



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

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

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

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

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

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

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

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

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

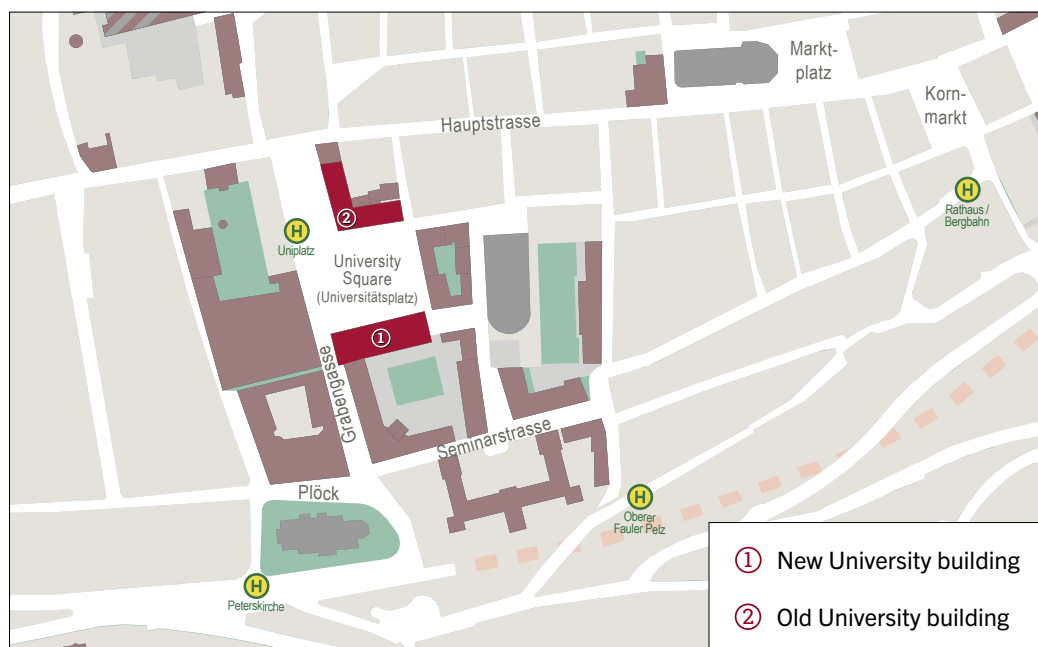
**Prof. Dr. Jadranka Gvozdanović**  
President of ISHL, Chair of ICHL26

# 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	Coffee break				
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	Lunch break				
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	Break				
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	Coffee break		
10:30–11:00	Robbeets		
11:00–11:30	Berge		
11:30–12:00	Knapen		
12:00–12:30	Miyamoto		
12:30–13:30	Lunch break		
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	Break		
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	<i>Coffee break</i>				
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.	Šefčík	Bru	Bossuyt & Daveloose	Pronk
12:30–13:30	<i>Lunch break</i>				
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	<i>Break</i>				
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	<i>Coffee break</i>		
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	<i>Lunch break</i>		
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	<i>Break</i>		
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	<i>Coffee break</i>				
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	<i>Lunch break</i>				
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	<i>Break</i>				
16:00–16:30	Pounder	McCarley	Rosenkvist	Pompei	Ebert et al.
16:30–17:00				Gisborne & Truswell	Klævik-Pettersen
17:00–17:30	<i>Business meeting, then conference dinner</i>				
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	<i>Lunch break</i>			
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	<i>Break</i>			
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	<i>Lunch break</i>
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	<i>Break</i>
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 für die Deutsche Sprache 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](http://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 9<sup>th</sup> 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 9<sup>th</sup> to 8<sup>th</sup> 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 9<sup>th</sup> to 8<sup>th</sup> century BC. The latter phase is immigration from lower Nagdong River basin to Fukuoka Plains at 7<sup>th</sup> 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 6<sup>th</sup> to 5<sup>th</sup> 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 6<sup>th</sup> to 5<sup>th</sup> 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 \**ŋwə* supports dating the PTB/Sinitic split to after 5.6K YBP. Rice was first domesticated in the lower Yangtse 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 \**jaŋ*, 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<sup>h</sup>it* (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 rarely 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](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](mailto:bridget.drinka@utsa.edu))  
 Terttu Nevalainen ([terttu.nevalainen@helsinki.fi](mailto:terttu.nevalainen@helsinki.fi))  
 Gijsbert Rutten ([g.j.rutten@hum.leidenuniv.nl](mailto: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/](http://www.un.org/dppa/decolonization/))
- UNHCR: The UN Refugee Agency ([www.unhcr.org/](http://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 María 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 14<sup>th</sup> and 19<sup>th</sup> 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.

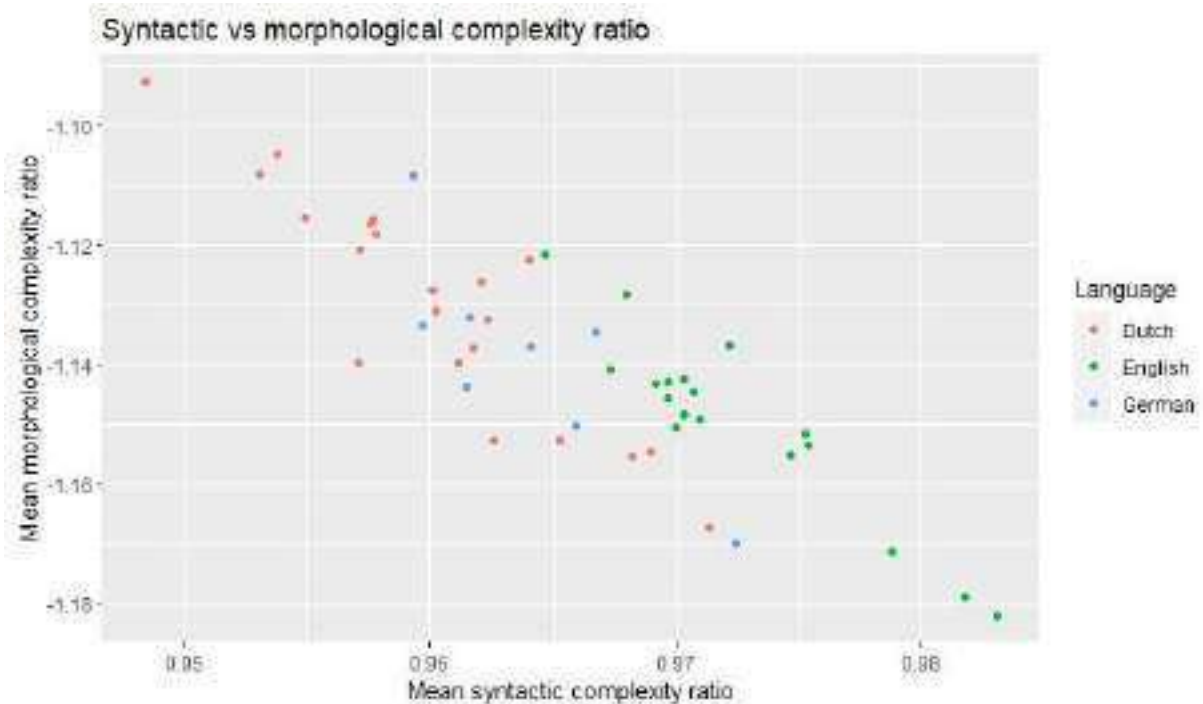


Figure 1: Syntactic vs morphological complexity ratio

## 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 16<sup>th</sup> and 17<sup>th</sup> 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.<sup>1</sup> Secondly, the emergence of national ideologies in the 19<sup>th</sup> century accompanied by a search for the language variety with sufficient cultural weight to represent the nation (such as the language of the 16<sup>th</sup> 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.

---

<sup>1</sup> 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 14<sup>th</sup> century to become almost categorical at the beginning of the 15<sup>th</sup> century. Shortly after, however, the long forms declined rapidly and disappeared in the 16<sup>th</sup> 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-17<sup>th</sup> 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 (17<sup>th</sup> C on) and then English (19<sup>th</sup> 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)<sup>1</sup> [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

<sup>1</sup><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 Jensen<sup>1</sup>, Isabell Landwehr<sup>2</sup>, Barbara McGillivray<sup>3</sup> and Stefania Degaetano-Ortlieb<sup>2</sup>

<sup>1</sup>Springer Nature Group, <sup>2</sup>Saarland University, <sup>3</sup>King'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 Shkenca 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). BAHF: 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 Rönchen<sup>1</sup>, Oscar Billing<sup>1</sup>, and Tilo Wiklund<sup>2</sup>

<sup>1</sup>Department of Linguistics and Philology, Uppsala University, Sweden

<sup>2</sup>Chief 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.](#)
- [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](mailto:eva.zehentner@es.uzh.ch)  
Ilaria De Cesare (Universität Potsdam), [decesare@uni-potsdam.de](mailto: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](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  $se_{\text{passive}}$  V subject structure to  $se_{\text{imp}}$  V object<sub>(former subject)</sub> 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; Wełna 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-Kończyk 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](mailto: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](mailto: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. "Altlatein *cortumiō* 'Geländeausschnitt', Idg. \**kr-tomh*<sub>1</sub>-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](mailto:riccardo.ginevra@unicatt.it) / [rginevra@chs.harvard.edu](mailto: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 miśráṁ 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.), *Mímesis, 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. *AIΩN-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*<sup>2</sup> 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.
- LIV<sup>2</sup> = Rix, Helmut and Kümmel, Martin (2001). *Lexikon der indogermanischen Verben: Die Wurzeln und ihre Primärstammbildungen*. 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*

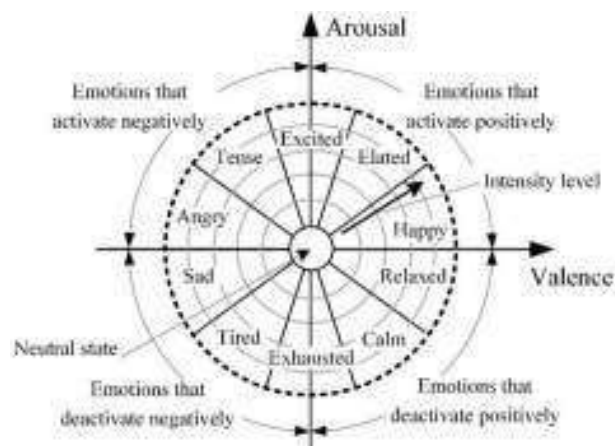
One of the major issues in researching the history of emotions has been the question whether our ancestors did feel completely different or indeed very much the same as we do today. The positions voting for the latter may be labelled, on the one hand, “universalist” – claiming a set of emotions as “basic” 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*. 2<sup>nd</sup> 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*. 2<sup>nd</sup> edition. Tübingen / Basel: A. Francke Verlag UTB.



Silvia Zampetta<sup>+</sup>, Erica Biagetti<sup>\*</sup>, Nicolò de Rossi<sup>\*</sup>, Martina Giuliani<sup>#</sup>, Chiara Zanchi<sup>\*</sup>, Silvia Luraghi<sup>\*</sup>

<sup>\*</sup>University of Pavia, <sup>+</sup>University of Bologna, <sup>#</sup>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](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](mailto: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 *σκούζα* ‘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 \*h<sub>2</sub>eǵd- ‘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](mailto:laura.grestenberger@oeaw.ac.at))

Iris Kamil, University of Edinburgh ([iris.kamil@ed.ac.uk](mailto:iris.kamil@ed.ac.uk))

Viktoria Reiter, University of Vienna ([viktoria.reiter@univie.ac.at](mailto: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*]<sub>n-s</sub> ‘king’: [[\**basil-eú*]<sub>n-j</sub>]<sub>v-ō</sub> ‘am/act as king’ → Modern Greek *stóx-os* ‘target’ [[*stox*]<sub>n-év</sub>]<sub>v-o</sub> ‘to aim at’; Pre-Proto-Algonquian [\**api*]<sub>v-hm</sub>]<sub>n</sub> ‘sitting place, seat’, \**net*-[[*api*]<sub>v-hm</sub>]<sub>n-ena-n</sub> ‘where we sit; our sitting place’ → Proto-Algonquian \**net*-[[*api*]<sub>v-hm</sub>]<sub>v?</sub>-*ena-n* (*ma-hi*) ‘we sit over there’ (Oxford 2014: 14-15)
2.  $a \rightarrow v$ : Gm. *Kraft* ‘strength’: [[*kräft*]<sub>n-ig</sub>]<sub>a</sub> ‘with strength, strong’; [[[*kräft*]<sub>n-ig</sub>]<sub>a-en</sub>]<sub>v/T[-fin]</sub> ‘to strengthen’ → *Pein* ‘pain’: [[[*pein*]<sub>n-ig</sub>]<sub>v-en</sub>]<sub>T[-fin]</sub> ‘to torture’ (\**pein-ig* ‘painful’)
3.  $a \rightarrow v/ptcp$ : Sanskrit *ásva-* ‘horse’: [[*asv*]<sub>n-ín</sub>]<sub>a-</sub> ‘possessing horses’ → √*yaj* ‘sacrifice’: [*yāj-ín*]<sub>ptcp-</sub> ‘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 *\*-eh<sub>1</sub>-* 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-* (<*\*-eh<sub>2</sub>-*) declension: */-ā-/* < *\*-eh<sub>2</sub>-ye* (with loss of the intervocalic glide, subsequent merging of the vowel sequence and eventual reanalysis of the resulting piece as a *v<sup>0</sup>-derivative*): e.g., *curāmus* ‘cure’ (cf. *curā* ‘cure’). The *-ē-* conjugation developed mostly from the stative suffix *-ē-* (<*\*-eh<sub>1</sub>-*) and from causatives in *\*-eye-* (with *o*-grade of root): */-ē-/* < *\*-eh<sub>1</sub>-*; e.g., *sedēmus* ‘we sit’ (<*\*sed-eh<sub>1</sub>-*; 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<sup>w</sup>en-ye-*). *ě-*, *-ī-*. The pieces */-ě-/* and */-ĩ-/* developed from actional/aspectual *\*-e*, *\*-ye* (*legimus* <*\*leg<sup>y</sup>-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 *\*-eh<sub>2</sub>-ye* but also from de-adjectival factitive with the suffix *\*-h<sub>2</sub>: 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 *v<sup>0</sup>-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<sup>h</sup>rā-* ‘beautiful’ < *śub<sup>h</sup>-* ‘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<sup>h</sup>rā-* (see above), 24.8% are compound adjectives (that is, the *bahuvrīhi* type termed by Indian grammarians) such as *híraṇya-pāṇi-* ‘having gold hands’, 13.7% are prefixed adjectives such as *su-vīra-* ‘heroic’ < *vīrá-* ‘hero’, 9.8% are denominative adjectives such as *pítṛiya-* ‘paternal’ < *pítṛ-* ‘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 [1896<sup>1</sup>]. *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 3<sup>rd</sup> 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)

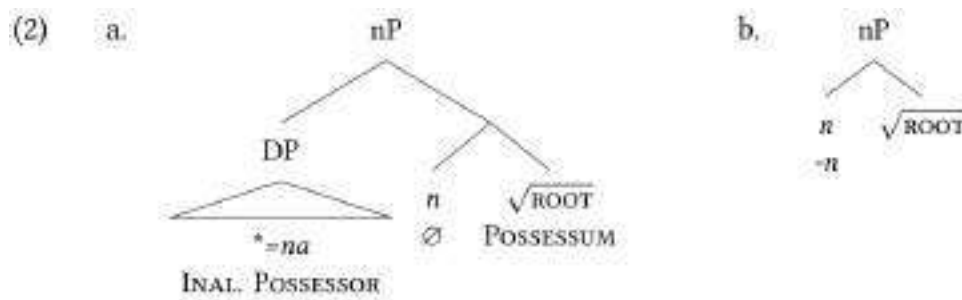
PMP *\*haRəzan* ‘ladder’ > *era-ʔ*, *era-k* [Amarasi] ‘stairs’, *ela-k* [Molo] ‘ladder’

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<sub>AKK</sub>-Anchattung*

(1d) *(das) Chatten der Freunde<sub>GEN</sub>* ‘(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): Reanalysis<sup>2</sup> 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 Nahuatl 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 Nahuatl 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 Nahuatl 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-Nahuatl 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 Nahuatl 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)<sup>2</sup> 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. Germei. 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](mailto:chingduang.yurayong@helsinki.fi) > Lexical Indianisation of Suvarṇabhūmi; Typological Sinicisation of Korea

2. Sami Honkasalo (University of Helsinki) – [sami.honkasalo@helsinki.fi](mailto: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](mailto:puiyu.szeto@unive.it) > Typological and lexical Sinicisation of Lingnan
4. Charlotte Ye Xu (Nanjing University) – [charlottexuye@gmail.com](mailto:charlottexuye@gmail.com) > Lexical Sinicisation of Korea and Japan

#### *Historiography*

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

#### *Archaeology*

1. Youngchan Oh (Ewha Womans University) – [ycoh@ewha.ac.kr](mailto: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](mailto: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](mailto:zkdwlsk0111@snu.ac.kr) > Trade of glass beads in India and Southeast Asia
4. Cheunsoo Park (Kyungpook National University) – [chspark@knu.ac.kr](mailto: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](mailto: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 1<sup>st</sup> and 2<sup>nd</sup> 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](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](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 Ἀρσοί, 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. \*mayuŋxa-.

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-æŋ*, I *cær-ʒyn-æŋ* ‘I will live’) must have its origin in nominal compounds \*X-čānāh ‘desiring X<sub>N</sub>’ (whence deverbal ‘(be) wanting to X<sub>V</sub>’ > ‘X<sub>V</sub>-FUT’; Kim fthc. A); given the derivational isolation of \*-čānah- 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, 1<sup>st</sup>-5<sup>th</sup> 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).

3<sup>rd</sup> person and 1<sup>st</sup> person > 3<sup>rd</sup> person and 1SG > 3<sup>rd</sup> 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.     ɔæmən           wetiaj           mæ-s-u  
      I               to sleep       IPRF-sleep.PRS-1SG  
      “I am sleeping” or “I am on the edge of falling asleep”
2.     ɔemæ sipatæke=man       æs-e=ne  
      we   clothes=1PL       buy.PST-3PL=tobe.1PL  
      ɔistæ   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: *Našrīye-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 *bidän* (cognate of Standard Persian *budan* ‘to be’) and *šidän* (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>män kitab bu-xun-um.</i><br/>I book IPFV-read<sub>1</sub>-1SG<br/>‘I am reading a book. / I read books.’</p> <p>b. <i>ägär xeyli gäp bi-zän-um</i><br/>if much word IPFV-hit<sub>1</sub>-1SG<br/><i>män=ä järimä bu-kun-id.</i><br/>I=DDO fine IPFV-do<sub>1</sub>-3<br/>‘If I talk too much, he is (definitely) going to fine me.’</p> <p>c. <i>ägär vaxt=im bi-šid</i><br/>if time=POSS:1SG IPFV-be<sub>1,3</sub><br/><i>kitab=ä mu-xun-um.</i><br/>book=DDO EVT-read<sub>1</sub>-1SG<br/>‘If I (hypothetically) have time, I will read the book.’</p> | <p>a. <i>nun=mun=ä mu-xor-än.</i><br/>bread=POSS:1PL=DDO IPFV-eat<sub>1</sub>-3PL<br/>‘They eat / are eating our bread.’</p> <p>b. <i>umru borun bə-riz-id.</i><br/>today rain MOD-flow<sub>1</sub>-3<br/>‘Today it is going to rain.’</p> <p>c. <i>ayri bi-šin-äd</i><br/>separate MOD-sit<sub>1</sub>-3<br/><i>ayri mi-šin-äd.</i><br/>separate IPFV-sit<sub>1</sub>-3<br/>‘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 (as 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 eíche 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) *Pað 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-í 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](mailto:o.krasnoukhova@hum.leidenuniv.nl), [shahar.shirtz@gmail.com](mailto:shahar.shirtz@gmail.com), [annemarie.verkerk@uni-saarland.de](mailto: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<sup>hi</sup>-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](mailto: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](mailto: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*),<sup>1</sup> 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.<sup>2</sup>

- (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),

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

<sup>2</sup> 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](mailto: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 7<sup>th</sup> – 13h30-17h**

**Organisers:**

**Étienne Baudel** – École des hautes études en sciences sociales (Paris) ([etienne.baudel@gmail.com](mailto:etienne.baudel@gmail.com))

**Aleksandra Jarosz** – Nicolaus Copernicus University in Toruń ([aljarosz@umk.pl](mailto:aljarosz@umk.pl))

**Georg Orlandi** – Osaka University ([georg.orlandi@gmail.com](mailto: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 motodzuita 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 20<sup>th</sup> 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 narzeca 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  $\phi$  at some point before becoming *h*. This change can be attested by historical documents written in Korean and Chinese (Tawata 2010 and Li 2015).

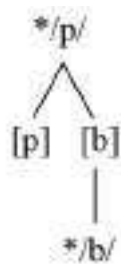
It is important to note, however, that this change did not necessarily occur in all dialects of the language – some still retained *p*. The word for ‘nose’ in Yoron dialect is *pana* (Kiku and Takahashi 2005), for example, while the same word is *hana* in Naha (Uchima and Nohara 2006). The same phenomenon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

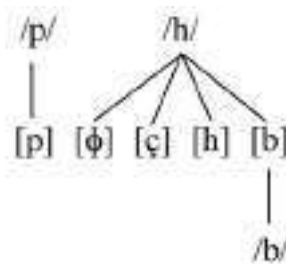
- (1) sampin ‘jasmine tea’: sammin ‘calculation’

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.

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

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

	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.

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

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

	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 «Hyakuniñisshu»*. 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 (Luo') 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 *mâjka*, *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’  
*pàdi* ‘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<sup>hi</sup>-*) 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)<sup>↓</sup>(pàà) ‘to arrive’  
(3) /gò =<sup>↓</sup>i bwà / → gò =<sup>↓</sup>i bwà ‘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  $/b̥$ -, 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  $/b̥$ -/ (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  $[b̥]$ , while younger speakers may produce a devoiced  $[p]$ , 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  $*b > /b̥$ - 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  $/b̥$ -/  $>$   $[p^h]$ -/ 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.
- Gehrman, 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*´ *í:ken*, and for *thí:ken* ‘that’, they use *tho*´ *í:ken*. Mohawk *kèn*´ *en* or *ken*´ is the proximal locative adverbial ‘here’, and *tho* is the distal locative adverbial ‘there’. The word *í: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*: *í: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-iij</i>	way (argument), way-GEN

Case forms	<i>naa-d-[a]x-</i> CASE	‘the N-CASE on the close side [of]’	<i>zam-iŋ-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=güü.]*  
 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=güü.*  
 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 Chukotko-Kamchatkan, Chukotko-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 (Lezgian, 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]<sub>i</sub>. On u {[dva rada]<sub>\*i,j</sub>}/[ta dva rada]<sub>i,\*j</sub> 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]<sub>i</sub>. {[Deo reči]<sub>\*i,j</sub> / [Taj deo reči]<sub>i,\*j</sub>} 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<sub>i</sub> je napisao rad. {[Neko]<sub>\*i,k</sub> / [Taj neko]<sub>i,\*k</sub>} 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 Frajzngier (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”; **Stavinski**, 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 15<sup>th</sup> 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 *\*h<sub>2</sub>eu-* '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-* (Wodtko 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 *\*h<sub>1</sub>éd=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 1<sup>st</sup> 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 *\*h<sub>1</sub>éd=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 Papiament-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](mailto: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** plest 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 12<sup>th</sup> 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  $\theta$ -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 Trips<sup>1</sup> & Tom Rainsford<sup>2</sup>

<sup>1</sup>University of Mannheim, <sup>2</sup>University 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 Cassarà<sup>1</sup>, Achim Stein<sup>1</sup>, Chantal van Dijk<sup>1,2</sup>, Holger Hopp<sup>2</sup>

<sup>1</sup>University of Stuttgart, <sup>2</sup>TU 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 Jäger<sup>1</sup>, Robert Forkel<sup>2</sup>, Johann-Mattis List<sup>2,3</sup>

<sup>1</sup> Institute of Linguistics, University of Tübingen

<sup>2</sup> Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

<sup>3</sup> 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<sup>1,2</sup>, Erica Biagetti<sup>2</sup>, Chiara Zanchi<sup>2</sup>, Silvia Luraghi<sup>2</sup>

<sup>1</sup> University of Bergamo

<sup>2</sup> 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 Rzymiski<sup>1</sup>

<sup>1</sup> 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 Forkel<sup>1</sup>, Simon Greenhill<sup>2</sup>

<sup>1</sup> Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

<sup>2</sup> 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 Dellert<sup>1</sup>, Verena Blaschke<sup>2</sup>

<sup>1</sup> Institute of Linguistics, University of Tübingen

<sup>2</sup> 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 ea. 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 Blum<sup>1</sup>, Johann-Mattis List<sup>1, 2</sup>

<sup>1</sup> Dep. of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology Leipzig

<sup>2</sup> 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](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 Mertner<sup>1</sup>, Matías Guzmán Naranjo<sup>1</sup>

<sup>1</sup> 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>   | <i>va-ka-wok-a...</i>  |
|     | 3PL.SBJ-PST-eat-PST  | 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* <ftrength> 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 (*-cip+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(†) 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 18<sup>th</sup> and 19<sup>th</sup> 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.<sup>1</sup>

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.

---

<sup>1</sup> 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<sup>th</sup> this alarum, and the noise?* [1 Henry VI, 1.6]
- (2) *My dearest love, Duncan comes<sup>s</sup> 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 14<sup>th</sup>, 15<sup>th</sup> and 16<sup>th</sup> 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 ὅδ’ 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 *chéng rán***

Modern Chinese *chéng rán* 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 chéng rán 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 *Chéng rán huàjù yǒu yúlè de gōngnéng, dàn shì tā bù yīng línwéi fàn hòu tiándiǎn.*  
 although drama have recreation NOMZ function but it NEG should reduce.to  
 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 *chéng rán* and a SC *chéng rán* developed from a separated clause *chéng rán* ‘(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: “Chéng rán, 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 *chéng rán* 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, chéng rán 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 19<sup>th</sup> 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 (16<sup>th</sup>-18<sup>th</sup> 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 17<sup>th</sup>- and 18<sup>th</sup>-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 17<sup>th</sup> and 18<sup>th</sup> 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](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**á*, *sabr**ía*) 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 20<sup>th</sup> 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 19<sup>th</sup> 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 19<sup>th</sup> century, up until the beginning of the 20<sup>th</sup> 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 18<sup>th</sup> 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 19<sup>th</sup> 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. I: 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.

## Elektronic 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, potomu peredъ  
knjazemъ na srokъ ne stalъ.*  
CONJ old.GEN.SG accuse.PST.3SG CONJ PREP  
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 (10<sup>th</sup> to 20<sup>th</sup> 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).

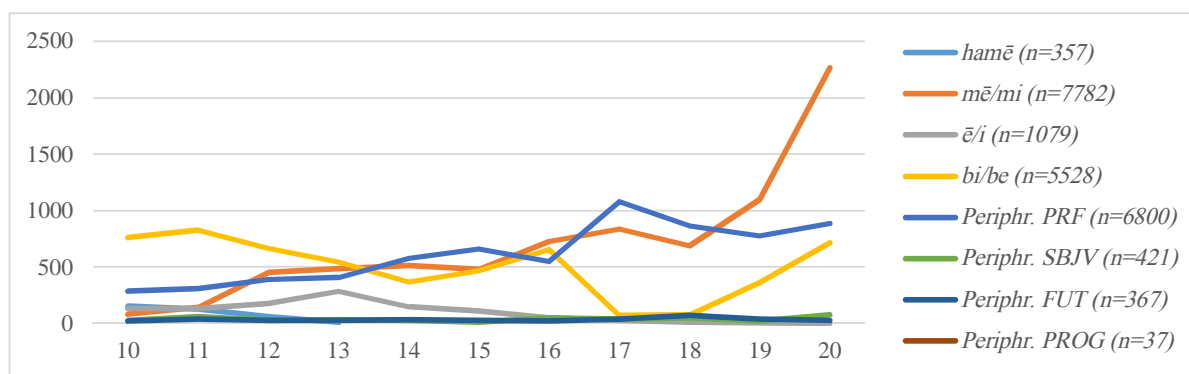


Figure 1. The token frequency of aspect-mood markers in the data set (by century)

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<sup>e</sup>-XVII<sup>e</sup> 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 mən̄í-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 mætətɛæ 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 (2<sup>nd</sup>)*, 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](https://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*, 2<sup>nd</sup> 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* ~ *guek*, 2pl *zuek* (~ *zuok*). Although essentially this system is found since the first extensive documentation of the language (in the 16<sup>th</sup> 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 *guek* 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 *guek* 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 *guek*, *zuek* involves the affixation of a demonstrative.
- b) The 1pl form *guek* is actually attested (in a Basque gloss to an 11<sup>th</sup> c. Latin document) before Sp. *nosotros* ‘we’ and *vosotros* ‘you.pl’ became grammaticalized (12<sup>th</sup>-15<sup>th</sup> 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 20<sup>th</sup> 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 6<sup>th</sup>-8<sup>th</sup> 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 (12<sup>th</sup>-13<sup>th</sup> c.).

### 4. Conclusion

This paper argues that the emergence of the new pronoun forms may therefore be set within a time span around the 10<sup>th</sup> 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 3<sup>rd</sup> 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<sup>a</sup>. 1923-1925. *Morfología vasca*. Euskaltzaindia: Bilbo.
- de Jonge, Bob & Dorien Nieuwenhuysen. 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* ([søit]) *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 NEG<sub>1</sub> easy NEG<sub>2</sub>  
 ‘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 17<sup>th</sup> century preachers use more biblical features than early 17<sup>th</sup> century preachers – there is significant synchronic and diachronic variation throughout the 17<sup>th</sup> 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<br>hād | ME Compound<br>child hōd | contemporary English derivative/suffix<br>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 (1969<sup>2</sup>) 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<br>gaydar<br>jewdar<br>humordar      | (4) | -which from sandwich<br>fishwich<br>hamwich<br>veggiewich                 | (6) | glut- from gluten<br>glutamine<br>glutamate<br>glutaminase |
| (3) | -gasm from orgasm<br>wargasm<br>nerdgasm<br>shoegasm | (5) | Franken- from Frankenstein<br>frankenfood<br>frankenplant<br>frankenscene | (7) | heli- from helicopter<br>heliport<br>helibus<br>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 (4<sup>th</sup>-6<sup>th</sup> 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 /en/ ‘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 \*h<sub>1</sub>en ‘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 /enbu/ ‘to fill’ □ Modern Tosk *mbush*, whereas in Malsia it is *bush*.

The Modern Tosk and Gheg preposition *ne* ‘in’ is derived from Old Albanian (15<sup>th</sup>-16<sup>th</sup> CE) *nde*. 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 ta* /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 *ta* /tã:/ ‘it’. The Modern Tosk equivalent *te* /tə/ is cited by Kortland (2010) as being derived from an earlier \**tom*/\**tām* (Chronology can be posited as \**tom* > *ta* > *te*).

The creation of the Modern Tosk preposition *nga* /ŋga/ ‘where, where from’ is described in Forston (2010: 455) as “... *nga* is shortened from \**en-ka* ‘where, from where’, with \**en* ‘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 (7<sup>th</sup>-2<sup>nd</sup> 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 πι. 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 Gjuhes Se Sotme Shqipe*. Prishtine, Kosove. Rilindja Prishtine.
- Fortson, Benjamin W. IV. 2010. *Indo European Language and Culture*. An Introduction Second Edition. Malden, MA. Wiley-Blackwell.
- Gjinari, Jorgji. 1989. *Dialektet e Gjuhes Shqipe*. Akademia e Shkencave e RPS te Shqiperise. Instituti I Gjuhesise dhe I Letersise. Tirane, 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. Tirane, Albania.
- . 1974. *E folmja e Hotit*. In Dialektologjia Shqiptare II. Tirane, Albania.
- . 1975. *E folmja e Kelmendit*. In Dialektologjia Shqiptare II. Tirane, Albania.

**Uncovering lost paths in the Congo rainforest:  
A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu**

Sara Pacchiarotti<sup>1</sup>, Natalia Chousou-Polydouri<sup>2</sup>, Jean-Pierre Donzo<sup>1,3</sup>, Guy Kouarata<sup>1</sup>, Lorenzo Maselli<sup>1,4</sup>, and Koen Bostoen<sup>1</sup>

<sup>1</sup>UGent Centre for Bantu Studies, Ghent University, Belgium

<sup>2</sup>Department of Comparative Language Science, University of Zurich, Switzerland

<sup>3</sup>ISP Gombe, Kinshasa, DRC

<sup>4</sup>Service 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.<sup>2</sup> 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  $\Delta$  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 ( $\Delta\Delta > DD^h$ )<sup>3</sup> in Indic,<sup>4</sup> the right plosive is non-aspirated plosive in Iranian:

- i.  $D^h + T > \Delta\Delta > DD^h$  (Indic)
- ii.  $D^h + T > \Delta\Delta > DD/\text{Ð}\text{Ð}$  (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  $\Delta Z$  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

<sup>1</sup> IE  $ts > OIA ts$ , Av.  $Os$ ; IE  $d^h s > OIA ts$ , Av.  $Oz$ ; IE  $ss > OIA ts$ , Av.  $Os$ ; IE  $Ks > OIA ks$ , Av.  $Oš$ ; IE  $g^h s > OIA ks$ , Av.  $Ož$ ; II  $šs > OIA ks$ , Av.  $Oš$ .

<sup>2</sup> 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).

<sup>3</sup> We assume, similarly to D. G. Miller (1977), that voice was a primary quality, not aspiration.

<sup>4</sup> 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.
- LIV<sup>2</sup> = 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. 1996<sup>3</sup>. *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).<sup>1</sup> It appears that the concept of dialect in this period is not geographic, but cultural, more akin to register.<sup>2</sup> For example, Ionic forms are often described as 'poetic',<sup>3</sup> 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' ("ἀηδές πάνυ" (*Ecl.* 339; *Ecl.* 332.)) forms, some of which are 'so wrong that not even Menander uses [them]' ("οὕτως ἀδόκιμον ὡς μηδὲ Μένανδρον αὐτῷ χρῆσασθαι" (*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.<sup>4</sup> 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.

<sup>1</sup> This contradicts Swain's (1996: 53) observation that 'all other dialects, including Homeric Greek, are firmly rejected.'

<sup>2</sup> 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.

<sup>3</sup> This idea is also found in Strabo 1.2.6 and Hermogenes *On Style* 2. 319f.

<sup>4</sup> 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.