen et al. | D4 | W03.7 | Using simulated data to evaluate models of Indo-European vocabulary evolution |
| Rosenkvist, Henrik | Rosenkvist | D4 | 50 | Structural ambiguity and reanalysis – the case of Swedish <i>fortsatt</i> |
| Roth, Kerstin | Roth | D3 | 79 | Rhetoric, stylistic and argumentative strategies of German language female authors in the 17th century |
| Roth, Theresa | Roth | D1 | W05.5 | Etymologies and emotions: Historical linguistics as a key to emotion categories |
| Round, Erich | Round et al. | D2 | 196 | The natural stability of 'unnatural' morphology |
| Rüdiger, Jan Oliver | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |
| Russell, Kerri | Russell | D2 | 212 | Properties of Complex Compounds in Old Japanese |
| Rutten, Gijsbert | Drinka et al. | D2 | W02 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Rutten, Gijsbert | Drinka et al. | D2 | W02.1 | Macro-level social motivations for language change: Contact, migration, and globalization |
| Rzyski, Christoph | Rzyski | D2 | W14.3 | From Old Data to Fresh Phylogenies – A Linguistic Data Journey in the Times of CLDF |
| Sæbø, Lilja | Sæbø & Grossman | D5 | W11.6 | A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis |
| Saiz Sánchez, Marta | Saiz Sánchez | D1 | 88 | The periodization of the Pre-Classical French through the study of <i>nennil</i> and <i>non</i> in grammars, remarks and treatises (15th–17th centuries) |
| Salaberri, Iker | Ariztimuño & Salaberri | D1 | 152 | A new perspective on the evolution of mood and negation markers in Proto-Basque |
| Salaberri, Iker | Salaberri | D5 | 30 | Towards an account of the emergence, evolution and variability of emphatic negative coordination in Indo-European, part 2: A diachronic perspective |
| Salmons, Joseph | Salmons | D2 | W02.10 | Verticalization and the historical sociolinguistics of language maintenance |
| Salvesen, Christine Meklenborg | Salvesen | D3 | 109 | Tracing the origins of resumption in Swedish |
| Santamaria, Andrea | Santamaria | D2 | 263 | The Greek suffix <i>-θ-</i> and the Caland System |
| Sapp, Christopher | Sapp et al. | D5 | 207 | Another look at Noun-Genitive vs. Genitive-Noun in Early New High German |

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------------------|----------------------------|-----|-------|---|
| Satō, Tomomi | Satō & Bugaeva | D4 | W10.1 | On stative/active intransitive split within tripartite alignment: A case of Kuril Ainu |
| Schack, Jørgen | Jensen & Schack | D4 | 84 | Adverbs ending in <i>-(l)ig</i> '-ly' and <i>-(l)igt</i> '-ly' in Danish |
| Schäfer, Lea | Schäfer | D3 | 103 | Dramatic texts as a source of stigmatization from below |
| Schlechtweg, Dominik | Schlechtweg | D4 | W03.5 | The LSCD Benchmark – A testbed for diachronic word meaning tasks |
| Schulte im Walde, Sabine | Maurer et al. | D4 | W03.3 | Quantifying Changes in English Noun Compound Productivity and Meaning |
| Schulte, Michael | Schulte | D2 | 125 | A re-assessment of Early Runic Metrics |
| Schützler, Ole | Schützler | D5 | 35 | Third-person verb inflection in Shakespeare's dramatic texts |
| Šefčík, Ondřej | Šefčík | D2 | 70 | Bartholomae's law revisited and remodelled |
| Serangeli, Matilde | Serangeli | D2 | 267 | Rumpled chicken come home to roost. From [TO CARD – IMPURITY] to [TO PURIFY/HEAL (someone) – from DISEASE]. Evidence from Anatolian, Ancient Greek, and Old Indic |
| Seržant, Ilja | Seržant | D2 | W04.6 | Ambiguity avoidance and DOM |
| Shamseddinov, Abdurahman | Shamseddinov & Authier | D1 | 31 | Contact-driven grammaticalization and drift of new terminal tenses from go-periphrasis in Azeri and Kryz (East Caucasian) |
| Shcherbakova, Olena | Shcherbakova et al. | D2 | 96 | Diachronic pathways of definite articles distribution |
| Shimabukuro, Moriyo | Shimabukuro | D4 | W10.2 | Debuccalization of *p in the Naha dialect of the Ryukyuan language |
| Sidwell, Paul | Sidwell | D1 | W01.7 | Austroasiatic dispersal: sea levels and estuarine environments in late Neolithic Mainland SEAsia |
| Sigurðardóttir, Sigríður Sæunn | Sigurðardóttir | D2 | 220 | From complex to simple prepositions in Icelandic: The case of <i>á bak við</i> to <i>bakvið</i> 'behind' |
| Sigurðardóttir, Sigríður Sæunn | Eyþórsson & Sigurðardóttir | D2 | 126 | Micro-level conflict in the productivity of anticausativization strategies: Evidence from the history of Icelandic |
| Sims-Williams, Helen | Sims-Williams | D1 | 147 | Rehabilitating 'non-proportional' analogy |
| Sitchinava, Dmitri | Sitchinava | D5 | 286 | A panchronic corpus of Old East Slavic and Russian: bringing together Slavic historical and modern corpus resources |
| Smirnova, Elena | Smirnova | D2 | W04.1 | The role of ambiguity at different stages of diachronic change |
| Smith, John Charles | Smith | D5 | PL7 | Fifty years of ICHL, 1973–2023 |
| Sobolev, Andrey N. | Sobolev | D2 | W02.5 | Contact as a major Motivation for Linguistic Change in the History of Balkan Slavic |
| Sobotka, Piotr | Kisiel & Sobotka | D2 | 51 | The functional interpretation of semantic and syntactic shifts in the domain of North Slavic "conversive" preposition-pronominal constructions |
| Somers, Joren | Cluyse et al. | D1 | 208 | Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis |
| Somers, Joren | Elens et al. | D1 | 213 | The Alternating Behavior of 'Like' in Old Norse-Icelandic: Facts or Fiction |
| Sonnenhauser, Barbara | Mendoza et al. | D4 | 99 | Anchoring patterns in emerging complement clauses in Slavic |
| Sonnenhauser, Barbara | Widmer & Sonnenhauser | D1 | W07.4 | Linguistic convergence in the Ancient Near East |
| Souag, Lameen | Souag | D1 | W07.3 | Prehistoric language contact in Berber |

| Participant | Authors' abbrev. | Day | No. | Title |
|------------------------|--------------------------|-----|-------|---|
| Sowada, Lena | Sowada | D2 | W02.7 | Language use in Alsace from 1914 to 1919. Private texts between official legislation and individual identity construction |
| Stein, Achim | Cassarà et al. | D4 | W13.3 | Marked vs. unmarked unaccusativity with alternating verbs: Linking diachronic and experimental data |
| Sternefeld, Leah | Sternefeld | D5 | 123 | What is <i>ke</i> and if so how many? – The Persian modal particle <i>ke</i> and its diachronic development |
| Stratton, James | Stratton | D2 | 115 | Where did <i>wer</i> go? Searching for s-curves in lexical change from Old English to Middle English |
| Strauss, Silvie | Strauss | D2 | 235 | Paradigmatic redundancy in the complement system of Basque |
| Struik, Tara | Cassarà et al. | D4 | W13 | New methods for old languages: the comparability of data |
| Struik, Tara | Cassarà et al. | D4 | W13.0 | New methods for old languages: the comparability of data |
| Suleymanov, Murad | Suleymanov | D5 | W08.7 | Semantic Shift and Morphosyntactic Convergence of Tense-Aspect-Mood Categories in Alazan Persian |
| Swanenvleugel, Cid | Swanenvleugel | D5 | 290 | The Sardinian substrate lexicon and its Mediterranean comparanda |
| Tan, Tamisha | Tan | D4 | W06.4 | 'Inalienable' nominalisers across Meto |
| Tan, Tamisha L. | Tan | D2 | 133 | The Lost Cause: Inflection Class in Amarasi |
| Teich, Elke | Degaetano-Ortlieb et al. | D4 | W03 | Computational models of diachronic language change |
| Tian, Zuoyu | Amaral et al. | D4 | W03.6 | Model evaluation for diachronic semantics: A view from Portuguese and Spanish |
| Tieku, Enock Appiah | Tieku | D1 | 281 | Drivers of Diversity in the Construal of Quantity in the World's Languages |
| Tikhonov, Aleksej | Tikhonov et al. | D2 | 143 | Pronoun history and information structure in 18th century non-religious Russian texts |
| Tomaschek, Fabian | Nieder & Tomaschek | D5 | 174 | Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics |
| Torres-Latorre, Aina | Torres-Latorre | D2 | 176 | Synthetic or analytical: factors which explain the formal variation of future and conditional in Old Catalan |
| Tresoldi, Tiago | Tresoldi et al. | D1 | 165 | A Phylogenetic Study of the Cariban Family: Combining Linguistic and Archaeological Data |
| Tresoldi, Tiago | Lindgren & Tresoldi | D3 | 198 | The Charition Mime: Decoding the "Indian Language" through Typology and Entropy |
| Trips, Carola | Trips & Rainsford | D4 | W13.2 | How to use Yang's Principles to model acquisition in diachrony. The case of psych verbs |
| Truswell, Robert | Gisborne & Truswell | D4 | 222 | Contact and the origins of headed <i>wh</i> -relatives in Hungarian |
| Ulman, Vít | Ulman | D2 | 253 | Genesis of the Japanese Compound Particles |
| van Beek, Lucien | van Beek | D1 | W05.1 | Clouds or Arrows? Conceptual Metaphors and the Etymology of Homeric Greek <i>kertoméō</i> 'to mock; taunt' |
| van Beek, Lucien | Kölligan & van Beek | D1 | W05 | Conceptual metaphors in a comparative and diachronic perspective |
| van Dam, Kellen Parker | van Dam | D5 | 41 | Internal subgrouping of Northern Naga based on Bayesian phylogenetic analysis |
| Van de Velde, Freek | Nijs et al. | D2 | W02.4 | An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German |
| van Kemenade, Ans | van Kemenade | D4 | PL5 | Word order change, architecture and interfaces: Evidence from V2 word orders and their loss in the history of English |

| Participant | Authors' abbrev. | Day | No. | Title |
|------------------------|-------------------------|-----|-------|---|
| Västerdal, Ida | Västerdal | D2 | 77 | A case of Verschärfung in the Swedish dialect from Stora Rågö in Estonia |
| Verkerk, Annemarie | Verkerk et al. | D3 | 242 | Exploiting phylogenetic modeling to uncover directionality in the emergence of universals |
| Verkerk, Annemarie | Krasnoukhova et al. | D4 | W09.2 | Negated but similar – Negation in the domains of locative, existential, and possessive predication: The case of Indo-European |
| Vincent, Nigel | Börjars & Vincent | D1 | 52 | Auxiliary, light or lexical: the history of GO verbs |
| Vincent, Nigel | Vincent et al. | D4 | PL6 | Linguistic models (with a focus on morphosyntactic change) |
| Visser, Lourens | Visser | D1 | 111 | Adverbs of degree from Old to Early New High German |
| Voigtmann, Sophia | Voigtmann | D2 | 142 | Where do all the NPs go? – A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900 |
| Vydrin, Valentin | Perekhvalskaya & Vydrin | D5 | W11.7 | Tonal density and its correlation with the types of tonal systems: Diachronic aspects |
| Walkden, George | Vincent et al. | D4 | PL6 | Linguistic models (with a focus on morphosyntactic change) |
| Walkden, George | Björnsdóttir et al. | D1 | 203 | The rise of raising in Early Modern English |
| Werner, Martina | Werner | D4 | W06.5 | When verbal complexes become nouns via infinitive nominalization: A parallel to the verbal domain or category-individual? |
| Westergaard, Lennart | Westergaard & Boye | D3 | 90 | On semantic change in grammaticalization: Why it is never metaphoric |
| Westergaard, Lennart | Westergaard | D5 | 92 | The long and winding road of the Danish evidential <i>vel</i> – from epistemic modality via concessivity to evidentiality |
| Wichers Schreur, Jesse | Wichers Schreur | D2 | 221 | Differential Place Marking and the reconstruction of the Proto-Nakh system of spatial cases |
| Wichmann, Søren | Jonjić et al. | D2 | 144 | Isoglosses and distributions of features – Analyses of the <i>Dialectological Atlas of the Russian Language</i> |
| Widmer, Paul | Widmer & Sonnenhauser | D1 | W07.4 | Linguistic convergence in the Ancient Near East |
| Wieczorek, Aleksandra | Wieczorek | D5 | 269 | Discontinuous noun phrases containing adjective or adjective-like modifiers in Middle Polish texts. Preliminary research conducted on an experimental dependency treebank |
| Wiemer, Björn | Mendoza et al. | D4 | 99 | Anchoring patterns in emerging complement clauses in Slavic |
| Wier, Thomas | Wier | D1 | W07.5 | Language Contact in the Ancient Caucasus: the View from Kartvelian |
| Wiklund, Tilo | Rönchen et al. | D4 | W03.7 | Using simulated data to evaluate models of Indo-European vocabulary evolution |
| Willis, David | Darling et al. | D2 | 180 | The Diachrony of Person-Number Marking of Subjects in Celtic |
| Wolfe, Sam | Wolfe | D4 | 78 | Parallel Phases in the History of French |
| Wolfer, Sascha | Engelberg et al. | D3 | PL4 | Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language |
| Wolfsgruber, Anne | Wolfsgruber | D2 | W04.4 | Text-type specific conventions, subordinate environments and ambiguity (avoidance) in Medieval Spanish passive <i>se</i> -constructions |
| Yurayong, Chingduang | Yurayong et al. | D1 | W07.6 | An archaeolinguistic approach to Indianisation and Sincisation of languages in Eastern Eurasia |
| Zampetta, Silvia | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |
| Zanchi, Chiara | Zampetta et al. | D1 | W05.6 | <i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin |

| Participant | Authors' abbrev. | Day | No. | Title |
|--------------------|-------------------------|------------|------------|--|
| Zanchi, Chiara | Brigada Villa et al. | D2 | W14.2 | Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages |
| Zehentner, Eva | Zehentner & De Cesare | D2 | W04 | Ambiguity (avoidance) as a factor in language change |
| Zeng, Xiuwei | Zeng | D5 | 164 | Ditransitive GIVE-construction in three Hainan Min-Chinese: Interaction between inherited structures and contact-induced changes |