

January 20<sup>th</sup>, 2023

### Automating Comparative Reconstructions: Case Study in Austronesian and Ongan

Although comparative reconstruction has always been one of the key endeavors of linguistics, there exists no widely accepted method for evaluating its applications (Michalove, 1998). Instead, evaluation is conducted through debate, often spanning decades, as in the case of Altaic, Nostratic, and, more recently, Dene-Yeniseian. Previous attempts to introduce quantitative measures for genetic relatedness are heuristics for estimating similarity, usually either by calculating the average phonetic distance for each putative word-pair (Downey et al., 2008; Kondrak, 2003) or by computing the proportion of cognates between the two wordlists (Chang et al., 2015; Atkinson & Gray, 2003). Since none of these previous attempts engage with diachronic change directly, most researchers agree that, while they are useful when manual reconstruction is not feasible, traditional methods are still the gold standard (Kiparsky, 2015).

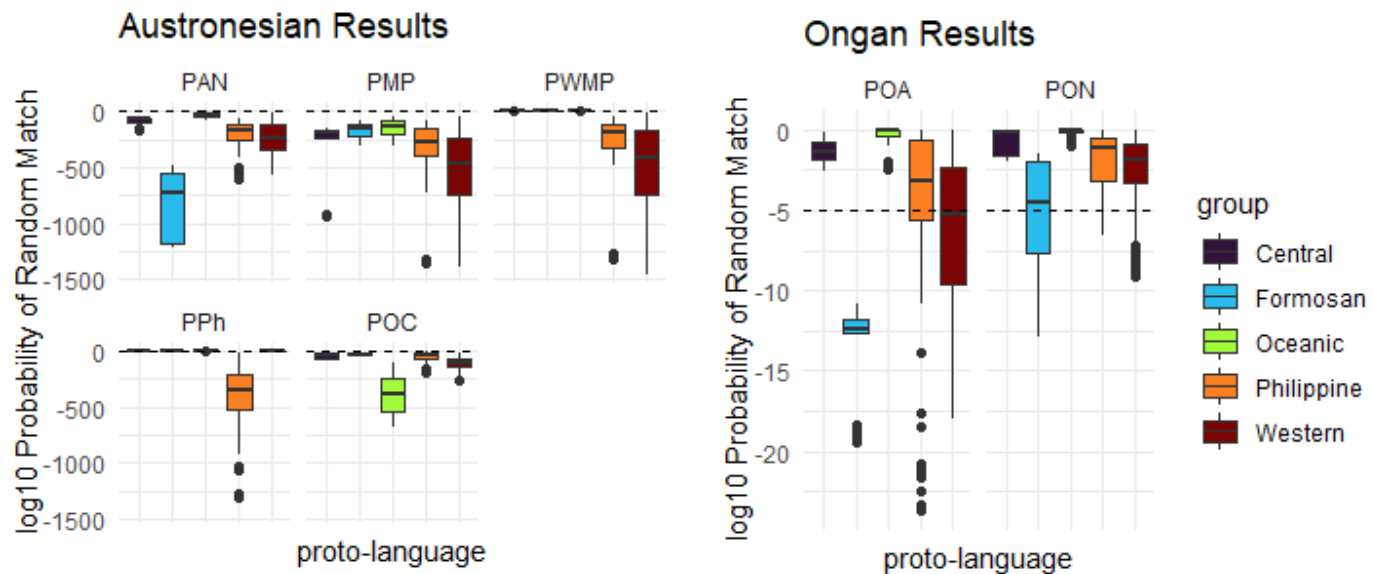
I present a probabilistic framework for evaluating comparative reconstruction attempts. The series of transformations – sound changes, borrowing, semantic change, etc – serves as the input to the framework's evaluation function. The output is the estimated probability that a randomly generated wordlist merits a reconstruction from the mother language using the same number of transformations or fewer than required by the daughter language. Thus, the framework evaluates reconstruction attempts themselves rather than the original dataset, setting it apart from previous quantitative measures.

The framework was incorporated into a simulated annealing learning algorithm, where reconstructions from a mother wordlist to a daughter wordlist were suggested stochastically with a bias toward decreasing the probability of a random match. The algorithm was tested on a genetically diverse sample of Austronesian languages and 5 Austronesian proto-languages. Figure 1 presents the probability of a random match in automated reconstructions from the proto-languages to the 5 Austronesian groups tested, as defined in the Comparative Austronesian Dictionary (Trussel & Blust, 2010). The results are in line with general knowledge in the Austronesian field. For the 237 comparisons between an Austronesian proto-language and a direct descendant, the algorithm always found a reconstruction with a probability of a random match below the chosen cut-off of reliability at .0001. The probability of a random match appears to be strongly correlated with the time depth of the reconstruction.

The case study was further extended to evaluate the putative Ongan-Austronesian connection (Blevins, 2007), a hypothesis not generally accepted in the field (Blust, 2014). Figure 2 presents the probability of a random match in reconstructions from proto-Ongan and proto-Ongan-Austronesian to the 5 Austronesian groups. In reconstructions from proto-Ongan-Austronesian to the Austronesian languages, the results are mixed with the algorithm finding probabilistically non-arbitrary reconstructions to 26 of the 74 of the Austronesian languages tested. Reconstructions from proto-Ongan-Austronesian to the Ongan languages are similarly mixed, with some extremely convincing and others not at all. In general, the results with respect to the Ongan-Austronesian hypothesis appear promising, but not conclusive.

This research is meant to introduce a framework for objective debate surrounding comparative reconstructions and controversial language groupings. The framework can also be used to reason about the comparative method more broadly. For example, the results of the case study reveal that the probability of a random match is mostly determined by the number of borrowings posited, as well the phonotactic complexity of the daughter language. The effect of individual sound changes on reconstruction arbitrariness is measurable but comparatively minor. Future implementations of the framework can be extended to other types of diachronic transformation, e.g. semantic change, morphological change, etc.

January 20<sup>th</sup>, 2023



Figures 1 & 2 : The log probability that a randomly generated wordlist merits a reconstruction of the same size or smaller than the one generated automatically by a simulated annealing algorithm for 74 Austronesian languages and 5 widely accepted Austronesian proto-languages (Figure 1) and 2 putative Ongan proto-languages (Figure 2). PAN = proto-Austronesian; PMP = proto-Malayo-Polynesian; PWMP = proto-West-Malayo-Polynesian; PPh = proto-Philippine; POC = proto-Oceanic; POA = proto-Ongan-Austronesian; PON = proto-Ongan.

#### References:

- Atkinson, Q., & Gray, R. (2005). Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology*, 54(4), 513-526.
- Blevins, Juliette (2007). A long lost sister of proto-Austronesian? Proto-Ongan, mother of Jarawa and Onge of the Andaman Islands. *Oceanic Linguistics*, 46(1), 155-198.
- Blust, Robert (2014). Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics*, 53(2), 300-391.
- Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, 91(1), 194-244.
- Downey, S., Hallmark, B., Cox, M., Norquest, P., & Lansing, J. (2008). Computational Feature-Sensitive Reconstruction of Language Relationships: Developing the ALINE Distance for Comparative Historical Linguistic Reconstruction. *Journal of Quantitative Linguistics*, 15(4), 340-369.
- Kiparsky, P. (2015). New perspectives in historical linguistics. *The Routledge Handbook of Historical Linguistics*.
- Kondrak, G. (2003). Phonetic Alignment and Similarity. *Computers and the Humanities*, 37(3), 273-291.
- Michalove, P., Georg, S., & Ramer, A. (1998). Current Issues In Linguistic Taxonomy. *Annual Review of Anthropology*, 27(1), 451-472.
- Trussel, Stephen & Robert Blust (2010). *Austronesian Comparative Dictionary*.  
<https://www.trussel2.com/ACD/introduction.htm>

Sam Mirelman

Research Fellow, Department for the Study of Antiquity, University of Geneva  
Sam.Mirelmann@unige.ch

Translation as Royal Legitimation: The Concepts of “Source” and “Target” Language in Sumerian-Akkadian Royal Inscriptions from the Old Babylonian Period (2000–1600 BC)

Bilingual Sumerian-Akkadian royal inscriptions from the Old Babylonian period (2000–1600 BC) are amongst the earliest examples of literary translation in world history. The concept and purpose of translation in this early corpus remains largely unexplored, despite its importance to historical linguistics. For example, in most instances translation involves the transference of meaning from an original “source” language to a “target” language. It remains an open debate whether in Sumerian-Akkadian royal inscriptions from the Old Babylonian period, one language is the original “source” and the other language is the “target”.

Sumerian is perhaps the earliest language recorded in writing. The date of its earliest attestation is disputed, due to the fact that early cuneiform writing is almost entirely logographic. Assuming that such early logographic writing does indeed represent Sumerian, the Sumerian language is attested from the late fourth millennium BC. The period during which Sumerian died out as an everyday language is also disputed, although this development almost certainly took place around the late third millennium BC. From the Old Babylonian period (2000–1600 BC) until the late first millennium BC, translation between Sumerian and Akkadian became commonplace. It is in this period that Sumerian became a clear marker of antiquity and prestige, a function which has been compared to the role of Latin in Medieval Europe. By contrast, Akkadian was the vernacular language of Mesopotamia. However, various dialects and registers of Akkadian may be identified, including a literary dialect.

In the 18<sup>th</sup> century BC the king of Babylon, Hammurabi, began to compose (or have composed for him) bilingual Sumerian-Akkadian inscriptions (these inscriptions are “virtual” bilinguals, because the Sumerian and Akkadian versions are written on different physical objects, typically clay nails). This is an important phenomenon, as it is certain in this instance that the Sumerian versions of such inscriptions were newly composed for the king, in a period during which Sumerian was certainly a language known exclusively amongst scholars and priests. The Sumerian in such inscriptions features several markers of language contact with Akkadian. Indeed, one may argue that in such inscriptions the Akkadian is the “source” language and the Sumerian is the “target”. However, if this is indeed the case it would go against the expected direction of translation, from the prestige language of Sumerian to the vernacular Akkadian. Indeed, it would suggest that Sumerian functioned as a means of conveying the impression of being a source language, due to its symbolic role as a marker of antiquity and authority. Thus, in the case of this very early example of literary translation, translation does not seem to function as a means of conveying meaning from one language to another. Instead, Sumerian-Akkadian translation functions as a means of conveying prestige and authority, and as a means of asserting royal legitimation.

In the case of Sumerian-Akkadian translation described above, one may arguably use the term “pseudo-translation” (Touy 1995: 40–52), or the concept of language contact through translation (e.g. Kranich 2014). The specific problems involved in the study of written, as opposed to oral, language contact must also be considered (Adams, Janse and Swain eds. 2002; Lavidas 2022).

**References**

- Adams, J. N., M. Janse and S. Swain (eds.). 2002. *Bilingualism in Ancient Society: Language Contact and the Written Text*. Oxford: Oxford University Press.
- Kranich, S. 2014. “Translations as a Locus of Language Contact.” In J. House (ed.), *Translation: A Multidisciplinary Approach*. London: Palgrave Macmillan, 96–115.
- Lavidas, N. 2022. *The Diachrony of Written Language Contact*. Brill’s Studies in Historical Linguistics 15. Boston/Leiden: Brill.
- Toury, G. 1995. *Descriptive Translation Studies and Beyond* (2nd Edition 2012). Amsterdam: John Benjamins Publishing.

## Changes in the Polish address practices after the Second World War

After the Second World War, as a result of the Yalta and Potsdam conferences, Poland fell into the Soviet sphere of influence. The new political reality had a great impact on social life. The sociocultural changes involved radical changes in discourse practices, and in the national discourse system in general (Duszak, 2006; cf. Fairclough, 1992). They also affected Polish politeness: a transition from 'traditional Polish' politeness to 'more egalitarian' politeness (cf. Huszcza, 2005). There were changes in address practices (cf. Betsch, 2019). The aim of the study is to conduct an analysis of the changes in the Polish address practices which occurred within the four decades after the Second World War.

The theoretical framework used in the analysis is a combination of historical sociopragmatics and the Discourse-Historical Approach. Historical sociopragmatics, focusing on the interaction between specific aspects of social context and particular historical language use (Culpeper, 2011: 4), can account for the role of address forms in relation to the discourse in which they appear and the context of situation. The Discourse-Historical Approach (Wodak, 2002; Wodak et al., 2009), which is one of the main approaches to Critical Discourse Studies, is an interdisciplinary, problem-oriented approach which goes beyond the linguistic dimension and includes the historical, political and sociological dimensions in the analysis and interpretation of a specific discursive event (Wodak and Reisigl, 2015: 583; Reisigl, 2018). As such, the DHA could greatly contribute to historical sociopragmatics (Culpeper, 2010: 87).

The data used in the analysis come from Polish films depicting the then reality in Poland, produced in the years 1945-1989.

## References

- Betsch, M. 2019. Address forms in grammars and textbooks of West and South Slavic languages in the eighteenth and nineteenth century. In: Paternoster, A., Fitzmaurice, S. (Eds.), *Politeness in Nineteenth-Century Europe*. John Benjamins Publishing Company, Amsterdam, pp. 39-74.
- Culpeper, J. 2010. Historical sociopragmatics. In: Jucker, A. H., Taavitsainen, I. (Eds.), *Historical Pragmatics*. Walter de Gruyter, Berlin/New York, pp. 69-94.
- Culpeper, J. 2011. Historical sociopragmatics. In: Culpeper, J. (Ed.), *Historical Sociopragmatics*. John Benjamins, Amsterdam/Philadelphia, pp. 1-8.
- Duszak, A. 2006. Why 'New'Newspeak?: axiological insights into language ideologies and practices in Poland. In: Mar-Molinero, C., Stevenson, P. (Eds.), *Language Ideologies, Policies and Practices*. Palgrave Macmillan, London, pp. 91-103.

Fairclough, N. 1992. *Discourse and Social Change*. Polity Press, Cambridge.

Huszcza, R. 2005. Politeness in Poland: from 'titlemania' to grammaticalised honorifics. In: Hickey, L., Steward, M. (Eds.), *Politeness in Europe. Multilingual Matters*, Clevedon/Buffalo, pp. 218-233.

Reisigl, M. 2018. The discourse-historical approach. In: Flowerdew, J., Richardson, J. E. (Eds.), *The Routledge Handbook of Critical Discourse Studies*. Routledge, London, pp. 44-59.

Wodak, R. 2002. The discourse-historical approach. In: Wodak, R., Meyer, M. (Eds.), *Methods of Critical Discourse Analysis*. Sage Publications, London, pp. 63-94.

Wodak, R., R. De Cillia, M. Reisigl, K. Liebhart, 2009. *The Discursive Construction of National Identity*. Edinburgh University Press, Edinburgh.

Wodak, R., M. Reisigl, 2015. Discourse and racism. In: Tannen, D., Hamilton, H. E., Schiffrin, D. (Eds.), *The Handbook of Discourse Analysis*. Wiley Blackwell, Chichester, pp. 576-596.

## A case of *Verschärfung* in the Swedish dialect from Stora Rågö in Estonia

Ida Västerdal, PhD student, Gothenburg University, Sweden

From the 13<sup>th</sup> century until the 1940's, the northwestern coast of Estonia was inhabited by a Swedish population, speaking diverse Swedish dialects with many peculiar linguistic traits. There, on the small island of Stora Rågö, the local dialect was characterized by the insertion of consonant clusters *-ggv-* or *-ggj-* between vowels: e.g. *fruggver* 'wives' (singular *fru*), *liggjer* 'scythes' (singular *li*). Clearly, these stop insertions are almost identical in form to the insertions caused by the sound law known as *Verschärfung*, appearing in Old Norse and Gothic, as well as in modern Faroese. A study presented in Västerdal (2021) show that there are also close similarities regarding phonetic context and conditions for insertion to take place. It is therefore claimed that the stop insertions in the dialect of Stora Rågö (SR) is a fourth case of Germanic *Verschärfung*.

The term *Verschärfung*, or *Holtzmann's law*, describes a Proto-Germanic sound change found in Old Norse (ON) and Gothic (goth.). The sound change involved glides that developed into *-ggv-/-ggj-* in ON and *-ggw-/-ddj-* in goth., e.g. ON *tryggr*, goth. *triggws* 'trustworthy'; ON *tveggja*, goth. *twaddje* 'two' (Petersen 2002).

A later case of *Verschärfung* occurred in modern Faroese, where it is called *skerping*. In Faroese, stop insertions *-gv-* and *-ggj-* first appeared between two vowels in disyllabic words, e.g. *rógva* 'row', *nýggjur* 'new', but subsequently spread to monosyllabic words (Thráinsson et al. 2004).

Besides the similarities in the form of the stop insertions, there are also significant similarities regarding phonetic context between the stop insertions in SR and Faroese. In the latter, *-gv-* follows after the diphthongs /uu, ou/ and *-ggj-* after /ei, ui, ai, oi/, which are all monophthongized in this context (Thráinsson et al. 2004). In SR, *-ggv-* follows after /u, ʉ/ or a diphthong ending in /ʉ/. Equivalently, *-ggj-* follow after /i/ or a diphthong ending in /i/. All diphthongs preceding insertions are monophthongized before *-ggv-* and *-ggj-* (Västerdal 2021). At times, the similarities between SR and Faroese (Far.) are striking: e.g. SR *sjoggin*, Far. *sjógvín* 'the sea'; SR *biggar*, Far. *bíggjar* 'villages'. The results of the *Verschärfung* development in SR and Faroese seem closely related, yet there are no recordings of any contact between inhabitants of Stora Rågö and the Faroe Islands. Therefore it must be presumed that *Verschärfung* has developed spontaneously in both varieties, just as it has already been suggested that *skerping* in modern Faroese is a similar but separate development from the Old Norse *Verschärfung*.

In this presentation, the SR stop insertions and their phonetic context will be described and compared to the Old Norse, Gothic and Faroese equivalents. There will also be a brief discussion on the possible origins as well as the social and historical context that tentatively contributed to the development of *Verschärfung* in SR.

### References

- Petersen, Hjalmar P. 2002. *Verschärfung* in Old Norse and Gothic. In: *Arkiv för nordisk filologi* 117, pp. 5–26. Lund: Arkiv för nordisk filologi.
- Thráinsson, Höskuldur, Hjalmar P. Petersen, Jógvan í Lon Jacobsen & Zakaris Svabo Hansen. 2004. *Faroese. An overview and reference grammar*. Tórshavn: Føroya Fróðskaparfelag.
- Västerdal, Ida. 2021. *Verschärfung* in the Estonian-Swedish dialect of Stora Rågö. In: *Arkiv för nordisk filologi* 136, pp. 209–230. Lund: Arkiv för nordisk filologi.



## Parallel Phases in the History of French

This talk will present new evidence from the history of French to evaluate the Parallel Phases Hypothesis. A fundamental issue in historical syntax concerns whether certain changes cluster and why. One recent proposal, the Parallel Phases Hypothesis (Poletto 2006; 2014; 2015), suggests that phases – semi-autonomous units of syntactic structure – change in a uniform fashion. Based on evidence from Old Italian, Poletto shows that head and phrasal movement are lost concomitantly at the CP,  $\nu$ P, and DP edges, providing an apparently uniform account for the loss of Verb Second, scrambling, and a number of word-order operations licensed in the Old Italian nominal expression.

This talk will assess whether the predictions of the hypothesis extend to a closely related language, French; this is a necessary first step in assessing the validity of the hypothesis as, if it holds up crosslinguistically, the hypothesis has the potential to shed light on why, in a number of languages, several major word-order patterns change quite radically in a short period of time, yielding modern languages with significantly distinct typological footprints to their early counterparts (see, for example, Walkden 2014 on early Germanic, Wolfe 2018 on early Romance, and Willis 2007 on early Celtic).

Examination of Old and Middle French data from the *Base de Français Médiéval*, supplemented with a hand-annotated corpus of Renaissance and Classical French texts reveals that – as per the predictions of the hypothesis – Verb Second-related movement to the CP, object scrambling to the  $\nu$ P edge, as well as widespread Adjective-Noun orderings and N-to-D movement to the DP edge are all lost in tandem as fully productive operations between 1400 and 1525. Moreover, the corpus data suggest that Poletto's hypothesis extends further to the discourse-pragmatic and syntactic features associated with the phrasal elements undergoing movement: while a wide range of objects can be focalised or topicalised to both the CP and  $\nu$ P edge in Early Old French, by 1225 such objects show a strong tendency to either be focal or discourse 'old'. Moreover, in Middle French texts, focal objects at both the CP and  $\nu$ P edges are heavily restricted, but QPs remain stable in both contexts. More generally, we suggest that at the CP,  $\nu$ P, and DP edge, reanalysis of a maximally general movement operation proceeds first as a movement operation restricted by discourse-pragmatic status, which in turn is reanalysed as an operation where only categories bearing particular categorial features may move (e.g. [+Focus] > [+Q]). This progression is observable in the loss of Verb Second, the loss of scrambling, and the restrictions on adjective movement to the left periphery of the extended nominal expression.

The talk will conclude with the proposal that the Parallel Phases Hypothesis can be subsumed within the Input Generalization Principle proposed by Roberts (2007). Under this approach, the acquirer – unless receiving strong evidence to the contrary – assigns similar featural makeup to syntactic heads perceived as forming a uniform class. In the case under examination the class in question is that of phase-heads.

Giusti, G. 2015. *Nominal syntax at the interfaces: a comparative analysis of languages with articles*. Newcastle: Cambridge Scholars • Poletto, C. 2006. Old Italian Scrambling: The Low Left Periphery of the Clause. In P. T. Brandt & E. Fuss (eds), *Form, Structure, and Grammar: A Festschrift Presented to Günther Grewendorf on Occasion of His 60th Birthday*, 209–229. Berlin: Akademie • Poletto, C. 2014. *Word Order in Old Italian*. Oxford: OUP • Roberts, I. 2007. *Diachronic syntax*. Oxford: OUP • Walkden, G. 2014. *Syntactic reconstruction and Proto-Germanic*. Oxford: OUP • Willis, D. 2007. Historical syntax. In *The syntax of Welsh*, 286–337. Cambridge: CUP • Wolfe, S. 2018. *Verb Second in Medieval Romance*. Oxford: OUP.



## Abstract for ICHL 26 in Heidelberg, 4–8 Sept. 2023

Kerstin Roth (TU Dresden)

### Rhetoric, stylistic and argumentative strategies of German language female authors in the 17<sup>th</sup> century

In the history of the German language, the 17<sup>th</sup> century is known for its so called *Sprachgesellschaften* and their aim to evolve and practice a well-formulated and written German language. This can be seen as an unofficial start for the standardization of the language, followed by the production of dictionaries and other literature that had the aim of showing how one could and should write in “good” German.

One question, however, is if and how women participated in this evolution? Who were they and why did they use the German language as they did?

This presentation focusses on different female authors from the 17<sup>th</sup> century who explain in their texts why they wrote in German, whom they addressed and which impact on the evolution of the German language this could have had.

Rhetoric, stylistic and argumentative strategies will build the center of this presentation. This is why the focus will lie on so called *Paratexts*, yet poems or letters in which the vernacular German is discussed will also be treated. *Paratexts* are texts that introduce a work to the addressees. In these texts (introductions, poems, etc.) the authors explain their aims. We like to show that the choice of German is a choice in the direction of an interested, but not (always) academically educated audience. The use of the vernacular German can also be seen as an approach to evoke intimacy between the authors and their public. The choice of the language is a choice between close and distant (vernacular vs. f. ex. Latin) or between a non-academic and an academic world.

Examples for this approach are texts from the scientists Maria Cunitz and Maria Sibylla Merian along with the poets Sophia Elisabeth Brunner and Johanne Charlotte Unzer. This study is based on results of cooperative research with researchers from the field of early modern literature. We intend to show that interdisciplinary work is needed to fully understand the language and literature of the 17<sup>th</sup> century.

### References

Corinna Dziudzia & Sonja Klimek (2022) (Hgg.): *Gelehrte Frauen der Frühaufklärung. Einsame ‚Wunderthiere‘ oder vernetzte Akteurinnen?* Wiesbaden.

Claudia Opitz & Ulrike Weckel (1998) (Hgg.): *Ordnung, Politik und Geselligkeit der Geschlechter im 18. Jahrhundert.* Göttingen.

Kerstin Roth & Katharina Worms (t. b. in print): Selbstkonzeptionen von Autorinnen der frühen Neuzeit in ihren Paratexten. Sophia Elisabeth Brenner, Maria Cunitz, Maria Sibylla Merian und Johanne Charlotte Unzer. In: Peter Heßelmann (Hg.): *Simpliciana 2023.*

Katharina Worms (t. b. in print): „die Freyheit, von Liebe und Weine zu scherzen“: Rhetorik der Emanzipation in Johanne Charlotte Unzers anakreontischer Lyrik. In: Annina Klappert (Hg.): *Textfiguren der Emanzipation. Autorinnen des europäischen 17. und 18. Jahrhunderts wieder (und wieder) lesen.* Hannover.

## Early New High German preposed adverbial clauses: integration and discourse functions

Barthe Bloom, Universität Potsdam, bloom@uni-potsdam.de

The study investigates the use of preposed adverbial clauses in Early New High German (henceforth: ENHG) narratives and evaluates whether the choice between integration, resumption and juxtaposition is determined by local and global discourse factors.

In ENHG, preposed adverbial clauses are positioned either i) before an argument, as in (1), ii) adjacent to a resumptive element, like in (2) and (3); or iii) directly adjacent to the finite verb of their host sentence, exemplified in (4) (König and van der Auwera, 1988).

- (1) Da der swartz ritter das gewar wart, **er** greiff yne mit dem helm  
lit. 'And when the black knight noticed that, HE GRABBED him by the helmet.' (Pontus)
- (2) ob es in die warheit nit gesaht het, **so** muß es sterben  
lit. 'If it hadn't told him the truth, SO MUST IT die.' (Tristrant)
- (3) vnd da alle ding bereyt waren **da** gieng sie zû dem Peter  
lit. 'And when all things were ready, THEN WENT SHE to Peter.' (Magelone)
- (4) Do der künig das verstûnde sprach er zum graffen  
lit. 'When the king understood that, SPOKE HE to the count.' (Fortunatus)

The four constructions illustrated above represent different degrees of integration, reaching from simple juxtaposition (1) to resumption (2) and (3) to full integration (4). This continuum is thought to be reflective of the diachronic development of the position of adverbial clauses (König and van der Auwera, 1988; Axel, 2004; Lötscher, 2006), i.e., from juxtaposed via resumption to complete integration.

This diachronic scenario predicts that at least one of the resumptive constructions was paradigmatically related to the integrating construction (Zehentner, 2019) at the time the use of the integrating construction was taking over, i.e., in ENHG (Axel, 2004). Moreover, it is expected that the juxtaposition construction has been ousted by resumption at this time, and found its own niche.

To test these predictions, a data set of 1500 preposed adverbial clauses from seven ENHG narratives – Pontus, Melusine, Wigalois, Wilhelm, Tristrant, Fortunatus, and Magelone – is investigated. The data indicate that the two resumptive constructions are associated with distinct discourse functions, and that resumption with *so* does not show systematic similarity to the integration construction.

First, the constructions are found to differ with regard to local discourse functions. For example, concessives and conditionals are normally resumed by *so*, whereas the temporal adverbials occur primarily in the integrating construction or with *da*. Juxtaposition has not developed its own niche here.

Second, the constructions' global discourse functions are compared. Specifically, it is hypothesized that the resumptives are associated with different narrative speed (Genette, 1983; Packard, 2008). The *so*-construction is relatively frequent in direct speech segments compared to the other three patterns. As such, it is associated with scenes, which are characterized by isochrony. In contrast, *da*-resumption likely condenses more story time into less narration, since the adverbial clauses that occur in this construction often summarize earlier events.

How this relates to the use of integration and juxtaposition presents the final part of the puzzle. In order to support the diachronic development, the juxtaposition construction should have a specialized function, while the integrating construction is expected to show a tight similarity to one of the resumptive constructions.

## Primary sources

**Fortunatus** (1509). Schmitt, L. E. and R. Noll-Wiemann (eds.) 1974. *Fortunatus: Von Fortunato und seynem Seckel auch Wünschhütlein*. Hildesheim/New York: Georg Olms Verlag.

**Magelone** (1535). Müller, J.-D. 1990. *Romane des 15. und 16. Jahrhunderts. Nach den Erstdrucken mit sämtlichen Holzschnitten*. Frankfurt am Main: Deutscher Klassiker Verlag.

**Melusine** (1474). Müller, J.-D. 1990. *Romane des 15. und 16. Jahrhunderts. Nach den Erstdrucken mit sämtlichen Holzschnitten*. Frankfurt am Main: Deutscher Klassiker Verlag.

**Pontus** (1450-1500). Schneider, K. 1961. *Pontus und Sidonia: in der Verdeutschung eines Ungenannten aus dem 15. Jahrhundert / kritisch hrsg. von Karin Schneider*. Berlin: E. Schmidt.

**Tristrant** (1484). Elsner, H. (ed.). 1989. *Tristan und Isolde (Augsburg bei Antonius Sorg 1484) mit einem Nachw. von Helga Elsner*. Hildesheim: Georg Olms Verlag.

**Wigalois** (1519). Melzer, H. 1973. *Wigalois / mit eine Vorw. von Helmut Melzer*. Hildesheim: Georg Olms Verlag.

**Wilhelm** (1481 & 1491). Podleiszek, F. 1936. *Volksbücher von Weltweite und Abenteuerlust / bearbeitet von Dr. Franz Podleiszek*. Leipzig: Verlag von Philipp Reclam jun.

## References

Axel, K. (2004). The syntactic integration of preposed adverbial clauses on the German left periphery: A diachronic perspective. In H. Lohnstein and S. Trissler (Eds.), *The Syntax and Semantics of the Left Periphery*, pp. 23–58. Berlin: Mouton de Gruyter.

Genette, G. (1983). *Narrative discourse: An essay in method*, Volume 25. Ithaca: Cornell University Press.

König, E. and J. van der Auwera (1988). Clause integration in German and Dutch conditionals, concessive conditionals, and concessives. In J. Haiman and S. Thompson (Eds.), *Clause combining in grammar and discourse*, pp. 101–133. Amsterdam: John Benjamins.

Lötscher, A. (2006). Linksperiphere Adverbialsätze in der Geschichte des Deutschen. Pragmatische Aspekte eines grammatischen Wandels. *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 127(3), 347–376.

Packard, S. (2008). Two-dimensional narrative velocity. *Journal of Literary Semantics* 37(1), 55–67.

Zehentner, E. (2019). *Competition in Language Change: The Rise of the English Dative Alternation*. Berlin: De Gruyter Mouton

### Adverbs ending in *-(l)ig* ‘-ly’ and *-(l)igt* ‘-ly’ in Danish

In Modern Standard Danish, adverbs based on adjectives typically end in a *-t*, cf. *gør det omhyggeligt!* ‘do it carefully’, where the adverb *omhyggeligt* ‘carefully’ is based on the adjective *omhyggelig* ‘careful’. In the 19<sup>th</sup> century, this was not so; back then, adverbs based on adjectives would typically occur without such a *-t*, e.g. *hun indøvede stykket omhyggelig* ‘she practiced the play carefully’.

On the expression side, the adverbs without a final *-t* as well as the ones ending in a *-t* coincide with adjectival gender forms (e.g. *omhyggelig-Ø* ‘careful-C’ vs. *omhyggelig-t* ‘careful-N’). Therefore, they have often been analyzed as adjectives used as adverbs (e.g. Mikkelsen 1911; Diderichsen 1946). However, following Skautrup (1947, 1953) and Brink (2018), we consider them adverbs derived from adjectives (Schack & Jensen, *subm.*; Jensen & Schack, *subm.*). Historically, adjective-based adverbs with and without final *-t* are preceded by other derivational forms, most notably ones with the suffix *-e* (e.g. *listige* ‘cunningly’, *grumme* ‘badly, ugly’) and – if the adjective itself ends with the derivational suffix *-(l)ig* – by forms with the suffix *-(l)igen* (e.g. *retteligen* ‘properly’, *lystigen* ‘merrily’).

During the 17<sup>th</sup> and 18<sup>th</sup> centuries, adverbs ending in *-e* and in *-(l)igen* recede and give way to adverbial forms with no particular expression marker (e.g. *omhyggelig*) and later still to the adverbial forms ending in *-t* (e.g. *omhyggeligt*).

This diachronic outline concerns all adverbs based on adjectives, e.g. derived adjectives with the suffixes *-bar*, *-som*, (*mærkbar* ‘tangible’, *langsom* ‘slow’), compositional adjectives with the highly productive elements *-fuld*, *-løs* (*sorgfuld* ‘mournful’, *tankeløs* ‘thoughtless’), as well as simplex adjectives without a suffix (*smuk* ‘beautiful’, *streng* ‘strict’). However, in the literature on Danish, most attention has been devoted to adverbs ending in *-(l)ig*.

The adjectival suffix *-lig* is a cognate to the English adverbial suffix *-ly*. In the course of history, a number of adjectives historically with other endings have joined the adverbs ending in *-lig* (cf. Falk & Torp 1900; Skautrup 1947, 1953). In grammars of Modern Danish, adverbs ending in *-lig(t)* and *-igt(t)* are traditionally treated as one issue, and we follow this tradition.

In this paper, we present a study of the change from *-(l)ig* to *-(l)igt* in recent history. The study is based on two corpora, both representing written language. One corpus consists of literature and letters from the 19<sup>th</sup> century; the other of texts from the year 2017.

The study documents how forms without *-t* are replaced by forms with a final *-t*. The study also reveals that the replacement does not take place at an even pace in all adverbial functions. Manner adverbs without *-t* are significantly more frequent in the 19<sup>th</sup> century than forms without *-t*; in 2017, manner adverbs exclusively occur with *-t*. Temporal adverbs change from forms with *-t* being much more frequent than forms without *-t* in the 19<sup>th</sup> century to the opposite situation in 2017. As regards adverbs of degree, in the 19<sup>th</sup> century, they almost exclusively occur without *-t*; in 2017, the ratio of forms with and without *-t* is fifty-fifty.

This pattern supports ideas of adverbs being a morpho-syntactic more diverse lexical class than traditionally assumed.

**References**

- Brink, L. (2018): Adjektiv eller adverbium? *Nordiske Studier I Leksikografi* 14, 56–62.
- Diderichsen, P. (1946): *Elementær Dansk Grammatik*. København: Gyldendal.
- Falk, H. & A. Torp (1900): *Dansk-norskens syntax i historisk fremstilling*. Kristiania: Aschehoug & Co.
- Jensen, E.S. & J. Schack (subm.): Ubøjede adverbialer uden og med *-t* og historien bag. In: *Ny forskning i grammatik* 30.
- Mikkelsen, K. (1911): *Dansk ordføjninglære med sproghistoriske tillæg*. København: Lehmann & Stages Forlag.
- Schack, J. & E.S. Jensen (subm.): Uden og med *-t*. Adverbialer dannet af adjektiver på *-ig* og *-lig*. In: Schoonderbeek mfl. (red.). *19. Møde om Udforskningen af Dansk Sprog (MUDS)*.
- Skautrup, P. (1947): *Det danske sprogs historie. Bind 2*. København: Gyldendal.
- Skautrup, P. (1953): *Det danske sprogs historie. Bind 3*. København: Gyldendal.

### Adjectival typology in four ancient Indo-European languages

Scholars usually agree that the part of speech (PoS) system of Proto-Indo-European (PIE) is similar to that in Latin, Greek, Sanskrit and Hittite and, in each case, it is based on three major classes: nouns, verbs and adjectives or schematically [N, A, V] (implicitly, Nussbaum 2022). Still, adjectives show the same endings as nouns, comparative morphology is lacking in Hittite, Tocharian, Armenian and Albanian, and most PIE adjectives show the “recent” inflection in *\*-e/o-*. Thus, many scholars argued that the adjective is a recent category (Wackernagel 2009: 466, Lehmann 1974: 208, Comrie 1997: 101): that is, PIE or, at least, pre-PIE could be a language “without adjectives” or “with only two major lexical classes”, nouns and verbs (i.e. [N, V]). However, this idea was further developed into two opposite directions: some scholars claimed that quality concepts were merged with nouns in PIE and, thus, PIE was a language with “noun-like adjectives” or a language of type [(NA) V] (Balles 2006, 2009); others claimed that quality concepts were merged with verbs in PIE and, thus, PIE was a language with “verb-like adjectives” or a language of type [N (AV)] (Alfieri 2016, 2021; Bozzone 2016). In the former scenario, the adjective class arose from appositional nouns added to the feminine motion: i.e. *\*[noun]-ieh<sub>2</sub>/ih<sub>2</sub>-Agr*, where Agr means “agreement” (Brugmann 1888: 420; Fritz & Meier-Brügger 2020: 225); in the latter, it arose from derived nouns built on verbal roots of (nearly) quality meaning added to the feminine motion: i.e. *\*[verb-NM]-ieh<sub>2</sub>/ih<sub>2</sub>-Agr*, where NM means “nominalizer”. The talk aims to show that, if a functional-typological definition of the adjective is accepted, the latter view is more plausible than to the former.

The starting point for this research is Croft’s definition of PoS (2001: 67ff.). In his view, the “adjective” is not a language-specific, formal class, but a zone of cognitive space defined in terms of semantics and syntax. In practice, the “adjective” is defined as the most typical Quality Modifier construction that is found in any language. This definition is applied to 4 ancient IE languages, namely Rig-Vedic Sanskrit, Homeric Greek, Classical Latin and Hittite. Thus, a sample of texts is gathered for each language – 51 hymns of the *Rig-Veda*, the first book of Homer’s *Iliad* and *Odyssey*, Sallust’s *De coniuratione Catilinae*, and an anthology of Hittite texts –, all the Quality Modifiers in each text are collected (between 800 and 1000 in each language) and their internal structure is analysed. The results are the following (the data on RV Sanskrit and Homeric Greek come from Alfieri 2016, 2021, Alfieri & Gasbarra 2021, while the data on Latin and Hittite are presented here for the first time).

Basically, the same six construction types code the ADJECTIVE slot in each language, namely (RV Sanskrit is quoted as an example for all IE languages): 1) the simple adjective or [adjective]-Agr, i.e. Skt. *kr̥ṣṇá-* ‘black’; 2) the deverbal adjective or [verb-NM]-Agr, i.e. Skt. *tap-ú-* ‘hot’, *mah-ánt-* ‘big’ < *tap-* ‘become/make hot’, *mah-* ‘be/make big’; 3) the denominative adjective or [noun-ADJ]-Agr, i.e. Skt. *pítṛ-īya-* ‘paternal’ < *pítár-* ‘father’, where ADJ means “adjectivalizer”; 4) the prepositional adjective or [preposition-ADJ]-Agr, i.e. Skt. *paramá-* ‘most distant’ < *párā* ‘away’; 5) the prefixed adjective that is, a nominal stem attached to a prefix or PRE-[...]-Agr, i.e. Skt. *su-víra-* ‘having good heroes’ < *víra-* ‘hero’; 6) the compound adjective or [...]<sub>N</sub>-[...]<sub>N</sub>-Agr, i.e. Skt. *híraṇya-pāṇi-* ‘golden-palmed’. However, the frequency of each construction type is far different from a language to another (Tab. 1):

	RV Skt.	Hom. Gk.	Hittite	Latin
[adjective]-Agr	7.6%	48.1%	65.7%	80.4%
[verb-NM]-Agr	45.9%	12.9%	25.7%	6.0%
[noun-ADJ]-Agr	10.4%	10.5%	0.8%	6.5%
[preposition]-Agr	2.1%	0.3%	7.9%	0.5%
Pre-[...]-Agr	14.4%	12.1%	0%	6.5%
[...] <sub>N</sub> -[...] <sub>N</sub> -Agr	19.9%	15.6%	0%	0%



Tab. 1 shows that the most frequent Quality Modifier construction is [adjective]-Agr in Latin, Homeric Greek and Hittite, which are *specialized* languages of type [N, A, V] in Hengeveld's terms (1992), while it is [verb-NM]-Agr in RV Sanskrit, which is a *rigid* language with verb-like adjectives in Hengeveld's terms and falls into type [N (AV)]. The easiest way to interpret the difference between RV Sanskrit and the remaining IE languages is to claim that PIE was a language “without” adjectives or a language of type [N (AV)], in which quality concepts were coded “verbally” and the most typical “adjective” was \*[verb-NM]-Agr; this PoS system is preserved in RV Sanskrit bar minor changes, while a previously neglected typological change of type [N (AV)] → [N, A, V] came about in the prehistory of Greek, Latin and Hittite, although the change occurred independently and with a different timing in each branch of the IE family.

- Alfieri, Luca. 2021. Parts of speech comparative concepts and Indo-European linguistics. Alfieri L., Arcodia G. & Ramat P. (eds.), *Linguistic categories, language description and linguistic typology*, 313-366. Amsterdam-Philadelphia: John Benjamins.
- , 2016. The typological definition of the (apparently historical) notion of root. *Archivio Glottologico Italiano* 102(1): 129-169.
- & Gasbarra, Valentina. 2021. The adjective class in Homeric Greek and the parts of speech change in the Indo-European languages. L. Biondi, F. Dedé & A. Scala (eds.), *Change in Grammar. Triggers, Paths and Outcomes*, 7-26. Alessandria: Ed. dell'Orso.
- Balles, Irene. 2009. Zu den i-stämmigen Adjektiven des Lateinischen. Lühr R. & Ziegler S. (eds.): *Protolanguage and Prehistory*. Akten der XII. Fachtagung der Indogermanischen Gesellschaft, Krakau, 11.-15. Oktober 2004. Wiesbaden 2009.
- , 2006. *Die altindische Cvi-Konstruktion: Form, Funktion, Ursprung*. Bremen: Hempen.
- Bozzone, Chiara. 2016. The Origin of the Caland System and the Typology of Adjectives. *Indo-European Linguistics* 4: 15-52.
- Brugmann, Karl & Delbrück, Bertold. 1887-1916. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. Vol. II.1: Brugmann K. 1888 [1906<sup>2</sup>]. *Lehre von den Wortformen und ihrem Gebrauch. Das Nomen*. Strassburg: Trübner.
- Comrie, Bernard. 1997. La famiglia linguistica indoeuropea: prospettive genetiche e tipologiche. A. Giacalone Ramat & P. Ramat (eds.), *Le lingue indoeuropee*, 95-122. Bologna: Il Mulino.
- Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.
- Fritz, Matthias & Meier-Brügger, Michael. 2021<sup>10</sup>. *Indogermanische Sprachwissenschaft*. Berlin-Boston: Mouton De Gruyter.
- Hengeveld, Kees. 1992. *Non-verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton.
- Lehmann, Winfried P. 1974. *Proto-Indo-European Syntax*. Austin: University of Texas Press.
- Nussbaum, Alan J. 2022. Derivational properties of “adjectival roots” (expanded handout). Malzahn M., Fellner H.A. & Illés T. (eds.), *Zurück zur Wurzeln. Struktur, Semantik und Funktion der Wurzel im Indogermanisch*. Akten der 15. Fachtagung der Indogermanischen Gesellschaft vom 13. bis 16. September 2016 in Wien, 205-224. Wiesbaden: Reichert.
- Wackernagel, Jakob. 2009. *Lectures on Syntax: with Special Reference to Greek, Latin, and Germanic*. Translated and edited with notes and bibliography by David Langslow. Oxford: Oxford University Press [Id. 1920-1924<sup>1</sup>. *Vorlesungen über Syntax mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch*. 2 vols. Basel: Birkhäuser].



## The periodization of the Pre-Classical French through the study of *nennil* and *non* in grammars, remarks and treatises (15<sup>th</sup>-17<sup>th</sup> centuries)

### Key words

Pre-Classical French, Periodization, Historical Pragmatics, Negation markers, Remarqueurs

### Abstract

The periodization of Pre-classical French is relatively recent compared to the other generally accepted periods for French (Medieval, Classical and Modern French). Based on morphosyntactic criteria, Combettes (2003) isolates the period 1500-1650 (Preclassical French), where the language does not function as in the 15th century (Middle French), nor as in the late 17th century (Classical French). Combettes and Marchello-Nizia (2010) revisit the periodization of Pre-Classical French, taking 1550 as the initial boundary and 1650 as the final boundary. More recent studies (Ayes-Bennett & Caron 2016; Amatuzzi *et al.* 2020) have analysed the accuracy of the commonly accepted final boundary (1650), identifying a break around 1620-1630 which implies the culmination or the beginning of a number of morphosyntactic changes. Only in a decade, the French language underwent an upheaval that can be identified in oral corpora and which was also perceived in French language treatises, remarks and grammars of the time, as demonstrated by Ayres-Bennett and Caron (2016).

The aim of this presentation is to contribute to the reflection on the periodization of Pre-Classical French by analysing the evolution of the use of the disagreement markers *nennil* and *non* between 1450 and 1700. *Nennil*, the disagreement marker complementary to the agreement marker *oïl/ouï* of Medieval French, disappears during this period in favour of *non*, which will be paired with *oui* from Classical French onwards. After a brief presentation of the distribution and the pragmatic and interactional functioning of the markers *nennil*, *non* and *non* + verb substitute (*non feray*, *non est*), I will analyse the comments made about these markers in the 33 publications collected in the *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue (XIVe-XVIIe s.)* (Colombat & Fournier 2011).

Remarks such as those made by Vaugelas (1647) and Mauger (1659) will help me to date the transition from the use of medieval *nennil* to modern *non*, especially with regard to the complementarity with *oui* and the language level.

« [*&c* eft bien placé] apres *oïy*, & *non*, comme *Oïy Madame, Non Madame, il ne fe voit rien &c.* »

[*&c* is well placed after *ouï* and *non*, as in *Oïy Madame, Non Madame, il ne fe voit rien &c*]  
(Claude Favre de Vaugelas, 1647, *Remarques sur la langue française utiles à ceux qui veulent bien parler et bien écrire*)

« *Nenny*, eft vn mot vulgaire; & jamais on ne s'en fert en écriuant; tant il eft de bas alloy. »

[*nenny* no is a vulgar word and is only used in writing because it has little value]

(Claude Mauger, 1659, *Grammaire française avec des augmentations*)

My results will support the conclusions of other authors on the date of transition from Pre-Classical to Classical French between 1620 and 1630. The study of pragmatic phenomena is innovative in the sense that periodization research generally analyses phonetic and/or morphosyntactic changes, at least for French. Also, this study of a corpus of grammars, remarks and treatises of the French language of the period complements other quantitative works.

## References

- Amatuzzi, Antonella, *et al.* 2020. « Changement linguistique et périodisation du français (pré)classique: deux études de cas à partir des corpus du RCFC », *Journal of French Language Studies* 30, 301-326.
- Ayres-Bennett, Wendy & Philippe Caron, 2016. « Periodization, Translation, Prescription, and the Emergence of Classical French », *Transactions of the Philological Society* 114(1), 339-390.
- Colombat, Bernard & Fournier, Jean-Marie (dir.). 2011. *Grand Corpus des grammaires françaises, des remarques et des traités sur la langue française*, Paris, Classiques Garnier Numérique.
- Combettes, Bernard (ed.). 2003. *Évolution et variation en français préclassique : études de syntaxe*, Paris, Champion.
- Combettes, Bernard & Christiane Marchello-Nizia, 2010. « La périodisation en linguistique historique : le cas du français préclassique », in : Combettes, Bernard. *et al.* (eds.), *Le Changement en français*, Berne, Peter Lang, 129-41.

## Just a bystander? Semantic change in the English simple tenses

Juliette Kayenbergh, Hendrik De Smet

As the English progressive construction *BE Ving* grammaticalized from a stylistically employed device without clear grammatical meaning into an obligatory grammatical aspect marker expressing ongoingness, incompleteness and dynamicity (e.g. Kranich 2010), it entered into a paradigmatic relation with the simple form. Accordingly, Bybee (1994) argues that when the progressive grammaticalized, the zero-marked simple form became obligatorily associated with the expression of non-progressive meaning – a process she refers to as “the grammaticization of zero”. However, the timing of obligatorification of *BE Ving* for the expression of progressive meaning remains unclear. Indeed, Kranich (2010: 165) points out that the history of the simple tense forms has been largely ignored in the literature, leaving a major blind-spot in our understanding of the development of aspectual marking in English. The aim of our research is to investigate the impact of the rise of *BE Ving* on the meanings of the simple forms in order to test Bybee’s (1994) hypothesis.

The hypothesis is tested against the history of the simple forms and their alternates in the *BE Ving* construction for three verbs: *sleep* (mostly expressing a ‘undirected activity’ (Croft 2012)), *drink* (mostly expressing an ‘activity’ or ‘accomplishment’) and *stand* (mostly expressing a ‘state’) (Vendler 1957). Data are collected from the EEBO, CLMET and BNC corpora, covering Early Modern, Late Modern and Present-day English. The data consist of random samples capped at 250 instances per verb per period.

As a first step, we establish how the interpretations of simple forms, which range from imperfective (including habitual and progressive) to perfective, arise pragmatically from other elements in the context, such as adverbial modifiers and relations established to other profiled events. For example, both the adverbial *now* in (1) and the framing subordinate clause in (2) encourage a progressive reading, while in (3) the sequence of events and the explicit endpoint to the *sleep* event jointly trigger a perfective interpretation.

- (1) Beyond the fear of ling’ring punishment, Aspasia now, within her lover’s arms, Securely **sleeps**[.] (CLMET, 1726-49)
- (2) Vnto this Pertynax therefore wente Letus, and Electus, with a few of theyr complices aboute mydnyghte, whyles all men **slepte**. (EEBO, 1556)
- (3) And so [I] wente to my bedde full ryght Where I **slepte** styll and merely tyll foure of the clocke after mydnyght Than vp I rose by the candell light (EEBO, 1506)

As a next step, we exploit the presence of these pragmatic interpretative cues as proxies to the likely intended aspectual reading of any given instance. This way, we can gain insight into any semantic change in the simple tense and its timing. If Bybee’s (1994) hypothesis holds, the simple tense is expected to shift away from contexts favouring or imposing a progressive interpretation as *BE Ving* gains in frequency.

At the time of writing, the analysis has been completed for *sleep*. Preliminary results show striking differences between present and past tense simple forms. In the present tense, the simple form over time increasingly refers to habitual situations, whereas the share of episodic progressive meaning declines. Although this is in line with expectations, we still find PDE examples that express ongoingness in the present-tense simple form, suggesting that contrary to received wisdom the simple tense does not obligatorily express non-progressive meaning even in PDE. The past tense casts even more doubt on the initial hypothesis, revealing an essentially stable coexistence of the simple form and *BE Ving* in the expression of progressive meaning. Results thus paint a far more complex picture than the literature suggests. If anything, obligatorification of the non-progressive interpretation in simple forms is a slow and contextually sensitive process.

## References

- Bybee, Joan L. 1994. Asymmetries in tense and aspect systems. In William Pagliuca (ed.), *Current Issues in Linguistic Theory*, vol. 109, 235-254. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/cilt.109.02byb>.
- Croft, William. 2012. *Verbs: aspect and causal structure* (Oxford Linguistics). Oxford [England] ; New York: Oxford University Press.
- Kranich, Svenja. 2010. *The Progressive in Modern English: a Corpus-Based Study of Grammaticalization and Related Changes* (Language and Computers 72). Amsterdam New York, NY: Rodopi.
- Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66(2). 143-160. <https://doi.org/10.2307/2182371>.

### **On semantic change in grammaticalization: Why it is never metaphoric**

We understand metaphor as a “conscious or voluntary shift in a word’s meaning” (cp. Matisoff 1991: 384).

Researchers like Heine et al. (1991) and Matisoff (1991) argued that metaphor is a fundamental mechanism of semantic change in grammaticalization. This idea has however been contested from the very start. For instance, Bybee et al. (1994: 24-25) argued that if metaphor would be a relevant mechanism under grammaticalization, we would expect to see semantic leaps, but these are rarely if ever found. Semantic change in grammaticalization tends to progress gradually involving small-scale reanalyses (in the sense of Hansen 2021).

The marginal position of metaphor in grammaticalization is in sharp contrast to the fact that metaphor is often considered to be one of the most important mechanisms in lexical change (e.g., Geeraerts 2015: 422; Blank 1997: 157). This remarkable difference between semantic change of lexical and of grammatical elements has to our knowledge received little attention, and it calls for an explanation.

Juge (2007) argued that metaphor plays a marginal role in grammaticalization because (as opposed to for instance pragmatic inferencing) it presupposes a high degree of speaker awareness of the metaphorically employed unit. Awareness is incompatible with the mechanisms of grammaticalization, since these are subconscious, he argues. We do not disagree, but Juge’s account begs the question why the mechanisms of grammaticalization are necessarily subconscious – or, at least, evade awareness – and why this is not the case with lexical semantic change. This question represents a fundamental challenge to linguistic change as both lexical semantic change and grammaticalization of lexical items have the same point of departure, namely lexical units.

In this paper, we offer an answer to the question. Our account takes its point of departure in a revised version of Boye & Harder’s (2012: 21) definition of grammaticalization. According to the revised definition, grammaticalization consists in the conventionalization of attentionally backgrounded status. This crucially entails that the input to grammaticalization is attentionally backgrounded, and we argue that this restriction on the input is what makes grammaticalization incompatible with metaphorically employed expressions. From a speaker as well as a hearer perspective, metaphor demands attention or awareness (cp. Juge 2007: 45): for a source concept to structure a target concept, speaker and hearer must pay attention to the internal structure of the concepts at play. Furthermore, novel metaphors will typically attract attention in that they are atypical ways of expressing oneself for which there arguably must be a reason based on the relevance maxim.

Our argument thus looks as follows: 1. A precondition for being grammaticalized is contextually being attentionally backgrounded. 2. Metaphor draws attention to the metaphorically used expressions. 3. Therefore, the metaphorical unit cannot lose the competition for attention and be conventionalized as having ancillary status.

We furthermore argue that based on our account, we can explain why also metonymy in a narrow sense (that is, as a conscious and voluntary process) seems to be marginal in grammaticalization. One of the consequences of our argument is thus that we should re-evaluate lumping pragmatic inferencing and metonymy narrowly defined together under the heading metonymy.

We argue that other approaches to grammaticalization are incapable to account for the same facts.

### **References**

Blank A. (1997). *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*. Tübingen: M. Niemeyer Verlag.

- Boye, K. & P. Harder (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, 88(1), 1-44.
- Bybee, J., R. Perkins & W. Pagliuca (1994). *The Evolution of Grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Geeraerts, D. (2015). How words and vocabularies change. In Taylor, J. (ed.) *The Oxford Handbook of the Word*. Oxford: Oxford University Press, 416-430.
- Hansen, M-B. M. (2021). In defense of a pragmatic view of reanalysis. *Journal of Historical Syntax*, 5, 1-34.
- Juge, M. 2007. Metaphor and teleology do not drive grammaticalization. In J. Salmons & Sh. Dubenion-Smith (eds.), *Historical Linguistics 2005: Selected papers from the 17th International Conference on Historical Linguistics, Madison, Wisconsin, 31 July - 5 August 2005*. Amsterdam: John Benjamins, 33-48.
- Matisoff, J. A. 1991. "Areal and Universal Dimensions of Grammatization in Lahu". In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization II*. Amsterdam: John Benjamins, 383-453.

## Predicative possession in the languages of the Ethiopian area

### Abstract

In Ethiopian languages, *have*-possession and existential constructions of possession are the predominant predicative possession types represented by the two dominant languages namely Oromo and Amharic, respectively. Previous research has shown that the *have*-possession is a Cushitic feature (cf. Thomason 1983:216) whereas the existential construction of possession has no clear historical records for its origin in Semitic (Rubin 2005:57). However, the Amharic-type of existential constructions of possession is common in the whole Ethiosemitic. The goal of this paper is two-fold: (1) showing further predominance of the existential constructions of possession over the *have*-possession, and (2) giving an overview of the notional characteristics of predicative possession in the languages of the Ethiopian area. It depends on the data from informants, online corpora, native-speaker intuition being bilingual in Oromo and Amharic, and published grammars. While the *have*-possession is employed by a few Cushitic and Omotic languages, the existential constructions of possession have been widespread in the area.

On the other hand, the Ethiopian languages show more commonality with the notions coded by the predicative possession constructions. Not only the possessive notions that include the prototypical ownership and the non-prototypical kinship, part-whole, physical, abstract, etc. relations, but also the non-possessive domains such as attribution, experience, and location are expressed by the predicative possession. The paper shows that the Ethiosemitic influence tends to be accounted for predominance of the existential constructions of possession based on the synchronic analysis because such construction is observable in all Ethiosemitic that might have spread to others through Amharic, the lingua franca of the country. The distribution of these predicative possession constructions implies that the predicate structures and the associated notional domains probably add important information to the Ethiopian Language Area.

**Key words:** *have-possession, Ethiopian Language Area, notions, predicative possession, existential constructions*

### Reference

- Thomason, S. G. (1983). Genetic Relationship and the Case of Ma'a (Mbugu). *Studies in African Linguistics*, 14(2), 195–231.
- Rubin, A. D. (2005). *Studies in Semitic Grammatication*. Winona Lake, Indiana: Eisenbrauns.



### The long and winding road of the Danish evidential *vel* - from epistemic modality via concessivity to evidentiality

The meaning of the Modern Danish modal particle *vel* (cognate *well*) has been analysed as expressing an appeal to the addressee to verify the truth of the proposition (cp. Davidsen-Nielsen 1996: 286). For its Modern Swedish cognate *väl*, the same meaning has recently been rendered in terms of *engagement* (Bergqvist 2020: 471).

However, these analyses only address the contextualized meaning. In utterances like (1) and (2), it would be absurd to hold that the speaker appeals to the addressee for verification:

- (1) Jeg fik ca. 30 sting og det gjorde **vel** ondt i en måneds tid.  
'I got approx. 30 stitches and it hurt **VEL** for a month approximately'
- (2) Jeg har **vel** sikkert gjort forsøg på at falde ind i dialekten.  
'I have **VEL** certainly tried to adapt to the dialect'

*Vel* does not necessarily seek confirmation but marks the proposition as put forth without having evidence for it, i.e., as a guess or conjecture. Thus, *vel* is so to speak grammaticalized 'gut feeling' (distinct from epistemic possibility since it readily combines with epistemic necessity markers, cp. (2)). Based on this meaning the contextual function to seek confirmation emerges: If the addressee can (dis-)confirm the proposition, a cooperative move would be to do so. In my paper, I will present an account of the development of this meaning.

Based on corpus studies, starting its development a millennium ago, the semantic path of evidential *vel* appears to be reconstructable as follows:

- (3) 'good, in satisfactory manner > 'easily' > epistemic necessity > concessive > evidential conjecture

With this semantic path, the development of *vel* can be accounted for based on small-scale hearer-driven reanalyses in terms of Hansen (2021).

The development of concessives out of epistemic modality markers is fairly uncontroversial (cp. e.g., Sweetser 1990: 70-72; Bybee et al. 1994: 226-227, Squartini 2012). The central part of the development is the transition from concessive to evidential meaning (cp. Thurgood 1986: 217-218). In line with Aikhenvald (2004: 276; 2011: 610) who argues that one source of evidential meanings are so-called evidential strategies whereby non-evidential markers are used evidentially in particular contexts, I argue that the conjecture meaning emerges based on contextual meanings in concessive sequences as rendered in Couper-Kuhlen & Thompson (2000: 382): The second, conceding, move in these sequences is a reiteration and an acknowledgment of the validity of others' statements. As pointed out by Squartini (2012: 2123), the conceded proposition is thus reported. Therefore, such sequences provide latent, contextually given evidential meanings where the addressee puts forth a proposition as true relying on others' assessment and hence, crucially for *vel*, without having supportive evidence for the assessment herself. In a pragmatically driven reanalysis (Hansen 2021), this second aspect ('lack of evidence') is then conventionalized as the coded meaning of *vel*.

The exact nature of the evidential meaning can only be understood in taking into account the paradigmatic oppositions of *vel* diachronically. Being grammaticalized, *vel* enters into opposition with the other evidential modal particles *nok* and *vist* which respectively express that the evidence is only subjectively or intersubjectively available (cp. Hansen & Heltoft 2011: 1058-60). Thus, only through paradigmaticization (Lehmann [1982]2015: 174; Nørgård-Sørensen et al. 2011; Diewald & Smirnova 2012), *vel* semantically finds its current position in the paradigm of evidential modal particles.

## Reference

- Aikhenvald, A. Y. (2004) *Evidentiality*. Oxford: Oxford University Press, Incorporated.
- Aikhenvald, A. Y. (2011). The grammaticalization of evidentiality. In: Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 605–13. Oxford: Oxford University Press, 605-613.
- Bergqvist, H. (2020). Swedish modal particles as markers of engagement: Evidence from distribution and frequency. *Folia Linguistica*, 54(2), 469-496.
- Bybee, J. L., R. Perkins, W. Pagliuca (1994). *The Evolution of Grammar. Tense, Aspect and Modality in the Languages of the World*. Chicago/London: The University of Chicago Press.
- Couper-Kuhlen, E. & S. Thompson (2000) Concessive patterns in conversation. In: Couper-Kuhlen, E., Kortmann, B. (Eds.), *Cause, Condition, Concession, Contrast*. Berlin and New York: Mouton de Gruyter, 381-410.
- Davidson-Nielsen, N. (1996). Discourse particles in Danish. In: Engberg-Pedersen E., Fortescue M., Harder, P., Jakobsen, L.F., Heltoft L. (eds) 1996 *Content, Expression and Structure. Studies in Danish Functional Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 283-314.
- Gabriele D. & E. Smirnova. 2012. "Paradigmatic integration: the fourth stage in an expanded grammaticalization scenario." In Davidse, K., T. Breban, L. Brems and T. Mortelmans (eds.) [SLCS 130] *Grammaticalization and Language Change. New reflections*. Amsterdam: John Benjamins Publishing Company, 111-133.
- Hansen, E., & Heltoft, L. (2011). *Grammatik over det danske sprog*. Odense: Syddansk Universitetsforlag.
- Hansen, M-B. M. (2021). In defense of a pragmatic view of reanalysis. *Journal of Historical Syntax*, 5, 1-34.
- Lehmann, C. ([1982]2015). *Thoughts on grammaticalization. 3rd edition 2015* (Classics in Linguistics 1). Berlin: Language Science Press.
- Nørgård-Sørensen, J., L. Heltoft, & L. Schøsler (2011). *Connecting grammaticalisation*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Squartini, M. (2012). Evidentiality in interaction: The concessive use of the Italian Future between grammar and discourse. *Journal of Pragmatics*, 44(15), 2116-2128.
- Sweetser, E. (1990). *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Thurgood, G. (1986). The nature and origins of the Akha evidentials system. In W. Chafe and J. Nichols (eds.) *Evidentiality: the linguistic coding of epistemology*. Norwood, NJ: Ablex Publishing Corporation, 214-222.

## **Grammaticalization as conventionalization of discursively secondary status: Isolating what is unique to grammaticalization, and deconstructing the lexical-grammatical continuum**

While there is wide agreement about approximately what grammaticalization is, a more profound understanding faces two major challenges. One is that the presupposed distinction between grammatical and lexical is itself hard to get a grip on (e.g. Boye & Harder 2012: 1-6). The other challenge is that even with a theoretically anchored and well-defined distinction between grammatical and lexical, it is not clear that grammaticalization is a distinct type of language change rather than an epiphenomenon (e.g. Campbell 2001: 151; see also Section 4).

Boye & Harder (2012) offered a solution to the first challenge, arguing for an understanding of grammatical elements as defined by conventionalized discursively secondary status (roughly, attentional background status). However, they circumvented the second challenge. Rather than attempting to define grammaticalization as a diachronic phenomenon, they defined it in terms of its result, namely as “the diachronic change that gives rise to linguistic expressions that are by convention ancillary and as such discursively secondary” (Boye & Harder 2012: 22). This definition is problematic as it includes all changes under grammaticalization as long as the output is a grammatical (i.e. secondary by convention) element.

This paper has two aims: Firstly, it proposes a definition of grammaticalization which is still based on the understanding of grammatical elements in Boye & Harder (2012), but which targets the nature of grammaticalization as a diachronic phenomenon: Grammaticalization is the conventionalization of discursively secondary status.

Secondly, it discusses important implications of the proposed definition: 1) Grammaticalization basically applies to meaning. What is conventionalized as discursively secondary is basically a meaning; a sign is only ‘discursively secondary by convention’ by virtue of its meaning. 2) Grammaticalization is a special case of a well-known type of change, viz. conventionalization. 3) Grammaticalization covers a restricted range of phenomena. For instance, changes like phonological reduction and semantic bleaching that are often associated with grammaticalization are external to grammaticalization. 4) Grammaticalization is not a type of overall development like, for instance *go* > *gonna*, but a small and well-defined part of such larger changes. 5) Grammaticalization is a gradual process to the extent that conventionalization is a gradual process. However, 6) the lexical-grammatical continuum must be deconstructed into other continua, including the conventionalization continuum, a splitting continuum (e.g. Heine & Reh 1984: 57), and a discourse prominence continuum. 7) Degrammaticalization (understood as a process in which grammatical elements give rise to lexical ones) is infrequent because it requires attention to discursively secondary (i.e. attentionally backgrounded) elements.

The paper is theoretical in intent, but invokes both standard examples of grammaticalization (e.g. the grammaticalization of future tense auxiliaries) and non-standard examples (e.g. what Jespersen 1922 called ‘secretion’) to illustrate its points.

### **References**

- Boye, K. & P. Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88.1. 1–44.
- Campbell, L. 2001. What’s wrong with grammaticalization? *Language Sciences* 23.2–3. 113–161.
- Jespersen, O. 1922. *Language: Its nature, development and origin*. London: Allen & Unwin.

### **R Deletion in Brazilian Portuguese: Diachronic and Synchronic Evidence for Lexical Diffusion**

This paper provides diachronic and synchronic evidence that R deletion in coda position in Brazilian Portuguese (BP) is being implemented through Lexical Diffusion (Wang; Chen, 1977; Oliveira, 1997). R deletion in BP can occur in both nouns and verbs, either in middle or final position: *amor* > *amô*-- 'love', *cantar* > *cantá*-- 'to sing', *perfume* > *pe—fume* 'perfume', *perseguir* > *pe—seguir* 'to chase'. Final R in verbs is almost categorically deleted (95,4%, according to Oliveira (1983)). The first document to provide evidence for this deletion process is the Appendix Probi, a document from the 3<sup>rd</sup> century that provides a list of words that people were pronouncing or writing incorrectly according to Classic Latin. Item number 149 from that list reads “*persica non pessica*” ‘peach’. There is then evidence that in the 3<sup>rd</sup> century there were already words being written or pronounced without the R. Also, this document shows there was social stigma against that. Gil Vicente, a Portuguese playwright from the 15<sup>th</sup> century, also provides examples of the stigma that R deletion would carry. In his plays, when lower class people are speaking, they do not pronounce final Rs in verbs. Leite de Vasconcelos (1970) and Chaves de Mello (1971) also show examples of that. Oliveira (1997, p. 33), using data from interviews with native speakers, analyzes this process through the lenses of Lexical Diffusion and proposes that “all phonological changes are guided through Lexical Diffusion.” In addition to that, he proposes that social class is a strong factor determining R deletion. Other factors such as stress, word class and informal speech also contribute to that. Our study of this process investigated final R deletion in nouns only. Our goal was to understand which social or internal factors were favoring the R deletion or retention. We also aimed at understanding if the change was proceeding through Lexical Diffusion. Interviews with 30 native speakers were conducted, and a total of 2,606 occurrences of final R in nouns was gathered. Overall, our study showed deletion rate of 12%, and factors such as stress, word class, age group, and social class were relevant to the process. The data also showed evidence for Lexical Diffusion, as the change is spreading through the lexicon according to token frequency, with frequent words changing first. This claim is supported by Bybee (1995). In sum, our paper makes a connection with Historical Linguistics as it analyzes R deletion diachronically, but it also presents results of a synchronic experiment.

## Diachronic pathways of definite articles distribution

Definite articles are commonly suggested as an example of an unstable grammatical feature (Croft 1996, Wichmann & Holman 2009, Dediu and Levinson 2012, Greenhill et al. 2017). Frajzyngier (2008: 18) describes grammatical instability as a system “[i]n which given functional distinctions are neutralized in significantly extensive environments”, a situation which may cause articles to emerge and subsequently erode from a system. However, the properties promoting article emergence, loss, and persistence on a global scale remain unclear. Different linguistic features have been posited as the predictors of their distribution – gender marking and nominal classifiers, verb typology, topic and focus marking, verb-final word order, flexible word order, switch reference (Givón 1978, 1983), ergative alignment (Du Bois 1987, 2017), case marking (Comrie 1989), and perfective aspect (Leiss 2000). Empirical studies identified case, verb-final word order, flexible subject order, and ergative alignment as potentially robust predictors of the absence of definite articles (Evers 2020). These relationships have been explained in the light of 1) efficiency (case and flexible argument order can serve to mark definiteness and languages with these features might be less likely to gain another grammatical means with the similar function) and 2) the information status in languages with ergative alignment and verb-final languages (A arguments are prototypically definite, and clause-final arguments are prototypically indefinite). However, the relationships between the four predictors are complex: case marking is typically found in verb-final and flexible word order languages (e.g. Levshina 2021), and many ergative languages have case marking. An explicit causal approach is thus required to reveal the diachronic processes behind the cross-linguistic distribution of definite articles.

Here we disentangle the causal relationships underlying the synchronic distribution of definite articles on a global sample of 1232 languages matched for typological information from Grambank (The Grambank Consortium 2022) and the global EDGE tree (Bouckaert et al. 2022). We use logistic regression with phylogenetic correction within phylogenetic path analysis in *phylopath* package (van der Bijl 2018) to first establish the existing causal relationships between the predictors themselves and then to evaluate how and whether case, word order, and ergativity explain the distribution of definite articles.

We find that verb-final languages are more likely to maintain and gain case marking, and languages with case marking are likely to develop flexible ordering of core arguments. We use this causal model as the basis for several competing models predicting definite articles distribution. Out of the tested predictors of definite articles, only verb-final word order proves robust on a global sample: articles are more likely to be absent or disappear in verb-final word order languages. This suggests that word order dependent preferences for negotiating information status primarily shape the evolution of definite articles. The previously observed patterns of complementary distribution between 1) definite articles and 2) case, ergative alignment, and flexible word order are not supported by causal inference. Instead, these suggested relationships might be products of complex interactions between case, word order, and ergativity.

## References

- Bijl, Wouter van der. 2018. *phylopath*: Easy phylogenetic path analysis in R. *PeerJ*. PeerJ Inc. 6. e4718.
- Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate E Jones & Quentin Atkinson. 2022. Global language diversification is linked to socio-ecology and threat status [Preprint]. <https://osf.io/preprints/socarxiv/f8tr6/>.
- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80(1). 1–28.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.
- Croft, William. 2002. *Typology and universals*. Cambridge University Press.
- Dediu, Dan & Stephen C Levinson. 2012. Abstract profiles of structural stability point to universal tendencies, family-specific factors, and ancient connections between languages. *PLoS ONE* 7. e45198.
- Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.
- Du Bois, John W. 2017. Ergativity in discourse and grammar. In Jessica Coon, Diane Massam & Lisa D. Travis (eds.), *The Oxford Handbook of Ergativity*, 23–58. Oxford: Oxford University Press.
- Evers, Stephanie. 2020. *To 'the' or not to 'the': Cross-linguistic correlations between existing morphosyntax and the emergence of definite articles*. Buffalo, NY: The State University of New York at Buffalo PhD Thesis.
- Frajzyngier, Zygmunt. 2008. Grammaticalization, typology and semantics. In María José López-Couso & Elena Seoane (ed.), 61–102. Amsterdam: John Benjamins.
- Givón, Talmy. 1978. Definiteness and referentiality. *Universals of Human Language* 4. 291–330.
- Givón, Talmy. 1983. Topic continuity in discourse: the functional domain of switch reference. In John Haiman & Pamela Munro (eds.), *Switch Reference and Universal Grammar*, vol. 2, 51–82. Amsterdam: Benjamins.
- Greenhill, Simon J, Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C Levinson & Russell D Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences*. National Acad Sciences 114(42). E8822–E8829.
- Leiss, Elisabeth. 2000. Artikel und Aspekt. Die grammatischen Muster von Definitheit. In *Studia Linguistica Germanica* 55. Berlin: Walter de Gruyter.
- Levshina, Natalia. 2021. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 2791.
- The Grambank Consortium (eds.). 2022. Grambank 1.0. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://grambank.clld.org>.
- Wichmann, Søren & Eric W Holman. 2009. *Assessing temporal stability for linguistic typological features*. Munich: LINCOM Europa.



As Syntax Interfaces with Information Structure: Old Icelandic Non-Canonical Scrambled Orders

This talk reports some findings from my research on word order types, derived by Scrambling and found with double object constructions with one non-finite verb in the corpus of *Íslendinga Sögur* (1998). The account of Scrambling is launched in an integrated Minimalist syntactic framework (Eythórsson 1995; Haugan 2001; Thráinsson 2001; Epstein & Seely 2006; Wallenberg 2009; Cheng & Corver 2013; Engels & Vikner 2014) and provides convergent support to the claim that Scrambling is an optional displacement operation raising internal Arguments and Adjuncts out of their source positions into phrasally-adjoined targets in the left periphery of vP. Remarkably, while Conservation of C-Command (Wallenberg 2009: 132) bans movement of constituents across c-commanding functional heads, information-structural and semantic factors step in to instigate movement out of the VP domain. The given-before-new-information packaging strategy is the default option with canonical scrambled orders (Lambrecht 1994; Choi 1999; Hinterhölzl & Petrova 2009; Meurman-Solin, Lopez-Couso & Los 2012; Bech & Eide 2014; Biberauer & Walkden 2015), whereby the main findings can be boiled down to the following: pronouns scramble almost obligatorily; definite objects scramble quite freely; indefinite objects scramble only rarely; heavier objects usually remain in the unscrambled position.

Scrambled orders attested in non-canonical  $V_{fin}$ -IO(Dat)- $V_{non-fin}$ -DO(Acc),  $V_{fin}$ -DO(Acc)- $V_{non-fin}$ -IO(Dat),  $V_{fin}$ -IO(Dat)-DO(Acc)- $V_{non-fin}$  and  $V_{fin}$ -DO(Acc)-IO(Dat)- $V_{non-fin}$  constructions are in focus here, with “non-canonical” being reserved for scrambled orders that exhibit some syntactic structure deviations or information structure ambiguities, are of low frequency, and seem to be ‘non-optimal’ in terms of violating a syntactic constraint or overriding a principle of information structure, cf:

1) ... að hann mun þeim manni gefa dóttur sína, hvat manna sem hann er, ef að bana yrði orminum, (Ragn 819)

‘... that he will give his daughter in marriage to that man, whatever kind of man that be, if he would slay the dragon,’

2) ... ef þeir hefðu hlaupið frá mannum en þeir mættu grið gefa honum, (Heið 1387)

‘... if they had run away from the man so they might give him mercy,’

3) Nú skal veita svör þínu máli, að eg vil öllum yður grið gefa skipverjum. (Laxd 1564)

‘Now I shall give answers to your request that I will give mercy to all of you, shipmen.’

4) Vil eg það ráð þér gefa sem hverjum öðrum að hann leiti sér þess ráðuneytis ... (Fljót 723)

‘I’ll give that counsel to you as to anyone else that he should seek for himself this solution ...’

In 1) the in-situ DO *dóttur sína* escapes the definiteness effect while the scrambled IO *þeim manni* obeys definiteness but is arguably focussed. In 2) the ex-situ bare nominal DO *grið* is indefinite and the in-situ IO *honum* is an anaphoric pronoun that fails to obey the anti-focality effect. In 3) the non-constituent string *öllum yður* is scrambled along with the indefinite DO *grið*, as the nominal element of the IO *skipverjum* stays in-situ. Left-dislocation of the DO *það ráð* and the non-constituent pronominal element of the IO *þér* in 4) is sensitive to definiteness and anti-focality but the second element of the coordinate IO *sem hverjum öðrum* stays in situ.

I will argue that Scrambling in O(ld) Ice(landic) occurs on the Syntax-Information Structure Interface, and, by corollary that it can be thought of as a type of information packaging syntactic device. The studied interactions of word order and information structure can be explained as interface interactions that license scrambled orders on the basis of their syntactic, information-structural, and semantic properties. Base-generated word order and Scrambling of pronouns favour unmarked interpretation, Scrambling definite phrases is a less marked option than Scrambling indefinite phrases, Scrambling focussed phrases is more marked than Scrambling unfocussed phrases.

Among the conclusions are the following: Non-constituents may also be targeted by Scrambling in OIce; OIce Scrambling may evoke non-canonical information-structural effects: ex-situ XPs (and non-constituents) may be construed in terms of contrastive, non-presupposed, emphatic, focussed interpretations; A possible mismatch between the locus of default sentence focus and the position of the allegedly focalized expression can trigger Scrambling in OIce.



## References:

- Bech, Kristin, and Kristine Gunn Eide, eds. 2014. *Information Structure and Syntactic Change in Germanic and Romance languages*. Amsterdam, Philadelphia: John Benjamins.
- Biberauer, Theresa, and George Walkden, eds. 2015. *Syntax over Time: Lexical, Morphological and Information-Structural Interactions*. Oxford: Oxford University Press.
- Cheng, Lisa Lai-Shen, and Norbert Corver, eds. 2013. *Diagnosing Syntax*. Oxford: Oxford University Press.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, California: CSLI Publications.
- Engels, Eva and Sten Vikner. 2014. *Scandinavian Object Shift and Optimality Theory*. Houndmills, Basingstoke, Hampshire and New York: Palgrave Macmillan Ltd.
- Epstein, Samuel D. and T. Daniel Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.
- Eythórsson, Thórhallur. *Verbal Syntax in the Early Germanic Languages*. PhD diss., Cornell University, 1995.
- Haugan, Jens. *Old Norse Word Order and Information Structure*. PhD diss., Norwegian University of Science and Technology, 2001.
- Hinterhölzl, Roland, and Svetlana Petrova, eds. 2009. *Information Structure and Language Change. New Approaches to Word Order Variation in Germanic*. Berlin: Mouton de Gruyter.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Meurman-Solin, Anneli, Maria Jose Lopez-Couso, and Bettelou Los, eds. 2012. *Information Structure and Syntactic Change in the History of English*. Oxford: Oxford University Press.
- Thráinsson, Höskuldur. 2001. Object Shift and Scrambling. In *The Handbook of Contemporary Syntactic Theory*, edited by Marc Baltin and Chris Collins, 148-212. Oxford: Blackwell Publishing.
- Wallenberg, Joel. *Antisymmetry and the Conservation of C-command: Scrambling and Phrase Structure in Synchronic and Diachronic Perspective*. PhD diss., University of Pennsylvania, 2009.

## Obscenity as a Window into Slavic Linguistic History

Slavic obscenity has received some diachronic attention (e.g. Hamp 1968, Isačenko 1964, Uspenskij 1996), but its study has been marginalized in Slavic countries. Trubačev's (1964-73) translation and annotation of Vasmer (1950-58) omitted the core obscenities *ebat'* (older *eti*) 'fuck', *pizda* 'cunt' and *xuj* 'cock, prick', and *bljad'* 'whore'. Stavyc'ka (2008) provides some comparative Slavic material, but there has been no attempt to map out the patterns of retention and innovation in Slavic obscenity, and these display significant phenomena.

Slavic obscenity is remarkably conservative from an Indo-European standpoint, arguably more so than any other branch. The root (j)eb- is cognate with forms indicating copulation in Greek, Sanskrit, and Sogdian, but evidence from Luwian and Tokharian show that the original etymon began as a euphemism that was obscenified by contamination in dialectal Indo-European. While universally preserved in Slavic, the verbal root is in retreat as the *vox propria* for 'fuck' in most of West Slavic, especially Sorbian, Czech, and Kashubian, less so in Polish, and not in Slovak. Similarly, Common Slavic \**peizdā* began as a euphemism on the Indo-European dialectal level, with cognates in Albanian and Nuristani (Hamp 1968, Mallory and Adams 1997). Bulgarian and Kashubian have specific developments, and Sorbian shows competition. An old isogloss separates South Slavic *kur* 'cock' from North Slavic *xuj* 'prick', with Bulgarian being transitional.

Since obscenities are subject to euphemization, which euphemisms become contaminated and displace earlier obscenities, leaving the older obscenities to either become obsolete or shift meaning, it would appear that in Slavic, obscenities underwent a process of decontamination and recontamination. Evidence for this is suggested in old South Slavic and East Slavic texts (Vinodolski Zakon, Old East Slavic Birchbark Letters). Alternatively, evidence from the Birchbark Letters could suggest that in the Middle Ages the situation, at least for *pizda* and *eb-* was like that in, e.g., modern Romani, where the single lexical item, *mindž*, can be translated 'vagina/vulva' or 'cunt' depending on the context, e.g. medical or invective.

By contrast it appears that, unlike terms for sexual intercourse and female genitalia, the male member was subject to the usual processes of obscenified euphemism replacement (with items that correspond lexically to English 'prick' and 'cock' in the North and South, respectively) at some time during the break-up of Common Slavic, i.e. the early Middle Ages. Based on the evidence of the Vinodolski Zakon and the Birchbark Letters, it can thus be argued that Medieval Slavic *eti* and *pizda* were obscene only contextually and did not become restricted to obscenity until the early modern period (evidence argues for the same treatment of Russian *bljad'*). Still, the Common Slavic for the male member may already have been restricted to obscenity, whence its replacement by the obscenification of new euphemisms, *xuj* and *kur*, after the break-up of Common Slavic.

The history of core obscenities in Slavic thus illustrates the importance of studying obscenities in general. In the case of Slavic: 1) The remarkable conservatism of two out of three Slavic core obscenities suggests either decontamination and re-obscenification or a period characteristic of languages for which context determines obscenity; 2) Those languages with the strongest Germanic contact are most likely to innovate obscenities, consistent with facts of German (Stavyc'ka 2008); 3) The male member was treated differently from both female genitalia and sexual intercourse and in this regard Bulgarian shows connections to East Slavic, pointing to Macedonian's closer connection to the rest of South Slavic vis-à-vis Bulgarian.

## References

- Buck, Carl Darling. 1949. *A dictionary of selected synonyms in the principal Indo-European Languages*. Chicago: University of Chicago Press.
- Hamp, Eric P. 1968. Albanian *pidh*: Slavic *peizd'ǎ*. *International Journal of Slavic Linguistics and Poetics* 11.25-26.
- Isačenko A. V. 1964. Un juron russe du XVIe siècle. *Lingua Viget: Commentationes slavicae in honorem V. Kiparsky*, ed. by I. Vahros & K. Kahla, 68-70. Helsinki: Suomalaisen Kirjallisuuden Kirjapaino.
- Mallory, James P. and Douglas Q. Adams, 1997. *Encyclopedia of Indo-European Culture*. London/Chicago: Fitzroy Dearborn.
- Stavyc'ka, Lesja. 2008. *Ukraïnska mova bez tabu*. Kyiv: Krytyka.
- Trubačev, O.N., ed. and transl. 1964-1973. Maks Fasmer, *Ėtimologičeskij slovar's russkogo jazyka*. Moscow: Progress.
- Uspenskij, Boris. 1996. Mifologičeskij aspekt russkoj èkspressivnoj frazeologii. *Anti-mir russkoj kul'tury*, ed. by N. Bogomolov, 9-107. Moscow: Ladomir.
- Vasmer, Max. 1950-1958. *Russisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

### Anchoring patterns in emerging complement clauses in Slavic

Complement structures are considered to consist of a complement-taking predicate (CTP) and an element indicating the subordinated/complement status of the clause it introduces (Schmidtke-Bode 2014, Kehayov/Boye 2016). However, in actual discourse we regularly come across cases that do not allow an easy categorization of these elements (and their relation). This holds in particular for diachronic data. The challenge arises from the fact that the categorization of complementizer and CTP in a potential complementation structure is subject to a certain circularity. In standard cases this does not present a problem because we have strong paradigmatic and preference-based evidence for the functional loads of the respective units. In less obvious cases, however, it becomes more difficult. This is particularly true for the Slavic languages, which possess very few clear-cut indicators of subordination (for preliminary findings cf. Sonnenhauser 2021 on Sln. *naj*, Wiemer 2021; forthcoming).

In our presentation we aim to identify the impact of various types of predicates on the inference of cross-clausal discourse relations and the eventual emergence of structures with clausal complements in Slavic. Special attention is given to predicates which usually do not join the list of ‘classic’ CTPs. We focus on the relation of clausal structures containing optative / directive expressions such as Po. *niech*, Ru. *pust’*, Sln. *naj* to clauses preceding them. The categorization of a clause-initial element as an untypical complementizer might be based on the presence of a typical CTP in the preceding clause, whereas the identification of an untypical CTP might be triggered by a subsequent typical complementizer. Cf. example (1) with a directive-optative marker (DIR) like Russ. *pust’* and the regular CTP *poželat’* ‘wish’ and (2) with the atypical CTP *nie dość* ‘not enough’ and a standard complementizer (COMP) like Pol. *że*:

- (1) Russian (RNC; T.N. Tkačenko, 1995)  
*Ja poželala [ej skorej opraviti’sja] i [pust’ segodnjašnee DTP budet ej v nazidanie] [...]*  
‘I wished.CTP [her a fast recovery] and [may.DIR today’s accident be a lesson for her]’
- (2) Polish (PNC)  
*Nie dość, [że jest to bardzo niewygodne dla samych obywateli], to jeszcze urąga samej powadze sądu okręgowego, [...]*  
‘Not only.QUASI-CTP [that.COMP this is very inconvenient for the citizens themselves], it also hurts the very authority of the district court’

To avoid the pitfalls of circular argumentation we refrain from assumptions about the categorial affiliation of the elements in question. We explore the relation between predicates that have the potential of serving as anchors for cross-clausal discourse continuation and clause-initial elements that introduce non-first clauses in a chain of clauses, regardless of their mainstream classification. This includes an account of combinations of COMP elements (Po. *że*, Ru. *čto*, Sln. *da*) and DIR elements (Po. *niech*, Ru. *pust’*, Sln. *naj*), adjacent as well as non-adjacent.

Using sets of random samples from two periods (17<sup>th</sup>-19<sup>th</sup> c. vs. contemporary stage; see References) we aim to identify possible correlations between the (type of) predicate of a preceding and a following clause containing a DIR-element, with respect to the following criteria concerning the anchoring element: (i) lemma, (ii) semantic class, (iii) PoS, (iv) inflectional form. On this basis, we explore whether the identified anchoring patterns bear on structural features within the subsequent clause, such as the (non)-initial appearance of the DIR-element and the distance between this element and the finite verb. Initial position of DIR-elements has been considered favorable for them to function as complementizers, non-initial as favorable for AUX status.

### References

- Kehayov, Petar & Kasper Boye. 2016. Complementizer semantics – an introduction. In: Kasper Boye & Petar Kehayov (eds.), *Complementizer Semantics in European Languages*, 1–11. Berlin & Boston: De Gruyter Mouton.
- Schmidtke-Bode, Karsten. 2014. *Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization*. Jena (unpubl. PhD thesis).
- Sonnenhauser, Barbara. 2021. Slovene *naj*: an (emerging) clausal complementizer? In: Wiemer, Björn & Barbara Sonnenhauser (eds.). *Clausal complementation in South Slavic*. Berlin, Boston: De Gruyter Mouton, 443–475.
- Wiemer, Björn. 2021. A general template of clausal complementation and its application to South Slavic: theoretical premises, typological background, empirical issues. In: Wiemer, Björn & Barbara Sonnenhauser (eds.). *Clausal complementation in South Slavic*. Berlin, Boston: De Gruyter Mouton, 29–159.
- Wiemer, Björn (forthcoming). Between analytical mood and clause-initial particles – on the diagnostics of subordination for (emergent) complementizers. *Zeitschrift für Slawistik* 68–2 (2023).

### Corpora

Polish

<http://nkjp.pl/>

[https://korba.edu.pl/query\\_corpus/](https://korba.edu.pl/query_corpus/)

Russian

<https://ruscorpora.ru/new/>

Slovene

<http://nl.ijs.si/imp/>

<http://www.gigafida.net/>

## The anticausative alternation in Italian and Spanish: a historical corpus-based perspective

The causative alternation is the grammatical alternation whereby languages express externally caused (CAUSAL) vs. spontaneously occurring (NONCAUSAL) events. Typological studies have shown that the alternation is encoded via different patterns across languages, based on whether the causal event, the noncausal one, both, or neither are overtly marked (see Tubino-Blanco 2020). A cross-linguistically peculiarity of several languages of Europe is that, in this area, one finds an exceptional preference towards explicit marking of the noncausal event (Nichols et al. 2004), which goes under the name of anticausativization (Haspelmath 1987; Zúñiga & Kittilä 2019: 41–53). In many of these languages, as is common cross-linguistically (Bahrt 2021), anticausativization is formally expressed by the same marker of reflexivity. Examples from Italian and Spanish are given in (1) and (2). Notably, alongside the anticausative pattern, in these languages one also finds labile verb pairs, as in (1b-c).

Regarding the alternation between anticausative and labile patterns, it has been shown that individual verbs may trigger anticausativization more frequently than others, to the effect that verb meanings can be ranked based on their likelihood to receive anticausative marking (Haspelmath 1993, 2016). To explain these preferences, scholars have resorted to either verb semantics or frequency effects. Semantics-based accounts appeal to notions such as spontaneity (Haspelmath 1987) and claim that verbs lexicalizing events less likely to occur spontaneously more frequently trigger anticausative marking (Haspelmath 1993: 106). In frequency-based approaches, marking asymmetries mirror frequency asymmetries, based on the assumption that higher usage frequency items are more predictable and favor shorter coding (Haspelmath 2021). This means that verbs that more routinely occur in noncausal contexts are less likely to occur in the anticausative pattern (Haspelmath et al. 2014). These findings have also been confirmed by corpus data from French and Spanish (Heidinger 2019).

A yet unresolved issue with frequency-based approaches is that “we cannot measure the earlier frequencies” (Haspelmath 2016: 601). To overcome this issue, this work aims at testing whether claims advanced by the frequency-based approach are borne out by historical data and what the interplay is between semantic and frequency effects.

To do so, we undertake a contrastive exploration of the distribution of anticausative marking in historical corpora of Spanish and Italian. The choice of Italian and Spanish is also due to the fact that systematic diachronic studies on anticausatives in these languages are relatively few (e.g., Portilla 2007; Cennamo 2012, 2021). Data for Italian come from the MIDIA corpus whereas data for Spanish come from the CDH corpus, both including texts ranging from the 13<sup>th</sup> to the 20<sup>th</sup> century. In particular, based on the 20 verb meaning pairs list in Haspelmath et al. (2014), reported in (3), we have sampled (max) 500 tokens of each of the corresponding Italian and Spanish verbs. For each verb, we extract data regarding their token frequency in causal vs. noncausal contexts and on the encoding of the alternation (anticausative vs. labile). This will allow us to explore whether the observed frequency of the verbs under investigation in causal and noncausal contexts changes over time and what this reveals about asymmetries in the encoding of the anticausative alternation in Italian and Spanish. In addition, we also consider a number of additional factors that have been claimed, on synchronic grounds, to play a role in the choice of the anticausative vs. labile pattern, including semantic features of the subject (animacy, control) and aspectual properties of the verbs (Cennamo 2012, 2021; de Benito Moreno 2022: Chap. 4). By resorting to regression modelling techniques, we assess whether the role of these factors is stable across time or differences can be pinpointed at specific language stages and across languages.

The results of this work will contribute to showing how anticausativization comes about, and will offer the first in-depth empirical assessment of how anticausative markers spread through the verbal lexicon across time.



**Examples**

- (1) a. *Il ragazzo bruciò il cibo* CAUSAL  
 the boy burn.PST.3SG the food  
 ‘The boy burnt the food.’
- b. *Il cibo si bruciò* NONCAUSAL-ANTICAUSATIVE  
 the food REFL burn.PST.3SG
- c. *Il cibo bruciò* NONCAUSAL-LABILE  
 the food burn.PST.3SG  
 ‘The food burnt.’
- (2) a. *Juan rompió la mesa*  
 J. break.PST.3SG the table  
 ‘Juan broke the table.’
- b. *La mesa se rompió*  
 the table REFL break.PST.3SG  
 ‘The table broke.’
- (3) *boil, freeze, dry, wake up, go out/put out (fire), sink, melt, stop, turn, burn, fill, rise/raise, improve, rock, connect, gather, open, break, close, split*

**References**

- Bahrt, Nicklas N. 2021. *Voice syncretism*. Berlin: Language Science Press.
- Benito Moreno, Carlota de. 2022. *The Middle Voice and Connected Constructions in Ibero-Romance*. Amsterdam & Philadelphia: John Benjamins.
- Cennamo, Michela. 2012. Aspectual Constraints on the (Anti)Causative Alternation in Old Italian. *Transactions of the Philological Society* 110(3). 394–421.
- Cennamo, Michela. 2021. Anticausatives and lability in Italian and French. In Silvia Luraghi & Elisa Roma (eds.), *Valency over time*, 265–304. Berlin & New York: de Gruyter.
- Haspelmath, Martin. 1987. *Transitivity alternations of the anticausative type*. Köln: Institut für Sprachwissenschaft.
- Haspelmath, Martin. 1993. More on the typology of the inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (eds.), *Causatives and Transitivity*, 87–120. Amsterdam & Philadelphia: John Benjamins.
- Haspelmath, Martin. 2016. Universals of causative and anticausative verb formation and the spontaneity scale. *Lingua Posnaniensis* 58(2). 33–63.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics* 57(3).
- Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyacı. 2014. Coding causal–noncausal verb alternations. *Journal of Linguistics* 50(3). 587–625.
- Heidinger, Steffen. 2019. Reflexive and unmarked anticausatives in French and Spanish: Frequency of transitive use and undergoer overlap. *Langages* 216(4). 53–69.
- Nichols, Johanna, David A. Peterson & Jonathan Barnes. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8(2). 149–211.
- Portilla, Mario Chaves. 2007. Diacronía de las construcciones con pronombres reflexivos en español. *Filología y Lingüística* 33(1). 131–149.
- Tubino-Blanco, Mercedes. 2020. Causative-Inchoative in Morphology. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.
- Zúñiga, Fernando & Seppo Kittilä. 2019. *Grammatical Voice*. Cambridge: Cambridge University Press.

**Corpora**

- Morfologia dell’Italiano in DIACRONIA* (MIDIA): [www.corpusmidia.unito.it](http://www.corpusmidia.unito.it) [Accessed: 28/12/2022]
- Real Academia Española (2013): Corpus del Diccionario histórico de la lengua española (CDH): <https://apps.rae.es/CNDHE> [Accessed: 28/12/2022]



## The Chronicle of Lingbe, an Extinct Bantu Language of East Congo

Lingbe, aka Ngbele, was a Bantu language spoken in the Northeast of the Democratic Republic of Congo, near the Bomokandi river. UNESCO (Moseley 2010) and the Summer Institute of Linguistics (see Eberhard et al. 2022) consider the language to be extinct. Indeed, a recent questionnaire on the language, recorded in 2020, resulted in a 100-word list of a Nilo-Saharan language, possibly Mangbetu, rather than the Bantu language we were actually looking for.

We are presenting the chronicle of a death foretold: the extinction of Lingbe was already announced in the 1930s. Moeller (1936) writes that the Mangbele were fishermen, one of the “Bakango” peoples living at the Uélé and Bomokandi rivers. At Rungu and Wamba the Mangbele had already exchanged their language for Mayogo, an Ubangi language. Other Mangbele had been conquered by the Mangbetu and incorporated in Mangbetu and “Madjaga” communities. As mentioned, Mangbetu is a Nilo-Saharan, more specifically Central-Sudanic, language. The present paper explores historical sources from the early 20<sup>th</sup> century that document the language and its speakers as well as the on-going language shift. This allows us to study the context in which the language shift and ultimate extinction of Lingbe occurred.

Still, the language lived at least until the 1990s when the second author recorded a wordlist in Lingbe. As short as it may be, merely 67 words, it has already allowed the authors and colleagues to identify Lingbe as a Boan Bantu language, more specifically affiliated to the western Boan subgroup which includes Leboale, Kango and Ngelima. Interestingly, this is in line with the hypothesis forwarded by Moeller (who cites De Calonne-Beaufaict 1921) that they are closely related to the Ababua. The present paper compares the surviving linguistic material to the surrounding Bantu, Central-Sudanic and Ubangi languages, as such reconstructing as much as is possible on Lingbe’s early history and studying the process of language shift.

The present paper assembles all available documentation on Lingbe, be it historical sources, linguistic or other, and retrieves the utmost from the linguistic material. The research results offer unique insights into the history of a lost Bantu language.

### References:

- De Calonne Beaufaict, Adolphe. 1921. *Azande : introduction à une ethnographie générale des bassins de l’Ubangi-Uele et de l’Aruwimi*. Bruxelles: M. Lamertin.
- Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2022. *Ethnologue: Languages of the World. Twenty-fifth edition*. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com> .
- Moeller, A. 1936. *Les Grandes Lignes des migrations des Bantous de la Province-Orientale du Congo belge*. Brussels : IRCB, Section des sciences morales et politiques, Mémoires, T. VI.
- Moseley, Christopher (ed.). 2010. *Atlas of the World’s Languages in Danger, 3rd edn*. Paris, UNESCO Publishing.

## Dramatic texts as a source of stigmatization from below

Linguistic stigmatization has mostly been described as a "from above" process. However, stigmatizing grammarians must also have their structures from somewhere: either from their own observation and evaluation of language, or from stereotypes and stigmas that are already circulating within the speech community. Moreover, stigmatizations from the top must also descend to the bottom, where they must be incorporated into a broader discourse of linguistic registers and evaluative techniques.

Because fictional, literary language is typically a standard (written) variety, it frequently exploits the contrast with orality. This frequently involves orality features that are considered characteristic of certain groups of speakers, as we see in classical Greek drama, which clearly influenced modern European literatures.

In German literature we already find passages with fictionalizations of oral varieties in Middle High German texts. As a literary strategy, however, the opposition between written and oral language(s) becomes popular only from the early modern period. Especially within passages of direct speech and dramatic texts, oral varieties are adapted. Such oral variety implementations convey linguistic (and non-linguistic) stereotypes as for example the excessive use of diminution (by means of *-le*) and the apocope in the text of a Swabian peddler in the play "Der Eheteufel auf Reisen" (1822) by the Viennese author Josef Alois Gleich:

*Ach du lieber Herrgott, was ist das für ein Lebe-ø , ja, es wäre schon recht, ich wollte auch gern tanze-ø und gute Bißle esse, wenn wir nur auch das Geldle dazu hätte-ø. Aber mein Mann, das ist ein Lump ohnegleiche-ø – den ganzen Tag sitzt er in Branntweinhäusle, und ich kann nicht begreife-ø, wo er auf den Abend das Geldle hernimmt.* [Emphasis added]

'Oh, dear God, what kind of life is that, yes, it would be all right, I would also like to dance and eat good bits, if only we had the money for it. But my husband is an incomparable rascal - all day long he sits in a liquor house, and I can't understand where he gets the money for the evening.'

The lecture examines the extent to which such structures of intended orality serve stereotypes (in the definition according to Labov 1972 and Silverstein 2003) and can be used as a source of stigmatization from below using 200 German-language plays from the 16th to the 19th centuries. Selected phenomena (diminution, pronominal adverbs, variation within the VP, and verbal periphrasis) will be investigated and compared to previous studies on stigmatization processes in the history of German (esp. Davies & Langer 2006).

Davies, Winifred V. / Langer, Nils (2006): *The Making of Bad Language*. Frankfurt am Main: Peter Lang.

Labov, William (1972): *Sociolinguistic Patterns*. Conduct and Communication. Philadelphia: University of Pennsylvania Press.

Silverstein, Michael (2003): „Indexical Order and the Dialectics of Sociolinguistic Life.“ *Language and Communication* 23 (3): 193–229.

## Divergence and contact in Cappadocian concessive conditionals

In this talk we present preliminary results from an ongoing investigation into concessive conditionals (CCs) in Cappadocian, a near-extinct variety of Greek that was heavily influenced by Turkish when it was spoken in Asia Minor from Byzantine times until the Greek-Turkish population exchange in 1923–1924. We investigate whether Cappadocian CCs deviate from their Standard Modern Greek counterparts and, if so, whether the deviations can be attributed to language contact with Turkish.

CCs are a special type of conditionals which express not one antecedent  $p$ , but a set of antecedents that all lead to the same consequent  $q$ : ‘if  $\{p_1, p_2, p_3, \dots\}$ , then  $q$ ’ (König 1986; Leuschner 2006, 2020). Three quantificational strategies to evoke this set are distinguished (ibid.): scalar concessive conditionals (SCCs) mention an extreme value  $p_n$  and imply that  $q$  also holds for other, less extreme values (cf. English *even if  $p_n$* ); alternative concessive conditionals (ACCs) express a disjunction which exhausts the scale at hand (cf. English *whether  $p_1$  or  $p_2$* ); universal concessive conditionals (UCCs) express free-choice quantification over instantiations of a variable, often realized as an interrogative-like pronoun (cf. English *WH-ever*).

Under Haspelmath & König’s (1998) typology of CCs in European languages, Turkish qualifies as uniformly-coding, i.e. as a language that encodes all CCs as conditionals, while Standard Modern Greek is differentially-coding, i.e. a language in which only SCCs have overt conditional coding, while ACCs and UCCs have primarily quantificational, e.g. interrogative-like, coding. This typological divide makes Cappadocian an interesting case study.

We investigate the coding strategies of Cappadocian CCs in a corpus of 58 folktales from 11 villages (ca. 50,000 words, the largest corpus of Cappadocian to date). While Cappadocian CCs are differentially coded like their Standard Modern Greek counterparts, the actual coding is distinct between both varieties. In part, these differences are due to Turkish influence, as Turkish loan words are found in ACCs, e.g. *jáxot ... jákot ...* ‘whether ... or ...’ < Tr. *yahut* ‘or, else’ and sporadically in UCCs, e.g. *-dak* in *ótia-dak* ‘whatever’ < Tr. *dek* ‘until, as far as’. Mostly, however, Cappadocian CCs differ from their Standard Modern Greek equivalents in ways that cannot be attributed to Turkish. In Cappadocian SCCs, the focus particle *ke* ‘even’ invariably follows the conditional conjunction *an* ‘if’, whereas *ke* precedes *an* in SCCs in Standard Modern Greek (where *an ke* is purely concessive). And whereas Haspelmath & König (1998) suggest that Standard Modern Greek UCCs usually contain focus particles like *-dipote* ‘ever’ or *ke* ‘even’ and/or conditional *an*, Cappadocian UCCs lack any overt coding other than the WH-word in 68% of all instances.

Future studies should investigate whether these differences are a consequence of changes in Modern Greek, with Cappadocian preserving coding strategies from earlier stages of Greek due to its relative isolation from mainstream Greek since the defeat of the Byzantine Empire by the Seljuk Empire at the Battle of Manzikert in 1071. Future research should also try to account for SCCs in the Floïta dialect, which are introduced by *an ki* like exceptive conditionals. To our knowledge, concessive and exceptive conditionals are not coded identically in any other varieties of Greek nor, indeed, any other languages. We suggest this overlap can be explained either in terms of accidental homonymy or as scale/polarity reversal.

## References

- Haspelmath, Martin & Ekkehard König (1998): Concessive conditionals in the languages of Europe. In Johan van der Auwera (ed.): *Adverbial constructions in the languages of Europe*. Berlin: de Gruyter, 563–640. <https://doi.org/10.1515/9783110802610.563>
- König, Ekkehard (1986): Conditionals, concessive conditionals and concessives: Areas of contrast, overlap and neutralization. In Elizabeth Closs Traugott, Alice ter Meulen, Judy Snitzer Reilly & Charles A. Ferguson (eds.): *On conditionals*. Cambridge: Cambridge University Press, 229–246. <https://doi.org/10.1017/CBO9780511753466.013>
- Leuschner, Torsten (2006): *Hypotaxis as building-site: The emergence and grammaticalization of concessive conditionals in English, German and Dutch*. Munich: Lincom Europa.
- Leuschner, Torsten (2020): Concessive conditionals as a family of constructions. *Belgian Journal of Linguistics* 34, 234–246. <https://doi.org/10.1075/bjl.00049.leu>

### Gascon u-perfects and the analogical foregrounding of inflectional class

This paper combines historical corpus data and comparative reconstruction to elucidate the origin and spread of the thematic vowel /u/ found today throughout preterite and imperfect subjunctive forms (i.e. reflexes of Latin *perfectum* forms; morphomic distribution ‘PYTA’, Maiden 2018) of third-conjugation verb lexemes in Gascon (Gallo-Romance) varieties.

Modern Gascon verb inflection (Grosclaude & Nariò 1998, Romieu & Bianchi 2005, Massourre 2012) is typified by a three-way conjugational class distinction in most tense-aspect-mood categories, expressed principally via theme vowels. *Perfectum* reflexes are characterised by thematic /ɛ/ in the first conjugation, e.g. *cantè* [kan'tɛ] ‘sing.PRET.3SG’, *cantèsse* [kan'tese] ‘sing.IPF.SBJV.3SG’, continuing Latin -DEDĪ perfects; thematic /i/ in the fourth conjugation, e.g. *bastí* [bas'ti] ‘build.PRET.3SG’, *bastísse* [bas'tise] ‘sing.IPF.SBJV.3SG’, continuing Latin -IUĪ perfects; and thematic /u/ in the third conjugation, e.g. *venó* [be'nu] ‘sell.PRET.3SG’, *venósse* [be'nuse] ‘sell.IPF.SBJV.3SG’. While the early historical development of the first- and fourth-conjugation forms shows continuity with other Occitan varieties (Ronjat 1937, Skårup 1997, Wheeler 2012), thematic /u/ is unique to Gascon within southern Gallo-Romance.

Traditional accounts (Zauner 1896:444-446, Allières 1988:177, Massourre 2012:214) ascribe Gascon u-perfects to the analogical influence of the single item ‘be’, and historical reconstruction supports this view. Sound change only yields theme vowel /u/ in *perfectum* reflexes of ‘be’ (e.g. FUIT > *fo* /fo/ > modern *hó* /hu/ ‘be.PRET.3SG’), and there is no potential source of /u/ elsewhere in the paradigm: in Gascon, the past participle in -/yt/ < -ŪTUM remains distinct from the *perfectum* reflexes, whereas in northern Gallo-Romance, reflexes of Latin past participles in -ŪTUM exert analogical influence on *perfectum* reflexes (Pope 1952:370, 381).

While the initial emergence of u-perfects occurs at the pre-literary period and is not directly visible in textual evidence (Bourciez 1927), the chronology and direction of their subsequent geographical and lexical spread can be traced in data from the Linguistic Corpus of Old Gascon (LCOG; Field 2012, 2013) and the *Atlas linguistique de la Gascogne* (ALG; Allières 1971). The progressive lexical extension of u-perfects is associated with the levelling of root and stress alternations in former ‘strong’ perfects (those with root-stressed forms in a subset of persons), e.g. analogical *metó* [me'tu] ‘put.PRET.3SG’ replaces etymological *mes* [mes] < MĪSIT, a process largely complete by the late thirteenth century; and later, with reduction in distinctive root allomorphy characterising morphomic distribution patterns. In some varieties, u-perfects spread beyond the third conjugation and into all fourth conjugation lexemes; this development is visible in northern dialects from the thirteenth century onwards, and is associated with the introduction of a thematic ‘augment’ (Maiden 2004, Esher 2016), e.g. analogical *bastissó* [basti'su] ‘build.PRET.3SG’ replaces etymological *bastí* [bas'ti].

The Gascon developments are significant for theories of inflectional analogy since they provide evidence for analogical remodelling of multiple items on the model of a single, idiosyncratic item of high token frequency, a phenomenon rarely reported in existing literature (see e.g. Cowgill 1959:11, Fulk 2018:189, 209, 307–308 for possible examples in Germanic). A further noteworthy aspect is the historical tendency of Gascon inflectional systems to retain and reinforce conjugational class distinctions and morphomic stem contrasts based on thematic vowels, contrary to the tendency observed across other Occitan varieties, in which contrasts based on distinctive root allomorphy take precedence (compare Esher 2021a,b,2022).

## References

- Allières, Jacques. 1971. *Atlas linguistique de la Gascogne. Volume V: Le verbe*. Paris: CNRS.
- Allières, Jacques. 1988. Quelques énigmes de la conjugaison gasconne ou le verdict de l'aire.  
In Christian Abry (ed.), *Espaces romans*. Grenoble: Ellug. 175–192.
- Bourciez, Jean. 1927. *Recherches historiques et géographiques sur le parfait en Gascogne*.  
Bordeaux: Feret.
- Cowgill, Warren. 1959. The inflection of the Germanic *ō*-presents. *Language* 35(1). 1–15.
- Esher, Louise. 2016. Morphomic distribution of augments in varieties of Occitan. *Revue Romane* 51(2). 271–306.
- Esher, Louise. 2021a. Hypercorrection and velar-to-labial change in Occitan preterites, *Zeitschrift für romanische Philologie* 137(1). 61–98.
- Esher, Louise. 2021b. La diachronie du prétérit occitan à la lumière des sources textuelles. *Linguistica Pragensia* 31(2). 115–136.
- Esher, Louise. 2022. The historical development of imperfect indicative and conditional inflection in Pyrenean Romance. *Transactions of the Philological Society*.  
<https://doi.org/10.1111/1467-968X.12234>
- Field, Thomas T. 2012. Variation et diachronie: le témoignage du corpus électronique gascon.  
In Mario Barra-Jover, Guylaine Brun-Trigaud, Jean-Philippe Dalbera, Patrick Sauzet & Tobias Scheer (eds.), *Études de linguistique gallo-romane*. Paris: Presses universitaires de Vincennes. 21–32.
- Field, Thomas T. 2013. *The Linguistic Corpus of Old Gascon*. Research database.  
<<http://mllidev.umbc.edu/gascon/>>.
- Fulk, R. D. 2018. *A comparative grammar of the early Germanic languages*. Amsterdam: Benjamins.
- Grosclaude, Michel & Gilabèrt Nariò. 1998. *Répertoire des conjugaisons occitanes de Gascogne*. Orthez: Per Nòste.
- Maiden, Martin. 2004. Verb augments and meaninglessness in Romance morphology. *Studi di grammatica italiana* 22. 1–61.
- Maiden, Martin. 2018. *The Romance verb. Morphomic structure and diachrony*. Oxford: OUP.
- Massourre, Jean-Louis. 2012. *Le Gascon: Les mots et le système*. Paris: Champion.
- Pope, Mildred K. 1952. *From Latin to modern French*. Manchester: Manchester University Press.
- Romieu, Maurice & André Bianchi. 2005. *Gramàtica de l'occitan gascon contemporanèu*.  
Bordeaux: Presses Universitaires de Bordeaux.
- Ronjat, Jules. 1937. *Grammaire istorique des parlers provençaux modernes*, vol. III.  
Montpellier: Société des Langues Romanes.
- Skårup, Povl. 1997. *Morphologie élémentaire de l'ancien occitan*. Copenhagen: Museum Tusulanum Press.
- Wheeler, Max. 2012. Vies d'analogia i d'explicació en l'evolució del pretèrit feble de la conjugació-e romànica. *Estudis Romànics* 34. 7–36.
- Zauner, Adolf. 1896. Die Konjugation im Béarnischen. *Zeitschrift für Romanische Philologie* 20. 433–470.



### **The role of French in the Johnsons' correspondence**

Among the studies flourished in the field of historical sociolinguistics, mainly in the strand devoted to English history, during its Medieval and early modern phases, multilingual texts have been analysed using theories and models coming from contact linguistics, thus applying synchronic models and approaches to the study of the past (Eckert 2012, Nurmi & Phata 2004, among others).

This contribution aims at investigating the role and the usage of the French language in the correspondence between Otwell Johnson and his brother, John Johnson, merchants of the Staple in Calais. Their letters, along with those of their other brother and their partners, were preserved only because of legal reasons (Oldroyd 1998), and the entire collection was subsequently transcribed for biographical studies by Winchester (1955).

The letters written by Otwell Johnson to his brother display interesting usages of the French language, making them multilingual; although the writer makes predominant use of English, we can find French elements in 17 letters (out of 161). These non-English items can be broadly grouped into three different contact phenomena: language choice, inter-sentential code switching, and intra-sentential code switching (defined as in Ciccolone & Dal Negro 2021).

Applying an atomistic approach (Hernández-Campoy 2003) to the study of the contact phenomena, I will investigate the above-mentioned documents, trying to explore the nature of the switching forms they contain, from an intra-writer variation perspective (Auer 2015, Nevalainen & Raumolin-Brunberg 2003). After analysing the letters and the type of multilingualism in them, I will also try to evaluate the functions given to the French language in this specific context.

### **References:**

- Auer, A. (2015). Stylistic variation. In Auer, A., Schreier, D., & Watts, R.J. (eds.), *Letter Writing and Language Change*. Cambridge: Cambridge University Press, 133-155.
- Ciccolone, S., & Dal Negro, S. (2021). *Comunità bilingui e lingue in contatto. Uno studio sul parlato bilingue in Alto Adige*. Caissa.
- Eckert, P. (2012). Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41, 87-100.
- Hernández-Campoy, J.M. (2003). Exposure to contact and the geographical adoption of standard features: two complementary approaches. *Language in Society*, 32(2), 227-255.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical Sociolinguistics. Language Change in Tudor and Stuart England*. London: Longman Pearson Education.

Nurmi, A., & Phata, P. (2004). Social stratification and patterns of code-switching in Early English letters. *Multilingua*, 23(4), 417-456.

Oldroyd, D. (1998). John Johnson's letters: the accounting role of Tudor merchants' correspondence. *The Accounting Historians Journal*, 25(1), 57-72.

Winchester, B. (1955). *Tudor Family Portrait*. London: Jonathan Cape.

## Tracing the origins of resumption in Swedish

Modern Swedish is a strict Verb-Second language. However, the language regularly displays Verb-Third strings when an adverbial particle is inserted between a fronted adverbial and the finite verb (*adverbial resumption*) (1).

- (1) *Ändå så finns det där hatet där ute*  
 still SÅ exist that there hatred.DEF there out  
 ‘Still, there is such a hatred out there.’ (Bloggmix 2005)

This paper will provide a detailed account of the evolution of adverbial resumption through the history of the Swedish language, focusing on the resumptive particles *så* ‘so’, and *då* (2). The main difference between the two is that *då* must follow an element that conveys time, while there are no restrictions on the use of *så*. In Meklenborg’s terms, *så* is a generalized resumptive, while *då* is specialized (Meklenborg, 2020). *Så* is by far the more common (see *inter alia* Elmquist (1945); Ekerot (1998); Nordström (2010)).

Using the Korp corpus (Borin et al., 2012), we find that in Early Old Swedish, resumption is very common. It becomes less frequent in the following centuries, before surging in the 18th century.

- (2) *hauir bondæn alt synir . tha takær han ey mer æn*  
 have farmer.DEF already sons . tha take.PRS he NEG more than  
*en sun*  
 a son  
 ‘If the farmer already has sons, then he will not take more than one son’

(SL, 1203–1212)

Whereas resumption with *så* is the most prominent structure in Modern Swedish, the structure was marginal in Old Swedish. The dominating resumptive element in the earliest period of the Swedish language was *þa* ‘then’ (= *då*). Out of 3976 cases where a fronted CP is followed by a resumptive in Early Old Swedish, the resumptive is *þa* in 3969 cases, while only 7 cases contain the resumptive *swa* (= *så*). Resumption with *swa* is restricted to contexts where the initial CP conveys a condition or a comparison. *Swa* is therefore a specialized resumptive during this period.

In the 16th century, the picture changes completely. Out of 190 cases of fronted CP + resumptive, the resumptive is *så* in 162 cases. In the next centuries, the ratio of resumption with *då* drops further. Looking at the semantics of the fronted adverbial constituent, we find that resumption with *då* gets more and more restricted, while these resumption with *så* becomes more widespread. We can, in other words trace the evolution into specialized and generalized resumptives.

## References

- Lars Borin, Markus Forsberg, and Johan Roxendal. Korp – the corpus infrastructure of språkbanken. In *Proceedings of LREC 2012. Istanbul: ELRA*, pages 474–478, 2012. URL [http://www.lrec-conf.org/proceedings/lrec2012/pdf/248\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/248_Paper.pdf).
- Lars-Johan Ekerot. *Så-konstruktionen i svenskan. Konstruktionstypen "Om värdet tillåter, så genomförs övningen i funktionellt grammatisk perspektiv"*. PhD thesis, University of Lund, Lund, 1998.
- Axel Louis Elmquist. The use of resumptive så after non-adverbial expressions in swedish. *Scandinavian Studies*, 18:284–288, 1945.
- Christine Meklenborg. Resumptive Particles and Verb Second. In Rebecca Woods and Sam Wolfe, editors, *Rethinking V2*, pages 90–125. Oxford University Press, Oxford, 2020.
- Jackie Nordström. The swedish så-construction, a new point of departure. *Working Papers in Scandinavian Syntax*, 85:37–63, 2010.

**Old and new approaches to vernacular glosses in textual culture of medieval Poland: The case of embedded glosses**

Most of Old Polish texts are in fact fifteenth century copies with accompanying inscribed or reinscribed glosses (interlinear and marginal) introduced into the text. The aim of the present article is to show how to recognize these types of glosses in the text using elements of semantics, syntax, rhetoric and source studies. The majority of examples come from the most extensive monument of Polish and Slavic medieval prose and apocryphal literature in the Polish language: *Rozmyślanie przemyskie* [The Przemysl Meditation]. About several hundred glosses introduced into the text of *Rozmyślanie* have been recognized. The glosses have had an enormous impact on the form of the preserved copy.

In the only preserved copy of this apocryphal text all glosses are incorporated into the main body of the text, often haphazardly and in inappropriate places. Some of them were turned into headings of chapters by the last copyist.

During our paper we will show how to work with a text with a multi-layered structure (a "text within a text within a text"), what tools are most suited for proper recognition of glosses and how to analyse the degree of their integration with the main body of the text.

## The ordering of matrix clauses and subordinate causal clauses in the Old Bailey Corpus 1720–1913

The paper investigates changes in the order of finite adverbial causal clauses and matrix clauses in Late Modern English. Such clauses most frequently follow their main clause, as in (1), but they can also precede it, as in (2).

(1) *I did not much wonder at it, **because she had attempted to cut her Throat before** [...]* (OBC-1726)

(2) [...] ***because it was to be my place to take them**, I kept company with them all four for two hours* (OBC-1755)

Diessel (2005:465) argues that main clauses containing final adverbial clauses are cognitively privileged because they are easier to parse (Diessel 2005: 465–9); initial positioning is only possible for short subordinate clauses. Likewise, discourse pragmatic function plays a role, with initial position more likely if the reason presented is given rather than new. Therefore, *because* clauses, usually providing new causes, are less likely to be initial than clauses introduced by *as* and *since*, which tend to introduce known causes (Diessel 2005: 465–6; Chafe 1984: 442–4). Studies of adverbial clauses in earlier stages of English have broadly confirmed many of these results: Given-/newness plays a role for the positioning of causal clauses in Early Modern English (cf. Claridge & Walker 2001: 46–8; Pentrel 2017: 278 for temporal clauses in the 17th century), and Eitelmann (2016: 409), e.g., stresses the general validity of end-weight in Late Modern English.

The present paper analyses 1055 causal subordinate clauses from the 24-million word Old Bailey Corpus 2.0, which contains transcripts of trials from London's Central Criminal Court from 1720–1913 (Huber, Nissel and Puga 2016). Using multiple logistic regression, it explores the historical development of the influence of clause length and the given-/newness of the cause on the ordering of causal and matrix clause in a formative phase of English syntax, which saw the demise of *for* and the establishment of both *as* and *because* as causal conjunctions (Rissanen 1998). It is hypothesised that, under the uniformitarian hypothesis, these well-known factors had similar effects on clause position as in present-day English. The results feed into a diachronic constructional analysis which models the historical changes in the network of English causal clauses (Kanetani 2019).

### References

Chafe, Wallace. 1984. How People Use Adverbial Clauses. In Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society, February 17-20, 1984, 437–49. Berkeley Linguistic Society.

Claridge, Claudia & Terry Walker. 2001. Causal clauses in written and speech-related genres in Early Modern English. ICAME Journal (25). 31–64.

Eitelmann, Matthias. 2016. Support for end-weight as a determinant of linguistic variation and change. English Language and Linguistics. Cambridge University Press 20(3). 395–420.



Diessel, Holger. 2005. Competing Motivations for the Ordering of Main and Adverbial Clauses. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 43(3). 449–470.

Huber, Magnus and Magnus Nissel and Karin Puga. 2016. *Old Bailey Corpus 2.0*. hdl:11858/00-246C-0000-0023-8CFB-2.  
[http://fedora.clarin-d.uni-saarland.de/oldbailey/downloads/OBC\\_2.0\\_Manual%202016-07-13.pdf](http://fedora.clarin-d.uni-saarland.de/oldbailey/downloads/OBC_2.0_Manual%202016-07-13.pdf). (7 January 2023).

Kanetani, Masaru. 2019. *Causation and Reasoning Constructions*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Rissanen, Matti. 1998. Towards an integrated view of the development of English: Notes on causal linking. In Jacek Fisiak & Marcin Krygier (eds.), *Advances in English historical linguistics*, 389–406. Berlin & Boston: Mouton De Gruyter.

## The shape of grammaticalization: matching the bridging context scenario with patterns of frequency use

Heine's theory of contexts (2002) has been recognized as a satisfying account of a grammaticalization process, detailing the diachronic steps leading to and stemming from the drastic semantic shift or reanalysis that lies at the core of a grammaticalization process (Detges & Waltereit 2002). Other grammaticalization scholars have offered similar scenarios for the key steps of a grammaticalization process (Diewald 2002), introducing a critical context in which a new semantic inference can take place, while this inferential mechanism as a trigger for grammaticalization has been scrutinized on its own (Traugott 1988, Nicolle 1998, Nicolle 2011). From then, the form spreads over new contexts and becomes conventionalized through further constructional changes (Smirnova 2015).

On the other hand, it is known that language change is reflected, in terms of frequency of use, by a common template, that of the S-curve (Kroch 1989, Blythe & Croft 2012). Furthermore, this S-curve has been specifically associated with semantic expansion (Feltgen et al. 2017). Therefore, it is to be expected that the context-based scenario outlined by Heine and others in the one hand, and the S-curve pattern of frequency change on the other hand, should match on the level of their diachronic development; however, no robust parallel between the two accounts has been offered so far. Notably, frequency has been considered as a driving force for semantic change in grammaticalization processes (Bybee & Thompson 1997, Bybee 2006); nevertheless, whether the frequency needs to reach a 'critical threshold' for the shift to occur (or in which stage of the S-curve does the form start transitioning from the bridging context to the switch context) remains largely unclear. The goal of this contribution is precisely to remedy to this state of things.

To do so, I provide three empirical case studies of grammaticalizations in French, based on corpus data from the Frantext database (ATILF 1998-2023). These three case studies have been chosen to test my findings over three widely different grammaticalization scenarios. I thus explore a grammaticalization leading to the entrenchment of a schematic construction (Trousedale 2014), *une espèce de N* ('a kind of N'); a grammaticalization highlighted by a syntactic shift of the form (Fischer 2010), *mis à part*, which starts being fronted by the nineteenth century; and a grammaticalization driven by a calque over an already existing form (Cornillie 2019), *d'une façon ADJ* ('in a(n) ADJ way'), calqued from *d'une manière ADJ*. For each of these studies, I follow qualitatively the emergence of the grammatical meaning by manually parsing the occurrences, and I compare it with the S-curve of frequency rise extracted from the quantitative data.

In all of these three cases, the S-curve appears in the wake of the grammaticalization 'trigger', that is, as a consequence of the semantic shift. Several hints support this hypothesis; e.g., in the case of *une espèce de N*, the frequency over the types (i.e. how many fillers can enter the construction) also follows an S-curve, which would indicate a lexical diffusion (Ogura 2007); but then the token frequency of each individual fillers follows the same, identical S-curve, suggesting that they all 'register' the same semantic shift. The competition between *d'une façon* and *d'une manière* also follows two mirror S-curves, indicative of a transfer of a 'semantic load' between the two, similar to other competition processes (*en vs dans* in Fagard & Combettes 2013). The S-curve is thus mostly associated with stages 3 and 4 in Diewald's scenario, or with a diffusion over switching contexts in Heine's account. As Heine emphasizes, this also shows that the trigger of a grammaticalization does not guarantee its entrenchment in language use. A further actuation step is necessary. Incidentally, it also shows that frequency plays an ambivalent role in a grammaticalization scenario: it follows the semantic expansion and derives from it, seems necessary to entrench it, but does not seem to foster or facilitate it.

## References

- ATILF. (1998-2023). *Base textuelle Frantext* (En ligne). ATILF-CNRS & Université de Lorraine. <https://www.frantext.fr/>
- Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88(2), 269-304.
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733.
- Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. *Annual Meeting of the Berkeley Linguistics Society*, 23(1), 378-388.
- Cornillie, B. (2019). On the pace of syntactic elaboration from Latin calques: Evidence from Spanish, Dutch and English. *Belgian Journal of Linguistics*, 33(1), 82-107.
- Detges, U., & Waltereit, R. (2002). Grammaticalization vs. reanalysis: A semantic-pragmatic account of functional change in grammar. *Zeitschrift für Sprachwissenschaft*, 21(2), 151-195.
- Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. In Wischer, I. & Diewald, G. (ed.), *New reflections on grammaticalization*, John Benjamins Publishing, 103-120.
- Fagard, B., & Combettes, B. (2013). De en à dans, un simple remplacement?. *Langue française*, 178, 93-115.
- Fischer, S. (2010). *Word-order Change as a Source of Grammaticalisation*. John Benjamins Publishing.
- Heine, B. (2002). On the role of context in grammaticalization. In Wischer, I. & Diewald, G. (ed.), *New reflections on grammaticalization*, John Benjamins Publishing, 83-102.
- Kroch, A. S. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1(3), 199-244.
- Nicolle, S. (1998). A relevance theory perspective on grammaticalization, *Cognitive Linguistics*, 9(1), 1-35.
- Nicolle, S. (2011). Pragmatic aspects of grammaticalization. In Heine, B. & Narrog, H. (ed.), *The Oxford handbook of grammaticalization*, Oxford University Press, 401-412.
- Ogura, M. (2007). Lexical Diffusion and Complex Adaptive System. *Studies in Modern English*, 2007(23), 23-51.
- Smirnova, E. (2015). When secondary grammaticalization starts: A look from the constructional perspective. *Language Sciences*, 47, 215-228.
- Traugott, E. C. (1988). Pragmatic strengthening and grammaticalization. *Annual Meeting of the Berkeley Linguistics Society*, 14, 406-416.
- Trousdale, G. (2014). On the relationship between grammaticalization and constructionalization. *Folia Linguistica*, 48(2), 557-578.

***WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose.***

In Latin and ancient Greek linguistics, onomasiological studies on the lexicalizations of concepts have led to a productive line of research based on in-depth philological analysis of extensive textual evidence (e.g. Moussy 1999; Kaster 2001; Craig 2005; Moussy 2005; 2007; Zamboni 2008; Lecaudé 2010; Kutscher and Werning 2014). The main obstacle to large-scale onomasiological analyses is that they require specialized language resources organized by concepts (Black 2001) rather than by lemmas. The Historical Thesaurus of English (Kay 1984) has enabled studies on diachronic patterns in English lexicalization of various concepts (Wild 2010; Alexander 2018; Allan 2020). For ancient Greek and Latin, we can now use WordNet (Biagetti et al. 2021).

Building on the work done within the MultiWordNet project to automatically generate Latin synsets from dictionaries (Minozzi 2017), the Latin (<https://latinwordnet.exeter.ac.uk>) and Ancient Greek WNs (<https://greekwordnet.chs.harvard.edu>), directed by Dr William Short (University of Exeter), aim to create accurate machine-readable and machine-actionable representations of the semantic structures of Ancient Greek and Latin using the WN framework. Thanks to the same data structures and semantic descriptors (*synsets*, *semfields*) provided by the Princeton WN, they intend to model the idiosyncratic and unique features of the semantics of ancient languages, in terms that are compatible and comparable with models constructed for other languages, as well as integratable into larger computational systems. WNs for the ancient languages try to capture diachronic and generic variations in word meaning, and figurative structures that impact semantic development. Based on conceptual metaphor theory in cognitive linguistics (Lakoff and Johnson 1980), they capture both the metaphorical or metonymical relationships between word senses at the level of word meaning and large-scale figurative relations that may operate supra-lexically. In addition, they contain etymological information.

Linking onomasiological resources such as WNs to corpus data can help us answer quantitative questions on the lexicalization of concepts. In our study, we combine WNs for ancient languages and corpus data, with specific attention to Ancient Greek and Latin. For our analysis, we choose the lexical field of SEA, given its importance in the two cultures. For the Greeks and the Romans, the sea was a source of wealth and trade (e.g. Reed 2003; Wilkinson 2020; Boardman et al. 2021), and a key part of their military strategy (e.g. Harris 2017; Nash 2018), but also a place of mystery and danger (e.g. Berens 1979; Lindenlauf 2004; Nikoloska 2012; Beaulieu 2016).

We focus on Ancient Greek and Latin pairs of nouns (e.g. AGr. *thálassa*, *póntos*, *pélagos*, *háls* : Lat. *mare*, *pontus*, *pelagus*, *aequor* ‘sea’) and verbs (e.g. AGr. *pléō* : Lat. *navigo* ‘sail’) related to SEA. We enrich the Ancient Greek and Latin Dependency Treebanks (Celano 2019; v2.1 [http://perseusdl.github.io/treebank\\_data/](http://perseusdl.github.io/treebank_data/)) with semantic information from the Ancient Greek and Latin WNs. Our corpus includes the treebanked passages from Herodotus’s *Histories* and Caesar’s *De bello Gallico*. We then compare the quantitative data from the two languages to understand in what way Ancient Greek and Latin lexicalize concepts related to SEA, investigating the best way to add semantic information to the Ancient Greek and Latin Dependency Treebanks. This is a challenging task as WNs often provide a large amount of potentially relevant semantic information, and, even in a specific textual context, words can possess multiple senses, or their sense(s) can be ambiguous. Our study will cast new light on the use of WNs for crosslinguistic comparison for ancient Indo-European languages. It will also represent an attempt to digitally link semantic and syntactic information for classical languages, integrating WNs and treebanks (for modern languages see e.g. Kingsbury et al. 2002; Hajnicz 2014).

## REFERENCES

- Alexander, M. 2018. Lexicalization Pressure. In *International Conference of English Historical Linguistics XX*, August 2018, Edinburgh, plenary paper.
- Allan, K. 2022. Degrees of lexicalization in the history of English: antonymy and asymmetry. In *International Conference on Historical Linguistics*, Oxford, 1-5 August 2022. Book of Abstracts.
- Biagetti, E., Zanchi, C. & Short, W. M. 2021. Toward the creation of WordNets for ancient Indo-European languages. In *Proceedings of the 11th Global Wordnet Conference*, University of South Africa (UNISA). Global Wordnet Association, 258–266.
- Beaulieu, M-C. 2016. *The Sea in the Greek Imagination*. Philadelphia: University of Pennsylvania Press.
- Berens, E. M. 1979. *The Myths and Legends of Ancient Greece and Rome*. Boston: Longwood Press.
- Blank, A. 2001. Words and concepts in time: Towards diachronic cognitive onomasiology. In *Words in Time: Diachronic Semantics from Different Points of View*, 37–65.
- Boardman, J., G. R. Tsetschladze, A. Avram, and J. Hargrave. 2021. *The Greeks and Romans in the Black Sea and the Importance of the Pontic Region for the Graeco-Roman World (7th Century BC-5th Century AD): 20 Years on (1997-2017)*. *Proceedings of the Sixth International Congress on Black Sea Antiquities (Constanta - 18-22 September 2017)*. Oxford: Archaeopress.
- Celano G. G. 2019. The Dependency Treebank for Ancient Greek and Latin. In *Digital Classical Philology*, 279–298.
- Craig, W. 2015. The language of gender: lexical semantics and the Latin vocabulary of unmanly men. In M. Masterson, N. S. Rabinowitz, and J. Robson, (eds.), *Sex in antiquity: exploring gender and sexuality in the ancient world*, 461-481. London: Routledge.
- Hajnicz, E. 2014. Lexico-Semantic Annotation of Składnica Treebank by Means of PLWN Lexical Units. In *Proceedings of the Seventh Global Wordnet Conference*, 23–31.
- Harris, W. V. 2017. Rome at Sea: The Beginnings of Roman Naval Power. In *Greece and Rome* 64 (1), 14–26.
- Kaster, R. A. 2001. The Dynamics of ‘Fastidium’ and the Ideology of Disgust. In *Transactions of the American Philological Association*, 131, 143–89.
- Kay, C. 1984. The Historical Thesaurus of English. In R. R. K. Hartmann (ed.), *Lexeter '83 Proceedings*, 87-90. Tübingen: Niemeyer.
- Kingsbury, P., M. Palmer, and M. Marcus. 2002. Adding Semantic Annotation to the Penn Treebank. In *HLT '02: Proceedings of the second international conference on Human Language Technology Research*, 262–256.
- Kutscher, S. and D. A. Werning. 2014. *On ancient grammars of space: linguistic research on the expression of spatial relations and motion in ancient languages*. Berlin/Boston (Mass.): De Gruyter.
- Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. London: The University of Chicago Press.
- Lecaudé, P. 2010. *La notion de puissance: les équivalents latins du grec δύναμις*. Thèse dactylographiée soutenue le 11 décembre 2010 à l'Université de Paris-Sorbonne.
- Lindenlauf, A. 2004. The Sea as a Place of No Return in Ancient Greece. In *World Archaeology* 35 (3), 416–433.

- Minozzi, Stefano. 2017. Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval. In P. Mastrandrea (ed.), *Strumenti digitali e collaborativi per le scienze dell'antichità*. Venezia: Edizioni Ca' Foscari.
- Moussy, C. 1999. Les vocables latins servant à désigner le sens et la signification. In M. Baratin and C. Moussy (eds.), *Conceptions latines du sens et de la signification*, 13- 27. Paris: PUPS.
- Moussy, C. 2005. Champ lexical et polysémie: l'expression du pardon en latin. In *Journal of Latin Linguistics*, 9(1), 359–370.
- Moussy, C. 2007. Ambiguus, ambiguitas, anceps, utroqueuersus dans le vocabulaire de l'ambigüité. In C. Moussy and A. Orlandini (eds.), *L'ambigüité en Grèce et à Rome. Approche linguistique*, 57–64. Paris: PUPS.
- Nash, J. M. 2018. *Rulers of the Sea – Maritime Strategy and Sea Power in Ancient Greece 550-321 BC*. The Australian National University: PhD Thesis.
- Nikoloska, A. 2012. The Sea Voyage of Magna Mater to Rome. In *Histria Antiqua* 21, 365–371.
- Reed, C. M. 2003. *Maritime Traders in the Ancient Greek World*. Cambridge/New York: Cambridge University Press.
- Wild, K. 2010. Angelets, trudgeons and bratlings: the lexicalization of childhood in the Historical Thesaurus of the Oxford English Dictionary. In M. Adams (ed.), *Cunning passages, contrived corridors: Unexpected essays in the history of lexicography*, 289–308. Monza: Polimetrica.
- Wilkinson, Troy. 2020. *The Indian Ocean Trade and the Roman State*. University of Wales Trinity Saint David: MA Thesis.
- Zamboni, A. 2008. Sul lessico latino, tardo e neolatino dell'alienazione mentale. In R. Wright (ed.), *Latin vulgaire–latin tardif VIII: actes du VIIIe Colloque international sur le latin vulgaire et tardif, Oxford, 6-9 septembre 2006*. Hildesheim/New York: Olms-Weidmann.



## Where did *wer* go?

### Searching for s-curves in lexical change from Old English to Middle English

To refer to a male adult, speakers of Present Day English have several lexical items to choose from (e.g., *man*, *guy*, *dude*, *fella*, *bloke*, *gentleman*, and *geezer*). However, variation within this semantic field is not new. According to *The Thesaurus of Old English*, there were at least 25 lexical items which denoted ‘male adult’ in Old English (e.g., *ceorl*, *guma*, *man*, *wer*) which could occur in referentially comparable contexts, as in (1). To examine the evolution of this onomasiological set from Old English and Middle English, the present study uses variationist quantitative methods, addressing two research questions. First, what was the distribution of third-person male adult nouns referents in Old English and Middle English? In other words, which variants were most frequent, in which contexts did they occur, and how did their ranking change over time? Second, is there any evidence to suggest that these variants were conditioned, constrained, or influenced by any attested intra- or extralinguistic factors of variation?

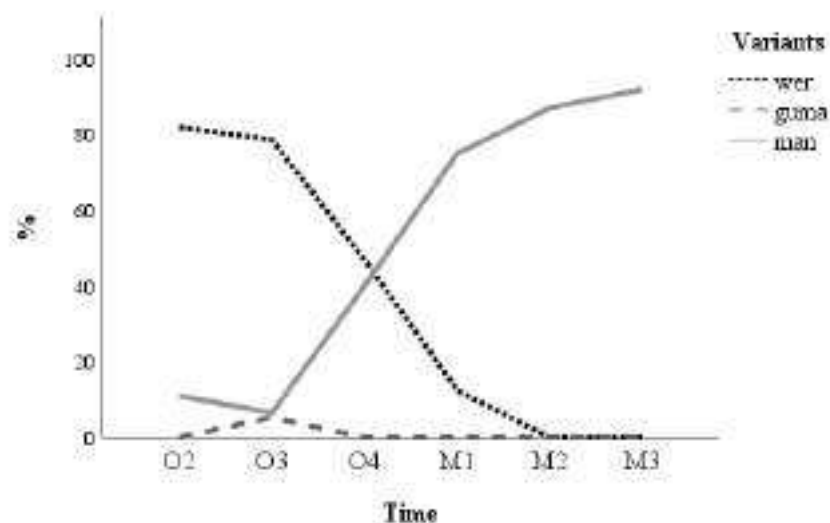
The *Helsinki Corpus of English Texts* (Rissanen et al., 1991) was used as the principal source of linguistic data. A list of third-person male adult noun referents was compiled using previous literature (e.g., Kleparski, 2005), dictionaries (e.g., Bosworth & Toller, 1882), and thesauruses (*The Historical Thesaurus of English*, *A Thesaurus of Old English*). Search queries were run to find instances of these variants in the corpus, which were subsequently downloaded and manually inspected for the removal of any functionally non-equivalent tokens (e.g., vocatives of address). Socio-historical context, as well as comparisons with Latin-based originals for translated texts, when available, were used to ensure only tokens with overt male referents were included in the envelope of variation. Each token was coded for both intra- (e.g., alliteration) and extra-linguistic factors (e.g., text type, provenance, time, text ID).

Results from the analysis demonstrate a significant shift in frequency from the favored variant *wer* in Old English to *man* in Middle English, a type of lexical replacement which coincides with collocational and frequency changes. As *wer* decreases, *man* takes on the former function of *wer*, with the diachronic shift in frequency following a prototypical s-curve distribution (Figure 1). Multivariate analyses using Rbrul (Johnson, 2009) found text type and text provenance to significantly constrain lexical choices, with certain variants (e.g., *rinc*, *scealc*, *knizt*) occurring more frequently in verse texts than prose texts on account of alliterative and metrical requirements.

Findings from the study are interpreted in the context of ongoing scholarship on lexical variation and change. First, the shift from *wer* to *man* illustrates a clear example of lexical replacement. As *wer* is replaced, it temporarily retreats to use as part of the semantic field ‘married man’ which later too is usurped by a competing lexical item *husband*. Second, although linguistic change does not have to follow an s-curve pattern, the shift from *wer* to *man* follows an s-curve trajectory, which is often used as a diagnostic for lexical replacement. Third, while work using apparent time data (Chambers, 1995) or short periods of time (Grieve et al., 2017) point to the applicability of s-shaped trajectories for lexical change, the present analysis of the semantic field of third-person male adult noun referents over approximately six hundred years adds a diachronic dimension to this discussion. Finally, in line with synchronic work on lexical variation, the present study shows how intra- and extra-linguistic factors harmoniously affect lexical choices and subsequently change diachronically.

- (1) (a) Ond on ðone ylcan dæg Crist gereorde fif ðusenda **wera** of fif hlafum ond of twam  
 and on that same day Christ fed five thousand men of five bread and of two  
 fixum, eac wifum ond cildum þara wæs ungerim  
 fish also women and children which was uncountable  
 ‘And on that same day, Christ fed five thousand men, with five loaves of bread and two  
 fish. In addition, he also fed women and children, of which there were many’  
 [Old English Martyrology, 950-1050]
- (b) on þære fyrde wæron þe ferdon fram Egipte, sixhund þusend **manna** butan wifum  
 in the army were which traveled from Egypt six-hundred thousand men except women  
 7 cildum  
 and children  
 ‘In that army, there were 600,000 men who travelled from Egypt, and that number  
 does not include women and children’ [Ælfric’s Letter to Sigeward, 1050-1150]
- (c) ðonne onwæcneð eft wineleas **guma**  
 then awakens again friendless man  
 ‘Then the man without any friends woke up’ [The Wanderer, 950-1050]

**Figure 1.** Frequency of *wer*, *guma*, and *man* from Old English to Middle English



### Selected References

- Bosworth, Joseph & Toller, Northcote. (1882). *An Anglo-Saxon Dictionary: Based on the Manuscript Collections of the Late Joseph Bosworth*. Oxford: Clarendon Press.
- Chambers, Jack. (1995). The Canada-US border as a vanishing isogloss: the evidence of chesterfield. *Journal of English Linguistics*, 23(1-2), 155-166.
- Grieve, Jack., Nini, Andrea & Guo, Diansheng. (2017). Analyzing lexical emergence in Modern American English online. *English Language & Linguistics* 21(1), 99–127.
- Kleparski, Grzegorz. (2005). Towards the semantics of Middle English synonyms of MAN. *Studia Anglica Resoviensia* 3, 88-95.
- Rissanen, Matti, Kytö, Merja., Kahlas-Tarkka, Leena, Kilpiö, Matti., Nevanlinna, Saara, Taavitsainen, Irma, Nevalainen, Terttu & Raumolin-Brunberg, Helena. (1991). *Helsinki Corpus of English Texts: Diachronic and Dialectal* (Helsinki).

### The PRESENTATIVE > DEMONSTRATIVE Grammaticalization Pathway in Arabic

In their recent survey of lexical sources for the grammaticalization of demonstrative forms, Heine et al. (2020) identify the development of imperative verbs of perception to demonstratives as a cross-linguistically attested pathway. As a component of that discussion, they refer to the proposed origin of the French demonstrative *ce* in an earlier presentative construction *ecce ille* ‘Behold that!’ (cf. van Gelderen 2011), noting that, “while not a verb, *ecce* has a meaning similar to that of the imperative form of a perception verb” (Heine et al. 2020: 420) and proceeding with their analysis on that basis. In this presentation, I move to expand on this latter observation by adducing additional corroborating data from varieties of Arabic, and to discern in finer detail the relation between PRESENTATIVE > DEMONSTRATIVE and PERCEPTION VERB > DEMONSTRATIVE grammaticalization pathways. In so doing, I identify conceptual characteristics of the proposed diachronic sources that are especially consonant with established precepts of synchronic demonstrative function, which speak to the broader viability of a PRESENTATIVE > DEMONSTRATIVE grammaticalization pathway beyond the specific contexts of the languages here considered.

The material for this discussion comes primarily from the review and synthesis of three Arabic demonstrative series previously proposed in to originate in presentative predications, but not till now considered under a unified, concentrated lens. These include the Classical Arabic *hāḏā* proximals (Magidow 2013), the Egyptian Arabic *dawwa* proximals/*dukha* distals (Leddy-Cecere 2021), and Libyan Arabic *āhwa* proximals (Pereira 2008). Extending beyond earlier accounts, I will show that each of the above – while distinct from the others in time, place and etymology – may be plausibly analyzed as originating in a topic-dislocated presentative structure of the type ‘Behold it, X!’/‘X, behold it!’, through processes of grammaticalization (desemanticization, extension and decategorialization – cf. Heine 2007) and rebracketing. Classical Arabic *hāḏā* and Egyptian *dawwa/dukha* represent “reinforcing” developments that incorporate and modify an inherited demonstrative element in the grammaticalizing source construction, while Libyan *āhwa* does not.

Building on the observation preliminarily voiced by Heine et al., I assert that the linkage between PRESENTATIVE > DEMONSTRATIVE developments like these and those deriving demonstratives from imperative verbs of perception consists explicitly not in the *semantic* dimension of their source meanings, but rather in the shared *pragmatic* status of those sources as directive speech acts that implore an addressee to attend to a given demonstratum. This directive nature thus contributes to the deictic component of the demonstrative function – the identification/demonstration to the addressee of an index – while the inherited pronominals involved in all examples considered here contribute toward the demonstrative function’s classificatory component – assisting the addressee’s ultimate attachment of that index to a referent (cf. Nunberg 1993).

As a direction for future research, I will briefly consider the implications of these findings for a novel interpretation of two further Arabic demonstrative series – the Classical Arabic *ḏālika* and North Fertile Crescent Arabic *hāk* distals – as grammaticalizing from dative presentative/offertive constructions (‘Here you go, X!’). While more data remains to be gathered (particularly in the Classical Arabic case, for which a rich textual record survives), the entailment of a directive to attend in a directive to take/receive supports the viability of such a pathway in light of the PRESENTATIVE > DEMONSTRATIVE cases already considered, and further work in this area may serve to illuminate aspects of the interflow between person- and distance-oriented distinctions in demonstrative development.

## References

- van Gelderen, E. (2011). “The grammaticalization of agreement.” In H. Narrog & B. Heine (eds.), *The Oxford Handbook of Grammaticalization*, 491-501. Oxford: Oxford University Press.
- Heine, B. (2007). “Grammaticalization.” In B. Joseph & R. Janda (eds.), *The Handbook of Historical Linguistics*, 575-601. Oxford: John Wiley & Sons, Ltd.
- Heine, B., T. Kuteva, H. Long, H. Narrog & F. Wu. (2020). “Where do demonstratives come from?”. *STUF* 73(3), 403-434.
- Julia, M. (2020). “About Latin *Ecce* ‘Behold! Lo! See! There!’ and Some Ancient Languages Presentatives. *International Journal of Language and Linguistics* 8(1), 17-23.
- Leddy-Cecere, T. (2021). “Interrogating the Egypto-Sudanic Arabic connection.” *Languages* 6, 123.
- Magidow, A. (2013). *Towards a Sociohistorical Reconstruction of Pre-Islamic Arabic Dialect Diversity*. Austin: The University of Texas at Austin doctoral dissertation.
- Nunberg, G. (1993). “Indexicality and Deixis.” *Linguistics and Philosophy* 16(1), 1-43.
- Pereira, C. (2008). *Le parler arabe de Tripoli (Libye): Phonologie, morphosyntaxe et catégories grammaticales*. Paris: INALCO Doctoral dissertation.

### Syntactic change and DLM in German: a corpus study

This study argues that the variation in the placement of relative clauses (RC) in German can be explained by the principle of dependency length minimization (DLM), which states that languages tend to place syntactically and semantically related words close together (e.g., Gibson, 2000, Futrell et al., 2015). Although Gildea & Temperley (2010) ascribe only a weak effect of DLM to German grammar, my diachronic corpus study will show that we can see a strong effect of DLM by looking at the RC position in the history of written German. I will testify to a two-stage change: The syntactic complexity reflected by a high number of intraposed RC first increases in the 17th and 18th centuries due to the growing standardization of the written language, which favors verb-final structures and reaches its peak at the end of the 18th century (Admoni, 1967), before it decreases continuously so that extraposition becomes the most frequent word order.

According to Hawkins (1994, 2004), word order variation in sentence production can be explained by processing economy constraints. Specifically, he states that, given alternative word orders within a construction, the one that allows faster recognition of the immediate constituents is expected to be preferred. Gibson (1998, 2000) takes a similar approach concerning sentence comprehension. He establishes that the memory cost is higher the more incomplete syntactic dependencies one has to keep track of during sentence processing. In German word order, auxiliary and main verb are placed discontinuously in a sentence. According to DLM, there should be a tendency to reduce the distance between the verbs by outsourcing heavy NPs behind the finite verb to the end of the sentence to avoid an overload of the working memory capacity.

Addressing the question of whether DLM plays a role in explaining the changing RC position in German, I investigate the placement of heavy NPs and PPs with relative clause modifiers from newspaper texts from 1600 until Present-day German. The relevant structures are intraposition where the RC is placed adjacent to its head noun (1) and extraposition where the nominal head appears preverbally and the RC occurs at the right edge of the sentence (2).

- (1) Von deß Orators Leuten **sollen** 2. [<sub>PP</sub> vnter einem tumult/ [so die Türcken deß Nachts in jhrem Losament angefangen]] **vmbkommen sein** (1609: Relation)  
‘Two people of the orator are said to have been killed during a tumult that the Turks started in their accommodation at night.’
- (2) sonst **weren** in Spannia auch [<sub>NP</sub> Ampassatores vom König Matthiasen vnd Hertzogen von Savoia] **angelangt** / [so beim König schon Audienz gehabt] (1609: Relation)  
‘Furthermore, ambassadors of King Matthias and of the Duke of Savoy had also arrived in Spain, who had already had an audience with the king.’

An intraposed word order with a long dependency (i.e., a high amount of language material between the verbs) leads to high working memory costs and therefore runs into the risk of processing difficulties. Differently, extraposition can facilitate sentence processing because the dependency length (DL) between the verbs is minimized. Complicating matters, however, is that the DL between RC and its antecedent is increased with extraposition, thus, presenting potentially competing motivations.

To measure syntactic complexity, I consider the factors (i) RC length, (ii) distance between the discontinuous verbs, and (iii) distance between the RC and its antecedent. My results show a strong effect of DLM on the development of German word order preferences: both DLs (between RC and its antecedent and between the separated verbs) decrease over time significantly. On the one hand, extraposition becomes the most frequent word order which leads to reduced distances between the discontinuous verbs over time. On the other hand, the DL between the extraposed RC and the antecedent decreases because it becomes the norm that no more than two words can occur here. This can be explained by the processing pressure in spoken modalities that influence the written standard in newspapers over time increasingly.

## References

- Admoni, Wladimir G. 1967. Der Umfang und die Gestaltungsmittel des Satzes in der deutschen Literatursprache bis zum Ende des 18. Jahrhunderts. *Beiträge zur Geschichte der deutschen Sprache und Literatur*, 89, 144–199.
- Futrell, Richard, Kyle Mahowald, & Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33), 10336-10341.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1–76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95–126.
- Gildea, Daniel, and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science* 34, 286–310.
- Hawkins, John A. 1994. A Performance Theory of Order and Constituency. *Journal of Linguistics*, 33(1), 171–212.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.



## Spread the German new(s): third-person reflexive *zich* in 17<sup>th</sup>-century Dutch newspapers

In Middle Dutch, reflexivity was commonly expressed by using personal pronouns (third person singular: *haar* ‘her’ or *hem* ‘him’; plural: *haar* ‘them’ or *hen/hun* ‘them’), sometimes followed by *zelf* ‘self’ (e.g. Mooijaart & Van der Wal, 2011: 46). Although attested from the 10<sup>th</sup> century onwards, especially in the eastern parts of the Low Countries, the specific reflexive form *zich*(*zelf*), currently used in Standard Dutch in both singular and plural third person, only became the standard third-person reflexive pronoun during the seventeenth century (e.g. Van der Sijs, 2021: 429).

This reflexive pronoun is generally accepted as being a grammatical borrowing from the neighbouring High German *sich* (e.g. Postma, 2011: 139; Van der Sijs, 2021: 429-430), yet the factors determining its expansion in the Dutch speaking world still remain subject of discussion (as summarised in Bennis, 2005). On the one hand, research suggests diffusion from above, through formal written texts, in particular religious texts from Germany (e.g. Hermodsson, 1952; Van der Wal & Van Bree, 2008: 214-215; Nobels, 2013: 115-121). On the other hand, Boyce-Hendriks’s study (1998: 209-224) suggests dissemination from below, through the informal spoken language of the many German immigrants settling in the Low Countries in the sixteenth and seventeenth centuries.

In this presentation, we intend to contribute to this ongoing debate by presenting the results of a study testing a third hypothesis: diffusion via newspapers. Newspapers as a genre came into existence in Germany in the early seventeenth century and quickly found their way to the Low Countries (Van Oostendorp & Van der Sijs, 2019: 23-76). Although we do not have exact circulation figures, circumstantial evidence indicates that Dutch newspapers, which were set in Dutch and not in Latin, were meant for and read by many people, from all social classes (Van Groesen, 2016). Newspapers can therefore be seen as the first mass medium (Van Oostendorp & Van der Sijs, 2019: 61), and could thus have served as the ideal vehicle for the dissemination of *zich*, especially given their close connection to German newspapers (e.g. Der Weduwen, 2017).

We test this hypothesis using the new Couranten Corpus (2022), a corpus containing Northern Dutch newspapers from all years between 1618 to 1700 (ca. 19 million words). Within these newspapers, we examine what third-person singular and plural reflexive constructions were used: the previously mentioned uses of personal pronouns and *zich* (both with and without *zelf*), Low German *sick* (Van der Wal, 2018), and possessive pronouns followed by *eigen* ‘own’ or *zelf* ‘self’ (see Weijnen, 1965: 49). We track the distribution of these constructions both over time as well as regarding the place of origin of the article that contains it, focussing in particular on German places of origin. Not only did newspapers originate in Germany, Dutch newspapers also contained many news reports from German speaking areas, and, at the start of the century, news reports were sometimes directly translated from German (Van Oostendorp & Van der Sijs, 2019: 29-31). Although we do not know the exact extent of this German ‘borrowing’, it is safe to assume that early news reports from German speaking areas were either translated from German, or came from a correspondent who lived in Germany and had heard the news from a German source, which both could have significantly influenced the use of *zich*, especially given the presumed lack of editing of incoming news reports (Demske, 2022). Comparing those news reports to domestic news and news reports from other language areas, as well as changes in their distributions over time, will thus provide valuable insight into (1) the influence of German on the use of *zich*, and (2) the possible role of newspapers in spreading this new form.



## References

- Bennis, H. (2005). Een Duitse expansie. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 121(3), 201-210.
- Boyce Hendriks, J. (1998). *Immigration and linguistic change: a socio-historical linguistic study of the effect of German and southern Dutch immigration on the development of the northern Dutch vernacular in 16th/18th century Holland*. Madison. (unpublished dissertation).
- Couranten Corpus (version 1.0) (May 2022) [Online Service]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-u9>
- Demske, U. (2022). Variation Across Newspapers in Early Modern German: Degrees of Syntactic Complexity. *Journal of Historical Syntax*, 6(13-18). <https://doi.org/doi:10.18148/hs/2022.v6i13-18.136>
- Der Weduwen, A. (2017). *Dutch and Flemish Newspapers of the Seventeenth Century, 1618-1700 (2 Vols.)*: Brill. <https://doi.org/https://doi.org/10.1163/9789004341890>
- Hermodsson, L. H. (1952). *Reflexive und intransitive Verba im älteren Westgermanischen*. Uppsala: Almqvist und Wiksell.
- Mooijaart, M., & Van der Wal, M. J. (2011). *Nederlands van Middeleeuwen tot Gouden Eeuw* (2nd rev. ed.). Nijmegen: Uitgeverij Vantilt.
- Nobels, J. (2013). *(Extra)Ordinary letters: A view from below on seventeenth-century Dutch*. University of Leiden. Utrecht: LOT.
- Postma, G. (2011). Language contact and linguistic complexity—the rise of the reflexive pronoun *zich* in a fifteenth-century Netherlands border dialect. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical Change: Origins, Nature, Outcomes* (pp. 139-159): Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199582624.003.0007>
- Van der Sijs, N. (2021). *Taalwetten maken en vinden: het ontstaan van het Standaardnederlands*. Gorredijk: Sterck & De Vreese.
- Van der Wal, M. J. (2018). Early Modern migrants in a language contact setting: Characteristics of the Dutch Heusch correspondence (1664–1665). *Journal of Historical Sociolinguistics*, 4(2), 253-280. <https://doi.org/doi:10.1515/jhsl-2017-0029>
- Van der Wal, M. J., & Van Bree, C. (2008). *Geschiedenis van het Nederlands*. Houten: het Spectrum.
- Van Groesen, M. (2016). Reading Newspapers in the Dutch Golden Age. *Media History*, 22(3-4), 334-352. <https://doi.org/10.1080/13688804.2016.1229121>
- Van Oostendorp, M., & Van der Sijs, N. (2019). *Een mooie mengelmoes: meertaligheid in de Gouden Eeuw*. Amsterdam: Amsterdam University Press.
- Weijnen, A. A. (1965). *Zeventiende-eeuwse taal* (5th ed.). Zutphen: N.V. W.J. Thieme & Cie.

*The use of “thanks” and “to thank” in Old Saxon and Old High German*

Modern German counts on different nouns, verbs, adjectives, and prepositions to express gratitude. Many of these elements, such as the noun *Danke* (thanks) and the verb *danken* (to thank) were already common as early as in the 8th century (Pfeifer 1993). However, the use of these nouns and verbs in expressions of thankfulness have been studied only synchronically and in comparative studies (Zborowski 2005; Siebold 2021). The goal here is to provide a qualitative analysis of the expression of thanking with *Danke* and *danken* at the earlier stage of the history of German, filling a long-standing gap in the literature. More specifically, this investigation seeks to answer the following research questions:

1. How were *Danke* and *danken* used in Old Saxon and Old High German?
2. What can the use of *Danke* and *danken* tell us about historical expressions of thanking in Old Saxon and Old High German?

To answer these questions, I examined all the instances of *Danke* and *danken* in a corpus of religious and secular Old Saxon and Old High German texts taken from the *Referenzkorpus Altdeutsch* (Reference corpus Old German). The targeted structures of this investigations are as showed in (1), (2), (3), and (4):

- (1) *ôlat sagde themu the these uuerold giscôp*  
 thank said to the one that this world created  
 ‘He said thank to the one who created this world’  
 (Hêliand, XLIX, 4091)
- (2) *endi gode **thancode**, sagde them ôlat*  
 and god thanking, said (he) to him thank  
 ‘He was thanking god and said thank to him’  
 (Hêliand, LVI, 4633)
- (3) *er nú ana wánc hábet fora góte **thanc***  
 he now without doubt has before god thanc  
 ‘He has to say thanks to god now without doubts’  
 (Evangelienbuch, 20, 6)
- (4) *thir **thánkon** mit wórton joh mit wérkon*  
 you I thank with words and with deeds  
 ‘I thank you with words and deeds’  
 (Evangelienbuch, 24, 91)

The data were extracted using the lemma search function offered by the online corpus. The frequency of the instances and the addressers and addressees involved in the expressions of thanking with *Danke* and *danken* were analyzed.

Regarding the first question, a total of 76 instances of *Danke* and *danken* were found, and most of them were in religious texts (71 in total). The data show that the words *Danke* and *danken* were mostly used to express gratitude, but there were instances in which they were found with the meanings of “reward/to reward” and “praise/to praise.” These results show how, Old Saxon and Old High German

speakers relied on the same linguistic elements to carry out different speech acts and how, probably, these speech acts may have been perceived similar to each other.

Turning now to the second research question, the analysis of the addressers and addressees involved in these instance shows how, in the texts selected, and independently of their genre (religious vs. secular), the expression of thanking (but also praising, and reward) with *Danke* and *danken* could only be found when the addresser was in lower social position than the addressee. These results seem to suggest that, in Old Saxon and Old High German, these types of expression could be carried out only when addressees and addressers had an asymmetrical relationship.

In conclusion, these results suggest that use of *Danke* and *danken* in expressions of thanking (but also in praising and rewarding) were restricted to specific communicative settings in which the social status of the speakers played a major role.

### **Selected references**

Pfeifer, Wolfgang. 1993. *Etymologisches Wörterbuch des Deutschen. Digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache.*

<https://www.dwds.de/d/wb-etymwb>.

Siebold, Kathrin. "Implicit and explicit thanking in Spanish and German." *Linguistic Insights-Studies in Language and Communication* 132 (2012): 155-172.

Zborowski, Piotr. 2005. *Dankbarkeit vs. Höflichkeit und sprachliche Routine. der Dankakt im Schwedischen verglichen mit Polnisch und Deutsch.* Wydawnictwo Naukowe UAM.

## What is *ke* and if so how many?

### The Persian modal particle *ke* and its diachronic development

**Keywords:** Modal particles, discourse markers, common ground, uncontroversality, grammaticalisation, pragmaticisation, Indo-Iranian, Turkic

In this paper, I will argue that the polysemous Persian conjunct and relative pronoun *ke* (PIE \*k<sup>w</sup>ís/\*k<sup>w</sup>ós) has acquired the additional function of a modal particle in colloquial Persian. This stands in direct contrast to previous accounts which have rendered the particle a focus particle or an emphatic marker (Oroji and Rezaei 2013; Sadat-Tehrani 2002), claims that can be refuted based on the examples and restrictions of *ke* presented here. Due to their impalpable nature, modal particles (henceforth MPs) and discourse markers (DMs) were long neglected in linguistic research but have increasingly gained traction in recent years. While both MPs and DMs express the attitudes of the speaker towards a proposition - for which they have also been called "Würzwörter" ('words that add spice') in German - and do not change its truth value, MPs differ from DMs in that they are often well integrated in a phrase and have scope over only the proposition they appear in. The MP at hand *ke* can both appear after the topic as well as at the right periphery. Akin to the German MP *doch*, the basic properties of *ke* can be summarized as follows: ADVERSITY, (also sometimes referred to as CORRECTION (Döring 2016)), UNCONTROVERSIALITY (also described as COMMON GROUND/BACKGROUND) (Grosz 2016; Diewald 2006), and what I refer to as RELEVANCE/SALIENCE, such that:

$[[ke]](p) = p$  and the current question  $q$  stands in conflict with  $p$  which the speaker renders uncontroversial/part of the common ground but is retrieved for purposes of salience (cf. (Grosz 2016))

Consider the following examples: Person A: Shall I make lamb curry for Ali? Person B:

- |  |   |
|--|---|
| <p>(1) Ali <i>ke</i> gusht nemikhore.<br/>Ali <i>ke</i> meat eat:NEG.3SG.PRS<br/>(But) Ali doesn't eat meat.</p> | <p>(2) Ali isst doch gar kein Fleisch.<br/>Ali eat.3SG <i>doch</i> no meat.<br/>(But) Ali doesn't eat meat.</p> |
|--|---|

In the example given, *ke* highlights the adversity of  $p$  toward the current question  $q$  (eating a dish containing meat) and that person B thinks  $p$  should be known to A (common ground) but that  $p$  was apparently not salient enough or momentarily forgotten (relevance). In a quest to answer the hitherto unanswered question of how *ke* might have acquired the function of a MP, I will propose a diachronic development from a conjunction through a process of grammaticalisation along the path of (referential function) → (text-connective function) → (discourse function) as suggested by Diewald (2006) based on (Traugott 1995, 1999). I will argue that this development was especially facilitated by the deictic function of conjuncts as described by Hentschel (1986) and Diewald (2006) and the coordination of Persian subjunctive sentences in the form [A co][B] (Haspelmath 2004).

As shall become clear, the existence of an Old Turkish emphatic particle *är-ki* seems to seriously challenge this theory at first as it has been argued that the Turkish modal particle *ki* is derived from this OldkTurkish particle (Karakoc 2010) thereby insinuating that the Persian MP is in fact borrowed from Turkish and not vice versa. However, this theory can be dismissed on the basis of further, comparative evidence from the North-Afghan Badākhshāni dialect of Persian as well as due to the strong anchoring of second position MPs in other Indo-Iranian languages such as Pashto *kho*, Urdu *to* and Marathi *tər* (Bayer 2018; Deo 2022). Not only are these particles almost identical in function to *ke* but, coincidentally, are all also used as conjuncts denoting "but" and "then"/"so" respectively. This strongly suggests similar paths of grammaticalisation triggered perhaps by an Indo-Iranian predisposition for this kind of development. Even if one is to dismiss a development of MPs *out of* conjuncts, one cannot deny the intricate relationship that exists between the two, a matter worthy of further investigation.

## References

- Bayer, Josef (2018): 'A Comparative Note on the Bangla Particle "to" and the German Particle "doch"', *Personal Essays, Poems, and Scholarly Papers* .
- Deo, Ashwini (2022): 'Could be stronger: Raising and resolving questions with Hindi =to', *Language* **98**(4), 716–748.
- Diewald, Gabriele (2006): 'Discourse particles and modal particles as grammatical elements', *Approaches to Discourse Particles* pp. 403–425.
- Döring, Sophia (2016): *Modal Particles, Discourse Structure and Common Ground Management. Theoretical and Empirical Aspects..* Phd, Humboldt-Universität zu Berlin.
- Grosz, Patrick G (2016): Discourse particles. In: D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann and T. E. Zimmerman, eds, *The Wiley Blackwell Companion to Semantics*. Wiley, Oxford.
- Haspelmath, Martin (2004): Coordinating constructions. In: M. Haspelmath, ed., *Typological Studies in Language*. Vol. 58, John Benjamins, Amsterdam/Philadelphia, pp. 3–39.
- Hentschel, Elke (1986): *Funktion und Geschichte deutscher Partikeln: Ja, doch, halt und eben*. Max Niemeyer Verlag, Tübingen.
- Karakoc, Birsal (2010): Mutmaßungen über die Etymologie des türkischen Suffixes KI. In: H. Boetschoten and J. Rentzsch, eds, *Turcology in Mainz*. 82 edn, Wiesbaden, pp. 101–112.
- Oroji, Mohammad Reza and Amir Rezaei (2013): 'Exploring 'ke' as a Focus Particle in Persian from both Form and Function Points of View', *Australian Journal of Linguistics* **33**(1), 76–84.
- Sadat-Tehrani, Nima (2002): 'The indifference-ke construction in modern conversational Persian', *Linguistica atlantica* **24**, 43–69.
- Traugott, Elizabeth Closs (1995): Subjectification in grammaticalization. In: S. Wright and D. Stein, eds, *Subjectivity and Subjectivisation*. Cambridge University Press, Cambridge, pp. 31–54.
- Traugott, Elizabeth Closs (1999): The Rhetoric of Counter-Expectation in Semantic Change: A Study in Subjectification. In: A. Blank and P. Koch, eds, *Historical Linguistics and Cognition*. Mouton de Gruyter, Berlin, New York, pp. 61–89.

## The development of future-referring constructions (in Indo-European languages)

### Abstract

The investigation of the evolution of grams encoding futurity both in individual languages or language families (cf., e.g., Fleischman 1982, Botne 1998, Whaley 2000) and in typological comparison (e.g., Ultan 1978, Dahl 1985, Bybee & Dahl 1989, Heine & Kuteva 2002: 331) has focused on their grammaticalization from various source constructions with etyma such as ‘will’, ‘have’, ‘become’ and motion verbs (‘go’, ‘come’). According to the “source determination” hypothesis, the origin of a gram determines its syntactic and semantic restrictions in its further development. Another frequent source for future-referring verb forms is hypoanalysis, i.e., a process by which peripheral and/or contextually conditioned functions of a construction, or a categorial opposition, become its inherent property due to a reduction in the inherited functional range (Croft 2000: 126f.); typical cases are North Slavic perfective present (PFV.PRS) > pfv. future (Wiemer 2020: 275f.) and the Latin futures in *-ē-* (e.g., *legēs* ‘you will say’), and *-b-* (continuing the SBJV of the root *\*b<sup>h</sup>uH-* ‘be’, e.g., 3SG *amā-bit* ‘will love’) and the future of the copula (*erō, eris, erit* etc.), which are based on a preceding subjunctive (Meiser 2006: 199, Weiss 2020: 441f.). This raises the question whether these two types of futures show systematic differences in their grammatical behaviour, e.g., regarding the frequent function of futures to also encode deontic, epistemic and other kinds of modality.

Using a dataset from Indo-European languages comprehensive over time and space, the paper will discuss the following hypotheses: (a) Contrary to the assumption of source determination, morpho-syntactic and semantic/pragmatic restrictions related to the source construction only pertain at an intermediate stage of the development of futures, whereas fully developed futures overcome these and display a characteristic range of functions (first of all, modal and illocutionary functions) independent of their source; (b) Futures arisen from hypoanalysis differ from futures based on grammaticalization in that they are stable with respect to their initial restrictions (e.g., regarding aspect); (c) The diachronic relation between future meaning and modal, especially epistemic, function is not unidirectional, i.e., either may precede the other. This applies to both types of futures; (d) Futures either stay what they are, or they disappear, but they do not develop into anything else (a “post-future” stage). In particular, if epistemic readings develop as a “sub-function” of futures, they do not oust the future reading as the default meaning.

Assumption (a) is largely confirmed by our dataset. Hypothesis (b) is motivated from parallel observations concerning, e.g., English futures (*going to* contracts to *gonna*, the latter is no longer available as reference to physical movement; *will* no longer codes volition, apart from archaic formula) and PFV.PRS > pfv. future in North Slavic: while the future reading is salient (as a default), non-deontic modal readings are still widely available. The latter, in turn, are among the dominant readings of PFV.PRS in South Slavic, while South Slavic futures based on WANT no longer code volition (see Engl. *will*) and are now morphologically or lexically distinct from WANT (see (1)). Concerning (c), languages differ as for whether their well-established futures are used for epistemic judgment referring to the moment of speech or not, regardless of the diachronic pathway of the future: for instance, Germ. *werden* and Span. *tener* do allow for such usage (see (2)), while Engl. *will*, futures in North Slavic and in Lithuanian do not. Moreover, Balkan Slavic has epistemic uses of future markers (see above), also in combination with *da*-clauses (= irrealis marking; see (3)), but the chronological relation to future meaning requires clarification, also in comparison to Engl. *will* (cf. Ziegeler 2006) and High Alemannic (vs Standard German) *werden* (*Schweiz. Idiotikon* 16, 1344, 1346-7, cf. <https://www.idiotikon.ch/woerterbuch/idiotikon-digital>). As for (d), no IE language shows a “post-future” stage for a “surviving” future, unless as a suppletive form in another paradigm (e.g., Span. 2SG.PRS *eres* ‘you are’ continuing Latin *eris* ‘you will be’; Lausberg 1972: 3.251, Penny 1993: 181; 2014: 191). We also discuss whether the lack of post-future stages might be an IE. feature.



### Examples

- (1) South Slavic futures based on WANT: Bulg. *šte*, Mac. *ke*, Srb.-Cr. inflected *ć-u*, *ć-eš*, *ć-e...* – distinct from contemporary WANT: Bulg. *iskam*, Mac. *sakam*, Srb.-Cr. *hoć-u*, *hoć-eš*, *hoć-e...*
- (2) Germ. *Er wird gerade seinen Vortrag halten* ‘He must be having his lecture right now’  
(personal knowledge)
- (3) Bulg. *Šte da ima poveče ot pedeset*. ‘S/He must be older than fifty.’  
(Tomić 2006: 476)

### References

- Botne, Robert. 1998. The evolution of future tenses from serial ‘say’ constructions in Central Eastern Bantu. *Diachronica* 15-2, 207-230.
- Bybee, Joan L. & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13, 51-103.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford, New York: Basil Blackwell.
- Fleischman, Suzanne. 1982. *The Future in Thought and Language. Diachronic evidence from Romance*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 1998. The semantic development of old presents: new futures and subjunctives without grammaticalization. *Diachronica* 15-1, 29-62.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.
- Lausberg, Heinrich. 1972. *Romanische Sprachwissenschaft. III. Formenlehre*. 2nd ed. Berlin, New York: Walter de Gruyter.
- Meiser, Gerhard. 2006. *Historische Laut- und Formenlehre der lateinischen Sprache*. 2. Aufl. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Penny, Ralph. 1993. *Gramática histórica del Español*. Barcelona: Editorial Ariel.
- Penny, Ralph. 2002[2014]. *A History of the Spanish Language*. 2nd ed. Cambridge: Cambridge University Press.
- Tomić, Olga Mišeska. 2006. *Balkan Sprachbund Morpho-syntactic Features*. Dordrecht: Springer.
- Ultan, Russell. 1978. The nature of future tenses. In: Greenberg, Joseph H., Charles A. Ferguson & Edith Moravcsik (eds.). *Universals of Human Language. Vol. 3. Word Structure*. Stanford: Stanford University Press, 83-123.
- Weiss, Michael L. 2009. *Outline of the Historical and Comparative Grammar of Latin*. Ann Arbor: Beech Stave Press.
- Whaley, Maryka Lynn. 2000. *The evolution of the Slavic ‘be(come)’-type compound future*. Ohio: The Ohio State University (PhD thesis).
- Wiemer, Björn. Grammaticalization in Slavic. In: Bisang, Walter & Andrej Malchukov (eds.): *Grammaticalization Scenarios: Cross-linguistic Variation and Universal Tendencies. Vol. 1: Grammaticalization scenarios from Europe and Asia*. Berlin, Boston: De Gruyter Mouton, 249-307.
- Ziegeler, Debra. 2006. Omnitemporal will. *Language Sciences* 28, 76-119.



## **A re-assessment of Early Runic Metrics**

This paper on early runic metrics systematically excludes inscriptions of the transitional period and the Viking Age. Thus, the focus is on the metricity of the unsyncopated, linguistically archaic runic inscriptions of the older period AD 150-500/550. This has the advantage of systematically excluding syncopated or partially syncopated inscriptions such as the Eggja stone and the Blekinge inscriptions. Due to their linguistic status, these inscriptions will have to be subjected to a separate study.

The author takes a minimalist standpoint by first taking the position of the *Advocatus Diaboli*. This means that metrical criteria of the older runic inscriptions should speak for themselves without being directly derived from or equated with later language stages with their elaborated metrical systems. In other words, the present approach attempts to avoid constrictions and direct comparisons with the Old Germanic literary languages, especially Old Norse and Old High German. This is so because the language typology of Ancient Nordic cannot directly be equated with Old Norse in terms of syllabic metrics (cf. Pascual 2016 on Old English). Sievers' typology (Sievers 1893) is also rejected here as a straightforward tool for analysis, since it is oriented towards the Old Germanic literary languages. This raises the basic question of scientifically verifiable regularities of early runic metrics that meet criteria of validity and reliability.

The leitmotif of this analysis is the fusion and interaction of language, sentence rhythm and alliterative metrics. Not entirely unexpectedly, the Germanic long line emerges as the basic unit of Ancient Nordic metrics, since it can be relatively reliably verified in a small group of older runic inscriptions. Finally, this unit is identified as the proto-long line of North-West Germanic metrics.

### **Keywords**

Early Runic metrics, Germanic long-line, proto-long-line, Germanic verse, older runic inscriptions, Sievers' metrical types, formulaicity, resolution, heaviness requirement.

## Micro-level conflict in the productivity of anticausativization strategies Evidence from the history of Icelandic

The objective of this paper is, first, to argue that different morphosyntactic strategies used to form anticausatives are productive in different periods of Icelandic, a language characterized by rich inflection. Second, we show how shifts in anticausativization strategies coincide with factors in other domains of grammar (cf. Cennamo 2022). Our analysis is carried out on the basis of the diachronic variation exhibited by selected predicates in historical corpora (e.g. ONP, RMH). Expanding on Ottosson (2013), we identify five different causative-anticausative patterns in Old and Modern Icelandic (1a–e); the case frames for the relevant predicates are given within square brackets.

- (1) a. Weak-strong alternation: caus. *sökkti* ‘sank’ [Nom–Acc] : anticaus. *sökk* ‘sank’ [Nom]  
 b. *na*-verbs: caus. *braut* ‘broke’ [Nom–Acc] : anticaus. *brotnaði* ‘broke’ [Nom]  
 c. Case-Preserving Anticausativization (CPA): caus. *hvessti* ‘sharpened’ [Nom–Acc] :  
 anticaus. *hvessti* ‘got windier’ [Acc]  
 d. Lability: caus. *lokaði* ‘closed’ [Nom–Dat] : anticaus. *lokaði* ‘closed’ [Nom]  
 e. *st*-predicates: caus. *opnaði* ‘opened’ [Nom–Acc] : anticaus. *opnaðist* ‘opened’ [Nom]

Two of these strategies, (1a) and (1b), ceased being productive already in pre-Old Icelandic (before 1150 AD). Their unproductivity is reflected in the fact that occasionally some predicates may form an anticausative in more than one way (2a), or occur with double marking simultaneously (2b):

- (2) a. OIcel *sökk-ti-st* ‘sunk’, a weak form with *-st* for an older strong form *sökk* in (1a)  
 b. ModIcel *brot-na-ði-st* ‘broke’, with both *-na-* and *-st* instead of *brot-na-ði* in (1b)

The remaining three strategies (1c–e) show productivity to a varying degree in Modern Icelandic, engaging in a “micro-level conflict” against each other. We posit a hierarchy of strategies for the modern language such that the *st*-strategy (1e) is selected provided the relevant *st*-form does not already have another function (reflexive, reciprocal and denominals). The productivity of this strategy is clearly boosted by the ubiquity of the *st*-suffix with all kinds of verb formations. In cases where the *st*-form is dispreferred or unavailable, a different strategy (CPA or lability) is selected. The strategy of labile verbs (1d) seems to be gaining ground, having previously been rather limited (witness new verbs in the semantic domain of technology like *hlaða* ‘charge’ and *starta* ‘start’). In some cases its rise may be “accidental”, due to a common change called Nominative Substitution, whereby an oblique subject is replaced by a nominative (e.g. Svavarsdóttir 1982). However, even in Modern Icelandic, the emergence of new oblique subjects can still be triggered by CPA (1c).

In conclusion, the diachrony of anticausativization strategies in Icelandic suggests that in this highly inflected language, morphological marking with anticausatives ((1a–c), (1e)) is preferred over an unmarked option (1d). Nevertheless, various conflicting factors create a micro-level tension in the grammar, affecting the productivity of the different anticausativization strategies. Thus, Nominative Substitution replaces oblique subjects and the otherwise very productive *st*-suffix reaches a point of saturation due to its multifunctionality, leaving the labile option as the most viable one in Modern Icelandic.

## References

- Cennamo, Michela. 2022. Paths of change in the diachrony of anticausatives in Romance: the rise and fall of labiality. Paper presented at ICHL25, Oxford, 1–5 August 2022.
- ONP = *Ordbog over de danske sprog*. Copenhagen: Den Arnamagnæanske Kommission (eds). <<https://onp.ku.dk/onp/onp.php>>
- Ottosson, Kjartan. 2013. The anticausative and related categories in the Old Germanic languages. In Folke Josephson & Ingmar Söhrman (eds.), *Diachronic and Typological Perspectives on Verbs*, 329–381. Amsterdam: John Benjamins.
- RMH = Risamálheild [Gigaword Corpus]: Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. *Proceedings of LREC 2018*, 4361–4366, Japan. <<https://malheildir.arnastofnun.is/?mode=rmh2022>>
- Svavarsdóttir, Ásta. 1982. „Þágufallssýki“: Breytingar á fallnotkun í frumlagssæti ópersónulegra setninga [“Dative Sickness”: Changes in the use of subject case in impersonal sentences]. *Íslenskt mál og almenn málfræði* 4:19–62.

### The totalizing function of the Vedic particle *cid*

The interpretation of particles is often exceedingly difficult, especially in extinct languages (cf. Goldstein 2019:269–271). These problems become even more grave in an attempt to reconstruct the functions of a certain particle in Proto-Indo-European.

The enclitic particle  $*=k^h id$  can be safely reconstructed for Proto-Indo-European. Its most productive reflexes are attested in Indo-Iranian but other Indo-European branches exhibit reflexes as well (cf. Dunkel 2014:448–451). In order to reconstruct not only the form but also the functions of this particle properly, thorough synchronic examinations of its reflexes in the oldest stages of the languages in which it is reflected are necessary, in particular those in which it is used productively.

This paper is concerned with a detailed analysis of one reflex, namely the particle *cid* in the Rigveda, the oldest Indo-Iranian text. It is attested there 691 times. In the Rigveda, *cid* can fulfill a number of functions, among others that of an additive focus particle ‘even, also’ (e.g. Grassmann 1873:454f., Lühr 2017:283–285). This paper will concentrate on one function of this particle which until now has not received proper treatment in the literature, namely its totalizing function.

It is a well-known fact that the particle *cid* is cliticized to interrogative proforms in order to form indefinites, e.g. *kás cid* ‘some, any, every’. However, some scholars assume that also *cid* alone possesses such a function (e.g. Gonda 1954–1955:281). Thus, *cid* itself has also been translated as ‘all’ (e.g. Grassmann 1873:455). I will argue that although such a translation of *cid* is adequate in certain contexts such occurrences are to be differentiated from those of the particle in indefinite proforms. For even though indefinites formed by interrogatives and additive particles like *cid* are typologically widespread, additives themselves usually do not function as indefinites or universal quantifiers (König 2017:40; Ying 2017:218–226). I assume that instead of being a quantifier *cid* possesses a function which according to Forker (2016:84–86) is also attested for additive particles in other languages and may be called ‘totalizing’. This means that it emphasizes that all elements of a set are referred to, but I will argue that in spite of the resemblance to universal quantification, *cid* is not a genuine quantifier. I will show that this function can be identified after numerals, universal quantifiers, demonstratives, the pronominal adjective *anyá-* ‘other’ and possibly after multiplicative adverbs.

I will also discuss another context where I assume this function of *cid*, namely after *purá* ‘before, of old’. When *purá* occurs with a verb in the perfect or present it expresses a norm or habit in the past (Mumm 2004:55–61). There, I consider it to emphasize that the event or state denoted by the predicate is true for the entire time span that is loosely defined by *purá*. In this case, *cid* can be translated as ‘always’ but again, unlike the English adverb, it does not quantify over a set of time points because the habituality is already expressed without the particle.

The results of the synchronic investigation of Vedic *cid* have consequences for the reconstructed semantics of Proto-Indo-European  $*=k^h id$ . Dunkel (2014:451) assigns it an additive and a generalizing function. However, both the Vedic and typological data suggest that the generalizing function is not part of the semantics of the particle itself but that it only occurs in combination with interrogative proforms. In order to corroborate this assumption and to determine whether  $*=k^h id$  also possesses a totalizing function further synchronic analyses of its reflexes are necessary.

## References

- Dunkel, George E. 2014. Lexikon der indogermanischen Partikeln und Pronominalstämme: Band 2. Lexikon. Universitätsverlag Winter, Heidelberg.
- Forker, Diana. 2016. Toward a Typology for Additive Markers. *Lingua* 180, 69–100.
- Goldstein, David, 2019. Discourse Particles in LSJ: A Fresh Look at  $\gamma\epsilon$ , In: Stray, Christopher, Clarke, Michael, Katz, Joshua T. (Eds.), Liddell and Scott. The History, Methodology, and Languages of the World's Leading Lexicon of Ancient Greek. Oxford University Press, Oxford, pp. 268–287.
- Gonda, Jan. 1954–1955. Notes on the Indo-European  $k''i$ - and  $k''o$ -pronouns. *Lingua* 4, 241–285.
- Grassmann, Hermann. 1873. Wörterbuch zum Rig-Veda. F.A. Brockhaus, Leipzig.
- König, Ekkehard. 2017. Syntax and Semantics of Additive Focus Markers from a Cross-linguistic Perspective: A Tentative Assessment of the State of the Art, In: De Cesare, Anna-Maria, Andorno, Cecilia (Eds.), Focus on Additivity. Adverbial Modifiers in Romance, Germanic and Slavic Languages. John Benjamins Publishing Company, Amsterdam, Philadelphia, pp. 23–43.
- Lühr, Rosemarie. 2017. Stressed and Unstressed Particles in Old Indic, In: Bayer, Josef, Struckmeier, Volker (Eds.), Discourse Particles. Formal Approaches to their Syntax and Semantics. Walter de Gruyter GmbH, Berlin, Boston.
- Mumm, Peter-Arnold. 2004. Altindisch *sma*. *International Journal of Diachronic Linguistics* 1, 19–86.
- Ying, Zhang, 2017. The Categorization of *Dou* (都) in Chinese: A Study from a Cross-linguistic Perspective. *Bulletin of Chinese Linguistics* 10, 214–234.

## **The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective**

Two central questions regarding language evolution are a) how representations of linguistic structures emerge in individuals and b) how emergent linguistic structures spread throughout communities and became conventionalised (Author 1a). The concept of “protolanguage” (Tallerman 2012) or “early languages” (Heine & Kuteva 2007) is often evoked to bridge the gap between fully complex linguistic structures and an earlier stage of structured communication. However, questions a) and b) hold for the evolution of protolanguage just as much as they do for fully complex modern language. This paper discusses theoretical models of the interaction of two important dimensions and their potential for shedding light on these two questions from a usage-based perspective: 1) the individual cognitive level, in which processes of automation and entrenchment lead to the emergence of structured linguistic representations, and 2) the community level, in which processes of ritualisation and conventionalisation lead to the diffusion and stabilisation of usage patterns within communities of practice (Schmid 2020; Author 2).

First, regarding the level of individuals, domain-general cognitive processes such as automation and analogy are central to how structures emerge, are represented, and are stored in the minds/brains of individuals. The parallelism of dialogic interaction (Du Bois 2014) invites speakers to share (ad-hoc) structures and repeat them in future usage events. Frequent repetition promotes structures to be entrenched and automated as complex constructions that, over time, emancipate from their original concrete referential uses. That is, assisted by further domain-general capacities like chunking and metaphor, more schematic representations with procedural (i.e. grammatical) functions may emerge. Entrenchment and its associated effects have been implicated in the emergence of structure in language learning (Schmid 2016), language change (Hilpert 2017) and language evolution (Author 1b, Author 1a).

Secondly, community-wide processes such as conventionalisation and ritualisation refer to the establishment, negotiation and diffusion of community-wide, regularised practices (Schmid 2020; Author 2). Processes of conventionalisation and ritualisation depend on processes of social transmission interacting with individual processes of entrenchment, and as such represent an important part of explaining how communicative structures came to be adopted both in proto(linguistic) communities. Conventionalisation also represents the foundation for cumulative culture (Tomasello 1999), which enables the cumulative accretion of changes and increasing structuration of communicative systems. This in turn enabled the gradual change from protolanguage and early languages to modern human language (Heine & Kuteva 2007).

The channel of interaction between entrenchment processes and conventionalisation processes is usage, whereby rough alignments between mental form–meaning pairings and communal linguistic norms arise. In summary then, the present paper will present the interaction of the individual cognitive dimension on the one hand, and the community level on the other, as well as their underlying mechanisms, as crucial building blocks of a usage-based approach to the evolution of language.

**References:**

- Du Bois, J.. 2014. Towards a dialogic syntax. *Cognitive Linguistics* 25(3). 359–410.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction*. Oxford University Press.
- Hilpert, M. (2017). *Frequencies in diachronic corpora and knowledge of language*. In M. Hundt, S. Mollin, & S. E. Pfenninger (eds.), *The changing English language: Psycholinguistic perspectives* (pp. 49-68). Cambridge University Press.
- Schmid, H.-J.. (Ed.). (2016). *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge*. De Gruyter.
- Schmid, H.-J.. (2020). *The dynamics of the linguistic system: usage, conventionalization, and entrenchment*. Oxford University Press.
- Tallerman, M. (2012). Protolanguage. In M. Tallerman & K. R. Gibson (eds.), *The Oxford handbook of language evolution* (pp. 479-491). Oxford University Press.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.



## The development of number strengthening in German declensional classes. A diachronic-dialectal corpus study

Keywords: Historical morphology, diachronic dialectology, plural marking, corpus linguistics, German

The present study addresses the implementation of overt marking of number in the plural forms of German nouns belonging to declensional classes which originally lacked number distinctivity. Initially, the phonological reduction of final syllables (cf. Braune 2018: 248-249, 265) had led to a formal overlapping of singular and plural forms within several of the subclasses of the original vocalic stems, most importantly within the paradigms of the neuter *a*-stems, see OHG/MHG nom./acc.sg. *wort* – nom./acc.pl. *wort* ‘word – words’, as well as within the feminine *ō*-stems, see OHG nom./gen./acc.sg. *gēba* – nom./acc.pl. *gēba*, MHG nom.-acc.sg. *gebe* – nom./acc.pl. *gebe* ‘gift – gifts’ (cf. Braune 2018: 250, 265; Klein et al. 2018: 73; Ronneberger-Sibold 2013: 19). In addition, there was massive overlapping within the paradigms of the consonantal *n*-stems in Old and Middle High German, where the oblique cases in the singular coincided with the plural forms, all ending in *-(e)n*, see OHG gen./dat./acc.sg. *zungūn* – nom.-acc.pl. *zungūn* (cf. Braune 2018: 282), MHG gen./dat./acc.sg. *zungen* – nom.-acc.pl. *zungen* ‘tongue – tongues’ (cf. Klein et al. 2018: 73). In the following periods, an expansion of overt plural markers took place reaching its peak in Early New High German. This reinforcement of number distinctivity is generally referred to as *Numerusprofilierung*, i.e. the strengthening of number distinctions, and is one of the most significant processes in the history of German nominal morphology. It consists of various strategies aiming to create number distinctivity. Two basic kinds of strategies can be distinguished:

- 1) changes in the plural paradigm to create number distinctivity, e.g. various overt plural markers are assigned to originally indistinctive forms, sometimes accompanied by the process of *Umlaut*, as in NHG sg. *Wort* – pl. *Worte/Wörter* ‘word – words’
- 2) changes in the singular paradigm as in the case of the *n*-stems where the oblique cases of the singular were identical with the plural ones as in MHG *herze* ‘heart’, gen./dat.sg. *herzen* – nom.-acc.pl. *herzen* (cf. Klein et al. 2018: 73) and the *n*-suffix is reanalyzed as a marker of plurality and removed from the singular paradigm cf. NHG sg. *Herz* – pl. *Herzen*

On the basis of the reference corpora of the historical periods of the German language, the reference corpus for Middle High German, *Referenzkorpus Mittelhochdeutsch* (REM), covering the period from 1050 to 1350 and the reference corpus for Early New High German, *Referenzkorpus Frühneuhochdeutsch* (REF), covering the period from 1350 to 1650 (cf. Klein/Dipper 2016; Dipper/Kwekkeboom 2018), the present study explores the rich meta-annotation provided in the corpora to investigate the diachronic and dialectal spread of plural morphology within members of declensional classes that were originally number indistinct. This investigation is part of a larger study that aims to account for the precise diachronic and dialectal representation of the individual processes pertaining to the overall phenomenon of *Numerusprofilierung* in German. It also serves to explore the methods of retrieving and compiling the relevant data from the different reference corpora to generate a multidimensional map of the spread of these processes over time and space.

## References

- Braune, Wilhelm (2018): Althochdeutsche Grammatik. I. Laut- und Formenlehre. 16. Aufl. Neu bearbeitet von Frank Heidermanns. Berlin/Boston: De Gruyter.
- Dipper, Stefanie/Kwekkeboom, Sarah (2018): Historische Linguistik 2.0. Aufbau und Nutzungsmöglichkeiten der historischen Referenzkorpora des Deutschen. In: Kupietz, Marc/Schmidt, Thomas (ed.): Korpuslinguistik. Berlin/Boston: De Gruyter, p. 95-124.
- Enger, Hans-Olav (2014): Reinforcement in inflection classes. Two cues may be better than one. *Word Structure* 7.2, p. 153–181.
- Klein, Thomas/Dipper, Stefanie (2016): Handbuch zum Referenzkorpus Mittelhochdeutsch. *Bochumer Linguistische Arbeitsberichte* 19. Bochum: Ruhr-Universität Bochum.
- Klein, Thomas/Solms, Hans-Joachim/Wegera, Klaus-Peter (2018): Mittelhochdeutsche Grammatik. Pt. 2. Flexionsmorphologie. Vol. 1. Substantive, Adjektive, Pronomina. Berlin/Boston: De Gruyter.
- Ronneberger-Sibold, Elke (2013): Einführung in die historische Morphologie des Deutschen. In: *Jahrbuch für ungarische Germanistik*, S. 13-34.

## Corpora

- Klein, Thomas/Wegera, Klaus-Peter/Dipper, Stefanie/Wich-Reif, Claudia (2016): Referenzkorpus Mittelhochdeutsch (1050–1350), Version 1.0, <https://www.linguistics.ruhr-uni-bochum.de/rem/>. ISLRN 332-536-136-099-5.
- Wegera, Klaus-Peter/Solms, Hans-Joachim/Demske, Ulrike/Dipper, Stefanie (2021). Referenzkorpus Frühneuhochdeutsch (1350–1650), Version 1.0, <https://www.linguistics.ruhr-uni-bochum.de/ref/>. ISLRN 918-968-828-554-7.

**The Lost Cause: Inflection Class in Amarasi**

**1. Overview** While the origin and development of inflection classes has long been an area of interest for historical linguistics, the majority of research in this area has focused solely on Indo-European languages (Maiden 2005, Collier 2013, Kaye 2015). With this gap in mind, this paper presents a novel perspective on the source of these classes by building on the observation that many Timoric languages appear to have undergone low-level innovations which have given rise to morphological inflection classes (which cannot be reconstructed for Proto-Austronesian).

In particular, this paper investigates an incipient inflection class system in Amarasi (Central Malayo-Polynesian: West Timor), which has two distinct paradigms of prefixal subject agreement that take the shapes CV (1a) and C (1b). The distribution of these prefix sets is partly phonologically predictable as in (1c) on the basis of regular phonotactic constraints (e.g. a ban on CCC clusters and cross-morpheme hiatus, and a dispreference for quadrisyllabic words).

(1) a. Syllabic Agreement (CV)	b. Asyllabic Agreement (C)	c. Phonotactic Distribution																																																								
<table border="0"> <tr> <td></td> <td>SG</td> <td>PL</td> <td></td> </tr> <tr> <td>1</td> <td><i>ku/?u-</i></td> <td><i>ta-</i></td> <td>(IN.)</td> </tr> <tr> <td></td> <td></td> <td><i>mi-</i></td> <td>(EX.)</td> </tr> <tr> <td>2</td> <td><i>mu-</i></td> <td><i>mi-</i></td> <td></td> </tr> <tr> <td>3</td> <td><i>na-</i></td> <td><i>na-</i></td> <td></td> </tr> </table>		SG	PL		1	<i>ku/?u-</i>	<i>ta-</i>	(IN.)			<i>mi-</i>	(EX.)	2	<i>mu-</i>	<i>mi-</i>		3	<i>na-</i>	<i>na-</i>		<table border="0"> <tr> <td></td> <td>SG</td> <td>PL</td> <td></td> </tr> <tr> <td>1</td> <td><i>k/?-</i></td> <td><i>t-</i></td> <td>(IN.)</td> </tr> <tr> <td></td> <td></td> <td><i>m-</i></td> <td>(EX.)</td> </tr> <tr> <td>2</td> <td><i>m-</i></td> <td><i>m-</i></td> <td></td> </tr> <tr> <td>3</td> <td><i>n-</i></td> <td><i>n-</i></td> <td></td> </tr> </table>		SG	PL		1	<i>k/?-</i>	<i>t-</i>	(IN.)			<i>m-</i>	(EX.)	2	<i>m-</i>	<i>m-</i>		3	<i>n-</i>	<i>n-</i>		<table border="0"> <tr> <td>SHAPE</td> <td>PREFIX</td> <td>3SG</td> <td>EXAMPLE</td> </tr> <tr> <td>#CC</td> <td>CV</td> <td><i>na-skau</i></td> <td>'invites'</td> </tr> <tr> <td>#V</td> <td>C</td> <td><i>n-inu</i></td> <td>'drinks'</td> </tr> <tr> <td>σσσ+</td> <td>C</td> <td><i>n-marine</i></td> <td>'is happy'</td> </tr> </table>	SHAPE	PREFIX	3SG	EXAMPLE	#CC	CV	<i>na-skau</i>	'invites'	#V	C	<i>n-inu</i>	'drinks'	σσσ+	C	<i>n-marine</i>	'is happy'
	SG	PL																																																								
1	<i>ku/?u-</i>	<i>ta-</i>	(IN.)																																																							
		<i>mi-</i>	(EX.)																																																							
2	<i>mu-</i>	<i>mi-</i>																																																								
3	<i>na-</i>	<i>na-</i>																																																								
	SG	PL																																																								
1	<i>k/?-</i>	<i>t-</i>	(IN.)																																																							
		<i>m-</i>	(EX.)																																																							
2	<i>m-</i>	<i>m-</i>																																																								
3	<i>n-</i>	<i>n-</i>																																																								
SHAPE	PREFIX	3SG	EXAMPLE																																																							
#CC	CV	<i>na-skau</i>	'invites'																																																							
#V	C	<i>n-inu</i>	'drinks'																																																							
σσσ+	C	<i>n-marine</i>	'is happy'																																																							

**2. Two classes** However, **disyllabic #C** verbs appear to be idiosyncratically and lexically-specified to take one of the two prefix sets. Edwards (2020) observes that 75% of these verbs take C prefixes (2a, 3a) while 25% take CV prefixes (2b, 3b). Semantically-unrelated (near-)minimal pairs as in (2, 3) show that prefix set selection for verbs of this shape is not outwardly phonologically-conditioned, suggesting the existence of two conjugation classes.

(2) a. In <b>n</b> -reku.	b. In <b>na</b> -reku.	(3) a. Ho <b>m</b> -hani.	b. Ho <b>mu</b> -hana.
3SG 3SG.C-strike	3SG 3SG.CV-ruin	2SG 2SG.C-dig	2SG 2SG.CV-cook
'It strikes (x o'clock).'	'S/he ruins, besmirches.'	'You dig.'	'You cook.'

**3. Analysis** I propose that the distribution of CV prefixes originates in verbs which took the PMP causative *\*pa-* prefix, and that the current system is the result of an interaction between historical antepenultimate vowel syncope and phonotactic constraints on cluster formation. Comparative data from neighbouring Rote languages (Tamelan 2021, Edwards 2021) shows that stem-initial #CC clusters in Amarasi originate from the application of antepenultimate vowel syncope to known derivational prefixes (4). This syncope clearly also derived the C prefix set from the original CV forms (e.g. 3SG *\*na-CVCV* > *n-CVCV*). Crucially missing from our comparanda is the highly productive and well-attested PMP causative prefix *\*pa-*, which should have produced Amarasi *\*h-*. Despite allowing many typologically unusual sonority-falling clusters like /fk/, /ft/, /mt/, Amarasi shows a complete absence of #hC. I argue that the descendants of these missing *\*pa-C* > *\*hC* clusters are verbs which synchronically take CV prefixes.

	<b>PMP</b>		<b>Rote</b>		<b>Amarasi</b>		
	Prefix	Function	Root	Verb	Meaning	Verb	Meaning
(4)	<i>*ka-</i>	achieved state	<i>*bəntəŋ</i>	<i>na-ka-bete</i>	'is tense, tight'	<i>na-kbeet</i>	'is stiff, tight'
	<i>*ma-</i>	stative	<i>*lapaR</i>	<i>na-ma-laʔa</i>	'is hungry'	<i>na-mnaha</i>	'is hungry'
	<i>*ta-</i>	spontaneous action	<i>?*belaj</i>	<i>na-ta-mbele</i>	'flies'	<i>na-tpene</i>	'flies'

I propose that the phonotactically illicit *\*hC* was repaired by deleting the overt featural content of *\*h* but retaining its representation in phonological structure as a ghost consonant  $\emptyset_C$  (Piggott 1991; Kiparsky 2003). Recalling how #CC verbs require CV prefixes (1c), I posit that this silent etymological  $\emptyset_C$  on #C verbs created a covert #CC cluster that similarly blocks C prefixes. Thus, a verb like *na-hana* 'cooks' (3b) goes back to PMP *\*pa-panas* 'make hot' > *\*na-ha-hana* > *na- $\emptyset_C$ hana*. Strong evidence for this analysis comes from i) the existence of verbs which show an (anti)-causative alternation as expressed only by a change in prefix set (5), captured straightforwardly by a derivational history as in (6); and ii) the retention of a small set of fossilised forms where exceptionally unsyncopeated causative *ha-* alters with stative *ma-* (Edwards 2020: 445), confirming the existence of *\*pa-* in Proto-Amarasi (7).

	<b>C Prefix Set</b>		<b>CV Prefix Set</b>			<b>Stative Noun</b>		<b>Causative Verb</b>	
	Verb	Meaning	Verb	Meaning		Verb	Meaning	Verb	Meaning
(5)	<i>n-ʔate</i>	'serves (ITR.)'	<i>na-ʔate</i>	'enslaves'		<i>mainuan</i>	'open'	<i>n-hainua-b</i>	'opens'
	<i>n-mae</i>	'is ashamed'	<i>na-mae</i>	'shames s.o.'	(7)	<i>maʔekiʔ</i>	'smooth'	<i>n-haʔeki</i>	'smoothens'
	<i>n-peaʔ</i>	'breaks (ITR.)'	<i>na-peaʔ</i>	'breaks s.t.'		<i>maʔkafaʔ</i>	'light'	<i>n-haʔkafa</i>	'lightens'
	<i>n-punu</i>	'rots, decays'	<i>na-punu</i>	'makes rot'		<i>maputuʔ</i>	'hot'	<i>n-haputu</i>	'heats up'

- (6) a. PMP *\*buRuk* 'rotten' > *\*na-punu* →<sub>syncope</sub> *n-punu* 'rots'
- b. PMP *\*pa-buRuk* 'makes rotten' > *\*na-ha-punu* →<sub>syncope</sub> *\*na-h-punu* →<sub>CC red.</sub> *na-punu* 'makes rot'

## References

- Collier, Scott James. 2013. *The evolution of complexity in Greek noun inflection*. Guildford: University of Surrey dissertation.
- Edwards, Owen. 2020. *Metathesis and Unmetathesis: Parallelism and Complementarity in Amarasi, Timor*. Berlin: Language Science Press.
- Edwards, Owen. 2021. *Rote-Meto Comparative Dictionary*. Canberra: The Australian National University Press.
- Kaye, Steven. 2015. *Conjugation class from Latin to Romance: Heterocclisis in diachrony and synchrony*. Oxford: University of Oxford dissertation.
- Kiparsky, Paul. 2003. Finnish noun inflection. In Diane Nelson & Satu Manninen (eds.), *Generative approaches to Finnic and Saami Linguistics*, 109–161. Stanford, CA: CSLI Publications.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. In Geert Booij & Jaap Marle (eds.), *Yearbook of Morphology 2004*, 137–175. Dordrecht: Springer.
- Piggott, Glyne L. 1991. Empty onsets: evidence for the skeleton in prosodic phonology. *McGill Working Papers in Linguistics/Cahiers linguistiques de McGill* 7. 41–71.
- Tamelan, Thersia. 2021. *A grammar of Dela; an Austronesian language of Rote in Eastern Indonesia*. Canberra: The Australian National University dissertation.

### West Germanic 2.sg. *-st* Revisited: The Role of Supervescence

The Uniformitarian Principle restrains historical linguists who are reconstructing a proto-language not only from positing typologically unattested linguistic configurations (e.g., a verb system with 2 tenses in the indicative vs. 22 in the subjunctive, or a phonological system with no low vowels at all), but also from positing typologically unattested sociolinguistic situations: e.g., a system that lacks variation, or one without any stigmatized forms. When a proto-language with stigmatized forms (due, e.g., to hypercorrection) divides into several descendant languages such that each inherits those stigma-bearing items, all the ingredients are in place for subsequent loss of the stigma to allow the forms at issue to appear late (at a time subsequent to the initial stage of attestation) in the written texts of many or even all of the descendants. This delayed appearance of once-stigmatized but later-accepted forms can be called *supervescence* [a haplogogized blend of *super*+(*fer*)*vescence*], because it indeed involves, as it were, the bubbling up to the surface of formerly submerged forms. This phenomenon is already well-established for the attested stages of languages' histories: e.g., certain taboo-words have taken centuries to appear in print.

Joseph 2006, 2012, 2013 adduces evidence for plausible cases of such “bubbling up” in the histories of Germanic and Indo-Iranian languages, emphasizing that supervescence obviates an appeal to convergent Sapirian “drift” as an explanation for independent parallel developments in related languages. Janda & Joseph 2023 point out that supervescence (as one innovation, later multiply inherited) not only is more economical than “co-drift” (as multiple innovations, often across all descendant languages), but also is nearer to Sapir's own 1921 notion of drift — which includes several types, one of them involving the persistence and spread of patterns from a proto-language into its descendants (a parallel to pattern-persistence and -spread within one language). It must be stressed that distinct distributions and chronologies of eventually supervescent forms in different descendants of a common ancestral language are expected, because destigmatization itself is governed by sociolinguistic conditions that are unique to each descendant language.

Here, we argue that the appearance of 2.sg. *-st* (< earlier *-s*) in multiple West Germanic [WGmc.] languages is more comprehensible when analyzed as an instance of the supervescence of a once-stigmatized Proto-WGmc. [PWGmc.] form that was originally due to hypercorrection, not as an instance of co-drift whereby *-s* independently and repeatedly became *-st*. Crucially, WGmc. *-st* in OE and OHG has frequently been treated as a case of independently convergent development (cf., e.g., Greenberg 1957, Sihler 1986, Ringe 2002, Ringe & Taylor 2014), but *-st* actually appears in at least one stage of all five major WGmc. languages, hence also in both Old and Middle Frisian, Old (but not Middle) Dutch, and Middle Low German [though not (earlier) Old Saxon]. We agree fully with Greenberg 1957 (and other specialist works, such as Wilmanns 1906, Campbell 1962, Brunner 1965, Paraschkeow 2003, Braune & Reiffenstein 2004) that the pivotal hypercorrection was the addition of a /t/ after a verb-final *-s* that preceded a 2.sg. pronoun starting with *þ...* (or the like, depending on the language) — cf. OHG *gilaubist thū* < *gilaubistū* < *gilaubis thū* ‘believest thou’, e.g. — except that, unlike those authors, we locate this reanalysis in PWGmc. Lausberg 1972 cites a Romansch parallel: *chantast (tu)* ‘singingest thou’ < *cantas tu*.

Sihler 1986, Ringe 2002, and Ringe & Taylor 2014 treat the inversion context *...-s + þū* as a red herring in the rise of 2.sg. *-st* in OE: they view that configuration as syntactically minor, and analyze the analogical model of OE (plus OHG) preterite-present verbs already ending in *-st* (like OE *wāst* ‘thou knowest’) as rendering inversion structures irrelevant. However, we show, first, that the syntactic environments where the specifically relevant inversion occurs are some of the most basic in OE, and, second, that the preterite-presents (and similar verbs) in *-st* can have been extremely relevant to the reanalysis of *-s* as *-st* without having been the SOLE reason for it. Further, given the predominant view that PWGmc. was already a V2 language, we can project the hypercorrection-related inversion back into the proto-language, along with the various verbs already marked with 2.sg. *-st* — thereby further strengthening the case for supervescence.

### References

- Braune, [Theodor] Wilhelm & Ingo Reiffenstein. 2004. *Althochdeutsche Grammatik, Band I: Laut- und Formenlehre*. Tübingen: Max Niemeyer Verlag. (Sammlung kurzer Grammatiken germanischer Dialekte, A: Hauptreihe, 5. [Fifteenth edition; first edition (by Braune alone) 1886.]
- Brunner, Karl. 1965. *Altenglische Grammatik, nach der Angelsächsischen Grammatik von E[duard] Sievers*. Tübingen: Max Niemeyer Verlag. (Sammlung kurzer Grammatiken germanischer Dialekte, A: Hauptreihe, 3.) [Third edition; first edition 1942 (based on Sievers' grammar: third edition 1898, first edition 1882).]
- Campbell, Alistair. 1962. *Old English Grammar*. Oxford, UK: Clarendon Press. [Second, corrected hardback edition (first hardback edition 1959); paperback edition 1983, reprinted 1991.]
- Greenberg, Joseph H[arold]. 1957. "The Problem of Linguistic Subgroupings". Chapter 4, pp. 46-55, in his *Essays in Linguistics*. Chicago, IL: University of Chicago Press.
- Janda, Richard D. & Brian D. Joseph. 2023. "Drift as a Centenarian Vagabond: Initially Floated in 1921 but Still an Unmoored Concept". Paper presented at the 35th Annual Meeting of the North American Association for the History of the Language Sciences (NAAHoLS); Denver, CO; January 6-7.
- Joseph, Brian D. 2006. "On Projecting Variation Back into a Proto-Language, with Particular Attention to Germanic Evidence". On pp. 103-118 in Thomas D. Cravens (ed.), *Variation and Reconstruction*. Amsterdam: John Benjamins. (Current Issues In Linguistic Theory, 268.)
- Joseph, Brian D. 2012. "A Variationist Solution to Apparent Copying Across Related Languages". On pp. 151-164 in Lars Johanson & Martine Robbeets (eds.), *Copies vs. Cognates in Bound Morphology*. Leiden: Brill.
- Joseph, Brian D. 2013. "Demystifying Drift — A Variationist Account". On pp. 43-66 in Martine Robbeets & Hubert Cuyckens (eds.), *Shared Grammaticalization, with Special Focus on the Transeurasian Languages*. Amsterdam: John Benjamins.
- Lausberg, Heinrich. 1972. *Romanische Sprachwissenschaft III: Formenlehre*. Berlin: Walter de Gruyter & Co. (Sammlung Göschen, 7199.) [Second, revised edition; first edition 1962 (Sammlung Göschen, 1199).]
- Lühr, Rosemarie. 1984. "Reste der athematischen Konjugation in den germanischen Sprachen". On pp. 25-90 in Jürgen Untermann & Béla Brogyányi (eds.), *Das Germanische und die Rekonstruktion der indogermanischen Grundsprache: Akten des Freiburger Kolloquiums der Indogermanischen Gesellschaft; Freiburg [im Breisgau], 26.-27. Februar 1981*. Amsterdam: John Benjamins. (Current Issues in Linguistic Theory, 22.)
- Paraschkewow, Boris. 2003. "Zur Polygenese des -t in der Verbalendung -st". *Zeitschrift für germanistische Linguistik* 31.3:382-385.
- Ringe, Donald A., Jr. 2002. "Syncopated Present Indicative Forms in Old English". On pp. 125-156 in Alfred R. Wedel & Hans-Jörg Busch (eds.), *Verba et Litterae — Explorations in Germanic Languages and German Literature: Essays in Honor of Alfred L. Lloyd*. Newark, DE: Linguatext, Ltd.
- Ringe, Donald A., Jr. & Ann A. Taylor. 2014. *A Linguistic History of English, Volume II: The Development of English*. Oxford, UK: Oxford University Press. (Oxford Linguistics.)
- Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York, NY: Harcourt, Brace, & World. [Hardback edition; paperback edition 1955.]
- Sihler, Andrew. 1986. "Germanic 2nd-Person Endings in -st". *Münchener Studien zur Sprachwissenschaft* 47:193-222.
- Wilmanns, [Franz] Wilhelm. 1906. *Deutsche Grammatik: Gotisch, Alt-, Mittel- und Neuhochdeutsch, Band 3: 3. Abteilung, Flexion; 1. Hälfte, Verbum*. Strasbourg: Verlag Karl J. Trübner. (First and second edition [sic].)



### From *de* to *ke*: functional transfer of a topic shift marker from Turkish to Cappadocian Greek

This presentation discusses how the particle *ke* is used as a topic shift marker in Cappadocian Greek (CG), which I argue is the result of contact-induced language change. CG is a cluster of closely related, critically endangered dialects of the Greek language that were spoken in the Turkish region of Cappadocia until the Greek-Turkish Population Exchange of 1924 (Janse 2020). Due to intensive influence of the Turkish language, CG became a textbook example of language contact cited by many scholars as a case of heavy borrowing (Thomason & Kaufman 1988: 215-22; Thomason 2001: 74; Winford 2003: 83-4; 2005: 402-9; Matras 2020: 231-4).

A salient example of this strong Turkish influence is the extension of the functional range of the CG proclitic particle *ke*. In CG, *ke* functions as a coordinating (‘and’), additive (‘also’) and scalar-additive (‘even’) particle, as was already the case in Ancient and Medieval Greek (Beekes & Van Beek 2010: 615). However, a new function appears in the following examples:

- (1) *k’ ekíno ke lex: “írta na vró ta tría güzélja.”*  
 and DEM.NOM **and** say.PRS.3SG  
*k’ ekín ke lex: “aderé éxo éks peðjá.”*  
 and DEM.NOM **and** say.PRS.3SG  
 ‘And he says, “I came to find the three Fair Ones.”  
 And she says, “Now I have six sons.”’ (Dawkins 1916: 306-7)

Alongside a first *k(e)*, which functions here as a coordinating particle, a second, seemingly redundant *ke* is introduced, usually positioned between the topic and the verb of the sentence. I argue that this specific use of *ke* marks a topic shift, which is the result of ‘functional transfer’ (FT) from the Turkish enclitic particle *dA*. FT is defined by Siegel (2012: 189) as “applying the grammatical functions of a morpheme from one language (the Source Language) to a morpheme in another language (the Recipient Language)”. More specifically, the transfer from *dA* to *ke* is a case of ‘type II’ (2012: 194-8), meaning that “[t]he functional range of a grammatical item or construction is extended and/or reduced” (Jennings & Pfänder 2018: 91).

In this case, the functional range of the Greek (RL) proclitic particle *ke* is extended and incorporates the function of topic shift marker associated with the Turkish (SL) enclitic particle *dA*, based on already existing similarities with Greek *ke*: both are used not only as a coordinating particle, but also as a (scalar-)additive particle (Göksel & Kerslake 2005: 110; Dawkins 1916: 605). Additionally, Turkish *dA* also functions as a topic shift marker, unlike Ancient and Medieval Greek *ke*:

- (2) *Ben sinema-ya git-ti-m. Ahmet de tiyatroya git-ti-Ø.*  
 I cinema-DAT go-PFV-1SG Ahmet **TOP** theatre-DAT go-PFV-3SG  
 I went to the cinema. **As for** Ahmet, he went to the theatre.

It is this function of topic shift marker that is included into the functional range of the CG particle *ke*. The analysis of this specific use of the particle was carried out in a corpus of 58 CG folktales (ca. 50,000 words, the largest CG text collection to date). It should be noted that topic shift marker *ke* is only found in combination with the verb *lé(γ)ο* ‘say’. This can be explained by the distribution of the Turkish model *dA*, which introduces a new topic “without changing the direction of the discourse” (Göksel & Kerslake 2005: 513). Without this continuity or connection between two events or situations, other topic shift markers, like the clitic *ise* would be used (Göksel & Kerslake 2005: 401). Subsequently, since a reported dialogue (introduced by the verb *lé(γ)ο* ‘to say’) between constantly shifting subjects is a prototypical example of switching topics without changing the action in the predicate, this could explain why the use of CG *ke* is seemingly restricted to this specific verb.

This paper contributes to the research on language contact by investigating the mechanisms by which the Greek dialect in Cappadocia – which was isolated from the rest of the Greek-speaking world from the 11th century onwards – underwent significant changes under the strong influence of the surrounding Turkish *superstratum*. Additionally, it contributes to the overall knowledge of this underdocumented Greek dialect at the brink of extinction.



## References

- Beekes, R.S.P. & Van Beek, L. (2010). *Etymological dictionary of Greek*. Brill.
- Dawkins, R.M. (1916). *Modern Greek in Asia Minor*. Cambridge University Press.
- Göksel, A. & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. Routledge.
- Janse, M. (2020). Ἑλληνιστὶ γινώσκεις; (Acts 21:37): The survival of Cappadocian Greek. *Cursor: Zeitschrift für Freunde der Lateinischen Sprache und europäischen Kultur*, 16, 49-57.
- Jennings, W. & Pfänder, S. (2018). *Inheritance and Innovation in a Colonial Language: Towards a Usage-Based Account of French Guianese Creole*. Palgrave-McMillan.
- Matras, Y. (2020). *Language contact* (2nd ed.). Cambridge University Press.
- Siegel, J. (2012). Two Types of Functional Transfer in Language Contact. *Journal of Language Contact* 5: 187-215.
- Thomason, S.G. & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. University of California Press.
- Thomason, S.G. (2001). *Language contact. An introduction*. Edinburgh University Press.
- Winford, D. (2003). *An introduction to contact linguistics*. Blackwell.
- Winford, D. (2005). Contact-induced change: classification and process. *Diachronica*, 22(2), 373-427.

## A computational approach to detect discourse traditions and register differences: a case study on historical French

Historical sociolinguists have demonstrated the crucial role of register/genre in mediating the spread of innovations throughout language communities (Nevalainen and Raumolin-Brunberg 2017). However, the traditional conceptualization of genre has been challenged by the concept of Discourse Traditions (Kabatek 2005), henceforth DTs.

The core idea in the DTs framework is that language is not a monolithic object and one cannot dispense with the impact of textual traditionality to study the evolution of individual phenomena (Kabatek 2005). Additionally, the detection of DTs represents a challenge for quantitative corpus linguistics, as each texts can allow for global or internal classifications (Kabatek 2013: 19). Although previous research has discussed distinctive classification features for textual genre, descriptions might be biased by a researcher's particular interest or object language. It is therefore worthwhile to explore whether thorough philological analysis can be complemented by bottom-up generated classifications.

The first goal of our paper is to leverage on the popularity of computational models for the semantic representation of words and texts, so-called 'vector space models' (Boleda 2020), for the unsupervised, bottom-up identification of DTs. Document-based vector space models represent a document's content by means of a frequency profile (i.e., vector) of the terms occurring therein. Afterwards documents can be compared by calculating a similarity value based on those frequency vectors. The rationale is that the co-occurrence of certain terms in a document will be correlated with certain DTs. For this endeavor we explore a corpus of 1400+ historical French theater plays dated between 1600 and 1930 (Author 2023). This corpus is annotated in terms of sub-genres (e.g., comedy, tragedy, pastoral, etc.), which might correlate with different registers.

The second goal is to verify how the automatic classification of documents improves or complements a traditional genre classification provided by the corpus metadata. Building on previous work on Spanish (Author *in press*), we evaluate this comparison by checking the impact of both classifications on a case study of syntactic alternation in French, namely the distribution and change in inverted (1) and clefted (2) interrogatives.

- (1) *Aimez-vous voyager?*  
'Do you like to travel?'
- (2) *Est-ce que vous aimez voyager?*  
'Is it that you like to travel?'

By including these two different operationalizations of DTs in a logistic regression, we show how the bottom-up classification (a) improves the overall fit of the regression model, (b) reveals unattested differentiation within theater texts, and (c) functions as a principled approach to distinguishing 'change from above' and 'change from below'. Overall, the proposed approach evidences the relevance of computational-semantic methods for historical (socio)linguistic research.

## References

Author 2023.

Authors in press.

Boleda, Gemma. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*. 6(1). 213–234.

GITHE. 2015. Codea+ 2015. *Corpus de documentos españoles anteriores a 1800*.

- Kabatek, Johannes. 2005. Tradiciones discursivas y cambio lingüístico. *Lexis: Revista de lingüística y literatura* 29(2). 151–177.
- Kabatek, Johannes. 2013. ¿Es posible una lingüística histórica basada en un corpus representativo? *Iberoromania* 77(1).
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 2017. *Historical Sociolinguistics : Language Change in Tudor and Stuart England*. London: Routledge.

## Where do all the NPs go? – A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900

Although in modern German, it is highly marked to place an NP in the postfield,<sup>1</sup> the phenomenon is not as rare as expected in early New High German (1650-1900) data (ex. A).

- A. ...weil er [...] von den meisten Medicis [genennet wird]<sub>RSB</sub> **ein Schmid aller Kranckheiten**.  
 ... as he ... by the most doctors called is a forger of all sicknesses.  
 "...as he is called a forger of sicknesses by most physicians." (Abel 1699, 225)

However, studies concerned with extraposition in diachrony treat the placement of NP as a marginal phenomenon that can nearly exclusively be explained by the length of the NP (Ebert 1980, Sapp 2014) or pragmatic factors like givenness (Light 2011).

Although it is not mentioned as such in these studies, both explanations can be linked to processing difficulties which are resolved by extraposition. Processing difficulties can be rather objectively investigated using Information Density, namely Surprisal (ID; Shannon 1948). Levy and Jaeger (2007; 1) define ID as the "amount of information per unit comprising the utterance". It is calculated as the likelihood with which a word occurs in a context ( $P(\text{word}) = -\log_2(\text{word}|\text{context})$ ). More expected combinations of words result in lower surprisal values and, thus, in lower perceiving difficulties (Hale 2001), as low surprisal values reduce the impact of the working memory (Levy & Jaeger 2007, Hale 2001, Levy 2008). We claim that the surprisal value of NPs is also relevant for their placement in the postfield. Therefore, we propose that NPs with high surprisal values are more likely to be extraposed.

To investigate this claim, we built a corpus of medical and theological texts from 1650 to 1900 taken from the Deutsches Textarchiv (DTA, BBAW 2019). We manually extracted extraposed and embedded NP and the sentence brackets using WebAnno (Eckart de Castilho et al. 2016). Then, we calculated a 2-Skip-Bigram-Language Model (Guthrie et al. 2016) to gain surprisal values for every word in the context. These surprisal values were used to calculate the mean Skipgram surprisal on lemmata for every annotated NP. Furthermore, we determined the length of the NP, the text genre (medical vs. theological), and the Orality Score (COAST, Ortmann & Dipper 2022) since extraposition is claimed to be more likely in conceptionally oral texts (Koch & Oesterreicher 2007) and the time of publication, the period. To determine the most influential factor for extraposition, logistic regression was performed with R (The R Core Team 2022).

As a result, we find that extraposition is indeed linked to high surprisal values ( $z=-2.44$ ,  $p<.05$  \*) and that length is not significant ( $z=-0.48$ ,  $p<.63$ ), in contrast to the aforementioned literature. However, both the genre ( $z=-2.58$ ,  $p<.001$ \*\*) and the interaction between Orality Score and the period ( $z=-2.68$ ,  $p<.001$ \*\*) are more significant. That suggests an influence of genre and a change over time. The latter is furthermore supported by a slightly significant result for the interaction between length and period ( $z=-1.75$ ,  $p<.1$ ).

Following Speyer (2015: 499), we suggest that there are more processing capacities available behind the right sentence bracket because the main verb is eventually processed at this point. Thus, there is no uncertainty about the constituent function of the extraposed phrases, which causes further strain on the working memory. This leaves more capacities to process lexical difficulties, represented by the surprisal values. In our corpus, the effect is more pronounced than the influence of length. Furthermore, we detect indications of language change in the interactions and an influence of the genre, suggesting a difference in writing style that could yield further investigations.

<sup>1</sup> The postfield is the position behind the right sentence bracket (RSB) and the RSB is the position late in the clause where verbal material, which is distributed over two positions in the clause in German, occurs (Wöllstein 2014).

**Example taken from:**

Abel, H. (1699). *Wohlerfahrner Leib-Medicus Der Studenten*. Leipzig: Groschuff.

**References:**

- BBAW (2019). Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Berlin-Brandenburgische Akademie der Wissenschaften; <http://www.deutschestextarchiv.de/>. [last accessed: 2023-01-19]
- Ebert, R. P. (1980). Social and stylistic variation in early new high german word order: The sentence frame (>satzrahmen<). 102. Jahresband, 357–398.
- Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A. and Biemann, C. (2016): A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In *Proceedings of the LT4DH workshop at COLING 2016*, Osaka, Japan.
- Guthrie, D., B. Allison, W. Liu, L. Guthrie, and Y. Wilks (2006). A closer look at skip-gram modelling. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.
- Hale, J. 2001. A probabilistic Early parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics*.
- Koch, P. & Oesterreicher, W. 2007. Schriftlichkeit und kommunikative Distanz. *ZGL* 35, 346-375.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106(3), 1126–1177.
- Levy, R. & Jaeger, F. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19. 849-856.
- Light, C. (2011). The information structure of subject extraposition in early new high german. In S. Müller (Ed.), *Proceedings of the HPSG 2011 Conference*
- Ortmann, K. and S. Dipper (2022a). Coast (conceptual orality analysis and scoring tool). <https://github.com/rubcompling/COASTcoast-conceptual-orality-analysis-and-scoring-tool> [last accessed: 2023-01-18]
- R Core Team (2022). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
- Sapp, C. D. (2014). Extraposition in middle and new high german. *The Journal of Comparative Germanic Linguistics* 17(2), 129–156.
- Speyer, A. (2015a). Auch früher wollte man informieren – Zum Einfluss der Informationsstruktur auf die Syntax in der Geschichte des Deutschen. *Zeitschrift für germanistische Linguistik* 43(3), 485–515.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* 27(3), 379 – 423
- Wöllstein, A. (2014). *Topologisches Satzmodell* (2 ed.). Heidelberg: Winter.

### **Pronoun history and information structure in 18th century non-religious Russian texts**

The registers - situational and context-dependent language varieties (Biber & Conrad, 2019) - that dominated Russian in the diachronic point of view are religious original and translated texts, legends, and chronicles, which are also often bound to religious canons<sup>1</sup> (Azimova & Johnston, 2012; Comrie, 2018; Vakulenko, 1989). An apparent change came with the age of Enlightenment, which began in Central and Western Europe around the middle of the 17th century. The movement came to the Russian-speaking area in the late 17th-early 18th century and started the process of register development in the written language. The range of different registers in written language became broader and included, among other things, travelogues, letters from ordinary people, diaries, and others (Anciferova, 2012; Kotkov & Pankratova, 1964; Mayorov, 2006). This development period of Russian could be understood as the transition phase to the modern Russian language. The changes that take place at this stage should not be understood as selective changes that happened at a specific moment in the history of the language. Before a specific phenomenon becomes noticeable or written down, there must first be a tendency in the development of language history that leads to this change (Lieb, 1970, p. 43).

In transitional stages of language development, pronouns are sensitive and can indicate more profound linguistic changes. Languages can create new pronouns and pronoun functions over time. Among others, there may be denotational semantic reasons, such as the need to convey refined gender and sex distinctions (Lakoff, 1975). Overall, however, changes in pronouns can also affect the syntax, morphology, and information structure of the sentence or phrase. The historical corpus linguistics and digital humanities methods demonstrate the flexible nature of pronouns. For the history of Russian, Meyer (2011) demonstrated by statistical corpus analysis a shift from a consistent null-subject language to a partial null-subject language over a period of about 700 years (cf. Roberts & Holmberg, 2010 for the typology of null-subject languages). Among others, this change weakened overt pronouns & function from demonstrative, referent-introducing, and contrastive to coreferent, which apparently affected 1st and 2nd person before 3rd person. An appropriate comparative empirical analysis must be based on diachronic corpora in which pronouns are annotated in a unified manner across languages for their syntactic roles, information structural categories (focus, givenness, and topic), as well as coreferential properties and links to their discourse antecedents.

The talk aims at providing the first comprehensive, corpus-based investigation of the historical developments of the grammar and use of pronouns in 18th-century Russian non-religious texts. Focusing mainly on the history of subject pronouns, it uses a combination of grammar- and discourse-oriented methods to investigate how pronominal forms and their functions change over time. The analysis discussed during the talk involves compiling, annotating, and evaluating a corpus for Russian. Developing a general workflow that includes all phases of text processing for corpus linguistic investigations and would be projectable to other languages and language development stages is central to our approach to historical corpus data. Furthermore, explicit text selection principles are developed, allowing for a more objective look at language history and intended to redefine language diachronic definitions of past centuries.

---

<sup>1</sup> An exception are the birch bark letters from the 11th - 15th century, which represent the spoken register in written form and are now available in a large amount. However, they usually only contain a few lines, phrases, or words (Schaeken, 2019; Zaliznjak, 2004; Zaliznjak, 2006). Texts in other registers also exist, but they are not frequent and only in rare cases originate from a non-religious context.

## References:

- Anciferova, O. «Журнал Путешествия...» Демидова Как Памятник Русского Языка Второй Половины XVIII Века (Doctoral Dissertation). Vol. Dissertation. Ekaterinburg: Ural Federal University, 2012.
- Azimova, N., and B. Johnston. 'Invisibility and Ownership of Language: Problems of Representation in Russian Language Textbooks'. *The Modern Language Journal (Boulder, Colo.)* 96, no. 3 (2012): 337–49.
- Biber, Douglas, and Susan Conrad. *Register, Genre, and Style*. Cambridge: University Press, 2019.  
<https://doi.org/10.1017/9781108686136>.
- Comrie, Bernard. 'Russian'. In *The World's Major Languages / Editor, Bernard Comrie*, Third edition. London: Taylor and Francis, 2018.
- Kotkov, S., and N. Pankratova. *Istočniki Po Istorii Russkogo Narodno-Razgovornogo Yazyka XVII-Načala XVIII Veka*. Moskva: Nauka, 1964.
- Lakoff, Robin Tolmach. *Language and Woman's Place Text and Commentaries*. New York, 1975.
- Lieb, Hans-Heinrich. 'Sprachstadium und Sprachsystem: Umriss einer Sprachtheorie'. Stuttgart u.a.: Kohlhammer, 1970.
- Mayorov, A. *Региональный Узус Деловой Письменности XVIII Века (По Памятникам Забайкалья)*. Diss. Диссертация... Доктора Филологических Наук/АП Майоров, 2006. Moskva: V.V. Vinogradov Russian Language Institute of the Russian Academy of Sciences, 2006.
- Meyer, Roland. *The History of Null Subjects in North Slavonic. A Corpus-Based Diachronic Investigation*. Regensburg, 2011.
- Roberts, I., and A. Holmberg. 'Introduction: Parameters in Minimalist Theory'. Edited by T. Biberauer and et al. *Parametric Variation: Null Subjects in Minimalist Theory.*, no. Cambridge University Press (2010): 1–57.
- Schaeken, Jos. *Voices on Birchbark: Everyday Communication in Medieval Russia / Jos Schaeken*. Studies in Slavic and General Linguistics ; v. 43. Leiden, Boston: BRILL, 2019.
- Vakulenko, T. *Формирование и Функционирование Административноканцелярской Лексики в Русском Языке XVIII Века*. Kyiv: Kyiv Pedagogical Institute A.M.Gorky, 1989.
- Zaliznjak, Andrej Anatol'evič. 'Berestjanye Gramoty — Bescennyj Istochnik Svedenij o Drevnej Rusi i Ee Jazyke'. *Lekcii Laureatov Demidovskoj Premii (1993-2004)*, 2006, 229–32.
- . *Drevnenovgorodskij dialekt*. 2. izd., Pererabot. s učetom materiala nachodok 1995-2003 gg. Jazyk, semiotika, kul'tura. Moskva: Jazyki Slavjanskij Kul'tury, 2004.

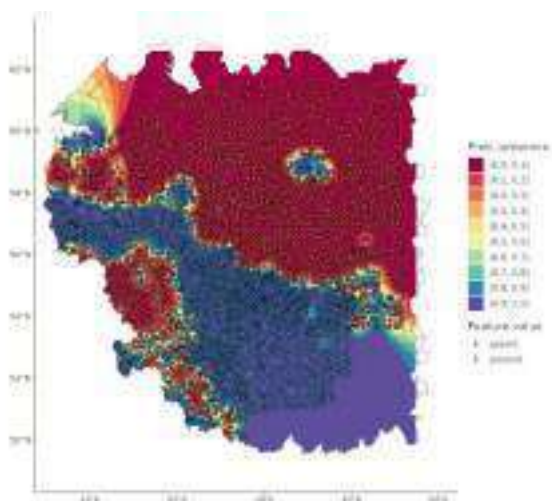


### **Isoglosses and distributions of features – Analyses of the *Dialectological Atlas of the Russian Language***

The Dialectological Atlas of the Russian Language (abbreviated DARJa based on its Russian title) represents more than 4 decades of data collection, from 1938 onwards, and was published in Moscow during 1986-2005. It contains 313 maps, each corresponding to a linguistic feature, and covers 4196 locations. In 2015-16, researchers at Kazan Federal University extracted linguistic features and their values directly from the physical maps and created Excel files giving the values of features across locations (Isaev et al. 2016). We have processed these materials further, georeferencing the map of locations covered and manually extracting the latitude and longitude location of every location.

Thus, the DARJa data are now amenable to systematic, quantitative analyses. For instance, it is possible to define dialect areas in more principled ways than was hitherto possible. For instance, Zaxarova and Orlova (2004: 166) present a map of 28 dialect zones. A rather similar map can be generated from the DARJa data by computing Hamming distances based on the features present in different locations, classifying the locations in a UPGMA tree and cutting this tree into  $k = 28$  clusters. This approach is principled and also versatile, since  $k$  can be any number.

The focus in this talk is on two issues of dialectological method, namely how to draw isoglosses computationally and how to measure the similarity of two distributions of feature values. After having binarized all features by taking each feature value as present/ absent we extract isoglosses, as follows. First, we fit a thin plate spline (Franke 1982) to each binarized feature, and produce a spatial interpolation on the region in question (following Wieling et al. 2011 and Guzmán Naranjo and Becker 2021). An example of this can be seen in Figure 1. This figure shows the spatial distribution of so-called *Akan'e* (weakening of unstressed *o*). If the probability of the interpolation is rounded to 0 and 1 we are left with clearly delimited regions that allow for extracting the isoglosses by applying an edge detection procedure to the map. Finding features which have a similar spatial distribution is also performed using the interpolated values. Here we use correlation distances between values across locations. Using these methods for drawing isoglosses and comparing distributions we go on to analyze the interplay between the many phonological, morphological, syntactical, and lexical features in DARJa.



**Figure 1:** Map of interpolated probabilities for *Akan'e* (weakening of unstressed *o*). The color scheme shows a probability  $P$  of presence = 1 as blue vs.  $P = 0$  as red, with the intermediate color range representing interpolated values.

## References

- Franke, Richard. 1982. Smooth interpolation of scattered data by local thin plate splines. *Computers & Mathematics with Applications* 8.4, 273–281.
- Guzmán Naranjo, Matías & Laura Becker. 2021. Statistical bias control in typology. *Linguistic Typology* 26, 605–670. <https://doi.org/10.1515/lingty-2021-0002>
- Isaev, Igor', Valerii Solov'ev, Farid Salimov, Aleksandr Piljugin & Venera Bairaševa. 2016. Creating a database of Russian dialects and prospects for dialectometric studies. *Herald of the Russian Academy of Sciences* 86(6): 448-453.
- Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PloS ONE* 6, e23613.
- Zaxarova, Kapitolina F. & Varvara G. Orlova. 2004. *Dialektnoe členenie russkogo jazyka*. Moskva: URSS.

## Hearsay in Historical German Newspapers (1740–1840)

The century ‘around’ 1800 was a central period in the history of European society. Between the late Enlightenment and the revolutions of 1848, a series of changes fueled by rapidly growing industrialization and urbanization took place, as well as advances in mass education, widespread politicization of the masses, and increasingly pressing requests for democratic policies.

This period was pivotal for the evolution of newspapers. Because of booming alphabetization and the blossoming of the bourgeoisie, newspapers garnered a growing readership who, in time, became increasingly exigent. This led to the professionalization of journalism and journalists. However, this important period and its impact on the language of newspapers have not been fully investigated in the field of the history of the German language.

In the paper, I will delineate the first results of a research project on the connections among changes in society, changes in the text genre ‘report,’ and changes in the formulation of reported speech in German newspapers in the century ‘around’ 1800.

The focus of the paper is the lexical and grammatical marking of information source of hearsay in reported speech. Following Wiemer (2010: 59), hearsay is understood as synonymous with reportive evidentiality, that is, “reported information with no reference to who it was reported by” (Aikhenvald 2018: 12). Hearsay markers in German comprise — but are not limited to — the reportive modal verb *sollen* (a) (Wiemer 2010: 81; Smirnova & Diewald 2013) and indirect speech with the generic 3SG subject pronoun *man* (b) (Jäger 2010: 179), as well as constructions with the verb *hören* (*hear*; Whitt 2009: 1088–1889) and with the reportive particle *angeblich* (*allegedly*; Wiemer 2010: 92).

- a. Sonst **sollen** Se. Königl. Majestät in Ruppin am See, 3 der größten Häuser erkaufte haben (BN 4.10.1740: 2)  
[Furthermore, His Royal Majesty **is said to** have purchased 3 of the largest houses in Ruppin on the lake]
- b. **Man sagt**, daß der General Brocon [...] das Commando über die Truppen [...] führen werde; (BN 2.7.1740: 1)  
[**It is said** that General Brocon will command the troops]

### 1. Research hypothesis:

Due to the changes delineated above, newspaper reports became more transparent in handling information and information source (Schröder 2017: 169–172); it is expected that this led to a decrease in the frequency of hearsay and an increase in quotative strategies, that is, “reported information with an overt reference to the authorship of the quoted source” (Aikhenvald 2018: 12).

### 2. Research object:

Reported speech, understood as any form of direct or indirect reporting of or reference to linguistically encoded content produced in another communicative situation in the *Wiener Zeitung* (WZ) and the *Berlinische Nachrichten von Staats- und gelehrten Sachen* (BN) (1740–1840), with a focus on hearsay.

### 3. Corpus:

Factual reports (*Kurzbericht*, *faktizierende Meldung*, *Erlebnisbericht*, Wille 2020: 150–155) in the WZ and BN from the years 1740, 1745 – 1770, 1775 – 1800, 1805 – 1830, 1835 (two issues per year).

### 4. Methodology:

Production of a taxonomy of hearsay markers present in the WZ and BN (1740–1840);

Quantitative investigation of the diachronic changes in frequency of different hearsay markers in comparison to quotative strategies in reported speech.

### 5. Results:

As expected, in the analyzed time span, hearsay markers became progressively rare. In the WZ issues from 1740 to 1775, between 38% and 47% of all instances of reported speech are hearsay; that is, they do not contain any indication of the information source. In the 19<sup>th</sup> century, a steady decrease in the frequency of hearsay is observable: In the 1835 issues of the WZ, only 7.5% of all reported speech is hearsay. This diachronic development can be explained fully in light of the changes in media and society in the century ‘around’ 1800.

However, it remains to be discussed why not all markers of hearsay decline at the same rate. Whereas constructions with the generic 3SG subject pronoun *man* decline from being one of the most frequent formulations of reported speech in 1740–1775 (16–20%) to disappearing almost entirely (from 3.2% in 1830 to 0% in 1835), the frequency of the reportive *sollen* remains constant, at ca. 6%.

## References:

- Aikhenvald, A.Y., 2007. Information source and evidentiality: what can we conclude? *Italian Journal of Linguistics* 19, 209–227.
- Aikhenvald, A.Y., 2018. Evidentiality: The framework. In: Aikhenvald, A.Y. (Ed.) *The Oxford handbook of evidentiality*. Oxford University Press, Oxford, 1–43.
- Cornillie, B., Marín Arrese, J., Wiemer, B., 2015. Evidentiality and the semantics-pragmatics interface: An introduction. *Belgian Journal of Linguistics* 29, 1–17.
- Fabricius-Hansen, C., 2002. Nicht-direktes Referat im Deutschen - Typologie und Abgrenzungsprobleme. In: Fabricius-Hansen, C., Leirbukt, O., Letnes, O. (Eds.) *Modus, Modalverben, Modalpartikeln*. WVT Wiss. Verl. Trier, Trier, pp. 7–29.
- Habermann, M., 2019. Praktiken des Zitierens in der vormodernen Wissensliteratur. *Zeitschrift für germanistische Linguistik* 47, 102–136.
- Haß-Zumkehr, U., 1998. "Wie glaubwürdige Nachrichten versichert haben": Formulierungstraditionen in Zeitungsnachrichten des 17. bis 20. Jahrhunderts. Narr, Tübingen.
- Jäger, A., 2010. Reported speech constructions and grammaticalization of hearsay as evidentiality: a cross-linguistic survey. *STUF Sprachtypologie und Universalienforschung* 63, 177–195.
- Letnes, O., 2011. Über werden und sollen als modale und/oder evidentielle Marker. In: Diewald, G., Smirnova, E. (Eds.) *Modalität und Evidentialität. = Modality and evidentiality*. Wiss. Verl. Trier, Trier, 109–123.
- Meyer, H.F., 1972. Berlinische Nachrichten von Staats- und gelehrten Sachen (1740-1874). In: Fischer, H.-D. (Ed.) *Deutsche Zeitungen des 17. bis 20. Jahrhunderts*. Verlag Dokumentation, München, 103–114.
- Oggolder, C., 2016. Politik, Gesellschaft, Medien. Österreich zwischen Reformation und Erstem Weltkrieg. In: Karmasin, M., Oggolder, C. (Eds.) *Österreichische Mediengeschichte. Band 1: Von den frühen Drucken zur Ausdifferenzierung des Mediensystems (1500 bis 1918)*. Springer, Wiesbaden, 9–27.
- Reisner, A., Schiemer, A., 2016. Das Wien(n)erische Diarium und die Entstehung der periodischen Presse,. In: Karmasin, M., Oggolder, C. (Eds.) *Österreichische Mediengeschichte. Band 1: Von den frühen Drucken zur Ausdifferenzierung des Mediensystems (1500 bis 1918)*. Springer, Wiesbaden, 87–112.
- Schröder, T., 2017. Information und Meinung. Presstextsorten vor der Trennungsnorm. In: Pfefferkorn, O., Riecke, J., Schuster, B.-M. (Eds.) *Die Zeitung als Medium in der neueren Sprachgeschichte. Korpora - Analyse - Wirkung*. De Gruyter, Berlin/Boston, 165-175.
- Smirnova, E., Diewald, G., 2013. Kategorien der Redewiedergabe im Deutschen: Konjunktiv I versus *sollen*. *Zeitschrift für germanistische Linguistik* 41, 443-471.
- Squartini, M., 2018. Extragrammatical expression of information source. In: Aikhenvald, A.Y. (Ed.) *The Oxford handbook of evidentiality*. Oxford University Press, Oxford, pp. 273–285.
- Whitt, R.J., 2009. Auditory evidentiality in English and German: The case of perception verbs. *Lingua* 119, 1083-1095.
- Wiemer, B., 2010. Hearsay in European languages: towards an integrative account of grammatical and lexical marking. In: Diewald, G., Smirnova, E. (Eds.) *Linguistic realization of evidentiality in European languages*. De Gruyter, Berlin, 59-129.
- Wille, M., 2020. Die Tageszeitung des 18. Jahrhunderts auf dem Weg zum Massenmedium: Eine textlinguistische Untersuchung des „Hamburgischen Correspondenten“. Universitätsbibliothek, Gießen.

### Towards Quantifying Social Behavior in Language Contact

This talk investigates social dynamics in contact-induced change using Exemplar Theory (ET, Pierrehumbert 2001). Language contact theory attributes the outcomes of contact to social factors and structural factors but theories differ in the primacy attributed to either. While Van Coetsem (1988) mostly attributes the outcomes to structural considerations (e.g. presence or absence of contrasts), Thomason & Kaufman (1988) mostly attribute outcomes to social ones (e.g. intensity of language contact). Situationally, however, it is difficult to define the role of social factors as scholarship often conflates different types of social factors (e.g. “intensity”, social evaluation, etc.).

In this talk, I examine the influence of intensity in the implementation of sub-phonemic shift. I define intensity as the amount of interactions that one community has with another. I create an ET model based on a case-study of Mexican Mennonites (Plautdietsch-speakers) described in Burns (2022). In the Plautdietsch community, Spanish-Plautdietsch bilingualism is traditionally characteristic of males. Males are community liaisons to the world beyond the religious community and learn Spanish at a young age by accompanying older males. While Plautdietsch males become proficient bilinguals, Plautdietsch females remain functionally monolingual.

Northern dialects of Mexican Spanish are undergoing a sub-phonemic shift involving deaffrication of  $/tʃ/ > [ʃ]$ . Deaffrication is believed to have been active in the early 20<sup>th</sup> century (Brown 1989), which is around the time that Canadian Mennonites developed their settlements in Mexico (Burns 2016). While such a shift lacks structural consequences in Mexican Spanish, in Plautdietsch, it can potentially lead to the merger of  $/tʃ/$  and  $/ʃ/$ . Burns (2022) finds that a sample of Plautdietsch-speaking males born in Mexico around 1950 deaffricate, but females born around this time do not. Younger females, who have no knowledge of Spanish, exhibit some deaffrication (Burns 2022). This suggests that deaffrication entered the community through bilingual males, despite a dearth of lexical borrowings from Spanish, and recently expanded beyond this group to other members of the community.

I created an ET model (based on Pierrehumbert 2001) with static phonetic and phonological input in order to test how varying degrees of intensity, as defined above, impact the outcomes of contact in a community where only half of the members are expected to engage in bilingual interactions as outlined above for Plautdietsch. The script that ran 10,000 production–perception loops for gradiently defined tokens of  $[tʃ]$  and  $[ʃ]$ . Spanish-speakers categorized both as types of  $/tʃ/$ , whereas Plautdietsch speakers categorized tokens as either members of  $/tʃ/$  or  $/ʃ/$ . Conversational interactions were assigned between either Plautdietsch males and Plautdietsch females, Plautdietsch males and Spanish-speakers, or any community with itself (i.e. Plautdietsch males, Plautdietsch females, Spanish). This script was run 10 times, each time with random conversational assignment. Across all trials, Plautdietsch males and females started with the same grammatical input.

The outcomes of the model suggest that even when only half of a community is bilingual, sub-phonemic shift can occur, albeit delayed in the non-bilingual group. Initially, Plautdietsch males and females diverged as Plautdietsch males converge with the Spanish community, but Plautdietsch females eventually begin to shift. The ratio of conversational pairings indicates how quickly the community shifts (i.e. if Plautdietsch males have more pairings with Spanish speakers, the shift occurs faster). While the ratio of interactions impacts how quickly subsets of a community adopt the innovation, non-bilingual participants do not preclude the diffusion of innovation as Plautdietsch females and Spanish-speakers also shift. These results indicate when sup-phonemic shift does not occur in communities with bilingualism, other social and structural factors may be necessary explanatory factors.

### References

- Brown, Dolores. 1989. El habla juvenil de Sonora, México: la fonética de 32 jóvenes. *Nueva Revista de Filología Hispánica* 37(1). 43–82.
- Burns, Roslyn. 2022. Mechanisms of sub-phonemic change in prescriptive bilingualism: The case of Mexican Plautdietsch. *Glossa: a journal of general linguistics* 7(1). DOI: 10.16995/glossa.5744
- Burns, Roslyn. 2016. The Plautdietsch vowel shift across space and time. *Journal of Linguistic Geography* 3(2). 79–94.
- Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–158. Amsterdam: John Benjamins.
- Thomason, Sarah & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Van Coetsem, Frans. 1988. *Loan Phonology and the two Types of Transfer in Language Contact*. Dordrecht: Foris Publications.



### Rehabilitating ‘non-proportional’ analogy

The proportional model of analogical change attributes morphological productivity to the solution of a proportion like (a) below, involving ad-hoc generalisation of relationships between small sets of inflected words. This captures the fact that similarity begets similarity in language change: items that look or behave alike, sometimes in quite superficial ways, are liable to become even more alike. Yet many examples of morphological innovation do not appear to be based on any proportional model: contrast the Greek subjunctive *rhégnūtai* in (a) with the variant form *rhégnuētai* in (b), which seems to be formed simply by combining a stem and suffix. Such changes fit better within a rule-based approach to productivity, in which speakers perform a global analysis of linguistic data and distill it into a mental grammar capable of reconstituting surface forms as needed.

a. *phéretai* ‘it is carried’ (ind.) is to *phérētai* (subj.) as *rhégnutai* ‘it shatters’ (ind.) is to ***rhégnūtai***

b. ***rhégnu-ētai*** SHATTER-mediopassive.3sg.subj (cf. *phér-ētai* CARRY-mediopassive.3sg.subj)

Such phenomena led the Neogrammarians to concede that not all analogical processes are proportional (e.g. Osthoff 1897, Paul 1886:95). Attempts to reconcile ‘non-proportional’ and proportional analogy have since fallen into two camps: either all morphological change is treated as non-proportional within a rule-based approach (e.g. Kiparsky 1968), or all analogical change is treated as proportional (see e.g. Hill 2020, Garrett 2008, Fertig 2016). The former approach fails to capture the role of surface similarity in morphological change, while the latter relies on explaining away apparently non-proportional examples by finding either a proportional model or an alternative explanation such as dialect borrowing. Nonetheless both proportional and non-proportional analogy are still widely invoked in historical linguistics, and the theoretical conflict between them remains unresolved.

I will outline a new way of reconciling proportional and ‘non-proportional’ analogy by expanding analogical proportions to include any number of known surface forms, rather than being limited to the three forms of classical proportions (*‘a is to b as c is to x’*). Using computational procedures a local analysis of these forms is performed on the fly and used to predict an unknown inflectional form. In this framework, rules and analogies are seen as notational variants at opposite ends of a spectrum: while traditional rules are maximally general in scope, traditional analogical proportions are based on a single exemplar. This accommodates both types under a single umbrella, while retaining the insights that speakers look for similar models when producing morphological forms, and that the generalisations revealed by morphological innovations can be quite local and idiosyncratic (Joseph 2011). I will show how by integrating both types of analogy with a common formalism, they can be subjected to the same measures (e.g. of phonological/morphosyntactic similarity, type and token frequency) and subjected to a unified statistical analysis, to reveal factors that affect the probability of a putative analogical change actually taking place.

Fertig, D. (2016). Mechanisms of paradigm leveling and the role of universal preferences in morphophonological change. *Diachronica*, 33(4), 423-460.

Garrett, A. (2008). Paradigmatic uniformity and markedness. In J. Good (ed.), *Linguistic universals and language change*, 125-143. New York: OUP.

Hill, E. (2020). Analogy in inflectional change: Modification or whole-word replacement?. *Language* 96(1), 38-58.

Joseph, B. (2011). A localistic approach to universals and variation. In Peter Siemund (ed.), *Linguistic universals and language variation*, 404-425. Berlin: Mouton de Gruyter.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E. and Harms, R.T. (Eds.), *Universals in linguistic theory* (171-202). New York: Holt, Rinehart and Winston.

Osthoff, H. (1879). Kleine Beiträge zur Declinationslehre der Indogermanischen Sprachen. *Morphologische Untersuchungen*, 2, 1-147.

Paul, H. (1886). *Principien der Sprachgeschichte*. 2nd ed. Halle: Niemeyer.



## Aspectual uses of *saber* + infinitive in South American Spanish varieties: a corpus-based study

Silvina Espíndola Moschner, Malte Rosemeyer (Freie Universität Berlin)

*Keywords:* modality, aspect, grammaticalization

In addition to its well-attested modal meaning (ability), the construction *saber* ‘to know’ + infinitive can exhibit aspectual meanings in certain American Spanish varieties (cf. Kany 1945, Di Tullio 2006, Pfänder 2009, Duque Enríquez 2021). Thus, it can be used to express present or past habituality (1) or, in the preterite form, encode completed events (2).

- (1) a. *(El perro) En la calle sabe estar todos los días, salta a las personas, molesta.*  
‘(The dog) is usually in the street every day, it jumps on people, it annoys them.’  
b. *A veces sabía entrar a las 9:00 de la mañana y salía a las 3:00 de la mañana.*  
‘Sometimes I used to come at 9:00 in the morning and leave at 3:00 in the morning.’
- (2) *el barrio albaicín que supo ser asiento de las cortes de los monarcas ziríes en el siglo xi*  
‘the Albaicín neighborhood that was once the seat of the courts of the Zirid monarchs in the 11<sup>th</sup> century’

It is not well understood to what extent the grammaticalization processes leading to (1) and (2) can be considered independent or whether they are interrelated. Likewise, some scholars have assumed the meaning change illustrated in (1) to be induced by contact with Quichua, where the verb *yachay* has both the meanings of ‘to know’ and ‘to be used to’ (cf. Vázquez 1991, Duque Enríquez 2021). This study approaches these questions by comparing four American Spanish varieties (Argentinian, Uruguayan, Bolivian and Ecuadorian Spanish), all of which possess aspectual uses of *saber*.

We conduct a quantitative analysis of  $n = 6,000$  occurrences of *saber* + infinitive in the eSTenTen, a synchronic corpus of blogs and newspapers (Kilgarriff & Renau 2013). Using logistic regression analysis, we demonstrate that the functions of *saber* + infinitive can be predicted from a set of contextual properties, namely semantic features of the subject, predicate type and type of adverbial modification. While these proxies predict both habituality and completion readings, their effect is moderated by whether *saber* is inflected for imperfective or perfective aspect. Crucially, many of these examples, such as (3), can also be taken to express participant-external possibility, i.e. a possibility that does not depend on the agent, but on external circumstances.

- (3) *en la estancia la ema se saben ver (estos animales)*  
‘at the farm la ema one usually sees/can see [= it is possible to see] (these animals)’

Our analysis shows that there are differences in the degree to which the grammaticalization processes leading to (1) and (2) have been implemented in the varieties under study. Thus, while in some dialects (Argentinian, Uruguayan) both habituality and completion readings are attested, in others (Ecuadorian, Bolivian) habituality readings predominate. These findings allow us to postulate the grammaticalization cline in (a), i.e. that the semantic extension of *saber* from the expression of participant-internal to participant-external modality constituted a decisive step in the development of aspectual values, and (b) that the development of habituality and completion readings then followed different grammaticalization paths. The results also suggest that the development of habitual *saber* in Ecuador and Bolivia has been facilitated by contact with Quichua.

- (a) participant-internal possibility (ability) > participant-external possibility > aspectual values

**Corpus:**

SkE. Sketch Engine: Corpus query system. Corpus: esTenTen18. <https://www.sketchengine.eu/estenten-spanish-corpus/>

**References:**

- Aijmer, K. (2004). The semantic path from modality to aspect. *Corpus approaches to grammaticalization in English*, 13, 57-78. Amsterdam/Philadelphia: Benjamins.
- Bosque, I. (2000). ¿Qué sabe el que sabe hacer algo? Saber entre los predicados modales. *Palabras: Víctor Sánchez de Zavala in memoriam*, 303-324. Universidad del País Vasco/Euskal Herriko Unibertsitatea.
- Bybee, J., Perkins, B. & Pagluica, W. (1994). *The evolution of grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
- Di Tullio, A. (2006). Auxiliares y operadores aspectuales en el español rioplatense. *Signo y seña*, 15, 267-285.
- Duque Enríquez, P. (2021). El ‘saber habitual’: Una gramaticalización inducida por factores internos y externos en el español andino ecuatoriano. *Spanish in Context*, 18(3), 362-386.
- Heine, B. (2002). On the role of context in grammaticalization. *Typological studies in language* 49, 83-102. Amsterdam/Philadelphia: Benjamins.
- Hengeveld, K., & Olbertz, H. (2021). Habituals in Spanish. *Linguistics in Amsterdam*, 14(1), 81-104.
- Kany, C. (1945). *Sintaxis hispanoamericana*. Madrid: Gredos.
- Narrog, H. (2012). *Modality, subjectivity, and semantic change: a cross-linguistic perspective*. Oxford University Press.
- Linell, P. (2009). *Rethinking language, mind, and world dialogically*. IAP.
- Pfänder, S. (2009). *Gramática mestiza. Con referencia al castellano de Cochabamba*. Instituto boliviano de lexicografía y otros estudios. La Paz, Bolivia.
- Vatrican, A. (2015). La modalidad en la gramática: las construcciones saber/poder + infinitivo. *Revista Española de Lingüística*, 45(2), 115-141.
- Vázquez, H. (1991). *Reparos sobre nuestro lenguaje usual*. Núcleo del Azuay de la Casa de la Cultura Ecuatoriana “Benjamín Carrión”.
- Vendler, Z. (1957). Verbs and times. *The philosophical review*, 66(2), 143-160. Duke University Press.

Maria Flaksman  
LMU Munich

## LOST IN TRANSLATION: ONOMATOPOEIC WORDS IN OLD ENGLISH GLOSSES

Onomatopoeic (or imitative) words are words with iconic correlation between form and meaning. Iconicity – as opposed to arbitrariness – is a relationship of resemblance (Peirce, 1940). Onomatopoeic words (in some languages also termed ‘ideophones’), thus, are rough ‘copies’ or imitations of sounds they denote (e.g., English *meow*, *bang*, *buzz*, *plop*). But while onomatopoeia enjoys increasing popularity in research on modern languages (see Akita, 2013; Antilla, 1975; Dingemans, 2012; Hinton et al., 1994; Moreno-Cabrera, 2020; Voelz et al., 2001, etc.) it is rarely in focus of historical-comparative research (Carling et al., 2020; Georgescu, 2018; Koleva-Zlateva, 2008; Liberman, 2010). However, diachronic approach towards lexical iconicity yields significant results: exceptions from regular sound changes and otherwise inexplicable phonetic and/or semantic development are successfully explained through onomatopoeia (Campbell, 2013; Durkin, 2009; Hock, 1991; Lühr, 1988; Malkiel, 1990; Sadowski, 2001). It has also been suggested (Flaksman, 2017) that language change causes general *de-iconization* of imitative lexicons, which triggers new onomatopoeic coinage. Research on onomatopoeic words in ancient and reconstructed languages is rare (e.g., see Kozlova, 2013; Anderson, 1998) although evidence for their presence does appear in etymological dictionaries (e.g., Kroonen, 2013; Lehmann, 1986).

The *aim* of this talk is twofold: (1) to provide evidence for the existence of onomatopoeic words in Old English and (2) to discuss the specific problems related to translation of onomatopoeia.

This *comparative* research is based on the *material* of the two reference lists of onomatopoeic words from both Old English and Latin. The lists were comprised by the *method* of continuous sampling from the etymological dictionaries: de Vaan (2008) and Holthausen (1974) respectively. Words marked ‘onomatopoeic’, ‘expressive’, ‘(sound) imitative’ were selected (for example, L. *cuculus* ‘cuckoo’, *murmurare* ‘to rumble’, *crocire* ‘to croak’, *stridere* ‘to make a shrill sound’; OE *dynnan* ‘to make a noise, din’, *fneósun* ‘a sneezing’, *giellan* ‘to yell’). Words which originated as denotations of sounds were also added.

On the second stage of the research, the corpus of Old English glosses from Meritt (1945) was analysed for the presence of regular translation patterns of imitative words (e.g., *balare/bláetan* ‘to bleat’).

The talk will focus on the following *research questions*: (1) whether Latin onomatopoeic words were translated with Old English onomatopoeic words; (2) whether there were any cognate Latin/Old English pairs of words and (3) whether the Latin/Old English onomatopoeic word pairs are typologically similar (on the typology of onomatopoeic words – see Voronin, 2005).

## References

- Akita, K. 2013. Constraints on the semantic extension of onomatopoeia. In *Public Journal of Semiotics*, 5(1), 21–37.
- Anderson, E. 1998. *A Grammar of Iconism*. London: Associated University Press.
- Antilla, R. 1975. Affective vocabulary in Finnish: an(other) invitation. In *Ural-Altische Jahrbucher*, 47. 10–19.
- Bosworth, J., et al. *An Anglo-Saxon Dictionary Online*. ed. by T. N. Toller, Faculty of Arts, Charles University in Prague. URL: <https://bosworthtoller.com/> (accessed 2 Jan. 2023).
- Campbell, L. 2013. *Historical Linguistics. An Introduction*. Edinburgh: Edinburgh UP.
- Carling, G., E. Johansson, N., & Frid, J. 2020. *The evolutionary mechanisms of vocal iconicity. A study on basic vocabulary of the Indo-European family*. Abstract from SLE, 2020.
- de Vaan, M. 2008. *Etymological Dictionary of Latin and the Other Italic Languages*. Leiden: Brill.
- Dingemanse, M. 2012. advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10), 654–672.
- Durkin, Ph. 2009. *The Oxford Guide to Etymology*. Oxford: Oxford University Press
- Flaksman, M. 2017. Iconic treadmill hypothesis. the reasons behind continuous onomatopoeic coinage. In M. Bauer, A. Zirker, O. Fischer & Ch. Ljungberg (Eds.), *Dimensions of Iconicity* (pp. 15–38). Amsterdam: John Benjamins.
- Georgescu, S. 2018. Las ‘creaciones expresivas’ del *Diccionario crítico etimológico castellano e hispánico en contexto románico e indoeuropeo*. Un nuevo (y a la vez antiguo) enfoque. *Revista de Filología Románica*, 35, 119–139.
- Hinton, L., Nichols, J., & Ohala J. J. (Eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Hock, H.H. 1991. *Principles of Historical Linguistics*. Berlin: Mouton de Gruyter.
- Holthausen, F. 1974. *Altenglisches Etymologisches Wörterbuch*. Heidelberg: Carl Winter Universitätsverlag.
- Koleva-Zlateva, Zh. 2008. *Slavianskaya leksika zvukosimvolocheskogo proiskhozhdeniya* [Slavonic words of phonoiconic origin]. *Studia Slavica Hungarica*, 53/2, 381–395.
- Kozlova, T. O. 2013. Iconically expressible meanings in proto-Indo-European roots and their reflexes in daughter branches. In L. Elleström, O. Fischer & Ch. Ljungberg (Eds.), *Iconic Investigations* (pp. 311–330). Amsterdam: Benjamins.
- Kroonen, G. 2013. *Etymological Dictionary of Proto-Germanic*. Leiden-Boston: Brill.
- Lehmann, W. P. 1986. *A Gothic Etymological Dictionary*. Leiden: Brill.
- Lieberman, A. 2010. Iconicity and Etymology. In C. J. Conradie, R. J.M. Beukes, O. Fischer, & Ch. Ljungberg (Eds.), *Synergy* (pp. 243–258). Amsterdam: John Benjamins.
- Lühr, R. 1988. *Expressivität und Lautgesetz im Germanischen*. Heidelberg: Carl Winter Universitätsverlag.
- Malkiel, Ya. 1990. *Diachronic Problems in Phonosymbolism*. Amsterdam-Philadelphia: John Benjamins
- Meritt, H.D. 1945. *Old English glosses (a collection)*. Oxford: Oxford University Press.
- Moreno-Cabrera, J. C. 2020. *Iconicity in Language: An Encyclopaedic Dictionary*. Cambridge: Cambridge Scholars Publishing.
- Peirce, Ch. S. 1940. *The Philosophy of Peirce: Selected Writings* (ed. by J. Buchler). New York: Harcourt, Brace and Company.
- Sadowski, P. 2001. The sound as an echo to the sense. the iconicity of English gl- words’. In O. Fischer & M. Nänny (Eds.). *The Motivated Sign* (pp. 69–88). Amsterdam: Benjamins.
- Voeltz, E. F. K. & Kilian-Hatz, Ch. (Eds.). 2001. *Ideophones* [Typological Studies in Language 44]. Amsterdam-Philadelphia: John Benjamins.
- Voronin, S. V. 2005. *Iconicity. Glottogenesis. Semiosis: Sundry Papers*. St. Petersburg: St. Petersburg University Press.

## A new perspective on the evolution of mood and negation markers in Proto-Basque

Borja Ariztimuño & Iker Salaberri  
(University of the Basque Country)

Keywords: <reconstruction, morphosyntactic change, tense-aspect-mood, negation, Basque>

Recent decades have witnessed an increasing interest in historical Basque linguistics in general, and in reconstructing Proto-Basque (PB) in particular. This has led to meaningful advances in the reconstruction of PB phonology and morphology and, to a more limited extent, morphosyntax. Against that background, this paper addresses the evolution of PB markers of mood, modality and negation, which have not received much attention despite being well documented in textual records.

Early recorded stages of Basque present a complex picture concerning the expression of mood, modality and negation: some verbs are both marked with the suffix *-ke/-te* —which has various temporal, aspectual and modal uses— and accompanied by the modal auxiliary *ahal* (negative counterpart *ezin*), as in *Ayn fuerte eta galanto labradu-ric nola ecin ayn-beste mundu-an al ç-a-te-an* so strongly and beautifully build-PTCP that cannot so-much world-LOC can 3SG.PST-be-*te*-PST ‘Built so strongly and beautifully that it could not have been possible on Earth’ (Lazarraga 1567–1602, *apud* Bilbao et al. 2020: 76). By contrast, the same suffix seems to suffice as a modal marker when accompanied by the standard negator *ez*, cf. *Bat-a ez=pa-d-a nahi, ez=k-ita-ke-k gudu-ka ni eta hi* one-DEF NEG=SUB-3SG.PRS-PRS want NEG=1PL-AUX-*ke*-ALLC ‘If one of us doesn’t want to, we cannot fight each other, you and I’ (Oihenart 2003 [1657]: 256). In yet other cases the same functions are indicated only by means of *ahal/ezin*, compare *Ehor-c hura gayxteri-a-z ecin l-eça-n inbia* no.one-ERG 3SG evilness-DEF-INS cannot HYP-AUX-SUB envy ‘So that no one could envy her out of evilness’ (Etxepare 1980 [1545]: 94). All three kinds of verbal forms seem to be largely synonymous.

This state of affairs raises several questions: first of all, the aforementioned variation is suggestive of ongoing change, but it is unclear which marker, if any, originally served to indicate mood in the proto-language. Moreover, the numerous uses of the suffix *-ke/-te* —ranging from potential mood, through future reference, to doubt (conjecture) on the speaker’s part towards the truth of the proposition, among others (Lafon 1970)— provide few clues as to which use is original and what path of change subsequently ensued. According to Rebuschi (1984: 275–276) modal, conditional and conjectural uses are all derived from an initial predicative meaning, whereas Mounole (2014: 340–341), on the basis of cross-linguistically common paths of change (Bybee 1994: 265–266), argues that the shift must have been from modal to temporal. Inherent to these proposals is the view that the use of modal markers *ahal/ezin* is an innovation, i.e., they came to reinforce verbal forms with *-ke/-te* in order to avoid ambiguous modal-temporal readings.

In turn, it has been recently suggested that the suffix *-ke/-te* must have had an original negative meaning, which over time yielded temporal (future) reference (Ariztimuño & Salaberri 2022, from an earlier proposal by Ariztimuño 2016). Following this line of thought, here we put forward a three-stage path of change for the uses of *-ke/-te*:

**Stage #1:** *-ke/-te* is used as a marker of future tense, whereas the modal auxiliary *ahal/ezin* is the only means of indicating potential mood;

**Stage #2:** *-ke/-te* undergoes semantic bleaching and develops new meanings, including epistemic modality (conjecture). Consequently, verbal forms with *-ke/-te* become, in some contexts, ambiguous between a temporal and an epistemic modal meaning;

**Stage #3:** in order to avoid ambiguity, epistemic modal uses of *-ke/-te* are reinforced by the modal auxiliary *ahal/ezin*. As a consequence, *-ke/-te* is increasingly identified with and ultimately takes on a potential modal meaning it previously did not have.



This chain of stages (#1-3) is meant to account for two facts: (a) modal uses of *-ke/-te* seem to stem from an initial temporal meaning, judging by the fact that all other uses of this suffix (predicative, apodosis, conjecture i.e. epistemic modality, etc.) can also be derived by the same path; (b) the existence of rare potential verb forms with *ahall/ezin* but without *-ke/-te*—such as *Eta nehor-c ecin ihardets c-i-eço-yon hitz-ic* and no one-ERG cannot answer 3SG.PST-AUX-SBJ-3SG-PST word-PART ‘And no one could answer him a word’ (Leizarraga 1990 [1571]: 367)—in 16th-century Basque and the widespread prevalence of *ahall/ezin* suggest that the modal auxiliary must have been the original means of expressing mood (and negation) in the proto-language. Accordingly, this paper reconstructs a typologically uncommon path of change (temporal > modal) which is, however, supported by a careful interpretation of the data.

### Abbreviations

1 = 1st person; 2 = 2nd person; 3 = 3rd person; ALLC = allocutive; AUX = auxiliary; DEF = definite; ERG = ergative; HYP = hypothetical; INS = instrumental; ITER = iterative; LOC = locative; NEG = negator; PART = partitive; PL = plural; PRS = present tense; PST = past tense; PTCP = participle; SBJ = subjunctive; SG = singular; SUB = subordinator.

### References

- Ariztimuño, Borja. 2016. Euskal ezeztapena(k) ‘Jespersen Zikloen’ argitan [Basque negation(s) in light of Jespersen’s Cycles]. Talk held at the *Monumenta Linguae Vasconum mintegi irekiak* conference, University of the Basque Country, Vitoria-Gasteiz, 05/12/2016.
- Ariztimuño, Borja & Iker Salaberri. 2022. Reconstructing negation in a language isolate: Jespersen’s Cycle(s) and grammaticalization in Basque. Talk held at the *25th International conference in historical linguistics*, University of Oxford, Oxford, 03/08/2022.
- Bilbao, Gidor, Ricardo Gómez, Joseba A. Lakarra, Julen Manterola, Céline Mounole & Blanca Urgell. 2020. *Lazarraga eskuizkribuaren edizioa eta azterketa* [Edition and analysis of the Lazarraga manuscript], vol. 2: *Testua* [Text]. Bilbo/Bilbao: University of the Basque Country.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: tense, aspect and modality in the languages of the world*. London/Chicago: University of Chicago Press.
- Ettxepare, Bernard. 1980 [1545]. *Linguae vasconum primitiae* (ed. by Patxi Altuna). Bilbo/Bilbao: Euskaltzaindia, Ediciones Mensajero.
- Lafon, René. 1970. Le suffixe *-ke, -te* dans la conjugaison basque. *Bulletin de la Société Linguistique de Paris* 65(1). 184–212.
- Leizarraga, Joanes. 1990 [1571]. *Jesus Christ gure iaunaren testamentu berria, Othoizta ecclesiasticoen forma, Catechisme, Kalendrera, ABC edo christinoen instructionea* [New Testament of our lord Jesus Christ, Form of ecclesiastic prayers, Catechism, Calendar, ABC or christians’ teachings]. Bilbo/Bilbao: Euskaltzaindia.
- Mounole, Céline. 2014. Le verbe basque ancien: étude philologique et diachronique. *International Journal of Basque Linguistics and Philology* [ASJU] 48(1/2). 1–506.
- Oihenart, Arnaud. 2003 [1657]. *Euskal atsotitzak eta neurtitzak* [Basque proverbs and poems] (ed. by Patxi Altuna and Jose A. Mujika). Bilbo/Bilbao: Euskaltzaindia.
- Rebuschi, Georges. 1984. *Structure de l’énoncé en basque*. Paris: Société d’Études Linguistiques et Anthropologiques de France.

## Verified Computational Rule-based Historical Phonology in Standard ML and Isabelle/HOL

This paper introduces an implementation of rule-based phonology in Standard ML and a formal definition and verification of the core components of such phonology in Isabelle/HOL, an interactive theorem prover. This phonology is used to automatically derive modern reflexes of Spanish, Portuguese, Chinese and Sino-Korean from their ancestral etyma using one underlying model. The architecture of this program is as such: We first implement a featureful segmental inventory, which means that each segment is not merely a character literal but data types with feature information. Then we define the constituents of syllables and syllables themselves by gluing the segments into nested lists. Once syllables are defined, phonological words come naturally, as they can be implemented as lists of syllables. Finishing the definition of phonological words means that we can represent all of the etyma and reflexes in the languages that we are investigating in this project. The rest of the program deals with the definition of the operations on those data, namely sound changes, and convenient utilities to help us define all of the sound changes happened in the history of these four languages.

A sound change in our system is represented as function mapping a phonological word to another, which matches the intuition of a working linguist. One may stop here and start implementing all of the sound changes as recursive ML functions on lists, as the phonological words are so represented in our system, but this approach is tedious, certainly may result in a lot of boilerplate, for that many sound changes in world languages only differ in some details and are structurally similar; it is also error-prone, as the reduplication of similar routines in the code base usually is. Thus instead of writing those sound changes entirely by hand, most of the sound changes in our system are created through schema that decouples the problem into manageable modular pieces. Here is the concrete explanation: just like we have roughly three tiers in the representation of a phonological word: the segments (which themselves are products of features), syllables (that have constituents like onset, nucleus, and coda), and phonological words. Our strategy is to define utilities that would rewrite one tier at a time and compose them into an actual sound change.

These two components, the component that represents data and the component that rewrites data, constitutes the trusted kernel of the program; this kernel is what we are going to verify in Isabelle/HOL. Isabelle is a member of HOL family of theorem provers. It is based on Higher-Order Logic, which a battle-tested logical system that is more than enough to verify our system to secure the desired behaviors. This verified kernel is shared among all the languages whose history we implemented; the only two language dependent parts of our system are the etyma that are to be rewritten and the respective sound changes in those languages. It should be noted that although there are four languages in our project, we only need two sets of etyma: Latin for Spanish and Portuguese, Chinese for both Chinese and Sino-Korean. Even more, Spanish and Portuguese share all the sound changes in our system until their split in Medieval times. Spanish and Portuguese are chosen precisely because of their similarities, so that we can demonstrate another feature of our system: our program is able to output and parse reflexes coupled with their history represented a list of sound changes, which enabled us to: define sound changes that are shared among Spanish and Portuguese, apply them to the etyma, and store the results so that the now separate Spanish and Portuguese modules can deal with the Romance etyma respectively.

The entirety of this program is written in Standard ML '97 and is able to be compiled both by the SML/NJ and MLton compilers. The proof scripts are written in Isabelle/HOL 2022. The source code of this project will be released under the BSD-3-Clause license.

[Wil38] Edwin B. Williams. *From Latin to Portuguese: Historical Phonology and Morphology of the Portuguese Language*. University of Pennsylvania Press, 1938.



- [Llo87] Paul M. Lloyd. *From Latin to Spanish*. American Philosophical Society, 1987.
- [Pau90] Lawrence C. Paulson. “Isabelle: The Next 700 Theorem Provers”. In: *Logic and Computer Science* (1990), pp. 361–386.
- [Pau96] Lawrence C. Paulson. *ML for the Working Programmer*. Cambridge University Press, 1996.
- [Pen02] Ralph Penny. *A History of the Spanish Language*. Cambridge University Press, 2002.
- [Har03] Steven Lee Hartman. *Phono (Version 4.0): Software for Modeling Regular Historical Sound Change*. 2003. URL: <https://langhist.weebly.com/files/theme/ver40.pdf> (visited on 10/18/2022).
- [Hua05] José Ignacio Hualde. *The Sounds of Spanish*. Cambridge University Press, 2005.
- [Sch09] Axel Schuessler. *Minimal Old Chinese and Later Han Chinese: A Companion to Grammata Serica Recensa*. University of Hawai’i Press, 2009.
- [BS14] William H. Baxter and Laurent Sagart. *Old Chinese: A New Reconstruction*. Oxford University Press, 2014.

**Where and How?**  
**Request verb constructions in Ancient Greek**  
 Abstract

Ancient Greek shows different ways of expressing speech acts of request (cf. Dickey 1996 and 2016; Denizot 2011). One of the most common strategies consists of using a request verb followed by a structure that expresses the content of the request, in other words, what the speaker would like the addressee to perform. Various factors influence how to make a request and where to insert it, and a wide range of variation can be observed across languages. However, two elements always play a significant role in shaping the construction of the request: the relationship between speaker and addressee and the context in which the utterance is performed (i.e., communication setting, discourse structure, and topic).

This paper explores speech acts of request in Ancient Greek that indicate the usage of a request verb. Specifically, it aims at addressing the variation within the usage of request verbs in relation to where they occur (i.e., type of text and their position within the discourse) and how they are constructed (i.e., their corresponding syntactic constructions, their pragmatic functions, and the co-occurrence of discourse particles).

The main contribution of the present paper is a combination of diachronic and synchronic analyses focusing on different language stages (i.e., Classical and Postclassical Greek) and various types of texts. In particular, it integrates data from documentary papyri into the analysis of literary sources, which has typically been the focus of scholarly research. Documentary papyri provide evidence of the Greek language with a continuity of more than a millennium (4th cent. BCE – 8th cent. CE) and reflect language usages that are more oriented to everyday communication purposes instead of adhering to literary conventions and genre constraints (cf. Dickey 2011).

In addition to verbs such as αἰτέω / *aitéō* ('ask, demand'), ἐρωτάω / *erōtáō* ('ask about'), and παρακαλέω / *parakaléō* ('exhort/beg'), two verbs are particularly worth considering: δέομαι / *déomai* and ἀξιόω / *axióō*. They occur in literary sources of the Classical period with the meaning of 'to be in want or need' and 'think worthy of / esteem', respectively; later on, they are used in documentary papyri for introducing a request (cf. Di Bartolo 2021). Specifically, petitions of the Ptolemaic period are the first type of documentary texts that attest to the new meaning of these two verbs. Accordingly, δέομαι / *déomai* occurs in petitions addressed to kings (i.e., ἐντεύξεις / *enteúxeis*) and ἀξιόω / *axióō* in petitions addressed to officers (i.e., ὑπομνήματα / *hupomnēmata*) in order to express a request (cf. Baetens 2020; Di Bitonto 1967 and 1968). Later on, papyrus private letters also show the occurrence of both verbs in request speech acts.

The investigation also addresses questions related to language change, synchronic variation, and information structure, discussing different occurrences of the above-mentioned request verbs. Instances will be collected by means of secondary literature, TLG, the two papyrus databanks *DDbDP* and *Trismegistos*, and the linguistically annotated treebank corpus *PapyGreek*. The paper illustrates the changes within the range of the syntactic patterns of these verbs between the Classical and the Postclassical periods and considers the pragmatic function of discourse particles occurring with them. A comparison with analogous request constructions in other ancient Indo-European languages, such as Latin, will be drawn. Furthermore, the paper addresses the semantic shifts of δέομαι / *déomai* and ἀξιόω / *axióō* in the context of Greek petitions, explaining them in terms of Traugott's subjectification theory (cf. Traugott 1995).

**References:**

- Baetens, Gert. 2020. *I am wronged. Petitions and Related Documents from Ptolemaic Egypt (332 - 30 BC)*. Leuven: Trismegistos online publications.
- Di Bitonto, Anna. 1967. “Le petizioni al re”. *Aegyptus* 47. 5-57
- Di Bitonto, Anna. 1968. “Le petizioni ai funzionari nel periodo tolemaico”. *Aegyptus* 48. 53-107.
- Denizot, Camille. 2011. *Donner des ordres en grec ancien: étude linguistique des formes de l'injonction*. Mont-Saint-Aignan: Presses des l'Universités de Rouen et du Havre.
- Di Bartolo, Giuseppina. 2021. *Studien zur griechischen Syntax dokumentarischer Papyri der römischen Zeit*. (Papyrologica Coloniensia Vol. XLIV). Paderborn: Ferdinand Schöningh.
- Dickey, Eleanor. 1996. *Greek Forms of Address. From Herodotus to Lucian*. Oxford: Oxford University Press.
- Dickey, Eleanor. 2011. The Greek and Latin Languages in the Papyri. In Roger S. Bagnall (ed.), *The Oxford Handbook of Papyrology*, 149–169. New York / Oxford: Oxford University Press.
- Dickey, Eleanor. 2016. Emotional language and formulae of persuasion in Greek papyrus letters. In Ed Sanders and Matthew Johncock (eds.), *Emotion and Persuasion in Classical Antiquity.*, 237-262. Stuttgart: Franz Steiner.
- Traugott, Elisabeth C. 1995. “Subjectification in Grammaticalization”. In Dieter Stein and Susan Wright (eds.), *Subjectivity and Subjectivisation*, 31–54. Cambridge: Cambridge University Press.

## Evolving rhythms:

### A quantitative assessment of rhythmic alternation in the history of English

Stress-based languages such as English favor alternating rhythms made up of stressed and unstressed syllables (Selkirk 1984; Kelly & Bock 1988). At a basic level of rhythmic structure, this means that consecutive stressed syllables, i.e. ‘clashes’ (e.g. *búild bäck bétter*), and consecutive unstressed syllables, i.e. ‘lapses’ (e.g. *survival of the fittest*), are dispreferred. Non-optimal patterns typically trigger prosodic repair mechanisms such as pauses or prominence adjustments. However, Breiss and Hayes (2020) have demonstrated statistically that rhythmic optimization reaches beyond phonology, showing that clashes are not only repaired with prosodic means but outright avoided through syntactic (or lexical) choices (see also Schlüter 2005; Shih et al. 2015; Anttila 2016).

The proposed study extends this line of inquiry into diachrony by asking (a) whether a global preference for rhythmically optimized patterns can be detected throughout the history of English, and, if so, (b) whether it is possible to identify lexical or morpho-syntactic developments that have helped to stabilize or even improve rhythmic well-formedness.

To this end, the study investigates rhythmicity in the history of English, measured in terms of the occurrence probabilities of clashes and lapses in word bigrams sampled from Middle, Early Modern and Modern English texts. The data for the analysis come from the Penn-Helsinki Parsed corpora of English (Kroch & Taylor 2000; Kroch, Santorini & Delfs 2004; Kroch, Santorini & Diertani 2016). Quantitative analysis is carried out with R (R Core Team 2023), using linear and generalized additive models (GAMs) (Wood 2017). Apart from time period, various other predictor variables encoding prosodic and morpho-syntactic constituency will be taken into account.

Identifying clashes and lapses in the historical texts crucially depends on the correct interpretation of (the reflexes of) unstressed inflections and monosyllabic function words. The former are generally not pronounced as syllabic in Present-Day English, but their status in Middle English is not always obvious in the individual case (e.g. ME *makede* ‘made’). Monosyllabic function words generally exhibit low stress probabilities in Present-Day English, but might have been more prominent in earlier stages of the language (e.g. ME *thou schalt haue* ‘you will have’). To account for these complications in a systematic manner, evidence from contemporaneous metrical verse (e.g. Chaucer, Lydgate, Shakespeare, Spenser) will be used to calculate probability scores for inflectional syllabicity and monosyllable stress with the help of machine learning techniques such as conditional inference trees and hierarchical clustering (Levshina 2015).

Preliminary results suggest that overall rhythmicity has not changed much since Middle English. However, it can also be shown that the diachronic process of schwa loss (Minkova 1991) must have posed a major challenge to rhythmic well-formedness, as it would have significantly increased the occurrence of clashes, had it not been offset by various structural adjustments, including analyticization (e.g. ME *Gódes sóne* vs. PDE *son of Gód*). The results are interpreted in terms of language evolution (Croft 2000; Baumann & Ritt 2017): prosodic preferences act as a selective pressure tipping the balance in favor of rhythmically more optimal syntactic (or lexical) variants.

## References

- Anttila, Arto. 2016. Phonological effects on syntactic variation. *Annual Review of Linguistics* 2(1), 115–137. doi:10.1146/annurev-linguistics-011415-040845.
- Baumann, Andreas & Nikolaus Ritt. 2017. On the replicator dynamics of lexical stress: accounting for stress-pattern diversity in terms of evolutionary game theory. *Phonology* 34(3), 439–471.
- Breiss, Canaan & Bruce Hayes. 2020. Phonological markedness effects in sentence formation. *Language* 96(2), 338–370. doi:10.1353/lan.2020.0023.
- Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow, England; New York: Longman.
- Kelly, Michael H. & J. Kathryn Bock. 1988. Stress in time. *Journal of Experimental Psychology* 14(3), 389–403. doi:10.1037/0096-1523.14.3.389.
- Kroch, Anthony, Santorini, Beatrice, & Delfs, Lauren. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony, Santorini, Beatrice, & Diertani, Ariel. 2016. The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony & Taylor, Ann. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania. <http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4>.
- Levshina, Natalia. 2015. *How to do linguistics with R: data exploration and statistical analysis*. Amsterdam: John Benjamins. doi:10.1075/z.195.
- Minkova, Donka. 1991. *The history of final vowels in English: the sound of muting*. Berlin, New York: Mouton de Gruyter.
- R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schlüter, Julia. 2005. *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. Berlin; New York: Mouton de Gruyter.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.
- Shih, Stephanie, Jason Grafmiller, Richard Futrell & Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In Ralf Vogel & Ruben Vijver (eds.), *Rhythm in cognition and grammar: a Germanic perspective*. Berlin: De Gruyter Mouton, 207–233.
- Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC. doi:10.1201/9781315370279.

## The History of /pf/ in New Braunfels German: Another Case of Rule Inversion?

The status of rule inversion, i.e. “reversal of the input and output of a rule and complementation of the environment” (McCarthy 1991: 194), as a mechanism of language change remains controversial. While scholars like Vennemann (1972) have argued that it is a relatively common process, others, like McCarthy (1991), have contended that rule inversion is at best very rare. Additional examples of rule inversion would help resolve this controversy. This paper therefore addresses another potential example of rule inversion, involving the history of the affricate /pf/ in New Braunfels German (NBG), a critically endangered New World variety of German. According to Eikel (1954), which is based on data collected in the 1940s and 1950s, /pf/ did not appear word-initially in NBG, meaning that words beginning with [pf] in Standard German, e.g. *Pferd* ‘horse’, *Pfeffer* ‘pepper’, and *Pfirsich* ‘peach’, began with [f] in NBG. It did, however, appear word-medially and word-finally, e.g. in *Topf* ‘pot’. In light of the numerous similarities between Standard German phonology and NBG phonology discussed by Eikel, this situation indicates that a sound change from /pf/ to /f/ (deaffrication) had taken place.

The situation soon changed: Gilbert (1972: Map 103), which is based on data collected in the 1960s, notes that his informants pronounced words like *Pferd* with an initial [pf] (e.g. 100% of his informants produced an initial [pf] in *Pferd*). In other positions within the word, Gilbert’s informants used both [pf] and [f]. This indicates that the earlier sound change had been undone, as /f/ had been affricated in word-initial position to [pf].

In the data collected by members of the Texas German Dialect Project (TGDP; [www.tgdp.org](http://www.tgdp.org)) since 2001, the situation has changed again. According to Boas (2009), only 8% of his informants pronounced words like *Pferd* with an initial [pf]. This shows that the affrication process indicated by the Gilbert data has largely been undone, i.e. that /pf/ has again been deaffricated to /f/.

I argue that the best account of the NBG facts is a relatively straightforward sequence of sound changes, /pf/ > /f/ > /pf/. This account outperforms other possible analyses of the Texas German data, like the model of new dialect formation proposed by Trudgill (2004), which can account for the differences between the Eikel data and the Gilbert data, but not for the differences between the Gilbert data and the TGDP data (Boas 2009). These changes admittedly do not correspond precisely to the classical definition of rule inversion, since they take place in the same environment, and do not involve the “complementation of the environment,” as true rule inversion does, but they do involve the “reversal of the input and output of a rule.” While this particular development may therefore not be a clear-cut example of rule inversion, the reasoning here could be extended to cover other phonological phenomena in Texas German (e.g. the unexpected occurrence of front rounded vowels in some contexts), which may yet reveal such examples of rule inversion in Texas German.

### References

- Boas, Hans C. 2009. *The life and death of Texas German*. Durham: Duke Univ. Press.  
 Eikel, Fred. 1954. *The New Braunfels German dialect*. Dissertation, Johns Hopkins.  
 Gilbert, Glenn. 1972. *The linguistic atlas of Texas German*. Austin: UT Press.  
 Trudgill, Peter. 2004. *New-dialect formation*. Edinburgh: Edinburgh Univ. Press.  
 Vennemann, Theo. 1972. Rule inversion. *Lingua* 29: 209-242.

**Ditransitive GIVE-construction in three Hainan Min-Chinese:  
Interaction between inherited structures and contact-induced changes**

Xiuwei Zeng (The Chinese University of Hong Kong)

Ditransitive *GIVE*-constructions in Sinitic languages can be classified into two types: (1) the “canonical” [V-IO-DO] construction (i.e., “give me a book”), which is found in Mandarin, Northern Chinese and Southern Min-Chinese; and (2) [V-DO-IO] construction (i.e., “give a book me”), which is common in Southern Chinese (Hashimoto 1976). Hainan Min-Chinese is a variety of Southern Min-Chinese consisting of various dialects. It has undergone intense language contact with the indigenous Kra-Dai languages (Hlai and Be) and other Chinese varieties on the Hainan Island for more than one millennium (Liu 2006). Cao (2008) claims that Hainan Min differs from other Southern Min varieties in employing the [GIVE-DO-IO] construction, as well as using the verb /ʔio/ ‘take’ as a *GIVE* verb. Zhang (2011) further argues that this is the result from the omission of dative markers in prepositional dative constructions (i.e., “take a book (tə) me”) under the pressure of contact with [V-DO-IO]-type Cantonese.

This paper presents evidence from three Hainan Min dialects (Haikou, Qionghai, Gangmen) showing that [GIVE-IO-DO] (Southern Min) construction is strongly preferred, and that the [GIVE-DO-IO] (non-Southern Min Southern Chinese) construction is still in its infancy of development. By comparing ditransitive *GIVE*-constructions in the three Hainan Min dialects, with reference to other Southern Min varieties, early modern vernacular Southern Min texts, other Southern Chinese varieties and the indigenous Be language, I have identified the origin and the historical strata of *GIVE* verbs. In present-day Hainan Min, *bun* ‘distribute’, *ʔio* ‘take’ and *khi* ‘beg/give’ all coexist as *GIVE* verbs. I argue that Hainan Min inherited *khi* and [GIVE-IO-DO] construction from Southern Min, while the sememe {GIVE} of *bun* and *ʔio* were introduced through language contact with Hakka (Sinitic) and Be (Kra-Dai) respectively at different stages of historical development.

Previous studies have also proposed an implicational universal about ditransitive constructions in Sinitic languages: “**absence of [GIVE-IO-DO] constructions**  $\supset$  **absence of R-type GIVE verbs**” (Zhang 2011; Phua 2015; Phua and Xiang 2020). R-type *GIVE* verbs refer to *GIVE* verbs that introduce recipient argument (Li and Wu 2015; cf. Margetts and Austin 2010). It is further postulated that for a verb meaning *TAKE/HOLD/DISTRIBUTE* to be used as a ditransitive *GIVE* verb in the [GIVE-IO-DO] construction, it has to undergo DO-fronting and preposition incorporation to be a R-type *GIVE* verb (Li and Wu 2015; Xia 2017). These proposals are challenged by the Haikou and Gangmen dialects of Hainan Min, which use *bun* ‘distribute’ and *ʔio* ‘take’ as ditransitive *GIVE* verbs in the [GIVE-IO-DO] construction, as they do not allow *bun* and *ʔio* to introduce recipient argument without preposition in between. Preposition incorporation is either in progress (Haikou dialect), or has not begun (Gangmen dialect). I argue that these unusual syntactic behaviors of *bun* and *ʔio* are due to the fact that they did not undergo the development from *DISTRIBUTE/TAKE* verbs into *GIVE* verbs through DO-fronting and preposition incorporation. Instead, the polysemous patterns of *bun* (‘distribute’/ ‘give’) and *ʔio* (‘take’/ ‘give’) were copied into Hainan Min at the time of contact, despite that the corresponding morpheme of *ʔio* in Be only appears in [GIVE-DO-IO] construction. *bun* and *ʔio* replaced the native *GIVE* verb *khi* and occupied its position in the inherent [GIVE-IO-DO] constructions, making Hainan Min a rare exception to established patterns. The findings illustrate examples of polysemous pattern being transferred in contact situation regardless to the mismatch in syntactic structures, and it is mapped onto the inherent structures in the recipient language without causing change in word order.

**keywords:** ditransitive construction, language contact, Hainan Min-Chinese



## References

- Cao, Zhiyun. ed. 2008. *Hanyu fangyan dituji* (Linguistic Atlas of Chinese Dialects). Beijing: Commercial Press.
- Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: problems of typological diversity in Sino-Tibetan. *Computational Analyses of Asian and African Languages* 3:49–65.
- Li, Xuping and Yicheng Wu. 2015. Ditransitives in three Gan dialects: valence increasing and preposition incorporation. *Language Sciences* 50:66-77.
- Liu, Xinzhong. 2006. *Hainan minyu de yuyin yanjiu* (Phonetics and phonology of the Min dialect in Hainan). Beijing: China Social Sciences Press.
- Margetts, Anna and Peter K. Austin. 2007. Three Participant Events in the Languages of the World: Towards a Crosslinguistic Typology. *Linguistics* 45 (3), 393–451.
- Phua, Chiew Pheng and Mengbing Xiang. 2020. Liancheng xinquan kejiahua de jiyu dongci he jiyulei shuangjiwu jiegou (The GIVE verbs and ditransitive give-type construction in Liancheng Xinquan Hakka dialect). *Yuyanxue Luncong* 61:50-91.
- Phua, Chiew Pheng. 2015. *Shanggu hanyu yuge jushi yanjiu* (Study in dative constructions in archaic Chinese). Beijing: Commercial Press.
- Xia, Liping. 2017. Jufa kucang liebian: cong liandongshi dao jiyulei shuangjiwu jiegou (Syntactic inventory fission: from serial verb construction to ditransitive GIVE-construction). *Bulletin of Linguistic Studies* 18:25-46.
- Zhang, Min. 2011. Hanyu fangyan shuangjiwu jiegou nanbei chayi de chengyin: leixingxue jianjiu yinfa de xinwenti (Revisiting the alignment typology of ditransitive constructions in Chinese dialects). *Bulletin of Chinese Linguistics* 4 (2), 87–270.

## **A Phylogenetic Study of the Cariban Family: Combining Linguistic and Archaeological Data**

This study presents a preliminary linguistic phylogenetic analysis of the Cariban language family, a group of indigenous languages in Northern South America and Central Brazil. The family comprises approximately three dozen languages, most of which are spoken by small communities of a few hundred speakers (Hammarström et al., 2022). The languages are closely related in phonology and lexicon. However, except for a few shallow clades, most of the family's history is unknown and subject to many theories on its homeland and expansion routes (Meira & Franchetto, 2005; Gildea, 2012).

We follow the best practices in current computational historical linguistics (Hoffman et al., 2021; Greenhill et al., 2020; Jäger, 2019; Tresoldi et al., 2022), including initial analyses with neighbournets (Huson & Bryant, 2006) and Bayesian MCMC inference (Bouckaert et al., 2019) using different evolutionary models. We collected the data from reliable sources (Matter, 2020; de Tauste, 1680; Courtz, 2008; Largo, 2011; Ruiz Blanco 1888 [1690], von den Steinen 1892; Koehn & Koehn, 1986). It is organised into an independent, normalised, and open-access database in CLDF format (Forkel et al., 2018), carrying cognate assignments made by experts (Carvalho et al., *forth.*). We will discuss the results of our classification in the context of earlier classifications of the Cariban language family, including those by Derbyshire (1999), Meira (2006), Gildea (2012), and Meira et al. (2015), such as the statistical support of our findings for consensual and nearly consensual clades (e.g., Parukotoan, Pekodian) and for other clades proposed in the literature (e.g., the “Venezuelan” branch, Wayana-Apalai, Panare-Pemongan). Besides our maximum clade credibility (MCC) tree informed by linguistic and archaeological data, we will present our first phylogeographic inference models (Lemey et al., 2009).

This study is part of a larger initiative to analyse the linguistic history of South America. It will contribute analyses and insights into the evolution and relationships within the Cariban language family, including the location of its homeland, the date of its first expansions, and its migration movements. For general phylogenetic inference, it will contribute to the discussion on the solution and strategies for incorporating archaeological data, a necessary step when classifying language families without written records extending over multiple centuries. The classification based on lexical evidence will motivate the search for shared innovations in phonology and morphology, paving the way for the reconstruction of intermediate-level proto-languages using traditional and computer-assisted methods. It will likewise expand the foundations for research on language contact among South American native languages.

## References

- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019) “BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis”. *PLoS computational biology*, 15(4), e1006650.  
<https://doi.org/10.1371/journal.pcbi.1006650>
- Courtz, H. (2008). *A Carib grammar and dictionary*. Magoria books.
- de Tauste, F. (1680). *Arte y vocabulario de la lengua de los indios chaymas, cumanagotos, cores, parias, y otros diversos de la provincia de Cumana, o Nueva Andalucia*. Madrid: Bernardo de Villa.
- Derbyshire, D. (1999). “Carib”. In Dixon, R. M. W. & Aikhenvald, A. (eds.), *The Amazonian Languages*, 23-64. Cambridge: Cambridge University Press.
- Carvalho, F. O., Ferraz Gerardi, F., Aragon, C. & Tresoldi, T. (forth.). *Cariban Lexical Database*.
- Forkel, R., List, J.-M., Greenhill, S. J., Bank, S., Rzymiski, C., Cysouw, M., Hammarström, H., Haspelmath, M., Kaiping, G. A. & Gray, R. D. (2018). “Cross-linguistic Data Formats, advancing data sharing and reuse in comparative linguistics”. *Scientific Data*, 5:180205.  
<https://doi.org/10.1038/sdata.2018.205>
- Gildea, S. (2012). “Linguistic studies in the Cariban family”. In Campbell, L. & Grondona, V. (eds.), *The Indigenous Languages of South America: A Comprehensive Guide*, 441-494. Berlin, Boston: Berlin: Mouton.
- Hammarström, H., Forkel, R., Haspelmath, Martin & Bank, S. (2022). *Glottolog 4.7*. Leipzig: Max Planck Institute for Evolutionary Anthropology.  
<https://doi.org/10.5281/zenodo.7398962> . Available online at <http://glottolog.org>, accessed on January 17th, 2023.
- Hoffmann, K., Bouckaert, R., Greenhill, S. J. & Kühnert, D. (2021). “Bayesian phylogenetic analysis of linguistic data using BEAST”. *Journal of Language Evolution*, 6: 119–135.  
<https://doi.org/10.1093/jole/lzab005>
- Huson, D. H. & Bryant, D. (2006). “Application of Phylogenetic Networks in Evolutionary Studies”. *Molecular Biology and Evolution*, 23(2): 254-267.  
<https://doi.org/10.1093/molbev/msj030>
- Greenhill, S. J., Heggarty, P., & Gray, R. D. (2020). “Bayesian Phylolinguistics”. In Janda, R. D., Joseph, B. D. & Vance, B. S. (eds) *The Handbook of Historical Linguistics*, Volume II, pp. 226–253. Wiley-Blackwell: New Jersey.

- Jäger, G. (2019). “Computational historical linguistics”. *Theoretical Linguistics*, 45(3-4): 151-182. <https://doi.org/10.1515/tl-2019-0011>
- Koehn, E., & Koehn, S. (1986). “Apalai”. In D. S. Derbyshire & G. K. V. I. Pullum (Eds.), *Handbook of Amazonian languages*, Vol. 17: 33-127. Berlin: de Gruyter.
- Largo, W. (2011). *Una Gramática del Yukpa (Colombia)*. Privately published.
- Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. (2009). “Bayesian Phylogeography Finds Its Roots”. *PLoS Computational Biology* 5, e1000520. <https://doi.org/10.1371/journal.pcbi.1000520>
- Matter, F. (2020). *Comparative Cariban Database*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <https://cariban.clld.org/>, accessed on January 17th, 2023.
- Meira, S. (2006). “Cariban Languages”. In Brown, K. (ed.), *Encyclopedia of Language and Linguistics*, 199-203. 2nd edn. Amsterdam: Elsevier.
- Meira, S. and B. Franchetto (2005). The Southern Cariban languages and the Cariban family. *International Journal of American Linguistics* 71: 127-192.
- Meira, S., Birchall, J., Chousou-Polydouri, N. (2015). “A character-based internal classification of the Cariban family”. Talk. *48th Annual Meeting of the Societas Linguisticae Europaea*. Leiden, Netherlands. Available online at [https://www.academia.edu/15980095/A\\_character-based\\_internal\\_classification\\_of\\_the\\_Cariban\\_language\\_family](https://www.academia.edu/15980095/A_character-based_internal_classification_of_the_Cariban_language_family), accessed on January 17th, 2023.
- Ruiz Blanco, M. (1888 [1690]). *Arte y tesoro de la lengua cumanagota (Vol. 3)*. Leipzig, Germany: B.G. Teubner.
- von den Steinen, K. (1892). *Die Bakairí-Sprache: Wörterverzeichnis, Sätze, Sagen, Grammatik mit beiträgen zu einer Lautlehre der Karäibischen Grundsprache*. Leipzig: K. F. Koehler's Antiquarium.
- Tresoldi, T., Rzymiski, C., Forkel, R., Greenhill, S. J., List, J.-M. & Gray R. (2022). “Managing historical linguistic data for computational phylogenetics and computer-assisted language comparison”. In Berez-Kroeker, A. L., McDonnell, B., Koller, E. & Collister, L. B. (eds). *Open Handbook of Linguistic Data Management*. MIT Press. <https://doi.org/10.7551/mitpress/12200.003.0033>

### Evidence for a Chibcha-Jê connection

The genealogical composition of the South American continent is a major puzzle of historical linguistics. This talk discusses lexical and grammatical evidence for a genealogical relationship between Chibchan languages of Central America and northern South America and Macro-Jê languages, predominantly spoken in the Brazilian lowlands, south of the Amazon. The correspondences include seventeen grammatical morphemes and twenty-four lexical items (Pache 2023), such as corresponding suppletive forms – Proto-Chibchan \*<sup>n</sup>daʔ ~ \*taʔ ‘to go<sub>1</sub>’, \*<sup>m</sup>bã ‘to go<sub>2</sub>’, Proto-Northern Jê \*tẽ ‘to go<sub>1</sub>’, \*mõ ‘to go<sub>2</sub>’ – and (near-)homophones such as Proto-Chibchan \*siʔ ‘thorn, tooth (incisor)’, \*sih ‘meat, flesh’, Proto-Northern Jê \*-jĩ ‘thorn’, \*-jĩ ‘meat’ (Pache 2018, Nikulin 2020). Among the further evidence is an unusual parallel grammaticalization path in Chibchan and Macro-Jê languages, which may reflect variation that existed in the shared ancestor language (see Joseph 2012). For instance, Boruca, a Chibchan language of southeastern Costa Rica, has an inessive postposition *kabá* (Quesada Pacheco 2019: 103); its cognate counterpart in Rama (eastern Nicaragua) encodes benefactive and purposive meanings (Craig 1989: 207), as shown in (1).

Rama (Chibchan)

- (1) *jaŋ-kama*  
 what-PURPOSE  
 ‘why’ (Craig 1989: 206)

Both Boruca inessive *kabá* and Rama purposive *kama* derive from a Proto-Chibchan form \*ka<sup>m</sup>ba by regular sound change (see Pache 2018). The reflex of its Proto-Northern Jê counterpart, locative \*kãm ~ \*kãm (Nikulin 2020: 511) likewise indicates the inessive in Apinajé (Northern Jê, central Brazil) (Oliveira 2005: 147), and, in a particular construction illustrated in (2), purpose.

Apinajé (Northern Jê)

- (2) *ja kamã*  
 this PURPOSE  
 ‘for that, for this reason, that’s why’ (Oliveira 2005: 147)

Together, the correspondences in question suggest a genealogical connection between Chibchan and Macro-Jê languages. This also has implications for our understanding of migration and the populating of the Americas.

### References

- CRAIG, COLETTE G. 1989. A Grammar of Rama. Université de Lyon. Report to National Science Foundation, BNS 8511156. Unpublished manuscript.
- JOSEPH, BRIAN D. 2012. A variationist solution to apparent copying across related languages. Copies versus Cognates in Bound Morphology, ed. Lars Johanson and Martine Robbeets, pp. 151–64. Studies in Language, Cognition and Culture, 2. Leiden: Brill.
- NIKULIN, ANDREY. 2020. Proto-Macro-Jê: um estudo reconstrutivo. Ph.D. dissertation, Universidade de Brasília.
- OLIVEIRA, CHRISTIANE CUNHA DE. 2005. The Language of the Apinajé People of Central Brazil. Ph.D. dissertation, University of Oregon.
- PACHE, MATTHIAS. 2018. Contributions to Chibchan Historical Linguistics. Ph.D. dissertation, Universiteit Leiden.
- \_\_\_\_\_. 2023. Evidence for a Chibcha-Jê connection. International Journal of American Linguistics 89 (2).
- QUESADA PACHECO, MIGUEL ÁNGEL. 2008. Gramática de la lengua guaymí (ngäbe). Languages of the World/Materials, 474. Munich: LINCOM Europa.

### Epistemic modality out of ‘playfulness’: Modern Greek *pezi*

The emergence of epistemic constructions is a well-known and much-discussed issue, with various claims regarding the possible sources, the directionality of changes and the morphosyntactic properties of such modals and constructions (cf. e.g. Traugott & Dasher, 2002, Narrog, 2012, Hilpert, Cappelle & Depraetere 2021, among many others). The article aims to contribute to this trend of research by presenting the emergence of a novel epistemic construction in Modern Greek, which had escaped notice till now and is based -rather surprisingly- on the verb *pezo* (=play), as illustrated below:

- (1) O Mike Flanagan ... ***pezi na ine*** to mono atomo pu bori.  
 The Mike Flanagan ... plays that be the only person that can  
 “Mike Flanagan... may be the only person that is able to”.  
 (*watchandchill.gr*, retrieved 19/1/23)

Based on an examination of available corpora of Modern Greek (TextCorpora, Corpus of Modern Greek), as well as an online search, the article shows that:

- a) The epistemic construction ‘*pezi na*’ emerged recently (last 10-15 years), most probably in spoken varieties of Modern Greek, and has gained considerable token frequency in everyday registers
- b) The morphosyntactic properties of the construction follow closely well-known typological (and Modern Greek) trends, for instance the verb is attested only in 3<sup>rd</sup> person singular, it exhibits only two tense forms (past / non-past) etc.

As far as the diachronic origin of the construction is concerned, the article argues for a multiple-source causation, illustrating the inter-relationship between the epistemic ‘*pezi na*’ and specific constructions involving the very same verb with different but related semantics (‘*dhen pezete*’ = you cannot beat this! / ‘*pezete*’= something is at stake / ‘*pezete*’= something is not finalized or is not certain yet) that were already at place.

Drawing on the observations mentioned above, the article:

- a) Broadens our perspective on the possible sources of epistemic constructions cross-linguistically
- b) Discusses the complex inter-relationship between constructions and its possible outcomes, drawing on recent Construction Grammar insights (cf. e.g. Sommerer & Smirnova, 2020)
- c) Attempts to draw on its usage-based approach to answer the elusive actuation question: Why ‘*pezi na*’ develops into an epistemic construction in 21<sup>st</sup> c. Greece?

**References**

- Hilpert M., B. Cappelle & I. Depraetere (eds.), 2021. *Modality and Diachronic Construction Grammar*. Amsterdam: Benjamins.
- Narrog, H. 2012. *Modality, Subjectivity, and Semantic Change. A cross-linguistic perspective*. Oxford: Oxford University Press.
- Sommerer L. & E. Smirnova (eds.), 2020. *Nodes and Networks in Diachronic Construction Grammar*. Amsterdam: Benjamins.
- Traugott, E. & R. Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.



### Explaining the speed of lexical change in historical Dutch

Why are different words replaced by new synonyms at different rates? In some domains, new variants replace each other rapidly (e.g. ‘awesome’ or ‘lit’ to refer to something COOL), while other domains display more stability (e.g. ear). Research has shown that the speed of lexical change is influenced by word-related features, like frequency, word class, length or age of acquisition (Bochkarev, Solovyev & Wichmann 2014; Monaghan 2014; Pagel, Atkinson & Meade 2007; Wichmann & Holman 2013). In this paper, we analyze whether characteristics of concepts play a role as well. Taking our lead from Franco et al. (2019) who showed that concept characteristics such as familiarity, vagueness and affect-sensitivity influence the amount of synchronic lexical variation in the base dialects of Dutch, we test whether these characteristics affect the speed of diachronic change in Dutch as well.

The data we use come from the ‘Middelnerlands Woordenboek’ (Middle Dutch Dictionary: 1250-1550) and the ‘Woordenboek der Nederlandsche Taal’ (Dictionary of the Dutch Language: 1500-1976), two large dictionaries of historical Dutch. We extract data from the digitized versions of these dictionaries with the DiaMaNT tool (Depuydt & de Does 2018), a semantic historical computational lexicon for Dutch, zooming in on 279 concepts from two semantic fields: body parts and clothing terms. In particular, for each body part or clothing concept we record all the variants that are available as (historical) synonyms to express the concept, as well as the times at which they were used by relying on the citations available in the dictionary. For example, for the body part JAW, we record that it occurs with 5 variants between 1500 and 1550, including ‘kaak’ (the current Standard Dutch lexeme), ‘kinnebak’, ‘pellorijn’ and ‘kieuw’.

Next, we divide the dataset into 50-year periods. For each period, we calculate two types of information: (1) the number of variants in use at each time point, and (2) the proportion of variants in use at a given period that were also used during the previous period. Using this information, we can answer two research questions: (1) is the number of synonyms for the concept diachronically stable, or are there fluctuations (diachronic stability)?; (2) how quickly do variants disappear from the data, how quickly are they replaced with new synonyms (the rate of lexical replacement)?

Our hypotheses are that the factors that play a role in synchronic data, affect diachronic change as well, viz. familiarity, vagueness and affect-sensitivity. Moreover, we may find differences between the body part concepts and the clothing concepts as the former concepts have a higher degree of universality and may therefore only rarely be referred to with novel lexical items.

Preliminary results on the body part concepts indicate that there are some trends in the data that confirm the correlation between familiarity and diachronic stability on the one hand, and affect-sensitivity and diachronic stability on the other. For vagueness, the picture is less clear. Further data collection and analyses will take place in the coming months.

#### References

- Bochkarev, V., V. Solovyev & S. Wichmann. 2014. Universals versus historical contingencies in lexical evolution. *Journal of The Royal Society Interface* 11(101). 20140841. <https://doi.org/10.1098/rsif.2014.0841>.
- Depuydt, Katrien & Jesse de Does. 2018. The Diachronic Semantic Lexicon of Dutch as Linked Open Data. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Paris, France*.
- Franco, Karlien, Dirk Geeraerts, Dirk Speelman & Roeland Van Hout. 2019. Concept characteristics and variation in lexical diversity in two Dutch dialect areas. *Cognitive Linguistics*. De Gruyter Mouton 30(1). 205–242. <https://doi.org/10.1515/cog-2017-0136>.

- Monaghan, Padraic. 2014. Age of acquisition predicts rate of lexical evolution. *Cognition* 133(3). 530–534. <https://doi.org/10.1016/j.cognition.2014.08.007>.
- Pagel, Mark, Quentin D. Atkinson & Andrew Meade. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 449(7163). 717–720. <https://doi.org/10.1038/nature06176>.
- Wichmann, Søren & Eric W. Holman. 2013. Languages with longer words have more lexical change. In *Languages with longer words have more lexical change*, 249–282. De Gruyter Mouton. <https://doi.org/10.1515/9783110305258.249>.

## The diachronic study of Bangla case marking system

Syed Shahrier Rahman PhD

Professor, Department of Linguistics, University of Dhaka

[ss.rahman@du.ac.bd](mailto:ss.rahman@du.ac.bd)

Mithun Banerjee

Associate Professor, Department of Linguistics, University of Dhaka

[mithun.b@du.ac.bd](mailto:mithun.b@du.ac.bd)

### Abstract

Bangla case marking is one of the least focused areas of grammatical exploration. Since this language allows an extensive set of word order variations, case marking elements of Bangla grammar are used to license their distributions in the syntactic domain. However, a few case markers have different functionality than others create complexity and ambiguity when attempting to identify case features. It raises an important query: Is there any historical data that we have access to that might explain why the Bangla case markers correlate one-to-many correspondences? From a diachronic point of view, this article examines the missing linkages among the case marking components of the Bangla language and attempts to characterise their historical evolution.

The eight-case system of Old Indo-Aryan (OIA) has been mostly lost throughout the centuries by the Indo-Aryan languages (Butt & Ahmed, 2011). Direct (the result of the merger of nominative, accusative and dative) and Oblique (mostly deriving from the Old Indo-Aryan genitive) are the only two cases that have survived however, in some languages, including Bangla, isolated traces of some other oblique cases, such as instrumental, locative, or ablative, can still be found, sometimes even within the declension paradigm (Beekes, 1995). Those bound or free morphemes of varied provenance, including many that exist in this usage already in OIA, take over for the lost cases and replenish the inventory of case markers. It's worth noting that, towards the conclusion of the Middle Indo-Aryan (MIA) era, some (oblique) case forms may become indistinguishable from bare stems because of the degradation of the nominal inflexion. Old Bangla possessed an ergative construction in the perfect aspect (Chatterji, 1970, pp. 947-8), similar to the MIA ergative clause. In contrast, modern Bangla, however, has lost this pattern by allowing the same subject case-marking for its non-perfect and perfect counterparts. New postpositions, often added to the oblique case form, have seen a dramatic growth in usage and grammaticalisation in Bangla.

- a)       ami    toma=ke   golpo-ti                    bolte   chai  
 I.NOM you=ACC story-the.DET.DAT tell.INF want.PRS  
 "I want to tell you the story."
- b)       ama=r    toma=ke   golpo-ti                    bolte   hobe  
 I.GEN you=ACC story-the.DET.DAT tell.INF have.FUT  
 "I will have to tell you the story."
- c)       ami    toma=r    shonge    golpo-ti                    niye        alap        korlam  
 I.NOM you.GEN with.PostP story-the.DET.DAT about.PostP discuss.VN do.PST  
 "I discussed the story with you."

Because of this grammaticalisation, a new case can emerge from the combination of a postposition and the nominal stem or oblique case (Blake, 2001). A bulk of markers of genitive and dative containing *k-* and/or *r-*, can be found in Bangla. A detailed chronological analysis of the major turning points in the corresponding grammaticalisation scenarios, therefore, is necessary to understand the facts behind the case marking system of Bangla language.

**Keywords:** Case, Bangla language, Case marking elements, Grammaticalisation

### References

- Beekes, R. S. P. (1995). *Comparative Indo-European Linguistics. An Introduction*, Benjamins.  
 Blake, B. J. (2001). *Case*. Cambridge University Press.

Butt, M. & Ahment, T. (2011). The redevelopment of Indo-Aryan case systems from a lexical semantic perspective, *Morphology* ; 21, 3-4, 545-572 <https://dx.doi.org/10.1007/s11525-010-9175-0>

Chatterji, S. K. (1970[originally published in 1926]). *The Origin and Development of the Bengali Language*, George Allen and Unwin Ltd.

===

German V2-Argument Clauses from a Diachronic Perspective

This study investigates argument realizing verb-second clauses (arg.V2) in the Early New High German period (ENHG) on the basis of a corpus of narrative texts from the 15<sup>th</sup> and 16<sup>th</sup> c. (Pontus und Sidonia (1450), Melusine (1456), Wigalois (1472), Wilhelm von Österreich (1481), Tristrant und Isalde (1484), Huce Scheppel (1500), Fortunatus (1509), Schöne Magelone (1527), and Goldener Esel (1538)). I will argue that arg.V2 are licensed by two different pragmatical factors independently: *at-issueness* (Simons et al. 2010) and *mediated assertivity* (Reis 1997).

In ENHG (as well as in Present Day German (PDG)), argument clauses of verbs that denote an act of assertion can either be realized as asyndetic V2 clauses (a.) or as syndetic verb-end clauses (VE) (b.), b. being the canonical structure of subordinate clauses.

- a. sy sagten all [sy **wiβten** es nit] (Fortunatus)  
 they said all they KNEW it NEG  
 ‘They all said they didn’t know it’
- b. der schray (...) [das Fortunatus nit umb die ding **wiβt**] (Fortunatus)  
 he screamed (...) that Fortunatus NEG about the things KNEW  
 ‘He screamed (...) that Fortunatus didn’t know about these things’

The V2-structure is assumed to be a main clause phenomenon in German and cross-Germanic (Holmberg 2015), hence arg.V2 are often argued to come close to main clauses (Reis 1997, Gärtner 2002, Truckenbrodt 2006). Although the PDG counterparts of arg.V2 have been thoroughly investigated, their licensing conditions are still a matter of debate (Jacobs 2020, Djärv 2022). There are at least two theoretical approaches to the pragmatical licensing conditions of arg.V2 in German. Firstly, arg.V2 are claimed to be *mediated assertions* (Reis 1997, Gärtner 2002), as the truth value of the proposition is usually asserted by a matrix subject. Secondly, arg.V2 are assumed to mark *at-issue*-content (in the sense of Simons et al. 2010), that is, assertions that are relevant for the *Question under Discussion* (in the sense of Klein & von Steutterheim 1992) of a discourse (Antomo 2015).

The diachronic perspective on the licensing conditions of arg.V2 is often unconsidered. Since the formal distinction between dependent and independent clauses had already developed in the Old High German period (Axel 2007), former studies have primarily focused on arg.V2 in the Old and Middle High German period. Petrova (2020) has shown that there are noticeable parallels in the typology of matrix verbs of arg.V2 between these periods and PDG. However, diachronic frequency changes indicate that there is a stronger association between the discourse pragmatical status of a clause and the V2-structure in ENHG: In our corpus of ENHG narratives, argument clauses of verbs of saying have V2 in 48,8 % of cases, whereas narratives from the 18<sup>th</sup> and 19<sup>th</sup> c. (Deutsches Textarchiv) show arg.V2 in 34,78 % of cases.

In fact, the corpus data provide evidence that the different structures of argument clauses followed rather strict discourse-pragmatical principles after verbs of saying in ENHG – this is especially prominent in the *Fortunatus*, which is the only non-translated text in the corpus. I will show that VE is very much restricted to propositions that have a discourse antecedent and that are therefore presupposed (which is the case in b.). The function of V2 is twofold: On the one hand, V2 is accessible for *mediated assertions*, whereby it is not relevant if the narrator is committed towards the truth of the proposition or not (for example, in a., the proposition of the argument clause can immediately be identified as a lie). On the other hand, V2 marks argument clauses with *at-issue*-content that independently adds to the discourse. The proposition may even be presupposed in such contexts. These two factors can, but do not necessarily have to interact. The pragmatical two-sidedness suggests that the V2-structure can hardly be ascribed to one particular pragmatic function and the existing theories do not necessarily oppose one another.

## References

- Antomo, M. 2015. Abhängige Sätze in einem fragebasierten Diskursmodell. Göttingen: Georg-August-Universität Göttingen dissertation.
- Axel, K. 2007. Studies on Old High German Syntax - Left sentence periphery, verb placement and verb-second. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Djärv, K. 2022. On the interpretation and distribution of embedded main clause syntax: new perspectives on complex discourse moves. In *Glossa: a journal of general linguistics* 7(1). doi: <https://doi.org/10.16995/glossa.5752>
- Freywald, U. 2013. Uneingeleiteter V1-und V2-Satz. In J. Meibauer, M. Steinbach & H. Altmann (eds.), *Satztypen des Deutschen*, 317-337. Berlin: De Gruyter.
- Gärtner, H.-M. 2002. On the force of V2 declaratives. In *Theoretical Linguistics* 28, 33-42.
- Holmberg, A. 2015. Verb Second. In T. Kiss & A. Alexiadou (eds.), *Syntax – Theory and Analysis: An International Handbook of Contemporary Syntactic Research*, 342-383. Berlin, New York: Mouton de Gruyter.
- Jacobs, J. 2020. What kind of main clause phenomenon is V2 in German? In H. Lohnstein and A. Tsiknakis (eds.), *Verb Second*, 169-206. Boston/Berlin: De Gruyter.
- Klein, W. & C. von Stutterheim. 1992. Textstruktur und referentielle Bewegung. In *Zeitschrift für Literaturwissenschaft und Linguistik* 86, 67-92.
- Petrova, S. 2020. Variation and change in the licensing of dependent V2 in German. In H. Lohnstein & A. Tsiknakis (eds.), *Verb Second*, 251–276. Boston/Berlin: De Gruyter.
- Reis, M. 1997. Zum syntaktischen Status unselbstständiger Verbzweit-Sätze. In C. Dürscheid, K.-H. Ramers und M. Schwarz (eds.), *Sprache im Fokus. Festschrift für Heinz Vater zum 65. Geburtstag*, 121-144. Tübingen: Niemeyer.
- Simons, M., D. Beaver, J. Tonhauser & C. Roberts. 2010. What projects and why. *Proceedings of SALT 20*, 309–327.
- Truckenbrodt, H. 2006. On the semantic motivation of syntactic verb movement to C in German. In *Theoretical Linguistics* 32, 257–306.

# Semantic factors influencing the change in position of German adnominal genitives in the 17<sup>th</sup> to 19<sup>th</sup> centuries

Lisa Dücker (CAU Kiel)

In contemporary German, adnominal genitives usually follow the noun they modify: *das Haus der Frau* (the house the-GEN woman-GEN ‘the woman’s house’). This is especially true for common nouns; proper nouns have at the same time retained the ability of appearing before the head noun: *Marias Haus* (Maria-GEN house ‘Maria’s house’) but may also appear postnominally. This current word order in noun phrases with adnominal genitives is the result of a diachronic change that has been going on for several hundred years. In Old High German, genitive attributes usually appeared in the prenominal position but from the later Old High German period onwards, more and more adnominal genitives have changed to the postnominal position, starting with abstract common nouns. Since the 18<sup>th</sup> century, it has been mostly proper nouns that still appear prenominally (c.f. Demske 2001).

This syntactic change is well-studied and a number of morphological, syntactic, semantic and pragmatic factors have been found to influence this phenomenon (for an overview, see e.g. Ackermann 2021). With the exception of Ackermann’s (2021) study on onymic genitives, those factors have usually been studied in isolation, although it has, of course, been noted that properties such as the animacy of the attribute and the semantic relation between the attribute and the noun it modifies (e.g. possessive, partitive) interact with one another. With regards to the semantic relation between the nouns, it has additionally been proposed that the two word orders can correlate with distinct readings, especially with regards to subjective and objective genitives, meaning that *Lenas Entdeckung* (Lena-GEN discovery) will likely be read as a discovery made by Lena, while *die Entdeckung Lenas* (the discovery Lena-GEN) suggests the reading that Lena is being discovered (c.f. Eisenberg & Smith 2002).

In my talk I will analyse noun phrases containing adnominal genitives from three corpora: the Deutsches Textarchiv corpus, which contains written texts from the 17<sup>th</sup> to 19<sup>th</sup> centuries, the RIDGES corpus consisting of texts about herbology from the 15<sup>th</sup> to 19<sup>th</sup> centuries, and the SiGS corpus, which is made up of handwritten protocols of witch trials from the 16<sup>th</sup> and 17<sup>th</sup> centuries. I investigate three semantic properties of noun phrases containing a genitive attribute, namely the animacy of the attribute, the semantic relation between attribute and the head noun, and whether the adnominal genitive is a proper noun or not. These three factors will be analysed in isolation and in interaction as independent variables in a *random forest* model (Tagliamonte & Baayen 2012).

This analysis will show which adnominal genitives are the last to change to the postnominal position at a time when most genitives have already changed to the postnominal position. It will also show the importance of differences in text genre (c.f. Peter 2015). I propose that proper nouns and highly animate nouns (i.e. nouns denoting humans, superhumans, animals and plants), as well as nouns found to be in a subjective, objective, possessive or auctorial relation with the noun they modify will be found to occur in the prenominal position much later than common nouns in general. It is especially inanimate attributes and those that occur in a different kind of semantic relation with the head of the phrase that will appear mostly postnominally early on.

## References

- Ackermann, Tanja (2021): Pre- and postnominal onymic genitives in (Early) New High German. *Journal of Historical Linguistics* 11 (3), 499–533.
- Demske, Ulrike (2001): *Merkmale und Relationen. Diachrone Studien zur Nominalphrase des Deutschen*. Berlin: De Gruyter.



Eisenberg, Peter, and George Smith (2002): Der einfache Genitiv. Eigennamen als Attribute. In Corinna Peschel (ed.): *Grammatik und Grammatikvermittlung*. Frankfurt a.M.: Peter Lang, 113-126.

Peter, Klaus (2015): Zur Semantik pränominaler onymischer Genitivattribute. *Zeitschrift für germanistische Linguistik* 43 (2), 199–232.

Tagliamonte, Sali A. & R. H. Baayen (2012): Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24 (2), 135–178.

*Existential HAVE in Late Latin: insights on its diachrony in the passage to Romance*

This paper investigates the rise of existential HAVE in the transition from Latin to Romance and its paths of development. Existential HAVE constructions in Romance are usually viewed in the literature as stemming from the possessive scheme associated with the transitive verb of possession HAVE (Gaeta 2013; Ciconte 2015: 231; Cruschina 2015: 58, among others).

Early and Late Latin data, however, provide evidence for a different origin of these constructions. More specifically, it is shown that existential HAVE in Romance continues and further develops patterns which became available in Late Latin to ‘introduce a new entity or situation into the world of discourse’ within a spatio-temporal frame, starting from:

- (i) *the stative-locative meaning of the verb HAVE* (‘being in a (physical/abstract) state/place, location’) (... *ille geminus, qui Syracusis habet* ... (Plt. *Men.* 68-69) that.NOM twin.NOM who Syracuse.loc has ‘...that (other) twin, who is in Syracuse ...’),
- (ii) *its non-lexical uses*, witnessed by existential-like ‘impersonal’ constructions consisting of an adverb+the 3<sup>rd</sup> singular of the verb alternating with the verb BE, *esse* (e.g., *bene habet/est, recte habet/est* ‘it is good’, attested in Early (e.g., Plautus) and Classical authors (e.g., Cicero),
- (iii) *its functional equivalence with the copula esse* ‘be’, occurring in Early and Classical Latin (Baldi & Nuti 2010: 273, note 34; 278, 376; Pinkster 2015: 97; Ciconte 2015), attested also in equative clauses in Late Latin (*ubi omnia aequalia habent* (Orib. Syn. VII, 49, 10) where all.N.PL alike.N.PL have.PRS.IND.3PL ‘Where all these are alike’ (Luque Moreno 1998: 140)

Rare examples of existential-like HAVE are reported for Early Latin (Cato, III-II BC) (2), with the verb in the active impersonal form and the nominal in preverbal position in the accusative case (Baldi & Nuti 2010: 275): ... *nisi calicem pertusum cauum habeat* unless cup hole.ACC hollow have.SBJV.PRS.3S ‘... except that **there is a bowl with a pierced hole**’ (Cato, agr. 80,1).

Existential(-like) constructions with HAVE are well attested in Late Latin, in 4<sup>th</sup>-6<sup>th</sup> c. texts (e.g., *Itinerarium Egeriae, Mulomedichina Chironis, Palladius, Oribasius, Anthimus*), occurring with [–human], most typically inanimate, indefinite/non-specific pivots (exs. (1) – (2)) and are found also in spatial (*inde ad sanctam Teclam habebat de civitate forsitan mille quingentos passus* from-there to saint Tecla have.IMP.3SG from city one thousand five-hundred.M.PL.ACC steps.M.PL.ACC ‘From that place to the mountain of God it was perhaps four miles’ (*Itin. Eger.* 23,2) and temporal constructions (*Pater eius ... ex quo hinc profectus est habet annos XIII* father.NOM his.GEN from which.ABLfrom-here leave.PST.PTCP.M.SG.NOM be.PRS.IND.3SG **have.PRS.IND.3SG years.ACC** 14 ‘It has been fourteen years since his father left (from) here’ (Hist. Apoll. RA 31) (Svennung 1935: 475-477, 572-573; Leumann, Hofmann & Szantyr 1965: §221, c; Cennamo 2011: 177-179; Pinkster 2015: 97; Panayotakis 2016 and further examples and further references therein).

HAVE is in the third person singular ‘impersonal’ active (1) and passive (2) forms, with variability of the construction involving three syntactic domains:

- (i) **word order: post and pre-copular NP/pivot** (1a) *In Hebraeo ... non habet hunc numerum* (Hier. Ezech. 11. 297B) in Hebrew not have.PRS.IND.3SG this number.ACC ‘In Hebrew this number does not exist’ vs (1b)... *ibi ... altarium ... habet* (*Itin. Eger.* 4,4) there ... altar.ACC have.PRS.IND.3SG ‘... there ... there is an altar’; (2a) *Ibi habetur capella* (Pard. 369, y. 673) there have.MP. PRS.IND.3SG goat.NOM ‘There is a goat there’ vs (2b) ... *non alter habetur* (Comm. Ap. 374) not other.NOM have.MP.PRS.IND.3SG ‘there is nobody else’
- (ii) **±agr of the copula habere with the post-/pre-copular NP/pivot**: [+AGR]: (2a) *habetur capella*, (2b) *alter habetur*; [-AGR]: (2c): *habetur ... reliquias* (Diehl, ILCV 2013) have.MP.PRS.IND.3SG remnants ‘... there are remnants of ...’; (1c) ... *unde ergo habet zizania* (Vulg. *Matth.* 13, 26-7; Ciconte 2015: 231) whence therefore have.PRS.IND.3SG darnel.weeds.ACC ‘From where, therefore, are the darnel weeds?’

- (iii) **case-marking of the pre/post-copular NP/pivot: NOM vs ACC:** (2a) *habetur capella* vs (2d) *habetur ... tumulum* (Greg. Tur. glor. conf. 35) have. PRS.IND.3SG.MP tomb.ACC (Mikulová 2016: 158)

It will be shown that the data investigated point to the presence of ‘impersonal’ HAVE and lack of pivot agreement (in number /case) as the overt markers of a change in progress, leading to the subsequent reanalysis of the locative argument as a non-referential, unspecified argument, the abstract spatio-temporal argument of predication (as witnessed in early Italo-Romance), and ultimately to the new Romance existential construction (step III on the scheme in (3)), where the theme argument (y), the pivot, ‘takes the predicating function and is thus predicated of the unspecified (spatio-temporal) argument’ (x) (Bentley 2015: 152 and references therein)

Considering the diachronic steps in the changes in the logical structure of HAVE from verb of location/locative copula to an existential copula, illustrated in (3) (adapted from Bentley 2015: 151-152, Ciconte 2015: 231)

(3) I. Be/Have-Loc’ (location, theme) > II. Be/Have-Loc’ (x, theme) > III Be’/Have’ (x,y/pivot)

the Late Latin data investigated appear to witness an intermediate stage between steps I and II. *Habere* still occurs in its locative and copular functions. It does not consistently occur in the default third person singular, preceded/followed by a non-agreeing nominal, the theme, that is still an argument in the construction, and the locative phrase/adverb in the pattern is fully referential (unlike at stage II, where it has lost its locative meaning, as witnessed in some early Northern Italian texts) (Parry 2013; Bentley 2015: 152; Ciconte 2015).

#### References

- Baldi, Ph. & Andrea N. 2010. Possession. In Ph. Baldi & A. Nuti (eds), *New Perspectives on Historical Latin Syntax*, vol. 3. Berlin: Mouton De Gruyter, 239-387.
- Bentley, D. 2015. Predication and argument realization. In D. Bentley, F. M. Ciconte, S. Cruschina. 2015, 99–160.
- Cennamo, M. 2011. Impersonal constructions and accusative subjects in Late Latin. In A. Malchukov & A. Siewierska (eds), *Impersonal Constructions: a Crosslinguistic Perspective*. Amsterdam: J. Benjamins, 169-188.
- Ciconte, F. M. 2015. Historical context. In D. Bentley, F. M. Ciconte, S. Cruschina. 2015, 217–267.
- Cruschina, Silvio. 2015. Patterns of variation in existential constructions. *Isogloss 1/1*: 33–65.
- Gaeta, L. 2013. Existential constructions. In E. van Gelderen, M. Cennamo & J. Barðdal (eds.), *Argument structure in flux: the Naples/Capri Papers*. Amsterdam & Philadelphia: J. Benjamins, 477-509.
- Leumann, M., J. B. Hofmann, A. Szantyr. 1965. *Lateinische Grammatik*. vol. 1. München: Beck.
- Löfstedt, E.. 1933. *Syntactica*, vol 2. Lund: Gleerup.
- Mikulová, J. 2016. *Haberi* + passive perfect participle. *Glotta* 92: 152-174.
- Moreno, J.L. 1978. En torno al sintagma “*Haber* impersonal + sustantivo y sus orígenes latinos. *Revista Española de lingüística* 8.1: 125-147.
- Panayotakis, S. 2016. Expressions of time in early and late Latin: the case of temporal *habet*. In J. N. Adams & N. Vincent (eds), 202: 216.
- Parry, M. 2013. Variation and change in the presentational constructions of north-western Italo-Romance varieties. In E. van Gelderen, M. Cennamo & J. Barðdal (eds.), *Argument structure in flux: the Naples/Capri Papers*. Amsterdam & Philadelphia: J. Benjamins,
- Pinkster, H. 2015. *Oxford Latin Syntax I. The Simple Clause*. Oxford, OUP.
- Raimos Guerreira, A. 1998. Consideraciones sobre la expresión de la posesión externa en latín. In B. García Hernández (ed), *Estudios de Lingüística Latina*, vol 2, Madrid: Ediciones Clásicas, 673-688.
- Reichenkron, G. 1933. *Passivum, Medium und Reflexivum in den Romanischen Sprachen*. Berliner Beiträge zur Romanischen Philologie. III, 1. Jena & Leipzig: Gronau.
- Thielmann, Ph.1885. *Habere* mit dem Part. Perf. Pass. *Archiv für Lateinische Lexicographie* 2: 372-423.
- Svennung, J. 1935. *Untersuchungen zu Palladius und zur Lateinischen Fachund Volkssprache*. Uppsala: Almqvist & Wiksell.

## Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics

Maltese is a prime example of a language that emerged through extensive language contact, joining the two linguistic worlds of Semitic and Italo-Romance languages. Previous studies have demonstrated this on the basis of comparative methods (Comrie, 2009; Comrie & Spagnol, 2016; Lucas & Čéplö, 2020), mostly focusing on the non-concantenative and concantenative morphology of Maltese, with broken plurals such as *kelb-klieb* ‘dogs’ belonging to the first and sound plurals such as *fjura-fjuri* ‘flowers’ belonging to the second language family. The present study aims to extend earlier comparative studies by applying a computational method to the classification of a word’s origin.

To do so, we trained a simple two-layer neural network (NDL, Baayen et al. (2011)) to classify 2-phones from 2347 Tunisian Arabic nouns from Gugliotta and Dinarelli (2020) as Semitic and to classify 2-phones from 2347 Italian nouns from the **subtlex-it** corpus<sup>1</sup> as Non-Semitic. Subsequently, the trained network was tasked with the classification of 6511 Maltese singular and plural nouns from Nieder et al. (accepted) as belonging to the categories **semitic** vs. **non-semitic**.

The network achieved an overall high classification accuracy of 97% in the training data. When the network was required to classify Maltese nouns, which were unknown to the network, as Semitic vs. Non-Semitic, the overall classification accuracy was at 70.76%. When inspecting the classification in more detail, we find that nouns with a Non-Semitic origin show a smaller classification probability as Semitic than those with a Semitic origin (39.0% vs. 84.6%,  $\Delta = 45.6\%$ ,  $z\text{-value} = 34.34$ ,  $p < 0.001$ ). Moreover, we were interested how classification depended on the nouns’ morphological class (sound plural, broken plural, singular sound, singular broken).

While all nouns with a Semitic origin across all morphological classes were classified as Semitic, plurals with a Non-Semitic origin showed a higher classification probability as Semitic when they show a broken pattern. This indicates that Semitic nouns in Maltese are more similar to their Semitic relatives than Non-Semitic nouns to their Non-Semitic relatives in terms of their phonotactic characteristics.

Our results show that it is possible to classify Maltese nouns using a simple two-layer network with a training based on Tunisian and Italian nouns only. The network is sensitive to the phonotactics of individual languages and can use language-specific phonotactic knowledge to classify a language that is the result of extensive language contact.

## References

- Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. [Publisher: American Psychological Association]. *Psychological review*, 118(3), 438–481.
- Comrie, B. (2009). Maltese and the world atlas of language structures [Publisher: John Benjamins]. *Introducing Maltese Linguistics*, 3–12.
- Comrie, B., & Spagnol, M. (2016). Maltese loanword typology. In G. Puech & B. Saade (Eds.), *Shifts and patterns in Maltese* (pp. 315–330). De Gruyter Mouton. <https://doi.org/10.1515/9783110496376-012>
- Gugliotta, E., & Dinarelli, M. (2020). TArC: Incrementally and Semi-Automatically Collecting a Tunisian Arabish Corpus. *LREC*.
- Lucas, C., & Čéplö, S. (2020). Maltese. <https://doi.org/10.5281/zenodo.3744525>
- Nieder, J., Chuang, Y.-Y., van de Vijver, R., & Baayen, R. (accepted). A Discriminative Lexicon Approach to Word Comprehension, Production and Processing: Maltese Plurals. *to appear in Language*.

<sup>1</sup>downloaded from <http://crr.ugent.be/programs-data/subtitle-frequencies>

## From direct quotation to a chain of extended quotations: the history of Hungarian *úgymond* 'so/to/speak'

It is a well-known phenomenon of grammaticalization that the verb or verb construction of a quotative main clause serves as a source for new elements (e.g. Harris–Campbell 1995: 171, see also 168–172). The topic is quite well-researched for many languages (comprehensive analyses: Kuteva et al. 2019: 357–358, 375–388; Spronck–Casartelli 2021).

In present-day Hungarian there are two discourse markers which started from the quotation and arose from the verb *mond* 'say': *mondván* ('giving the reason') and *úgymond* ('so to speak'). Their history goes back approximately 600 years. At the 25th ICHL conference I discussed the element *mondván* (From quotative to causation – the history of Hungarian *mondván*). The present research focuses on the history of the marker *úgymond*. For this expression, the source used to be a collocation: the demonstrative pronoun *úgy* ('so') and the verb *mond* ('he/she says') fused and lexicalized as a quotative marker, which later acquired an extended discourse marker function (Dömötör 2015).

In the 15th century, and at the beginning of the 16th century (Old Hungarian period) *úgy mond* ~ *úgymond* occurred in two kinds of grammatical roles. On the one hand, as a collocation it was a part of a prototypical quotative main sentence (1–2). On the other hand, the fused form was also present as a grammaticalized quotative marker (3).

- (1) ő **úgy** **mond** **vala** ön+benne: „Ha en csak ő  
he so say-PRS3SG AUX.PST self+he.INE if I only he  
ruhá-já-t illet-end-em, megvigasz-om”  
dress-POS-ACC touch-FUT-3SG.DEF comfort-PRS3SG  
'He says to himself: If only I can touch his dress, I would be comforted' (MunC.15rb, 1466)
- (2) az szoror-ok [...] **úgy/** **mond-anak** **vala** öneki: „Mi teneked benne?”  
the sister-PL so say-PRS3PL AUX.PST she-DAT what you-DAT it.INE  
'The sisters told her: What do you care?' (MargL. 84, 1510)
- (3) ez gonosz füge+fá-ról mond-á ur-unk Jézus az igé-k-et:  
this ill fig+tree-DEL say-PST3SG lord-POSS3PL Jesus those word-PL-ACC  
„Íme **úgymond** három esztende-je vagyom, hogy gyümölcs-öt keres-ek [...]”  
lo QUOTDM three year-POSS3SG is that fruit-ACC seek-PRS1SG  
'Our Lord, Jesus told us these words about this ill fig tree: Lo, QUOTDM for tree years I have been looking for fruits [...]' (DignAp.11, 1521)

The quotative role of *úgymond* was attested during the Middle Hungarian period (4) and also in subsequent periods, and it also appears today. Meanwhile, the collocation did not lose its original function for a long time (until the 19th century).

- (4) mond-ja a fatens báty-ja: „No, megad-á **úgymond** nékem Jóczikné  
say-3SG.DEF the witness brother-POS3SG well give-PST3SG.DEF QUOTDM I.DAT Jóczik's wife”  
'The brother of the witness says: Well, Jóczik's wife QUOTDM clobbered me' (TMK, 1716)

Later, from the middle of the 20th century, the function of the quotative marker has extended and shifted towards attitude marking. It is used in present Hungarian primarily in this role, with the meaning 'so to speak'.

On the one hand, the discourse marker *úgymond* signals the speaker's attitude that they use a word or collocation which was taken from an unnamed outside source (possibly from the public discourse), but they distance themselves from that usage (5).

- (5) Minden **úgymond/** rossz anyá-nak megvan a maga történet-e  
each DM bad mother-DAT has-PRS3SG the own story-POSS3SG  
'Each, so to speak, bad mother has got her own story' ([hvg.hu/360/202239](http://hvg.hu/360/202239) 2022.10.11.)

On the other hand, the element can signal that the speaker reflects on their own language use. This usage of *úgymond* is the result of a further extension, as it goes beyond linking to the outside source. Among other things, the speaker can signal they use the following element unusually or inaccurately (6).

- (6) Amikor ezek-et a tartalm-ak-at **úgymond** leszerződte-t-ék, a jogtulajdonos-ok [...]  
when this-ACC the content-PL-ACC DM contract-PST-3PL.DEF the rightsholder-PL  
'When they, so to speak, contracted this contents, the rightsholders [...]' ([napidroid.hu/netflix](http://napidroid.hu/netflix) 2023.01.14.)

In my presentation, I investigate the steps of the functional extension of *úgymond*. I analyse the first, determining step (grammaticalization) and also focus on the relation between the different discourse marker functions.

---

## References

Dömötör Adrienne 2015. A *mondván* és az *úgymond*: diskurzusjelölő elemek keletkezése idéző szerkezetekből. [*Mondván* 'giving the cause' and *úgymond* 'so to speak': the emergence of discourse markers from quoting constructions.] Magyar Nyelv 111: 21–37.

Harris, Alice C. and Campbell, Lyle 1995: Historical syntax in crosslinguistic perspective. Cambridge, Cambridge University Press.

Kuteva, Tanja, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog and Seongha Rhee 2019. World Lexicon of Grammaticalization. Cambridge, Cambridge University Press.

Spronck, Stef and Casartelli, Daniela 2021. In a manner of speaking: how reported speech may have shaped grammar. *Frontiers in Communication* 6:624486. [www.frontiersin.org](http://www.frontiersin.org) 1–22.

## Cited sources

MunC. = Munnich codex (1466)

MargL. = Margaret legend (1510)

DignAp. = Booklet on the dignity of the saint apostles (1521)

TMK = Történeti Magánéleti Korpusz [Old and Middle Hungarian corpus of informal language use]



## Synthetic or analytical: factors which explain the formal variation of future and conditional in Old Catalan

Aina Torres-Latorre

In Romance languages, future and conditional are newly created verbal tenses. They are the result of grammaticalization processes of Latin periphrasis, mainly of the periphrasis CANTĀRE HABEŌ, which has been the successful one in most languages. During the Middle Ages, the grammaticalization process of the two verbal tenses was not yet complete in many Romance varieties, such as Portuguese, Castilian, Aragonese, Catalan, Occitan, or some Italian dialects. Two types of forms could be found: the so-called synthetic forms (*cantaré* ‘I will sing’) and the so-called analytical forms (*cantar-lo he* ‘I will sing it’). Analytical forms differed from synthetic forms because of the need of a weak pronoun between the infinitive and the auxiliary. The presence of a clitic within the verbal tense shows the absence of univerbation, one of the characteristics of grammaticalization (see Lehmann 1985, 2020). Hence, synthetic forms are more grammaticalized than analytical forms.

The aim of this work is to study the variation between synthetic and analytical forms in Old Catalan. The distribution between the two types of forms is not arbitrary. First, it follows syntactic-pragmatic restrictions. In medieval Romance, clitic placement was due to different syntactic environments: ones which entailed preverbal clitics and others in which pronouns were postverbal (for Old Catalan, Batllori *et al.* 2005, Francalanci *et al.* in press). Clitic placement followed the same principles with future and conditional (Sentí & Bouzouita 2022). In this case, preverbal clitics only appeared with synthetic forms (*lo cantaré* ‘I will sing it’), but postverbal pronouns could be used with analytical forms (*cantar-lo he* ‘I will sing it’) or with synthetic forms (*cantaré-lo* ‘I will sing it’). To study the variation between synthetic and analytical forms only those who appear in the same environments can be compared, that is, analytical forms and synthetic forms with postverbal clitics.

Then, in second place, the distribution of synthetic forms with postverbal clitics and analytical forms is not arbitrary either. Some studies based on Old Castilian (Bouzouita 2016a, 2016b) and Old Navarro-Aragonese (Primerano & Bouzouita submitted) suggest some morphological and syntactic factors which could explain a preference for synthetic forms instead of analytical forms, the most common ones in these languages. The proposed factors are: (i) verbal tense (if it is a future or a conditional), (ii) verbal conjugation, (iii) presence of a verbal syncope, (iv) presence of a non-finite verbal form after the future or conditional. This study intends to analyse the behaviour of these factors in Old Catalan, and to suggest differences between Catalan and the other languages studied. The mentioned factors will be reviewed based on data from the 11<sup>th</sup> century to the 16<sup>th</sup> century extracted from the *Corpus Informatizat del Català Antic* (CICA).



## Corpus

CICA = *Corpus Informatitzat del Català Antic*, Torruella, Joan (dir.), with Manuel Pérez Saldanya and Josep Martines: <http://www.cica.cat>.

## Bibliography

BATLLORI, Montserrat / IGLÉSIAS, Narcís / MARTINS, Ana Maria (2005): «Sintaxi dels clítics pronominals en català medieval», *Caplletra*, 38: 137-177.

BOUZOUITA, Miriam (2016a): «La posposición pronominal con futuros y condicionales en el código escurialense I.i.6: un examen de varias hipótesis morfosintácticas», in: Kabatek, Johannes: *Lingüística de corpus y lingüística iberorrománica*, Berlín: De Gruyter.

— (2016b): «La posposición pronominal con futuros y condicionales en el castellano medieval: ¿un caso de contacto de lenguas?», presented in the seminar *El orden de palabras en las lenguas iberorrománicas medievales*, 29<sup>th</sup> of July of 2016, University of Girona.

FRANCALANCI, Leonardo, Ana Maria MARTINS, Francisco ORDÓÑEZ & Josep E. RIBERA (in press): «Els pronoms febles (I): forma, posició i ordre», in Martines, Josep; Pérez Saldanya, Manuel; Rigau, Gemma (ed.): *Gramàtica del català antic*, Amsterdam: John Benjamins.

LEHMANN, Christian (1985): «Grammaticalization: Synchronic variation and diachronic change», *Lingua e stile*, 20: 303-318.

— (2020): «Univerbation», *Folia Linguistica Historica*, 41: 205-252.

PRIMERANO, Antoine & Miriam BOUZOUITA (submitted): «La gramaticalización de los futuros y condicionales en el navarroaragonés de los siglos XIII-XIV: una comparación con el castellano medieval», *Revue Romane*.

SENTÍ, Andreu & Miriam BOUZOUITA (2022): «The syntax of clitics in Old Catalan: Llibre dels Fets», *Catalan Journal of Linguistics*, 21: 1-31.

## Computational Anatolian phylogeny using maximum parsimony

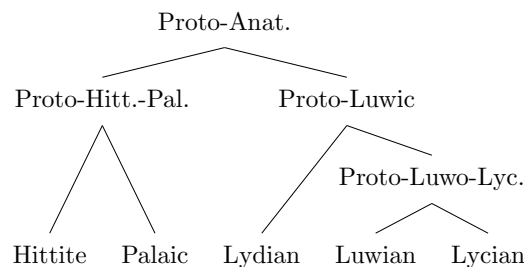
The Anatolian languages constitute an extinct branch of the Indo-European language family, attested across modern day Turkey from ca. the 19<sup>th</sup> cent. BCE to the 2<sup>nd</sup> cent. CE (Zinko 2017). Prominent members include Hittite, Luwian, Lydian, and Lycian.

Previous traditionally oriented work on the internal phylogeny of Anatolian has not reached a consensus. For example, in some studies (Oettinger 1979; Kloekhorst 2022), Palaic, Luwian, and Lycian form a clade, whereas other treatments (Melchert 2003; Rieken 2017) tend to assume a closer relationship between Lydian, Luwian, and Lycian to the exclusion of Palaic. It is consequently warranted to explore alternative methods for determining the topology of the Anatolian tree.

Recent studies applying computational methods to linguistic phylogenies have mostly operated with lexical cognate data as the sole input (e.g. Bouckaert et al. 2012; Chang et al. 2015; Ringe et al. 2002; Ringe et al. use some morphological characters but still principally base their analysis on lexical data). This MO is not viable for Anatolian, as the languages involved are too scarcely attested to allow for the compilation of a reliable cognate data set. Indeed, the material available for Carian, Sidetic, and Pisidian is too scanty for any analysis. Rather, Anatolian phylogeny must operate primarily with phonological and morphological data.

A solid candidate method for conducting computational phylogeny using non-cognate based data sets is *Maximum Parsimony*. This study employs PAUP\* (Swofford 2003) to infer an Anatolian tree on the basis of a data set consisting of 27 characters (12 phonological and 15 morphological) gathered from existing literature and additional research. The taxa involved are Hittite, Palaic, Lydian, Luwian, and Lycian. Considering that the characters used here are predominantly the result of the historical-comparative method, a root state is often possible to assign confidently. Consequently, our tree is rooted. Characters are assigned a weight from 1–4 on the basis of pre-established parameters (e.g. *unconditioned sound changes* are weighted 1 and *sporadic sound changes* are weighted 4). It should be noted that these parameters are unavoidably to some extent subjective, but we do not expect any strong objections from specialists.

Our analysis gives the following most parsimonious tree:



A bootstrap analysis (Felsenstein 1985) indicates that our tree is highly robust. However, considering that the innovations assumed for Proto-Hittite-Palaic under this topology are rather trivial, it may be most prudent to assume a polytomy between Hittite, Palaic, and Proto-Luwic, pending further evidence.

The method used in this study could be exploited for other scarcely attested extinct language families (e.g. Sabellic). An advantage over alternative methods is furnished by the transparency in the grounds on which trees are evaluated (cf. Hammarström et al. 2019: 236). Accordingly, a classically trained historical linguist and/or specialist on the language family at hand is given the opportunity to qualitatively assess the validity of the developments postulated by the analysis.

## References

- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., . . . Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, *337*(6097), 957–960.
- Chang, W., Hall, D., Cathcart, C., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language*, *91*, 194–244.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*(4), 783–791.
- Hammarström, H., Rönchen, P., Elgh, E., & Wiklund, T. (2019). On computational historical linguistics in the 21st century. *Theoretical Linguistics*, *45*(3-4), 233–245.
- Kloekhorst, A. (2022). Anatolian. In T. Olander (Ed.), *The Indo-European Language Family. A Phylogenetic Perspective* (pp. 63–82). Cambridge: Cambridge University Press.
- Melchert, H. C. (2003). The dialectal position of Lycian and Lydian within Anatolian. In M. Giorgieri, M. Salvini, M.-C. Trémouille, & P. Vannicelli (Eds.), *Licia e Lidia prima dell'Ellenizzazione. Atti del Convegno internazionale. Roma, 11-12 ottobre 1999* (pp. 265–272). Roma: Consiglio Nazionale delle Ricerche.
- Oettinger, N. (1979). Die Gliederung des anatolischen Sprachgebietes. *Zeitschrift für vergleichende Sprachforschung*, *92*, 74–92.
- Rieken, E. (2017). The dialectology of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 298–308). Berlin – Boston: De Gruyter Mouton.
- Ringe, D., Warnow, T., & Taylor, A. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society*, *100*(1), 59–129.
- Swofford, D. L. (2003). *PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods)*. <http://paup.phylosolutions.com> /. Sunderland, Massachusetts: Sinauer Associates.
- Zinko, C. (2017). The documentation of Anatolian. In J. Klein, B. Joseph, & M. Fritz (Eds.), *Handbook of Comparative and Historical Indo-European Linguistics. Vol. 1* (pp. 239–249). Berlin – Boston: De Gruyter Mouton.

## The Diachrony of Person-Number Marking of Subjects in Celtic

Mark Darling (University of Oxford), Marieke Meelen (University of Cambridge), David Willis (University of Oxford)

The system of marking person and number of verbal subjects in the Celtic languages has undergone considerable changes in the history of the language family. The earliest examples of the family – the continental Celtic languages of antiquity (Jordán Cólera 2019; Lambert 2003) and the early stages of Irish (Thurneysen 1998), Welsh (Evans 1964), Breton (Lewis and Piette 1990), and Cornish (Lewis 1990) from the medieval period – attest richly inflected verbal systems, with distinct verbal morphology for almost all persons and numbers. Alongside this, they attest null subjects: a non-NP subject of a verb is unexpressed in unmarked contexts.

This contrasts markedly with the situation found in later stages of the Celtic languages. Through the Middle Irish and into the Early Modern Irish period, both the verbal and the nominal morphology of the language becomes increasingly impoverished, and subject pronouns, which were already obligatory with the copula and the defective verb *ol* ‘says’ in later Old Irish, become required in ever more contexts. This development has traditionally been described as occurring first with verbs in the passive voice, with subject pronouns then spreading to be required in most contexts that had previously required a null subject. In modern standard Irish, most persons are expressed by the combination of an uninflected verbal form and a personal pronoun, while there remain inflected forms for some persons in some tenses, which require a null subject, e.g., *léim* ‘I read’ vs. *léann sé* ‘he reads’. Inflected forms are distributed to different persons of the verb in different tenses, and the amount of inflection preserved varies between dialects of the modern language. There are, however, questions to be answered as to the precise details of how the development takes place, and how the various types of pronouns and pronominal particles in the language participate in and are affected by the change from null subjects to overt pronominal subjects.

The developments undergone by Welsh are somewhat more complex than those found in Irish. While very little verbal morphology was lost over the history of the language – modern spoken Welsh retains distinctive personal endings for almost all persons of the verb, with syncretism being limited to the 1<sup>st</sup> and 3<sup>rd</sup> persons plural in the preterite and conditional – there has been considerable change regarding the null subject parameter in the language. In Middle Welsh an alternation can be observed between pronominal subjects in preverbal position, which must always be overt, and those in postverbal position, which may be either overt or null. Over the course of the development from Middle Welsh to modern spoken Welsh, this flexibility has been lost: subjects can only be post-verbal, and must always be overt, e.g., *gweles i* ‘I saw’, *gweloch chi* ‘you (pl.) saw’. Moreover, Welsh exhibits other interesting agreement properties, such as plural NPs triggering singular agreement on verbs, e.g., *gwelodd y plant* ‘the children saw (sg.)’ vs. *gwelon nhw* ‘they saw (pl.)’.

In this paper, we will use two newly developed historical corpora to examine the factors involved in the loss of null subjects in both Irish and Welsh. We will examine the extent to which Information Structure (IS) factors influence this development in both languages, as opposed to the development being purely morphosyntactic. For instance, it seems plausible that overt subject pronouns in Middle Welsh are more frequent in new-subject and/or topic-switch contexts. Do overt subjects spread from these environments to less IS-marked contexts? It is not clear that there is any correlation with verbal morphology. On the other hand, in Irish, it is far less clear that IS-factors play any role, and a correlation with morphological richness seems to be more evident. We will test these hypotheses against data from a selection of medieval and early modern texts from the two languages.

- Evans, D. Simon. 1964. *A Grammar of Middle Welsh*. Dublin: Dublin Institute for Advanced Studies.
- Jordán Cólera, Carlos. 2019. *Lengua y Epigrafía Celtibéricas*. 2 vols. Vol. 1. *Monografías de filología griega*. Vol. 29. Zaragoza: Universidad de Zaragoza.
- Lambert, Pierre-Yves. 2003. *La langue gauloise: description linguistique, commentaire d'inscriptions choisies*. 2 ed. Paris: Éditions Errance.
- Lewis, Henry. 1990. *Handbuch des Mittelkornischen*. Translated by Stefan Zimmer. Edited by Wolfgang Meid. *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck.
- Lewis, Henry, and J. R. F. Piette. 1990. *Handbuch des Mittelbretonischen*. Translated by Wolfgang Meid. Edited by Wolfgang Meid. *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft, Universität Innsbruck.
- Thurneysen, Rudolf. 1998. *A Grammar of Old Irish*. Translated by Daniel A. Binchy and Osborn Bergin. Dublin: Dublin Institute for Advanced Studies. 1946.

The loss of word-initial consonants in Kera'a – A challenge for phonological theory  
 Uta Reinöhl, Kirsten Culhane, Naomi Peck (all: University of Freiburg)

Kera'a (or Idu Mishmi), spoken in the districts Lower Dibang Valley and Dibang Valley of Arunachal Pradesh, shows an intriguing loss of most initial consonants in multi-syllabic words. This loss affects the progressive dialect Midu, while the consonants in question – non-aspirated stops, *h*-, nasals, glides – are retained in the more conservative dialect Mithu. For example, *aphe* 'bat' corresponds to *kaphe*, *ili* 'pig' to *bili*, *iku* 'dog' to *miku*, and *am<sup>w</sup>e* 'wild boar' to *yam<sup>w</sup>e*. While initial consonant loss has been noticed elsewhere in and outside of Trans-Himalayan (e.g., in Kiranti languages (Guillaume Jacques, personal communication), in at least 50 Australian languages (Hale 1964, Blevins 2001), in Sogeram languages of New Guinea (Don Daniels, personal communication)), it remains a challenge for the basic and common assumption in phonological theory that CV is the optimal syllable type given its wide-spread distribution in languages around the world, early acquisition in childhood, and the impression that it strikes an optimal balance between ease of articulation and signal clarity. Furthermore, consonants in word-initial position are generally considered to be 'prominent' cross-linguistically (Beckman 1998, Barnes 2002); for example, word-initial consonants undergo strengthening in a number of languages (Keating et al 2003).

A reduction of the consonant onset in word-initial position is thus unexpected for these and other theoretical and empirical reasons, and also because there is no obvious pathway of phonological change that would lead to it. For Australian languages, it has been argued that a stress shift from the first to the second syllable invited the consonant loss, but this account does not hold up for some of the Australian languages showing the change, including Ogunyjan and Oykangand–Olgol (e.g. Blevins 2001). For Kera'a, Reinöhl (2022) claims that a stress shift also does not play a role, but this claim still awaits a detailed empirical investigation. Since the change is ongoing rather than historical in Kera'a, we are able to draw on synchronic, acoustic data, comparing Midu and Mithu variants of the same words. Having explored the role of stress, we will consider alternative hypotheses for how to motivate this rare and theoretically important change. One of the alternative hypotheses that we will focus on is inter-vocalic lenition in connected speech, resulting from the prior loss of consonantal codas in Kera'a. This talk thus builds on Reinöhl (2022) studying initial consonant loss in Kera'a in greater phonological and phonetic detail, and examines the implications of the Kera'a data for one of the most wide-spread assumptions in phonological and specifically phonotactic theory.

Barnes, Jonathan. 2002. Domain-initial strengthening and the phonetics and phonology of positional neutralization. *Proceedings of the North East Linguistic Society* 32, 1-20.

Beckman, Jill N. 1998. *Positional faithfulness*. Doctoral Dissertation, University of Massachusetts Amherst.

Blevins, J. 2001. Where have all the onsets gone? Initial Consonant Loss in Australian Aboriginal Languages. In Jane Simpson, David Nash, Mary Laughren, Peter Austin, and Barry Alpher (eds.) *Forty years on: Ken Hale and Australian languages*. Canberra, The Australian National University: 481-492.

Hale, Kenneth. 1964. Classification of Northern Paman Languages, Cape York Peninsula, Australia: a research report, *Oceanic Linguistics* 3, 248-265.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. In *Phonetic interpretation: Papers in Laboratory Phonology VI*, ed. John Local, Richard Ogden, and Rosalind Temple, 145–163. Cambridge: Cambridge University Press.

Reinöhl, Uta. 2022. Locating Kera'a (Idu Mishmi) in Its Linguistic Neighbourhood: Evidence from Dialectology. In Mark W Post, Stephen Morey & Toni Huber (eds.), *Ethnolinguistic Prehistory of the Eastern Himalaya*, 232–263. Leiden: Brill.



## Anglo-Scandinavian Contact Influence on Verbs Entering the Causative Alternation

Keywords: Anglo-Scandinavian contact, labilization, causative alternation, corpus linguistics

The English causative alternation and which verbs participate in it in true syntactically labile fashion is a puzzle attempted from many angles and widely discussed (Levin & Rappaport Hovav 1994, 2012; cf. Schäfer 2009, a.o.). Diachronically, several derivational morphemes have contributed to the English set of causative-anticausative verb pairs (van Gelderen 2011, 2018; García García 2012, 2020; Ottósson 2013). One of them is the Germanic causativizer *-j* that is reflected in Old English (OE) as in *meltan* and *miltan* ‘(cause to) melt’ or *byrnan* and *bærnan* ‘(cause to) burn’ (cf. van Gelderen 2018, p. 85-96). However, while pairs like *fall* and *fell* or *sit* and *set* are still distinct in Present Day English (PDE), a number of these verb pairs became formally indistinct when the derivation had become both intransparent and unproductive (cf. Ottósson 2013, García García 2020, van Gelderen 2011). The resulting merged lexemes like *melt* and *burn* have thus become syntactically labile and entered the causative alternation (cf. Levin 1993), some of them in early Middle English already. In contrast to Visser’s (1963) account of 55 labile verbs in OE, McMillion (2006) documents a stark increase to over 800, of varying sources and derivations, until PDE. However, one possible factor in the labilization of English *-j* derived verbs entering the causative alternation however is scarcely discussed: Anglo-Scandinavian contact. Linguistic contact between these somewhat mutually intelligible and very closely related languages has been shown to have led to changes in the lexicon (Durkin 2014; Dance, Pons-Sanz & Schorn 2019). A number of verbs in the alternating verb classes of PDE (Levin 1993) show cognate influence of varying degrees, like *melt*, *burn* and *run*.

This work investigates whether the set of Germanic derived causative verbs in English has been influenced by contact with Old Norse cognate verb pairs. Specifically concerning the labilization of Germanic derived causative verbs in English, this work explores a possible correlation of the level of Old Norse contact influence on these lexical items and their beginning participation in the causative alternation during late Old English and Middle English. To this end a corpus analysis of OE derived causative verbs, their anticausative base verbs as well as their descendant lexemes in Middle English illustrates if and when these lexemes enter into the causative alternation in Middle English. Data are extracted from the *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor, Warner, Pintzuk & Beths 2003) and *The Penn-Helsinki Parsed Corpus of Middle English (2nd edition)* (Kroch & Taylor 2000). Work on Norse cognate verb pairs influences on these lexemes (Dance, Pons-Sanz & Schorn 2019) is considered in relation to the timing and extent of their labilization. If the analysis shows that verbs with significant Old Norse cognate influence show differences in timing or pace of labilization to lexemes without significant Old Norse influence, one might consider contact between Anglo-Scandinavian cognate verb sets as a factor in the labilization of these originally derived causative verbs. Formal and functional overlap between cognate verb pairs not only across their derivational relation, but also cross-linguistically due to the close genealogical relationship of the languages in contact might have led to increased ambiguity in a set of verbs already affected by derivational intransparency due to the loss of *-j*-causativization (cf. García García 2012, 2020; van Gelderen 2011; Ottósson 2013).

## References

- Dance, R., S. Pons-Sanz, and B. Schorn. (2019). *The Gersum Project: The Scandinavian Influence on English Vocabulary* (Cambridge, Cardiff, and Sheffield, 2019)
- Durkin, Philip. 2014. *Borrowed Words: A History of Loanwords in English*. Oxford/New York: Oxford University Press.
- García García, Luisa. 2012. Morphological Causatives in Old English: The Quest for a Vanishing Formation1. *Transactions of the Philological Society*, 110(1), 122–148. <https://doi.org/10.1111/j.1467-968X.2012.01287.x>
- García García, Luisa. 2020. The Basic Valency Orientation of Old English and the Causative Ja-Formation: A Synchronic and Diachronic Approach. *English Language & Linguistics* 24 (1). 153–77. <https://doi.org/10.1017/S1360674318000345>.
- Gelderen, Elly van. 2011. Valency changes in the history of English. *Journal of Historical Linguistics*, 1(1), 106–143. <https://doi.org/10.1075/jhl.1.1.05van>
- Gelderen, Elly van. 2018. *The diachrony of verb meaning: Aspect and argument structure*. New York: Routledge.
- Kroch, Anthony & Ann Taylor. 2000. *The Penn-Helsinki Parsed Corpus of Middle English, second edition (PPCME2)*. Philadelphia: University of Pennsylvania <https://www.ling.upenn.edu/ppche/ppche-release-2010/PPCME2-RELEASE-3/>
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, Beth & Malka Rappaport Hovav. 1994. A Preliminary Analysis of Causative Verbs in English. *Lingua* 92. 35–77.
- Levin, Beth & Malka Rappaport-Hovav. 2012. Lexicon Uniformity and the Causative Alternation. In M. Everaert, M. Marelj, & T. Siloni (Eds.), *The Theta System: Argument Structure at the Interface*. 150–176. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199602513.001.0001>
- McMillion, Alan. 2006. *Labile Verbs in English: Their Meaning, Behavior and Structure*. Dissertation. Stockholm: Department of English, Stockholm University. <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-994>
- Ottósson, Kjartan. 2013. The anticausative and related categories in the Old Germanic languages. In F. Josephson & I. Söhrman (Eds.), *Diachronic and Typological Perspectives on Verbs* (Vol. 134). 329–382. Amsterdam: Benjamins <https://doi.org/10.1075/slcs.134.12ott>
- Schäfer, Florian. 2009. The Causative Alternation. *Language and Linguistics Compass*, 3(2), 641–681. <https://doi.org/10.1111/j.1749-818X.2009.00127.x>
- Taylor, Ann., Warner, A., Pintzuk, S. and Beths, F. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. Electronic texts and manuals available from the Oxford Text Archive.
- Visser, Frederik Theodor. 1963. *An historical syntax of the English language*, vol. 1. Leiden: Brill.

## The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database

One of the less discussed limits of Comparative Method is the database requirements for establishing regularity of correspondences. We may consider a hypothetical example assessing relatedness among three languages – A, B, C. For Correspondence 1, there are 20 cognate sets where Langs A and B have /t/ and Lang C has /s/. For Correspondence 2, there are 2 sets where Lang A has /n/ and Lang B has /ŋ/, 2 different sets where Langs A and C have /n/, and no other [anterior] nasal sets. Correspondence 1, A /t/, B /t/ C /s/, is more securely based than Correspondence 2, A /n/, B /ŋ/, C /n/ and it is not clear that Correspondence 2 is sufficiently supported (Mailhammer 2015; Mailhammer & Harvey 2018).

The basic aim of the Comparative Method is to determine whether inheritance is better supported than chance or contact as explanations for similar forms (Harrison 2003; Weiss 2014). We propose that it is possible to statistically evaluate chance vs inheritance in cases where the database is limited, such as Correspondence 2. There are different mathematical approaches have been applied to assess correspondences in hypothesised distant genetic relationships (Ringe 1993; Oswald 1993).

We propose a further development of Ringe (1992), assessing how likely it is that sound correspondences are accidental using a binominal distribution formula. The key information needed for this method is how frequent on average each relevant phoneme is in each relevant position across all the languages in the sample. Frequencies are multiplied for each phoneme considered and the formula determines whether a match in Y languages across a sample of X languages is accidental. We propose two innovations. The first is assessing the frequency of phonemes across entire lexicons to avoid false positives (Baxter 1993). The second is to compare only identical matches, i.e. identical phonological forms that share a common meaning.

We exemplify this method using a database from Australian languages, which provide a good testing ground, as there are proposals for extensive genetic relationships among Australian languages (Harvey & Mailhammer 2017; Koch 2014), but there are limited numbers of potential cognates and establishing correspondences is problematic (Miceli & Round 2022). We assembled phonologically identical forms expressing 25 lexical concepts across Australian languages, and we examined full lexicons from 35 Australian languages to determine average frequencies of phonemes in all phonotactic positions.

Results show that widespread identities are very unlikely to be due to chance. For example, there are 27 languages where the form /pu/ conveys the meaning 'hit'. It can be ruled out that this match is accidental on a 0.000 level, irrespective of whether average or the highest attested frequencies for /p/ and /u/ are used in the calculation. This even holds if the frequencies of /p/ in initial and /u/ in second position were 50%, which is of course unrealistic across the lexicon of any language. Given that chance is not supported, the remaining hypotheses on shared forms, such as /pu/ 'hit', are contact or inheritance. We do not consider the contact vs inheritance choice here, but note that there are well-established criteria bearing on the choice: e.g. continuity vs discontinuity of attestation (Harvey & Mailhammer 2017); variation in probability of borrowing by part-of-speech class and semantic domain (Tadmor, Haspelmath & Taylor 2010; Tadmor 2009).

## References:

- Baxter, William. 1993. Response to Oswalt and Ringe. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 217–236. Amsterdam & Philadelphia: John Benjamins.
- Harrison, S. P. 2003. On the Limits of the Comparative Method. In Brian D. Joseph & Richard D. Janda (eds.), *The Handbook of Historical Linguistics*. Maldon, MA: Blackwell.
- Harvey, Mark & Robert Mailhammer. 2017. Reconstructing remote relationships: Proto-Australian noun class prefixation. *Diachronica* 34(4). 470–515.
- Koch, Harold. 2014. Historical relations among the Australian languages. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia: a comprehensive guide* (The World of Linguistics 3), 23–89. Berlin & Boston: Walter de Gruyter.
- Mailhammer, Robert. 2015. Etymology. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 423–441. London: Routledge.
- Mailhammer, Robert & Mark Harvey. 2018. A reconstruction of the Proto-Iwaidjan phoneme system. *Australian Journal of Linguistics* 38. 329–359.
- Miceli, Luisa & Erich Round. 2022. Where have all the sound changes gone? Examining the scarcity of evidence for regular sound change in Australian languages. *Linguistics Vanguard* 8(s5). 509–518. <https://doi.org/10.1515/lingvan-2021-0094>.
- Oswalt, Robert. 1993. A probabilistic evaluation of North Eurasian Nostratic. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 199–216. Amsterdam & Philadelphia: John Benjamins.
- Ringe, Don. 1993. Probabilistic evidence for Indo-Uralic. In Joseph C. Salmons & Brian D. Joseph (eds.), *Nostratic: Sifting the evidence*, 153–197. Amsterdam & Philadelphia: John Benjamins.
- Ringe, Donald. 1992. On calculating the factor of chance in language comparison. *Transactions of the American Philosophical Society* 82(1). 1–110.
- Tadmor, Uri. 2009. Loanwords in the World's Languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the World's Languages*, 55–75. Berlin/New York: De Gruyter Mouton.
- Tadmor, Uri, Martin Haspelmath & Bradley Taylor. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2). 226–246.
- Weiss, Michael. 2014. The Comparative Method. In Claire Bowerman & Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, 127–145. London: Routledge.

**‘Old presents’ and the layered history of the Andi verb**

In a seminal article on the typology of verbal inflection, Haspelmath (1998) points out that many anomalous features in the shape and behaviour of imperfective verb forms cross-linguistically can be explained as a side effect of grammaticalization. Once an innovative present tense grammaticalizes into the TAM system of a given language, the pre-existing formation whose central function it takes over – labelled an ‘old present’ – may become restricted to more peripheral roles (whether in terms of semantics or of lexical distribution) that have only their diachronic heritage in common. This observation predicts that at a particular moment in a language’s history, traces of numerous diachronic layers of present formation may be preserved side by side in its synchronic morphology and morphosyntax. The formal complexity of such a system thus provides clues to its development.

In this paper we present a verb system of just this kind in Andi, an understudied minority language of the East Caucasian family, and show that the unusual functional distribution of its morphological material makes sense as the result of a particularly multi-layered history, in which each successive imperfective formation has encroached upon the domain of the one preceding it.

Andi, belonging to the Avar-Andic branch of East Caucasian and spoken by approximately 20,000 people in a handful of villages in mountain Daghestan, is a largely unwritten language attested only since the late 19<sup>th</sup> century: our material is drawn from the two printed works comprising the Andi corpus (Magomedova & Alisultanova 2010 and Luke 2015), complemented by dialect descriptions and the results of recent fieldwork. A striking feature of Andi morphology is the division of its verb system into two formal zones based on distinct, lexically listed inflectional stems – neither of which, however, has an identifiable function in its own right, e.g. they do not straightforwardly encode tense or aspect. Instead, the longer of these two stems serves as the basis for a disparate range of somewhat peripheral verb forms, including certain specialized converbs, the negative (but not positive) imperative, the typologically notable ‘counterexpectation present’ (Maisak & Verhees 2020), the future, and the present habitual – but not the basic present itself, which uses Stem 1.

‘comb’	Stem 1 <i>roxo-</i>	Stem 2 <i>roxud-</i>
	<i>roxo-∅</i> AOR ‘combed’	<i>roxud-ja</i> FUT ‘will comb’
	<i>roxo-rado</i> PRS ‘combs, is combing’	<i>roxud-o</i> HAB ‘(generally) combs’
	<i>roxo-ddu</i> PF ‘has combed’	<i>roxud-abiddu</i> COUNTEREXP.PRS ‘still isn’t combing (!)’
	<i>roxo-r</i> MSD ‘(action of) combing’	<i>roxud-obʔ:ij</i> ANT.CVB ‘before combing’
	<i>rox-o!</i> IMP ‘comb!’	<i>roxud-os:ub!</i> NEG.IMP ‘do not comb!’

Table 1. Some examples of Gagatli Andi finite and non-finite verb forms based on Stems 1 and 2

We show that this complex synchronic situation can be explained by reconstructing a series of developments whereby each ‘new present’ takes over the central functions of the preceding one, while small sets of lexical items may resist the change. Thus the current PRS *roxorado* ousted what is now FUT *roxudja*, which once had a more general non-past value, as attested by the existence of an identical participial form referring to inherent characteristics, e.g. [*hinuk:u*] *dašdja* ‘openable [from inside]’ (Salimov 2010: 222); this value also survives on finite modal verbs, e.g. FUT *ʔidja* ‘may’.

However, *roxudja* was itself an innovation, marginalizing earlier non-past *roxudo*; the latter formation in fact still serves as the basic present for precisely one verb, meaning ‘go’. We give cross-dialectal evidence that *-dja* and the other Stem 2 forms are based on the formation underlying *-do* – some of them via an imperfective participle in *-dob*, which survives only as the suffix deriving ordinals.

Meanwhile, internal reconstruction and Avar-Andic parallels allow us to identify *-do* as an innovation itself: it is the grammaticalized present in *-o* of an iterative in *\*-id-*. And remarkably, a few exceptional verbs retain a present signalled by this *-o* directly. This means we can identify formations from at least *four* diachronic layers coexisting with basic present value for different verbs in Andi.

## References

Haspelmath, Martin (1998). The semantic development of old presents: new futures and subjunctives without grammaticalization. *Diachronica* 15.1: 29-62.

Luke (2015) = *Lukašdi bosam rerhanol'li'i xabar* [The good news according to Luke]. Moscow: Institut perevoda Biblii.

Maisak, Timur, and Samira Verhees (2020). *Counterexpectation present in Andi*. MS., HSE University, Moscow.

Salimov, Khangerej (2010). *Gagatlinskij govor andijskogo jazyka* [The Gagatli dialect of Andi]. Makhachkala: IJaLI.



## Quasi-Suffixaufnahme in Classical Armenian

The phenomenon of *Suffixaufnahme*, or case stacking, refers to a type of morphosyntactic agreement whereby a dependent noun or phrase shows case agreement with its head noun in addition to its regular, functional case marking. Prototypically, the dependent noun occurs in the genitive case, signifying appurtenance, but other cases may also be involved (PLANK 1995).

The phenomenon is most well known from its occurrence in Old Georgian (BOEDER 1995), Hurrian (WEGNER 2007:69–75), and Urartian (SALVINI AND WEGNER 2014:29–31), but also occurs in other languages of the Caucasus and ancient Middle East as well as in some languages of Australia (e.g. Lardil, cf. RICHARDS 2013); the vast majority of the languages concerned exhibit agglutinative morphology. Examples (1) and (2) illustrate this structure in Old Georgian and Hurrian, respectively.

- |     |   |     |   |
|-----|---|-----|---|
| (1) | <i>šəçevn-ita čmid-isa sameb-isa-jta</i><br>help-INS holy-GEN trinity-GEN-INS<br>“with the help of the Holy Trinity” (Sos 1980 no. 2) | (2) | <i>sen(a)=iffu=ue=ne=z asti=i=z</i><br>brother-1SG.POSS-GEN-CON-ERG wife-3SG.POSS-ERG<br>“my brother’s wife” (Mil. III 7) |
|-----|---|-----|---|

Agreement by *Suffixaufnahme* is not obligatory in all languages which possess the pattern, but may be used to disambiguate or in a limited subset of case combinations.

Indo-European languages are not commonly known to exhibit this particular agreement pattern, although limited parallels exist (e.g. in Slavonic, cf. CORBETT 1995). In Classical Armenian, however, a very similar type of agreement does occur: dependents of heads in the accusative marked by the direct object proclitic *z=* may optionally also receive the same proclitic marking regardless of their functional case. This type of agreement is most common with genitive dependents (3), but also extends to other cases (4) and even subordinate clauses (5).

- |     |  |
|-----|--|
| (3) | <i>covac’uc’anēr z=vardapetut’ean=n z=xorut’iwn</i><br>plunge-into.3SG.PST OBJ=teaching.GEN.SG=DET OBJ=depth.ACC.SG<br>“he plunged into the depth of the teaching” (Koriwn §111)   |
| (4) | <i>xoselov z=noc’anē z=amenayn č’arut’iwn</i><br>tell.INF.INS OBJ=3PL.ABL OBJ=all wickedness.ACC.SG<br>“relating all their (lit. from them) wickedness” (Elišē III.234)  |
| (5) | <i>varesc’ē z=tiezerakan išxanut’iwn=d z=or awandeał ē dma</i><br>use.3SG.AOR.SBJV OBJ=universal power.ACC.SG=DET OBJ=REL.NOM.SG give.PTCP be.3SG 3SG.DAT<br><i>y=Astucoy</i><br>from=God.ABL.SG<br>“he will use the universal power which was given to him by God” (Elišē II.130) |

This paper aims to explore two dimensions of this agreement phenomenon in Classical Armenian: firstly, a classification of its usage in 5<sup>th</sup>-century texts by types of cases and nominal phrases involved; and secondly, an attempt at an explanation of the development of this pattern.

Historically, the direct object proclitic is a preposition; other prepositions in Classical Armenian do not, however, show similar agreement by repetition in non-translated texts. Since Armenian nominal morphology rarely distinguishes NOM and ACC, an internal explanation based on the need or wish to disambiguate cannot be excluded. Yet, given the occurrence of *Suffixaufnahme* in the linguistic area, potential interference from Old Georgian or Urartian cannot be excluded *a priori* despite the limited evidence of other contact phenomena (cf. YAKUBOVICH 2010 on morphological interaction between Armenian and Urartian).



## References

- BOEDER, W. (1995) "Suffixaufnahme in Kartvelian," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 151–215.
- CORBETT, G.G. (1995) "Slavonic's Closest Approach to Suffixaufnahme: The Possessive Adjective," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 265–282.
- PLANK, F. (1995) "(Re-)Introducing Suffixaufnahme," in F. PLANK (ed.), *Double case: Agreement by Suffixaufnahme*, Oxford: Oxford University Press, 3–112.
- RICHARDS, N. (2013) "Lardil "Case Stacking" and the Timing of Case Assignment," *Syntax* 16 (1), 42–76.
- SALVINI, M. and WEGNER, I. (2014) *Einführung in die urartäische Sprache*, Wiesbaden: Harrassowitz.
- WEGNER, I. (2007) *Hurritisch. Eine Einführung*, second revised edition, Wiesbaden: Harrassowitz.
- YAKUBOVICH, I. (2010) "Morphological negation in Urartian," *Aramazd. Armenian Journal of Near Eastern Studies* 5 (1), 141–165.

### Can fortis stops spirantise without aspiration?

We have known for over 200 years that fortis stops can spirantise to fortis fricatives in phonological change, but it is not clear that we know *why*. Grimm (1822) showed clearly (and others had noticed earlier), for example, that the fortis stops (in bold) in the Latin words in (1) correspond to the fortis fricatives in Gothic, and that Latin preserves Proto-Indo-European stops while Gothic illustrates a Germanic innovation of fricatives.

- (1) Latin                      Gothic  
*pes, frater, canis*    *fotus, broþar, hunds*

A change of this type (something along the lines of an unconditioned: p, t, k > f, θ, x) has also been recognised in other languages, including: Greek, Proto-Iranian, Proto-Italic, High German and Liverpool English. The latter two cases preserve evidence that a fortis *affricate* stage can (or must) intervene between the fortis stop and fortis fricative stages, which would mean that the change should be understood as: p, t, k > pf, tθ, kx > f, θ, x (ignoring the precise place of articulation of the fricatives). This paper is intended as an exploration of what it might mean to say that we *understand* this type of change.

One crucial facet of ‘understanding a type of change’ is to be certain about the nature of the pre-change phonological state into which it can be innovated – any notion that some aspect of a pre-change state might *cause* a change clearly requires this. A major claim along these lines is that: aspiration is required for fortis stops to spirantise in this way. For example, Salmons (2021, 138) writes that “aspiration is often taken for granted as a, or the, motivation for” changes like this, echoing a long tradition, including Whitney (1884, 92), who wrote that “the spirants (f, th, and so on) are almost universally derived from the full mutes ... and they come especially from such mutes as were originally aspirated”. If this claim can be shown to be true, we could reasonably see it as a firm step in the direction of understanding the fortis-stop-to-fortis-fricative change. The claim has never been rigorously tested, however. I test it in this paper.

In order to work out if this claim is true, we need two things:

- (i) a phonetic and/or phonological rationale to link aspiration and affrication/spirantisation to allow us to argue that the claim is plausible
- (ii) a consideration of all (or, rather, many) cases of changes of the fortis-stop-to-fortis-fricative type, to check if the pre-change fortis stops were aspirated in every case

I first show that there is reason to think that condition (i) can be met. A number of such rationales have been proposed: e.g., Davis & Iverson (1995) consider how fission and spreading of place features can account for affrication, which lays the ground for deaffrication to fricatives; Scheer (1999) argues for the inherent incompatibility of the elements representing aspiration and occlusion in a single segment if the former is incorporated into a unitary segment, leading directly to a fricative; and Honeybone (2002) considers a misperception analysis (whereby postaspiration could be reanalysed as affrication, which, again, could allow for deaffrication).

I focus for the main part on (ii). This calls for an engagement with diachronic phonological typology. Honeybone (2016) argues that diachronic phonological typology is possible but complex, requiring both wide-ranging typological surveys *and* detailed analysis of instances of a change. Consonant with this, I argue that previous relevant typological surveys (Kümmel 2007, Cser 2003, Kirchner 1998) largely fit with the claim that fortis stops must be aspirated in order to be able to spirantise, but not completely. I then show that a detailed analysis of potential counterexamples, informed by an understanding of how laryngeal phonology (‘voicing’ and ‘aspiration’) works (following such work as Iverson & Salmons 1995), promises to remove these counterexamples on a principled basis, allowing us argue with some degree of certainty that fortis stops can only spirantise if they are aspirated.

## References

- Cser, András. 2003. *The typology and modelling of obstruent lenition and fortition processes*. Budapest: Akadémiai Kiadó.
- Davis, Garry & Iverson, Gregory. 1995. Segment organization in the High German Consonant Shift. *American Journal of Germanic Linguistics and Literatures* 7, 111-127.
- Grimm, Jacob. 1822. *Deutsche Grammatik*. 4 volumes. Second edition. Göttingen: Dieterichsche Buchhandlung.
- Honeybone, Patrick. 2002. *Germanic obstruent lenition: some mutual implications of theoretical and historical phonology*. PhD thesis, University of Newcastle upon Tyne.
- Honeybone, Patrick. 2016. Are there impossible changes?  $\theta > f$  but  $f \not> \theta$ . *Papers in Historical Phonology* 1, 316-358.
- Iverson, Gregory & Salmons, Joseph. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369-396.
- Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. PhD dissertation, University of California, Los Angeles.
- Kümmel, Martin Joachim. 2007. *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen*. Wiesbaden: Reichert.
- Salmons, Joseph. 2022. *Sound change*. Edinburgh: Edinburgh University Press.
- Scheer, Tobias. 1998. A theory of consonantal interaction. *Folia Linguistica* 32, 201-237.
- Whitney, William. 1884. *Language and the study of language: twelve lectures on the principles of linguistic science*. London: Trübner.

## Tonogenesis in Baltic and Slavic languages

Tijmen Pronk, Leiden University

Some Slavic and Baltic languages have contrastive tones. Lithuanian, Slovene and Serbian/Croatian/Bosnian/Montenegrin combine contrastive lexical stress with two or more tones. Latvian combines fixed word-initial stress with two or three tones, depending on the dialect. There are three hypotheses about the origin of these tones:

1. The tonal contrast was inherited from Proto-Indo-European
2. The tonal contrast is a shared innovation of Baltic and Slavic languages
3. The tonal contrast arose independently in Baltic and Slavic

There is no consensus among proponents of the last two hypotheses about how the Baltic and Slavic tones would have arisen. Three scenarios have been proposed in the literature:

- a. Tonogenesis through reanalysis of an earlier vowel length contrast
- b. Tonogenesis through reanalysis of an earlier phonation contrast
- c. Tonogenesis through reanalysis of phonetic cues for stress

There are typological parallels for each of these scenarios and all scholars agree that at least two of these mechanisms played some role in the evolution of tone in Baltic and Slavic.

The evaluation of the proposed scenarios is hampered by the fact that the presence of stress has caused changes to the realization of the tones, often depending on whether they occur in stressed or unstressed syllables, as well as the rise of additional tones. This complicates a direct comparison of most of the attested tones across different languages. Instead, it is necessary to peel off layers of innovations in the individual languages first. This is illustrated in the table below, which shows the reflexes of what is generally considered to be a single original tone in 1. medial syllables, 2. initial syllables that are always stressed and 3. initial syllables that sometimes stressed and sometimes unstressed.

	Aukštaitian Lithuanian	Žemaitian Lithuanian	Latvian	Common Slavic	
1.	-i:-	-j:-	-j:-	-í-	denominal suffix
2.	sè:ti	sje:te	sé:t	sǎeti	'to sow'
3.	gì:v-	gì:v-	dzj:v-	zì:v-	'alive'

In this paper it will be argued that scenario 3b explains the attested distributions best: a contrast between modal and laryngealized syllables can be reconstructed for the common ancestor of the Baltic and Slavic and oppositions that were exclusively tonal arose only after Baltic and Slavic had become separate entities.

### Literature

- Dybo, V.A., S.L. Nikolayev & S.A. Starostin. 1978. A tonological hypothesis on the origin of paradigmatic accent systems. *Estonian Papers in Phonetics* 1978, 16-20.
- Jasanoff, J. 2017. *The Prehistory of the Balto-Slavic Accent*. Leiden/Boston.
- Kortlandt, F. 2010. Glottalization and tonogenesis in Athabaskan, Balto-Slavic and Germanic. *Suvremena Lingvistika* 69, 37-44.
- Olander, T. 2009. *Balto-Slavic Accentual Mobility*. Berlin.

- Pronk, T. 2013. On Indo-European Tones, Accentuation, and Ablaut. In: *Indo-European Accent and Ablaut* (G. Keydana, P. Widmer & T. Olander, eds.). Copenhagen, 149-176.
- Ratliff, M. 2015. Tonoexodus, Tonogenesis, and Tone Change. In: *Handbook of Historical Phonology* (P. Honeybone, J. Salmons, eds.). Oxford, 245-261.

## Metaphor, Overtness and Word Order Routinization

This talk argues for a central role of metaphor in constraining and changing clausal syntax in two important, interconnected ways. First, functor-argument metaphors require the overt expression of arguments (Reinöhl 2016, [anonymized reference]). For example, it is possible to say *Everyone was waiting at the hotel. Finally, Kim arrived*. By contrast, people do not use the functor *arrived* metaphorically without a goal argument: *Everything had been pointing to that conclusion all along. \*Finally, Kim arrived*. What they say is *Finally, Kim arrived at it*. Second, overt arguments are required for word order routinization ([anonymized reference]): covert constituents do not need to be ordered. This means that metaphor plays a special role in the conventionalization of word order, because it ensures that in each use of the functor, a decision about the ordering of the functor and its necessarily overt argument(s) must be made.

Our talk builds on cross-linguistic as well as experimental research into metaphor-driven argument overtness, and expands this research with evidence that functor-argument metaphors also feed into word order routinization. This research is supported by Indo-Aryan historical data and framed in accounts of linguistic routinization and automatization (Bybee 2003, Haiman 1994). Hitherto discussion of linguistic routinization has focussed on processes associated with grammaticalization, and we see the effect of metaphor most clearly there.

Reinöhl (2016) shows how the New Indo-Aryan postpositional phrases grammaticalized from various parts of speech, including spatial nouns and adverbs. While these could optionally take arguments in Old and Middle Indo-Aryan, those arguments had become obligatory by New Indo-Aryan, a syntactic shift accompanying semantic reanalyses of functor-argument metaphors. Reinöhl argues that these semantic reanalyses could only occur when the functor and argument were overt, were adjacent, and stood in a particular order. This overtness condition and ordering pattern persists to this day.

Work on grammaticalization has noted the connection between the grammaticalization of lexical items and the rigidification of word order. Lehmann (2002 [1982]) views these as parallel developments. Others (Hopper 1996, Bybee 2003) describe string routinization as a facilitating mechanism for the fixing of word order. However, a crucial missing piece in this story is an explanation for why string routinization should necessarily go hand-in-hand with grammaticalization. We offer, as this missing piece, the requirement that metaphor-creating arguments need to be overt. The role of functor-argument metaphors in establishing bridging cases for grammaticalization is well-known (Svorou 1988, Meillet 1975[1912]). But in precisely these cases, the argument to the functor must be overt. When overt and adjacent, and thus a potential formal whole, functor and argument can then semantically develop greater gestalt properties (Himmelman 1997).

Metaphor has already been shown as fundamental in how we use language. More recently it has been shown to drive argument overtness. Here we argue that it also has a central role in the routinization of word order.

## References:

- Bybee, Joan. 2003. "Mechanisms of Change in Grammaticization: The Role of Frequency", in: B. D. Joseph & J. Janda (eds.) *The Handbook of Historical Linguistics*. Oxford: Blackwell, 602-623.
- Haiman, John. 1994. "Ritualization and the Development of Language", in: William Pagliuca (ed.). *Perspectives on Grammaticalization*. Amsterdam/Philadelphia: Benjamins, 3-28.
- Himmelmann, Nikolaus. 1997. *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Struktur*. Tübingen: Max Niemeyer.
- Hopper, Paul J. 1996. "Some recent trends in Grammaticalization", *Annual Review of Anthropology* 25, 217-236.
- Lehmann, Christian. 2002 [1982]. *Thoughts on Grammaticalization* (2nd, revised edition). (Arbeitspapiere Seminar für Sprachwissenschaft 9). Erfurt: Seminars für Sprachwissenschaft der Universität Erfurt.
- Meillet, Antoine. 1975 [1912]. "L'évolution des formes grammaticales", in: Antoine Meillet. *Linguistique Historique et Linguistique Générale*. Paris: Champion, 130-148.
- Reinöhl, Uta. Mai 2016. *Grammaticalization and the Rise of Configurationality in Indo-Aryan*. Oxford: Oxford University Press.
- Svorou, Soteria. 1988. *The experiential basis of the grammar of space. Evidence from the languages of the world*. Buffalo: SUNY dissertation.



**Complexity in counting systems:  
early systems vs. modern numerical ones**

**Brigitte L.M. Bauer**  
**The University of Texas at Austin**  
**Max Planck Institute for Psycholinguistics**

Numerical numerals are at the cross-roads of linguistics, mathematics, cognitive sciences, archaeology, and anthropology. Yet counting or quantification is not necessarily numerical nor is it necessarily language-bound. Infants, for example, recognize (change in) quantity and speakers of languages with few numerals easily identify large quantities.

A strong trend today is the spreading of a counting system that is numerical, decimal, and based on arithmetical operations, especially addition and multiplication. Numerical counting systems with high upper limits—such as decimal systems—are easily qualified in the literature as “complex”. Yet the criteria to identify “complexity” often remain implicit or may be open for discussion.

While the decimal numerical system continues to spread globally, it reduces numeral variation even if many languages in the past and today feature residues of earlier counting or measure systems. The Latin measures of length, for example, were based on body parts, with La. *pes* ‘foot’ equaling sixteen *digiti* (‘fingers’) or four *palmi* (‘palms’), and the *gradus* ‘step’ equaling 2.5 *pedes*, and five *pedes* equaling one *passus* ‘pace’. Similarly in today’s Indo-European languages, residues of earlier quantification systems relate quantity to commodity, each with their own (base) units for example, cf. Engl. *stone* (weight), *pint* (volume, liquid), or UK’s pre-1971 monetary *pound* system; Fr. *pouce* (length), *muid* (volume, dry/liquid), and so forth.

These residues allow to identify the main features of earlier systems, which ultimately may have their origins in non-numerical systems, such as the one based on tokens in the early stages of the agricultural revolution in the Near East.

In this talk, I discuss the spread of the decimal system identifying its various manifestations and evaluating its main characteristics against early systems of quantification as we know them from residues of the type mentioned above. This comparison will provide data and insights to assess the concept of “complexity” in counting systems.

### The classification of South Cushitic.

In his seminal work on the classification of African languages, Greenberg (1963) has South Cushitic as one of the primary branches of Cushitic. This proposal has been immensely influential in the interpretation of the history of East Africa. On the basis of this classification, the assumption is that the presence of South Cushitic in Tanzania is ancient (Ehret 1980), and pre-dates the entry of Bantu and Nilotic peoples (Ambrose 1998). As a consequence, (South) Cushitic has been linked to the Savannah Pastoral Neolithic cultural complex that is recognised in archaeology, and it has been proposed that the introduction of agriculture and cattle-keeping in Tanzania can be attributed to the South Cushitic speakers. Ehret's (1980) reconstruction of South Cushitic lexicon and phonology has been extremely influential in recognising linguistic contact in many of East Africa's Bantu and Nilotic languages despite the fact that this reconstruction has been severely criticised (Philippson 2013). Serious doubts on the classification of South Cushitic as primary branch have been raised though. Hetzron (1980) has pointed to many grammatical resemblances between South Cushitic and East Cushitic languages that argue for inclusion of South Cushitic within East Cushitic. In Tosco's (2000) Cushitic overview this uncertainty is represented by the fact that South Cushitic figures differently in his genetic trees: as primary branch of Cushitic for the classical view and as primary branch within East Cushitic reflecting Hetzron's suggestions; the issues are discussed in detail in Kießling (2001). Kießling and Mous (2003) provide an extensive lexical and phonological reconstruction of the four South Cushitic Tanzanian languages that are still spoken; while Kießling (2002) is a detailed morphological reconstruction of these languages. Kießling and Mous (2003) pointed out wider Cushitic parallels where they could but this did not enable them to suggest a position of Tanzanian Cushitic in the Cushitic tree. The challenges are the lexical innovations that Tanzanian Cushitic must have undergone and for many of these no other languages could be suggested as sources. In addition, the other languages that were classified as South Cushitic do not offer much for reconstruction: the languages Aasáx and Qwadza are obsolete and the data on these are unreliable because they were collected from rememberers rather than speakers (Kruijsdijk 2023); Ma'á, often characterised as a mixed language, is Bantu, and not Cushitic, and some of the original Cushitic lexicon after language shift survives in a parallel register which also contains words from a variety of other sources including manipulated words from the basis Bantu vocabulary (Mous 2003). The last suggested member, Dahalo, is now considered to be (marginally) East Cushitic rather than South Cushitic (Tosco 1989, Tosco and Blazek 1994).

I propose that Tanzanian Cushitic is a primary branch of Cushitic after all. I also argue that the earliest South Cushitic expansion into Tanzania was followed by two others that have left their (lexical) impact on Tanzanian Cushitic. The latest is the pre-Oromo influence on Tanzanian Cushitic. The suggestion for such language contact showing transfer from pre-Oromo lexical and morphological material featured already in Kießling and Mous (2003), and was recently substantiated by Rapold (2023). There is plenty of reconstructed Tanzanian Cushitic material that is clearly Cushitic but did not undergo the Oromoid innovations. An earlier expansion is formed by speakers of the Dullay-Yaaku subgroup (see Hayward 1978 that this is a subgroup). Recently Sands and Tosco (2022) have argued that early Dullay-Yaaku speakers must have been in contact with Hadza (a language isolate and in the area of Tanzanian Cushitic). I provide further evidence for this intrusion by showing Dullay-Yaaku influence on proto-Tanzanian Cushitic while the Tanzanian Cushitic proto lexicon also contains Cushitic lexical evidence that pre-dates Dullay-Yaaku. The consequences for the interpretation of East Africa's history are far-reaching: There was not one migration of Cushitic speakers into Tanzania but at least three. For all Cushitic lexical transfer into Bantu and Nilotic languages of Tanzania and Kenya, the source needs to be considered.

## References

- Ambrose, Stanley H. 1998. Chronology of the Later Stone Age and Food Production in East Africa. *Journal of Archaeological Science* 25:377–392.
- Appleyard, David 2006. *A comparative dictionary of the Agaw languages*. Cologne: Rüdiger Köppe.
- Ehret, Christopher 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin: Dietrich Reimer.
- Greenberg, Joseph 1963. *The Languages of Africa*. Bloomington: Indiana University
- Hayward, Richard J. 1978. The Qawko Dialects and Yaaku. *Abbay: Documents pour Servir à la histoire de la Civilisation Éthiopienne* 9: 59-70.
- Hetzron, Robert. 1980. The Limits of Cushitic. *Sprache und Geschichte in Afrika* 2:7-126.
- Kießling, Roland. 2001. South Cushitic links to East Cushitic. In *New data and New Methods in Afroasiatic Linguistics*, Andrzej Zaborski (ed.), pp. 95-102. Wiesbaden: Harrasowitz.
- Kießling, Roland. 2002. *Die Rekonstruktion der südkuschitischen Sprachen (West-Rift): von den systemlinguistischen Manifestationen zum gesellschaftlichen Rahmen des Sprachwandels*. (KuS 19). Cologne: Rüdiger Köppe.
- Kießling, Roland and Maarten Mous. 2003. *The lexical reconstruction of West Rift (Southern Cushitic)*. Cologne: Rüdiger Köppe.
- Kruisdijk, Iris 2023. The position of Asá and Qwadza within Cushitic. To appear in *Selected papers from the 53rd Annual Conference on African Linguistics*.
- Mous, Maarten. 2003. *The making of a mixed language: the case of Ma'a/Mbugu*. Amsterdam: John Benjamins.
- Philippon, Gérard 2013. Southern Cushitic loans in North-Eastern Bantu – a reconsideration of the evidence. In *Proceedings of the 5th International Conference on Cushitic and Omotic Languages, Paris, 16-18 avril 2008*, edited by Marie-Claude Simeone-Senelle and Martine Vanhove, pp. 81-93. Cologne: Rüdiger Köppe.
- Rapold, Christian 2023. Pre-Oromo links to Tanzanian Cushitic. Ms.
- Sands, Bonny and Mauro Tosco 2022. Early East African and Peripheral East Cushitic: Foragers and pastoralists in early East Africa. Paper presented on December 14 2022 in the Rift-Valley Network Seminar Series. <https://zenodo.org/record/7440446#.Y8lKq0HMJmM>
- Tosco, Mauro. 1989. The Classification of Dahalo: Another Perspective. Paper, read at the “2nd Symposium on Cushitic and Omotic Languages”, Torino.
- Tosco, Mauro. 2000. Cushitic Overview. *Journal of Ethiopian Studies* 33: 87-121.
- Tosco, Mauro & Václav Blažek, 1994. Between South and East Cushitic: Reconsidering the Position of Dahalo. Paper, read at the 9th Afrikanistentag, 20.9.94, Köln.

## Demonstrative modifiers in Middle Hungarian: a complex picture of renewal

**BACKGROUND:** The definite article grammaticalized in Old Hungarian to systematically encode the definiteness of the noun phrase, and the source of the article was the distal demonstrative *az* 'that'. The demonstrative system, quite atypically (although attested elsewhere, cf. van Gelderen 2011: 207–208), renewed via two different strategies: one involving reinforcement and one involving doubling. In Hungarian, even two pairs of reinforced demonstratives started to be used along with the old demonstrative modifier, namely *ezen/azon* 'this/that' and *imez/amaz* 'this/that', while a determiner doubling construction also emerged in the 16th century. In the latter case, the old demonstrative pronoun, agreeing in number and case, joined to the noun phrase already determined by the definite article.

**EMPIRICAL RESEARCH:** The reinforcement strategy and the determiner doubling strategy did not appear simultaneously, but following one another, which raises a couple of questions. Queries in historical corpora allowed me to specify numerically the change of ratio between the various strategies. Furthermore, corpus study will clearly show how the use of doubling constructions increased through centuries to arrive at its present day dominance in frequency. Data have been drawn from the Old Hungarian corpus, on the one hand, which is mainly composed of religious texts and from two Middle Hungarian corpora, on the other hand. Middle Hungarian sources include both Bible translations and private documents, thus the empirical research could consider diachronic processes, and variation between registers and individual sub-corpora as well.

In order to approach the most intriguing question as to why two different strategies emerged to renew the old system, the distributional properties of the modifiers as well as the semantic and pragmatic aspects of their uses have also been examined.

**ANALYSIS:** In the talk, I propose a syntactic structure for each of the patterns and also aim to model the syntactic change from one construction to the other. Reinforced demonstratives take over the construction type of the old demonstrative strategy, while the definite article emerges as the result of a reanalysis, one that corresponds to the so called Head Preference Principle (van Gelderen 2008 and 2011). At the same time, the determiner doubling strategy can be analyzed as an adjunction, which is supported by word order peculiarities (Author 2014) and by remnant constructions featuring two copies of the demonstrative modifier in apposition, rather than the combination of a demonstrative and an article. However, demonstratives in doubling constructions have been further reanalyzed and integrated into the DP domain, in accordance with the so-called 'Specifier Incorporation', (cf. the universal economy principles in: van Gelderen's 2008)

As a result of the closer examination of distributional phenomena, reinforcement strategies turn out not to have covered all the possible functions (see Himmelmann 1996 and Diessel 1999 for the possible contexts of uses). The reinforced pronouns *ezen/azon* 'this/that' were originally identity markers (meaning 'the same'). Their use only gradually expanded into the general anaphoric use, and never bleached enough to also express an exophoric (extra-linguistic deictic) function. The other pair of reinforced pronouns (*imez/amaz*) remained restricted to special contexts, first of all to the so called recognitional use. That is to say, neither of them encoded exophoric deixis. It was only the determiner doubling construction, emerged somewhat later, but completely neutral with respect to all the pragmatic contexts, that filled the gap. Thus the newest strategy of all could naturally replace the original construction, and at the same time spread into the pragmatic contexts in which the reinforced demonstrative were used.

**REFERENCES**

- Author 2014.** ‘The DP-cycle in Hungarian and the functional extension of the noun phrase.’ In: É. Kiss, K. (ed.): *The Evolution of Functional Left Peripheries in Hungarian Syntax*. Oxford. 56-82.
- Diessel H. 1999.** *Demonstratives. Form, Function and Grammaticalization*. Amsterdam.
- Gelderen, E. van 2008.** Linguistic cycles and Economy Principles. The role of Universal Grammar in language change. In: Eythórsson, Thórhallur (ed.): *Grammatical Change and Linguistic Theory. The Rosendal Papers*. Amsterdam, 245-264.
- Gelderen, E. van 2011.** *The Linguistic Cycle. Language Change and the Language Faculty*. Oxford.
- Himmelmann, N. 1996.** ‘Demonstratives in narrative discourse: a taxonomy of universal uses.’ In: Fox, B. (ed.): *Studies in anaphora*. Amsterdam, 205-254.

## The natural stability of ‘unnatural’ morphology

Models can serve as powerful tools for uncovering how a simple change process may lead to striking emergent outcomes [1,2], and likewise, how small revisions in assumptions can lead to dramatic shifts in how a system is predicted to behave [3]. Here we use simple modelling to demonstrate that analogical changes in inflectional systems can be expected to have a particularly counterintuitive, yet empirically well-supported, long-term effect: namely, inflectional classes’ and stem alternation patterns’ resilience against levelling, even as they undergo constant analogical change [4]. We first underscore why this resilience is so surprising. Next, we explain why analogical reasoning in inflectional systems is expected to trigger changes based not only on similarity but also on dissimilarity. Finally, we implement the predicted change processes in a simple iterated evolutionary model and show that their long-term consequence is to support not only the ongoing coalescence and coherence of morphological classes but also their resistance to complete collapse and levelling.

In inflectional systems, idiosyncratic morphological class systems such as inflectional classes and stem alternation patterns are empirically ubiquitous [5, 6] yet they present a stark theoretical challenge. Relative to their absence, these systems (i) incur a learning cost [7], yet (ii) offer no clear functional benefit [8]. Given that they are constantly undergoing analogical changes [4], standard evolutionary logic predicts that they *ought to disappear*—and indeed, early modelling work has implied this conclusion [9,10]. Yet in real languages and families, stem alternations and inflection class systems routinely persist across millennia, even as other, more functionally motivated inflectional phenomena collapse and fade. So, what makes these systems so resilient?

Recent work in experimental and computational psychology [11] has emphasised that inferential reasoning occurs within an *inductive context*, which is appropriate to the situation at hand, and which shapes the inferences that are more or less likely. For instance, people reason differently when items are related by physical distances, versus via a taxonomic tree. Carefully applying the notion of inductive context to inflectional systems reveals a significant implication: analogy is expected to licence both similarity-enhancing and dissimilarity-enhancing inferences. When these two types of inference are implemented in an iterative model, they give rise to two dynamic forces: one of attraction and one of repulsion. As in many attraction-repulsion systems in nature [12,13], this dynamic leads the system to self-organise into areas of internal coherence (i.e., morphological classes) while maintaining difference across them (i.e., avoidance of total levelling).

In conclusion, inflectional classes and other ‘morphomic’ categories [14, 15] have long been held in suspicion within certain theoretical circles [16], and have even earned the label ‘unnatural’ [17]. On the contrary, however, here we demonstrate that they are natural phenomena *par excellence*: they are natural (emergent) outcomes of natural (rational) inference, and they just so happen to be naturally tenacious survivors of aeons of unceasing analogical change.

## References

1. Boyd, R. & P. J. Richerson . 1988. *Culture and the evolutionary process*. Chicago: University of Chicago Press.
2. Smaldino, P. E. 2017. Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read & A. Nowak (eds), *Computational social psychology*, 311-331. New York: Routledge.
3. Guest, O. & A. E. Martin. 2021. How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789-802.
4. Maiden, M. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
5. Bach, X. 2018. *The origins of inflectional classes*. D.Phil dissertation, University of Oxford.
6. Herce, B. 2020. *A typological approach to the morphome*. Doctoral dissertation, University of Surrey.
7. Clark, E. V. 1998. Morphology in language acquisition. In A. Spencer & A. Zwicky (eds), *The handbook of morphology*, 374-389. Oxford: Blackwell.
8. Anderson, S. R. 2015. Dimensions of morphological complexity. In M. Baerman, D. Brown & G. G. Corbett (eds), *Understanding and measuring morphological complexity*, 11-26. Oxford: Oxford University Press.
9. Ackerman, F. & R. Malouf. 2015. The No Blur Principle effects as an emergent property of language systems. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA.
10. Esher, L. 2015. Approaches to the function and emergence of metamorphomic structure. Presentation given at Décembrettes 9. Toulouse, December 2015.
11. Kemp, C. & J. B. Tenenbaum. 2009. Structured statistical models of inductive reasoning. *Psychological review*, 116(1), 20.
12. Lennard-Jones, J. E. 1931. Cohesion. *Proceedings of the Physical Society*, 43(5), 461–81.
13. D’Orsogna, M. R., Chuang, Y. L., Bertozzi, A. L., & Chayes, L. S. 2006. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical review letters*, 96(10), 104302.
14. Aronoff, M. 1994. *Morphology by itself*. Cambridge: MIT Press.
15. Round, E. R. 2015. Rhizomorphomes, meromorphomes and metamorphomes. In M. Baerman, D. Brown & G. G. Corbett (eds), *Understanding and measuring morphological complexity*, 29-52. Oxford: Oxford University Press.
16. Bermúdez-Otero, R. & A. R. Luís. 2016. A view of the morphome debate. In A. R. Luís & R. Bermúdez-Otero (eds), *The morphome debate*, 309–340. Oxford: Oxford University Press.
17. Aronoff, M. 2016. Unnatural kinds. In A. R. Luís & R. Bermúdez-Otero (eds), *The morphome debate*, 11-32. Oxford: Oxford University Press.



### Diachronic stability of case functions: oblique in Romani dialects

It is well known that case marking is susceptible to change as a result of internal development and language contact (Johanson 2009). In this study I show how different functions of a case form are lost and preserved in Romani, an Indo-Aryan language that has been spoken in Europe since the Middle Ages (Matras 2002).

Structurally, the Romani case system is similar to those found in other new Indo-Aryan languages (NIA) and consists of several layers (Masica 1991). The first layer distinguishes between direct and oblique forms, and the second layer of case markers is attached to the oblique form, cf. Table 1.

Table 1. Case marking in Romani (Kalderash dialect)

Case	<i>manú</i> ‘person’		<i>bakró</i> ‘sheep’	
	SG	PL	SG	PL
DIR	<i>manú</i>	<i>manú</i>	<i>bakró</i>	<i>bakr-é</i>
OBL	<i>manu</i> és	<i>manu</i> én	<i>bakr-és</i>	<i>bakr-én</i>
ABL	<i>manu</i> és-tar	<i>manu</i> én-dar	<i>bakr-és-tar</i>	<i>bakr-én-dar</i>
DAT	<i>manu</i> és-ke	<i>manu</i> én-ge	<i>bakr-és-ke</i>	<i>bakr-én-ge</i>
...				

In most NIA languages, the oblique serves exclusively as a base form for secondary cases and cannot be used independently. In Romani, however, the oblique has several distinguished functions, as it marks i) animate direct objects, ii) possessor, iii) the recipient of the verb ‘give’, and iv) the experiencer of certain verbs (‘feel pain’, ‘like’) (Matras 2002: 85–87). This range of functions is arguably inherited from Middle Indo-Aryan (MIA) stage, as Romani oblique affixes are continuation of the MIA oblique forms (Beníšek 2009).

The goal of the study is to establish how the different functions of the oblique are preserved across Romani dialects. The data come from the Romani Morpho-Syntax database (RMS; <https://romani.humanities.manchester.ac.uk/rms/>) which contains questionnaire-based elicited data on various Romani dialects from 120 locations in Europe. Table 2 summarizes the distribution of case marking among the aforementioned functions of the oblique in the dataset.

Table 2. Case marking in the contexts typical for the oblique in RMS.

Function	OBL	DAT	LOC	Other
Experiencer, ‘like’	12,5%	25%	—	64% = DIR
Experiencer, ‘feel pain’	45%	4%	24%	23% = possessive pronouns
Recipient, ‘give’	50%	50%	28%	—
Possessor	58%	2%	30%	—
Animate direct objects	88%	—	—	not marked

I suggest that the three main factors which determine the stability of the oblique marking are i) type- and token frequency of the functions in speech, ii) the availability of other marking for similar semantic roles, and iii) case marking in contact languages. The lexically determined oblique marking (‘like’, ‘feel’, ‘give’) is less stable than the construction related oblique marking (possessor). This is especially clear with the verb ‘like’ which is often borrowed together with the argument structure of the source language. As usually the semantic roles of ‘recipient’ and ‘experiencer’ are marked in Romani with the dative, there is no wonder that the same marking is found alongside the oblique in these contexts. When the dative marking is additionally supported via language contact, the oblique marking can become obsolete (as it happens in the Romani dialects of Eastern Europe). Finally, the function of the oblique related to the differential object marking is very frequent in speech and cannot be replaced by other cases. That is why the oblique is usually preserved unless the differential object marking is lost altogether (sometimes together with the inflectional case systems as it happens in the Romani dialects of Italy).

**References**

- Beníšek, Michael. 2009. Middle Indo-Aryan Ablative and Locative Markers in Romani. *Indo-Iranian Journal* 52, 335–362.
- Johanson, Lars. 2009. Case and contact linguistics // *The Oxford handbook of case*. Oxford University Press. 494–501.
- Masica, C. 1991. *Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2002. *Romani. A linguistic introduction*. Cambridge: Cambridge University Press.

## The Charition Mime: Decoding the “Indian Language” through Typology and Entropy

The 2nd century CE papyrus P.Oxy III.413 records a Greek play titled “Charition” (Χαρίτιον) written by an anonymous author. The play, a parody of Euripides’ “Iphigenia in Tauris,” features a Greek maiden held captive in India and rescued by her brother by getting the local king and people drunk with wine (Page, 1992; Tsitsiridis, 2005; Webb, 2008; Whitmarsh & Thompson, 2013; Crevatin, 2009). Linguistically, the most exciting aspect of this work is the inclusion of large sections of dialogue in an “Indian” language. This language was initially identified as Dravidian, usually as proto-Kannada, by European and Indian scholars (Hultzsch, 1904; Rice, 1926; Sastri, 1926). However, this interpretation was later dismissed by most Western philologists following Barnett (1926).

There is documented contact between Greek-speaking Egypt and India at the time (Salomon, 1991; Seldeslachts, 1998), and the “barbaric” language does not appear to be gibberish. Hultzsch (1904) and Rice (1926) noted that there would not be much profit in accurately transcribing a gibberish language (including at least one apparent correction of a misspelling), and some phonological and morphological patterns seem compatible with natural languages. They even identified some words as related to modern Kannada, such as κοττως (ultimately a causative related to ಕುಡು [kuḍi], “to drink”) and κονζει (ultimately related to ಕೊಂಚ [koñca], “a little”). Recent proposals have attempted to connect the language to either Kannada or Tulu (Varadpande, 1981; Shivaprasad Rai, 1985; Upadhyaya, 1996), with scholarly work not necessarily impressive (cf. Hall, 2010). No modern study has provided a solid linguistic assessment of whether the language is natural or related to known languages by using expected practices, such as identifying regular sound correspondences.

Our work combines typological examination, linguistic knowledge of Dravidian languages, and quantitative analysis to uncover the nature of the unknown language. The typological assessment considers known features of Dravidian languages (Krishnamurti, 2003) and proposed translations (e.g., Varadpande, 1981) to review the alleged word correspondences in light of the comparative method. The statistical assessment experiments with two approaches, involving a review of the proposed English translations and the production of translations in other languages using automatic methods (Vaswani et al., 2017). In the first approach, the resulting corpus is employed for comparing information measures, such as perplexity (Brown et al. 1992), cross-entropy (Murphy, 2012), and bits-per-character (Shannon, 1951), comparing the unknown language to the translations aggregated by language families. In the second approach, we experiment with classifier models (Pedregosa et al., 2012) operating over skip n-grams (Manning & Schütze, 1999).

With such a multidisciplinary approach, we aim to provide more concrete answers on whether the unknown language in the “Charition” play displays characteristics of natural languages, specifically those of the Dravidian family, and evaluate the reliability of proposed translations. Our methods apply to other undeciphered or partially deciphered ancient texts. At the same time, for this play, they could provide valuable linguistic data for the study of the diachronic development of the Dravidian language family.

## References

- Barnett, L. D. (1926). "The Alleged Kanarese Speeches in P. Oxy. 413." *The Journal of Egyptian Archaeology*, 12(1/2), 13–15. <https://doi.org/10.2307/3854174>
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., Lai, J. C., & Mercer, R. L. (1992). "An estimate of an upper bound for the entropy of English." *Computational Linguistics*, 18(1), 31-40.
- Crevatin, F. (2009). "Nota a P.Oxy. III, 413." *Zeitschrift Für Papyrologie Und Epigraphik*, 171, 199–200.
- Hall, E. (2010). "Iphigenia in Oxyrhynchus and India: Greek tragedy for everyone". In Tsitsirides, S. (ed.), *Parachoregema*. University of Crete Press.
- Hultsch, E. (1904), "Remarks on a papyrus from Oxyrhynchus." *Journal of the Royal Asiatic Society*, 1904: 399-405.
- Krishnamurti, B. (2003). *The Dravidian Languages*. Cambridge Language Surveys. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486876>
- Manning, C. D. & Schütze, H., *Foundations of Statistical Natural Language Processing*, MIT Press: 1999.
- Murphy, K. (2012). *Machine Learning: A Probabilistic Perspective*. MIT.
- Page, D. L. (1992). *Select Papyri*. Loeb classical library. Greek authors. Volume 3. Harvard University Press. ISBN 0674993977, 9780674993976.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., et al. (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research*, 12: 2825-2830.
- Shivaprasad Rai, P. (1985). "Sariti: A 2000 Year Old Bilingual Tulu-Greek Play", *International Journal of Dravidian Linguistics*, 14/2: 320-30,
- Rice, E. P. (1929). "On the proposed identification of Kanarese in Oxyrh. Pap. No 413". In Powell, J. U. and Barber, E. A. (eds.), *New Chapters in the History of Greek Literature*, 2: 215-22.
- Salomon, R. (1991). "Epigraphic Remains of Indian Traders in Egypt". *Journal of the American Oriental Society*, 111(4), 731–736. <https://doi.org/10.2307/603404>
- Sastri, R. Shama (1926). "A Greek farce with Old Kannada passages", *Annual Report of the Mysore Archaeological Department*. Bangalore: Mysore Archaeological Department.

- Seldeslachts, E. (1998). “Translated Loans and Loan Translations as Evidence of Graeco-Indian Bilingualism in Antiquity”. *L’Antiquité Classique*, 67: 273–299.
- Shannon, C. E. (1951). “Prediction and entropy of printed English”. *Bell System Technical Journal*, 30(1): 50-64.
- Tsitsiridis, S. (2005). “Greek Mime in the Roman Empire (P.Oxy.413: Charition and Moicheutria)”. *The Classical Quarterly* 55 (2): 435-446, 2005.
- Varadpande, M. L. (1981). *Ancient Indian And Indo-Greek Theatre*. New Delhi: Abhinav Publications.
- Upadhyaya, U. P. (1996). *Coastal Karnataka*. Rashtrakavi Govind Pai Samshodhana Kendra.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Webb, R. (2008). *Demons and Dancers: Performance in Late Antiquity*. Harvard University Press.
- Whitmarsh, T., & Thomson, S. (2013). *The Romance between Greece and the East*. Cambridge University Press.

### Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure

Although some stem alternation patterns (aka. 'morphemes') in Romance (N, L, PYTA, see Maiden 2018) have been quite substantially described and analysed, others remain underexplored. Here we focus on a pattern of alternation that involves irregular stem shortening (S) in parts of the paradigm (e.g. It. *d-ire* 'say.INF' [vs *dic-iamo*], or *f-a* 'do.3SG.PRS.IND' [vs *fac-iamo*]). Because these short stems have never been subject to a systematic pan-Romance investigation (but see Malkiel [1977], Maiden [2004:237], Mariño Paz [2019], and Dubert García [2021]), we explore their paradigmatic and lexical domain across Romance, and their likely historical origin and motivation.

An initial qualitative inspection (in Maiden et al.'s 2010 ODRVM database) identified the reflexes of *faciō*, *dīcō*, *habeō*, *sapiō*, *possum*, and *volō* as the verbs that show these alternations most frequently. For these we coded, across the 70+ varieties in the database, the paradigmatic distribution of short stems in the paradigm, which yielded 2773 short-stem forms (18.67%), and 12082 long-stem forms (81.33%). Short stems were found to be most frequent in the cells 3SG.PRS.IND, 2SG.PRS.IND, 3PL.PRS.IND, and 2SG.IMP. A quantitative phylogenetic reconstruction of the ancestral states (i.e. presence of a short or long stem in a particular cell in a particular verb) at different points in time along the family tree found that the likelihood of short stems increases through time quite early in the history of the family. Results suggests that various short stems are likely ancestral to Proto-Western-Romance (i.e. Romance minus Sardinian and Balkan).

We propose an explanation for the timing of the emergence of short stems and their lexical and paradigmatic domain. The first part of the explanation relies on the well-known relation between length of expression and frequency of use (e.g. Zipf 1935, Bybee 2006, Gahl et al. 2012). The cells and verbs in which short stems are most common are all extremely frequent: among the 10 most frequent cells and among the 20 most frequent verbs respectively in Latin (Delatte et al. 1981). Having shorter forms for the expression of very common lexical and morphosyntactic meanings is an adaptive property for the efficient transfer of information, which would provide a motivation to prefer these in situations of competition (note that short stems would have appeared accidentally in the paradigm from regular sound change in some forms like *fa*<*fac* 'do.2SG.IMP' and *di*<*dīc* 'say.2SG.IMP').

The exact domain for the spread of short stems, however, must have been influenced, in addition, by the paradigmatic domains of extant stem alternations. The domain of short stems corresponds closely to those cells that partake in N alternations (i.e. stem-vowel differences related to stress), but not in L alternations (i.e. stem-final consonant alterations resulting from palatalization). N-L is an area of the paradigm, hence, within which a single stem would have been expected, whereas stem differences could easily exist with other parts of the paradigm. This links with the timeline of the better-known Romance morphemes. While PYTA is ancestral to all of Romance (i.e. it was present already in Classical Latin), those known as L and N emerged later. The sound changes that generated them all took place in Western-Romance but not always in the varieties that split before. The morphomic niche for the short-stem allomorphy, thus, is the direct result of the cross-classification of the domains of L and N, which means that it must have followed (and not preceded) the emergence in Romance of the L and N morphemes.

## References

- Bybee, Joan. 2006. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Delatte, Louis, Étienne Evrard, Suzanne Govaerts & Joseph Denooz. 1981. *Dictionnaire fréquentiel et index inverse de la langue latine*. L.A.S.L.A, Liege.
- Dubert García, Francisco. 2021. Las formas reducidas de «facēre» en el noroeste de la península ibérica. Convergencia y divergencia geolingüísticas." *Studia linguistica romanica* 6: 163-191.
- Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language* 66, 4: 789-806.
- Herce, Borja. 2019. Morpheme interactions. *Morphology* 29, 1: 109-132.
- Maiden, Martin. 2004. When lexemes become allomorphs: on the genesis of suppletion. *Folia Linguistica* 38: 227-256.
- Maiden, Martin. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford: Oxford University Press.
- Maiden, Martin, John Charles Smith, Silvio Cruschina, Marc-Olivier Hinzelin & Maria Goldbach. 2010. Oxford online database of Romance verb morphology. Available at: <http://romverbmorph.clp.ox.ac.uk/>.
- Malkiel, Yakov. 1977. Editorial Post-Script: Old Spanish far, fer, fazer. *Romance Philology* 31, no. 2: 257-262.
- Mariño Paz, Ramón 2019. O cambio *dizes > dis, diz > di, dizen > din* nas fontes do galego medio. Xosé A. Fernández Salgado, Aquilino S. Alonso Núñez (eds.). *Non haberá illa, pro hai o nome. Homenaxe a Antón Palacio*. Vigo: Universidade de Vigo, 97-113. <https://secretaria.u.vigo.gal/uv/web/publicaciones/public/show/342>.
- Zipf, George Kingsley. 1935. *The Psycho-Biology of Language: An Introduction to Dynamic Philology*. Boston, MA: Houghton Mifflin Company.



### Contact-induced change of Negative Indefinites – the case of Meadow Mari

Since the remarks about the borrowing of negative indefinite pronouns and adverbs (in short: negative indefinites (NI)) in Haspelmath's seminal 1997 book, research on contact-induced change of NI has greatly advanced. Newer research pertains not only to the question how and to what extent markers of NI, i.e., their morphology, can be borrowed, but also how language contact can shape the functional distribution of series of NI (e.g. Elšík & Matras 2006; Karjalainen 2019; Lucas 2013, 2020). In a recent monograph, Breitbarth et al. (2020: Ch. 7) argue, based on van Coetsem's (1988, 2000) model of language contact, that the outcome of contact-induced change of NI depends on the psycholinguistic dominance relations of bilingual speakers in contact situations: if speakers transfer material from an L2 into their dominant language, then transfer of form is usually what is found, whereas in situations of long-lasting bilingualism, the functions of NI may also structurally converge. In this talk I discuss an intriguing case of contact-induced change of NI in Meadow Mari which was likely caused both by direct borrowing of form and structural convergence, combining the two possible developments.

Meadow Mari is a Uralic language spoken by roughly 320,000 people (according to the 2021 All-Russian population census), largely in the Mari El Republic and adjacent areas, as well as in the Ural Mountains. It is situated in the Volga-Kama area, an area of intense historical language contact between Uralic (Mari, Udmurt) and Turkic (Tatar, Chuvash) languages, as well as later intense contact of those languages with Russian. The Uralic and Turkic languages of the area show a large number of lexical, phonological, and morphological convergences which are regarded as outcomes of this contact (e.g. Berezcki 1984; Hesselbäck 2005; Wintschalek 1993). The research of syntactic convergence in the area is still at a nascent stage, however.

NI in Mari, Chuvash, and Russian on the one hand show a morphological parallel; in all three languages, NI are prefixed with *ni-*, a borrowing from Russian in Chuvash and Mari (Egorov 1964; Sibatrova 2021). However, the NI of Mari and Chuvash show additional, structural convergences to the exclusion of Russian which have so far remained unnoticed. For example, in both languages they are licit as standard of comparison (1–2) whereas this is not possible in Russian (3). The data suggests that this structural convergence between Mari and Chuvash arose independently of the borrowing of form from Russian.

(1) Chuvash

*Väl axal'-ten mar [nikam-ran ästa letčik] šutlan-nă.*  
 3SG simple-ABL NEG nobody-ABL skilled pilot consider-prt.perf  
 'Not for no reason was he considered an unsurpassed pilot.' (Chuv.-Rus.-Corpus)

(2) Meadow Mari

*Kö tide saska-m kočk-eš, tudo [nigö deč vijan da patər] lij-eš.*  
 who this fruit-ACC eat-3SG 3SG nobodyfrom strong and powerful be-3SG  
 'He who eats this fruit will be stronger and more powerful than anyone else.' (Mari Corpus)

(3) Russian

*Zdes' prijatn-ee ži-t' čem \*nigde / gde-libo v mire.*  
 here comfortable-CMPR live-INF than nowhere where-ever in world.PREP  
 'Here it is better to live than anywhere in the world.' (elicited)

In this talk I present ongoing research into the mechanisms of contact-induced change of NI. Based on corpus and elicited language data from Mari, Chuvash, and Russian, showing both convergent and divergent structural patterns of NI, I argue that the present-day distribution of the Mari NI can be attributed to multi-layered language contact with Chuvash and Russian, involving both syntactic convergence and morphological borrowing. Besides sketching the development of the Mari NI series, the talk will also discuss possible challenges that systems of negative indefinites as found in Mari and Chuvash can pose for typologies of negative indefinites.

## References

- Bereczki, Gábor. 1984. Die Beziehungen zwischen den finnougriischen und türkischen Sprachen im Wolga-Kama-Gebiet. *Nyelvtudományi Közlemények* 86, 307–314. [The relations between the Finno-Ugric and Turkic languages of the Volga-Kama Area]
- Breitbarth, Anne, Christopher Lucas & David Willis. 2020. *The History of Negation in the Languages of Europe and the Mediterranean. Volume II: Patterns and Processes*. Oxford: Oxford University Press.
- Chuv.-Rus.-Corpus = Dvujazyčnyj korpus čuvašskogo jazyka. <https://ru.corpus.chv.su/> [Bilingual Russian-Chuvash Corpus]
- Egorov, V. G. 1964. *Etimologičeskij slovar' čuvašskogo jazyka*. Čeboksary: Čuvaškoe knižnoe izdatel'stvo. [Etymological dictionary of Chuvash]
- Elšik, Viktor & Yaron Matras. 2006. *Markedness and Language Change. The Romani Sample*. Empirical Approaches to Language Typology 32. Berlin & New York: Mouton de Gruyter.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Hesselbäck, André. 2005. *Tatar and Chuvash Code-copies in Mari*. Uppsala: Uppsala Universitetet.
- Karjalainen, Heini. 2019. Borrowing morphology: The influence of Russian on the Veps system of indefinite pronouns. In: Sofia Björklöf & Santra Jantunen (eds.) *Multilingual Finnic – Language contact and change*. Helsinki: University of Helsinki, 55–87.
- Lucas, Christopher. 2013. Negation in the history of Arabic and Afro-Asiatic. In: Dave Willis, Christopher Lucas & Anne Breitbarth (eds.) *The History of Negation in the Languages of Europe and the Mediterranean. Volume I: Case Studies*. Oxford: Oxford University Press, 399–452.
- Lucas, Christopher. 2020. Contact and the expression of negation. In: Christopher Lucas & Stefano Manfredi (eds.) *Arabic and contact-induced change*. Berlin: Language Science Press, 643–667.
- Mari Corpus = Meadow Mari Corpus of the Giellatekno/Divvun Korp infrastructure: [https://gtweb.uit.no/u\\_korp/?mode=mhr](https://gtweb.uit.no/u_korp/?mode=mhr)
- Sibatrova, Serafima. 2021. *Marijsko-russkie jazykovye kontakty: morfologija marijskogo jazyka*. Joškar-Ola: MarNIIJaLI. [Mari-Russian language contact: Morphology of Mari]
- Van Coetsem, Frans. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.
- Wintschalek, Walter. 1993. *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal*. Wiesbaden: Harrassowitz. [Areal linguistics exemplified by syntactic correspondences in the Volga-Kama Area]

### The Rise of Raising in Early Modern English

Modern English distinguishes between two structures, raising and control, that are surface-identical yet, nonetheless, differ structurally in crucial ways. These two structures are exemplified in (1):

- (1) a. David<sub>i</sub> wants [PRO<sub>i</sub> to like syntax]. [Control]  
 b. David<sub>i</sub> seems [<sub>t</sub> to like syntax]. [Raising]

In (1a), *want* is a control predicate, which means that the DP *David* does not raise from within the infinitival clause; instead there is a PRO subject. In contrast, *seem* is a raising verb that requires *David* to raise to the specifier of the highest TP. The difference between the two structures is that raising verbs, unlike control verbs, do not assign a theta-role to their external argument. Raising and control represent two subclasses of verbs that select non-finite complements in Modern English. The question is what makes a predicate either raise or control? Many attempts have been made to capture the semantic distinction between the two predicate classes (e.g. Stiebels, 2007). However, there are several confounds, including that some predicates in Modern English vacillate between raising and control complements. Furthermore, the conditions that license either raising or control may vary cross-linguistically and some languages do not even encode the distinction (Polinsky, 2013). Therefore, the question is what conditions license the emergence of such a distinction between raising and control predicates in the first place?

The dramatic change in the history of English complementation has been extensively studied (De Smet, 2013; Los, 2005; Rohdenburg, 2006). However, prior accounts have mostly left out the raising-control distinction. The only existing prior diachronic study is Higgins's (1990) investigation of the development of *promise* and *happen* which led him to conclude that raising verbs emerged from control verbs. Hitherto, however, it has been unclear what triggered the conditions for this change. In this paper, we analyze the distribution of high-frequency raising and control verbs using the Penn-Helsinki Parsed Corpus of Early Modern English (Kroch, Santorini & Delfs, 2016). This period is of particular interest for this distinction since at the time non-finite complements were well established in the grammar and it seems like there were more subtle argument structure distinctions emerging with inevitable variation (e.g. Fanego, 2004).

Our pilot study focuses on the ten most frequent verbs in Early Modern English which (in present-day English) take raising and control complements respectively. The complement-taking predicate was coded by hand based on PPCEME data. Figure 1 (overleaf) shows that there is a massive rise in frequency of to-infinitive complements overall during this period, but that this rise is driven by a very small number of verbs – particularly *seem* and *appear* – which are prototypical raising verbs. This indicates that raising as a structure emerges around or before 1500 and spreads dramatically during the Early Modern period on a lexically specific basis. In our paper we will also investigate the diagnostic structural properties of control and raising verbs (e.g. expletive subjects) and test models of lexically-specific syntactic change based on the Tolerance Principle (Yang, 2016; Irani, 2019).

**References**

De Smet, Hendrik. 2013. *Spreading patterns: diffusional change in the English system of complementation*. Oxford: Oxford University Press

Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change. *Diachronica* 21(1), 5–25.

Higgins, F. Roger. 1990. By mishap out of control: on the meaningful descent of raising in English. Ms., University of Massachusetts at Amherst.

Irani, Ava. 2019. *Learning From Positive Evidence: The Acquisition Of Verb Argument Structure*. PhD dissertation, University of Pennsylvania.

Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2016. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania. CD-ROM, first edition, release 3 (<http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3>).

Los, Bettelou. 2005. *The rise of the to-infinitive*. Oxford: Oxford University Press.

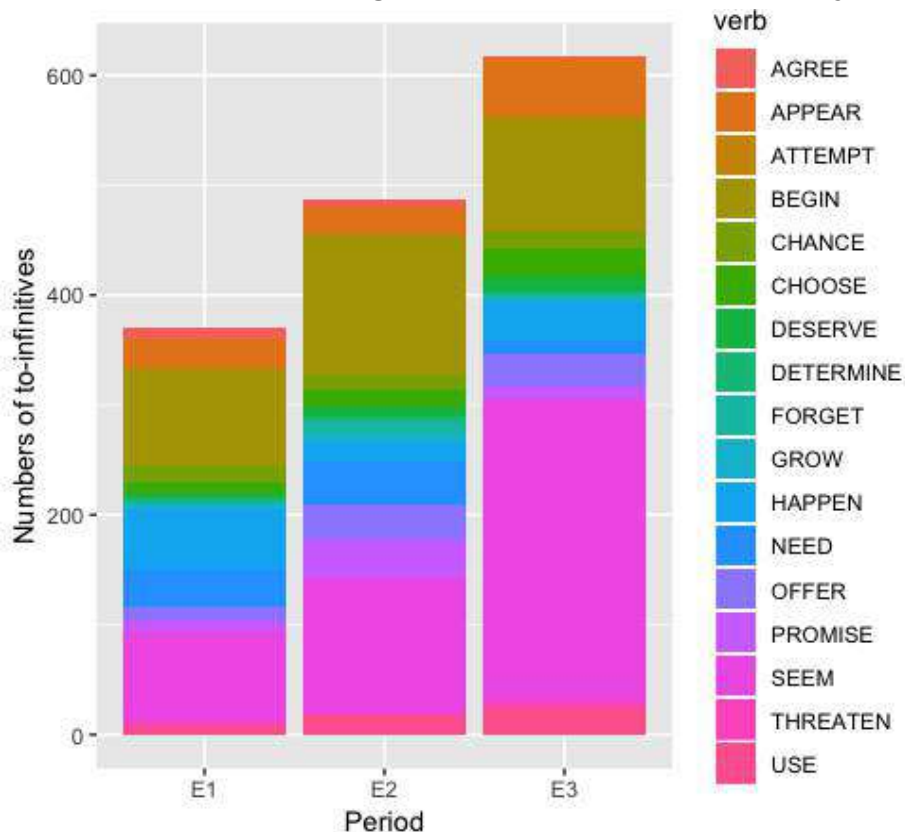
Polinsky, Maria. 2013. Raising and control. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 577–606. Cambridge: Cambridge University Press.

Rohdenburg, Günter. 2006. The role of functional constraints in the evolution of the English complementation system. In Christiane Dalton-Puffer, Nikolaus Ritt, Herbert Schendl & Dieter Kastovsky (eds.), *Syntax, style and grammatical norms: English from 1500–2000*, 143–166. Frankfurt: Lang.

Stiebels, Barbara. 2007. Towards a typology of complement control. In Barbara Stiebels (ed.), *Studies in Complement Control*, 1–80. ZAS Working Papers in Linguistics 47. Berlin: ZA.

Yang, Charles. 2016. *The Price of Linguistic Productivity*. MIT Press.

**Figure 1: the 10 most frequent raising and control verbs in PPCEME by period**



### Conservative pressure on the progressive: the passival

Keywords: passive construction, progressive construction, grammaticalization, Early Modern English, functional load, idiolect

This talk discusses the internal reorganization of the progressive construction in Early Modern English (ca. 1500-1700), with special attention for the so-called ‘passival’, which was active in form but passive in meaning (*the house was building* ‘the house was being built’). I will argue that, while the passival was on the increase in Early Modern English, it ‘needed’ to disappear in order to make further progress of the progressive possible. As such, it may be seen as a ‘false turn’ in a grammaticalization process that was eventually resolved by resourceful language users.

While the progressive construction started to increase its scope in the late Middle English period (e.g., Kranich 2010), its grammaticalization was completed only with the introduction of the progressive passive (*the house was being build*) in the late eighteenth century (Hundt 2004: 101). This introduction has been characterized as coming ‘out of the blue’ rather late (Anderwald 2016: 201–202). I will argue that its lateness is due to the passival being the more natural development out of the more nominal origins of the progressive, which originated in part as a prepositional gerundial construction of the type *he was on hunting*. While the preposition gradually eroded, the original nominal characteristics still shimmer through when in Early Modern English the verb’s object was occasionally expressed by *of* NP (e.g. *He was killing of this man* instead of *he was killing this man*) instead of as a direct object. More generally, nominalizations typically do not formally differentiate agent and patient (cf. the famous Latin example *amor matris* where the mother could be both subject or object of the love). In that respect, the passival is a natural functional expansion of the gerundial origin of the progressive. However, as the progressive became more and more verbal, the inclusion of ever more animate and inanimate contexts created competition within the progressive construction, with the passival blocking progress in two different domains. Passivals with human agents, which were occasionally attested, as in (1), had difficulty gaining ground because of obvious confusion with active progressives.

- (1) *The Place where they say the Virgin Mary Prayed for St. Stephen, while he was Stoning.*  
 ‘The place where they say the Virgin Mary prayed for St. Stephen, while he was being stoned.’ (Nathaniel Crouch 1672. EMMA)

Second, active progressives with inanimate subjects remained virtually restricted to intransitives and passivals.

Data for the analysis come from the analysis of 15 prolific authors from the EMMA corpus (Petré et al. 2019), collectively constituting a corpus of more than 20 million words. While the unstable nature of the passival has been hinted at in earlier work (e.g., Hundt 2004), the EMMA corpus allows for a more fine-grained and also individual-level analysis. Such an analysis shows that the functional block of the passival is a consistent cognitive reality across most (but not all) authors. However, where this cognitive dissonance within the progressive construction remained inconspicuous in early generations, it came more and more to the fore in later generations as the passival became more prevalent, eventually leading to its demise in favour of the progressive passive, in spite of the latter’s semi-artificial introduction by professional writers.

### References

Anderwald, Lieselotte 2016. *Language between Description and Prescription: Verb Categories in Nineteenth-century Grammars of English*. Oxford: Oxford University Press.

- Kranich, Svenja. 2010. The progressive in Modern English: A corpus-based study of grammaticalization and related changes. *The Progressive in Modern English*. Leiden: Brill.
- Hundt, Marianne 2004. The passival and the progressive passive – a case study of layering in the English aspect and voice systems. In Hans Lindquist and Christian Mair (eds.), *Corpus Approaches to Grammaticalisation in English*. Amsterdam: Benjamins, pp. 79–117.
- Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing and Odile A.O. Strik 2019. Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME journal* 43: 83–122.



### Another look at Noun-Genitive vs. Genitive-Noun in Early New High German

It has long been recognized that historical varieties of German exhibit variation between prenominal vs. postnominal genitives (GN vs. NG orders), with an overall development away from GN and toward NG under various conditions (Behaghel 1923). In broad strokes, retention of GN order is associated with possessive genitives (as opposed to subjective, objective, partitive, or explicative genitives), genitives denoting persons, and single-words, proper names, or pronouns. By 1700, near the beginning of the Modern German era, GN order accounts for only 10% of adnominal genitives, and GN becomes restricted to proper nouns by 1750 (Niehaus 2016). In Early New High German (ENHG; 1350-1650) there is often variation between GN and NG within these conditions even within individual texts. Despite several examinations of this variation and change (e.g., Ebert 1988; Lunt Lanouette 1990, 1998; Pickl 2020), this topic remains surprisingly under-researched.

In this paper, we explore the hypothesis that in at least some varieties of ENHG, the grammar of adnominal genitives includes two properties not found in Modern German:

- (1) Adnominal genitives are generated in the Specifier position of the NP.
- (2) NG order is derived by an optional rule that extraposes the genitive DP or a subpart of it and which is sensitive to the length/weight of the genitive phrase.

Property (1) accounts for examples like (3), in which pre-nominal genitives appear to the right of attributive adjectives, and property (2) accounts for instances in which the pre-nominal genitive is modified by a post-nominal phrase.

- (3) *eyn besunder* [<sub>DP</sub> *Rulands*] *streitgesel* (Fierrabras 196 [1533])  
a certain Ruland-GEN battle-companion ‘a certain combatant of Ruland’
- (4) [<sub>DP</sub> *Josephs* <sub>TPP</sub>] *sun* [<sub>PP</sub> *von aramathea*] (Karrenritter 472 [1430])  
Joseph-GEN son of Aramathea ‘Joseph of Aramathea’s son’

Taken together, these properties yield the tendency for NG order with longer genitive DPs, but GN with one-word genitives, as well as the split construction illustrated in (4).

We present data from an ongoing corpus-based study of GN vs. NG variation in ENHG. The corpus will ultimately consist of 60 texts, with one text per 50-year bin from 10 dialects, representing a variety of genres. Texts in the corpus are constituency parsed according to the Penn annotation system (e.g., Kroch 2020). Preliminary data have been extracted from the first four completely parsed texts, yielding the following results thus far:

- By text: GN varies widely from 84% in *Karrenritter* (1430) to 48% in *Fierrabras* (1533) to just 7% in *Geistlicher Mai* (1529). More texts need to be analyzed to tease apart the effects of time, dialect, and genre.
- Proper vs. common noun genitive: In texts other than *Geistlicher Mai*, proper nouns appear in GN order more frequently (67-83%) than common nouns.
- Genitive type: for both GN and NG, the vast majority of adnominal genitives are possessives, thus the effect of this factor is inconclusive so far.
- Length: One-word genitives most strongly favor GN (67% excluding *Geistlicher Mai*) and longer genitives increasingly disfavor it (down to 19% for five or more words, again excluding *Geistlicher Mai*.)

We suggest that the Modern Standard German pattern in which only personal possessives in *-s* (which are probably no longer genitives, see Fuß 2011) and possessive adjectives occur prenominally and full genitive DPs occur only in the order NG, results from a reanalysis triggered by the increased application of the extraposition rule in the late ENHG period.



## References

- Behaghel, Otto. 1923. *Deutsch Syntax: Eine geschichtliche Darstellung*. Volume 4. Heidelberg: Carl Winter.
- Ebert, Robert Peter. 1988. Variation in the position of the attributive genitive in sixteenth century German. *Monatshefte* 80: 32-49.
- Fuß, Eric. 2011. Eigennamen und adnominaler Genitiv im Deutschen. *Linguistische Berichte* 225: 19-42.
- Kroch, Anthony. 2020. *The Penn Parsed Corpora of Historical English*. University of Pennsylvania. <https://www.ling.upenn.edu/hist-corpora/>
- Lunt Lanouette, Ruth. 1990. *Prenominal specifiers in the Early New High German Noun Phrase*. Unpublished dissertation, Princeton University.
- Lund Lanouette, Ruth. 1998. The attributive genitive in the history of German. *Germanic linguistics: Syntactic and diachronic*, ed. by Rosina L. Lippi-Green and Joseph Salmons, 85-102. Philadelphia: John Benjamins.
- Niehaus, Konstantin. 2016. *Wortstellungsvarianten im Schriftdeutschen: Über Kontinuitäten und Diskontinuitäten in neuhochdeutscher Syntax*. Heidelberg: Winter.
- Pickl, Simon. 2020. Polarization and the emergence of a written marker: A diachronic corpus study of the adnominal genitive in German. *Journal of Germanic Linguistics* 32: 145-182.

## Latin *placēre* as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis

Traditionally, the Latin verb *placēre* ‘like, please’, which is a two-place predicate licensing a nominative and a dative argument, is analyzed as a Nom-Dat verb, which means that it is the nominative that is considered to be the subject, whereas the dative is assumed to be the object. However, examples like the ones in (1) below show that either order, Dat-Nom as in (1a) and Nom-Dat as in (1b), is acceptable in Latin. The fact that both word order patterns are equally fine has been observed in the field, but it has generally been attributed to what is termed ‘free word order’ (Devine & Stephens 2006, Spevak 2010, inter alia).

- (1) a. (Ov. *Tr.* 4,10,19)  
 At **mihi iam puero caelestia sacra** placebant  
 and I.DAT even boy.DAT mystic.NOM service.NOM like.IMPF.3PL  
 “And I, even as a boy, liked the mystic services”
- b. (Cic. *Orat.* 2, 42, 179)  
**Qui ordo tibi** placeat inquit Catulus  
 what.NOM arrangement.NOM you.DAT like.SBJV.PRS.3SG say.PRS.3SG Catulus.NOM  
 “What arrangement would please you, said Catulus [...]”

The present paper advocates a more radical approach, namely in terms of *alternating predicates*. Such structures have also been shown to exist in Germanic (Barnes 1986 for Faroese, Allen 1995 for Old English, Barðdal 1998 for the history of the Mainland Scandinavian languages, Barðdal 2001 for Modern Icelandic, Barðdal, Eythórsson & Dewey 2019 for Modern German) and in Romance (Illoaia 2022 for Romanian), and they may also exist in Baltic, Slavic, Hittite and Sanskrit (cf. Barðdal 2023: Ch. 3). Alternating predicates systematically occur with two diametrically opposed argument structures: a Dat-Nom argument structure and a Nom-Dat argument structure. As a consequence, (1a) would contain a dative subject and a nominative object, whereas the opposite is true for (1b), which would contain a nominative subject and a dative object.

Here we focus on one specific verb in Latin, *placēre* ‘like, please’, confining our analysis to occurrences of this verb as a two-place predicate, which licenses a nominative and a dative argument. The data are drawn from the *LatinISE* corpus, which stretches a period of approximately 500 years, from the Archaic Period (3rd–2nd century BCE) up to the Late Latin period (7th century). Our dataset comprises 350 occurrences of *placēre*, annotated for (pro)nominativity, person, definiteness, length, and animacy.

In line with Eythórsson & Barðdal (2005), Barðdal & Eythórsson (2012, 2018), and Barðdal (2023), we define subject as the leftmost argument of the argument structure. This definition is based on a generalization across a range of diagnostics which have been successfully applied to various Germanic languages and the behavior of the arguments relative to these. Unfortunately, research on the modern linguistic concept of subject is still in its early stages within the Latin scholarship. The issue was first dealt with by Michaelis (1992), later to be taken up by Baños Baños (2003) and Fedriani (2009, 2014).

More recently, Barðdal et al. (2023) have shown that several subject tests identified for the Germanic languages may be successfully applied to Latin and Ancient Greek. On this basis, we discuss the behavior of the two arguments of *placēre* with regard to these subject tests and show that either argument, the nominative or the dative, passes the subject tests in Latin. We focus in particular on data involving word order, raising-to-object, raising-to-subject, and control infinitives.

Barðdal et al. (2023) further document that ordinary nominative subjects in Latin precede the object in ca. 70% of the cases, thereby establishing a baseline against which to compare the statistics obtained for alternating predicates. We compare our Latin word order statistics with corresponding statistics from Old English (Allen 1995) and Old Norse-Icelandic (Elens, Somers & Barðdal 2023), arguing that Latin *placēre* ‘like, please’, shows the same distributional properties as alternating predicates in the Early Germanic languages.

## References

- Allen, Cynthia L. 1995. *Case Marking and Reanalysis: Grammatical Relations from Old to Early Modern English*. Oxford: Oxford University Press.
- Baños Baños, José Miguel, 2003. *Paenitet* y los verbos impersonales de sentimiento en latín: sintaxis y pragmática del acusativo personal [*Paenitet* and the impersonal verbs of feeling in Latin: syntax and pragmatics of the personal accusative]. *Praedicatoria: Estudios sobre complementación verbal en griego y latín*, ed. by José Miguel Baños Baños, Concepción Cabrillana, María Esperanza Torrego & Jesús de la Villa, 51–77. Santiago de Compostela: Universidad de Santiago de Compostela.
- Barðdal, Jóhanna. 1998. Argument structure, syntactic structure and morphological case of the impersonal construction in the history of Scandinavian. *Scripta Islandica* 49: 21–33.
- Barðdal, Jóhanna. 2001. The perplexity of Dat-Nom verbs in Icelandic. *Nordic Journal of Linguistics* 24: 47–70.
- Barðdal, Jóhanna. 2023. *Oblique Subjects in Germanic: Their Status, History and Reconstruction*. To appear with Mouton de Gruyter.
- Barðdal, Jóhanna, Eleonora Cattafi, Leonardo Biondo, Laura Bruno & Serena Danesi. 2023. Non-nominative subjects in Latin and Ancient Greek: Applying the subject tests on Early Indo-European languages. *Indogermanische Forschungen* 128.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. “Hungering and lusting for women and fleshly delicacies”: Reconstructing grammatical relations for Proto-Germanic. *Transactions of the Philological Society* 110(3): 363–393.
- Barðdal, Jóhanna & Thórhallur Eythórsson. 2018. What is a subject? The nature and validity of subject tests. *Non-Canonically Case-Marked Subjects: The Reykjavík–Eyjafjallajökull Papers*, ed. by Jóhanna Barðdal, Na’ama Pat-El & Stephen Mark Carey, 257–273. Amsterdam: John Benjamins.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1): 107–170.
- Barnes, Michael. 1986. Subject, nominative and oblique case in Faroese. *Scripta Islandica* 37: 13–46.
- Devine, A. M., & Laurence D. Stephens. 2006. *Latin Word Order: Structured Meaning and Information*. Oxford: Oxford University Press.
- Elens, Wannes, Joren Somers & Jóhanna Barðdal. 2023. The alternating behaviour of ‘like’ in Old Norse-Icelandic: Facts or Fiction. Ms. Ghent University.
- Eythórsson, Thórhallur & Jóhanna Barðdal. 2005. Oblique subjects: A common Germanic Inheritance. *Language* 81(4): 824–881.
- Fedriani, Chiara. 2009. The “Behaviour-Before-Coding-Principle”: Further Evidence from Latin. *Archivio Glottologico Italiano* XCIV: 156–184
- Fedriani, Chiara. 2014. *Experiential Constructions in Latin*. Leiden: Brill.
- Ilioiaia, Mihaela. 2022. Toward canonical marking of core arguments in Romanian? The case of *plăcea*. A presentation delivered at ICHL 25, Oxford.
- Michaelis, Laura A. 1992. On Deviant Case-Marking in Latin. *Advances in Role and Reference Grammar*, ed. by Robert D. Van Valin, 311–373. Amsterdam: John Benjamins.
- Spevak, Olga. 2010. *Constituent Order in Classical Latin Prose*. Amsterdam: John Benjamins.

## The diachronic development of future markers in Chinese

Linguistic analyses of future marking distinguish two different viewpoints: 1) an objectivistic view, i.e., the *branching future* concept, according to which the future is a kind of modality (Portner 2009: 236), and 2) a subjective perspective, in which future tense simply refers to a time following speech time (Reichenbach 1947, Bochnak 2019). Within the cartographic approach, future tense is hosted in a functional projection separate from the projections hosting different kinds of modality.

(1) ModP<sub>epist</sub>>TP(Past)>TP(fut)>AspP<sub>habit</sub>>ModP<sub>volition</sub>>AspP...>ModP<sub>obligation</sub>>ModP<sub>permission/ability</sub> (modified from Cinque 2004)

Chinese does not have morphological tense marking; temporal and aspectual distinctions are expressed analytically. However, future is the most regularly expressed temporal (and/or aspectual) concept in Archaic Chinese. Future markers by default appear in complementary distribution with other aspecto-temporal markers, and in Archaic Chinese they permit a purely temporal reading. This is shown in example (2) with the future marker *jiāng*, which is semantically similar to the Pre- and Early Archaic future marker *qí* 其, frequently appearing in oracular predictions (Djamouri 2009).

(2) 公 將 以 某 日 薨 (Lüshi chunqiu LAC/EMC)  
gōng jiāng yǐ mǒu rì hōng  
duke FUT YI such.and.such day pass.away  
'The duke will pass away on such-and-such day.'

In Early Middle Chinese, new future markers grammaticalize from the lexical verbs *dāng* 當 'correspond to', which first develops into a deontic auxiliary 'should', and from the volitional verb *yù* 欲. Both markers include modal, besides their temporal readings. Similar to *jiāng*, they appear in complementary distribution with other aspecto-temporal markers in the TP layer. When they appear in combination with future *jiāng*, DANG and YU either have to be analyzed as pre-modal verbs, or the combination functions as a disyllabic future marker.

(3) a. 若干 百年 當 至 于 闐 國。  
Ruògān bǎi nián dāng zhì yútiàn guó  
Several hundred year DANG arrive Khotanstate  
'After several hundred years it will/should arrive in Khotan.'  
b. 『欲云何作?』  
yù yúhé zuò  
YU how do  
'How will you (do you want to) do it?'

The only syntactic difference between EMC future DANG and YU, and LAC *jiāng* is the position of negation. Since DANG and YU are verbal heads, NEG has to precede them, but it has to follow the aspecto-temporal adverb *jiāng*. Meisterernst (2020) proposed two different functional projections hosting future tense and deontic modality for LAC and EMC. Contrastingly, we propose one unified zone within TP (following Ramchand and Svenonius 2014), which can be targeted by either a root modal necessity marker or a by future marker; epistemic necessity is hosted in a higher projection. In LAC, this zone could be occupied by either a modal negator or a future marker; modal verbs were confined to the lexical, i.e., the *vP* layer. The situation changes in EMC, when true deontic modals emerged, which were hosted in the TP layer in the same zone as modal negators and future markers in LAC.

(4) [CP ModP<sub>epist</sub> [TP TP<sub>zone</sub> FUT/ASP/NEG<sub>Mod</sub>/MOD<sub>deont</sub> [<sub>vP</sub> VP<sub>zone</sub> (MOD<sub>circum</sub>) *vP*]]]

Similar to what Ramchand and Svenonius propose, the markers in the *vP* external zone are characterized by a relative independence of the aktionsart feature of the *vP*.

## References

Bochnak, Ryan M. 2019. Future reference with and without future marking. *Language and Linguistics Compass* 13,1. <https://doi.org/10.1111/lnc3.12307>

- Djamouri, Redouane. 2009b. Jiaguwen zhong 'qi' de yongfa 甲骨文中“其”的用法 [Use of *qi* in the Oracle Bone Inscriptions]. In Feng Li & Yang Yonglong (eds.), *Hanyu shiti wenti lishi yanjiu* 漢語時体問題歷史研究 [Diachronic Studies on Tense and Aspect in Chinese], Beijing: Yuwen chubanshe.
- Meisterernst, Barbara. 2020. A syntactic approach to the grammaticalization of the modal marker dāng 當 in Middle Chinese, *Journal of Historical Syntax*, 4,5. 2020-05-29.
- Portner, Paul (2009). *Modality*. Oxford: OUP
- Ramchand, Gillian and Svenonius, Peter. 2014. Deriving the Functional Hierarchy. *Language Sciences* 46: 152-174.
- Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: Dover Publication.

## Properties of Complex Compounds in Old Japanese

This project investigates complex compounds in Old Japanese (OJ), the language of 8<sup>th</sup> century Japan. Properties of compounds are well studied for modern Japanese, but less so for earlier stages of the language, and this project will add to our understanding of how the Japanese language developed over time.

Mithun (1984) describes noun incorporation as a syntactic morphological process. In her framework there are 4 types of noun incorporation. This framework is hierarchical:

- 1) lexical compounding where the noun is assigned the role of patient, location, or instrument by the incorporating verb (IV)
- 2) the manipulation of case, where the incorporated noun (IN) loses its argument status and another noun receives it
- 3) the manipulation of discourse structure, where the IN is part of information structure representing known (and not new) information
- 4) classificatory noun incorporation, where the IN narrows the scope of the IV but can be accompanied by an NP which classifies the IN

OJ shows evidence of types 1 and 4.

In addition, Rosen (1989) claims there are 2 separate word formation processes.

- 1) argument structure of the IV changes, so that the verb loses an argument
- 2) argument structure is unaltered

Rosen claims that a language may have one or the other process, but not both. OJ, however, shows both processes.

The data for the initial study were extracted from the Oxford Corpus of Old Japanese (OCOJ), a linguistically annotated corpus of the language of 8<sup>th</sup> century Japan, the earliest attested stage of the Japanese language. The data were then further annotated with information about orthography, verbal inflection, syntactic structure, whether the noun adjacent to a verb was bound or free, whether *rendaku* (sequential voicing) occurred, and whether the incorporated noun was modified.

Criteria for determining whether or not nouns are incorporated in OJ are 1) positive evidence for incorporation; 2) positive evidence against incorporation; and 3) absence of evidence against incorporation.

Unlike Modern Japanese, the data show that OJ has both noun and phrasal incorporation; incorporated nouns in OJ can be modified by adjectives, possessives, and relative clauses. A variety of semantic roles can be assigned to the incorporated noun, including both arguments and adjuncts. Subjects of intransitive verbs can be incorporated, and, finally, incorporated structures are not always nominalized and do not always occur in light verb constructions.

- Frellesvig, Bjarke. 2010. *A History of the Japanese Language*. Cambridge: Cambridge University Press.
- Frellesvig, Bjarke, Stephen Wright Horn, Kerri L. Russell, & Peter Sells. n.d. The Oxford Corpus of Old Japanese. <http://vsarpj.orinst.ox.ac.uk/corpus/index.html>
- Kinuhata, Tomohide. 2010. 上代語の名詞抱合について [On incorporated nouns in Old Japanese]. 語文, 92/93, 34-44.
- Kageyama, Taro. 1982. "Word formation in Japanese". *Lingua* 57, 212-258.
- Martin, Samuel E. 1987. *The Japanese language through time*. New Haven: Yale University Press.
- Mithun, Marianne. 1984. "The evolution of noun incorporation". *Language* 60, 847-894.
- Omodaka Hisataka, editor. 1967. 時代別国語大辞典：上代編 [A dictionary of the Japanese language by periods: Old Japanese volume]. Tokyo: Sanseidō.
- Rosen, Sara Thomas. 1989. "Two type of noun incorporation: A lexical analysis." *Language* 65, 2, 294-317.
- Russell, Kerri L. 2012. "Noun incorporation in Old Japanese (上代日本語における名詞抱合について)" presented at 通時コーパスプロジェクト・オックスフォード大VSARPJプロジェクト合同シンポジウム(NINJAL Diachronic Corpus Project – Oxford VSARPJ Project Joint Symposium Corpus Based Studies of Japanese Language History).
- Russell, Kerri L., John Whitman, and Yuko Yanagida. 2014. "Was Old Japanese a polysynthetic language?" presented at International Symposium on Polysynthesis in the World's Languages 国際シンポジウム「世界の言語における複統合性」National Institute for Japanese Language and Linguistics, 20-21 February, 2014
- Shibatani, M. and T. Kageyama. 1988. Word formation in a modular theory of grammar: Post-syntactic compounds in Japanese. *Language*, 64(3):451-484.
- Stekauer, Pavol, Salvador Valera, and Livia Kortvélyess. 2012. *Word-formation in the World's languages: A typological survey*. Cambridge: Cambridge University Press.



## The Alternating Behavior of ‘Like’ in Old Norse-Icelandic: Facts or Fiction

In a recent article, Sigurðsson & Viðarsson (2020) claim that the verb *líka* ‘like’ in Old Norse-Icelandic is an alternating Dat-Nom/Nom-Dat verb, as opposed to Modern Icelandic where this same verb is uncontroversially a non-alternating Dat-Nom verb. The difference between the two is that alternating verbs may instantiate two diametrically opposite argument structures, i.e. Dat-Nom and Nom-Dat, while non-alternating Dat-Nom verbs only instantiate one of these, the Dat-Nom argument structure (Bernóðsson 1982, Barnes 1986, Jónsson 1997–98, Barðdal 1999, 2001, Barðdal, Eythórsson & Dewey 2014, 2019, Platzack 1999, Wood & Sigurðsson 2014, Somers & Barðdal 2022, inter alia).

Sigurðsson & Viðarsson (2020) base their claims on both language internal Old Norse-Icelandic evidence and on the comparative evidence from the other early Germanic languages (Fischer & van der Leek 1983, Allen 1986, 1995, Barðdal 1998, Eythórsson & Barðdal 2005). Starting with the comparative evidence, it has been argued for Old English that *lician* is an alternating verb in that language (Allen 1995: 141) and the same has been argued for *galeikan* in Gothic (Eythórsson & Barðdal 2005: 833).

Turning to the language internal evidence for an alternating analysis of *líka* in Old Norse-Icelandic which Sigurðsson & Viðarsson introduce, this consists of data involving two subject tests, i) control infinitives and ii) word order. Sigurðsson & Viðarsson present several examples of control infinitives with the verb *líka* where it is indeed the nominative and not the dative that is left unexpressed in such structures. This they take as conclusive evidence that the nominative behaves syntactically as a subject and the dative as an object. Likewise, Sigurðsson & Viðarsson also present a handful of examples involving word order distribution, which are incompatible with a Dat-Nom analysis of the argument structure of *líka*, and call instead for a Nom-Dat analysis of the relevant structures.

While we agree with Sigurðsson & Viðarsson on their analysis of the relevant control infinitives, in that there is no doubt that these examples show that it is indeed the nominative that is left unexpressed and not the dative, we still call into question the relevance of their data set. As they acknowledge themselves, all their examples of control infinitives are from translated texts, and the same applies to their word order examples. In general, translated examples may well be taken to speak for authenticity, but for these particular examples, we argue that the relevant translations are word-for-word glosses of the Latin verb *placere* which sometimes means ‘like’ and sometimes ‘please’, depending on its argument structure (cf. Cluyse, Somers & Barðdal 2023).

Therefore, in order to shed light on this issue, we present word order statistics for *líka* in Old Norse-Icelandic. The data have been extracted from three different sources: i) the Saga Corpus, ii) the Icelandic Text Archive, and iii) the Dictionary of Old Norse Prose (ONP), resulting in a dataset containing approximately 200 occurrences of *líka*. It turns out that there is a major divide between native and translated texts. This means that in texts originally written in the Old Norse-Icelandic vernacular, *líka* consistently occurs with a Dat-Nom order, unless the nominative contains a demonstrative pronoun, then the Nom-Dat word order is preferred. This suggests that *líka* could only instantiate the Dat-Nom argument structure construction in texts originally written in Old Norse-Icelandic, with the Nom-Dat word order representing topicalizations. In contrast, in the translated texts, the proportions between Dat-Nom and Nom-Dat with *líka* are more even and are not tied to word class. Therefore, since any “alternating” behavior of *líka* is confined to translated texts, we conclude that this seeming behavior is a translation effect.

Our alternative analysis of the data involving Old Norse-Icelandic *líka* above makes a certain prediction, namely that the existing alternating analysis of Old English *lician* and Gothic *galeikan* may be equally faulty as the analysis provided by Sigurðsson & Viðarsson, as most if not all instances in Old English and Gothic are also translations. In other words, our analysis predicts that the apparent alternating behavior of ‘like’ in Old English and Gothic may also be a translation effect.

**References:**

- Allen, Cynthia L. 1986. Reconsidering the history of *like*. *Journal of Linguistics* 22. 375–409.
- Allen, Cynthia L. 1995. *Case Marking and Reanalysis: Grammatical Relations from Old to Early Modern English*. Oxford: Oxford University Press.
- Barðdal, Jóhanna. 1998. Argument structure, syntactic structure and morphological case of the impersonal construction in the history of Scandinavian. *Scripta Islandica* 49. 21–33.
- Barðdal, Jóhanna. 1999. The dual nature of Icelandic psych-verbs. *Working Papers in Scandinavian Syntax* 64. 78–101.
- Barðdal, Jóhanna. 2001. The perplexity of Dat-Nom verbs in Icelandic. *Nordic Journal of Linguistics* 24. 47–70.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2014. Alternating predicates in Icelandic and German: A sign-based construction grammar account. *Working Papers in Scandinavian Syntax* 93. 50–101.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1). 107–170.
- Barnes, Michael. 1986. Subject, nominative and oblique case in Faroese. *Scripta Islandica* 37. 13–46.
- Bernódusson, Helgi. 1982. Ópersónulegar setningar [Impersonal sentences]. University of Iceland Master's Thesis.
- Cluyse, Brian, Joren Somers & Jóhanna Barðdal. 2023. Latin *placere* as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis. Ms., Ghent University.
- Eythórsson, Thórhallur & Jóhanna Barðdal. 2005. Oblique subjects: A common Germanic Inheritance. *Language* 81(4). 824–881.
- Fischer, Olga C. M. & Frederike C. van der Leek. 1983. The demise of the Old English impersonal construction. *Journal of Linguistics* 19. 337–368.
- Icelandic Text Archive = Stofnun Árna Magnússonar. (n.d.) Leit. In *Íslenskt Textasafn*. <http://corpus.arnastofnun.is/leit.pl>.
- Jónsson, Jóhannes Gísli. 1997–98. Sagnir með aukafallsfrumlagi [Verbs selecting for oblique subjects]. *Íslenskt mál* 19–20. 11–43.
- ONP = University of Copenhagen. (n.d.). Líka vb. In M. Arvidsson, S. Battista, H. Degnbol, P. Ellyton, Þ. Helgadóttir, B. C. Jacobsen, E. Þ. Jóhannsson, A. S. Kjeldsen, J. E. Knirk, J. Lindholm, E. Rode, C. Sanders, A. Sigurðardóttir & T. Wills (Eds.), *ONP: Dictionary of Old Norse Prose*. <https://onp.ku.dk/onp/onp.php?o98918>.
- Platzack, Christer. 1999. The subject of Icelandic psych-verbs: A minimalistic account. *Working Papers in Scandinavian Syntax* 64. 103–116.
- Wood, Jim & Halldór Ármann Sigurðsson. 2014. *Let*-causatives and (a)symmetric DAT-NOM constructions. *Syntax* 17(3). 269–298.
- Saga Corpus = Stofnun Árna Magnússonar. (n.d.). Líka. In *Saga Corpus*. <https://malheildir.arnastofnun.is/?mode=fornrit#?lang=en>.
- Sigurðsson, Einar Freyr & Heimir van der Feest Viðarsson. 2020. Frá skiptisögn til ósamhverfrar aukafallssagnar: Um líka í fornu máli [From an alternating to a non-alternating oblique subject verb: About *líka* 'like' in Old Icelandic]. *Orð og tunga* 22. 39–68.
- Somers, Joren & Jóhanna Barðdal. 2022. Alternating Dat-Nom/Nom-Dat verbs in Icelandic: An exploratory corpus-based analysis. *Working Papers in Scandinavian Syntax* 107. 83–110.

## The rise of *do*-support during Scots anglicisation: Insights from the *Parsed Corpus of Scottish Correspondence*

Lisa Gotthard, University of Edinburgh

This paper explores the rise of *do*-support in Scots, and investigates whether it exhibited similar functions to English ‘intermediate’ *do* (as analysed by Ecay (2015)) before regularising into its current function. In doing so, this study provides more insight into syntactic change in Scots during the period of *anglicisation*, starting in the 16th century, when English became favoured over Scots in writing. While *anglicisation* is usually discussed in terms of lexis and orthography (e.g. Devitt (1989); Meurman-Solin (1993b); Kniezsa (1997)), the 16th to 18th century has been obscure within Scots syntax research; aided by the new *Parsed Corpus of Scottish Correspondence* (PCSC; Gotthard 2022), this gap now begins to be filled.

*Do*-support is the mandatory insertion of the auxiliary *do*, which has historically been bleached of its semantic meaning and today has a strictly morpho-syntactic function: carrying tense and agreement features when the transfer of these features between the subject and main verb is interrupted and no other auxiliary is present (E.g., *I do not eat cake*). English *do*-support is extensively researched (e.g., Ellegård 1953; Denison 1985; Ecay 2015; Garrett 1998; Kroch 1989; Nurmi 2011; Poussa 1990; Tieken-Boon van Ostade 1990; van der Auwera and Genee 2002; Warner 2002), but the feature in Scots has received less attention. It has been suggested that Scots *do*-support is a transfer from English, supported by that the feature emerged in Scots during the height of anglicisation (Gotthard 2019, 2022, 2023; Meurman-Solin 1993a), and that it remains variable in more traditional dialects (Jamieson 2015; Jonas 2002; Smith 2000), but it could also have developed independently, from the causative *do* auxiliary inherited from Old English. Based on observations already made by Ellegård (1953), Ecay (2015) identifies an ‘intermediate’ *do* auxiliary in English pre-1575; an agentive marker which merges in a lower syntactic position than post-1575 *do*. This leads to the question of whether Scots *do* follows the same path, or is adopted with the same function as post-1575 English *do* – if a candidate feature emerges in the receiving language in its fully grammaticalised form, then it is more likely to be a transferred feature (e.g. Pa-Tel 2013).

In order to investigate (i) the emergence and trajectory of *do* in Scots, and (ii) whether Scots *do*-support is a plausible outcome of anglicisation, proportions of affirmative and negative declarative *do* in the PCSC are calculated across different syntactic contexts. The likelihood of the feature being an anglicisation outcome is evaluated by assessing the results against criteria for contact-induced change (e.g., Thomason and Kaufman 1988; Thomason 2001; Pa-Tel 2013; Robbeets and Cuyckens 2013; Poplack and Levey 2010). It is found that Scots *do* emerges towards the end of the 16th century, and remains at low proportions (around 20%) until ca. 1700 when it increases more dramatically, and initially behaves largely consistent with what Ecay (2015) observed for ‘intermediate’ *do* in English. The social context and timing of the rise of Scots *do* suggest that it is a contact-induced change, but the fact that the auxiliary shows ‘intermediate’ *do* qualities warrants further discussion; if this is truly an intermediate stage in the grammaticalisation of *do* then the analysis is compromised, but this *do* auxiliary may also be as different type of *do*, spreading northward from English into Scots.

## References

- Denison, D. (1985). The origins of periphrastic do: Ellegård and visser reconsidered. In Eaton, R., Fischer, O., Koopman, W. F., and van der Leek, F., editors, *Papers from the 4th international conference on English historical linguistics* 4th International Conference on English Historical Linguistics, pages 45–60. Amsterdam/Philadelphia: John Benjamins Publishing.
- Devitt, A. J. (1989). *Standardizing written English: diffusion in the case of Scotland, 1520-1659*. Cambridge: Cambridge University Press.
- Ecay, A. (2015). *A multi-step analysis of the evolution of English do-support*. PhD thesis, University of Pennsylvania.
- Ellegård, A. (1953). *The auxiliary do: The establishment and regulation of its use in English*. Stockholm: Almqvist and Wiksell.
- Garrett, A. (1998). On the origin of auxiliary do. *English Language and Linguistics*, 2:283–330.
- Gotthard, L. (2019). Why do-support in Scots is different. *English Studies*, 100(3):1–25.
- Gotthard, L. (2022). *Syntactic change during the anglicisation of Scots: Insights from the Parsed Corpus of Scottish Correspondence*. PhD thesis, University of Edinburgh.
- Gotthard, L. (2023). Subject-verb agreement and the rise of do-support during the period of anglicisation of Scots. To appear in proceedings of ICEHL21, Leiden 2021.
- Jamieson, E. (2015). An investigation of verb raising in the Shetland dialect of Scots. Master's thesis, University of Edinburgh.
- Jonas, D. (2002). Residual V-to-I. In Lighfoot, D., editor, *Syntactic effects of morphological change*, pages 251–270. Oxford: Oxford University Press.
- Knieszsa, V. (1997). The origins of scots orthography. In Jones, C., editor, *The Edinburgh History of the Scots Language*, pages 24–46. Edinburgh: Edinburgh University Press.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language variation and change*, 1:199–244.
- Meurman-Solin, A. (1993a). Periphrastic do in sixteenth- and seventeenth century Scots. In Meurman-Solin, A., editor, *Variation and change in early Scottish prose*, pages 259–276. Helsinki: Suomalainen Tiedeakatemia.
- Meurman-Solin, A. (1993b). *Variation and change in early Scottish prose*. Helsinki: Suomalainen Tiedeakatemia.

- Nurmi, A. (2011). The rise and regulation of periphrastic do in negative declarative sentences: a sociolinguistic study. In Kastovsky, D. and Mettinger, A., editors, *The History of English in a Social Context: A Contribution to Historical Sociolinguistics*, pages 339–362. New York: de Gruyter Mouton.
- Pa-Tel, N. (2013). Contact or inheritance? criteria for distinguishing internal and external change in genetically related language. *Journal of Language Contact*, 6(2):313–28.
- Poplack, S. and Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In *Language and Space – An international handbook of linguistic variation: Volume 1 – Theories and methods*. Berlin: Mouton de Gruyter.
- Poussa, P. (1990). A contact-universals origin for periphrastic do with special consideration of old english-celtic contact. In Adamson, S., Law, V., Vincent, N., and Wright, S., editors, *Papers from the 5th international conference on English historical linguistics*, volume 65. Amsterdam/Philadelphia: John Benjamins Publishing.
- Robbeets, M. and Cuyckens, H. (2013). *Shared Grammaticalization: With Special Focus on Transeurasian Languages*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Smith, J. (2000). 'you ø na hear o' that kind o' things': Negative do in Buckie Scots. *English World Wide*, 21(2):231–259.
- Thomason, S. G. (2001). *Language Contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, S. G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tieken-Boon van Ostade, I. (1990). The origin and development of periphrastic auxiliary do: A case of destigmatisation. *NOWELE. North-Western European Language Evolution*, 16(1).
- van der Auwera, J. and Genee, I. (2002). English do: on the convergence of languages and linguists. *English Language and Linguistics*, 6:283–307.
- Warner, A. (2002). What drove do? In Kay, C., Horobin, S., and Smith, J., editors, *New perspectives on English Historical Linguistics*, volume I, pages 229–242. Amsterdam/Philadelphia: John Benjamins Publishing.

## Exploring language variation and change in the distant past.

The modelling of ‘prehistorical’ language features through comparative analyses and reconstructions offers a unique insight into language dynamics in the distant past (e.g., Dunn et al. 2011; Carling & Cathcart 2021). However, to analyse the particulars of the process of language change in its socio-historical embeddings, we are constrained to analyse written documentation. But not all records are suitable for this endeavour; historical sociolinguistic research has mostly focused on well-attested post-medieval European languages (e.g., Nevalainen & Raumolin-Brunberg 2003), and could be expected to find major limitations for ancient languages (Winter 1999). Then, how far back in time can we go to study variation and change in languages? What are the temporal limits for assessing the *uniformitarian principle* (Bergs 2012)?

While sociolinguistic variation has indeed been accounted for in the study of languages going back at least to classical antiquity (e.g., Adams 2013), this presentation explores the study of language variation and change ca. 4 millennia before present. Using corpora of Old Babylonian epistolary texts (cf. Hernáiz 2020), significant observations can be made regarding phonological and morphological variables in Akkadian, one of the earliest recorded languages. Two case studies will be discussed: the deaffrication of sibilants and gender syncretism in the demonstrative/pronominal paradigm.

Despite clear limitations, the characteristics of the continuously growing record of Akkadian provides documented data (i.e., not reconstructed) that illustrate the co-existence of variants of linguistic variables in communities of the distant past. The analysis of these variables offers a higher resolution image of language change in the context of socio-political fluctuations, population movements and contact between lectal varieties.

### Key words

Language variation and change, syncretism, deaffrication, ancient languages, Akkadian

### Bibliography

Adams, J. N. (2013). *Social variation and the Latin language*. Cambridge University Press.

Bergs, A. (2012). The Uniformitarian Principle and the risk of anachronisms in language and social history. In Hernández-Campoy, J. M. & Conde-Silvestre, J.C (eds). *The handbook of historical sociolinguistics*, 83-101.

Carling, G., & Cathcart, C. (2021). Reconstructing the Evolution of Indo-European Grammar. *Language*, 97(3), 561-598.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, 473(7345), 79-82.

Hernáiz, R. (2020). *Studies on linguistic and orthographic variation in Old Babylonian letters*. Ugarit-Verlag.

Nevalainen, T., & Raumolin-Brunberg, H. (2003). *Historical sociolinguistics: language change in Tudor and Stuart England*. Longman.

Winter, W. (1999). Sociolinguistics and dead languages. In Ernst Håkon Jahr (ed). *Language Change. Advances in Historical Sociolinguistics*, Mouton de Gruyter. 67-84.



## When change fails: evidence from French

**INTRODUCTION:** In this paper I focus on a series of pronouns found in French between the 12<sup>th</sup> and the 16<sup>th</sup> centuries. These pronouns pose a theoretical challenge as they bear a strong morphology (*moy, toy, soy*, etc.), which differentiates them from clitics, yet they appear in a derived position, like clitics. This construction only concerns infinitival complements as exemplified in (1) and (2).

- (1) *Vous auriez tort de moy faire desplaisir.*  
 you would.have wrong to me make.INF displeasure  
 “You would be wrong to make me unhappy.”
- (2) *et pour cuider qu’il deust soy appaisier.*  
 and for believe.INF that-he must REFL appease.INF  
 “and to believe that he had to appease himself.”

This construction has been described (Moignet 1970, Pearce 1990, de Kok 1993, Roberts 1997) but it is yet to receive a formal analysis. The sporadic occurrences of clitics in the immediate periphery of the infinitive are traditionally attributed to a preference for the use of pre-infinitival strong pronouns when clitic climbing is not possible (Moignet 1970, Robert 1997). This hypothesis is supported by the obligatoriness of clitic climbing until ca. 1600 (Martineau 1990). Nevertheless, recent research has shown that clitics could cliticise on infinitives from the earliest periods on (Olivier 2022), which challenges this view. The objectives of this paper are (a) to present a quantitative report of the evolution of such pronouns, (b) to characterise their nature, (c) to provide a formal analysis of the construction and (d) to account for their loss.

**METHODOLOGY:** I have created and manually tagged a corpus of legal texts from 1150 to 1856. The construction under focus is attested between the 12<sup>th</sup> and the 16<sup>th</sup> centuries, therefore I will limit the discussion to this timeframe. I counted 66 occurrences. This is extremely low compared to the number of enclisis and proclisis (Table 1); and it dismisses the traditional hypothesis that clitics are banned in this context.

ENCLISIS	PROCLISIS	CLITIC CLIMBING	PRE-INFINITIVAL PRONOUN
190	486	1,414	66

**Table 1:** Distribution of pronouns in the corpus (12<sup>th</sup>-16<sup>th</sup> c.)

In restructuring clauses, the pronoun is almost always a clitic and climbs. I have counted two occurrences where the pronoun does not climb and remains strong, see example (2). This construction is extremely rare. In non-restructuring clauses, there is no particular environment that appears to favour the presence of a pre-infinitival pronoun instead of a clitic.

**ANALYSIS:** DP objects can precede and follow infinitives in Old French (de Kok 1993: 261). The pronouns under focus are strictly pre-infinitival, which evidences that they appear in a derived position. I propose that their hybridity (they are not clitics but they appear in a derived position) can be explained alongside Cardinaletti & Starke’s (1999) tripartition, namely that they are Weak Pronouns (WP). Being phrases, WPs target a specifier, which I take to be that of vP. This construction is present for a limited amount of time (4 centuries) and in low quantities (Table 1). Put informally, it exposes a change that is not *successful* in the diachrony. Failed changes are reported cross-linguistically (Postma 2010, Meyer 2020, Ringe & Yang 2022), and they are cases where a new construction is either rapidly reanalysed, or where it is challenged by another construction. I will propose that the latter is true for French, with the rise of proclisis. Infinitival proclisis appears only a century after we first observe the construction under focus here, which rapidly overshadowed the innovative use of WPs in the 12<sup>th</sup> century. The two constructions are linearly identical and clitics were favoured during acquisition. Further, clitic climbing becomes optional in restructuring clauses towards the end of the 16<sup>th</sup> century (Olivier 2022) which substantially increases the acquirers’ exposure to proclisis. This pivotal moment connects to the loss of WPs in the corpus.



**SELECTED REFERENCES:**

- Cardinaletti, A., & Starke, M. (1999)** The typology of structural deficiency. *Clitics in the languages of Europe*, 145-233.
- Chomsky, N. (2004)** Beyond Explanatory Adequacy.
- Déchaine, R.M. & Wiltschko, M. (2002)** Decomposing pronouns. *Linguistic Inquiry* (33)3, 409-442.
- De Kok, A. (1985)** *La place du pronom personnel régime conjoint en français: une étude diachronique* (No. 23). Rodopi.
- de Kok, A. (1993)** Sur l'origine de l'emploi des formes toniques du pronom. In *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, New Brunswick, 14-18 August 1989* (Vol. 106, p. 249). John Benjamins Publishing.
- Hirschbühler, P., & Labelle, M. (1994)** L'évolution des propositions infinitives négatives en français. *Linx. Revue des linguistes de l'université Paris X Nanterre*, (5), 59-90.
- Hirschbühler, P., & Labelle, M. (2000)** Evolving Tobler-Mussafia effects in the placement of French clitics. *Amsterdam studies in the theory and history of linguistic science series 4*, 165-182.
- Labelle & Hirschbühler (2005)** Changes in clausal organization and the position of clitics in Old French. *Grammaticalization and parametric variation*, 60-71.
- Moignet, G. (1965). *Le Pronom personnel français: essai de psycho-systématique historique..* (Vol. 9). C. Klincksieck.
- Martineau, F. (1990)** *La montée du clitique en moyen français: une étude de la syntaxe des constructions infinitives*. University of Ottawa (Canada).
- Olivier, M. (2022)** *A Corpus Study of Clitic Placement with Infinitives in the Diachrony of French* (Doctoral dissertation, Ulster University).
- Olivier, M. (2022)** Diachronie de la proclise et de l'enclise avec l'infinitif en français médiéval (12e-15e siècles). *Studia linguistica romanica*, (8), 10-35.
- Pearce, E. H. (1990)** *Parameters in Old French Syntax: Infinitival Complements: Infinitival Complements* (Vol. 18). Springer Science & Business Media.
- Postma, G. (2010)** The impact of failed changes. *Continuity and change in grammar*, 269-302.
- Roberts, I. (1997)** Restructuring, head movement, and locality. *Linguistic inquiry*, 423-460.
- Roberts, I. (2010)** *Agreement and head movement: Clitics, incorporation, and defective goals* (Vol. 59). MIT Press.
- Yang, C. (2016)** *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

## Usage-based evolutionary models reveal context-specific word order change in Indo-European

In the early days of Indo-European studies, Schleicher (1868) published his famous reconstruction of a fable in Proto-Indo-European to demonstrate that is possible to reconstruct all aspects of a language, from lexicon to morphology, domains where the comparative method has been applied widely and successfully, but also syntax. After the foundational work of Delbrück (1893-1900) and others (Wackernagel 1892; Brugmann 1925), the interest in syntactic reconstruction ceased for a couple of decades. In the 1970s, both Lehmann (1974) and Friedrich (1975) proposed reconstructions of basic word order patterns in Proto-Indo-European following the Greenbergian framework of conditional word order universals. Recently, computational phylogenetic methods have been applied to model the diachronic dynamics and reconstruct syntactic traits and other grammatical features (Greenhill et al. 2010, 2017; Dunn et al. 2011; Carling and Cathcart 2021, and more).

Proto-Indo-European word order was very likely flexible to some extent and allowed non-basic word order for emphasis and to mark information-structural properties (Viti 2014; Lühr 2015). Studies of documented word order changes highlight the importance of synchronic variation as a precondition for change (England 1991; Harris and Campbell 1995; Ross 2007; Heine 2008). Therefore, we propose a new approach to infer the evolutionary dynamics of word order under different pragmatic conditions. Instead of coding word order as an abstract type, we take observed instances in specific pragmatic contexts as a starting point.

To control for pragmatics and information-structure, we extracted a set of 46 sentences in 36 modern Indo-European languages from a parallel corpus (Levshina 2016, with additional data collected by the authors to enhance the coverage of Indo-European languages). Our sentence sample includes different types of subjects and objects (pronouns, nouns, and object clauses) to cover a wide range of constructions that are common in naturalistic speech. We use Bayesian phylogenetic comparative methods to infer transition rates between the states of a binary feature that encodes the order of object and verb. The follow-up analysis investigates whether some of these contexts are more prone to change than others.

The long-term probability of being in one state or the other varies between sentences, with some sentences having a higher probability for OV, while others have a higher probability for VO. This suggests the co-existence of different word order patterns in Proto-Indo-European.

By applying k-means clustering on the mean posterior rates, we identified sets of sentences that evolve in a similar way: one cluster contains verbs of speech and mental verbs with complement clauses which are almost exclusively VO in all modern Indo-European languages, even in those with basic OV order. A second cluster encompasses verbs with object pronouns which tend to precede the verb in many languages of the Romance and Slavic branch that otherwise prefer post-verbal nominal objects. The last cluster contains mostly nominal objects.

Our study did not provide evidence for the initial hypothesis that pragmatic factors lead to more variation and therefore faster rates of change. It is likely that our sentence sample was too small to have sufficient data for word order variation conditioned by pragmatic factors. Specific constructions that share semantic and structural properties can still be identified based on their distinct rates of change. This suggest that these factors play a major role in the evolution of word order.

## References

- Brugmann, Karl. 1925. *Die Syntax Des Einfachen Satzes Im Indogermanischen*. Berlin; Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111586304>.
- Carling, Gerd, and Chundra Cathcart. 2021. “Reconstructing the Evolution of Indo-European Grammar.” *Language* 97 (3): lan.0.0253. <https://doi.org/10.1353/lan.0.0253>.
- Delbrück, Berthold. 1893-1900. *Vergleichende Syntax Der Indogermanischen Sprachen*. Grundriss Der Vergleichenden Grammatik Der Indogermanischen Sprachen 3–5. Strassburg: Karl J. Trübner.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russel D. Gray. 2011. “Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals.” *Nature* 473: 79–82.
- England, Nora C. 1991. “Changes in Basic Word Order in Mayan Languages.” *International Journal of American Linguistics* 57 (4): 446–86.
- Friedrich, Paul. 1975. *Proto-Indo-European Syntax: The Order of Meaningful Elements*. Journal of Indo-European Studies. Butte: Montana College of Mineral Sciences.
- Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray. 2010. “The Shape and Tempo of Language Evolution.” *Proceedings of the Royal Society B: Biological Sciences* 277 (1693): 2443–50.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. 2017. “Evolutionary Dynamics of Language Systems.” *Proceedings of the National Academy of Sciences of the United States of America* 114: E8822–29.
- Harris, Alice C., and Lyle Campbell. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge Studies in Linguistics 74. Cambridge: Cambridge University Press.
- Heine, Bernd. 2008. “Contact-Induced Word Order Change Without Word Order Change.” In *Language Contact and Contact Languages*, edited by Peter Siemund and Noemi Kintana, 33–60. Amsterdam; Philadelphia: Benjamins.
- Lehmann, Winfred P. 1974. *Proto-Indo-European Syntax*. Austin; London: University of Texas Press.
- Levshina, Nathalia. 2016. “Verbs of Letting in Germanic and Romance: A Quantitative Investigation Based on a Parallel Corpus of Film Subtitles.” *Languages in Contrast* 16 (1): 84–117. <https://doi.org/10.1075/lic.16.1.04lev>.
- Lühr, Rosemarie. 2015. “Traces of Discourse Configurability in Older Indo-European Languages?” In *Perspectives on Historical Syntax*, edited by Carlotta Viti, 203–32. Studies in Language Companion Series 169. Amsterdam & Philadelphia: Benjamins.
- Ross, Malcolm. 2007. “Calquing and Metatypy.” *Journal of Language Contact* 1: 116–43. <https://doi.org/10.1163/000000007792548341>.
- Schleicher, August. 1868. “Eine Fabel in Indogermanischer Ursprache.” *Beiträge Zur Vergleichenden Sprachforschung Auf Dem Gebiete Der Arischen, Celtischen Und Slawischen Sprachen* 5 (2): 206–8. <http://www.jstor.org/stable/23458804>.
- Viti, Carlotta. 2014. “Reconstructing Syntactic Variation in Proto-Indo-European.” *Indo-European Linguistics* 2 (1): 73–111. <https://doi.org/10.1163/22125892-00201004>.
- Wackernagel, Jakob. 1892. “Über Ein Gesetz Der Indogermanischen Wortstellung.” *Indogermanische Forschungen* 1: 333–436.



## References

- Berthele, Raphael, Matthew Whelpton, Åshild Næss & Pieter Duijff. 2015. Static spatial descriptions in five Germanic languages. *Language Science* 49:82–101.
- Friðjónsson, Jón G. 2005. Kerfisbundnar breytingar á notkun nokkurra forsetninga í íslensku [Changes in the use of some prepositions in Icelandic]. *Íslenskt mál* 27:7–40.
- Grestenberger, Laura. 2015. Number marking in German measure phrases and the structure of pseudo-partitives. *The Journal of Comparative Germanic Linguistics* 18:93–138.
- Jónsson, Jóhannes Gísli. 2017. Avoiding genitive in Icelandic. In Höskuldur Thráinsson et al. *Syntactic variation in Insular Scandinavian*, 141–163. Amsterdam: John Benjamins.
- Kjartansson, Helgi Skúli. 1979. Eignarfallsflótti: Uppástunga um nýja málvillu [Genitive-Avoidance: A proposal on a new linguistic error]. *Íslenskt mál* 1:88–95.
- ONP = *Ordbog over ded norrøne prosasprog*. Copenhagen: Den Arnamagnæanske Kommission (eds). Access at: <https://onp.ku.dk/onp/onp.php>
- RMH = Risamálheild [Gigaword Corpus]: Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. *Proceedings of LREC 2018*, 4361–4366, Japan. Accessed at: <https://malheildir.arnastofnun.is/?mode=rmh2022>
- ROH = *Ritmalssafn Orðabókar Háskólans*. Stofnun Árna Magnússonar í íslenskum fræðum (eds.). Access at: <http://ritmalssafn.arnastofnun.is/>
- Rögnvaldsson, Eiríkur. 2021. *Hliðiná*. A blog post on topics in Icelandic language use. Published January 9th, 2021. Accessed at: <https://uni.hi.is/eirikur/2021/01/09/hlidina/>

Differential Place Marking and the reconstruction of the Proto-Nakh system of spatial cases

Within the East Caucasian language family, the Nakh branch (often assumed to be the first to have split off from the protolanguage) has often been described as typologically dissimilar to all other languages of the family (collectively known as Daghestanian) with respect to its marking of spatial relations. Where Daghestanian languages of various branches typically display two-slot systems, with one suffix marking location (in, on, under, through, at, near) and another marking direction (allative, ablative, essive) (Daniel & Ganenkov 2012), the Nakh languages have been analysed as having fewer, monomorphemic spatial cases and as making use of more postpositions (e.g. Nichols 2011).

Two recent advancements have shed new light on the Nakh data. Firstly, reanalysis of the Tsova-Tush data (the third Nakh language besides Chechen and Ingush) has allowed the recognition of a Daghestanian-style two-slot system of spatial cases, (Author, forthc.). See Table 1 for 12 of the 33 (combinations of) spatial cases in Tsova-Tush. Secondly, the Tsova-Tush data displays clear features of Differential Place Marking. The notion of Differential Place Marking identifies splits in the coding of locative, allative or ablative roles depending on subclasses of nouns, in particular place names (toponyms), inanimate common nouns and human nouns (Haspelmath 2019).

	Goal	Location	Source	Compatible nouns
'near, at'	-go	-go-ħ	-go-ren	animates
'among, in'	-lo	-lo-ħ	-lo-ren	liquids, masses, collections
'in'	-i	-i-ħ	-i-ren	rooms, buildings, containers, place names
Default	-∅	-ħ	-ren	other

Table 1: Tsova-Tush spatial cases (excerpt)

Based on these findings, two questions arise:

- Should a two-slot system be reconstructed for Proto-Nakh, and if so, are the morphemes cognate with those found in Daghestanian languages?
- Is Differential Place Marking an innovation in Tsova-Tush, or is it inherited from Proto-Nakh?

This paper has three goals. (1) It puts the Tsova-Tush data in typological and areal perspective, and concludes that it shows clear parallels with the Daghestanian-style system, and furthermore that it obeys the typological universals concerning Differential Place Marking (i.e. there is less phonetic material in spatial suffixes on place names compared to other nouns, there is more phonetic material in spatial suffixes on animate nouns, and more phonetic material in non-spatial cases on place names). (2) It re-evaluates the Chechen and Ingush data, concluding that they show clear traces of a former two-slot system with many cognate morphemes (see Chechen comparative case *-l*, allative case *-ie/-ga/-a*); (3) It reconstructs the spatial case system for Proto-Nakh, with clear cognates to Daghestanian languages (as reconstructed by Alekseev (1997)), but concludes that the Differential Place Marking features of Tsova-Tush are secondary.

By answering the above questions, this paper aims to be an important case-study in (1) the grammaticalisation of locative markers; (2) the history and internal branching of the East Caucasian family, where often cognate sets of morphemes are established (e.g. Desheriev 1963:436), but attempts at reconstruction are few and far between; and (3) the recent topic of Differential Place Marking in Caucasian languages, where related phenomena such as Differential Subject Marking have been observed previously (see Arkadiev 2017).

Alekseev, Mixail. 1999. "Reconstruction of the Proto-East Caucasian locative morphemes". In: van den Berg, Helma (ed.) *Studies in Caucasian linguistics: selected papers of the eighth Caucasian colloquium*. Leiden: Research School of Asian, African and Amerindian Studies (CNWS), p. 116-124.

Arkadiev, Peter. 2017. Multiple ergatives. From allomorphy to differential agent marking. *Studies in Language* 41(3): 717-780. <https://doi.org/10.1075/sl.41.3.06ark>

Daniel, Michael & Ganenkov, Dmitry. 2008. "Case Marking in Daghestanian: Limits of Elaboration", in: Malchukov, Andrej L. & Spencer, Andrew (eds) *The Oxford Handbook of Case*. Oxford Academic. p. 668-685. <https://doi.org/10.1093/oxfordhb/9780199206476.013.0047>

Desheriev, Yunus. 1963. *Sravnitel'no-istoričeskaja grammatika naxskix jazykov [Historical comparative grammar of the Nakh languages]*. Grozny: Čečeno-ingušskoe knižnoe izdatel'stvo.

Haspelmath, Martin. 2019. Differential place marking and differential object marking. *STUF* 72(3):313-334. <https://doi.org/10.1515/stuf-2019-0013>

Nichols, Johanna, 2011. *Ingush grammar*. University of California Press

Author (forthcoming). *Intensive language contact in the Caucasus: the case of Tsova-Tush*.



### Contact and the origins of headed *wh*-relatives in Hungarian

Headed *wh*-relatives are a feature of Standard Average European (Haspelmath, 1998, 2001). De Vries (2002) found that 40% of Indo-European (IE) languages had such relative clauses, but only 2.3% of the non-IE languages in his sample did. The non-IE languages that do include Hungarian, Finnish, and Georgian; (1) is an example from Hungarian, from (Comrie 1998: 60).

- (1) A      fiú,      akit      láttam  
 The    boy    who.ACC    I saw      ‘the boy I saw’

Comrie (1998) and Hendery (2012) explain the cross-linguistic distribution of headed *wh*-relatives in terms of contact. However, the precise mechanisms of this contact-induced distribution are obscure. In particular it is hard to learn functional vocabulary like *which*: there are fluid pairings between category and denotation which give rise to persistent ambiguity. This underspecified nature of function-word meaning makes change likely, but direct borrowing difficult. Instead, it seems probable that Hungarian developed a precursor of headed *wh*-relatives through contact. Headed *wh*-relatives then emerged as Hungarian followed a recurring pathway found extensively in IE languages.

In IE languages, the Proto-Indo-European indefinite/interrogative pronouns *\*kwi-/kwo-*, which did not head relative clauses, are the source of IE *wh*-relative forms. There is a pathway from conditional to correlative (Belyaev & Haug, 2014, 2020) and on to headed relative (Haudry, 1973) with multiple possible pathways through the semantic space (Gisborne & Truswell, 2018). Conditional-correlative constructions, formed around indefinite-interrogative pronouns (a class of words found in interrogatives, conditionals, and other dependent contexts, Haspelmath, 1997), are therefore the source of headed *wh*-relatives in the IE daughter languages, where they have developed through parallel evolution. Word-order conditions the change. Conditional protases are topics (Haiman, 1978) and the indefinite/interrogative pronoun is focused. Kiparsky (1995) argues that early Indo-European (Vedic, Hittite, Greek) had the structure (TOPIC)-(FOCUS)-Clause: the emergence of conditional-correlatives involves topicalization of the conditional clause and, typically, fronting of the indefinite/interrogative pronoun.

Examples of conditional-correlatives are found in modern Hungarian: (2) is from Lipták (2009: 27).

- (2) Amelyik      kutya közel jön hozzám,      azt      elkergetem  
 REL.which    dog    close    comes to.me      that.ACC      chase.away

‘Which(ever) dog comes close to me, I’ll chase it away’ = ‘If a dog comes, close to me I’ll chase it away’

However, structures like (2) were not possible in proto-Hungarian, which Kiss (2013) reconstructs as strict SOV/head-final, with grammaticalized discourse roles for S and O. S is always the (primary) topic and O the focus or secondary topic. For patterns like (2) to emerge in Hungarian, a word-order change was necessary. Kiss (2013) argues that the SOV of proto-Hungarian developed into (TOPIC)-(FOCUS)-V-X\* in Old Hungarian, giving Hungarian a word-order type consistent with the early IE languages that developed conditional correlatives.

Contact-driven change depends on both the complexities of the sociolinguistic context, and on what can plausibly be transferred from one language to another in bilinguals and bilingual use. Some contact-driven changes seem more plausible than others. As noted above, functional vocabulary is hard to learn. Similar abstract forms to *wh*-relatives appear resistant to borrowing: Sorbian has co-opted its demonstratives to certain functions of definite marking without having all the properties of definite articles (Heine & Kuteva 2005: 71-73). And yet, contact induces word-order change (Heine & Kuteva, 2005). We argue that Kiss’ word-order change provides the relevant context for the development of a headed *wh*-relative in Hungarian. It is also possible that Hungarian borrowed left-adjoined conditional-correlatives formed on indefinite-interrogative pronouns. In either case, given the necessary pre-conditions, the potential pathway of change exists, making it possible for *wh*-relatives to emerge in Hungarian without having to borrow abstract, underspecified forms. The account developed here allows us to develop a plausible understanding of the role of contact in the diffusion of areal phenomena by understanding the context in which ‘replica’ development (Heine and Kuteva, 2005) is enabled.

**References**

- Belyaev, O., & Haug, D. (2014). *The genesis of wh-Based correlatives: From indefiniteness to relativization*.
- Belyaev, O., & Haug, D. (2020). The genesis and typology of correlatives. *Language*, 96(4), 874–907. <https://doi.org/10.1353/lan.2020.0065>
- Comrie, B. (1998). Rethinking the typology of relative clauses. *Language Design*, 1, 59–86.
- De Vries, M. (2002). *The syntax of relativization*. Universiteit van Amsterdam.
- Gisborne, N., & Truswell, R. (2018). *Parallel evolution of relative clauses in Indo-European*.
- Haiman, J. (1978). Conditionals are topics. *Language*, 54(3), 564–589.
- Haspelmath, M. (1997). *Indefinite pronouns*. Oxford University Press.
- Haspelmath, M. (1998). How young is standard average European. *Language Sciences*, 20(3), 271–287.
- Haspelmath, M. (2001). The European linguistic area: Standard Average European. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 1492–1510). De Gruyter.
- Haudry, J. (1973). Parataxe, hypotaxe et corrélation dans la phrase latine. *Bulletin de La Société Linguistique de Paris*, 68, 147–186.
- Heine, B., & Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge University Press.
- Hendery, R. (2012). *Relative clauses in time and space: A case study in the methods of diachronic typology*. John Benjamins.
- Kiparsky, P. (1995). Indo-European origins of Germanic syntax. In A. Battye & I. Roberts (Eds.), *Clause structure and language change* (pp. 140–169). Oxford University Press.
- Kiss, K. É. (2013). From Proto-Hungarian SOV to Old Hungarian Top Foc V X\*. *Diachronica*, 30(2), 202–231.
- Lipták, A. (2009). The landscape of correlatives: An empirical and analytical survey. In *Correlatives cross-linguistically* (pp. 1–46). John Benjamins Publishing Company.

**Areal alignment and the loss of ATR harmony in Riverine Bua languages (Chad)**

Several linguistic areas have been identified in Africa, one of the most discussed in recent years being the Macro-Sudan Belt (MSB; Güldemann, 2008, 2010, 2018; Clements and Rialland 2008). Areas of intermediary size have also been identified (e.g. Central Africa within the MSB, cf. Dryer 2009, Idiatov 2018, Güldemann 2018b: 457; Rolle, Lionnet & Faytak 2020, a.o.). Such linguistic areas suggest a form of areal pressure, i.e., languages have a strong tendency to adapt their linguistic profile to the area where they are spoken, or where their speakers migrate into (cf. Güldemann 2020, Rolle et al. 2020 for examples of such changes).

The goal of this talk is to illustrate the inner workings of areal alignment, with a detailed case study of Bua languages (southern Chad), focusing on the changes that affect the vowel systems. Bua languages form a tight family of 9 languages, traditionally classified in the “Adamawa” grouping within Niger-Congo. The family is divided into two markedly different branches: Riverine languages (Lua, Ba, Tun) and Inland languages. Inland languages have a robust ATR contrast and harmony and no interior (i.e. central and other non-peripheral) vowels, while Riverine languages have no ATR contrast or harmony, and have interior (mostly central and front rounded) vowels. This difference is illustrated with the inventories of most Inland languages (and Proto-Bua, as I will show) in (1), and Lua (Riverine) in (3).

(1) Proto-Bua (and most Inland)	(2) Loss of ATR	(3) Central vowels > Lua (Boyeldieu 1985)
ATR: + *i      *u	*i      *u	i   ị   u <i>high</i>
– *ɪ      *ʊ		
+ *e      *o	*e      *o	e   ə̣   o <i>mid</i>
– *ɛ      *ɔ	*ɛ   *a   *ɔ	ɛ   a   ɔ <i>low</i>
–      *a		

Based on a comparative Bua database of about 600 terms (Author et al. 2023), I show that:

1) Proto-Bua had a 9-vowel system very close to that of Inland languages, with robust ATR contrast and harmony – a 2IU system in Casali’s (2008) typology (step (1) above).

2) Riverine languages are doubly innovative: firstly, they lost the ATR contrast by merging the [-ATR] high vowels \*ɪ and \*ʊ with the [+ATR] mid vowels /e/ and /o/ respectively, and reinterpreted ATR harmony as height harmony in a rectangular vowel system (step (2) above), and secondly, they independently innovated central vowels (step (3)). This change confirms the tendency for ATR attrition to be caused by the loss of [-ATR] high vowels [ɪ ʊ], which are known to be perceptually confusable with either [i u] or [e o] (Casali 2003: 342; Rose 2018, a.o.). The confusability with [e o] was the phonetic precursor to the merger that took place in Riverine languages – a confusability that is still noticeable in present-day Inland languages, as I will show.

3) These changes are the result of areal alignment: the resulting vowel systems of Riverine languages are almost identical to those of neighboring languages: Laal (isolate), East Chadic, and Sara-Bongo-Bagirmi (Central Sudanic) languages. Conversely, the Inland languages (and by extension proto-Bua) have a general phonological profile that is unexpected in the area.

I will conclude with preliminary remarks and questions about (i) the relationship between ATR and interior vowels in the MSB; (ii) the sociolinguistic underpinnings of areal alignment – notably the crucial role of small-scale multilingualism (Lüpke 2016; DiCarlo et al. 2019); and (iii) what these diachronic changes tell us about the linguistic history of the region, notably about the time depth of the Central African ATR-deficient / Interior Vowel zone (Rolle et al. 2020), which the speakers of the distant ancestor of Bua languages (likely to have had ATR harmony) must have crossed in a distant past on their way from Nigeria to their current location.

## **References:**

- Author et al. 2023. [redacted for review purposes]
- Boyeldieu, Pascal. 1985. *La Langue lua ('niellim'), groupe boua, Moyen-Chari, Tchad: phonologie, morphologie, dérivation verbale*. (Descriptions de langues et monographies ethnolinguistiques, 1.) Paris: SELAF.
- Boyeldieu, Pascal, Raimund Kastenholtz, Ulrich Kleinewillinghöfer, and Florian Lionnet. 2018. The Bua Group languages (Chad, Adamawa 13): A comparative perspective. In Rajja Kramer and Roland Kießling (eds.), *Current Approaches to Adamawa and Gur Languages*, Afrika und Übersee Supplements, Vol. 34. Cologne: Köppe. Pp. 53-126.
- Casali, Roderic F. 2003. [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7:307-382.
- Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2. 496-549. <http://onlinelibrary.wiley.com/doi/10.1111/j.1749-818X.2008.00064.x/abstract>.
- Di Carlo, Pierpaolo, Jeff Good, and Rachel A. Ojong Diba. 2019. Multilingualism in rural Africa. *Oxford Research Encyclopedia of Linguistics*. Retrieved 2 Aug. 2022, from <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-227>.
- Dryer, Matthew S. 2009. Verb-object-negative order in central Africa. In Norbert Cyffer, Erwin Ebermann & Georg Ziegelmeyer (eds.), *Negation patterns in West African languages and beyond*. Typological Studies in Language 87, 307–362. Amsterdam: John Benjamins.
- Güldemann, Tom. 2008. The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In Heine, Bernd and Derek Nurse (eds.), *A linguistic geography of Africa*. Cambridge: Cambridge University Press, 151-185
- Güldemann, Tom. 2010. Sprachraum and geography: linguistic macro-areas in Africa. In Lameli, Alfred, Roland Kehrein and Stefan Rabanus (eds.), *Language and space: an international handbook of linguistic variation*, volume 2: language mapping. Handbooks of Linguistics and Communication Science 30,2. Berlin: Mouton de Gruyter, 561-585, Maps 2901-2914.
- Güldemann, Tom. 2018. Language contact and areal linguistics in Africa: introduction. In Güldemann, Tom (ed.), *The languages and linguistics of Africa*. The World of Linguistics 11. Berlin: Mouton de Gruyter, 445-446.
- Güldemann, Tom. 2020. Changing profile when encroaching on hunter-gatherer territory: towards a history of the Khoe-Kwadi family in southern Africa. In Güldemann, Tom, Patrick McConvell and Richard Rhodes (eds.), *The languages of Hunter-Gatherers*, 114-146. Cambridge: Cambridge University Press.
- Idiatov, Dmitry 2018. An areal typology of clause-final negation in Africa: Language dynamics in space and time. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 115–163. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110607963-005.
- Lüpke, Friederike. 2016. Uncovering small-scale multilingualism. *Critical Multilingualism Studies* 4(2): 35-74.
- Rolle, Nicholas, Florian Lionnet, and Matthew Faytak, 2020. Areal patterns in the vowel systems of the Macro-Sudan Belt. *Linguistic Typology* 24(1): 113-179. DOI: 10.1515/lingty-2019-0028
- Rose, Sharon. 2018. ATR Vowel Harmony: new patterns and diagnostics. *Proceedings of the 2017 Annual Meeting on Phonology*. DOI: 10.3765/amp.v5i0.4254.

## Morphosyntactic variation in Swahili: Tracing descriptions past and present

Swahili is a Bantu language spoken by some 100 million people across East Africa. In addition to its high speaker numbers and geographic spread, Swahili is one of the few African languages for which we have written materials from over a century ago. There are Swahili manuscripts dating from the 18<sup>th</sup> century – many of which record text and language which is likely to be older. Interest in this written tradition in part gave rise to descriptions of Swahili by European linguists, missionaries and grammarians towards the end of the 19<sup>th</sup> century. Another key feature of the descriptive accounts of Swahili from this period was the interest in (dialectal) variation in the language, with Swahili spoken along the coast from Southern Somalia to northern Mozambique, on a number of islands and into the ‘mainland interior’ of East Africa.

Many of these early accounts focused on lexical and phonological differences. However, there are also descriptions of grammatical differences on which the present talk focuses. We revisit a number of these earlier sources with a view to examining the variation described, compare these accounts to present-day variation, and to develop a more comprehensive picture of morphosyntactic variation and change in Swahili (cf. Nurse and Hinnebusch 1993). Morphosyntactic variation remains an under-examined aspect of Swahili, and the talk aims to show how the study of this variation in the early sources contributes to a better understanding of the interacting processes of language change, dialect contact and convergence, and effects of standardisation in the history of Swahili.

Early materials are taken from a number of key sources. We consider the data in Steere’s (1870) seminal *Handbook of the Swahili Language* which is based primarily on the Zanzibari dialect Kiunguja - the dialect on which ‘Standard Swahili’ was based and developed in the twentieth century. We also examine also Velten’s (1901) *Safari za Wasuahili* ‘Travels of the Swahili people’, which is a key reference for this time period, Sacleux (1909) which draws on descriptions of ten Swahili dialects (although focusing on Kiunguja), and Stigand’s (1915) comparative study which provides a more explicit account of dialectal variation in Swahili, primarily in phonology and the lexicon. Finally we consider a set of more recent sources in Lambert (1957, 1958a,b) which examine the Chijomvu, Kingare, Chifundi and Vumba dialects of the southern Kenyan coast. We examine the following domains of variation in these sources:

- Verbal morphology – e.g. subject and object agreement, plural marking; behaviour of monosyllabic verb stems such as the retention of infinitive *ku-* prefix
- Nominal domain and nominal dependents - e.g. variation in class 1/2 prefixes *m-* ~ *mu-*; variation in class 1/2 possessive agreement *y-ake* ~ *w-ake*, noun class membership, personal pronouns *mimi* ~ *mie*;
- Tense-aspect-mood distinctions – e.g. near and distant past tense *-li-/-liki-*;
- Negation strategies – e.g. the use of the negative word *hapana*
- Locative formation – e.g. noun class prefixes, *vahali* ~ *mahali*, class 16 concord *va-*

We use these sources and examples of variation to contribute to the discussion on the historical development of Swahili, and pathways of grammatical change more broadly, including in contexts of high linguistic diversity, language contact and multilingualism. The talk aims to provide a better understanding of morphosyntactic variation in Swahili, Swahili dialects, and how the dialects have changed over time. We show that many of the areas in which this earlier variation was described are similar to those found in present day Swahili, or in Bantu languages more widely. We also pay attention to issues of authorship and voice, as well as how processes of standardisation have impacted on the Swahili seen in the present day.

## References

- Lambert, H. E. 1957. *Ki-Vumba, a dialect of the southern Kenya coast*. (Studies in Swahili Dialect, II.) Kampala: East African Swahili Committee.
- Lambert, H. E. 1958a. *Chi-Jomvu and ki-Ngare: Sub-dialects of the Mombasa Area*. (Studies in Swahili Dialect 3). Kampala: East African Swahili Committee.
- Lambert, H. E. 1958b. *Chi-Chifundi: a Dialect of the Southern Kenya coast*. (Studies in Swahili Dialect, V.) Kampala: East African Swahili Committee.
- Nurse, Derek and Thomas J. Hinnebusch. 1993. *Swahili and Sabaki: A Linguistic History*. University of California Press.
- Sacleux, C. 1909. *Grammaire des dialectes swahilis*. Paris: R. P. du Saint Esprit
- Steele, Edward. 1919. *A handbook of the Swahili language as spoken at Zanzibar*. London: Society for promotion Christian knowledge.
- Stigand, Chauncy Hugh. 1915. *Dialect in Swahili: A Grammar of Dialectic Changes in the Kiswahili Language*. Cambridge University Press. Cambridge
- Velten, Carl. 1901. *Safari za Wasuaheli*. Gottingen: Vandenhoeck & Ruprecht.



### Outliers in variation and change: atypical users of the variants of negation in Old and Middle Hungarian

According to Weinreich et al. (1968: 188)’s classic observation, „Not all variability and heterogeneity in language structure involves change; but all change involves variability and heterogeneity.” The present paper aims to discuss the history of sentence negation in Hungarian, with a focus on variability and heterogeneity that **precede** change. This variation can be investigated in negative sentences that contain Verb Modifiers (VM); the basic features of the variants are summarized in the table below. The coexistence of the variants can be broken down into three periods: a) stable variation with the predominance of the A-variant (from the first written documents until the 19<sup>th</sup> century); b) the radical frequency growth of the M-variant (during the 19<sup>th</sup> century); c) stable variation with the predominance of the M-variant (since the 20<sup>th</sup> century).

word order	VM – NEG – V	NEG – V – VM
<b>example</b>	<i>el nem megyek</i> away <sub>[VM]</sub> not go.1sg 'I am not going away'	<i>nem megyek el</i> not go.1sg away <sub>[VM]</sub> 'I am not going away'
<b>structure</b> (É. Kiss 2014)	Negative particle: adjoined to the verb in a lower structural position (PredP); verb modifier: precedes the negated verb	Negative particle: merged into a higher structural position (NegP), eliciting verb movement; verb modifier: follows the negated verb
<b>shorthand</b>	<b>A</b> (djunction-based)- <b>variant</b>	<b>M</b> (ovement-based)- <b>variant</b>

The period that precedes change is investigated based on three groups of texts: a) Old Hungarian codices; b) Old and Middle Hungarian informal texts; c) 16<sup>th</sup> century (=early Middle Hungarian) translations of the New Testament. The overall distribution of the two variants seems to be fairly similar in the three groups, but each group features an outlier (see the table below). The paper addresses the questions 1. whether the outliers differ in the same way from the majority, or they are different from each other as well; 2. whether these can be seen as the forerunners of the 19<sup>th</sup>-century change.

group	Old Hungarian codices	16 <sup>th</sup> -century translations of the New Testament	Old and Middle Hungarian informal texts
<b>rate of the M-variant</b>	22.8% (N=3472)	24.3% (N=745)	15.4% (N=3125)
<b>outlier</b>	the Hussite codices (15th century)	Károlyi's translation (1590)	Kolozsvár witch trials (1572-1592)
<b>M-var. in the outlier</b>	71.7%	43.3%	55.2%
<b>source of data</b>	<a href="http://omagyarkorpusz.nytud.hu/en-intro.html">http://omagyarkorpusz.nytud.hu/en-intro.html</a>		<a href="https://tmk.nytud.hu/">https://tmk.nytud.hu/</a>

The tentative answers to the questions can be summarized as follows:

- a) The Hussite subgroup may represent an innovative dialect, being on the way to generalizing the use of the M-variant. However, this dialect disappeared. b) The more frequent use of the M-variant in Károlyi's translation may be Károlyi's idiolectal trait. Still, it cannot be excluded that as a highly influential text, it had a (long-term, indirect) impact on the choice of patterns. c) Kolozsvár subgroup: although the unmarked word order is VM – V in neutral sentences, this group displays a persistent use of the V – VM pattern in this context. The reasons for this are not fully understood yet, but it might be the case that the choice between the A- and the M-variant is also influenced by a factor that is operative in neutral sentences as well, that is, independent of negation.
2. There is one important feature all these outliers share both with each other and with the rest of the sources: if clause types that are attested frequently enough are ordered on a scale based on the frequency of the M-pattern, these scales are almost the same: reason / main > complement > relative > conditional (> *until*-clauses). This corroborates the suggestion (Author 2017, 2022) that the variants differed in their pragmatic load in the first period of stable variation, the M variant being more emphatic; the outliers in Old and Middle Hungarian were early overusers of the emphatic variant, but this did not lead to a change in the entire population of speakers at that time.



## References

É Kiss, Katalin 2014. The evolution of functional left peripheries in the Hungarian sentence. In *The evolution of functional left peripheries in Hungarian syntax*, ed. Katalin É. Kiss, 9–55. Oxford: OUP.

Weinreich, Uriel & Labov, William & Herzog, Marvin I. 1968. Empirical foundations for a theory of language change. In: *Directions for historical linguistics*, eds Winfred P. Lehmann & Yakov Malkiel 95–195. Austin: University of Texas Press.

---

Author 2017. A magyar tagadó mondatok szórendje és a konstansráta-hipotézis [Word order in Hungarian negative sentences and the Constant Rate Hypothesis]. In: *Nyelvelmélet és diakronia 3.*, eds Katalin É. Kiss & Attila Hegedűs & Lilla Pintér, 91–110. Budapest – Piliscsaba: PPKE BTK.

Author 2022. Contrast maintenance in syntax: a highly atypical case of the Negative Cycle. [Manuscript]

## **Towards a new classification of Western Bantu languages using non-lexical data**

The present contribution aims at revisiting Nurse and Philippson's 2003 paper wherein they propose a new classification of Bantu languages based on the study of 30 phonological and morphological criteria. According to the authors, classifications based solely on the study of the lexicon can lead to errors because of the imperfect notion of similarity (cognate) used to classify languages, possible borrowings and the rapid evolution of vocabulary. Therefore, Nurse and Philippson proposed the first non-lexically based historical classification of 80 Bantu languages across the entire family. While Nurse and Philippson's approach and methodology were innovative, they encountered some problems: (i) the 80 Bantu languages selected for their study inadequately covered the North-Western area with only 8 languages spoken in zone A, 1 in zone B and 4 spoken in zone C, (ii) the available data lacked both in quantity and quality, and (iii) the study lacked the use of advanced classificatory techniques.

In order to address these problems, we propose here to work on a new sample of languages and on the selection of new non-lexical data. For this paper, we have decided to focus first on the study of Western Bantu languages that are spoken in Cameroon, Gabon, Equatorial Guinea (EG), Congo, DRC, Angola, Namibia, Zambia and Botswana by building a database containing approximately 100 languages from zones A, B, C, D, K, R, H, L. We selected morphological criteria from both the nominal and verbal domains as well as phonological criteria to be used to classify these languages. Such criteria include but are not limited to the presence/absence of gender categories, patterns of syncretism within nominal paradigms, singular/plural correspondences, verbal derivation, tense-aspect morphology, and sound changes.

Brown et al. (2023) tests the historical informativity of these criteria focusing solely on 32 languages in Zone A and B spoken in Cameroon, Equatorial Guinea and Gabon with promising results (i.e., Northwest Bantu). The classification produced therein consists of strong genetic groupings that correspond to what is found in the lexically based classifications in Grollemund 2012 and Grollemund et al. 2015—namely among A70-80-90 languages and the West Coastal languages (B20-50-60 in our sample). Even more revealing was the emergence of larger grouping containing languages from Zone A10-20-30 and B10-30. The majority of these languages are spoken along the coast of Cameroon, EG and Gabon. It is therefore possible that the non-lexical criteria considered for this study have uncovered a contact relationship among these languages that the lexical data failed to show.

The resulting phylogenetic classification of the Western Bantu languages based on non-lexical parameters has also proven to be enlightening. For example, a clear division between the Forest Bantu languages (Zones A, B, C, parts of H and parts of D10-20-30) and the rest of the sample emerges when looking at the expression of perfectivity. Forest languages almost exclusively exploit the suffix *-i* for perfectivity (and related temporal categories) while the others utilize the suffix *-ile*. Furthermore, phonological evidence further divides Forest Bantu in two groups. We find that North-western languages (Zones A, B20 and some C languages) have the reflexes  $\emptyset$  and *k* for the PB consonants *\*k* and *\*g* respectively. These results reveal the utility of considering non-lexical data in doing language classification.

## References

Brown, Braden, Rebecca Grollemund, Gérard Philippson and Derek Nurse. Publication expected in January 2023. Towards a new classification of North-Western Bantu languages using non-lexical data. In: *Bantoid and Bantu in Cameroon: an historical re-assessment*, Grollemund, R., Nurse, D. and Watters, J. (eds). Berlin/Boston, De Gruyter Mouton.

Grollemund Rebecca. 2012. *Nouvelles approches en classification: Application aux langues bantu du Nord-Ouest*. Thèse en Sciences du Langage (Université Lumière Lyon 2).

Grollemund, Rebecca, Simon Branford, Koen Bostoen, Andrew Meade, Chris Venditti & Mark Pagel. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences (PNAS)* 112.13296-301.

Grollemund, Rebecca and Gérard Philippson. Publication expected in 2023. A new phylogenetic classification of North-Western Bantu languages based on consonant correspondences. In: *Bantoid and Bantu in Cameroon: an historical re-assessment*, Grollemund, R., Nurse, D. and Watters, J. (eds). Berlin/Bo ston, De Gruyter Mouton.

Nurse Derek & Gérard Philippson. 2003 Towards a historical classification of the Bantu languages. In: *The Bantu Languages*, eds Nurse D & Philippson G (Routledge, London-New-York), pp 164-181.

### **An evolutionary loner in Southern African Bantu: The classification of Yeyi**

Bantu is Africa's largest language family and part of the world's largest phylum, i.e., Niger-Congo. The relatedness of Bantu languages has been recognized since nearly two centuries, and their internal classification is better understood today than ever before. Nonetheless, there are still several important uncertainties obscuring not only the diachronic linguistic processes that gave rise to Bantu diversification, but also the migrations of ancestral Bantu speakers and other population dynamics underlying them. One of the persisting mysteries of Bantu genealogy is the unclassifiable Yeyi language, a minority language spoken in northwestern Botswana and northeastern Namibia. While the Bantu origin of Yeyi is straightforward and undisputed, it is not known what its closest relatives are, even not to which major Bantu branch it belongs (Gowlett, 1997; Seidel, 2005, 2009). This lack of genealogical resolution severely limits our understanding of the deep history of the Yeyi speech community and leaves us with questions on their origin and timing of their migration into their current habitat, and the deeper connections they might once have had with other language communities.

One of the possible explanations why it is so complicated to identify Yeyi's closest Bantu relatives is its incorporation of a large degree of linguistic influence from neighboring Khoisan languages. Khoisan languages, previously analyzed as a single phylum but more recently and accurately as (at least) three separate language families (Güldemann, 2014), are characterized by their use of click phonemes. These typologically rare consonants not only occur in Khoisan languages, but also in certain Bantu languages spoken in Southern Africa. They are therefore seen as a clear indicator of Bantu-Khoisan contact (Pakendorf et al., 2017). Yeyi has the largest click inventory of all attested Bantu languages (Fulop et al., 2003; Sommer & Voßen, 1992), which suggests extensive Khoisan influence. This is also seen in the presence of Khoisan loanwords, and the borrowing of certain bound verbal affixes (Gunnink, 2022).

In this paper, we intend to shed new light on the classification of the Yeyi language, with an aim to better understand its genesis and historical development. Using a lexicon-based, Bayesian phylogenetic approach, we will provide a new Bantu classification with the explicit objective of clarifying the position of Yeyi. We therefore include all Bantu languages of the wider geographic region, including even more far-flung languages that have previously been hypothesized to bear some relationship to Yeyi. The resultant linguistic phylogeny will show the (potential) linguistic affiliations of Yeyi. These will subsequently be studied in more detail in order to identify specific lexical, phonological or morphological innovations that Yeyi may share with its putative sister languages. The role of non-Bantu contact, particularly in terms of Khoisan influence, will also be taken into account. Together, they will provide new insights into the history of the Yeyi language and its speakers.

Fulop, S. A., Ladefoged, P., Liu, F., & Vossen, R. (2003). Yeyi clicks: Acoustic description and analysis.

*Phonetica*, 60, 231–260.

Gowlett, D. F. (1997). Aspects of Yeyi diachronic phonology. In W. H. G. Haacke & E. E. Elderkin (Eds.), *Namibian languages: Reports and papers* (pp. 235–263). Rüdiger Köppe.

- Güldemann, T. (2014). 'Khoisan' linguistic classification today. In T. Güldemann & A.-M. Fehn (Eds.), *Beyond 'Khoisan': Historical relations in the Kalahari basin* (pp. 1–40). John Benjamins Publishing Company.
- Gunnink, H. (2022). *Morphological Khoisan influence in the Southern African Bantu language Yeyi*. 43(1), 3–35. <https://doi.org/10.1515/jall-2022-8892>
- Pakendorf, B., Gunnink, H., Sands, B., & Bostoen, K. (2017). Prehistoric Bantu-Khoisan language contact: A cross-disciplinary approach. *Language Dynamics and Change*, 7(1), Article 1.
- Seidel, F. (2005). The Bantu languages of the Eastern Caprivi: A dialectometrical analysis and its historical and sociolinguistic implications. In *South African Journal of African Languages* (Vol. 4, pp. 207–242).
- Seidel, F. (2009). Layered language genesis in the 'catch basin' of the Linyanti and Okavango swamps: The case of Yeyi. In *Sprache und Geschichte in Afrika* (Vol. 20, pp. 231–264).
- Sommer, G., & Voßen, R. (1992). Schnalzwörter im Yei (R.41). *Afrika Und Übersee*, 75, 1–42.

## Diachronic Null Subject Use across Latin American Spanish: Comparing Corpora<sup>1</sup>

It has been widely noted that varieties of Latin American Spanish such as Dominican Spanish (Toribio 2000) increasingly allow both overt and null subject pronouns in the same contexts (opposed to the use of the same term to describe 3rd person referential split in earlier literature), resulting in their labelling as partial null subject languages (PNSLs). Their existence has challenged the NSP as formulated by Rizzi (1982, 1986) and suggested the occurrence of null subject reduction. Given Latin American Spanish's rich history of language contact over the last five centuries, the observed decrease in use of null subjects is likely a result of contact-induced simplification. Following Trudgill's (2011) sociolinguistic typology, this contact would be characterized as short-term adult second-language acquisition which is consistent with the scenario for the African slaves brought over to the Americas by the Spanish from the 16th century onward. Indeed, there are numerous non-standard varieties of Spanish spoken by Afro-Hispanic communities throughout Latin America. These Afro-Hispanic Languages of the Americas (AHLAs) have been described as the result of "conventionalized advanced second languages," much in line with Trudgill's proposal (Sessarego 2013, 2017). The L2-difficulty of the rules surrounding null subject use would have led adult learners to overproduce overt subject pronouns. Their errors would have then become nativized in the speech of the next generation. Similar to the Jespersen cycle for negation, the partial status of null subjects in varieties such as Dominican Spanish could reflect language in change as the variety moves from a null subject language to a non-null subject language á la van Gelderen's (2011) Subject Cycle. What is needed then is a robust diachronic analysis of null subjects.

To achieve this, I constructed a corpus of 57 texts from eight countries over four centuries to evaluate the diachronic rate of null subjects in Latin American Spanish. The corpus is supplemented by transcriptions of previous fieldwork in current AHLA varieties as well as historical texts written in AHLA vernacular. The aim was/is to determine whether the corpus supports the hypothesis that the rate of overt subject pronouns has risen significantly since contact with African L2-speakers. Broadly, we expected to see the highest variation in the form of more overt subject pronouns in varieties with the highest Afro-Hispanic populations. However, despite an attempt to balance the corpus by genre (equal literary and non-literary texts), preliminary data from the Dominican Republic, Bolivia, Panama, and Spain (as a control) has shown much variation, leaving no clear diachronic patterns (Figure 1). In order to see if something similar to but independent of genre could be affecting the pronominal data, Rosemeyer's (2019) method of measuring orality levels in Portuguese plays was adapted for use in this corpus. This ORSCORE was found to have a significant ( $p=.001$ ) relationship with overtness rates (Figure 2<sup>2</sup>). With a view to account for this obstruction, we set out to compare forms of the private verbs *creer* 'to believe/think' and *pensar* 'to think' which should lessen the influence of orality as its frequency in a text is one of the five criteria used to measure orality. Unfortunately, our corpus is too small to restrict the data to one lexical item, so we are now in the process of sifting through data from the *CORDIAM* and *CORDE/CREA* corpora. We hope to tease apart this orality effect to find underlying diachronic trends.

<sup>1</sup> For consideration as a talk or poster.

<sup>2</sup> There are two outlying texts with very high ORSCORES (1.77 and 2.35) that might seem to skew the data. However, the best-fit line remains the same when they are excluded.

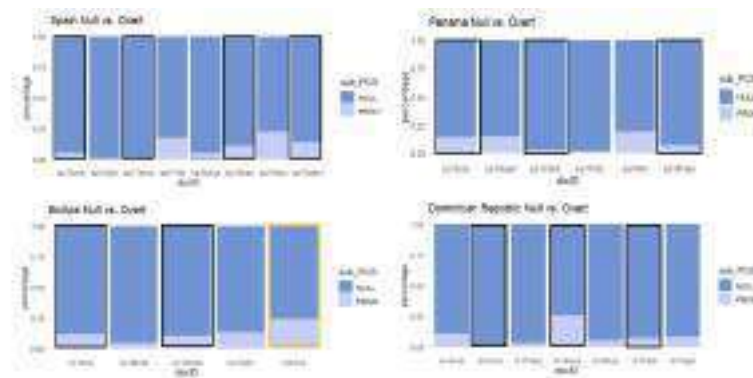
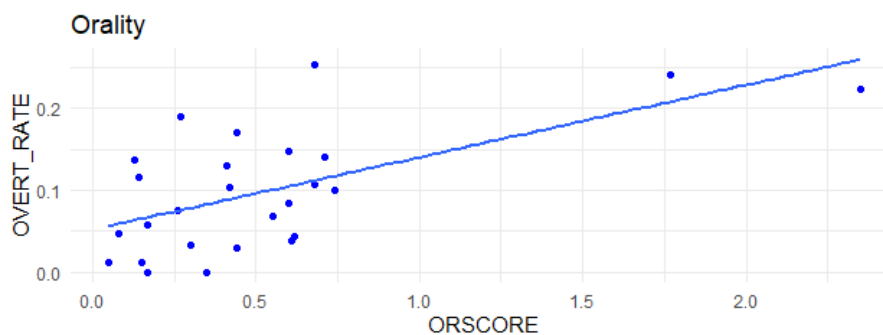
Figure 1:<sup>3</sup> Null vs. Overt Subject Pronouns

Figure 2: Degree of orality plotted against proportion of overt subject pronouns per text



## References:

- [CORDIAM] Academia Mexicana de la Lengua, *Corpus Diacrónico y Diatópico del Español de América*, [www.cordiam.org](http://www.cordiam.org)
- Gelderen, Elly van. 2011. *The Linguistic Cycle*. Oxford: OUP.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>>
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <<http://www.rae.es>>
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17: 501–57.
- Rosemeyer, Malte. 2019. "Actual and apparent change in Brazilian Portuguese wh-interrogatives." *Language Variation and Change* 31(2): 165–191. CUP.
- Sessarego, Sandro. 2013. "Afro-Hispanic Contact Varieties as Conventionalized Advanced Second Languages". *IBERIA* 5 (1): 99-125.
- Sessarego, Sandro. 2017. "Revisiting the Null Subject Parameter: New Insights from Afro-Peruvian Spanish". *Isogloss* 3 (1): 43-68.
- Toribio, Almeida J. 2000. "Setting parametric limits on dialectal variation in Spanish". *Lingua: International Review of General Linguistics* 110 (5): 315–341.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: OUP.

<sup>3</sup> The docID corresponds to the country, century, and title of a text (e.g. dr16ent = Dominican Republic, 16<sup>th</sup> century, Entremés). The document genre is bolded, and the golden bolded text is a transcript from Afro-Bolivian vernacular.



## Persistence and Change of Colexifications in Indo-European

Languages differ in how they organize their lexicon: two given concepts can be expressed by two different word forms in one language, but by the same form in another language. This encoding of different concepts with the same word form is known as colexification (François 2008). Studies have shown that highly associated concepts are colexified more frequently across languages (Xu et al. 2020), and that colexification patterns are less genetically stable than the phonological form of vocabulary, at least in Europe (Gast & Koptjevskaja-Tamm 2022). Colexification patterns can also be borrowed, as various examples of areally common colexifications show (e.g., Schapper 2022 on the colexification of BONE and STRENGTH in Melanesia or Segerer & Vanhove 2022 on the colexification of color terms in Africa).

The present study is concerned with colexification in the Indo-European language family, and its relationship to lexical material. It aims to answer the question of how strongly a colexification is tied to the phonological material its concepts are expressed with, and if it is likely to persist when phonological material is replaced. To investigate this, we use vocabulary lists from the CLICS database (Rzymiski et al. 2021), and etymology data from IELex (Chang et al. 2015) in order to split the vocabulary items into cognate sets.

When relating colexifications to etymological data, there are two main options:

- (i) Colexifications are independent of etymology and are equally found among unrelated lexemes denoting the same concept.
- (ii) Colexifications are dependent on etymology and are more frequently found with lexemes of a certain etymology.

(i) suggests that colexifications arise, persist, and spread independently of phonological substance, and that, more generally, the development of patterns is unrelated to the development of matter. (ii) on the other hand, would suggest that colexifications arise rarely and are then bound to a certain phonological form. In this scenario colexifications mostly spread and persist together with phonological matter.

As it has been shown that certain colexifications are more common cross-linguistically even among unrelated languages in different areas, we expect that colexifications in the Indo-European family are largely (but not completely) independent from the etymology of their word forms.

In a further stage of this study, we plan to use phylogenetic methods in order to investigate the history of certain colexifications in the Indo-European language family. We also plan to conduct a similar study on Austronesian languages in order to compare language families that differ strongly in terms of culture and geographical environment.

## Literature

Chang, Will, David Hall, Chundra Cathcart, and Andrew Garrett. "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis." *Language* (2015): 194-244.

François, Alexandre. "Semantic maps and the typology of colexification." *From polysemy to semantic change: Towards a typology of lexical semantic associations* 106 (2008): 163.

Gast, Volker, and Maria Koptjevskaja-Tamm. "Patterns of persistence and diffusibility in the European lexicon." *Linguistic Typology* 26, no. 2 (2022): 403-438.

Rzymiski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast et al. "The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies." *Scientific data* 7, no. 1 (2020): 13.

Schapper, Antoinette. "Baring the bones: The lexico-semantic association of bone with strength in Melanesia and the study of colexification." *Linguistic Typology* 26, no. 2 (2022): 313-347.

Segerer, Guillaume, and Martine Vanhove. "Areal patterns and colexifications of colour terms in the languages of Africa." *Linguistic Typology* 26, no. 2 (2022): 247-281.

Xu, Yang, Khang Duong, Barbara C. Malt, Serena Jiang, and Mahesh Srinivasan. "Conceptual relations predict colexification across languages." *Cognition* 201 (2020): 104280.

### PARADIGMATIC REDUNDANCY IN THE COMPLEMENT SYSTEM OF BASQUE

While most languages do not seem to have more than four types of complement clauses (Noonan 2007: 147), in Basque at least seven have to be distinguished. There is, however, a large overlap in the distribution of finite and non-finite types, which raises the question if Basque acquired finite complement clauses as a result of contact with Romance languages.

Basque has been in increasingly intense contact with Romance languages for more than two millennia, but only the last 500 years are documented in texts. Thus most instances of Romance influence cannot be observed directly but have to be inferred.

The Basque inventory of complement types consists of finite indicative and subjunctive clauses, as well as verbal nouns (VN) inflected in different cases and showing different morphosyntactic behaviour, and the infinitive-like perfective participle. Table 1 is a very rough representation of the mapping between types of matrix predicates and types of complement clauses. It shows that finite clauses are not only restricted to two classes of matrix predicates but also generally replaceable with a non-finite alternative without or with only a subtle change in meaning. The only exception to this are some utterance and cognition verbs like *esan* ‘say’ or *pentsatu* ‘think’, for which it is cross-linguistically not uncommon to take only direct speech complements (Dixon 2006: 28). Finite complements in Basque, thus, strictly speaking add nothing to the system besides a syntactically more integrated alternative to direct speech.

utterance, cognition, propositional attitude		manipulation	intention	implicative	immediate perception	phasal	modal
indicative		subjunctive					
VN in structural case		purposive VN, infinitival interrogative		VN in structural case		inessive VN	
						infinitive	

**Table 1:** Mapping between types of matrix predicates and finite and non-finite types of complement clauses.

This synchronic redundancy suggests that one class of complement clauses intruded into contexts occupied by the other. There are two possibilities:

**Hypothesis A: Finite types spread into contexts occupied by non-finite ones.** Having non-finite rather than finite complement clauses would fit the typological profile of Basque as a head-final language. In fact, OV languages tend to have preverbal complement clauses (Schmidtke-Bode/Diessel 2017: 10) and preverbal complement clauses are more commonly non-finite than finite (ibid.: 12-19). In the Romance contact languages, on the other hand, finite complement clauses are frequently used, and a subjunctive with similar functions as in Romance started to develop in Basque only at the beginning of the writing tradition (Mounole 2014: 328). Finite complementation in general seems, however, to be quite old since it was already well established at the time of the first texts, and the complementisers used are neither borrowed nor calqued from Romance.

**Hypothesis B: Non-finite types spread into contexts occupied by finite ones.** This would mean that the complex polypersonal forms of finite verbs are being replaced by simpler forms, a phenomenon that can also be observed in some other domains, like the potential mood being replaced by analytic constructions with *ahal izan* ‘be able’. While this is a plausible scenario for the subjunctive forms, which are very rarely found in non-embedded clauses and can thus almost completely be dispensed with when substituted by other subordinate forms, the indicative forms are less dispensable since they are still used in independent clauses.

Thus, a combination of both scenarios seems most plausible, namely a spread of finite complement clauses (whether they had developed independently or as a consequence of contact) replicating the Romance model and, probably more recently, a gradual substitution of subjunctive complements by non-finite forms.

**REFERENCES**

- Dixon, R. M. W. (2006): “Complement Clauses and Complementation Strategies in Typological Perspective.” In: Dixon, R. M. W. / Aikhenvald, Alexandra (eds.): *Complementation: A cross-linguistic typology*. Oxford: Oxford University Press. 1-48.
- Noonan, Michael (2007): “Complementation.” In Shopen, Timothy: *Language Typology and Syntactic Description*, vol. 2. Cambridge: Cambridge University Press. 52-150.
- Mounole, Céline (2014): *Le verbe basque ancien: étude philologique et diachronique*. ASJU 48. Bilbo: UPV/EHU.
- Schmidtke-Bode, Karsten / Diessel, Holger (2017): „Cross-linguistic patterns in the structure, function, and position of (object) complement clauses.” *Linguistics* 55(1): 1-38.

## Doing Conversation Analysis in Latin: The Case of Hedging

Rules underlying conversation and linguistic phenomena specific to interaction, have received a lot of attention, especially in the areas of conversation analysis<sup>1</sup> and pragmatics of conversation<sup>2</sup>: we now know that, while superficially messy, conversation follows a set of underlying rules which all speakers know and are in principle expected to adhere to: rules about when to take the conversational floor; desirable and undesirable responses; strategies to convey one's undivided attention to one's co-interactant, etc. When these rules are intentionally or unintentionally breached, for instance when an undesirable response is given—such as declining an invitation or providing an uninformative response—the situation calls for various mitigating strategies to avoid giving offence. One of these strategies is *hedging*—modifying one's commitment to the truth of one's statement.

In this paper, I shall look at hedging behaviour in Latin. Hedging has received some attention in Latin, notably in terms of politeness and language characterization,<sup>3</sup> but less from the point of view of Conversation Analysis. Using such sources as Plautus, Terence, Cicero and Petronius, and methodology developed within Conversation Analysis, I will address the following questions:

- Which expressions are used as hedges in Latin? Are they similar to hedges in other languages?
- In what contexts do they occur—i.e., which antecedents trigger them?
- Does the choice of hedging depend on the context of production (addressee, purpose of production, time) or genre (comedy, correspondence, novel)?
- What, if anything, does hedging tell us about spoken Latin? In other words, is Conversation Analysis applicable to Latin sources?
- Do the sources in Latin reflect hedging behaviour which has been shown to obtain on other languages?

Cross-cultural and cross-linguistic research has shown that hedging behaviour depends on cultural and cognitive factors. To gain a better understanding of hedges (linguistic devices) and hedging (communicative strategy), it is important that large bodies of data in different languages and cultures be studied. Latin, a large-corpus language, provides a wealth of material to study this phenomenon in detail and thus enrich our understanding of cognitive underpinnings of hedging, of cultural differences and commonalities, and of the interaction of

---

<sup>1</sup> See, for instance, Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language* 50, 696–735.

<sup>2</sup> See, for instance, Haugh, M. (2012). Conversational interaction. In K. Allan & K. Jaszczolt (Eds.), *The Cambridge Handbook of Pragmatics* (Cambridge Handbooks in Language and Linguistics, pp. 251-274). Cambridge: Cambridge University Press.

<sup>3</sup> Barrios-Lech, Peter (2016). *Linguistic Interaction in Roman Comedy*. Cambridge, Cambridge: University Press: 142–165;

Mencacci, Francesca (2022). 'Text as Interaction: Vt Mihi (Quidem) Videtur as a Hedging Device in Latin Literary Texts.' In *Politeness in Ancient Greek and Latin*, edited by Luis Unceta Gómez and Łukasz Berger, 103–126. Cambridge: Cambridge University Press.

hedges and genre. This paper is intended as a contribution to the growing cross-cultural body of research on hedging.

## It Ain't Over till It's Over. Bilingualism and language decay in Sicilian inscriptions

Inscriptions and ancient sources both indicate that Sicily has been characterized by multilingualism since before the classical period, when Greek, Punic, and indigenous languages were spoken in several areas of the island, and often coexisted in some areas. However, Sicilian indigenous languages disappeared with the end of the classical period and Punic survived only (and very scarcely) only until 1<sup>st</sup> c. BCE circa. Greek, on the other hand, is the only language attested in inscriptions throughout the entire Sicily, and it remained so until the Roman conquest, when Latin spreads across the island. During the Roman period, and until the late antiquity, Greek/Latin bilingualism was common in Sicily, although with important diatopic differences: e.g., in Syracuse (Korhonen 2012) and surrounding areas Greek clearly remained the most prestigious variety until at least the imperial period, while it is likely that rural areas were predominantly Latin-speaking (Korhonen 2016). Finally, and famously, Greek must have undergone language death in western Sicily during the early Middle Ages.

While it is relatively uncomplicating determining when a language no longer appears in inscriptions, establishing when it ceased to be spoken is a rather complicated task. Even if a language is not inscribed on a stone, it might still be present in some parts (e.g., strata) of a society, and not anymore used in inscriptions because of the increasing loss of prestige, or because of political reasons. In these cases, it is sometimes possible to determine the presence of these not explicitly represented language through the presence of language contact. For giving a famous Sicilian example, it has been convincingly suggested that behind the bilingual (Greek-Latin) inscription ISic000470 there is a Punic speaker (Susini 1968; Tribulato 2011). In other words, while a language is undergoing decay, and before it completely disappears as a spoken variety, it might still be detected through language contact. For what concerns ancient Sicily, I argue that we can access this information through what I call “implicitly bilingual” inscriptions, i.e., monolingual Greek inscriptions that show phenomena of language contact.

Despite the great importance of the topic of language decay and death, there are very few works on Greek language death (e.g., JANSE 2003), and even fewer on Punic and Sicilian indigenous languages (partially, POCETTI 2012 and MARCHESINI 2012), and no-one has yet exploited bilingual inscriptions to investigate the reasons for language death and language vitality. In this paper, I will use “implicitly bilingual” inscriptions to assess the dynamics of language decay and language death between Punic and Greek and between indigenous languages and Greek, and to assess the possibility that Punic has been spoken longer than it is attested. Furthermore, I will adopt this model to analyze Latin inscriptions from Western imperial Sicily that show phenomena of language contact with Greek, in order to assess the Greek language decay in areas that we know Greek was the minority language.

AMADASI GUZZO, M. G. 1990, *Iscrizioni fenicie e puniche d'Italia*, Roma

JANSE, M. 2003, Introduction: Language death and language maintenance. Problems and prospects. In: Janse, M. & Tol, S. (eds.), *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches*, Oxford, xi–xvii.

KORHONEN, K. 2011, Language and Identity in the Roman Colonies of Sicily. In: Sweetman, R. (ed.) *Roman Colonies in the First Century of Their Foundation*, Oxford, 7–31.

KORHONEN, K. 2016, *Questioni del bilinguismo siciliano antico in lunga durata*, «Linguarum Varietas» 6: 115–120.



MARCHESINI, S. 2012, The Elymian language. In: Tribulato, O. (ed.) *Language and Linguistic Contact in Sicily*, Cambridge, 95–114.

POCETTI, P. 2012, Language relations in Sicily: evidence for the speech of the Σικανοί Σικελοί and others. In: Tribulato, O. (ed.) *Language and linguistic contacts in ancient Sicily*, Cambridge, 49–94.

SUSINI, G. 1967. *Il lapicida romano*. Roma.

TRIBULATO, O. 2011, *The stone-cutter's bilingual inscription from Palermo (CIL X 7296 = IG XIV 297): a new interpretation*, «ZPE» 177: 131–40.

### The emergence of word-initial voiced stops in Proto-Hungarian

This presentation aims to bring new solutions to the old problem of the emergence of word-initial voiced stops in the prehistory of Hungarian. Word-initial voiced stops are frequent in modern Hungarian, but it is well-known that these arose during the relatively late prehistoric period, during the existence of Hungarian as an independent branch (Abaffy 2003: 122). It has been assumed that they emerged through sporadic changes  $*p > b$ ,  $*t > d$  (but not  $k > *g$ ) in word-initial position in some inherited words, and became more frequent with the influx of Turkic and Iranian loanwords (Abaffy 1994: 16–18; Bárczi 1958: 113). However, it is problematic that in some cases Hungarian shows voiced stops even though the attested Turkic data points to voiceless stops (for example, Hu *gödény* ← reconstructed Western Old Turkic *\*güdän* or *\*kütän*, cf. Middle Turkic *kutan* ‘a kind of bird’; Hu *dara* ← Turkic *\*dari* or *\*tari* cf. Old Turkic *tarig* ‘crops, millet, sowing’; Ligeti 1986: 193–194; Róna-Tas & Berta 2011: 1074–1075, 1078), and the more scarce Iranian loans also show a confusing picture, as corresponding voiced stops are found in some loans but not in others: Hu *bűz* ‘smell’ ← Iranian *\*bauda-* (> Ossetic *bud*, *bodæ* id.), Old Hungarian *kazdag* ‘rich’ ← Iranian (Alanic) *\*gæzdig* (> Ossetic *qæznyg*, *ğæzdug* id.) (Abaffy 1994).

It is also problematic that it is not known under which conditions the word-initial voiced stops arose in Hungarian in the first place, as these are not regular reflexes of any Proto-Uralic phonemes (contrary to word-internal voiced stops that developed regularly from the clusters of nasal and stop, Sammallahti 1988: 520). However, it has been claimed in presentations of Hungarian historical phonology that some inherited words have an irregular voiced reflex (Abaffy 2003: 118; Gerstner 2018: 107). The situation is somewhat similar as with the related Permic branch of the Uralic family, where similar voicing of has taken place (for example Proto-Permic *\*bur* < Proto-Uralic *\*para* ‘good’; Proto-Permic *\*d̥jn* < Proto-Uralic *\*t̥iŋi* ‘base of a tree’); however, in Permic this process seems to be a regular, conditioned change, voicing assimilation caused by word-internal voiced consonant (Csúcs 2005: 154–158), whereas in Hungarian the situation is often described as sporadic or no conditions are not given (Maticsák 2020: 360; Bereczki 2003: 69–70; Abaffy 2003: 118). It has been suggested that word-internal voiced consonants caused the voicing of word-initial stop (MSzFE: 105; UEW: 374–375; Honti 2017: 15), similarly as in Permic, but the situation is complicated as numerous counter examples can be found. The voicing of stops has been sometimes been considered a shared, areal innovation of Proto-Hungarian and Proto-Permic, but the evidence is inconclusive (Csúcs 2021: 44).

Most of the alleged examples of the sporadic voicing in inherited vocabulary are problematic etymologies, showing also irregular vocalism and semantic mismatch in many cases; for example, Hungarian *bal* ‘left’ allegedly reflects Proto-Uralic *\*palV* (MSzFE: 105; UEW: 351–352), but the only cognate is Udmurt Ud *pal’l’an* ‘left’ that shows irregular vocalism; Hu *bőr* ‘skin’ is allegedly a reflex of Proto-Uralic *\*perV* ‘bark’ (MSzFE: 110–111; UEW: 374), but *ő* is not the regular reflex of Proto-Uralic *\*e*. Some words have competing etymologies: for example, Hungarian *dug* ‘to squeeze’ has been considered both a loan from Turkic *\*dig-* ~ *\*tig-* ‘to squeeze’ (Róna-Tas & Berta 2011: 303–306) and a reflex of Proto-Uralic *\*tuŋki-* (MSzFE: 135; UEW: 537–538). Because of this, the idea that Proto-Uralic stops can be reflected by Hungarian voiced stops in anlaut is very doubtful.

In this presentation, all the relevant Uralic etymologies and loanwords will be discussed, and the aim is to determine when and under what conditions the word-initial voiced stops became possible in Proto-Hungarian.

## References

- Abaffy, Erzsébet 1994: Ósmagyar jövevényszavaink vállalása egykorú fonémarendszerünkről. In: István Szathmári & Erzsébet E. Abaffy & Éva B. Lőrinczy (eds.), *Bárczi Géza emlékkönyv*. Budapest. 9–20.
- Abaffy, Erzsébet 2003: Ósmagyar kor. Hangtörténet. In: Jenő Kiss & Ferenc Pusztai (eds.), *Magyar nyelvtörténet*. Budapest. 106–128, 301–351, 596–609.
- Bárczi, Géza 1958: *Magyar hangtörténet*. Budapest.
- Bereczki, Gábor 2003: *A magyar nyelv finnugor alapjai*. Budapest.
- Csúcs, Sándor 2005: *Die Rekonstruktion der permischen Grundsprache*. Budapest.
- Csúcs, Sándor 2021: A másodlagos permi-magyar nyelvi érintkezések kérdése. In: László Klima & Attila Türk (eds.), *Párhuzamos történetek*. Budapest. 41–50.
- Gerstner, Károly 2018: Hangtörténet. In: Jenő Kiss & Ferenc Pusztai (eds.), *A magyar nyelvtörténet kézikönyve*. Budapest. 103–128.
- Honti, László 2017: *A magyar és a nyugati ótörök szókészleti kapcsolatairól*. Budapest.
- Ligeti, Lajos 1986: *A magyar nyelv török kapcsolatai a honfoglalás előtt és az Árpád-korban*. Budapest.
- Maticsák, Sándor 2020: *A magyar nyelv eredete és rokonsága*. Budapest.
- MSzFE = Gyögy Lakó 1971–1978: *A magyar szókészlet finnugor elemei*. Budapest.
- Róna-Tas, András & Árpád Berta 2011: *West Old Turkic. Turkic loanwords in Hungarian*. Wiesbaden.
- Sammallahti, Pekka 1988: Historical phonology of the Uralic languages with special reference to Samoyed, Ugric and Permic. In: Denis Sinor (ed.), *The Uralic languages*. Leiden. 478–554.
- UEW = Károly Rédei 1986–1991: *Uralisches Etymologisches Wörterbuch*. Wiesbaden.

## The Evolution of Spatial Orientation Systems in Mayan and Nuristani

The diversity of spatial orientation systems in the world's languages has fascinated linguists, anthropologists and cognitive scientists alike (cf. Levinson 1998) and it has been taken as one of the few compelling arguments in favor of linguistic relativity (cf. Levinson 1996: 195-196).

Nevertheless, we still know little about how complex geomorphic orientation systems evolve diachronically or about the lexical sources that end up as elements of their paradigms. Palmer's (2015: 210) hypothesis that "a correlation will exist between a language's system of absolute spatial reference and the topography of the language locus" would suggest that certain environments favor the development of certain kinds of spatial systems. If this is the case, then it should be possible to compare the development of spatial orientation systems within language families that, e.g., have members both in mountainous and in flat environments.

In two case studies, we trace the evolution of spatial orientation paradigms from a variety of constructions involving motion verbs in Mayan languages and from adverbial formations in Nuristani languages. The fact that diverse sources lead to similar outcomes in unrelated languages spoken in similar environments, whereas related languages spoken in different environments do not develop the same amount of semantic distinctions, lends credence to the idea that the physical environment can under certain circumstances have a direct impact on linguistic structures and that linguistic coordinate systems "are constructed in response to the environment" (Palmer 2015: 210).

## References

- Levinson, Stephen. 1996. Relativity in spatial conception and description. In: Gumperz, John & Stephen Levinson, *Rethinking Linguistic Relativity* (Studies in the Social and Cultural Foundations of Language 17), 177-202. Cambridge: Cambridge University Press.
- Levinson, Stephen. 1998. Studying spatial conceptualization across cultures: Anthropology and cognitive science. *Ethos* 26(1). 7-24.
- Palmer, Bill. 2015. Topography in language: Absolute frame of reference and the topographic correspondence hypothesis. In: De Busser, Rik & Randy LaPolla (eds.), *Language Structure and Environment. Social, Cultural and Natural Factors* (Cognitive Linguistic Studies in Cultural Contexts 6), 179-226. Amsterdam: John Benjamins.

## **Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture**

A wide variety of work in both the Greenbergian typological tradition (Greenberg 1963, 1978; Hawkins 1979; Dryer 1989; Nichols 1992) and in the comparative-syntax paradigm (Baker 2001, 2008; Longobardi and Guardiano 2009; Roberts 2019) has established that different surface/typological features are frequently *correlated*; individual feature values can favour or disfavour others. Such correlations arise over historical time in a dynamical process in which both “vertical” genetic descent and “horizontal” contact processes play a role, and it has been suggested that the present-day global distribution of language types is reflective of the stationary state of this stochastic process (Greenberg 1978; Maslova 2000; Kauhanen et al. 2021; Jäger and Wahle 2021). This makes possible, among other things, the estimation of linguistic rates of change using a variety of techniques (Dediu and Cysouw 2013; Murawaki and Yamauchi 2018; Kauhanen et al. 2021).

What has received far less attention so far is the question of how *geographical* patterns of linguistic features relate to the purely linguistic properties of those features. Although Kauhanen et al. (2021) put forward a method for inferring rates of change of features from their present-day geospatial distributions, their method is restricted to statistically independent features. In this paper, using a set of word order features as our data, we show that feature correlations and geospatial distributions in fact stand in a systematic relationship. We show that typologically dispreferred types (such as the disharmonic combination OV & prepositions) tend to be surrounded by a greater variety of types than typologically preferred types (such as the harmonic combination OV & postpositions), which favour more uniform geographical environments. In particular, languages of dispreferred type are often found geographically “sandwiched” between languages of preferred type.

This *Sandwich Conjecture* is operationalized in terms of a notion of *neighbourhood entropy*, an information-theoretic measure of the extent of typological variability in a language’s immediate geographical neighbourhood. We apply this method to 28 word order feature pairs harvested from WALS (Dryer and Haspelmath 2013), and use an unrelated (and demonstrably uncorrelated) feature (presence/absence of voicing contrast) as a control feature. Measuring the typological correlation of two features with the usual  $\phi$  coefficient (cf. Jäger and Wahle 2021), measuring “sandwichness” with neighbourhood entropy, and using permutation tests to factor out random noise, we then show that a statistically significant linear relationship exists between feature correlations and geospatial patterning: the more two features are correlated purely typologically (i.e. simply by virtue of the number of languages exhibiting the various possible feature combinations), the more sandwiched the corresponding geospatial distribution of types is. Hence, feature correlations are reflected in the geospatial patterning of said features.

This finding suggests a possible mechanism for the diachronic stability of dispreferred types, an otherwise surprising fact given that disharmonic feature combinations should be expected to resolve one way or the other into harmony over sufficient historical time. The Sandwich Conjecture explains this paradoxical stability by proposing that dispreferred types are preferentially found in typologically rich geographical environments, meaning that ample opportunities for horizontal transfer (contact) exist to sustain disharmony.

In addition to the above quantitative results, we present converging evidence from proof-of-concept computer simulations in which the same finding emerges.

## References

- Baker, Mark. 2001. *The Atoms of Language*. New York: Basic Books.
- . 2008. “The Macroparameter in a Microparametric World.” In *The Limits of Linguistic Variation*, edited by Theresa Biberauer, 351–73. John Benjamins.
- Dediu, Dan, and Michael Cysouw. 2013. “Some Structural Aspects of Language Are More Stable Than Others: A Comparison of Seven Methods.” *PLOS ONE* 8: e55009.
- Dryer, Matthew S. 1989. “Large Linguistic Areas and Language Sampling.” *Studies in Language* 13: 257–92.
- Dryer, Matthew S., and Martin Haspelmath, eds. 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>.
- Greenberg, Joseph H. 1963. “Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press.
- . 1978. “Diachrony, Synchrony, and Language Universals.” In *Universals of Human Language*, edited by Joseph H. Greenberg, 1:61–91. Stanford, CA: Stanford University Press.
- Hawkins, John A. 1979. “Implicational Universals as Predictors of Word Order Change.” *Language* 55 (3): 618–48.
- Jäger, Gerhard, and Johannes Wahle. 2021. “Phylogenetic Typology.” *Frontiers in Psychology* 12: 682132.
- Kauhanen, Henri, Deepthi Gopal, Tobias Galla, and Ricardo Bermúdez-Otero. 2021. “Geospatial Distributions Reflect Temperatures of Linguistic Features.” *Science Advances* 7: eabe6540.
- Longobardi, Giuseppe, and Cristina Guardiano. 2009. “Evidence for Syntax as a Signal of Historical Relatedness.” *Lingua* 119 (11): 1679–1706.
- Maslova, Elena. 2000. “A Dynamic Approach to the Verification of Distributional Universals.” *Linguistic Typology* 4: 307–33.
- Murawaki, Yugo, and Kenji Yamauchi. 2018. “A Statistical Model for the Joint Inference of Vertical Stability and Horizontal Diffusibility of Typological Features.” *Journal of Language Evolution* 3: 13–25.
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.

## Exploiting phylogenetic modeling to uncover directionality in the emergence of universals

Extensive investigations into word order and hierarchical universals have uncovered well-supported cross-linguistic tendencies for certain features of languages to go hand in hand (Greenberg 1963, 1966, Keenan & Comrie 1977, Kozinsky 1981, Hawkins 1983, Dryer 1992, Jäger and Wahle 2021, Authors). Many linguists hold that such universals are to be accounted for in terms of diachrony, i.e. in the words by Bickel et al. (2015: 29), "most – perhaps even all – statistical universals are not really synchronic in nature, but are rather the result of underlying diachronic mechanisms that cause languages to change in preferred or 'natural' ways". However, we still know very little about these 'diachronic mechanisms'.

Here, we initiate an investigation into possible mechanisms through which correlated change proceeds in a strict quantitative, phylogenetic context. We consider for a set of highly supported word order universals (including the correlation pairs identified by Dryer 1992) and highly supported hierarchical universals (including well-known dependencies in the (morphological) marking of person, number and gender) which of the two features changed first. For example, in the well-supported universal "In languages with prepositions, the genitive almost always follows the governing noun" (Greenberg's 1963 Universal no. 2), which word order feature changed first: the order of adposition and noun or the order of genitive and noun?

The data that we use comes from Grambank (Skirgård et al. 2022), a large typological database featuring morphosyntactic data on 2400+ languages. The global language tree we use has been created by Bouckaert et al. (2022). We use two methods to investigate directionality in correlated evolution. First, Nunn and Cooper's (2015) so-called "species-pairs evolutionary lag test" (SPELT), which explicitly tests whether change in one feature lags behind change in another feature. Second, ancestral state estimations at various time depths are contrasted to show which feature changed to the state predicted by the universal first. This is done using *BayesTraits* (Meade and Pagel 2022).

We envision a number of possible outcomes of this study. First of all, we may find that there are no global patterns at all: to go back to Greenberg's (1963) Universal no. 2, we may find that it is (1) sometimes the order of genitive and noun that changes first, and the order of adposition and noun that changes second, and (2) sometimes the other way around. Secondly, we may find that there is some evidence for directed change, but only for a subset of the universals that we investigate. Thirdly, we may identify strong patterns of directionality, possibly unveiling what are the central pivot features (potentially including the order of object and verb, Dryer 1992) that trigger change in related features. Confirming or rejecting this third hypothesis would advance our understanding of the causal factors in grammatical change and also point toward explanations for cross-linguistic tendencies that all syntactic theories should be able to account for.



## References

Authors <>

Bickel, Balthasar, Alena Witzlack-Makarevich, Taras Zakharko, and Giorgio Iemmolo. 2015. 'Exploring Diachronic Universals of Agreement: Alignment Patterns and Zero Marking across Person Categories'. In *Agreement from a Diachronic Perspective*, edited by Jürg Fleischer, Elizabeth Rieken, and Paul Widmer, 29–51. Berlin: De Gruyter Mouton.

Bouckaert, Remco, David Redding, Oliver Sheehan, Thanos Kyritsis, Russell Gray, Kate E Jones, and Quentin Atkinson. 2022. 'Global Language Diversification Is Linked to Socio-Ecology and Threat Status'. Preprint. SocArXiv. <https://doi.org/10.31235/osf.io/f8tr6>.

Dryer, Matthew S. 1992. 'The Greenbergian Word Order Correlations'. *Language* 68 (1): 81–138.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of grammar*. Cambridge: The MIT Press.

Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. (Janua Linguarum, Series Minor, 59.) The Hague: Mouton.

Hawkins, John A. 1983. *Word Order Universals*. San Diego: Academic Press.

Jäger, Gerhard, and Johannes Wahle. 2021. 'Phylogenetic Typology'. *Frontiers in Psychology* 12: 682132.

Keenan, Edward L., and Bernard Comrie. 1977. 'Noun Phrase Accessibility and Universal Grammar'. *Linguistic Inquiry* 8 (1): 63–99.

Kozinsky, Isaak (1981). *Nekotorye grammaticheskie universalii v podsistemax vyrazenija subjektno-objektnyx otnosenij*. [Some grammatical universals in subsystems of expression of subject-object relations.] Doctoral dissertation, Moskovskij gosudarstvennyj universitet.

Meade, Andrew, and Mark Pagel. 2022. *BayesTraits V4.0.0*. Available at <http://www.evolution.reading.ac.uk/BayesTraitsV4.0.0/BayesTraitsV4.0.0.html>

Nunn, Charles L., and Natalie Cooper. 2015. 'Investigating Evolutionary Lag Using the Species-Pairs Evolutionary Lag Test (SPELT)'. *Evolution* 69 (1): 245–53.

Skirgård, Hedvig, Hannah J. Haynie, Damián E. Blasi, Harald Hammarström, Jeremy Collins, Jay J. Latache, Jakob Lesage, Tobias Weber, Alena Witzlack-Makarevich, Sam Passmore, Angela Chira, Luke Maurits, Russell Dinnage, Michael Dunn, Ger Reesink, Ruth Singer, Claire Bownen, Patience L Epps, Jane Hil, Outi Vesakoski, Martine Robbeets, Noor Karolin Abbas, Daniel Auer, Nancy A. Bakker, Giulia Barbos, Robert D. Borges, Swintha Danielsen, Luise Dorenbusch, Ella Dorn, John Elliott, Giada Falcone, Jana Fischer, Yustinus Ghanggo Ate, Hannah Gibson, Hans-Philipp Göbel, Jemima A. Goodall, Victoria Gruner, Andrew Harvey, Rebekah Hayes, Leonard Heer, Roberto E. Herrera Miranda, Nataliia Hübler, Biu H. Huntington-Rainey, Jessica K. Ivani, Marilen Johns, Erika Just, Eri Kashima, Carolina Kipf, Janina V. Klingenberg, Nikita König, Aikaterina Koti, Richard G. A. Kowalik, Olga Krasnoukhova, Nora L. M. Lindvall, Mandy Lorenzen, Hannah Lutzenberger, Tânia R.A. Martins, Celia Mata German, Suzanne van der Meer, Jaime Montoya, Samamé Michael Müller, Saliha Muradoglu, Kelsey Neely, Johanna Nickel, Miina Norvik, Cheryl Akinyi Oluoch, Jesse Peacock, India O.C. Pearey, Naomi Peck, Stephanie Petit, Sören Pieper, Mariana Poblete, Daniel Prestipino, Linda Raabe, Amna Raja, Janis Reimringer, Sydney C. Rey, Julia Rizaew, Eloisa Ruppert, Kim K. Salmon, Jill Sammet, Rhiannon Schembri, Lars Schlabach, Frederick W.P. Schmidt, Amalia Skilton, Wikaliler Daniel Smith, Hilário de Sousa, Kristin Sverredal, Daniel Valle, Javier Vera, Judith Voß, Tim Witte, Henry Wu, Stephanie Yam, Jingting Ye 葉婧婷, Maisie Yong, Tessa Yuditha, Roberto Zariquiey, Robert Forkel, Nicholas Evans, Stephen C. Levinson, Martin Haspelmath, Simon J. Greenhill, Quentin Atkinson & Russell D. Gray 2022. "Grambank Reveals the Importance of Genealogical Constraints on Linguistic Diversity and Highlights the Impact of Language Loss." SocArXiv. December 4. [doi:10.31235/osf.io/grh45](https://doi.org/10.31235/osf.io/grh45).

Periphrastic perfects reflect the lexical semantic distinctions of their auxiliaries

Compound perfect constructions commonly develop from resultative constructions, often with auxiliaries glossed as ‘be’ or ‘have’. Some formally similar constructions, however, exhibit important features not typically associated with perfects. In Portuguese, for instance, the ‘have’ + participle construction with a present tense auxiliary marks recent habitual action: *Tenho lido muitos artigos* ‘I have been reading many articles’ (not ‘I have read many articles’).

This paper contrasts the Portuguese periphrasis with *ter* (< Latin TENĒRE) + perfective participle with other Romance compound perfects formed with reflexes of Latin HABĒRE (cf. Harris 1982) and contextualizes it with regard to the effects lexical semantics on grammaticalization patterns involving semantically similar source material. I argue that previous analyses of auxiliaries have erroneously equated lexical verbs of similar meanings and consequently overlooked certain inference patterns that contribute to semantic change in grammaticalization (cf. Traugott & Dasher 2002). I also address the ongoing need for greater terminological care in this area.

While scholars such as Bybee et al. (1994) tend to treat sources of auxiliaries as equivalents if they can be glossed with a single English verb, Juge has argued (2002) that such an approach incorrectly predicts nearly identical outcomes of the grammaticalization of seemingly similar constructions that in fact result in significantly different structures. In the case of the Portuguese resultative construction, the key factor is the distinction between Latin HABĒRE and TENĒRE, both commonly glossed ‘have’.

These two verbs were not exact synonyms in Latin and did not yield exact synonyms in the Romance languages (cf. Harre 1991). While both indicate possession, TENĒRE and its reflexes also mean ‘hold’ and ‘keep’. The latter sense is a key factor in the development of Portuguese *ter* as an auxiliary. Certain predicates, when paired with a verb meaning ‘keep’, favor an iterative reading. If I assert that I keep my grass cut, for example, someone may conclude that I mow it periodically. For this reason, the label ‘iterative’ fits Portuguese *ter* + participle better than ‘perfect’.

Indeed, the term ‘perfect’ itself presents certain complications. First, it is often conflated with the aspectual term ‘perfective’, although some languages, such as Catalan, clearly show that these are orthogonal categories. Therefore in many cases the term ‘anterior’ is preferable to ‘perfect’ (cf. Bybee et al. (1994), among others; e.g., English future anterior *She will have arrived*).

Cruse (1986) suggests analyzing what he calls a lexeme’s sense-spectrum, or the collection of senses it encodes. For example, Spanish *mismo* has a range of senses that correspond to some of the senses of the English lexemes *same*, *very*, *right*, *oneself*, and *exactly*. This approach facilitates the identification of more fine-grained semantic relations that shape patterns of lexicalization and grammaticalization and discourages treating similar lexemes in different languages as being more similar than they are, which is an especially common problem in cases involving verbs of motion (cf. Juge 2007).

A widely recognized characteristic of verbs that become grammaticalized is that such lexemes are often highly polysemous before becoming grammaticalized and that they then show greater polysemy as a result of undergoing grammaticalization. The nature of the polysemy patterns shown by such lexemes, however, is still not well understood. Close analysis of the lexical semantics of grammaticalization—including interactions with factors like pragmatic inferencing—in familiar languages with well-attested histories allows typologists and historical linguists to more accurately apply insights gained from the examination of constructions in these languages to those in more poorly documented languages.

**References/**

- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.
- Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.
- Dias da Costa, Albano. 1976. Periphrastic expressions in Portuguese. In Jürgen Schmidt-Radefeldt (ed.), *Readings in Portuguese linguistics*, 187-243. Amsterdam: North-Holland Publishing.
- Emanatian, Michele. 1992. Chagga ‘come’ and ‘go’: Metaphor and the development of tense-aspect.” *Studies in Language* 16.1–33.
- Harre, Catherine. 1991. *Tener plus past participle: A case-study in linguistic description*. London: Routledge.
- Harris, Martin. 1982. The ‘past simple’ and the ‘present perfect’ in Romance. In Nigel Vincent & Martin Harris (eds.), *Studies in the Romance verb*, 42-70. London: Croom Helm.
- Juge, Matthew L. 2002. Polysemy patterns in semi-lexical items in Spanish and Catalan. *Studies in Language* 26(2). 315-335.
- Juge, Matthew L. 2007. Metaphor and teleology do not drive grammaticalization. In Joseph C. Salmons & Shannon Dubenion-Smith (eds.), *Historical linguistics 2005: Selected papers from the 17th International Conference on Historical Linguistics, Madison, 31 July-5 August 2005*, 33-48. *Current Issues in Linguistic Theory*, 284. Amsterdam: John Benjamins.
- Parkinson, Stephen. 1988. Portuguese. In Martin Harris & Nigel Vincent (eds.), *The Romance languages*, 131-169. New York: Oxford University Press.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

**Keywords/**

grammaticalization  
 lexical semantics  
 perfect  
 anterior  
 periphrasis

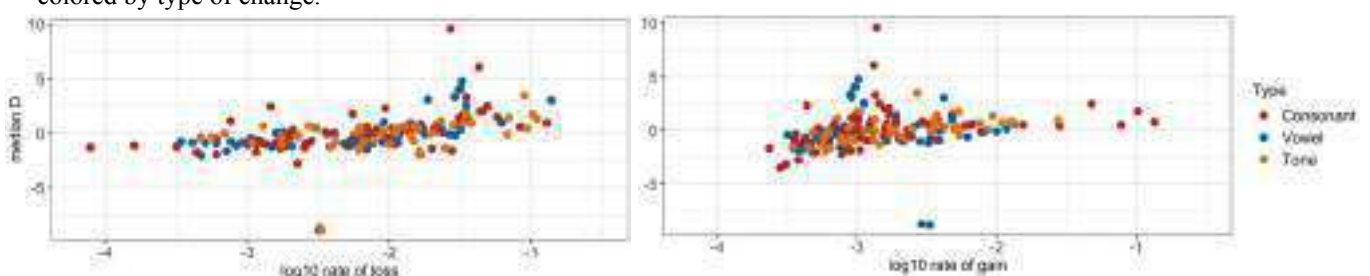
## Is tone change more rapid and irregular than segmental change? - A Mixtec case study

Despite the abundance of tonal languages around the world (Yip 2002), the diachrony of tone is still poorly understood, especially when compared to segmental sound change (Campbell 2021). This is even more so regarding tone change *per se* as opposed to tonogenesis (Ferlus 2004, Dockum 2019, among others). This *lacuna* has contributed to the assumption that tones are inherently unstable and can change unpredictably (Ratliff 2015). In this talk, I address the questions of whether tones change faster than segments and whether they can be used to inform subgrouping in the Mixtec languages of southern Mexico (Otomanguean). All Mixtec languages exhibit complex systems of lexical and grammatical tone that have to be reconstructed to the proto-language (Dürr 1987, Swanton & Mendoza Ruíz 2021) and most probably all the way back to proto-Otomanguean (Rensch 1976, Campbell 2021). As such, these languages provide an ideal case study for testing assumptions about tone change.

I created a database of tonal and segmental sound changes across a sample of 42 Mixtec languages. The changes were identified based on cognate sets derived from a 209-item basic vocabulary list. All entries were converted to IPA and standardized with regard to the tone notation for consistent identification of sound changes. For each cognate set, I reconstructed a proto-form (both tones and segments) applying the comparative method and incorporating previous reconstructions where available (Josserand 1983, Dürr 1987, Swanton & Mendoza Ruíz 2021). I established tone correspondences and tone changes across the 42 languages of the sample applying the comparative methods as with segments. The results of this work are stored in multiple, interlinked databases that can be expanded and re-used for other research questions in the future. Based on a posterior distribution of phylogenetic trees from a previous study (AUTHOR et al. submitted), I calculated phylogenetic signal with the metric  $D$  (Fritz & Purvis 2010) and estimated rates of gain and loss with a Hidden Markov Model (Beaulieu et al. 2013) for each segmental and tonal change identified.

The results, summarized in Fig. 1, show that tone change in Mixtec does not behave differently from segmental change in any significant way. Many tone changes carry phylogenetic signal and can thus contribute to our understanding of the internal structure of this language family just like segmental changes. Tones also do not change faster or slower than segments overall, exhibiting similar transition rates as segments. These two measures suggest that tone change operates much the same way as segmental change and should be investigated on a par with segmental change.

Fig. 1: Median  $D$  (phylogenetic signal, y-axis) and rates of gain and loss (change rate, x-axis) per sound change colored by type of change.



## References

- AUTHOR et al. submitted. [Title withheld to ensure blind review]
- Beaulieu, Jeremy, O'Meara, Brian & Donoghue, Michael. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. *Systematic biology*, 62(5), p.725–737.
- Campbell, Eric W. 2021. Why Is Tone Change Still Poorly Understood, and How Might Documentation of Less-studied Tone Languages Help? In: Epps, Law, Pat-El (eds.). *Historical Linguistics and Endangered Languages*. Routledge
- Dockum, Ricker. 2019. *The tonal comparative method: Tai tone in historical perspective*. Doctoral dissertation: Yale University
- Dürr, Michael. 1987. A preliminary reconstruction of the Proto-Mixtec tonal system. *Indiana* 11. 19–61.
- Ferlus, Michel. 2004. The origin of tones in Viet-Muong. In *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society 2001* (pp. 297–313). Arizona State University Programme for Southeast Asian Studies Monograph Series Press.
- Fritz, Susanne & Andy Purvis. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, 24(4), p.1042–1051.
- Josserand, J. Kathryn. 1983. Mixtec dialect history. Tulane University PhD dissertation.
- Ratliff, Martha. 2015. Tonoexodus, tonogenesis, and tone change. In: Honeybone, Patrick & Joseph Salmons (eds.). *The Oxford Handbook of Historical Phonology*. Oxford University Press. pp. 245-261.
- Rensch, Calvin R. 1976. *Comparative Otomanguean Phonology*. Bloomington: Indiana University Press
- Swanton, Michael & Juana Mendoza Ruíz. Observaciones sobre la diacronía del tono en el Tu'un Savi (mixteco) de Alcozauca de Guerrero. In Arellanes & Guerrero (eds.), *Estudios lingüísticos y filológicos en lenguas indígenas mexicanas: Celebración de los 30 años del Seminario de Lenguas Indígenas*. México, D.F: Universidad Nacional Autónoma de México
- Yip, Moira. 2002. *Tone*. Cambridge University Press.

### **Beyond the paradigm: Change and expansion in Thai pronominal reference**

This paper examines the historical development of personal pronouns in Thai. Building on work examining etymology and evolving dictionary definitions of Thai pronouns beginning in the 19<sup>th</sup> century (Author xxx), this study expands its scope to all available pedagogical and reference materials about Thai, ranging from the 17<sup>th</sup> century to the mid-20<sup>th</sup> century. We show how the set of pronouns has expanded, and how both the denotative meanings and the social context of their usage has changed over the centuries.

Multiple synchronic studies have explored the social dimensions of Thai pronouns (Cooke 1965, Palakornkul 1972, Saisuwan 2016). Saisuwan, for example, theorizes pronoun selection along dimensions of formality, power, and social distance. Thai is notable for its lack of a single pronominal paradigm, or even set of paradigms; many options are available for every cell in a conventional pronoun paradigm, each with distinct socioindexical meaning. Lexical items that can behave pronominally in Thai include proper names, occupational titles, social status terms, and kinship terms, in addition to a rich assortment of ‘true’ pronouns, including both ones inherited from Proto-Tai and ones borrowed from Khmer, Pali, Chinese, and English.

Here we focus on materials produced for second-language learners of Thai, because they often provide more direct explanation of social dimensions of pronoun usage. Some primary material from early written Thai texts (13<sup>th</sup>-15<sup>th</sup> centuries) is also surveyed in order to help establish which terms were in use before European contact. Works in our survey include grammatical descriptions (*e.g.*, de La Loubere 1693, Low 1828, Buell 1840, Jones 1842, Pallegoix 1850, Noss 1954), dictionaries (*e.g.*, Jones 1833, Caswell 1846, Pallegoix 1854, Bradley 1873, Michell 1892, Hays 1894, Pallegoix and Vey 1896, Cartwright 1907, Virajbhakaya 1924, McFarland 1939, Haas 1947, 1964), textbooks and primers (*e.g.*, McFarland 1900, Frankfurter 1900, Cartwright 1906, Haas & Subhanka 1945, Noss 1964).

Reasons the history of Thai pronouns makes an interesting historical question include:

(1) A pronoun paradigm is robustly reconstructible for Proto-Tai, with reflexes across the full geographic spread of the Tai branch of the Kra-Dai family (Matsuyama 1962, Benedict 1975, Strecker 1984). This reconstructed paradigm has a three-way number distinction and three-way person distinction, plus clusivity distinction in the 1<sup>st</sup> person, and animacy distinction in the 3<sup>rd</sup> person. Thai retains just three of these in unmarked usage, indicating that Thai pronouns reflect relatively recent changes in the social structure of Thai society.

(2) Some facets of Thai pronominal usage beyond the paradigm are found cross-linguistically in Southeast Asia (*e.g.* Cooke 1965, comparing Thai, Burmese and Vietnamese). This indicates language contact as a major factor in the process of pronominal expansion, but we must also distinguish more recent areal phenomena from older ones.

Tracing changing social factors in the history of Thai pronoun usage helps us better understand how this kind of large typological shift can take place – from a closed pronoun paradigm to a relatively open system that relies heavily on situational context and awareness of one’s social relationship to everyone they interact with.

**Selected references**

Author. 2021. xxxxx

Benedict, Paul K. 1975. Austro-Thai language and culture with a glossary of roots. HRAF Press.

Cooke, Joseph R. 1965. Pronominal reference in Thai, Burmese, and Vietnamese. Ph.D. dissertation, University of California, Berkeley.

de La Loubere, Simon. 1693. A new historical relation of the kingdom of Siam [translated by A.P. Gen. R.S.S].

Low, James. 1828. A grammar of the T'hai, or Siamese language. Baptism Mission Press.

Matsuyama, Osamu. 1962. Kyôtsû taigo no ninshô daimeishi no taikai ni tsuite [On the system of personal pronouns in Common Tai]. Tokyo Gaikokugo Daigaku Ronshu 9:1-8.

Palakornkul, A. 1972. A socio-linguistic study of pronominal strategy in spoken Bangkok Thai. PhD dissertation, University of Texas at Austin.

Saisuwan, Pavadee. 2016. Male femininity in Thai among men who identify. PhD dissertation, Queen Mary University of London.

Strecker, David. 1984. Proto-Tai personal pronouns. Ph.D. dissertation, University of Michigan.



**The Postil Time Machine: “God help those who have begun writing down these books in Lithuanian”  
(BP1591 II 77,2–4)**

JOLANTA GELUMBECKAITĖ (Frankfurt/Main), ØYVIND EIDE (Köln), MORTIMER DRACH (Frankfurt/Main), CHRISTIAN CHIARCOS (Augsburg), MAX IONOV (Köln)

The *Postil Time Machine* (PosTiMe, 2021–2025) is a novel research project on knowledge transfer in 16th-century Europe; specifically, on the situation in Lithuania Minor, the Lithuanian-speaking part of the Duchy of Prussia. It aims to reveal the intratextual and intertextual relationships of the Old Lithuanian Lutheran postils – the *Wolfenbüttel Postil* (WP1573/1574) and the *Bretkūnas Postil* (BP1591) – and their relationship to their Latin and German translation sources and models. These postils are collections of sermons for the Sundays and feast days of the liturgical year (Gelumbeckaitė 2017). Early Modern postils as “the applied distillation of Christianity delivered on a regular basis by the clergy to the laity” (Frymire 2010: 443) were broadly influential texts offering references not only to Biblical texts and the Church Fathers, but also the classics, theologians and historians, bridging scholarly and popular culture. The main objective of PosTiMe is the genetic and historical linguistic study of these Lithuanian texts and their presentation in a digital *time machine*-type publication (cf. Kaplan 2013).

Previous work on Old Lithuanian in the Digital Humanities was focused mainly on digitizing Old Lithuanian texts and making them searchable (e.g. in TITUS and the *Senieji raštai* database). More recently, linguistically deeply annotated corpora have been created (SLIEKKAS, CorDon). The *Postil Time Machine* follows this corpus-linguistic tack while contributing new data, a systematic philological edition of its texts, and a visualization of inter-textual relations.

Specifically, this interdisciplinary linguistic-philological and technological project consists of the following tasks: (1) the identification of the construction principles and translation strategies of the Lithuanian texts, (2) the study of the textual relations of the postils both with the translation sources and among each other, (3) the detection of intra- and intertextual references as well as the alignment with the so-identified originals using machine translation methods (which necessitates the creation of a language technologies stack for Old Lithuanian), (4) the linguistic-historical interpretation of their contents in the form of a linguistically deeply annotated reference corpus, (5) the modeling as a graph (and the development of a Linked Open Data interface in RDFa based on and connected to the TEI documents, cf. Tittel et al. 2018 and Chiarcos & Ionov 2018) as well as the representation of these structures as interactive visualizations, and (6) the implementation of a platform to make the research results searchable, traversable and generally accessible. The scholarly digital edition of the postils is carried out in accordance with the principles formulated by Patrick Sahle, who stresses that “a digitised edition is not a digital edition” (Sahle 2016: 33).

## Literature

- Chiarcos, Christian, Maxim Ionov. 2018. *Linking the TEI: Approaches, Limitations, Use Cases*. Paper presented at DH2019, Utrecht, Niederlande.
- Frymire, John M. 2010. *The Primacy of the Postils. Catholics, Protestants, and the Dissemination of Ideas in Early Modern Germany*. Leiden, Boston: Brill (Studies in Medieval and Reformation Traditions 147).
- Gelumbeckaitė, Jolanta. 2017. Predigtkultur in Litauen: Corpus der altlitauischen Postillen. In: Assel, Heinrich, Johann Anselm Steiger, Axel E. Walter (Hg.). *Reformatio Baltica: Kulturwirkungen der Reformation in den Metropolen des Ostseeraums*. Berlin, Boston: De Gruyter, 573–585.
- Kaplan, Frédéric 2013. “How to build an information time machine”. Presented at TEDxCaFoscariU, Venice, Italy.
- Sahle, Patrick. 2016. “What is a Scholarly Digital Edition?” In: Driscoll, Matthew James, Elena Pierazzo (eds). *Digital Scholarly Editing. Theories and Practices*. Open Book Publishers.
- Tittel, Sabine, Helena Bermúdez-Sabel, Christian Chiarcos. 2018. “Using RDFa to link text and dictionary data for Medieval French.” In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 7–12. Miyazaki, Japan.

## Databases and Corpora

- CorDon, <https://titus.fkdg1.uni-frankfurt.de/cordon/menu/eng/start.html>
- Senieji raštai, <http://seniejirastai.lki.lt/home.php>
- SLIEKKAS, <https://titus.uni-frankfurt.de/sliekkas/index.html>
- TITUS, <https://titus.uni-frankfurt.de/indexd.htm>

**‘Chained to the rhythm’:  
Using agent-based simulation to model the evolution of  
stress pattern diversity in English**

English is characterized by a relative heterogeneity of stress patterns (e.g. *léntil* vs. *hotél*, *ábject* vs. *intáct*, *íncrease* [N] vs. *incréase* [V]). We lay out a usage-based explanation for the historical evolution of stress pattern diversity in languages such as English and present it in the form of an agent-based model. We find that the predictions derived from such a model are in line with diachronic corpus data.

In stress-based languages such as English, physiological and cognitive constraints (Lehiste 1970; Pitt & Samuel 1990; Peelle & Davis 2012) favor an alternating rhythm made up of sequences of stressed and unstressed syllables (Hayes 1984; Selkirk 1984; Schlüter 2005). These preferences can affect lexical stress diachronically by biasing words or entire word classes towards those patterns which most successfully produce alternating rhythmic sequences in combination with other words in language use (e.g. *the íncrease wórries us* vs. *híkes incréase the cóst of living*; cf. Kelly & Bock 1988, Kelly 1989). In evolutionary terms (Croft 2000), the rhythmic preferences operating at the level of phrasal phonology exert a selective pressure on lexical stress, constantly testing the viability of a pattern within its usage context.

We choose agent-based simulation (Wilensky & Rand 2015) as a method for probing this line of argumentation. The agent population in our model is made up of constituent types (i.e. a proxy for lexical items) defined by linguistic attributes, notably stress pattern, syllable weight and morpho-syntactic class. In each round of the simulation, a predetermined number of agents are probabilistically selected to occur and interact with one another within one of a range of possible syntagmatic contexts to form a rhythmic phrase. The phrase is evaluated with respect to prosodic criteria (rhythmic alternation and weight-to-stress) and the agents are rewarded or penalized accordingly. These payoffs continuously update the agents' fitness attribute, which in turn determines the agents' chances of successfully reproducing into the next generation.

The simulation suggests that stress pattern diversity will stably establish itself if the occurrence contexts of polysyllables also include monosyllabic material at a sufficiently high rate. In such a setting, diverse rather than uniform lexical stress patterns will reduce the probability of rhythmically suboptimal clashes and lapses. This prediction matches diachronic data from the Penn-Helsinki Parsed corpora of English (Kroch & Taylor 2000; Kroch, Santorini & Delfs 2004; Kroch, Santorini & Diertani 2016).

## References

- Croft, William. 2000. *Explaining language change: an evolutionary approach* (Longman Linguistics Library). Harlow, England ; New York: Longman.
- Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15(1), 33–74.
- Kelly, Michael H. 1989. Rhythm and language change in English. *Journal of Memory and Language* 28(6), 690–710. doi:10.1016/0749-596x(89)90004-1.
- Kelly, Michael H. & Bock, J. Kathryn. 1988. Stress in time. *Journal of Experimental Psychology* 14(3), 389–403. doi:10.1037/0096-1523.14.3.389.
- Kroch, Anthony, Santorini, Beatrice, & Delfs, Lauren. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony, Santorini, Beatrice, & Dierani, Ariel. 2016. The Penn Parsed Corpus of Modern British English (PPCMBE2). Department of Linguistics, University of Pennsylvania.
- Kroch, Anthony & Taylor, Ann. 2000. The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Department of Linguistics, University of Pennsylvania.  
<http://www.ling.upenn.edu/ppche-release-2016/PPCME2-RELEASE-4>.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, Mass: MIT Press.
- Peelle, Jonathan E. & Davis, Matthew. 2012. Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology* 3. doi:10.3389/fpsyg.2012.00320.
- Pitt, Mark A. & Samuel, Arthur G. 1990. The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance* 16(3), 564-573.
- Schlüter, Julia. 2005. *Rhythmic grammar: the influence of rhythm on grammatical variation and change in English*. Berlin; New York: Mouton de Gruyter.
- Selkirk, Elisabeth. 1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.
- Wilensky, Uri & Rand, William. 2015. *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Cambridge, Mass.: MIT Press.

## **Fall of the jers: A multi-factorial analysis of the sound change progression in the Old Novgorodian birchbark texts**

The graphics of the birch bark texts produced in Novgorod from the 11th to the 13th centuries show effects of the fall of the jers in the weak position in that the jers in word-initial and word-medial syllables of word forms are rendered inconsistently, or not at all (Janin / Zaliznjak 1993; Zaliznjak 2004). According to Eom et al. (2004), the progression of the underlying sound change process, as reflected in these texts, is associated with the process of optimization of word structure. However, it is still an open question whether and how the position of weak jers in the phonological word influenced the progression of their fall (Zaliznjak 2004: 63). This paper investigates the progression of this process in birch bark texts as a function of structural and usage-based factors.

The data were coded and analyzed with respect to the following variables: (estimated) time of text creation; jer position in the phonological and morphological word; length of the phonological and morphological word (in syllables); status of the morpheme containing the weak jer; alternation of the target form with inflectional forms having strong jers in the word paradigm; nature and frequency of the resulting consonant clusters (if present in the lexicon). A multifactorial analysis of the data was performed using logistic regression with the random variables “text or author” and “time” (cf., Baayen, 2008; Szmrecsanyi, 2013).

The analysis demonstrates that the fall first affects the jers in the initial syllables of morphological words, then the jers in the word-medial syllables of morphological words, and finally the jers in the proclitics. In addition to the random variable “time”, the writing or non-writing of the weak jers in the word-initial and the word-medial syllables is determined by (a) the status of the morpheme containing the jer (suffixes are affected earlier than prefixes and prepositions), (b) the resulting consonant cluster’s frequency and (c) the cluster’s phonological makeup. The progression of the fall of jers can thus be explained by an interaction of structural – both morphological and phonological – and usage-based factors.

### **References**

- Baayen, Harald R. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.
- Eom, Jeehyeon / Werner Lehfeldt / Peter Meyer. 2004. Der Jerausfall und seine Folgen für die Wortstruktur der slavischen Sprachen. *Zeitschrift für Slavische Philologie* 63, 249-309.
- Janin, Valentin L. / Zaliznjak, Andrej A. 1993. *Novgorodskie gramoty na bereste (iz raskopok 1984-1989 gg.)*. Moskva.
- Szmrecsanyi, Benedikt. 2013. The great regression: Genitive variability in Late Modern English news texts. In: Borjars, Kersti, David Denison, and Alan Scott (eds.) *Morphosyntactic categories and the expression of possession*, 59–88. Amsterdam: Benjamins.
- Zaliznjak, Andrej A. 2004. *Drevnenovgorodskij dialekt*. Moskva: Jazyki slavjanskoj kultury.

## Genesis of the Japanese Compound Particles

The Japanese compound particles (e.g., *ni tai shite* ‘towards’, *ni totte* ‘for’, *wo motte* ‘by means of’, etc.) represent a salient feature of the modern standard Japanese language. They typically follow the pattern case particle + verb in a converbal (gerund) form and fulfill functions similar to the secondary prepositions in European languages, such as the English prepositions *about*, *against*, *along*, etc.

Ex.: *sensei-ni-mukat-te*                                    *shitsurei-na koto-o*    *iwa-nai-de kudasai*.  
 Teacher-**DAT-head.towards-CVB**    rude-ADN matter-ACC say-NEG-CVB please  
 ‘Please don’t say rude things **to** your teacher.’

However, this has not always been the case. There is a significantly lower number of these compounds present in older texts, which means that most of them clearly grammaticalized during the historical period. In this paper, their genesis throughout history will be tracked using the extant textual evidence and the major factors contributing to their proliferation will be discussed. Some of those compound particles were clearly influenced by certain Chinese constructions. The key to their genesis seems to be the traditional Japanese reading of Chinese texts - the so-called *kambun kundoku*. This seems to be true not only for compound particles containing morphemes of Chinese origin, but also for some purely Japanese compounds. The Chinese expressions on which they were patterned (some of which are used as coverbs even nowadays in Modern Chinese) will be presented and the grammaticalization clines explained with ample examples. The aim is to present a balanced view of the compound particles taking into account both the contact influence and the internal factors of the development of Japanese.

### Selected bibliography

- AIKHENVALD, Alexandra Y., et al. Grammars in contact: A cross-linguistic perspective. *Grammars in contact: A cross-linguistic typology*, 2007, 4: 1.
- ENDO, Mitsuaki. Language contact between Chinese and Japanese. *The Oxford Handbook of Chinese Linguistics*, 2015, 215.
- FRELLESVIG, Bjarke. *A history of the Japanese language*. Cambridge University Press, 2010.
- HARBSMEIER, Christoph. Thesaurus Linguae Sericae: An Historical Encyclopaedia of Chinese Conceptual Schemes. *Online at <http://tls.uni-hd.de/>. Accessed (1.09. 2018).*
- HEINE, Bernd; KUTEVA, Tania. On contact-induced grammaticalization. *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”*, 2003, 27.3: 529-572.
- PO-CHING, Yip; RIMMINGTON, Don. *Chinese: A comprehensive grammar*. Routledge, 2015.
- PULLEYBLANK, Edwin G. *Outline of classical Chinese grammar*. Ubc Press, 2010.

## Exploring the sources of animacy distinctions

**Keywords:** Diachronic morphology, Animacy Hierarchy, source- and result-oriented explanations, noun inflection

Animacy is a semantic category that influences the grammatical structure of languages in different ways. Its manifold effects have been linked to the functioning of a typological hierarchy, called either Nominal or Referential or Animacy Hierarchy, which has been claimed to be a true universal (Kiparsky 2008), not just a widely held typological generalization. Nevertheless, recent research on the sources of several animacy markers (Cristofaro 2013, 2019) has shown that at least some of them do not arise as morphemes specifically signaling animacy distinctions or that the function that they serve synchronically is preconditioned by the circumstances of their origin (in the Indic language Maithili, for instance, the pluralizer *lokain* is used only with human nouns presumably as a consequence of its original lexical meaning being ‘people’).

This source-oriented approach reveals that the association of certain morphemes with animacy can be regarded as secondary or contingent, a mere accident of their historical evolution. If this were the case of every morphological marker of animacy, the significance of the Animacy Hierarchy as a typological generalization (or even universal, be it absolute or statistical) would somehow be compromised, at least at a certain level of analysis (see Cristofaro & Zúñiga 2018). Some of the examples that we will examine in this paper are certainly suggestive of grammaticalization (or regrammaticalization) processes having led to the present animacy-based values (as occurred with the emergence of specialized inflectional markers such as the nominal dative singular case in *-ovi* in Slovak and other Slavic languages, formerly a common ending of a specific declension class).

However, the documented (or else presumed) existence of shared paths of change across languages as well as the convergent development of animacy-based innovations (materialized in similar asymmetries in the expression of plurality, differential object marking, animate first ordering in syntax, etc.) appear to demand a more comprehensive explanation, beyond the specific one referring to the history of each individual marker. Actually, this convergence in the outcomes, captured by such generalizations as the Animacy Hierarchy itself, suggests interpreting animacy as a fundamental cognitive (and linguistic) property, which is also evidenced by cognitive and psychological experiments (Jones et al. 1991, New et al. 2007, Trompenaars et al. 2021).

The diachronic coherence in the rise of animacy distinctions is a tendency identifiable both in morphological markers of the category (animacy as a feature) and in its behavior as a factor conditioning the selection of values of other features (animacy as a condition; cf. Santazilia 2020). Moreover, the grammatical consequences of the manifestation of animacy can ultimately be explained in terms of identical selective pressures and analogous evolutionary responses (Haspelmath 2019: 15), in much the same way as some biological traits (the wings of birds, bats, insects and old pterosaurs, for instance) are considered to have arisen for similar functional reasons, in spite of the fact that they have rather different evolutionary histories. Thus some general, overarching principles may turn out to be responsible for the attested structural similarities among languages (which would explain why universally preferred constructions may emerge even without any preconditions in the grammaticalization source, see Seržant & Rafiyenko 2021).



## References

- Cristofaro, Sonia. 2013. The referential hierarchy: reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries. Studies in memory of Anna Siewierska*, 69–93. Berlin & New York: De Gruyter Mouton.
- Cristofaro, Sonia. 2019. Taking diachronic evidence seriously: result-oriented vs. source-oriented explanations of typological universals. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne M. Michaelis & Ilja A. Seržant (eds.), *Explanation in typology. Diachronic sources, functional motivations and the nature of the evidence*, 25–46. Berlin: Language Science Press.
- Cristofaro, Sonia & Fernando Zúñiga. 2018. Synchronic vs. diachronic approaches to typological hierarchies. In Sonia Cristofaro & Fernando Zúñiga (eds.), *Typological hierarchies in synchrony and diachrony*, 3–27. Amsterdam & Philadelphia: John Benjamins.
- Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne M. Michaelis & Ilja A. Seržant (eds.), *Explanation in typology. Diachronic sources, functional motivations and the nature of the evidence*, 1–23. Berlin: Language Science Press.
- Jones, Susan S., Linda B. Smith & Barbara Landau. 1991. Object properties and knowledge in early lexical learning. *Child development* 62 (3). 499–516.
- Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalization. In Jeff Good (ed.), *Linguistic universals and language change*, 23–53. Oxford: Oxford University Press.
- New, Joshua, Leda Cosmides & John Tooby. 2007. Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 104 (42). 16598–16603.
- Santazilia, Ekaitz. 2020. The two faces of animacy. *Studies in Language* 44 (4). 812–830.
- Seržant, Ilja A. & Dariya Rafiyenko. 2021. Diachronic evidence against source-oriented explanation in typology: Evolution of prepositional phrases in Ancient Greek. *Language Dynamics and Change* 11. 167–210.
- Trompenaars, Thijs, Teresa A. Kaluge, Rezvan Arabi & Peter de Swart. 2021. Cognitive animacy and its relation to linguistic animacy: evidence from Japanese and Persian. *Language Sciences* 86 (101399): 1–17.

***Tone split and tone replacement:  
toward the three-tone system of the ‘Western’ SBB Languages  
(Central Sudanic, Central Africa)***

Pascal Boyeldieu (UMR 8135, CNRS, INaLCO, EPHE)

**Abstract**

*Sara-Bongo-Bagirmi (SBB)* languages represent a group of some 40 African languages that are scattered between Lake Chad in the North-West, and Lake Albert in the South-East, thus covering parts of Chad, Sudan, South-Sudan, the Central African Republic, and the Democratic Republic of Congo. The *SBB languages* form the largest sub-branch of Greenberg’s (1963) *Central Sudanic (Nilo-Saharan phylum)* and are geographically inserted among various languages or linguistic groups such as Chadic, Adamawa, Ubangian, other Central Sudanic, Eastern Sudanic, and Arabic.

While languages in the East display two-tone systems directly reflecting the historical \*SBB configuration, a large subgroup of ‘western’ languages (\*OCC for *occidental*) later innovated in developing a new tone in the high frequencies. More recently the western languages Yulu and Gula Koto independently underwent a splitting of their low tone, thus creating a now phonologised extra-low level. The main correspondences between the different types of systems are summarised in Table 1. In order to avoid possible ambiguities (e.g. H tone, but in which system?), tone levels will from now on be symbolised by numbers (0/1/2/3) as indicated below:

*SBB & eastern languages 2 tones	>	*OCC & western languages 3 tones	>	Yulu & Gula Koto 4 tones
High (2)		High (3)		High (3)
Low (1)		Mid (2)		Mid (2)
		Low (1)		Low (1)
				X-Low (0)

Table 1. Tone systems and tone level numbering

If it has been necessary for the following to mention the fourth tone of Yulu and Gula Koto, this paper is nevertheless specifically concerned with the emergence of a third tone level in the ‘western’ sub-branch, a change that affected a fair number of languages and played an important role in the history of this linguistic group. Moreover this transformation has followed different ways according to the grammatical category – verbs or nouns – it affected and therefore offers some particular interest for the general understanding of tone change.

*Verbs*

Most reconstructible \*SBB verbs display an original VCV form (with current reflexes of a VCV, sometimes CV or CVCV shape) that is associated with a tone pattern defined by two tonal components (e.g. 12 for *àpá* or *àpā*, 31 for *ápà*, etc.). In all languages the most frequent verb form – usually labelled as *definite aspect*, *aorist* or *perfective* – undergoes tone alternations conditioned by the nature of the subject and determining several tone classes.

Let us illustrate this specific behaviour with the case of Modò, an eastern language that is clearly the best representative of the common \*SBB situation under the respect of the verb morphology. Modò has four verb classes resulting from the different combinations of two tone highs in two-syllable patterns : A (12), B (22), C (11), and D (21). These tone patterns remain unchanged in the presence of a lexical subject. However a subject index of 3<sup>d</sup> person lowers the first component of the tone pattern (1, 2 > 1) while a 1<sup>st</sup>-2<sup>d</sup> person subject index raises this initial component (1, 2 > 2). Consequently the four-way contrast of the isolated verbs is reduced to two when they are preceded by a personal index:<sup>1</sup>

---

<sup>1</sup> The same patterns appear with plural indexes, although with a different distribution.

Class	A	B	C	D
Lexical subject	12	22	11	21
3 <sup>d</sup> subj. index		12		11
1 <sup>st</sup> -2 <sup>d</sup> subj. index		22		21

Table 2. Tone classes of Modo

Also each class is identified by three alternating tone patterns (e.g. : class A = 12//12/22, class B = 22//12/22).

Now, in all other languages, the lexical subject for some reason was aligned with the 3<sup>d</sup> subject index pattern so that, in the eastern languages at least, the verbs were, in all contexts, reduced to two classes, namely AB (12/22) and CD (11/21).

In the ‘western’ languages, the same happened for the reflexes of \*SBB\*A and \*SBB\*B that merged in \*OCC\*AB (\*12/\*22). Yet the contrast in the reflexes of \*SBB\*C and \*SBB\*D2 was not only preserved but even reinforced by enlisting a new \*OCC\*31 tone pattern and generating, by analogy with the \*11/\*21 alternation of \*OCC\*C, a specific \*21/\*31 alternation for \*OCC\*D verbs as shown in the correspondences displayed in Table 3:<sup>2</sup>

*SBB	*A *12/*12/*22	*B *22/*12/*22	*C *11/*11/*21	*D1 *?	*D2 *21/*11/*21
*OCC	*AB *12/*22		*C *11/*21	*D *21/*31	

Table 3. Morphological function of tone level 3 at the \*OCC level

Appendix A. displays some lexical illustrations of \*SBB\*C and \*SBB\*D2 reflexes.

### Nouns

In nominal comparative series, ‘western’ languages may show, in a quite irregular way, one or the other of four patterns involving a level 3 tone and reconstructible, at an \*OCC level, as \*31, \*32, \*13 or \*23. Reflexes of the latter may occur while other ‘western’ languages preserve the tonal reflex of an original \*SBB and/or \*OCC formula.

Some lexical illustrations of such irregularities are given in Appendix B.

The best explanation one can offer for such a situation is the following: the ‘western’ languages having undergone an important lexical replacement, probably due to contact with (an) up to now unidentified three-tone language(s), numerous nouns defined in terms of three contrastive tones integrated as such the \*OCC language system. With the possible contribution of the tone changes simultaneously affecting verbs, these new patterns contaminated the older nominal lexicon in form of free variants that were progressively eliminated, each ‘western’ language – or subgroup of ‘western’ languages – finally preserving one or the other of the two competitive tone patterns, inherited from either \*SBB or \*OCC.

### Outcomes

In verbs, the \*OCC class \*D formula \*31 represents a genuine reflex of \*SBB \*21. Moreover, it has to be emphasised that the shift \*SBB \*21 > \*OCC \*31 was constrained by the morphological alternation role played by the tones in the frame of class \*D verbs. With class \*C verbs the same original tone pattern remained unchanged: \*SBB \*21 > \*OCC \*21.

In nouns, the \*OCC formulas \*31, \*32, \*13, and \*23 are new patterns that do not result from any prior unit. In case they affect nouns dating to the original \*SBB vocabulary, they strictly represent replacements of the previous tone patterns still attested in the eastern languages. In this latter case the ‘western’ form of a noun combines older segments with new tones, thus revealing a partial loss and discontinuity in its history since its \*SBB origin.

Nevertheless, the two processes, probably reinforcing each other, jointly contributed to the emergence of a new \*OCC system that led to the current ‘western’ languages.

<sup>2</sup> The \*SBB\*D1 formula calls for comments that I shall not mention in this abstract.

The presentation is directly inspired by a historical comparative study of the SBB tone systems (Boyeldieu 2000). It will provide comparative data (also available in Boyeldieu, Nougayrol & Palayer 2006: <https://llacan.cnrs.fr/SBB/>) and highlight the linguistic and geographical conditions of the tonal change.

Boyeldieu, Pascal. 1996. Mutation et réfection dans l'expansion des registres tonals : l'exemple des langues sara-bongo-baguirmiennes. *Revue de Phonétique Appliquée* (Université de Mons-Hainaut), 120, p.163-178.

———. 2000. *Identité tonale et filiation des langues sara-bongo-baguirmiennes (Afrique centrale)*. Köln: Rüdiger Köppe Verlag (SUGIA, Beiheft 10).

Boyeldieu, Pascal, Pierre Nougayrol & Pierre Palayer. 2006. *Lexique comparatif historique des langues sara-bongo-baguirmiennes*. Paris : CNRS-LLACAN. Electronic resource: <https://llacan.cnrs.fr/SBB/>

Greenberg, Joseph H. 1963. Languages of Africa. *International Journal of American Linguistics*, 29, 1, part II. [2<sup>d</sup> ed.: 1966. *The Languages of Africa*. Indiana University, Bloomington, La Haye: Mouton]

---

	Eastern languages			*SBB*C (*11//*11/*21)				
	C Modo 11//11/21	CD Bongo (1)1/(2)2 <sup>1</sup>	CD Baka 11/21	C Yulu 11=00/21=20	ABC Gula Koto 1(1)/2(2)	CD Na 11/31	CD Sar 11/31	
V/069	òlè//òlè/ólè	yà/yá <sup>1</sup>	òyò/óyò	ùlè/ùlè	òy/òy	ò/ò	òy/òy	‘die’
V/071	òpò//òpò/ópò	òp/óp <sup>1</sup>	ànò/ánò	òpò/òpò	òp/òp	òpò/òpò	òp/òp	‘eat’
V/152	àtò//àtò/átò	òtù/òtù <sup>1</sup> ‘rot’		ààcè/ààcè	àt/àt	átù/átù	òtè/òtè	‘smell (bad)’

	Eastern languages			*SBB*D2 (*21//*11/*21)				
	D Modo 21//11/21	CD Bongo (1)1/(2)2 <sup>1</sup>	CD Baka 11/21	D Yulu 21/31	D Gula Koto 10/31	CD Na 11/31	CD Sar 11/31	
V/112	úpò//úpò/úpò		òfò/òfò	òfò/òfò				‘kill’
V/115	ádí//ádí/ádí	èdì/èdì <sup>1</sup>	(edj)	āādā/āādā	èd/èd <sup>1</sup>	àri/àri	èdà/èdà	‘rain’
V/111	átì//átì/átì	ètì/ètì <sup>1</sup>	ècì/ècì		èt/èt <sup>1</sup>	àtì/àtì		‘sneeze’

Appendix A. Lexical illustrations of \*SBB\*C and \*SBB\*D2 reflexes

	Eastern languages			*SBB*11	*OCC*11 (or *31)			
	Modo 11	Bongo 11	Baka 11	Yulu 11=00	Gula Koto 11	Na 11	Sar 11	
N/0017	kòmò	kòmò	kòmò	(kām(ə) !?)	kām	kām	kùm	‘eye, face’
N/0462	kòtò	kòtò	(mokoto)	kòt(ə)				‘sterility, sterile’
N/0536		kùdi		kēj(ə)	kèɖ	kùjè	kàdè	‘worm, grub, maggot’
N/0029	dòkòròsò (dò-kò ròsò ?)		(kési)		<sup>31</sup> kéc’	kòkè	kòsə	‘iron, hoe’
N/0019	(gáli !?)	gèl		gääl(ə)	gèl	<sup>31</sup> gáli	gèl	‘left (side)’
N/0353	tibò			vööv(ə)		<sup>31</sup> júbù	<sup>31</sup> ndúbè	‘bellows’

	Eastern languages			*SBB*12	*OCC*12 (or *OCC*31/*32/*13)			
	Modo 12	Bongo 12	Baka 12	Yulu 12=02	Gula Koto 11	Na 12	Sar 12	
N/0083	máá	máhá	(mbasa)	máàs(ə)	màs	màsā	màsə	‘ <i>Tamarindus indica</i> ’
N/0094	bàďú	bòďú	(uəɖu)	bääď(ə)	váď	bārū	bòr	‘warthog’
N/0107	ǫú	hiǫú	sùǫú		kùǫ	bū	yibə	‘oil, (grease)’
N/0112	kánó				<sup>32</sup> kún	<sup>32</sup> kúnū	<sup>31</sup> kón	‘nose’
N/0058	(tòkpè !?)	tíkí			tít	tihī, <sup>13</sup> tihí	tī	‘bowels’
N/0097	(kòwé ?)	kùlǫhí			cic	kòkī	<sup>13</sup> kòsə	‘cucumber’

	*SBB*22			*OCC*22 (or *OCC*31/*32/*23)				
	Eastern languages	Modo 22	Bongo 22	Baka 22	Yulu 22	Gula Koto 22	Na 22	
N/0196	kénzè	kínjí	kénzè	kēēnj(ə)	kānz	kānjē	kānjə	‘fish’
N/0198	népé	níhí	éfé	nēēp(ə)	lēhē	nōhē	nāā	‘moon’
N/0203	kúpó	kúhú	kófó	kōōf(ə)	kōhō	kōhō	kōō	‘seed’
N/0227		hídó	(sida)	sūūj(ə)	<sup>32</sup> sód̄		yāda	‘ <i>Anogeissus leiocarpus</i> ’
N/0433	yóri			<sup>31</sup> sóór(ə)		<sup>23</sup> kīró	<sup>23</sup> yāró	‘dirt’
N/0223	kólógbé	higé	(sige)	<sup>32</sup> síg(ə)	<sup>32</sup> kéd̄	<sup>32</sup> kóbē	<sup>32</sup> yégā	‘rat, mouse’

Appendix B. Irregular occurrences of level 3 tones in ‘western’ languages



**Talk: Reconstructing the Kugama tone system**

**Author: Lora Litvinova**

The paper presents the first results of the internal reconstruction of the tone system of Kugama, an Adamawa language of Nigeria. The Kugama tone system distinguishes three tone levels: High (H), Mid (M), and Low (L). I argue that the three-level tone system in Kugama originates from a two-level tone system. The current three levels have emerged primarily through the phonologization of an earlier phonetic raising of H and L in certain environments. Thus, the H level of the current three-level tone system goes back to the phonetic raising of H before a tautosyllabic L. For instance, this is suggested by the following synchronic tonal rule: M becomes H before a floating <sup>L</sup>, as in (1). The H tone of the two-level system that did not precede a tautosyllabic L was reinterpreted as M in the present-day Kugama tone system.

- (1)
- |         |                  |  |                     |
|---------|------------------|--|---------------------|
| /vélí   |                  |  | ˈɲōōkī/             |
| vélí    | L                |  | ɲōōkī  <sup>1</sup> |
| husband | GEN <sup>2</sup> |  | daughter            |
| ‘groom’ |                  |  |                     |

The L tone of the two-level system was equally phonetically raised before a raised H and subsequently this phonetically raised L was also reinterpreted as M. For example, nouns with an MH<sup>L</sup> tone pattern in their Free Form (main allomorph) have an LM tone pattern in their Construct Form 1 (a secondary allomorph): *kǎm* |kǎm<sup>L</sup>| ‘body hair’ and *kǎm* ‘body hair.CF1’, as in (2).

- (2)
- |                   |  |           |
|-------------------|--|-----------|
| /kǎm              |  | púūpī/    |
| kǎm               |  | púūpī     |
| body_hair.CF1     |  | whiteness |
| ‘white body hair’ |  |           |

I will discuss other relevant phonetic and phonological properties of the current Kugama tone system as well as provide further evidence for the proposed internal reconstruction.

---

<sup>1</sup> Vertical bars |a| represent the morphophonological level.

<sup>2</sup> CF1 - Construct Form 1, GEN - Genitive

## VSO orders in the *Egeriae* and *Antonini Placentini itineraria*; new evidence for the evolution towards Old Romance inversion systems

It is well documented that the Old Romance languages featured more subject-verb inversion than their modern descendants. Although many modern varieties of Italo-Romance and Ibero-Romance allow inversion with unaccusative verbs, inversion with transitive verbs is heavily constrained and not freely available outside particular constructions, such as for examples wh-questions (Rizzi 1990). In most documented varieties of Old Romance on the other hand, inversion is found with all types of verbs without distinction, a fact that has prompted some historical Romanists to consider these varieties 'V2 systems' because of the similarity with Modern Germanic V2 languages (Vance 1997; Poletto 2014; Wolfe 2018).

The origin of the Old Romance inversion systems is poorly understood, but their wide diatopic distribution suggests internal development within the Latin/Romance-family. Evidence has been sought in Late Latin, and one text in particular, the *Itinerarium Egeriae*, has attracted much attention. In this late 4th century text, VSO order is quite widespread, leading Ledgeway to the conclusion that the position of the verb is already the same as in Old Romance (Ledgeway 2017)

I will present data showing that this claim is too strong. A full quantitative and qualitative analysis of *Itinerarium Egeriae* reveals that the VSO orders are not the result of very high verb movement, but rather arise through a very low position of the subject. Furthermore, I will add data from a complete analysis of a text that has not featured prominently in the debate on the evolution of word order, namely the late 6th century *Itinerarium Antonini Placentini*. While VSO-order is quantitatively even more robust in this text than in *Egeria*, closer qualitative analysis shows that inversion is almost exclusively found with intransitive verbs and passives:

- (1) *Illic currit fluvius Asclepius*  
there runs river.NOM Asclepius.NOM  
'There the river Asclepius runs'
- (2) *Super his locis descendit ros sicut pluvia*  
Over this places.ABL descends dew.NOM as rain.NOM  
'Over these lands a dew falls like a rain'
- (3) *In qua etiam synagoga posita est trabis*  
In which indeed synagogue placed is beam  
'In this synagogue a beam is placed'

The combined evidence of these textual witnesses suggests that, while we need not abandon the hypothesis of a Romance-internal evolution of inversion, we should reconsider its diachrony. The new data from the youngest text reveals a new alignment system for the arguments of the verb which is sensitive to theta-roles, with thematic arguments liberally appearing in postverbal position, while agent arguments (in transitives) favour the preverbal position. If correct, this analysis entails that generalised inversion in Old Romance cannot have developed as early as the late 4th century, but must rather be postponed until at least the 7th century, thereby adding a new important piece to our understanding of the diachrony of word order in Romance.

## References

- Ledgeway, Adam (2017). "Late Latin Verb Second: The Sentential Word Order of the *Itinerarium Egeriae*". In: *Catalan Journal of Linguistics*. Catalan Journal of Linguistics 16. Ed. by Jaume Mateau Fantanals and Renato Oniga, pp. 163–216.
- Poletto, Cecilia (2014). *Word Order in Old Italian*. Oxford: Oxford University Press.
- Rizzi, Luigi (1990). "Speculations on verb second". In: *Grammar in Progress*. Ed. by Joan Mascaró and Marina Nespór. Walter de Gruyter, pp. 375–386.
- Vance, Barbara S (1997). *Syntactic Change in Medieval French : Verb-Second and Null Subjects*. Dordrecht: Kluwer.
- Wolfe, Sam (2018). *Verb Second in Medieval Romance*. Oxford: Oxford University Press.

## The Greek suffix -θ- and the Caland System

Andrea Santamaria (University of Bologna)

Via Zamboni, 32 – 40123 Bologna (BO), Italy

andrea.santamaria3@unibo.it

The present work addresses the distribution of the Greek verbal suffix -θ-, its connection to the Caland System and its development from the Homeric Greek up to the Byzantine period.

The presence of forms showing a Caland behavior in Greek was recognized as early as Wackernagel (1897), only few years after the related observations of Caland (1892; 1893) himself on Indo-Iranian languages. After more than a century of studies, Greek proved to match a high number of Caland suffixes attested elsewhere (cf. Ritsch 1974<sup>2</sup>: 66; Meißner 1998; Bichlmeier 2014). These suffixes commonly contribute to deriving adjectives (e.g. ἐρυθρός ‘red’ < \**h<sub>1</sub>rud<sup>h</sup>-r<sup>o</sup>-*), nouns (e.g. ἔρευθος ‘redness’ < \**h<sub>1</sub>reud<sup>h</sup>-e/os-*) and verbs (e.g. ἐρέυθω ‘I make/become red’ < \**h<sub>1</sub>reud<sup>h</sup>-eh<sub>1</sub>-*). Among them, the suffix -θ- was tentatively associated with some Caland formations (e.g. Nussbaum 1976: 90 and Rau 2009: 152-153 fn. 80). Still, no definitive evidence was put forward in favor its attribution to the Caland System nor its behavior *vis-à-vis* the other Caland suffixes was clearly detected.

Since Benveniste (1935: 188-210), a suffix -θ- was recognized in a wide range of forms, such as adjective abstracts (e.g. πλῆθος ‘multitude, quantity’), *nomina agentis* (e.g. τένθης ‘gourmand’), verbs (mostly θω-presents such as βρίθω ‘I am heavy’) and adverbs (e.g. μίνυνθα ‘a short time’). The adjective abstract μέγεθος ‘greatness’ was analyzed as deriving a property concept adjective and would therefore speak in favor of a ‘marginal’ Caland \**-d<sup>h</sup>o-* (cf. Nussbaum 1976: 90). On the other hand, few verbal forms such as λήθω ‘I hide’ were traced back to constructions employing the function verb \**d<sup>h</sup>eh<sub>1</sub>-* (cf. Hackstein 2002 and Schutzeichel 2014). However, as a complete investigation of forms in -θ- and their Indo-European relatives is still pending, both the typology of derived stems and their position within the Caland System remain opaque.

Thus, the present work focuses on (a) the typology, from both a morphological and a semantic perspective, of stems showing a derivative in -θ- and their diachronic developments; (b) the Indo-European cognates of -θ- and the role they played within the Caland System. Accordingly, I argue for the following classes:

a) Verbs deriving property concept roots showing both other Caland formations and, sometimes, root aorists. This group is the oldest one. Moreover, a PIE suffix \**-d<sup>h</sup>-* (perhaps ultimately traceable back to a grammaticalized \**d<sup>h</sup>eh<sub>1</sub>-*) with such roots would find parallels outside Greek (e.g. Gr. πλήθω ‘I am full’, Av. *frāda-* ‘thrive’ and Lat. *plēbēs* ‘common people’). Semantically, these verbs are commonly the inchoative member of a causative/inchoative alternation based on the opposition of derived stems (e.g. βρίθω ‘I am laden with’ : βριάω ‘I make strong’). Although such a typology of alternation was largely neutralized in Homeric Greek (cf. πύθω ‘I cause to rot’), it is revealingly traceable back to the Caland System.

b) Deverbal and denominal verbs in -θ- without Caland cognates and analogical to Caland forms in -θ-. In fact, up to the Byzantine period, new (inchoative) presents in -θω were mostly built on thematic stems either from *s*-stem nouns (e.g. τελέθω ‘I come to end’ : τέλος ‘coming to pass’) or *e*-graded presents (e.g. φλεγέθω ‘I burn up’ : φλέγω ‘I burn sth.’). In Homeric Greek, few θ-aorists are opposed to thematic aorists (e.g. ἔσχεθον : ἔσχον ‘I had’) and could therefore be directly related to the emergence of θ-passives.

c) Re-derived nouns. The vast majority are abstracts in -θεσ- analogical to μέγεθος. They are variously shaped from non-Caland property concept roots (e.g. ὄχθος ‘eminence’), verbs (e.g. κέλευθος ‘road, path’) or θω-presents (e.g. λήθος ‘forgetfulness’). Nevertheless, these nouns often belong to new ‘paradigms’ derived *einzel-sprachlich* through suffixes previously employed within the Caland System (e.g. ἔσθος ‘garment’, ἑάνος ‘fine robe’ and ἔννυμι ‘I put clothes on another’). The same behavior is to be found for *nomina agentis* in -θης and *nomina actionis* in -θη, which can partner with abstracts in -θεσ- too (e.g. ἐσθής ‘raiment’ and λήθη ‘forgetting’). Unsurprisingly, such suffixes were also employed to derive new adjectives from nouns in -θ- (e.g. λαθραῖος ‘secret’).

### Literature

Benveniste, E. 1935. *Origines de la formation des noms en indo-européen*. Paris.

Bichlmeier, H. 2014. Caland System and Greek. In: G. Giannakis, ed., *Encyclopedia of Ancient Greek Language and Linguistics*. Pp. 256-260. Leiden-Boston.

- Bozzone, C. 2016. The Origin of the Caland System and the Typology of Adjectives. *IEL* 4: 15-52.
- Caland, W. 1892. Beiträge zur Kenntniss der Avesta, no. 19. *KZ* 31: 266-268
- Caland, W. 1893. Beiträge zur Kenntniss der Avesta, no. 26. *KZ* 32: 592
- Hackstein, O. 2002. Uridg. \*CH.CC > \*C.CC. *HS* 115/1: 1-22.
- Hackstein, O. 2012. When Words Coalesce: Chunking and Morphophonemic Extension. In H.C. Melchert, ed., *The Indo-European Verb*. Pp. 87-104. Wiesbaden.
- Kölligan, D. 2018. Funktionsverbgefüge und Sekundärwurzeln. In E. Rieken *et alii*, ed., *100 Jahre Entzifferung des Hethitischen. Morphosyntaktische Kategorien in Sprachgeschichte und Forschung*. Pp. 219-237. Wiesbaden.
- Lehmann, W.P. 1942. The Indo-European *dh*-determinative in Germanic. *Language* 18: 125-132.
- Magni, E. 2010. L'evoluzione semantico-funzionale dell'elemento -θ- nella morfologia verbale del greco. In I. Putzu *et alii*, ed., *La morfologia del greco tra tipologia e diacronia*. Pp. 266-28. Milano.
- Meißner, T. 1998. Das „Calandsche Gesetz“ und das Griechische - nach 100 Jahren. In W. Meid, ed., *Sprache und Kultur der Indogermanen : Akten der X. Fachtagung der Indogermanischen Gesellschaft, Innsbruck, 22.-28. September 1996*. Pp. 237-354  
Innsbruck
- Meißner, T. 2006. *S-stem Nouns and Adjectives in Greek and Proto-Indo-European. A Diachronic Study in Word Formation*. Oxford.
- Nussbaum, A.J. 1976. *Caland's "Law" and the Caland System*. PhD Dissertation, Harvard University.
- Rau, J. 2009. *Indo-European Nominal Morphology. The Decades and the Caland System*. Innsbruck.
- Ritsch, E. 1974<sup>2</sup> [1937<sup>1</sup>]. *Wortbildung der homerischen Sprache*. 2nd ed. Berlin-New York
- Schutzzeichel, M. 2014. *Indogermanische Funktionsverbgefüge*. Münster.
- Tucker, E. 1990. The Creation of Morphological Regularity: Early Greek Verbs in -έω, -άω, -όω, -ύω and -ίω. Göttingen.
- Wackernagel, J. 1897. *Vermischte Beiträge zur griechischen Sprachkunde*. Basel
- Willi, A. 2018. *Origins of the Greek Verb*. Cambridge.

## Continuative relative clauses in Greek documentary papyri

Although relativization and its connected phenomena have received great scholarly attention in Ancient Greek, the most extensive examinations of Greek relative clauses concern the Archaic (Probert 2015), and the Classical period (Perna 2013, Faure 2021). Within the Post-Classical period, some consideration has been paid to New Testament Greek (Du Toit 2016, Hayes 2018): in particular, a recent contribution by Du Toit (2022) investigates the characteristics and function of continuative relative clauses in the New Testament.

However, this type of clauses has not yet been studied in the Post-Classical Greek papyri, where relative constructions are still largely unexplored (Kriki 2013; Bentein and Bağrıaçık 2018; Bentein and Cattafi forthcoming), despite the sociolinguistic potential of these documents in the investigation of morphosyntactic variation (Bentein 2019; Bentein, Cattafi and La Roi forthcoming).

The aim of this paper is therefore to analyse the behavior of continuative relative clauses in Greek documentary papyri from Egypt, by taking into account letters, petitions and contracts from the first to the eighth century AD.

First, I will discuss how the papyrological examples relate to the general concept of continuative relative clauses as pointed out in linguistic studies (cf. Lehmann 1984, *inter alia*) and in the other periods of Greek. For instance, in these texts, very different constructions such as (1) from a petition and (2) from a contract can be broadly ascribed to the “continuative” category.

- (1) Συρίων γενόμενος δεκάπρωτος [ἀπὸ τῆς αὐτῆς] κώμης Θρασὼ ἀναπίσας μου τὸν ἄνδρα Καῆτ ὀνόματι ποιμένιν αὐτοῦ τὰ πρόβατα — ὅστις [ἀδίκως τὰς τοῦ] προκειμένου ἀνδρὸς αἴγας καὶ πρόβατα τὸν [ἀριθμὸν ἐξήκο]ντα συναπέσπασεν αὐτῷ

“Syrion having become decaprotus of the aforesaid village, persuaded my husband Ganis to pasture his flock—it was he who wrongfully removed into his own keeping my husband's goats and sheep 60 in number” (P. Sakaon 36, ll. 8-11 – III AD)

- (2) εἰς σπορὰν καὶ κατάθεσιν ὧν ἐὰν αἰρῶμαι φόρου ἀποτάκτου πυροῦ ἀρταβῶν τριάκοντα ἄνπερ φόρον μετρήσω τῷ Ἐπειφ μηνὶ τῆς (αὐτῆς) ζ' ἰνδικ(τίωνος) ἄνπερθέτως

“to sow and plant with whatever I choose, at a total rent of thirty artabas of wheat, which rent I shall measure out in the month of Epeiph of the same 7th indiction without delay” (P. Charite 7, ll. 11-15 – IV AD)

Second, I will highlight some linguistic aspects of continuative relative clauses in papyri, in particular their syntactic type and their relationship with the antecedent (Bentein and Bağrıaçık 2018), and the relative markers that introduce the clauses.

Finally, I will explore their communicative and stylistic meanings, such as the role of linking different textual units (cf. Tabachovitz 1943: 11 on the possibility of using relative clauses as discourse connectives), by looking at the type of documents where continuative clauses tend to appear and at the function they perform within the text.

## References

- Bentein, K. and M. Bağrıaçık. 2018. ‘On The Third Type of Headed Relative Clause in Post-Classical and Early Byzantine Greek’. *TPhS* 116 (3): 529–54.
- Bentein, K. 2019. ‘Dimensions of social meaning in post-classical Greek towards an integrated approach’. *JGL* 19 (2), 119-167.
- Bentein, K. and E. Cattafi (forthcoming). ‘Relative clauses’. In Giannakis, G., P. Filos, M. Janse, B. Joseph and I. Manolesou (eds.). *Encyclopedia of Greek language and linguistics*. Leiden: Brill.
- Bentein, K., E. Cattafi, and E. La Roi (eds.) (forthcoming). *(In)subordination in Post-Classical Greek: syntax, context and complexity*. Berlin: De Gruyter.
- Faure, R. 2021. *The Syntax and Semantics of  $\omega$ -Clauses in Classical Greek.  $\omega$ elatives, Interrogatives, Exclamatives*. Leiden: Brill.
- Hayes, M. E. 2018. *An Analysis of the Attributive Participle and the  $\omega$ relative Clause in the Greek New Testament*. New York: Peter Lang.
- Kriki, E. 2013. ‘Morfosintaktiki neoterismi sti glossa ton mi filologikon papiron: I anaforikes protasis’. PhD Dissertation, Thessaloniki: Aristotle University of Thessaloniki.
- Lehmann, C. 1984. *Der  $\omega$ relativsatz*. Tübingen: Gunter Narr Verlag.
- Perna, E. 2013. ‘La frase relativa in greco antico. Analisi sintattica basata sul dialetto attico di Platone’. PhD Dissertation, Padua: University of Padua.
- Probert, P. 2015. *Early Greek  $\omega$ relative Clauses*. Oxford: University Press.
- Tabachovitz, D. 1943.  *$\omega$ tudes sur le grec de la basse époque*. Uppsala: Skrifter.
- Toit, H. C. du. 2016. ‘Some Syntactic Features of Relative Constructions in the Greek New Testament’. *Stellenbosch Papers in Linguistics* 45: 49–75.
- Toit, H. C. du. 2022. ‘The Continuative Relative Clause in the Greek New Testament’. *Novum Testamentum* 64, 162-183.

## From distal demonstrative to resultative marker (through definite article): evidence from Basque

### 1. Introduction

Demonstratives are at the origin of a variety of grammatical morphemes: definite articles, complementizers, conjunctions, etc. (see Diessel 1999: 39, Kuteva *et al.* 134-147). Some of these grammaticalization processes have also been described for Basque: for instance, it is beyond dispute that the definite nominal inflection emerged through the grammaticalization of demonstratives (Manterola 2015). This paper explores a grammaticalization path that to our knowledge is yet to be described in other languages, namely the one leading from demonstrative to resultative marker, through the intermediate stage of definite article.

### 2. Resultative constructions in Modern Basque

Resultative constructions in Basque consist of a predicative construction containing a past participle with the definite article and the auxiliary verb, as exemplified in (1). They are similar to adjectival predicates, in which the adjective appears also with the determiner (2):

- |     |    |                               |             |    |  |    |  |     |     |                |      |
|-----|----|-------------------------------|-------------|----|--|----|--|-----|-----|----------------|------|
| (1) | a. | Mikel                         | iritsi-a    | da |  | b. | Mikel                                  | eta | Ane | iritsi-ak      | dira |
|     |    | Michael                       | arrived-the | is |  |    | Michael                                | and | Ann | arrived-the.PL | are  |
|     |    | ‘Michael has already arrived’ |             |    |  |    | ‘Michael and Ann have already arrived’ |     |     |                |      |
| (2) | a. | Mikel                         | handi-a     | da |  | b. | Mikel                                  | eta | Ane | handi-ak       | dira |
|     |    | Michael                       | big-the     | is |  |    | Michael                                | and | Ann | big-the.PL     | are  |
|     |    | ‘Michael is big’              |             |    |  |    | ‘Michael and Ann are big’              |     |     |                |      |

Resultative constructions contrast with perfect analytic construction in that the latter do not attach the definite article to the past participle, see examples in (3):

- |     |    |                       |         |    |  |    |                                |     |     |         |      |
|-----|----|-----------------------|---------|----|--|----|--------------------------------|-----|-----|---------|------|
| (3) | a. | Mikel                 | iritsi  | da |  | b. | Mikel                          | eta | Ane | etorri  | dira |
|     |    | Michael               | arrived | is |  |    | Michael                        | and | Ann | arrived | are  |
|     |    | ‘Michael has arrived’ |         |    |  |    | ‘Michael and Ann have arrived’ |     |     |         |      |

That the morphological origin of *-a* (plural *-ak*), the definite article, is to be found in the distal demonstrative *ba* ‘that’ is uncontroversial (Azkue 1923-1925, Trask 1997). It is however obvious that its modern functions range well beyond a simple definiteness marker, see examples in (2).

### 3. Historical data

In Old Basque, adjectives and nouns involved in predicative constructions were usually not marked with a definite article (3).

- |     |  |      |    |                    |  |
|-----|--|------|----|--------------------|--|
| (3) | Eihera   | hon  | da | dabileno           | (Oihenart, 1657, <i>Proverbes</i> , 206) |
|     | mill.the   | good | is | as.long.as.it.runs |  |
|     | “Le moulin est bon tandis que la meule se remue” |      |    |                    |  |

Likewise, the oldest Basque texts exhibit a language stage in which resultativeness may be expressed through sheer past participle forms, see (4):

- |     |   |      |          |          |     |  |
|-----|---|------|----------|----------|-----|--|
| (4) | Habia   | egin | deneko,  | xori-a   | hil | (Oihenart, 1657, <i>Proverbes</i> , 206) |
|     | cage.the  | done | AUX.TEMP | bird-the | die |  |
|     | “Pour lors que la cage a été faite, l’oiseau est venu à mourir” |      |          |          |     |  |

The spread of the definite article to contexts where no definite interpretation is allowed is observable on written records. This spread is most readily noticeable in predicative constructions (2), and is present already in texts of the 16<sup>th</sup> century (Manterola 2015).

### 4. The historical evolution of resultative phrases

The definite articles *-a/-ak* spread to resultative contexts only from the 16<sup>th</sup> century onwards (Mounole 2014). This paper argues that this spread occurred on the model of adjective predicates; the need for number marking may have played a role, since number is only morphologically overt in definite phrases.



**References**

- Azkue, Resurrección M<sup>a</sup>. 1923-1925. *Morfología vasca*. Euskaltzaindia: Bilbo.
- Diessel, Holger, 1999, “The morphosyntax of demonstratives in synchrony and diachrony”, *Linguistic Typology* 3, 1-49.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, Seongha Rhee, 2019, *World Lexicon of Grammaticalization*, 2nd edition. Cambridge: Cambridge University Press.
- Manterola, Julen. 2015. *Towards a history of Basque morphology: articles and demonstratives*. Doctoral dissertation, UPV/EHU.
- Mounole, Céline. 2011. *Le verbe basque ancien: étude philologique et diachronique*. Doctoral dissertation, UPV/EHU and Michel de Montaigne-Bordeaux III. [Published in *ASJU* 48, 2014 (2018)]
- Trask, Robert Lawrence. 1997. *The history of Basque*. Londres, New York: Routledge.

### **Areality through Migration: Investigating the Structure of Numeral Classifiers in the Eastern Himalayan Region Reveals Historic Contact Events**

Numeral classifiers exhibit striking distribution in certain hotspots around the world, serving as strong areal markers due to their highly recessive nature genetically and areally (Nichols, 2003:299). By investigating one such known hotspot, the Eastern Himalayan Region (henceforth: EHR ; often called North-East India in the prior literature) – I show that the surface proliferation of a grammatical feature can actually be the result of migration and distant contact events, rather than an indication of intense, sprachbund-like convergence as was previously assumed for this feature. In doing so, I problematize the notion of an ‘areal feature,’ and the methods used to define them.

The EHR is a complex linguistic zone with four distinct language families - Trans-Himalayan, Indo-Aryan, Austroasiatic and Kra-Dai - all with long-standing similarities in history, culture and ecological environment (Konnerth et. al, 2020). Previous work identifying the EHR as a linguistic area sees classifiers as a defining areal feature (Moral, 1997), with recent work (Chelliah and Lester, 2017 ; Cathcart et. al, 2020) pointing to contact as the explanation for the prominence of classifiers in the region, or their emergence in Eastern Indo-Aryan, respectively. By surveying 22 of the languages in Assam (the most populous state of the region), I find evidence to the contrary, and present the first detailed investigation of the spread of the feature in this region that reveals clues about the linguistic prehistory of the EHR, and South Asia more broadly.

By analyzing the structure of the classifier systems in the area, I find that contact is an unlikely explanation for the high frequency of classifier languages in the area. The premise of my analysis is the fact that classifiers here are not lexical borrowings from other languages in the region, but instead have cognates in their own distant linguistic relatives. Hence, a situation of language contact must be similar to that elucidated by Matras and Sakel (2007) where structural, instead of lexical borrowing takes place, replicating a *pattern* from the donor language: a manifestation of metatypy (Ross, 1996, 2007). The key elements I identified for this analysis were Noun (N), Numeral/Quantifier (Q), and Classifier (CL) as established by previous scholars (Jones, 1970; Allan, 1977). When examining the order of these key elements, I find that the structure of the classifier phrase in *every* language of the area is predicted by its genetic affiliation, reflecting inheritance, rather than contact. This argument is further supported by the fact that the languages of the same family outside of the area (such as Standard Thai, Odia and Vietnamese) do not show a difference in the order of key elements as those in the EHR. All of this points towards contact being an insufficient explanation for the perceived areality of numeral classifiers. Some interesting oddities remain – Eastern Indo-Aryan and Khasian do share the order of key elements, and Tibeto-Burman is an outlier in all senses – the order of its classifier phrase is distinct from all languages in the area as well as its own genetic relatives outside of the area.

What then, explains the distribution of numeral classifiers in this region? I suggest that what has been considered an areal feature of the EHR can be better explained by two distant contact events and the subsequent migration of those communities into the region. This hypothesis is supported by previous work in historical linguistics – Peterson (2010, 2017) highlights classifiers as a potential contact feature between Eastern Indo-Aryan and Proto-Munda. Moreover, Dockum (2016) points towards metatypy between Southwest Tai and parts of Austroasiatic, which could have been an earlier contact event that feeds into this. I weigh the linguistic data against contributions from allied fields such as archaeogenetics, anthropology, and history, to build a timeline of when speaker groups could have migrated into the EHR, and elucidate the kind of contact that has occurred in the region. Understanding how patterns of grammatical features in the EHR can tell us about its history allows us to build a better picture of how contact and migration shape the linguistic diversity we see in other such complex linguistic zones with high genetic diversity, such as the Daly River region, Western South America, or Mainland South-East Asia.

## References

- Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311. <https://doi.org/10.1353/lan.1977.0043>.
- Cathcart, Chundra A., Andreas Hölzl, Gerhard Jäger, Paul Widmer & Balthasar Bickel. 2020. Numeral classifiers and number marking in Indo-Iranian: A phylogenetic approach. *Language Dynamics and Change* 11(2). 273–325. <https://doi.org/10.1163/22105832-bja10013>.
- Chelliah, Shobhana & Nicholas Lester. 2016. Contact and convergence in the Northeast. In *The Languages and Linguistics of South Asia: A Comprehensive Guide*. Walter de Gruyter GmbH & Co KG.
- Dockum, Rikker 2016. Numeral classifiers in areal perspective: Khmer and Thai ‘syntactic borrowing’ revisited. *Austroasiatic Syntax in Areal and Diachronic Perspective. Workshop of the Austroasiatic Working Group*. September 5-7, 2016. [Slides]
- Jones, Robert B. 1970. Classifier Constructions in Southeast Asia. *Journal of the American Oriental Society*. American Oriental Society 90(1). 1–12. <https://doi.org/10.2307/598428>.
- Konnerth, Linda, Stephen Morey, Mijke Mulder, Mark W. Post & Kellen Parker van Dam. 2020. Languages and Peoples of the Eastern Himalayan Region and the North East Indian Linguistics Society: Taking stock. *Himalayan Linguistics* 19(2). <https://doi.org/10.5070/H91151002>.
- Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical Borrowing in Cross-Linguistic Perspective* (Empirical Approaches to Language Typology [EALT]). Berlin, New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110199192>.
- Ross, Malcolm. 1996. Contact-induced change and the comparative method: cases from Papua New Guinea. In *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford University Press, USA.
- Ross, Malcolm. 2007. Calquing and Metatypy. *Journal of Language Contact*. Brill 1(1). 116–143. <https://doi.org/10.1163/000000007792548341>.
- Nichols, Johanna. 2003. Diversity and Stability in Language. In *The Handbook of Historical Linguistics*, 283–310. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470756393.ch5>.
- Moral, Dipankar. 1997. North-East India as a linguistic area. *Monkmer Studies* 43-54.

Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2).

<https://doi.org/10.5070/H99223479>.

Peterson, John. 2017. Fitting the pieces together - Towards a linguistic prehistory of eastern-central South Asia (and beyond). *Journal of South Asian Languages and Linguistics*. Berlin, Germany: Walter de Gruyter GmbH 4(2). 211–257.

<https://doi.org/10.1515/jsall-2017-0008>.

***Rumpled chicken come home to roost.***

**From [TO CARD – IMPURITY] to [TO PURIFY/HEAL (someone) – from DISEASE]. Evidence from Anatolian, Ancient Greek, and Old Indic.**

The Modern Greek and Italian expression ο κόμπος φτάνει στο χτένι and *tutti i nodi vengono al pettine* (litt. *the knot(s) came to the comb*) clearly means *rumpled chicken come home to roost*. The main historical dictionaries of proverbs trace its origin back to the textile production in the antiquity, but the details are unknown (e.g. La Pucci 2007: 1023).

I advance the hypothesis that this proverb goes back to a phraseological collocation [TO COMB – IMPURITY (from wool)], that is, [TO CARD (wool)], which may reflect Proto-Indo-European (PIE) heritage, and which developed a metaphorical meaning already in antiquity. In particular, I argue for the development of a meaning [to COMB – a IMPURITY/DISEASE (from someone)], that is, [TO PURIFY/HEAL (someone or something) – from a DISEASE]. In other words, healing someone from a disease became like carding wool from impurities. This development is attested in the Anatolian and Ancient Greek branch of the Indo-European language family, although each branch has undergone a specific development starting from the same basic collocation. Anatolian shows a metaphorical narrowing into [TO COMB (down) – DISEASE – from a BODY PART] meaning [TO PURIFY/HEAL (someone) – from a DISEASE] (CTH 765.1, CTH 409.I). The very same collocation remains unattested in Ancient Greek. Instead, it shows a long simile between carding wool and the purification of the city of Athens from a mortal plague, together with a series of lemmata concerning processing wool (Aristoph. *Lys.* 574-86). In addition to that, in a controversial Old Indic passage, the collocation [TO CARD – X] is used metaphorically (1×, AVP 2.31.4, early 1<sup>st</sup> mill. BC) in the sense of ‘to prevent death/ill-minded event’. Although these differences, Anatolian as well as Indo-Iranian and Greek share the same Proto-Indo-European verbal root \**kse(-n)-* ‘to card, to comb’ in the above-mentioned passages: Luw. *kiša-* ‘to comb’ (+ *katta* ‘down’), Ved. *kṣan-* ‘to card’, Gk. *ζαίνω* ‘to comb’ (+ *κατὰ* ‘down’). I will conclude that it is possible to venture the hypothesis that the basic parallel between surviving an unfavourable event and the carding process was already established at a PIE level, although it is methodologically unlikely to infer the very existence of a PIE metaphor even from the attestation of its basic elements attested in the daughter languages (as regards this methodological issue see recently Melchert 2020). Should this hypothesis be correct, it would enrich the numerous metaphors from the semantic field of textile production attested in the PIE languages – for a comparative perspective on this topic see, most recently, Olsen 2018. This study integrates the increasing interest on the identification of cognitive metaphors within the Indo-European languages, with the main scope to separate those elements that can be traced back to a PIE stage from those that are the result of later historical events. After the seminal work of Lakoff & Johnson 1980, who investigated the persistent use of metaphorical language in all areas of human experience, only some attempts have been made so far within the Indo-European languages (most recently, e.g. Kölligan 2020, 2022, van Beek 2017).

**Bibliography**

- Beek, Lucien van (2017). Die Bildersprache des Rechts im Indogermanischen: Griechisch *θεῖα δίκη* und *δικὴν βλάπτειν*. In H. Bichlmeier and A. Opfermann (eds), *Das Menschenbild bei den Indogermanen. Akten der Arbeitstagung der Indogermanischen Gesellschaft in Halle* (31. März – 2. April 2011), 129–150. Hamburg: Baar.
- Kölligan, Daniel (2020). Seething anger: Latin *furor*. In: L. Repanšek, H. Bichlmeier, and V. Sadovski (eds.), *vácāmsi miśrā kṛṇavāmahai*. Proceedings of the international conference of the Society for Indo-European Studies and IWoba XII, Ljubljana 4–7 June 2019, 397–412. Hamburg: Baar.

Kölligan, Daniel (2022). Conceptual metaphors and historical linguistics – some case studies. Talk at the Università Cattolica del Sacro Cuore, Milan. March 22, 2022.

Lakoff, George and Mark Johnson (1980). *Metaphors we Live by*. Chicago: University of Chicago Press.

Melchert, Craig. 2020. Hittite *handā(i)*- ‘to align, arrange, etc.’ and PIE Metaphors for ‘morally right’. In M. Serangeli and Th. Olander (eds.), *Dispersals and Diversification. Linguistic and Archaeological Perspectives on the Early Stages of Indo-European*, 166-178.

Olsen, B. A. 2018. Notes on the Indo-European terminology of sheep, wool and textile production. In R. Iversen & G. Kroonen (eds), *Digging for words. Archaeolinguistic case stories from the XV Nordic TAG Conference held at the University of Copenhagen 16–18 April 2015*, 66–77. Oxford.

## Secondary lateral obstruents in South Cushitic and their significance for the linguistic history of East Africa

The precise position of South Cushitic within Cushitic (Afroasiatic) is a matter of controversy. Proposals include South Cushitic as (1) a primary branch of Cushitic (Greenberg 1963:48-49, Tosco 2020:292), (2) a coordinate branch of East Cushitic under an East–South Cushitic node (Ehret 1995:489-490, 2008:159), (3) a sub-branch of East Cushitic (Tosco 2000:109), (4) part of Lowland East Cushitic (Appleyard 2012:278), and (5) a sub-branch of (Southern) Lowland East Cushitic (Hetzron 1980:77, 101).

South Cushitic is also unique in Cushitic for having lateral obstruents, **ɬ** and **ɬʰ**. These phonemes have been reconstructed to Proto-Cushitic (Ehret 1995, 2008), from which South Cushitic is thought to have inherited them (cf. Bender (2020:138) for an alternative view). Mous (2012:347) identifies the presence of lateral obstruents in South Cushitic as problematic for establishing its position within East Cushitic: The further down in the tree it is posited, the more events of loss of lateral obstruents have to be assumed in the rest of Cushitic, and/or the more wide-ranging the waves obliterating these sounds must have been; hence the less likely such a classification is *a priori*.

In this paper I propose that the lateral obstruents in at least some South Cushitic lexical items are secondary, i.e. not inherited *as lateral obstruents* from Proto-Cushitic. In these instances, then, lateral obstruents are no obstacle for a lower-level classification of South Cushitic. I further argue that these items are historically connected to Pre-Oromo, one of the lowest branches in Cushitic. I trace the sound changes involved in their development and establish the relative chronology of some of these processes.

An example is Proto-South Cushitic **\*ɬaʔ** ‘to love, like, want’ (Kießling & Mous 2003:254) which is connected to Oromo **ja:l-** ‘to love’ < Proto-East Cushitic **\*geʔl-** ‘*id.*’ (Sasse 1979:36). The following diagram shows the development of this root from Proto-East Cushitic to Oromo and Proto-South Cushitic, based on sound changes taken from the literature (Black 1974, Sasse 1979, Kießling & Mous 2003) and the assumption of an additional, cross-linguistically supported sound change **j > ɬ** that links the East and South Cushitic strands of development:

	East Cushitic		South Cushitic	
	<b>*geʔl-</b>			
changes within Pre-Oromo				changes within Pre-Proto-S. Cush.
<b>ʔ &gt; ɬ</b>	<b>geʔl-</b>			
<b>g &gt; j / _ {i, e}</b>	<b>jeʔl-</b>			
<b>e &gt; a / [most contexts]</b>	<b>jaʔl-</b>		<b>jaʔl-</b>	
<b>ʔ &gt; V: / { _ C, C _ }</b>	<b>ja:l-</b>		<b>ɬaʔl-</b>	<b>j &gt; ɬ</b>
			<b>ɬaʔ-</b>	non-initial <b>l &gt; ∅</b>
Oromo	<b>ja:l-</b>		<b>*ɬaʔ-</b>	Proto-South Cushitic

I discuss whether the link between South Cushitic and Pre-Oromo thus established is one of common inheritance or language contact and conclude that lateral transfer from Pre-Oromo to South Cushitic is the most likely scenario. In the final part of the paper the implications of this late emergence of seemingly archaic phonemes are drawn for the history of Cushitic and its speakers in East Africa.



## References

- Appleyard, David L. (2012). Cushitic. In: Lutz Edzard (ed.), *Semitic and Afroasiatic: Challenges and opportunities*. (Porta linguarum orientaliun, Neue Serie, 24). pp. 199-295. Wiesbaden: Otto Harrassowitz.
- Bender, M. Lionel (Grover Hudson, ed.) (2020). *Cushitic lexicon and phonology*. (Research in African studies, 28). Berlin: Peter Lang.
- Black, Paul David (1974). Lowland East Cushitic: Subgrouping and reconstruction. PhD dissertation. New Haven: Yale University.
- Ehret, Christopher (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, tone, consonants, and vocabulary*. (University of California publications in linguistics, 126). Berkeley: University of California.
- Ehret, Christopher (2008). The primary branches of Cushitic: Seriating the diagnostic sound change rules. In: John D. Bengtson (ed.), *In hot pursuit of language in prehistory. Essays in the four fields of anthropology. In honor of Harold Crane Fleming*. pp. 149-160. Amsterdam: John Benjamins.
- Greenberg, Joseph H. (1963). *The languages of Africa*. (Publications, 25). Bloomington: Research center in anthropology, folklore, and linguistics, Indiana University.
- Hetzron, Robert (1980). The limits of Cushitic. *Sprache und Geschichte in Afrika* 2:7–126.
- Kießling, Roland & Maarten Mous (2003). *The lexical reconstruction of West-Rift Southern Cushitic*. (Cushitic language studies, 21). Cologne: Rüdiger Köppe.
- Mous, Maarten (2012). Cushitic. In: Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*. (Cambridge language surveys). pp. 342-675. Cambridge: Cambridge university.
- Sasse, Hans-Jürgen (1979). The consonant phonemes of Proto-East-Cushitic (PEC): A first approximation. *Afroasiatic linguistics* 7(1):1-67.
- Tosco, Mauro (2000). Cushitic overview. *Journal of Ethiopian studies* 33(2):87-121.
- Tosco, Mauro (2020). East Cushitic. In: Vossen, Rainer & Gerrit J. Dimmendaal (eds), *The Oxford handbook of African languages*. pp. 290-299. Oxford: Oxford University.

## Discontinuous noun phrases containing adjective or adjective-like modifiers in Middle Polish texts. Preliminary research conducted on an experimental dependency treebank

In my presentation, I will show examples of unusual – compared to the modern state – order of words in noun phrases containing an adjective or adjective-like modifier in the Polish language of the 17<sup>th</sup> and 18<sup>th</sup> centuries (e.g. *dziwna trafiła się awantura* ‘there was a **strange** riot’, lit. ‘**strange** there was a **riot**’; o *pożytkach z xiąg tłumaczenia **wynikających*** ‘about the **benefits** resulting from translation of books’, lit. ‘about the **benefits** from books’ translation **resulting**’; u *meگو widział mię **stryia*** ‘he saw me at **my** uncle’s’, lit. ‘at **my** he saw me **uncle’s**’). The uniqueness of the quoted and similar examples from Middle Polish texts lies in the fact that other sentence elements are relatively often placed between the components of nominal phrases, especially those elements that are not components of this nominal phrase. The elements of such discontinuous constructions can be so far apart as to make the sentence difficult to understand. Contrary to the current state of the language, such an order was very common in Middle Polish (cf. e.g. Ostaszewska 2002).

This phenomenon is known to historians of the Polish language, but the novelty of my research is that it is conducted on a dependency treebank (a syntactically annotated corpus) – the first such resource created for pre-modern Polish. Middle Polish Micro-Treebank contains 1,000 sentences selected from the Electronic Corpus of the 17<sup>th</sup>- and 18<sup>th</sup>-century Polish Texts (KorBa, [www.korba.edu.pl](http://www.korba.edu.pl)). The treebank was created as an experimental resource; a further expansion is planned. One of the aims of the study presented here was to show what possibilities of studying historical syntax are offered by such a resource.

A dependency tree is a graph that unambiguously reflects the dependencies between the components of a sentence. It reflects e.g. the left or right position of the dependent relative to its governor (e.g. *dziwna awantura* vs. *awantura dziwna*) and the distance between them. It is worth mentioning that the graphical visualization of trees shows certain features of word order. In the case of a “typical”, linear word order, graph edges (visualized as arrows) do not intersect (as in the phrase *Każdy kraj ma przymioty* ‘Each country has qualities’). In the case of inversion and discontinuity of noun phrases with an adjective, we get trees with crossing edges (*Opaci **dobry** uprowidowani **fundacyą*** ‘Abbots supplied with **good** funds’, lit. ‘Abbots **good** supplied **funds**’).



Such a syntactically annotated corpus allows for various studies on the relative position of a noun and its modifier(s). In my speech, I will present some possible analyses and their results, carried out on a sample of Middle Polish texts.

**References (selection)**

Berdicevskis, A., and Eckhoff, H. (2020). *A Diachronic Treebank of Russian Spanning More Than a Thousand Years*. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 5251–5256, Marseille, France. European Language Resources Association.

<https://aclanthology.org/2020.lrec-1.646/>

Gruszczyński, W., Adamiec, D., Bronikowska, R., Kieraś, W., Modrzejewski, E., Wieczorek, A., and Woliński, M. (2022). *The Electronic Corpus of 17th- and 18th-century Polish Texts*. *Language Resources and Evaluation* 56, 309–332.

Klemensiewicz, Z. (1968). *Zarys składni polskiej [An outline of the Polish syntax]*. Warszawa: PWN.

Mel'čuk, I. A. (1988). *Dependency syntax: theory and practice*. Albany: SUNY Press.

Ostaszewska, D. (Ed.) (2002). *Polszczyzna XVII wieku. Stan i przeobrażenia [Polish language of the 17<sup>th</sup> century. State and transformations]*, Katowice.

Wróblewska, A., and Rybak, P. (2019). Dependency parsing of Polish. *Poznań Studies in Contemporary Linguistics*, 55(2):305–337.

## The dominant-recessive hypothesis does not account for overlapping suppletion

Börjars & Vincent (2011) hypothesize that when two lexemes undergo suppletion, one is dominant and the other is recessive. The dominant lexeme is semantically more general and receives forms from the the recessive, which is less general semantically. For example, English *go* is more general than *wend*, and *go* is the lexeme that incorporated a form, *went*, of *wend*. This paper addresses several weaknesses of this proposal, especially as it relates to overlapping suppletion.

The first complication surrounding the dominant-recessive hypothesis (DRH) concerns the documentation of suppletion and its development. While many of the best-known cases of suppletion come from languages with ample historical documentation, not all examples of suppletion fit that model. Thus the DRH must be used with care so as not to become circular. A further issue with lower levels of documentation relates to situations in which some lexemes' forms survive only in suppletive paradigms. For example, the roots of English *am~be~was* are no longer represented elsewhere in the language, unlike the case of *go~went* vs *wend* mentioned above.

Cross-linguistically, separate survival of lexemes participating in suppletion is rare. Juge (1999, 2019) has identified a relatively small number of languages with overlapping suppletion, that is suppletion in which suppletive forms belong to two or more lexemes. For example, the Spanish copula *ser* 'be' and *ir* 'go' exhibit overlapping suppletion in the preterit and related paradigms (Table 1).

Such cases raise the question of how to gauge the semantic generality of the participating lexemes, in contrast with the relatively simple evaluation of verbs meaning 'go' and 'walk', for example, Juge (1999, 2019) has proposed that it is possible to correlate the semantic distance between lexemes with the non-overlapping, optionally overlapping, or non-optionally overlapping nature of the suppletive paradigms found in different cases. A key difference between these approaches is that the DRH is much more difficult to apply to situations in which the participating lexemes do not belong to the same semantic field (e.g, existence vs motion).

Furthermore, the DRH does not account for sound change as a source of suppletion. This reflects in part the traditional disregard of the various types of suppletion besides the best-known kind, incursion (cf. Juge 1999, 2013, 2019). In such cases, like the suppletive form of the present tense of the English copula, there is no evident role for semantics at all. Instances of analogically-created suppletion also challenge the DRH (along with ideas regarding the role of analogy as a regularizing process), as in the Galician (Fisterran dialect) form *iña* 'go (imperfect indicative)' created on analogy with *viña* 'come (imperfect indicative)' (Juge 2013).

The Galician cases raises a broader difficulty found not only in suppletion studies but also in grammaticalization research, namely faulty lexical semantic analysis, especially concerning what it means for one lexeme to be more general than another and how to apply such an evaluation to cross-linguistic analysis.

While the DRH may contribute to the analysis of some instances of suppletion, it must be combined with more detailed lexical semantic analysis, including measures of semantic distance, and contextualized among the multiple types of sources of suppletion.

	<i>ser</i> 'to be' /			<i>ir</i> 'to go' /		
	present	imperfect	preterit/	present	imperfect	preterit/
1s	soy	era	<b>fui/</b>	voy	<i>iba</i>	<b>fui/</b>
2s	eres	eras	<b>fuiste/</b>	vas	<i>ibas</i>	<b>fuiste/</b>
3s	es	era	<b>fue/</b>	va	<i>iba</i>	<b>fue/</b>
1p	somos	éramos	<b>fuimos/</b>	vamos	<i>ibamos</i>	<b>fuimos/</b>
2p	sois	erais	<b>fuisteis/</b>	vais	<i>ibais</i>	<b>fuisteis/</b>
3p	son	eran	<b>fuleron/</b>	van	<i>iban</i>	<b>fuleron/</b>

Table 1—Overlapping and non-overlapping suppletion (Juge 1999)

/

**References/**

BÖRJARS, KERSTI & NIGEL VINCENT. 2011. 'The pre-conditions for suppletion', in Alexandra Galani, Glyn Hicks & George Tsoulas (eds.), *Morphology and its interfaces*. Amsterdam: John Benjamins, 239-265.

JUGE, MATTHEW L. 1999. 'On the rise of suppletion in verbal paradigms', in Steve S. Chang, Lily Liaw & Josef Ruppenhofer (eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society*. Berkeley: BLS, 183-194.

JUGE, MATTHEW L. 2013b 'Analogy as a source of suppletion'. In Ritsuko Kikusawa & Lawrence Reid (eds.), *Historical linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25-30 July, 2011*. Amsterdam: John Benjamins, 175-197.

JUGE, MATTHEW L. 2019. The sense that suppletion makes: Towards a semantic typology on diachronic principles. *Transactions of the Philological Society*. 117(3). 390-414. doi: 10.1111/1467-968X.12175.

**Keywords/**

suppletion  
 lexical semantics  
 morphology  
 dominant-recessive hypothesis

## **The case of Italian *segunte*: an European instance of current change from verb to demonstrative?**

My proposal aims at discussing a potential instance of grammaticalization in current Italian, i.e. the change of the adjective *segunte* ‘following’ from an endophoric meaning to the function of proximal demonstrative.

There is no agreement in the literature regarding the development of demonstratives from lexical sources. Contra Diessel (2006), Heine et al. (2020) recently showed that demonstratives may originate at least from locative adverbs and from verbs, quoting many examples from studies on non-European languages. My analysis attempts to show that something very similar is occurring in current Italian, namely the potential origin of a new demonstrative from a verbal source.

*Segunte* is an adjective originating through transcategorization from the present participle of the verb *seguire* ‘follow’. It means “that comes immediately after in time, in space, in an ordering”. Given this lexical meaning, in its standard usage it has a mandatory cataphoric value, both as adjective and when it is nominalized, in the absence of a head noun. However, in my corpus – primarily consisting of a variety of Italian written by students (exam texts, chapters of theses, term papers, and e-mails), but also of some institutional communication texts – *segunte* also occurs either pointing to a previous referent (e.g. -Question: *Identifica il sintagma nominale nella frase seguente* (‘Identify the noun phrase in the following sentence’) -Answer (after the sentence): *Nella seguente frase...* (‘In the following sentence...’)) or without any endophoric values (e.g. *Nel seguente capitolo* (‘In the following chapter’), with reference to the chapter in which the PP occurs, not only at its very beginning).

Therefore, my hypothesis is that an innovation is taking place, and that it is in the direction of grammaticalization as a spatial deictic, specifically as proximal demonstrative, due to the presence of the feature [+ (immediately) PROXIMAL] in *segunte*.

Considering that in standard Italian the proximal demonstrative is *questo* ‘this’, my first reason is that in my data the equivalence [ART + *segunt-* (+ N)] ~ *quest-* (+ N) works everywhere. Moreover, the phenomenon fits diachronically into a process of desemantization of *segunte* that has already begun in its transcategorization from present participle to adjective (loss of causative value and intransitivization). Eventually, in my data there are bridge contexts, *segunte* appears frozen in prenominal position, there is extension of contexts (from immediately proximal to proximal deixis), and obligatory co-occurrence with the article, hence forming a phonological word.

A tentative explanation for the phenomenon is a restructuring of the demonstrative system, not unrelated to the weakening of the definite article.

## References

- Croft, W. (2000), *Explaining Language Change: an Evolutionary Approach*, London, Longman.
- Da Milano, F. (2010), “Dimostrativi, aggettivi e pronomi”, *Enciclopedia dell’italiano*, Treccani. ([https://www.treccani.it/enciclopedia/aggettivi-e-pronomidimostrativi\\_%28Enciclopedia-dell%27Italiano%29/](https://www.treccani.it/enciclopedia/aggettivi-e-pronomidimostrativi_%28Enciclopedia-dell%27Italiano%29/)).
- De Cesare, A. M. (2010), “Deittici”, in *Enciclopedia dell’italiano*, Treccani. ([https://www.treccani.it/enciclopedia/deittici\\_%28Enciclopedia-dell%27Italiano%29/](https://www.treccani.it/enciclopedia/deittici_%28Enciclopedia-dell%27Italiano%29/)).
- Diessel, H. (1999). *Demonstratives: Form, Function and Grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. (2006), “Demonstratives, joint attention, and the emergence of grammar”, *Cognitive Linguistics* 17(4). 463–489.
- Diessel, H. (2013a), “Distance Contrasts in Demonstratives”, in M. S. Dryer - M. Haspelmath (eds.), *The World Atlas of Language Structure Online* (<http://wals.info/chapter/41>).
- Diessel, H. (2013b), “Pronominal and Adnominal Demonstratives”, in M.S. Dryer - M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. (<http://wals.info/chapter/42>).
- Diessel, H. & Coventry K. R. (2020), “Demonstratives in Spatial Language and Social Interaction: An interdisciplinary Review”, *Frontiers in Psychology*, 11.
- Dixon, Robert M. W. (2003), “Demonstratives: A cross-linguistic typology”, *Studies in Language* 27, 61 – 112.
- Heine, B., T. Kuteva, H. Long, H. Narrog and Wu F. (2020), “Where do demonstratives come from?” *STUF* 73(3). 403-434.
- Ferrari, A. (2010a), “Anafora”, in *Enciclopedia dell’italiano*, Treccani, ([https://www.treccani.it/enciclopedia/anafora\\_%28Enciclopedia-dell%27Italiano%29/](https://www.treccani.it/enciclopedia/anafora_%28Enciclopedia-dell%27Italiano%29/)).
- Ferrari, A. (2010b), “Catafora”, in *Enciclopedia dell’italiano*, Treccani, ([https://www.treccani.it/enciclopedia/catafora\\_%28Enciclopedia-dell%27Italiano%29/](https://www.treccani.it/enciclopedia/catafora_%28Enciclopedia-dell%27Italiano%29/)).
- Giacalone Ramat, A. (1994), “Fonti di grammaticalizzazione. Sulla ricategorizzazione di verbi e nomi come preposizioni”, in Cipriano, P., Di Giovine, P. and Mancini, M. (eds.), *Miscellanea di studi linguistici in onore di Walter Belardi*, Roma, Il Calamo, vol. 2, 877-896.
- Giacalone Ramat, A. and Mauri, C. (2008), “From cause to contrast: a study in semantic change”, in Verhoeven, E., Skopeteas, S., Shin, Y.M., Nishina, Y. and Helmbrecht, J., *Studies on Grammaticalization* (eds.), Berlin, Mouton de Gruyter, 303-320.
- Heine, B., Claudi, U. and Hünnemeyer, F. (1991), *Grammaticalization. A Conceptual Framework*, Chicago, University of Chicago Press.
- Hopper, P. J. e Traugott, E. C. (1993), *Grammaticalization*, Cambridge, Cambridge University Press.
- Kortmann, B (1992), “Reanalysis completed and in progress: participles as source of prepositions and conjunctions”, in Kellermann, G. and Morrissey, M. D. (eds.), *Diachrony within Synchrony. Language History and Cognition*, Frankfurt am Main, Lang, 429-453.
- Lehmann, Ch. (1995), *Thoughts on Grammaticalization*, II Ausg., München, Lincom Europa.
- Szantyka, I. A. (2017), “Lo studio morfo-sintattico dei costrutti determinanti con il pronome dimostrativo variabile *quello*”, *Lublin Studies in Modern Language and Literature* 11, 96 – 111.
- Meillet, A. (1948), “L’*évolution des formes grammaticales*”, in Meillet, A., *Linguistique historique et linguistique générale*, vol. 1°, 2 éd., Paris, Champion, 130-148.
- Ramat, P. (2005), “Su degrammaticalizzazione e transcategorizzazione”, in Ramat, P., *Pagine linguistiche. Scritti di linguistica storica e tipologica* (ed.), Roma-Bari, Laterza, 89-105.
- Vanelli, L. (1989), “Dimostrativi e articoli: deissi e definitezza”, in G. L. Borgato, A. Zamboni (eds.), *Dialettologia e varia linguistica per Manlio Cortelazzo*, Padova, Unipress, 369 - 381.



**Prosody Reveals Syntactic Structure:**  
Secondary Predication in Metrical Finite Corpus Data

The mapping of syntax to prosody is regulated by correspondence requirements that hold between abstract syntactic structure and prosodic structure (Selkirk 2011; Elfner 2012; Ito and Mester 2013, among others). Preferentially, syntactic constituents map to prosodic constituents of the same level: morphosyntactic words ( $X^0$ ) map to prosodic words ( $\omega$ ), syntactic phrases (XP) map to prosodic phrases ( $\phi$ ), and clauses (CP/TP) map to intonation phrases ( $\iota$ ). Given the crucial interaction between syntax and prosody, prosodic structure can be used to identify and differentiate syntactic structure.

Cross-linguistically, secondary predicates tend to be marked by special prosody: they are either in prominent positions and/or are subject to isolation from their respective VPs. Depictives are thought to be more prosodically independent than resultatives (cf. Irimia 2012: 208 and references therein). A further distinguishing feature of secondary predicates is the tendency to occur in STAGE-LEVEL predicates expressing a non-permanent state (Carlson 1977; Simpson 2005; Casaretto 2020). Following Kratzer (1995), STAGE-LEVEL predicates have an extra argument position for events. Likewise, secondary predicates correspond to additional syntactic structure (Kratzer 2005; Irimia 2012) which maps to a separate prosodic domain.

This study examines the distribution of secondary predication across finite metrical corpora, including the works of Homer (Greek) and the RigVeda (Vedic Sanskrit). The central goal of the present study is to identify the diagnostics for secondary predication in Greek and Vedic. Our survey suggests that secondary predicates in Greek and Vedic tend to exhibit uniform prosodic behavior. In particular, secondary predicates are separated from postverbal nominals by (i) caesura, (ii) line break, (iii) the process of *enjambment* whereby syntactic units are broken across multiple prosodic domains at the expense of Selkirk (2011)'s MATCH constraints, or a combination of these strategies. The Greek (1) and Vedic (2) data below illustrate this point.

- (1) **ton** d' ōs oun enoēse podarkēs dīos  
 he.ACC but thus really see.AOR.ACT.3SG swift.NOM.SG.M divine.NOM.SG.M  
 Achilles // **gymnon**  
 Achilles.NOM.SG // naked.ACC.SG.M  
 “now as brilliant swift-footed Achilles saw him **naked**” (II 21.49-50)
- (2) purutrā vṛtró aśayad **vyàstah**  
 in.many.places Vṛtra.NOM.SG.M. lie.IMPF.3SG fling.apart.PTCP.NOM.SG.M.  
 “Vṛtra lay (there), **flung apart** in many places” (RV 01.32.7d)

Prosodic isolation of secondary predicates in Greek is accomplished by enjambment: the depictive/resultative APs and postverbal nominals are parsed in different lines (1). The Vedic data in (2) demonstrate an additional isolation strategy: line-finality and post verbal position. Our findings therefore lend further support to the importance of caesurae and line-boundaries in syntactic analyses of ancient metrical corpora (Hale and Kissonck 2021). The tendency to combine with STAGE-LEVEL (rather than INDIVIDUAL-LEVEL) predicates is also apparent—the secondary predicates *seeing him naked* (1) and *laying flung apart* (2) express transient properties and not permanent ones, as predicted. These facts set secondary predicates apart from attributive APs, which do not have complex syntactic structure corresponding to recursive  $\iota$  domains preserved via isolation strategies in finite metrical corpora.

**Selected references.** Carlson, G.N. (1977). “Reference to kinds in English”. PhD thesis, UMass Amherst. ■ Casaretto, A. (2020). “Onsecondary predicates in Vedic Sanskrit–Syntax and Semantics” IJDL. ■ Hale, M. and M. Kissock (2021). “n the syntax of comparative clauses in Vedic Sanskrit. . . like someone eating the foam off the water”. *Trends in South Asian Linguistics* 367. ■ Kratzer, A. (1995). *Stage-level and individual-level predicates*. ■ Kratzer, A. (2005). “Building resultatives”. In Maienborn, Claudia and Angelika Wöllstein, *Event arguments: foundations and applications*. ■ Selkirk, E. (1996). “The prosodic structure of function words”. In: *Signal to syntax*. ■ Selkirk, E. (2011). “The syntax–Phonology Interface”. *The handbook of phonological theory*.

**Quality vs. quantity: Contrast maintenance and tradeoff in Southwestern Tai vowels**

Contrastive vowel length is frequently found in languages in the Tai branch of the Kra-Dai family, and has been reconstructed to Proto-Tai (Pittayaporn 2009b). In many Tai languages, however, only one length contrast remains, between /a/ and /a:/. In this study, we explore the loss of the last standing phonemic length contrast in the Upper Chindwin dialect of Tai Khamti (TK), a member of the Southwestern Tai subgroup, which is spoken in Khamti Township, Sagaing Region, Myanmar. Draw on evidence from both reconstruction and original fieldwork, we show that this quantity contrast is undergoing a compensatory change toward a quality contrast in this dialect.

**Background.** Prototypical Southwestern Tai (SWT) languages have nine phonemic vowel qualities forming a schematic 3x3 grid of height and backness. Pittayaporn (2009a, 2009b) reconstructs 7 quality contrasts and 7 length contrasts in Proto-Tai, which became 9 quality contrasts and 4 length contrasts in its descendant Proto-Southwestern Tai (see **Tables 1a-b**), which can be interpreted as a historical tradeoff whereby a reduction in the number of quantity contrasts went hand-in-hand with an increase in the number of quality contrasts. However, most modern SWT languages preserve at least one length contrast, regardless of how many of the 9 quality contrasts they preserve.

*i, *i:	*u, *u:	*u, *u:
*e, *e:	*ɤ, *ɤ:	*o, *o:
	*a, *a:	

**Table 1a.** PT vowels per Pittayaporn 2009b.

*i, *i:	*u, *u:	*u, *u:
*e:	*ɤ:	*o:
*ɛ:	*a, *a:	*ɔ:

**Table 1b.** PSWT vowels per Pittayaporn 2009a.

In the other direction, Bangkok Thai, the most populous and best studied Tai language, has accrued quantity contrasts apparently without any requisite tradeoff toward fewer vowel qualities. Sukhothai Thai, a closely related historical lect reconstructed from surviving texts, had 7 contrastive length contrasts circa the 13<sup>th</sup>-15<sup>th</sup> centuries (Maspong 2015), and Bangkok Thai expanded length contrasts to all 9 simple vowels by the late 19<sup>th</sup> or early 20<sup>th</sup> century (Author xxxx). However, even within the duration contrasts of Thai, vowel quality may act as a secondary or redundant cue.

**Data.** Recordings of 9 TK speakers ages 24 to 78 were made at 5 locations in 2014-2015. Each speaker recorded a 436-item Southeast Asia wordlist (SIL 2002), with 1-3 repetitions for each lexical item. The lexical material and token counts vary somewhat between speakers, but the total corpus used in this study has approximately 2,000 tokens each of /a/ and /a:/.

<b>All speakers</b>	<b>/a/</b>	<b>/aa/</b>	<b>all V</b>	<b>LP5 (24F)</b>	<b>/a/</b>	<b>/aa/</b>	<b>all V</b>
all contexts	94	175	156	all contexts	107	153	144
sonorant codas	<b>107</b>	<b>147</b>	130	sonorant codas	<b>108</b>	<b>123</b>	120
obstruent codas	96	105	125	obstruent codas	100	95	121

**Table 2a-b.** Vowel duration in milliseconds for all 9 TK speakers vs. speaker LP5.

**Results.** On the dimension of quality, **Figure 1** plots F1 and F2 for the 9 Tai Khamti speakers, and shows varying degrees of divergence between the two vowel means, but with several speakers showing large vowel quality differences. On the dimension of quantity, **Table 2a** shows that across the sample population that the length contrast has been fully neutralized before obstruent codas, but there is on average a 40ms difference before sonorant codas. Further, **Table 2b** shows that for certain speakers, such as LP5, the length contrast is becoming neutralized before sonorant codas, with an average difference of just 15ms between /a/ and /a:/. Both lines of evidence, from duration and formants, suggest that the total loss of contrastive vowel length is complete or nearly so among some members of the Tai Khamti speaker community, which we predict will continue to progress in the coming years. At the same time, this contrast is apparently being maintained via cues from vowel quality. This represents an interesting case of compensatory change, unusual among the Southwestern Tai languages.

**References**

Author. xxxx.

Jonsson, Nanna L. 1991. Proto Southwestern Tai. Ph.D. dissertation, State University of New York at Albany.

Li, Fang Kuei. 1977. A handbook of comparative Tai. Ocean Linguistics Special Publications 15.

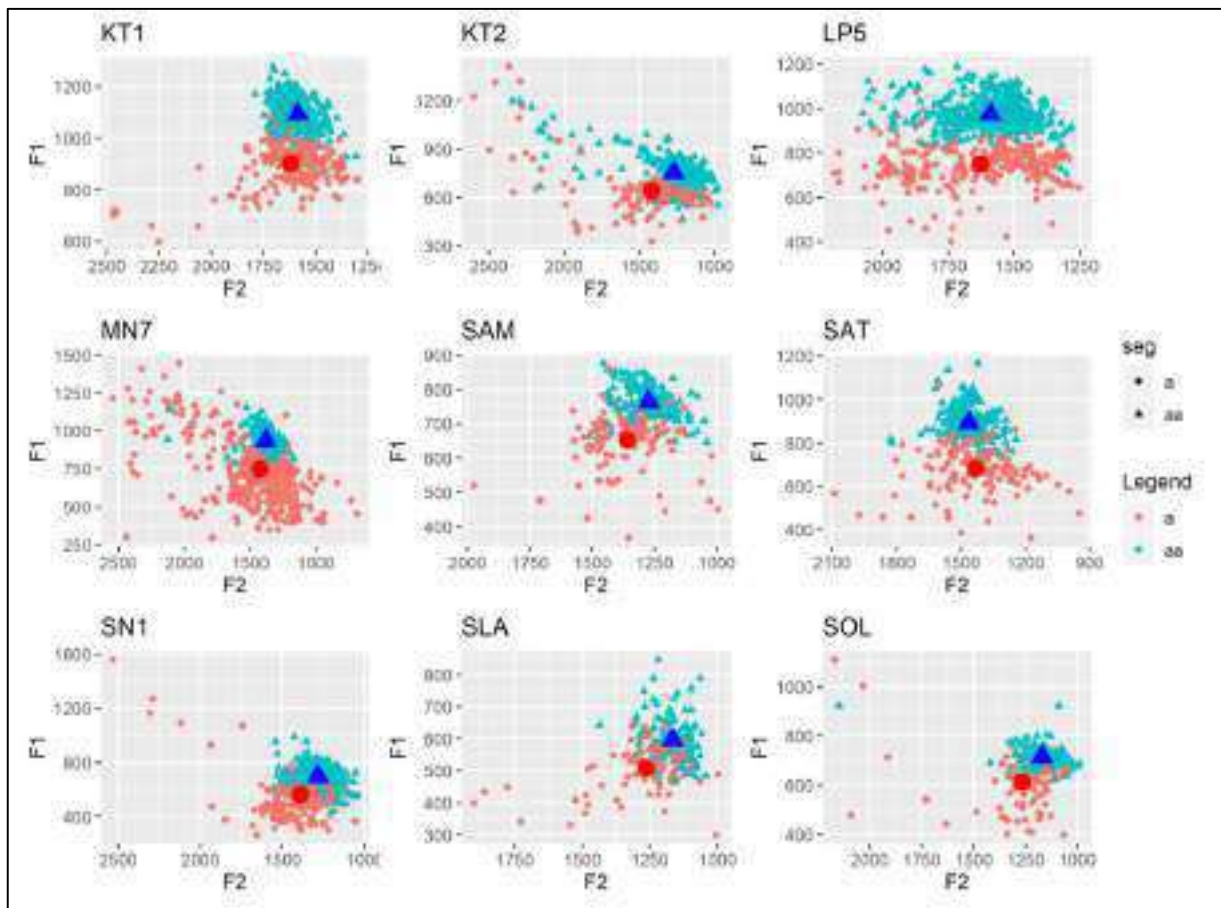
Maspong, Sireemas. 2016. Vowel length contrasts in the language of Sukhothai period. M.A. thesis, Chulalongkorn University. (in Thai)

Pittayaporn, Pittayawat. 2009a. Proto-Southwestern Tai revisited: a new reconstruction. *Journal of the Southeast Asian Linguistics Society* 2:121-143.

Pittayaporn, Pittayawat. 2009b. The phonology of Proto-Tai. Ph.D. dissertation, Cornell University.

Sarawit, Mary Elizabeth Sautter. 1973. The Proto-Tai vowel system. Ph.D. dissertation, University of Michigan.

Strecker, David. 1983. Proto-Tai vowels revisited: a comparison and critique of the work of Sarawit and Li. *Linguistics of the Tibeto-Burman Area* 7(2): 33-74.



**Figure 1.** Comparison of F1 and F2 formants, with means, of /a/ and /a:/ in 9 Tai Khamti speakers.

### **Morphologization of Phonological Processes as Integration**

The proposed paper presents a view of morphologization of phonological processes in which various degrees of integration of these processes into the morphology are possible, and follow a potential trajectory of increasing embeddedness. Debates on morphologization have focused on questions such as whether a phonological process needs to have become inactive, unproductive, or opaque as a precondition for morphologization (cf. e.g. Maiden 1991/2011). While this paper addresses the latter question likewise, the focus here is on how morphological systems integrate and absorb the phonological process and how they interact with the phonology thereafter. Different degrees of integration of phonological processes will be demonstrated with morphologization of vowel harmony in Turkish, consonant mutation in Welsh, fricative voicing in English, and German "umlaut".

In the following, we are assuming a modular grammar including a dynamic morphological component interfacing with lexicon, phonology, and syntax. Further, we assume that the phonology can only "see" phonological structures and only these can provide environments for phonological processes. If a process is restricted to specific morphological or lexical environments, the executive function over it belongs in the morphology and/or lexicon. The process itself will be applied by the phonology, but only under external orders.

As a first step towards morphologization of phonological processes, we can suggest that even prior to any loss of transparency/regularity or phonemicization of erstwhile allophones, there may develop an association between the phonological process and one or more morphological ones. Thus, for example, we would expect that an affinity between vowel-fronting and 'plural' would have been noted by speakers before the weakening of unstressed /i/ to schwa and thus loss of motivating environment in German, likewise between fricative voicing and 'plural' before fricative voicing became inactive as a phonological process in English. This "noting" could take the form of adding a redundant command to the phonology to the morphological operations creating plurals via suffixation etc. Such a scenario is preferable to one in which speakers are staring at the extra fronted vowels or voiced fricatives trying to find a use for them after the demise of productive phonological processes, i.e. wondering what to do with the resulting "junk" (Lass 1990).

While the cases of German umlaut and English fricative voicing could begin with "affinities" and be incorporated into already existing morphological processes such as affixation, ultimately potentially becoming the lone process corresponding to a particular morphological function, e.g. mapping onto 'plural', other pathways may involve rather more general associations with morphological functionality. In Turkish, for example, vowel harmony is observable in most words of two or more syllables, so that its value for associated morphological operations could not be much more concrete than 'I am a word(-form)'/ 'this is a word', i.e. a support for morphological structure in a general sense. Likewise with Welsh consonant mutation: the phonological process affected the entire obstruent series to begin with (Jackson 1953), so the syntactic contexts in which the initial consonant of some element of a construction was modified were particularly varied; here the contribution to the morphology would be something like 'I'm in construction' or 'this is a compound' etc. While Turkish vowel harmony may never progress beyond playing a supporting role in inflection and derivation, there is potential even for processes whose initial morphological function was very general to be "promoted" to the status of sole process in a morphological operation, as can be observed in Modern Welsh where in some dialects the nasal consonant alternation of a word-initial consonant has become the sole expression of 1 p.sg. possession ([kar] 'car', [vəŋar] 'my car' > [kar] 'car' ~ [ŋar] 'my car' (Jones 1998)). This is, of course, the highest degree of embedding into the morphology a once phonological process can undergo. In principle, then, a phonological process can be integrated first as a support to one or more morphological processes with only a very indirect relationship to morphological semantics, then in time become more closely associated with existing morphological semantic functions, eventually to figure as a morphological process with full status. However, long-term stability can be achieved at any of these stages.

**References Cited**

- Jackson, K. 1953. *Language and history in early Britain*. Cambridge: Harvard University Press.
- Jones, M. 1998. *Language obsolescence and revitalization: linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Clarendon Press.
- Lass, R. 1990. "How to do things with junk: Exaptation in language evolution." *Journal of Linguistics* 26:79-102.
- Maiden, M. 1991. *Interactive morphonology: Metaphony in Italy*. London: Routledge.

### The expression of predicative possession in Avestan

While the linguistic expression of the predicative possession has been widely investigated in the languages belonging to the major Indo-European branches such as Ancient Greek (Benvenuto&Pompeo 2012, Kulneff-Eriksson 1999), Latin (Baldi&Nuti 2010, Bolkestein 1983, 2001), Old Indian (Bauer 2000, Danesi&Barðdal 2018), in the case of the Ancient Indo-Iranian group these kinds of constructions have been observed by some scholars in previous studies, but have not been discussed in great depth. Regarding such earlier research, it is worth mentioning Benveniste’s well-known 1960 study where, on the basis of examples from many Indo-European languages including Old Persian and Avestan, the predicative constructions with dative and with genitive are classified respectively as «*prédicat de possession*», expressing “possession”, and as «*prédicat d’appartenance*», expressing “belonging”. Èdel’man’s 1975 paper and Bauer 2010 also deserve mention, since they contain a brief overview over the expressions of possession in the Iranian languages.

The aim of this paper is to provide a detailed account of the functional distribution of predicative possessive constructions in the Avestan language,

The investigation of the expressions of possession will be based on the situation documented in the Old Avestan texts. According to Kellens and Pirart (1988), the texts which are definitely Old Avestan are: the five *Gāθās* (Y. 28-35, 43-51, 53), the *Yasna Haptañhāiti* (Y. 35.2-41.5), two fragments, that is to say, the *Ahuna Vairiia* (Y. 27.13) and the *Airiiaman Išiia* (Y. 54.1); Y. 27.7 and Y. 56.1. The analysis will be focused on the two possessive constructions “verb to be plus dative” as in (1) and “verb to be plus genitive” as in (2), in order to identify the differences in their syntactic and pragmatic functions and in order to examine them from a semantic perspective.

(1) Av. Y. 62.1    *ušta buiiāṭ ahmāi naire*  
                           “Salvation be to this man”

(2) Av. Y. 43.7    *ciš ahī kahiiā ahī*  
                           “Who are you? Whose are you?”

In order to highlight the syntactic features of the two constructions we will take into account syntactic functions and the semantic roles of the Avestan genitive and dative, the relevant statistical tendencies regarding word order, focusing on the pragmatic role of the constituents.



## References

- Baldi, Philip&Pierluigi Cuzzolin (2005) Considerazioni etimologiche, areali e tipologiche dei verbi di «avere» nelle lingue indoeuropee. In: Sándor Kiss, József Herman, Luca Mondin&Giampaolo Salvi eds., *Latin et langues romanes: études de linguistique offertes à József Herman à l'occasion de son 80ème anniversaire*. Tübingen: 27-36.
- Baldi, Philip&Andrea Nuti (2010) Possession. In: Philip Baldi&Pierluigi Cuzzolin eds., *New perspectives on historical latin syntax constituent syntax: quantification, numerals, possession, anaphora* Berlin: 239-388.
- Bauer, Brigitte (2000) *Archaic syntax in Indo-European: the spread of transitivity in Latin and French* (Trends in linguistics. Studies and monographs; 125). Berlin.
- Benveniste, Emile (1960) “Être” et “Avoir” dans leurs fonctions linguistiques *Review of Bulletin de la Société de Linguistique de Paris* 55/1. 113-134.
- Benvenuto, Maria Carmela&Flavia Pompeo (2012) Expressions of predicative possession in Ancient Greek: "εἶναι plus dative" and "εἶναι plus genitive" constructions. *Review of AION. Annali dell'Università degli Studi di Napoli «L'Orientale»*. Dipartimento di Studi Letterari, Linguistici e Comparati 1: 77-103.
- Bolkestein, A. Machtelt (1983) Genitive and Dative Possessors in Latin. In: S. C. Dik ed., *Advances in functional grammar*. Dordrecht: 55-91.
- Bolkestein, A. Machtelt (2001) Possessors and experiencers in Classical Latin. In: Irène Baron, Michael Herslund&Finn Sørensen eds., *Dimensions of possession*. Amsterdam: 269-283.
- Cantera, Alberto (2012) *The transmission of the Avesta, (Iranica)*. Wiesbaden.
- Danesi, Serena&Jóhanna Barðdal (2018) Case marking of predicative possession in Vedic: The genitive, the dative, the locative. In: Barðdal, Jóhanna, Na'ama Pat-El&Stephen M. Carey eds., *Non-canonically case-marked subjects: the Reykjavík-Eyjafjallajökull papers*. Amsterdam, Philadelphia: 181–212.
- Èdel'man, Džoj I. (1975) Les verbes 'être' et 'avoir' dans les langues iraniennes. In: Françoise Bader ed., *Mélanges linguistiques offerts à Emile Benveniste*. Louvain: 151-158.
- Isačenko, Alexander (1974) On have and be Languages: A Typological Sketch. In: Michael S. Flier ed., *Slavic Forum: Essays in Linguistics and Literature*. The Hague-Paris: 43-77.
- Kellens, Jean&Eric Pirart (1988) *Les textes vieil-avestiques*. Wiesbaden.
- Kulneff-Eriksson, Karin (1999) *On 'have' in Ancient Greek. An investigation on εἶναι and the construction εἶναι with a dative as expressions for 'have'*, (Studia Graeca et Latina Lundensia 7). Lund.
- Skjærvø, P. Oktor (2009) Old Iranian: Avestan and Old Persian. In: Gernot Windfuhr ed., *The Iranian Languages*. London – New York.

## Computational Approaches for Romance Related Words Discrimination

### Abstract

Natural languages are living eco-systems, they are constantly in contact and, by consequence, they change continuously. Traditionally, the main problems in historical linguistics (“How are languages related?”, “How do languages change across space and time?”) have been investigated with comparative linguistics instruments. The main idea of the comparative method is to perform a property-based comparison of multiple sister languages in order to infer properties of their common ancestor. It is a time-consuming manual process that required a large amount of intensive work.

The identification of cognates is a fundamental process in historical linguistics, on which any further research is based. On the other hand, discriminating between lexical borrowings and inherited words is considered one of the most difficult and important tasks in HL (Jäger, 2019), for which “the computerised approach” is regarded as the appropriate solution even by classical linguists (Heggarty, 2012). We propose here computer-assisted methods for identifying cognates, for discriminating between cognates and borrowings, and for discriminating between inherited and borrowed Latin words.

Firstly, we introduce a method to automatically determine if a pair of words  $(u, v)$  are cognates or not, and we use it on a large database comprising the main Romance languages (Romanian, Italian, French, Spanish and Portuguese), applying it as well in subsequent tasks. Given an input pair of words, the initial task is to automatically determine if they are cognates or not. We developed a machine learning method for automatically producing the answer based on sequence alignment. To align pairs of words, we employed the Needleman Wunsch global alignment algorithm, which has been successfully used in natural language processing and computational biology. We used words as input sequences and a basic substitution matrix, which gives equal scores to all substitutions, disregarding diacritics (e.g., we ensure that  $e$  and  $\acute{e}$  were matched). For the machine learning part, we used an ensemble of methods. We applied our method to multiple data sets, showing that our approach improves on previous results, also having the advantage of requiring less input data, which is essential in historical linguistics, where resources are generally scarce. In the process of discriminating between cognate and borrowing, we tried to answer the following question: given a pair of words, are they cognates, borrowings, or neither? For the automatic discrimination between inherited and borrowed Latin words, the best results were obtained by a system based on SVM using features extracted from the word-etymon pairs. We apply our method on both graphic and phonetic forms of the words.

### Keywords

Romance languages, cognates, borrowings, inherited words.

### Acknowledgements

This research is supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS/CCCDI UEFISCDI, project number 108, CoToHiLi, within PNCDI III.

### References

[1] Ciobanu, Alina Maria and Liviu P. Dinu. 2019. Automatic Identification and Production of Related Words for Historical Linguistics. *Computational Linguistics*, vol. 45, No. 4, pages 667–704.

- [2] Alina Maria Cristea, Liviu P. Dinu, Simona Georgescu, Mihnea-Lucian Mihai and Ana Sabina Uban, 2021. Automatic Discrimination between Inherited and Borrowed Latin Words in Romance Languages In: *Proc. EMNLP 2021(Findings)*, Punta Cana, 2021
- [3] Alina Maria Ciobanu and Liviu P Dinu, 2014. Automatic detection of cognates using orthographic alignment. In *Proc. ACL 2014*, June 22-27, 2014, Baltimore, MD, USA
- [4] Alina Ciobanu and Liviu P Dinu, 2015. Automatic discrimination between cognates and borrowings. In *Proc. ACL 2015*, July 26-31, 2015, Beijing, China
- [5] Gerhard Jäger. 2019. Computational Historical Linguistics. *Theoretical Linguistics* | Volume 45: Issue 3-4, 45: Issue 3–4.
- [5] Paul Heggarty. 2012. Beyond Lexicostatistics: How to Get More out of "Word List" Comparisons. In *Quantitative Approaches to Linguistic Diversity: Commemorating the Centenary of the Birth of Morris Swadesh*, pages 113–137. Benjamins

### Emergence of alternate argument alignment patterns in Northwest Kainji

Northwest Kainji (NWK, Nigeria; Benue-Congo) languages display at least two argument alignment patterns: accusative and neutral. Head nouns in NWK occur with either prefixed or suffixed adnominal gender marking in most noun phrases (see Hoffmann 1967 and Dettwelier 2015 for Dakakari/C’Lela [dri]; Bendor-Samuel et al. 1973 for Duka/Ut-Hun [uth]; Author (2007, 2019) for Ut-Ma’in [gel], and D. Heath 2020 for Us-Saare [uss]). Labeled as ambifixes by Arkadiev (2022, mention of Ut-Ma’in [gel]) and typed as “clitic-like” by Güldemann & Fiedler (2022, mention of Dakakari/C’Lela [dri]), these powerhouse morphemes manifest grammatical categories of number (singular/plural/mass), class membership (including derivational functions adding designations of humanness, animacy, size, and shape), and, in certain morphosyntactic configurations, they flag grammatical role. However, affix placement, i.e., where, and whether or not, a particular form of adnominal marker occurs, is governed by the morphosyntactic configuration of the noun phrase. Further, the argument alignment pattern that occurs in any given clause also depends on the internal structure of a particular NP, see (i) and (ii), agnostic of grammatical role. That is, the argument flagging function on the adnominal markers is present, but potentially secondary to other referential functions.

- (i) Unexpanded NPs display **accusative alignment** by means of adnominal marking:
  - a. Unexpanded **subject** NPs occur with an adnominal gender **suffix** (nominative), (1–2);
  - b. Unexpanded **object** NPs occur with an adnominal gender **prefix** (accusative), as in (3).
- (ii) NPs that are expanded in any way, i.e., contain modifiers, display **neutral alignment** (4–6).

Two alignment patterns also exist to varying degrees within the pronominal system. In some NWK, noun class agreement pronouns manifest accusative alignment with distinct forms for subject and object. Most personal pronouns display neutral alignment; however, first person singular pronouns have distinct forms for subject.

- (1) *sē* [kó:t-jǝ] *rwōn* *ōr-vastè*  
 then guinea.fowl-C7.NOM exit C5-last  
 ‘Then a guinea fowl exited last.’ (Ut-Ma’in [gel], Author 2019: 104)
- (2) *kàna* [kó:t-jǝ] *zǝ-t.è* *ōr-kjat* □  
 there guinea.fowl-C7.NOM say-PRF C5-difficult  
 ‘There a guinea fowl has said, “Difficult...”’ (Ut-Ma’in [gel], Author 2019: 104)
- (3) *a=b* *hjan* [*ū-kó:t*]  
 COND=2SG see C7.ACC-guinea.fowl  
 ‘If you see a guinea fowl, ...’ (Ut-Ma’in [gel], Author 2019: 104)
- (4) [*jà=t-ǝ=s-té=tǝ*] *āzgōsè*  
 fruit=C6-ASSOC=C4-tree=C6.DEF roll.out  
 ‘Those fruit rolled out.’ (Ut-Ma’in [gel], Author 2019: 102)
- (5) [*jà=t-ǝ=s-té=tǝ*] *fámǝ* *t-mǝǝré*  
 fruit=C6-ASSOC=C4-tree=C6.DEF resemble C6-mango.fruit  
 ‘Those fruits resemble mangoes.’ (Ut-Ma’in [gel], Author 2019: 102)
- (6) *é=kár-g-ǝs:* [*jà=t-ǝ=s-té=tǝ*]  
 C2.SUBJ=pick-PST-ITR fruit=C6-ASSOC=C4-tree=C6.DEF  
 ‘They gathered those fruits.’ (Ut-Ma’in [gel], Author 2019: 102)

All NPs that demonstrate traditionally understood accusative alignment, may better be understood as demonstrating König’s (2008: 8, 158) Type 2 Marked Nominative system, crucially because of the wide range of functions that the accusative (prefix) form fills. In contrast, the nominative form has a very restricted set of functions. In this paper I present the morphosyntax of these alignment systems across four NWK languages, survey the morphosyntactic conditions that determine when the alternate patterns arise in each language, and propose rationale for the development of the suffixed marked nominative pattern.

## References:

Author. (2007). ...

Author. (2019). ...

Arkadiev, Peter, and Yuri Lander. (2022). *Ambifixes: Cross-Linguistic Variation and Possible Origins*.

Presented at the Princeton Phonology Forum 2022. Princeton University. December 2.

Bendor-Samuel, John, Donna Skitch and Esther Cressman. (1973). *Duka sentence, clause and phrase*. *Studies in Nigeria Languages 3*. Zaria, Nigeria: Institute of Linguistics.

Dettweiler, Stephen H. 2015. C'Lela Grammar Portrait. Electronic ms.

<https://www.sil.org/resources/archives/63135>.

Güldemann, Tom, and Ines Fiedler. (2022). *Cliticization cycles: Adnominal gender-number affixes in Niger-Congo*. Presented at the 25th International Conference on Historical Linguistics,

University of Oxford. August 2.

Heath, David M. (2020). U̇s-Saarė Phonological and Grammar Sketch. Electronic ms.

<https://www.sil.org/resources/archives/87200>.

Hoffmann, Carl. (1967). An outline of the Dakarkari noun class system and the relation between prefix and suffix noun-class systems. *La classification nominale dans les langues négroafricaines*, ed. G. Manessy, 237-259. Paris: CNRS.

König, Christa. (2008). *Case in Africa*. Oxford: Oxford University Press.

## Drivers of Diversity in the Construal of Quantity in the World's Languages

How have almost all the world's cultures developed diverse but conceptually related systems for exact reference to quantity? Observations of numeral and lexicogrammatical number systems for quantification influence the following conclusions about the origin and diversity of exact quantification resources. Numerals originate from counting practices and evolve through recursion (Wiese, 2007). Number, as a category of grammar, develops from lexical resources such as numerals and deictic expressions (Aikhenvald, 2018; Corbett, 2000). To account for the diversity in distribution of quantification resources across natural language, some studies invoke the complexity trade-off thesis: complexity in a system of quantification trades off complexity in another. For example, languages with numeral classifiers are said to have low-limit cardinal numeral systems and lack facultative plural number marking (Greenberg, 1987; Croft, 1994; Aikhenvald, 2000). Sociocultural and ecological factors such as the influence of climate and agricultural practices (Divale, 1999; Epps et al., 2012), as well as cognitive constraints like the economy of expression constraints (Haspelmath and Karjus, 2017) also drive diversity. The theories about diversity, however, seems to be based on a non-representative sample of the world's languages and often do not integrate influences of genealogy and contact, the key drivers of language diversity in Historical Linguistics framework. The association between diversity in numeral system and the full range of diverse lexicogrammatical number marking resources have not been explored. This study investigates the drivers of diversity in the distribution of resources for construing exact quantity in a representative sample of the world's languages, considering trade-offs and genealogy influences.

The study takes a diachronic approach to test whether (1) numeral system complexity trades-off lexicogrammatical number complexity over time. Numeral system complexity is coded as a continuous variable by measuring the restrictedness of the numeral system, presence of numeral classifiers and the robustness of the base system on a scale of 0-1. Data on numerals and grammatical number are respectively obtained from Numeralbank (Barlow et al., 2020) and World Atlas of Classifier Languages (WACL) (One-Soon, 2022). Grammatical number data from Grambank Consortium (2021) and World Atlas of Language Structures (WALS) are used to measure lexicogrammatical number complexity on a scale of 0-1 considering the presence and diversity of number features (e.g., singular, dual).

The study examines this trade-off in four geographically distributed language families: Sino Tibetan, Austronesian, Pama-Nyungan and Atlantic Congo. The evolution of quantification is modelled in two ways: An independent model of the evolution of numeral system complexity and grammatical number complexity and a dependent model of evolution. The study uses the phylogenetics Generalised Least Squares method and maps the data on Bayesian Phylogenetic tree data of the respective families. The talk presents preliminary results that hint a family specific co-evolution pattern.

## References

Aikhenvald, A. Y. (2018). Number systems in grammar - position paper. *Language and Culture Research Centre: 2018 Workshop*

Aikhenvald, A. Y. (2000). *Classifiers: A Typology of Noun Categorization Devices*. Oxford & New York: Oxford University Press.

Barlow, R. & Beller, S. & Bender, A. & Bibiko, H. & Forkel, R. & Greenhill, S. & Gray, R. & Hammarström, H & Jordan, F. & Rzymski, C. & Verkert, A. (eds.) (2020). *Numeralbank*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (available online at <http://numerals.clld.org>).

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.

Croft, W. (1994). Semantic universals in classifier systems. *Word* 45.2: 145-171.

Divale, W. (1999). Climatic instability, food storage, and the development of numerical counting: A cross-cultural study. *Cross-Cultural Research*, 33, 341-368.

Epps, P., Bower, C., Hansen, C. A., Hill, J. H., & Zentz, J. (2012). On numeral complexity in hunter-gatherer languages. *Linguistic Typology*, 16, 41-109.

Greenberg, J. (1974). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Proceedings of the 11th International Congress of Linguistics*, Bologna / Florence, Aug-Sept 1972. Bologna: 1974: 17-37.

Haspelmath, M. & Karjus, A. (2017). Explaining asymmetries in number marking: singularatives, pluratives and usage frequency. *Linguistics*, 55(6): 1213-1235.

One-Soon, Hammarström, H, Allasonnière-Tang, Marc, (2022). Defining numeral classifiers and identifying classifier languages of the world. *Linguistics Vanguard*

The Grambank Consortium, (ed.). 2021. *Grambank 1.0*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://grambank.clld.org> (accessed 4 July 2020).

Wiese, H. (2007). The co-evolution of number concepts and counting words. *Lingua*, 117, 758-772.



# Proto-Malayo-Polynesian: Some Phonetic Evidence for \*l

Philippine languages are considered strong evidence for assertions about reconstructed Proto-Malayo-Polynesian (PMP). We address the phonetic articulation of Proto-Austronesian \*l in PMP. We show reflexes via reported articulatory data from nine Philippine languages including both the north and south Philippine languages (Olson et al. 2010). Audio recordings support our work in three languages (Olson et al. 2009, 2008), one of which has also undergone an ultrasound study (Mielke et al. 2011). We suggest that a novel articulation (an interdental approximant) was in common use as far back as PMP, minimally in allophonic distributions, but possibly as the default pronunciation. Considering this evidence should cause us to carefully reconsider how we reconstruct words within PMP and higher reconstructions. The alternative hypothesis is to find pathways for independent innovations around the periphery of an archipelago in nine different Philippine languages. The proposed historical articulation, an interdental approximant, is phonetically rare in the world's languages, but not unattested (Everett 1982; Harley 2012). It is therefore likely overlooked for its historical significance in reconstructing PMP. We suggest that the diverse reflexes (ɭ, l, d, Ø, ʔ, n, y, ɔ̃) for \*l in Austronesian languages are based on the most salient acoustic cues for the specific speech communities. Zorc (1975: 264-6) acknowledges the irregular correspondences involving liquids but leaves the diversity unexplained. One possibility, which we advocate for, is that these variations are all strategies to bring the tongue inside the mouth due to social pressures while maintaining an auditory or perceptual cue. By appealing to visual motivations for sound change we link sound change process to language as a multi-modal experience (Vigliocco 2014; Ambrazaitis & House 2017; Perniss 2018). Finally following Havenhill and Do (2018), we agree that linking historical change phenomena in primarily oral-languages to how they are visually experienced leads scholars to a deeper understanding of community communication practices through time.

## Correspondences for the interdental and PAN \*l

English	Butbut	Lubuagen	Majukayong	Minangali	Kagayanen	Kalagan	Southern Catanduanes Bicolano	Blust (1999)
[eng]	[kyb]	[knb]	[knb]	[knb]	[knb]	[knb]	[knb]	[knb]
three	tu'ɔ̃u	ti'ɔ̃u	tu'ɔ̃u	tuɔ̃u	'tallo	toɔ̃o	tuɔ̃u	*telu
moon	'h <sup>w</sup> uɔ̃an	'buɔ̃an	so'ɔ̃ag	soɔ̃ag	'buɔ̃an	boɔ̃an	buɔ̃an	*bulaN / *qiNas
path	'tʃaɔ̃an	'keɔ̃sa	'qaɔ̃sa	ʔaɔ̃sa	'daɔ̃an	daɔ̃an	daɔ̃a	*zalan

## References

- Ambrazaitis, Gilbert, and David House. 2017. "Multimodal Prominences: Exploring the Patterning and Usage of Focal Pitch Accents, Head Beats and Eyebrow Beats in

- Swedish Television News Readings.” *Speech Communication* 95: 100–113. doi:10.1016/j.specom.2017.08.008.
- Blust, Robert. 1999. “Subgrouping, Circularity and Extinction: Some Issues in Austronesian Comparative Linguistics.” In *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, edited by Elizabeth Zeitoun and Paul Jen-kuei Li, 31–94. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert, and Stephen Trussel. 2013. “The Austronesian Comparative Dictionary: A Work in Progress.” *Oceanic Linguistics* 52 (2). University of Hawai‘i Press: 493–523. <https://www.jstor.org/stable/43286361>.
- Everett, Daniel Leonard. 1982. “Phonetic Rarities in Pirahã.” *Journal of the International Phonetic Association* 12 (2): 94–96. doi:10.1017/S0025100300002498.
- Harley, Matthew. 2012. “Unusual Sounds in Nigerian Languages.” In *Advances in Minority Language Research in Nigeria*, edited by R. Blench and Stuart McGill, 38–65. Kay Williamson Educational Foundation: African Languages Monographs 5. Köln: R. Köppe.
- Havenhill, Jonathan, and Youngah Do. 2018. “Visual Speech Perception Cues Constrain Patterns of Articulatory Variation and Sound Change.” *Frontiers in Psychology* 9: Article 728. doi:10.3389/fpsyg.2018.00728.
- Mielke, Jeff, Kenneth S. Olson, Adam Baker, and Diana Archangeli. 2011. “Articulation of the Kagayanen Interdental Approximant: An Ultrasound Study.” *Journal of Phonetics, Speech Reduction*, 39 (3): 403–12. doi:10.1016/j.wocn.2011.02.008.
- Olson, Kenneth S., Emy T. Ballenas, and Nilo M. Borromeo. 2009. “Buhì’non (Bikol) Digital Wordlist: Presentation Form.” *Language Documentation & Conservation* 3 (2): 213–25. <http://hdl.handle.net/10125/4440>.
- Olson, Kenneth S, Glenn Machlan, and Nelson Amangao. 2008. “Minangali (Kalinga) Digital Wordlist: Presentation Form.” *Language Documentation & Conservation* 2 (1): 146–56. <http://hdl.handle.net/10125/1772>.
- Olson, Kenneth S., and Jeff Mielke. 2008. “Acoustic Properties of the Interdental Approximant in Kagayanen.” *The Journal of the Acoustical Society of America* 123 (5). Acoustical Society of America: 3460–3460. doi:10.1121/1.2934307.
- Olson, Kenneth S., Jeff Mielke, Josephine Sanicas-Daguman, Carol Jean Pebley, and Hugh J. Paterson III. 2010. “The Phonetic Status of the (Inter)Dental Approximant.” *Journal of the International Phonetic Association* 40 (2): 199–215.
- Perniss, Pamela. 2018. “Why We Should Study Multimodal Language.” *Frontiers in Psychology* 9: Article 1109. doi:10.3389/fpsyg.2018.01109.
- Vigliocco, Gabriella, Pamela Perniss, and David Vinson. 2014. “Language as a Multimodal Phenomenon: Implications for Language Learning, Processing and Evolution.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 369 (1651): 20130292. doi:10.1098/rstb.2013.0292.
- Zorc, David R. 1975. “The Bisayan dialects of the Philippines: subgrouping and reconstruction.” Ph.D. dissertation, Cornell University

## Hypotheses and scenarios in North Germanic Tonogenesis

The presence of a lexical tonal distinction in North Germanic (*accent 1* vs. *accent 2*) occasions the following two things to explain (the explananda):

- A) The origin of a lexical representation
- B) The origin of a lexical distinction

A lexical *representation* is here a lexical tone or a lexical foot (inducing a tonal effect). A lexical *distinction* is here a contrast that is expressed in the tonal structure, and which isn't predictable from non-lexical information. An example would be the tonal contrast between the monomorphemes 'ketchup' 'id.' (accent 1) and 'senap' 'mustard' (accent 2) in Swedish. One of the members in this pair contains some lexical property that causes the tonal melodic difference. That lexical property is part of the representation.

Lexical distinctions are tied to lexical representations and that gives us a logical order for the explananda. We should have an account for explanandum A in order to properly address explanandum B, while the reverse does not hold.

Nevertheless, research tradition seems to have given primacy to the explanation of B, the origin of a lexical distinction, over and above A, the origin of a lexical representation. This is apparent in the importance given to two historical changes that are taken to be crucial for the development of the contrast in simplex forms (ca 1000–1200 AD): cliticization of the definite article (*and hinn* > *and-en* 'the mallard'), and epenthesis before sonorants (*segl* > *segel* 'sail'), both resulting in disyllabic forms with accent 1, which come to contrast with accent 2, taken to previously dominate in polysyllables.

I discuss the consequences of both orderings among the explananda, i.e., B>A and A>B. I contend that the B>A stance at best provides a *scenario* for explanandum A, but not a proper hypothesis. A *hypothesis* for explanandum A should propose to explain 1) how the marked tonal contour comes into being, and 2) how it gets phonologized in the relevant forms. The B>A stance fails on the phonologization issue.

The A>B stance shifts attention away from minimal pairs in simplex forms to postlexical tonal patterns that exist and persist in all dialects. I argue that tonal patterns and changes that are apparent today admit a hypothesis for tonogenesis in North Germanic that comes out favourably by Ockham's razor.

## Selected references

- Bye, Patrik. forthc. The Evolution of Pitch Accent in North Germanic: Pragmatics, Phonology, Geography, and History. Pavel Iosad & Björn Köhnlein (eds.).
- d'Alquen, K. and Brown, R. 1992. The origin of Scandinavian accents I and II. I. Rauch, G.F. Carr, & R.L. Kyes (eds.), *On Germanic Linguistics: Issues and Methods*. Berlin, New York: Walter de Gruyter, 61–80.
- Elstad, Kåre. 1980. Some remarks on Scandinavian tonogenesis. *Nordlyd*, 3, 62–77.
- Goldshstein, Yonatan. 2020. Stødets naturlige historie. Y. Goldshstein, I. Schoonderbeek Hansen & T. Thode Hougaard (eds.), *18. Møde om Udforskningen af Dansk Sprog*, Århus, 151–172.
- Gårding, Eva. 1977. *The Scandinavian Word Accents*. Lund: Gleerup.
- Iosad, Pavel. 2016. Tonal Stability and Tonogenesis in (North) Germanic. I. Giles, L. Chapot, C. Coijmans, R. Foster & B. Tesio (eds.), *Beyond Borealism: New Perspectives on the North*. Norvik Press, 78–95.
- Kingston, John. 2011. Tonogenesis. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, & Keren Rice (eds.), *The Blackwell Companion to Phonology*, 1–30.
- Kock, Axel. 1901. *Die alt- und neuschwedische Akzentuierung*. Strassburg: Trübner.
- Köhnlein, Björn. 2016. Contrastive foot structure in Franconian tone-accent dialects. *Phonology* 33, 87–123.
- Liberman, Anatoly. 1982. *The Scandinavian Languages. Germanic Accentology*. Vol. 1. Minneapolis: University of Minnesota Press.
- Lorentz, Ove. 2002. Delayed peak and tonal crowding in Scandinavian tonogenesis. Ms., University of Tromsø.
- Morén-Duolljá, Bruce. 2013. The Prosody of Swedish Underived Nouns: No Lexical Tones Required. *Nordlyd* 40.1, 196–248.
- Myrberg, Sara & Tomas Riad. 2015. The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics* 38, 115–147.
- Oftedal, Magne. 1952. On the Origin of the Scandinavian Tone Distinction. *Norsk tidsskrift for sprogvidenskap* 16, 201–225.
- Perridon, Harry. 2006. On the origin of the Scandinavian word accents. *Zeitschrift für Dialektologie und Linguistik*, Beiheft 131, 91–106.
- Riad, Tomas. 1998. The Origin of Scandinavian Tone Accents. *Diachronica* XV, 63–98.
- Stroh-Wollin, Ulla. 2016. The emergence of definiteness marking in Scandinavian – new answers to old questions. *Arkiv för nordisk filologi* 131, 129–169.
- Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in northern Mon-Khmer. *Phonetica* 46, 60–79.
- Wetterlin, Allison & Aditi Lahiri. 2015. The diachronic development of stød & tonal accent in North Germanic. *Historical Linguistics 2013: Selected Papers from the 21st International Conference on Historical Linguistics, Oslo, 5–9 August 2013*, 53–67. Amsterdam: John Benjamins.

## A panchronic corpus of Old East Slavic and Russian : bringing together Slavic historical and modern corpus resources

A panchronic corpus is a resource representing texts of multiple different historical periods of a given language or a branch of a language group. Typically, most large-scale diachronic corpora capture a given language only within a single historical period (“Old”, “Middle” or “Modern” lect), which is the case, for example, with the COHA language of American English and the GRAC corpus of Ukrainian (both, roughly, encompassing the period of 1820s-2020s). Families of historical corpora may be further divided by centuries, which is the case with historical corpora of Polish (see eg. <https://spxvi.edu.pl/> for the 16th century, <https://sxvii.pl/> for the 17th century, <https://korba.edu.pl/> for 1600-1772).

On the other hand, panchronic corpora are essentially large-scale diachronic corpora encompassing the bulk of the known history of the lect in question. A good example of a panchronic corpus is the *Frantext* database (<https://www.frantext.fr/>) that includes the texts from the whole written history of French starting from the 9<sup>th</sup> to the 21<sup>st</sup> century. A useful tool, it allows for building queries for Old French, Middle French and Modern French alike, but its lemmatization and annotation heavily depends on the modern orthographic and grammatical standard and is far from being accurate even with high-frequency tokens. Another panchronic resource is *Corpus Corporum* of the Zurich University (Roelli 2014), representing different stages of Latin and built as merger of different pre-existing Latin databases.

Panchronic corpora can be used for statistical study of linguistic *phénomènes de longue durée* on different levels, including orthography, morphosyntax, grammaticalized constructions, and semantics. It is of particular use in studying the so-called submerged phenomena (see e. g. for Latin: Adams, Vincent eds. 2016) that are not reflected in written sources during a large timespan but are shared by earliest and latest attestations of the language.

The paper presents the experience in bringing together the existing corpus resources within the Russian National corpus for Old East Slavic (a common ancestor of Russian, Ukrainian and Belarusian), Middle Russian, and Modern Russian, as well as a separate corpus of Old East Slavic birchbark letters. This unified resource is now searchable as the Panchronic corpus within the Russian national corpus. The source corpora had been annotated using different morphological tagsets and lemma standards stemming to different historical dictionaries. Main issues in bringing together these resources are related to mapping correspondences between the Old East Slavic phonetic rendering of lemmas prior to the loss of the short vowels known as yers (ѣ and ѥ), and, further, between Middle and Modern Russian, using rule-based and neural network algorithms with manual post-correction. General phenomena of historical linguistics such as split and merger of different lemmas due to phonetic changes and semantic divergence are discussed within this context. The issue of unified annotation of changing and emerging grammar is also to be addressed, particularly within the context of grammaticalization of East Slavic aspect and animacy.

The panchronic corpus within the RNC is also annotated by semantic classes, using Modern Russian cognates; parallels between this solution and the approach in the Historical thesaurus of English (Kay et al. eds. 2009) are discussed in the talk. As many words changed their semantics drastically this approach has inherent setbacks and should be used with caution but they can be compensated by gains in research availability for the majority of lexicon. This is illustrated in the talk by an example of searching within the panchronic corpus of “lists of sins” (a literary tradition known both in literature and vernacular birchbark writing) using a simple semantic query of three abstract nouns with negative connotation in a row, that yields relevant results in Old East Slavic, Middle Russian, and Early Modern Russian tiers alike.

The technology can be further applied to both Ukrainian and Belarusian (building of a comprehensively annotated Old Ukrainian / Old Belarusian / Ruthenian corpus being a prerequisite) as well, and also to other Slavic languages.

## References

- J. N. Adams, N. Vincent (eds.) 2016. *Early and Late Latin Continuity or Change?* Cambridge, Cambridge University press
- C. Kay et al. (eds.) 2009. *Historical Thesaurus of English*. Glasgow : Glasgow University.
- Ph. Roelli, 2014. The Corpus Corporum, a new open Latin text repository and tool. *Archivum Latinitatis Medii Aevi* 72, 2014

## The Sardinian substrate lexicon and its Mediterranean comparanda

The Sardinian language has a large number of lexical items that cannot have been inherited from Latin, but likely go back to one or more languages spoken on Sardinia before the Roman conquest (Wagner 1932; 1997, 254–80). The rapid developments in ancient DNA studies, which have shown that there is a significant degree of genetic continuity of the Sardinian population from the Neolithic up to the present day (Chiang et al. 2018; Calò et al. 2021), have made conditions more favorable than ever to uncovering and contextualizing the island’s linguistic prehistory. Proposals for the affiliation of Sardinia’s hypothetical substrate language(s) have included Basque, Etruscan, Berber, and less well-defined concepts such as “Eurafrican” or “Hispano-Caucasian” (Hubschmid 1953; Blasco-Ferrer 1988; Pittau 1995; Argiolas 2020).

Here, I zoom in on one set of Sardinian words of presumed pre-Roman origin whose comparanda show a fascinating distribution, being found in the languages of the Iberian peninsula and Southern France, as well as in Greek and languages of the Balkans. Examples of this Mediterranean-wide distribution include: Sardinian *golóstju*, Basque *gorosti*, Lengadocian *agaloûssès*, Greek κήλαστρος ‘holly’; Sardinian *kèya*, Catalan *sitja*, Old Provençal *setgia*, Greek καίατα, καίαδᾶς ‘(storage) pit, silo’; Sardinian *kòsti*, Basque *gazitgar*, Old Provençal *agast*, Greek ἄκαστος ‘maple’ (Hubschmid 1953, 29, 38–39, 80–82).

Although individual Sardinian substrate words with a Mediterranean distribution have been discussed at various occasions in the past, they have not been treated as a defined group. By a detailed discussion of six words with a similar distribution, I aim to shed light on how they relate to each other and on their implications for the pre-Roman linguistic situation of Sardinia as well as that of the Mediterranean region in general. I will do this by addressing the following questions:

1. Do the comparisons withstand formal and semantic scrutiny?
2. If so, how can the observed distribution be accounted for in term of language distribution and/or contact?
3. What (pre-)historical events could have led to the attested situation, based on current insights from archeological and genetic studies?
4. What implications does the Sardinian substrate lexicon have for our understanding of the pre-Indo-European linguistic situation in the Mediterranean region in general, and in Sardinia in particular?

My results will not only shed more light on the linguistic prehistory of the Mediterranean as a whole, but also constitute an interesting case study on the methodology of substrate research.



- Argiolas, Valeria. 2020. “Un État Des Études Sur l’action Du Substrat/Adstrat ‘Libyque’ En Sarde: Problèmes et Méthodes.” *Asinag* 15: 31–61.
- Blasco-Ferrer, Eduardo. 1988. “Il Sostrato Paleosardo: Fine d’un Rebus.” *Zeitschrift Für Romanische Philologie* 127 (1): 63–110.
- Calò, Carla Maria, Giuseppe Vona, Renato Robledo, and Paolo Francalacci. 2021. “From Old Markers to next Generation: Reconstructing the History of the Peopling of Sardinia.” *Annals of Human Biology* 48 (3): 203–12.
- Chiang, Charleston W.K., Joseph H. Marcus, Carlo Sidore, Arjun Biddanda, Hussein Al-Asadi, Magdalena Zoledziewska, Maristella Pitzalis, et al. 2018. “Genomic History of the Sardinian Population.” *Nature Genetics* 50: 1426–34.
- Hubschmid, Johannes. 1953. *Sardische Studien*. Bern: A. Francke.
- Pittau, Massimo. 1995. *Origine e Parentela Dei Sardi e Degli Etruschi: Saggio Storico-Linguistico*. Sassari: Carlo Delfino.
- Wagner, Max Leopold. 1932. “Über Die Vorrömische Bestandteile Des Sardischen.” *Archivum Romanicum: Nuova Rivista Di Filologia Romanza* XV: 207–47.
- . 1997. *La Lingua Sarda. Storia, Spirito e Forma [1950]*. Edited by Giulio Paulis. Nuoro: Ilisso.

# CONTRIBUTIONS BY PARTICIPANT

Participant	Authors' abbrev.	Day	No.	Title
Al-Laith, Ali	Al-Laith et al.	D4	W03.1	A Diachronic Analysis of Using Sentiment Words in Scandinavian Literary Texts from 1870–1900
Alfieri, Luca	Alfieri & Pozza	D1	86	Adjectival typology in four ancient Indo-European languages
Alfieri, Luca	Alfieri	D4	W06.2	On adjectivalizers in Rig-Vedic Sanskrit
Amaral, Patrícia	Amaral et al.	D4	W03.6	Model evaluation for diachronic semantics: A view from Portuguese and Spanish
Andersen, Henning	Andersen	D2	W02.3	Macro-changes at the dawn of history: The Slavic Expansion
Ariztimuño, Borja	Ariztimuño & Salaberri	D1	152	A new perspective on the evolution of mood and negation markers in Proto-Basque
Arnold, Laura	Arnold	D5	W11.4	Tone splits from vowel height in the Austronesian language of Raja Ampat
Assenzi, Lucia	Assenzi	D5	145	Hearsay in Historical German Newspapers (1740–1840)
Auderset, Sandra	Auderset	D2	244	Is tone change more rapid and irregular than segmental change? – A Mixtec case study
Auderset, Sandra	Auderset et al.	D5	W11	The diachrony of tone: connecting the field
Authier, Gilles	Shamseddinov & Authier	D1	31	Contact-driven grammaticalization and drift of new terminal tenses from go-periphrasis in Azeri and Kryz (East Caucasian)
Banerjee, Mithun	Rahman & Banerjee	D2	169	The diachronic study of Bangla case marking system
Barðdal, Jóhanna	Cluyse et al.	D1	208	Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis
Barðdal, Jóhanna	Elens et al.	D1	213	The Alternating Behavior of ‘Like’ in Old Norse-Icelandic: Facts or Fiction
Bartolotta, Annamaria	Bartolotta	D1	W05.2	The right-left conceptual mapping in a comparative and diachronic perspective
Baudel, Étienne	Baudel et al.	D4	W10	The (Pre)History of the Languages of Japan – Current issues and prospects
Baudel, Étienne	Baudel et al.	D4	W10.6	The (Pre)History of the Languages of Japan – Current issues and prospects
Bauer, Brigitte L. M.	Bauer	D5	191	Complexity in counting systems: early systems vs. modern numerical ones
Belelli, Sara	Belelli	D5	W08.6	A historical-comparative glimpse on Laki dialects
Beniamine, Sacha	Round et al.	D2	196	The natural stability of ‘unnatural’ morphology
Benvenuto, Maria Carmela	Benvenuto & Bichlmeier	D4	278	The expression of predicative possession in Avestan
Berge, Anne	Berge	D1	W01.2	Prehistoric climate changes and their effects on the development of the Eskaleut languages
Biagetti, Erica	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin
Biagetti, Erica	Brigada Villa et al.	D2	W14.2	Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages
Bichlmeier, Harald	Benvenuto & Bichlmeier	D4	278	The expression of predicative possession in Avestan
Billing, Oscar	Billing & Elgh	D5	178	Computational Anatolian phylogeny using maximum parsimony

Participant	Authors' abbrev.	Day	No.	Title
Billing, Oscar	Rönchen et al.	D4	W03.7	Using simulated data to evaluate models of Indo-European vocabulary evolution
Bjørn, Rasmus G.	Bjørn & Kilani	D1	W07.0	Interactions at the dawn of history: An introduction to the workshop
Bjørn, Rasmus G.	Bjørn & Kilani	D1	W07	Interactions at the dawn of history: Methods and results in prehistoric contact linguistics
Björnsdóttir, Sigríður	Björnsdóttir et al.	D1	203	The rise of raising in Early Modern English
Bloom, Barthe	Bloom	D3	80	Early New High German preposed adverbial clauses: integration and discourse functions
Blum, Frederic	Blum & List	D2	W14.6	A computational evaluation of regularly recurring sound correspondences
Bogdanowska-Jakubowska, Ewa	Bogdanowska-Jakubowska & Bogdanowska	D3	75	Changes in the Polish address practices after the Second World War
Bogdanowska, Nika	Bogdanowska-Jakubowska & Bogdanowska	D3	75	Changes in the Polish address practices after the Second World War
Bonmann, Svenja	Bonmann et al.	D2	32	Towards a New Reconstruction of the Proto-Yeniseian Sound System
Börjars, Kersti	Börjars & Vincent	D1	52	Auxiliary, light or lexical: the history of GO verbs
Bossuyt, Tom	Bossuyt & Daveloose	D2	104	Divergence and contact in Cappadocian concessive conditionals
Bostoen, Koen	Gunnink et al.	D5	232	An evolutionary loner in Southern African Bantu: The classification of Yeyi
Bostoen, Koen	Bostoen et al.	D1	W07.2	Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction
Bostoen, Koen	Pacchiarotti et al.	D5	69	Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu
Boye, Kasper	Boye	D3	93	Grammaticalization as conventionalization of discursively secondary status: Isolating what is unique to grammaticalization, and deconstructing the lexical-grammatical continuum
Boye, Kasper	Vincent et al.	D4	PL6	Linguistic models (with a focus on morphosyntactic change)
Boye, Kasper	Westergaard & Boye	D3	90	On semantic change in grammaticalization: Why it is never metaphoric
Boyeldieu, Pascal	Boyeldieu	D2	255	Tone split and tone replacement: toward the three-tone system of the 'Western' SBB Languages (Central Sudanic, Central Africa)
Božović, Đorđe	Božović	D5	W11.1	Tone, stress and length interactions in Central Neo-Štokavian
Bradley, David	Bradley	D1	W01.5	Climate change and the dispersal of Proto-Tibeto-Burman
Brigada Villa, Luca	Brigada Villa et al.	D2	W14.2	Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages
Bronikowska, Renata	Bronikowska	D5	40	Middle Polish adverb-like predicates ending in -a compared to other adverbial and adjectival predicates – corpus-based approach
Brosig, Benjamin	Brosig & Dolgor	D2	W12.2	From spatial noun to medial demonstrative: the case of Khalkha Mongolian
Brown, Braden	Brown & Grollemund	D5	230	Towards a new classification of Western Bantu languages using non-lexical data
Bru, Mathilde	Bru	D2	72	'So wrong that not even Menander uses it!': the Atticist lexicographers on the Ancient Greek dialects
Brunner, Thomas	Brunner	D4	112	The ordering of matrix clauses and subordinate causal clauses in the Old Bailey Corpus 1720–1913

Participant	Authors' abbrev.	Day	No.	Title
Bugaeva, Anna	Satō & Bugaeva	D4	W10.1	On stative/active intransitive split within tripartite alignment: A case of Kuril Ainu
Calabrese, Andrea	Calabrese	D4	W06.1	Inflectional vocalic pieces in Latin verbal morphology: a synchronic and diachronic analysis
Camilleri, Maris	Camilleri	D4	W09.5	Parallels in the development from locative and existential predications to possessive structures in Arabic and Hebrew
Capano, Marta	Capano	D5	237	It Ain't Over till It's Over. Bilingualism and language decay in Sicilian inscriptions
Caso, Anabelle	Caso & Hale	D2	273	Secondary predication in metrical texts: syntax-prosody mapping in ancient Indo-European languages
Cassarà, Alessia	Cassarà et al.	D4	W13.3	Marked vs. unmarked unaccusativity with alternating verbs: Linking diachronic and experimental data
Cassarà, Alessia	Cassarà et al.	D4	W13	New methods for old languages: the comparability of data
Cassarà, Alessia	Cassarà et al.	D4	W13.0	New methods for old languages: the comparability of data
Cathcart, Chundra	Herce & Cathcart	D5	199	Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure
Cattafi, Eleonora	Cattafi	D5	264	Continuative relative clauses in Greek documentary papyri
Céline, Mounole	Manterola et al.	D1	58	The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments
Cennamo, Michela	Cennamo	D4	173	Existential HAVE in Late Latin: insights on its diachrony in the passage to Romance
Chankova, Yana	Chankova	D2	97	As Syntax Interfaces with Information Structure: Old Icelandic Non-Canonical Scrambled Orders
Cluyse, Brian	Cluyse et al.	D1	208	Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis
Coenen, Pascal	Coenen	D1	127	The totalizing function of the Vedic particle <i>cid</i>
Concu, Valentina	Concu	D2	122	The use of "thank" and "to thank" in Old Saxon and Old High German
Conradie, Jac	Conradie	D1	59	The Afrikaans auxiliary <i>het</i> 'have' from clitic to desinence
Cornillie, Bert	Inglese et al.	D4	100	The anticausative alternation in Italian and Spanish: a historical corpus-based perspective
Creissels, Denis	Creissels	D4	W09.4	'Be/have' verbs in historical perspective
Currie, Oliver	Currie	D3	63	The emergence of a Welsh biblical literary standard and the evidence of early modern manuscript sermons
Cuyckens, Huybert	Nijs et al.	D2	W02.4	An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German
Däbritz, Chris Lasse	Däbritz	D4	W09	"Your birch-bark bag has something" – Grammaticalization and diachrony of locative, existential and possessive predication
Däbritz, Chris Lasse	Däbritz	D4	W09.1	"Your birch-bark bag has something" – Grammaticalization and diachrony of locative, existential and possessive predication
Darling, Mark	Darling et al.	D2	180	The Diachrony of Person-Number Marking of Subjects in Celtic
Das, Patrick	Das	D5	266	Areality through Migration: Investigating the Structure of Numeral Classifiers in the Eastern Himalayan Region Reveals Historic Contact Events
Daveloose, Eline	Bossuyt & Daveloose	D2	104	Divergence and contact in Cappadocian concessive conditionals
Daveloose, Eline	Daveloose	D3	137	From <i>de</i> to <i>ke</i> : functional transfer of a topic shift marker from Turkish to Cappadocian Greek

Participant	Authors' abbrev.	Day	No.	Title
De Cesare, Ilaria	Zehentner & De Cesare	D2	W04	Ambiguity (avoidance) as a factor in language change
de Rossi, Nicolò	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin
De Smet, Hendrik	Kayenbergh & De Smet	D5	89	Just a bystander? Semantic change in the English simple tenses
De Smet, Hendrik	Felser	D2	W04.3	Structural ambiguity in language comprehension and production
de Vaan, Michiel	Neri & de Vaan	D2	W12.6	Origin and development of the Albanian demonstratives
de Vos, Machteld	de Vos	D4	121	Spread the German new(s): third-person reflexive <i>zich</i> in 17th-century Dutch newspapers
Dedvukaj, Lindon	Dedvukaj	D1	68	Reanalyzing the Historical Constructions of Albanian Prepositions
Degaetano-Ortlieb, Stefania	Jenset et al.	D4	W03.2	Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i>
Degaetano-Ortlieb, Stefania	Degaetano-Ortlieb et al.	D4	W03	Computational models of diachronic language change
Dellert, Johannes	Dellert & Blaschke	D2	W14.5	Configurable Language-Specific Tokenization for CLDF Databases
Demolin, Didier	Ricquier & Demolin	D3	102	The Chronicle of Lingbe, an Extinct Bantu Language of East Congo
Deng, Bingcong	Deng	D1	W01.6	Climate change reflected in early Sino-Tibetan borrowings for crops and animals
Deo, Ashwini	Vincent et al.	D4	PL6	Linguistic models (with a focus on morphosyntactic change)
Dereza, Oksana	Dereza et al.	D4	W03.8	Evaluating historical word embeddings: strategies, challenges and pitfalls
di Bartolo, Giuseppina	di Bartolo	D3	154	Where and How? Request verb constructions in Ancient Greek
Dinu, Liviu P.	Dinu et al.	D3	279	Computational approaches for Romance related words discrimination
Dockum, Rikker	Dockum & Lu	D2	245	Beyond the paradigm: Change and expansion in Thai pronominal reference
Dockum, Rikker	Dockum & Wang	D4	274	Quality vs. quantity: Contrast maintenance and tradeoff in Southwestern Tai vowels
Dockum, Rikker	Auderset et al.	D5	W11	The diachrony of tone: connecting the field
Dolgor, Guntsetseg	Brosig & Dolgor	D2	W12.2	From spatial noun to medial demonstrative: the case of Khalkha Mongolian
Dömötör, Adrienne	Dömötör	D2	175	From direct quotation to a chain of extended quotations: the history of Hungarian <i>úgymond</i> 'so to speak'
Drach, Mortimer	Gelumbeckaitė et al.	D3	247	The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian"
Drinka, Bridget	Drinka et al.	D2	W02	Macro-level social motivations for language change: Contact, migration, and globalization
Drinka, Bridget	Drinka et al.	D2	W02.1	Macro-level social motivations for language change: Contact, migration, and globalization
Dücker, Lisa	Dücker	D4	171	Semantic factors influencing the change in position of German adnominal genitives in the 17th to 19th centuries
Egedi, Barbara	Egedi	D2	194	Demonstrative modifiers in Middle Hungarian: a complex picture of renewal
Elgh, Erik	Billing & Elgh	D5	178	Computational Anatolian phylogeny using maximum parsimony

Participant	Authors' abbrev.	Day	No.	Title
Ellison, T. Mark	Reinöhl & Ellison	D5	190	Metaphor, Overtness and Word Order Routinization
Elter, W. Juliane	Elter	D5	183	Anglo-Scandinavian Contact Influence on Verbs Entering the Causative Alternation
Engelberg, Stefan	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language
Enrique-Arias, Andrés	Enrique-Arias	D2	W02.8	Political influence as a factor in morphosyntactic variation: demonstratives <i>este</i> and <i>aqueste</i> in medieval Aragonese
Esher, Louise	Esher	D4	105	Gascon <i>u</i> -perfects and the analogical foregrounding of inflectional class
Esher, Louise	Round et al.	D2	196	The natural stability of 'unnatural' morphology
Espíndola Moschner, Silvina	Espíndola Moschner & Rosemeyer	D1	148	Aspectual uses of <i>saber</i> + infinitive in South American Spanish varieties: a corpus-based study
Étienne, Baudel	Baudel	D4	W10.4	Reconsidering the classification of Hachijō: A glimpse from historical phonology
Eyþórsson, Þórhallur	Eyþórsson & Sigurðardóttir	D2	126	Micro-level conflict in the productivity of anticausativization strategies: Evidence from the history of Icelandic
Farina, Andrea	Farina et al.	D2	114	WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose
Feltgen, Quentin	Feltgen	D1	113	The shape of grammaticalization: matching the bridging context scenario with patterns of frequency use
Figura, Lisa	Figura	D4	W13.1	Dative Experiencer Psych Verbs in (Old) French
Flaksman, Maria	Flaksman	D3	150	Lost in translation. Onomatopoeic words in Old English glosses
Fonteyn, Lauren	Degaetano-Ortlieb et al.	D4	W03	Computational models of diachronic language change
Fonteyn, Lauren	Fonteyn et al.	D2	29	From ecological to lexical diversity: measuring vocabulary richness in historical corpora
Forkel, Robert	Jäger et al.	D2	W14	Exploiting Standardized Cross-Linguistic Data in Historical Linguistics
Forkel, Robert	Forkel & Greenhill	D2	W14.4	Phlorest: A Database of Consistent and Reusable Language Phylogenies
Franco, Karlien	Franco	D5	168	Explaining the speed of lexical change in historical Dutch
Fransen, Theodorus	Dereza et al.	D4	W03.8	Evaluating historical word embeddings: strategies, challenges and pitfalls
Fried, Mirjam	Vincent et al.	D4	PL6	Linguistic models (with a focus on morphosyntactic change)
Friedman, Victor	Friedman	D2	98	Obscenity as a Window into Slavic Linguistic History
Fries, Simon	Bonmann et al.	D2	32	Towards a New Reconstruction of the Proto-Yeniseian Sound System
Fromm, Nathalie	Fromm	D1	129	The development of number strengthening in German declensional classes. A diachronic-dialectal corpus study
Funk, Ekaterina	Tikhonov et al.	D2	143	Pronoun history and information structure in 18th century non-religious Russian texts
Gehrmann, Ryan	Auderset et al.	D5	W11	The diachrony of tone: connecting the field
Gelumbeckaitė, Jolanta	Gelumbeckaitė et al.	D3	247	The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian"
Gfeller, Kim	Gfeller	D1	234	Persistence and Change of Colexifications in Indo-European
Gholami, Saloumeh	Karim & Gholami	D5	W08	Filling in the diachronic gaps: the view of Old Iranian from the present

Participant	Authors' abbrev.	Day	No.	Title
Gholami, Saloumeh	Karim & Gholami	D5	W08.1	Filling in the diachronic gaps: the view of Old Iranian from the present
Gholami, Saloumeh	Gholami & Naghshbandi	D5	W08.5	Polyptoton for the purpose of emphasizing within Iranian languages
Gibson, Hannah	Gibson et al.	D4	227	Morphosyntactic variation in Swahili: Tracing descriptions past and present
Ginevra, Riccardo	Ginevra	D1	W05.3	Indo-European Poetics meets Cognitive Linguistics: an integrated approach to the comparative reconstruction of metaphoric and metonymic expressions
Gisborne, Nikolas	Gisborne & Truswell	D4	222	Contact and the origins of headed <i>wh</i> -relatives in Hungarian
Giuliani, Martina	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin
González Saavedra, Berta	Orqueda & González Saavedra	D2	W12	From and Towards Demonstratives: Grammaticalization Processes and Beyond
Gopal, Deepthi	Gopal et al.	D1	240	Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture
Gosemann, Laura	Gosemann	D4	118	Syntactic change and DLM in German: a corpus study
Gotthard, Lisa	Gotthard	D1	214	The rise of <i>do</i> -support during Scots anglicisation: Insights from the <i>Parsed Corpus of Scottish Correspondence</i>
Gotthard, Lisa	Björnsdóttir et al.	D1	203	The rise of raising in Early Modern English
Gray, Russell	Shcherbakova et al.	D2	96	Diachronic pathways of definite articles distribution
Gray, Russell	Verkerk et al.	D3	242	Exploiting phylogenetic modeling to uncover directionality in the emergence of universals
Grestenberger, Laura	Grestenberger et al.	D4	W06	Categorizers in diachrony
Grestenberger, Laura	Grestenberger et al.	D4	W06.0	Categorizers in diachrony
Grimm, Nadine	Grimm	D5	W11.5	A diachronic study of grammatical tone in northwestern Bantu
Grollemund, Rebecca	Brown & Grollemund	D5	230	Towards a new classification of Western Bantu languages using non-lexical data
Grossman, Eitan	Sæbø & Grossman	D5	W11.6	A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis
Gugán, Katalin	Gugán	D1	228	Outliers in variation and change: atypical users of the variants of negation in Old and Middle Hungarian
Gunnink, Hilde	Gunnink et al.	D5	232	An evolutionary loner in Southern African Bantu: The classification of Yeyi
Günther, Laura	Bonmann et al.	D2	32	Towards a New Reconstruction of the Proto-Yeniseian Sound System
Guzmán Naranjo, Matías	Mertner & Guzmán Naranjo	D2	W14.7	Exploring the Geographical Distribution of Missing Data Using Approximate Gaussian Processes
Guzmán Naranjo, Matías	Jonjić et al.	D2	144	Isoglosses and distributions of features – Analyses of the <i>Dialectological Atlas of the Russian Language</i>
Gvozdanović, Jadranka	Gvozdanović	D2	W02.6	Ideology, language choice and language change
Hakimov, Nikolay	Hakimov	D4	251	Fall of the jers: A multi-factorial analysis of the sound change progression in the Old Novgorodian birchbark texts
Hale, Mark	Caso & Hale	D2	273	Secondary predication in metrical texts: syntax-prosody mapping in ancient Indo-European languages



Participant	Authors' abbrev.	Day	No.	Title
Halfmann, Jakob	Halfmann & Korobzow	D3	239	The Evolution of Spatial Orientation Systems in Mayan and Nuristani
Hamans, Camiel	Hamans	D1	67	A revolution in the history of affix-formation
Hansen, Magnus Pharao	Hansen & Davletshin	D1	W07.1	Tracing borrowings in and out of proto-Nahuatl
Hartmann, Stefan	Pleyer et al.	D1	128	The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective
Harvey, Mark	Mailhammer & Harvey	D1	184	The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database
Haspelmath, Martin	Haspelmath	D2	W04.7	Ambiguity avoidance vs. expectation sensitivity as functional factors in language change and language structures: Beyond argument marking
Hasselbach-Andee, Rebecca	Hasselbach-Andee	D4	W06.3	One or All: The Development of Singulatives to Collectives in Semitic
Heggarty, Paul	Heggarty	D1	W01.8	Languages, ecology and climate change: Worldwide perspectives and the test-case of the Andes
Hengeveld, Kees	Hengeveld	D4	W09.3	The development of locative, existential and possessive predication from a functional perspective
Herce, Borja	Herce & Cathcart	D5	199	Stem shortening in Romance verbs: the 'S morpheme' at the intersection of token frequency and paradigmatic structure
Hernáiz, Rodrigo	Hernáiz	D5	215	Exploring language variation and change in the distant past
Hill, Eugen	Bonmann et al.	D2	32	Towards a New Reconstruction of the Proto-Yeniseian Sound System
Hirvonen, Johannes	Hirvonen	D5	200	Contact-induced change of Negative Indefinites – the case of Meadow Mari
Hofmann, Klaus	Ritt & Hofmann	D2	248	'Chained to the rhythm': Using agent-based simulation to model the evolution of stress pattern diversity in English
Hofmann, Klaus	Hofmann	D2	157	Evolving rhythms: A quantitative assessment of rhythmic alternation in the history of English
Holopainen, Sampsa	Holopainen	D4	238	The emergence of word-initial voiced stops in Proto-Hungarian
Honeybone, Patrick	Honeybone	D2	188	Can fortis stops spirantise without aspiration?
Hu, Hai	Amaral et al.	D4	W03.6	Model evaluation for diachronic semantics: A view from Portuguese and Spanish
Hualde, José Ignacio	Hualde	D2	57	The diachrony of Basque accentuation: comparative method and internal reconstruction
Hualde, José Ignacio	Manterola et al.	D1	58	The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments
Huang, Yang	Huang	D2	56	The Expression of Negation in Sabde Minyag
Huback, Ana Paula	Huback & Fontes Martins	D2	94	R Deletion in Brazilian Portuguese: Diachronic and Synchronic Evidence for Lexical Diffusion
Hudson, Mark	Hudson	D1	W01.10	Risk, resilience and the ecology of farming/language dispersals
Idiatov, Dmitry	Idiatov	D2	65	Vowel reduction to /i/ in functional morphemes in Northern Sub-Saharan Africa
Iezzi, Luca	Iezzi	D5	108	The role of French in the Johnsons' correspondence
Igartua, Iván	Igartua	D4	254	Exploring the sources of animacy distinctions
Inglese, Guglielmo	Inglese et al.	D4	100	The anticausative alternation in Italian and Spanish: a historical corpus-based perspective

Participant	Authors' abbrev.	Day	No.	Title
Ionov, Max	Gelumbeckaitė et al.	D3	247	The Postil Time Machine: "God help those who have begun writing down these books in Lithuanian"
Ishiyama, Osamu	Ishiyama	D2	W12.3	On the Development of Demonstratives into Personal Pronouns
Jäger, Gerhard	Jäger et al.	D2	W14	Exploiting Standardized Cross-Linguistic Data in Historical Linguistics
Janda, Richard	Janda & Joseph	D1	135	West Germanic 2.sg. <i>-st</i> Revisited: The Role of Superscence
Jarosz, Aleksandra	Baudel et al.	D4	W10	The (Pre)History of the Languages of Japan – Current issues and prospects
Jarosz, Aleksandra	Baudel et al.	D4	W10.6	The (Pre)History of the Languages of Japan – Current issues and prospects
Jenkins, Chris	Maurer et al.	D4	W03.3	Quantifying Changes in English Noun Compound Productivity and Meaning
Jensen, Eva Skafte	Jensen & Schack	D4	84	Adverbs ending in <i>-(l)ig</i> '-ly' and <i>-(l)igt</i> '-ly' in Danish
Joseph, Brian	Joseph	D1	W01.9	(Im)mobility, climate, and language: Towards a geoanthropology of the Balkans
Joseph, Brian	Janda & Joseph	D1	135	West Germanic 2.sg. <i>-st</i> Revisited: The Role of Superscence
Juge, Matthew	Juge	D1	270	The dominant-recessive hypothesis does not account for overlapping suppletion
Kaltenbach, Lena	Cassarà et al.	D4	W13	New methods for old languages: the comparability of data
Kaltenbach, Lena	Cassarà et al.	D4	W13.0	New methods for old languages: the comparability of data
Kamil, Iris	Grestenberger et al.	D4	W06	Categorizers in diachrony
Kamil, Iris	Grestenberger et al.	D4	W06.0	Categorizers in diachrony
Karim, Shuan Osman	Karim & Gholami	D5	W08	Filling in the diachronic gaps: the view of Old Iranian from the present
Karim, Shuan Osman	Karim & Gholami	D5	W08.1	Filling in the diachronic gaps: the view of Old Iranian from the present
Kauhanen, Henri	Gopal et al.	D1	240	Correlations between linguistic features are reflected in their geospatial patterning: Introducing the geo-typological Sandwich Conjecture
Kayenbergh, Juliette	Kayenbergh & De Smet	D5	89	Just a bystander? Semantic change in the English simple tenses
Kilani, Marwan	Bjørn & Kilani	D1	W07.0	Interactions at the dawn of history: An introduction to the workshop
Kilani, Marwan	Bjørn & Kilani	D1	W07	Interactions at the dawn of history: Methods and results in prehistoric contact linguistics
Kim, Ronald I.	Kim	D5	W08.3	Steppe Iranian in the <i>longue durée</i> : contact, relative chronology, and internal reconstruction
Kinuhata, Tomohide	Kinuhata	D4	W10.3	Reconstructing the Proto-Japonic demonstrative system
Kiparsky, Paul	Kiparsky	D1	PL1	The word-order cycle
Kirby, James	Kirby & Pittayaporn	D5	W11.3	Tone and voicing in Cao Bằng Tai: implications for tonal evolution and change
Kisiel, Anna	Kisiel & Sobotka	D2	51	The functional interpretation of semantic and syntactic shifts in the domain of North Slavic "conversive" preposition-pronominal constructions
Klævik-Pettersen, Espen	Klævik-Pettersen	D4	261	VSO orders in the Egeriae and Antonini Placentini itineraria; new evidence for the evolution towards Old Romance inversion systems
Klosa-Kückelhaus, Annette	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language

Participant	Authors' abbrev.	Day	No.	Title
Knapen, Martijn	Knapen	D1	W01.3	Seals and sea ice: the (possible) climatic background of Amuric influence on Ainu
Kölligan, Daniel	Kölligan	D1	W05.7	Conceptual metaphors and etymology
Kölligan, Daniel	Kölligan & van Beek	D1	W05	Conceptual metaphors in a comparative and diachronic perspective
Kölligan, Daniel	Kümmel et al.	D1	124	The development of future-referring constructions (in Indo-European languages)
Korobzow, Natalie	Halfmann & Korobzow	D3	239	The Evolution of Spatial Orientation Systems in Mayan and Nuristani
Korobzow, Natalie	Bonmann et al.	D2	32	Towards a New Reconstruction of the Proto-Yeniseian Sound System
Kozhanov, Kirill	Kozhanov	D2	197	Diachronic stability of case functions: oblique in Romani dialects
Krasnoukhova, Olga	Krasnoukhova et al.	D4	W09.2	Negated but similar – Negation in the domains of locative, existential, and possessive predication: The case of Indo-European
Kreidl, Julian	Kreidl	D5	W08.2	Bactrian influence on local languages of Eastern Afghanistan
Krielke, Marie-Pauline	Degaetano-Ortlieb et al.	D4	W03	Computational models of diachronic language change
Kübler, Sandra	Amaral et al.	D4	W03.6	Model evaluation for diachronic semantics: A view from Portuguese and Spanish
Kümmel, Martin Joachim	Kümmel et al.	D1	124	The development of future-referring constructions (in Indo-European languages)
Lahiri, Aditi	Lahiri	D2	PL2	Phonological grammars: Pertinacious constraints on change
Landwehr, Isabell	Jenset et al.	D4	W03.2	Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i>
Larrivèe, Pierre	Poletto et al.	D4	36	Learning how to count – a treebank analysis of V2 word order in two Medieval Romance languages through time
Leddy-Cecere, Thomas	Leddy-Cecere	D1	116	The PRESENTATIVE > DEMONSTRATIVE Grammaticalization Pathway in Arabic
Lindgren, Freja	Lindgren & Tresoldi	D3	198	The Charition Mime: Decoding the “Indian Language” through Typology and Entropy
Lionnet, Florian	Lionnet	D5	W11.2	Accent and tone: the double origin of the Paicî tone system
Lionnet, Florian	Lionnet	D2	225	Areal alignment and the loss of ATR harmony in Riverine Bua languages (Chad)
List, Johann-Mattis	Blum & List	D2	W14.6	A computational evaluation of regularly recurring sound correspondences
List, Johann-Mattis	Jäger et al.	D2	W14	Exploiting Standardized Cross-Linguistic Data in Historical Linguistics
Litvinova, Lora	Litvinova	D2	258	Reconstructing the Kugama tone system
Luján, Eugenio	Luján & Ngomo Fernández	D2	W12.7	From demonstratives to articles in the Celtic languages
Luraghi, Silvia	Brigada Villa et al.	D2	W14.2	Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages
Luraghi, Silvia	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin
Maiden, Martin	Maiden	D1	43	A morphological freeloader: Ibero-Romance <i>cabrer</i>
Mailhammer, Robert	Mailhammer & Harvey	D1	184	The Comparative Method on a shoestring: Evaluating chance vs inheritance with a limited database

Participant	Authors' abbrev.	Day	No.	Title
Majtczak, Tomasz	Majtczak	D4	W10.5	Old, Middle and New: Periodisation as a back-burnered topic in the diachronic research of Japanese
Manterola, Julen	Mounole & Manterola	D1	265	From distal demonstrative to resultative marker (through definite article): evidence from Basque
Manterola, Julen	Manterola et al.	D1	58	The history of the Basque pronoun <i>zuek</i> 'you.all' in relation to similar Romance developments
Markopoulos, Theodore	Markopoulos	D3	167	Epistemic modality out of 'playfulness': Modern Greek <i>pezi</i>
Marr, Clayton	Marr	D4	W03.4	A computerized investigation of Albanian diachronic phonology
Marten, Lutz	Marten	D3	PL3	Historical linguistics and <i>ubuntu</i> translanguaging: Towards a model of multilingualism, language change and linguistic convergence in the Bantu Linguistic Area
Marten, Lutz	Gibson et al.	D4	227	Morphosyntactic variation in Swahili: Tracing descriptions past and present
Maselli, Lorenzo	Bostoen et al.	D1	W07.2	Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction
Maselli, Lorenzo	Pacchiarotti et al.	D5	69	Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu
Maurer, Maximilian	Maurer et al.	D4	W03.3	Quantifying Changes in English Noun Compound Productivity and Meaning
Mazzola, Giulia	Rosemeyer et al.	D5	138	A computational approach to detect discourse traditions and register differences: a case study on historical French
Mazzola, Giulia	Inglese et al.	D4	100	The anticausative alternation in Italian and Spanish: a historical corpus-based perspective
McCarley, Gemma	McCarley	D4	233	Diachronic Null Subject Use across Latin American Spanish: Comparing Corpora
McCrae, John P.	Dereza et al.	D4	W03.8	Evaluating historical word embeddings: strategies, challenges and pitfalls
McGillivray, Barbara	Jenset et al.	D4	W03.2	Computational linguistic modelling of the temporal dynamics of scientific communication: a quantitative corpus study on the journal <i>Nature</i>
McGillivray, Barbara	Farina et al.	D2	114	WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose
Meisterernst, Barbara	Meisterernst	D1	211	The diachronic development of future markers in Chinese
Mendoza, Imke	Mendoza et al.	D4	99	Anchoring patterns in emerging complement clauses in Slavic
Mertner, Miri	Mertner & Guzmán Naranjo	D2	W14.7	Exploring the Geographical Distribution of Missing Data Using Approximate Gaussian Processes
Mesthrie, Rajend	Mesthrie	D2	W02.9	Macro sociohistorical forces, contact, convergence and the development of modern linguistic areas: insights from South Africa
Meyer, Peter	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language
Meyer, Robin	Meyer	D2	187	Quasi- <i>Suffixaufnahme</i> in Classical Armenian
Meyer, Roland	Tikhonov et al.	D2	143	Pronoun history and information structure in 18th century non-religious Russian texts
Miletic, Filip	Maurer et al.	D4	W03.3	Quantifying Changes in English Noun Compound Productivity and Meaning

Participant	Authors' abbrev.	Day	No.	Title
Mirelman, Sam	Mirelman	D3	74	Translation as Royal Legitimation: The Concepts of "Source" and "Target" Language in Sumerian-Akkadian Royal Inscriptions from the Old Babylonian Period (2000–1600 BC)
Mithun, Marianne	Mithun	D2	W12.1	Further Pathways Towards Demonstratives
Mofidi, Roohollah	Mofidi	D1	55	Competition in the aspect-mood domain: The standardization of a diachronic data set of New Persian
Mohammadirad, Masoud	Mohammadirad	D5	W08.4	Remarks on the category of copula in Gorani dialects
Mounole, Céline	Mounole & Manterola	D1	265	From distal demonstrative to resultative marker (through definite article): evidence from Basque
Mous, Maarten	Mous	D5	192	The classification of South Cushitic
Munteanu, Andrei	Munteanu	D3	73	Automating Comparative Reconstructions: Case Study in Austronesian and Ongan
Næss, Åshild	Næss	D2	W12.5	Demonstratives taking over discourse: the grammaticalisation of deictic clitics in Äiwoo
Neels, Jakob	Pleyer et al.	D1	128	The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective
Neri, Sergio	Neri & de Vaan	D2	W12.6	Origin and development of the Albanian demonstratives
Nevalainen, Terttu	Drinka et al.	D2	W02	Macro-level social motivations for language change: Contact, migration, and globalization
Nevalainen, Terttu	Drinka et al.	D2	W02.1	Macro-level social motivations for language change: Contact, migration, and globalization
Ngomo Fernández, Esteban	Luján & Ngomo Fernández	D2	W12.7	From demonstratives to articles in the Celtic languages
Nichols, Johanna	Nichols	D2	W02.2	Reconstructing prehistoric sociolinguistics from modern grammatical evidence
Nieder, Jessica	Nieder & Tomaschek	D5	174	Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics
Nijs, Julie	Nijs et al.	D2	W02.4	An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German
Ochs, Samira	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language
Olivier, Marc	Olivier	D1	217	When change fails: evidence from French
Ongenaë, Tim	Ongenaë	D2	66	Towards a Diachronic Account of P-lability in Latin: The Semantic Extension of the Active Intransitive as an Anticausative Strategy in Latin
Orlandi, Georg	Baudel et al.	D4	W10	The (Pre)History of the Languages of Japan – Current issues and prospects
Orlandi, Georg	Baudel et al.	D4	W10.6	The (Pre)History of the Languages of Japan – Current issues and prospects
Orqueda, Verónica	Orqueda & González Saavedra	D2	W12	From and Towards Demonstratives: Grammaticalization Processes and Beyond
Orqueda, Verónica	Orqueda & Pooth	D2	W12.8	Latin <i>ecce</i> : arguments in favor of its development from a PIE demonstrative
Pacchiarotti, Sara	Bostoen et al.	D1	W07.2	Pre-Bantu substrate in Batwa Bantu languages of the Congo rainforest: A comparative study of nasal-oral stop cluster reduction
Pacchiarotti, Sara	Pacchiarotti et al.	D5	69	Uncovering lost paths in the Congo rainforest: A new, comprehensive phylogeny of West-Coastal and Central-Western Bantu

Participant	Authors' abbrev.	Day	No.	Title
Pache, Matthias	Pache	D5	166	Evidence for a Chibcha-Jê connection
Pan, Tao	Pan	D3	38	<i>ille ego</i> and Recognitional Use of Demonstratives
Paterson, Hugh	Paterson, H.	D4	282	Proto-Malayo-Polynesian: Some Phonetic Evidence for */
Paterson, Rebecca	Paterson, R.	D1	280	Emergence of alternate argument alignment patterns in Northwest Kainji
Peck, Naomi	Reinöhl et al.	D2	181	The loss of word-initial consonants in Kera'a – A challenge for phonological theory
Perekhval'skaya, Elena	Perekhval'skaya & Vydrin	D5	W11.7	Tonal density and its correlation with the types of tonal systems: Diachronic aspects
Persohn, Bastian	Persohn	D1	37	When 'still' comes to signal a near past
Petré, Peter	Petré	D1	206	Conservative pressure on the progressive: the passival
Piccione, Mariapaola	Cassarà et al.	D4	W13	New methods for old languages: the comparability of data
Piccione, Mariapaola	Cassarà et al.	D4	W13.0	New methods for old languages: the comparability of data
Pierce, Marc	Pierce	D2	158	The History of /pf/ in New Braunfels German: Another Case of Rule Inversion?
Pinzin, Francesco	Poletto et al.	D4	36	Learning how to count – a treebank analysis of V2 word order in two Medieval Romance languages through time
Pittayaporn, Pittayawat	Kirby & Pittayaporn	D5	W11.3	Tone and voicing in Cao Bằng Tai: implications for tonal evolution and change
Pleyer, Michael	Pleyer et al.	D1	128	The Interaction of the Cognitive and Community Level in Language Evolution: A Usage-Based Perspective
Pompei, Anna	Pompei	D4	272	The case of Italian <i>segunte</i> : an European instance of current change from verb to demonstrative?
Pompeo, Flavia	Pompeo	D1	W05.4	New meanings and old constructions: the conceptualization of 'fearing' and 'protecting' in Old Persian in comparison with other Indo-Iranian languages
Potochnik, Thomas	Potochnik	D3	236	Doing Conversation Analysis in Latin: The Case of Hedging
Pounder, Amanda	Pounder	D4	275	Morphologization of Phonological Processes as Integration
Pozza, Marianna	Alfieri & Pozza	D1	86	Adjectival typology in four ancient Indo-European languages
Pronk, Tijmen	Pronk	D2	189	Tonogenesis in Baltic and Slavic languages
Rainsford, Tom	Trips & Rainsford	D4	W13.2	How to use Yang's Principles to model acquisition in diachrony. The case of psych verbs
Rapold, Christian	Rapold	D2	268	Secondary lateral obstruents in South Cushitic and their significance for the linguistic history of East Africa
Razguliaeva, Mariia	Tikhonov et al.	D2	143	Pronoun history and information structure in 18th century non-religious Russian texts
Reetz, Malika	Reetz	D4	170	German V2-Argument Clauses from a Diachronic Perspective
Reinöhl, Uta	Reinöhl & Ellison	D5	190	Metaphor, Overtness and Word Order Routinization
Reinöhl, Uta	Reinöhl et al.	D2	181	The loss of word-initial consonants in Kera'a – A challenge for phonological theory
Reiter, Viktoria	Grestenberger et al.	D4	W06	Categorizers in diachrony
Reiter, Viktoria	Grestenberger et al.	D4	W06.0	Categorizers in diachrony
Riad, Tomas	Riad	D2	283	Hypotheses and scenarios in North Germanic tonogenesis

Participant	Authors' abbrev.	Day	No.	Title
Ricquier, Birgit	Ricquier & Demolin	D3	102	The Chronicle of Lingbe, an Extinct Bantu Language of East Congo
Riegger, Chiara	Björnsdóttir et al.	D1	203	The rise of raising in Early Modern English
Ritt, Nikolaus	Ritt & Hofmann	D2	248	'Chained to the rhythm': Using agent-based simulation to model the evolution of stress pattern diversity in English
Ritt, Nikolaus	Ritt & Böhm	D2	W04.5	Sound changes tend to reduce morphotactic ambiguity
Robbeets, Martine	Robbeets	D1	W01	From climate change to language change
Robbeets, Martine	Robbeets	D1	W01.1	From climate change to language change
Rodríguez-Somolinos, Amalia	Rodríguez-Somolinos	D5	48	From inference to hearsay: the development of the French parentheticals <i>à ce qu'il paraît, comme il paraît, il paraît, paraît-il</i>
Roland, Pooth	Orqueda & Pooth	D2	W12.8	Latin <i>ecce</i> : arguments in favor of its development from a PIE demonstrative
Rönchen, Philipp	Rönchen et al.	D4	W03.7	Using simulated data to evaluate models of Indo-European vocabulary evolution
Rosenkvist, Henrik	Rosenkvist	D4	50	Structural ambiguity and reanalysis – the case of Swedish <i>fortsatt</i>
Roth, Kerstin	Roth	D3	79	Rhetoric, stylistic and argumentative strategies of German language female authors in the 17th century
Roth, Theresa	Roth	D1	W05.5	Etymologies and emotions: Historical linguistics as a key to emotion categories
Round, Erich	Round et al.	D2	196	The natural stability of 'unnatural' morphology
Rüdiger, Jan Oliver	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language
Russell, Kerri	Russell	D2	212	Properties of Complex Compounds in Old Japanese
Rutten, Gijsbert	Drinka et al.	D2	W02	Macro-level social motivations for language change: Contact, migration, and globalization
Rutten, Gijsbert	Drinka et al.	D2	W02.1	Macro-level social motivations for language change: Contact, migration, and globalization
Rzyski, Christoph	Rzyski	D2	W14.3	From Old Data to Fresh Phylogenies – A Linguistic Data Journey in the Times of CLDF
Sæbø, Lilja	Sæbø & Grossman	D5	W11.6	A Database of Tonogenetic Events (DTE) and what it can tell us about tonogenesis
Saiz Sánchez, Marta	Saiz Sánchez	D1	88	The periodization of the Pre-Classical French through the study of <i>nennil</i> and <i>non</i> in grammars, remarks and treatises (15th–17th centuries)
Salaberri, Iker	Ariztimuño & Salaberri	D1	152	A new perspective on the evolution of mood and negation markers in Proto-Basque
Salaberri, Iker	Salaberri	D5	30	Towards an account of the emergence, evolution and variability of emphatic negative coordination in Indo-European, part 2: A diachronic perspective
Salmons, Joseph	Salmons	D2	W02.10	Verticalization and the historical sociolinguistics of language maintenance
Salvesen, Christine Meklenborg	Salvesen	D3	109	Tracing the origins of resumption in Swedish
Santamaria, Andrea	Santamaria	D2	263	The Greek suffix <i>-θ-</i> and the Caland System
Sapp, Christopher	Sapp et al.	D5	207	Another look at Noun-Genitive vs. Genitive-Noun in Early New High German



Participant	Authors' abbrev.	Day	No.	Title
Satō, Tomomi	Satō & Bugaeva	D4	W10.1	On stative/active intransitive split within tripartite alignment: A case of Kuril Ainu
Schack, Jørgen	Jensen & Schack	D4	84	Adverbs ending in <i>-(l)ig</i> '-ly' and <i>-(l)igt</i> '-ly' in Danish
Schäfer, Lea	Schäfer	D3	103	Dramatic texts as a source of stigmatization from below
Schlechtweg, Dominik	Schlechtweg	D4	W03.5	The LSCD Benchmark – A testbed for diachronic word meaning tasks
Schulte im Walde, Sabine	Maurer et al.	D4	W03.3	Quantifying Changes in English Noun Compound Productivity and Meaning
Schulte, Michael	Schulte	D2	125	A re-assessment of Early Runic Metrics
Schützler, Ole	Schützler	D5	35	Third-person verb inflection in Shakespeare's dramatic texts
Šefčík, Ondřej	Šefčík	D2	70	Bartholomae's law revisited and remodelled
Serangeli, Matilde	Serangeli	D2	267	Rumpled chicken come home to roost. From [TO CARD – IMPURITY] to [TO PURIFY/HEAL (someone) – from DISEASE]. Evidence from Anatolian, Ancient Greek, and Old Indic
Seržant, Ilja	Seržant	D2	W04.6	Ambiguity avoidance and DOM
Shamseddinov, Abdurahman	Shamseddinov & Authier	D1	31	Contact-driven grammaticalization and drift of new terminal tenses from go-periphrasis in Azeri and Kryz (East Caucasian)
Shcherbakova, Olena	Shcherbakova et al.	D2	96	Diachronic pathways of definite articles distribution
Shimabukuro, Moriyo	Shimabukuro	D4	W10.2	Debuccalization of *p in the Naha dialect of the Ryukyuan language
Sidwell, Paul	Sidwell	D1	W01.7	Austroasiatic dispersal: sea levels and estuarine environments in late Neolithic Mainland SEAsia
Sigurðardóttir, Sigríður Sæunn	Sigurðardóttir	D2	220	From complex to simple prepositions in Icelandic: The case of <i>á bak við</i> to <i>bakvið</i> 'behind'
Sigurðardóttir, Sigríður Sæunn	Eyþórsson & Sigurðardóttir	D2	126	Micro-level conflict in the productivity of anticausativization strategies: Evidence from the history of Icelandic
Sims-Williams, Helen	Sims-Williams	D1	147	Rehabilitating 'non-proportional' analogy
Sitchinava, Dmitri	Sitchinava	D5	286	A panchronic corpus of Old East Slavic and Russian: bringing together Slavic historical and modern corpus resources
Smirnova, Elena	Smirnova	D2	W04.1	The role of ambiguity at different stages of diachronic change
Smith, John Charles	Smith	D5	PL7	Fifty years of ICHL, 1973–2023
Sobolev, Andrey N.	Sobolev	D2	W02.5	Contact as a major Motivation for Linguistic Change in the History of Balkan Slavic
Sobotka, Piotr	Kisiel & Sobotka	D2	51	The functional interpretation of semantic and syntactic shifts in the domain of North Slavic "conversive" preposition-pronominal constructions
Somers, Joren	Cluyse et al.	D1	208	Latin <i>placēre</i> as an alternating Dat-Nom/Nom-Dat verb: A radically new analysis
Somers, Joren	Elens et al.	D1	213	The Alternating Behavior of 'Like' in Old Norse-Icelandic: Facts or Fiction
Sonnenhauser, Barbara	Mendoza et al.	D4	99	Anchoring patterns in emerging complement clauses in Slavic
Sonnenhauser, Barbara	Widmer & Sonnenhauser	D1	W07.4	Linguistic convergence in the Ancient Near East
Souag, Lameen	Souag	D1	W07.3	Prehistoric language contact in Berber

Participant	Authors' abbrev.	Day	No.	Title
Sowada, Lena	Sowada	D2	W02.7	Language use in Alsace from 1914 to 1919. Private texts between official legislation and individual identity construction
Stein, Achim	Cassarà et al.	D4	W13.3	Marked vs. unmarked unaccusativity with alternating verbs: Linking diachronic and experimental data
Sternefeld, Leah	Sternefeld	D5	123	What is <i>ke</i> and if so how many? – The Persian modal particle <i>ke</i> and its diachronic development
Stratton, James	Stratton	D2	115	Where did <i>wer</i> go? Searching for s-curves in lexical change from Old English to Middle English
Strauss, Silvie	Strauss	D2	235	Paradigmatic redundancy in the complement system of Basque
Struik, Tara	Cassarà et al.	D4	W13	New methods for old languages: the comparability of data
Struik, Tara	Cassarà et al.	D4	W13.0	New methods for old languages: the comparability of data
Suleymanov, Murad	Suleymanov	D5	W08.7	Semantic Shift and Morphosyntactic Convergence of Tense-Aspect-Mood Categories in Alazan Persian
Swanenvleugel, Cid	Swanenvleugel	D5	290	The Sardinian substrate lexicon and its Mediterranean comparanda
Tan, Tamisha	Tan	D4	W06.4	'Inalienable' nominalisers across Meto
Tan, Tamisha L.	Tan	D2	133	The Lost Cause: Inflection Class in Amarasi
Teich, Elke	Degaetano-Ortlieb et al.	D4	W03	Computational models of diachronic language change
Tian, Zuoyu	Amaral et al.	D4	W03.6	Model evaluation for diachronic semantics: A view from Portuguese and Spanish
Tieku, Enock Appiah	Tieku	D1	281	Drivers of Diversity in the Construal of Quantity in the World's Languages
Tikhonov, Aleksej	Tikhonov et al.	D2	143	Pronoun history and information structure in 18th century non-religious Russian texts
Tomaschek, Fabian	Nieder & Tomaschek	D5	174	Classifying the origin of Maltese nouns – A cross-language approach employing phonotactics
Torres-Latorre, Aina	Torres-Latorre	D2	176	Synthetic or analytical: factors which explain the formal variation of future and conditional in Old Catalan
Tresoldi, Tiago	Tresoldi et al.	D1	165	A Phylogenetic Study of the Cariban Family: Combining Linguistic and Archaeological Data
Tresoldi, Tiago	Lindgren & Tresoldi	D3	198	The Charition Mime: Decoding the "Indian Language" through Typology and Entropy
Trips, Carola	Trips & Rainsford	D4	W13.2	How to use Yang's Principles to model acquisition in diachrony. The case of psych verbs
Truswell, Robert	Gisborne & Truswell	D4	222	Contact and the origins of headed <i>wh</i> -relatives in Hungarian
Ulman, Vít	Ulman	D2	253	Genesis of the Japanese Compound Particles
van Beek, Lucien	van Beek	D1	W05.1	Clouds or Arrows? Conceptual Metaphors and the Etymology of Homeric Greek <i>kertoméō</i> 'to mock; taunt'
van Beek, Lucien	Kölligan & van Beek	D1	W05	Conceptual metaphors in a comparative and diachronic perspective
van Dam, Kellen Parker	van Dam	D5	41	Internal subgrouping of Northern Naga based on Bayesian phylogenetic analysis
Van de Velde, Freek	Nijs et al.	D2	W02.4	An information-theoretic approach to morphological and syntactic complexity in Dutch, English and German
van Kemenade, Ans	van Kemenade	D4	PL5	Word order change, architecture and interfaces: Evidence from V2 word orders and their loss in the history of English

Participant	Authors' abbrev.	Day	No.	Title
Västerdal, Ida	Västerdal	D2	77	A case of Verschärfung in the Swedish dialect from Stora Rågö in Estonia
Verkerk, Annemarie	Verkerk et al.	D3	242	Exploiting phylogenetic modeling to uncover directionality in the emergence of universals
Verkerk, Annemarie	Krasnoukhova et al.	D4	W09.2	Negated but similar – Negation in the domains of locative, existential, and possessive predication: The case of Indo-European
Vincent, Nigel	Börjars & Vincent	D1	52	Auxiliary, light or lexical: the history of GO verbs
Vincent, Nigel	Vincent et al.	D4	PL6	Linguistic models (with a focus on morphosyntactic change)
Visser, Lourens	Visser	D1	111	Adverbs of degree from Old to Early New High German
Voigtmann, Sophia	Voigtmann	D2	142	Where do all the NPs go? – A corpus linguistic study on NP extraposition in German scientific writing from 1650 to 1900
Vydrin, Valentin	Perekhval'skaya & Vydrin	D5	W11.7	Tonal density and its correlation with the types of tonal systems: Diachronic aspects
Walkden, George	Vincent et al.	D4	PL6	Linguistic models (with a focus on morphosyntactic change)
Walkden, George	Björnsdóttir et al.	D1	203	The rise of raising in Early Modern English
Werner, Martina	Werner	D4	W06.5	When verbal complexes become nouns via infinitive nominalization: A parallel to the verbal domain or category-individual?
Westergaard, Lennart	Westergaard & Boye	D3	90	On semantic change in grammaticalization: Why it is never metaphoric
Westergaard, Lennart	Westergaard	D5	92	The long and winding road of the Danish evidential <i>vel</i> – from epistemic modality via concessivity to evidentiality
Wichers Schreur, Jesse	Wichers Schreur	D2	221	Differential Place Marking and the reconstruction of the Proto-Nakh system of spatial cases
Wichmann, Søren	Jonjić et al.	D2	144	Isoglosses and distributions of features – Analyses of the <i>Dialectological Atlas of the Russian Language</i>
Widmer, Paul	Widmer & Sonnenhauser	D1	W07.4	Linguistic convergence in the Ancient Near East
Wieczorek, Aleksandra	Wieczorek	D5	269	Discontinuous noun phrases containing adjective or adjective-like modifiers in Middle Polish texts. Preliminary research conducted on an experimental dependency treebank
Wiemer, Björn	Mendoza et al.	D4	99	Anchoring patterns in emerging complement clauses in Slavic
Wier, Thomas	Wier	D1	W07.5	Language Contact in the Ancient Caucasus: the View from Kartvelian
Wiklund, Tilo	Rönchen et al.	D4	W03.7	Using simulated data to evaluate models of Indo-European vocabulary evolution
Willis, David	Darling et al.	D2	180	The Diachrony of Person-Number Marking of Subjects in Celtic
Wolfe, Sam	Wolfe	D4	78	Parallel Phases in the History of French
Wolfer, Sascha	Engelberg et al.	D3	PL4	Empirical approaches to the dynamics of the lexicon – internet-based tools and research platforms at the Leibniz-Institute for the German Language
Wolfgruber, Anne	Wolfgruber	D2	W04.4	Text-type specific conventions, subordinate environments and ambiguity (avoidance) in Medieval Spanish passive <i>se</i> -constructions
Yurayong, Chingduang	Yurayong et al.	D1	W07.6	An archaeolinguistic approach to Indianisation and Sincisation of languages in Eastern Eurasia
Zampetta, Silvia	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin
Zanchi, Chiara	Zampetta et al.	D1	W05.6	<i>Calidum hoc est!</i> Metaphors of HOT and COLD in Sanskrit, Ancient Greek, and Latin

<b>Participant</b>	<b>Authors' abbrev.</b>	<b>Day</b>	<b>No.</b>	<b>Title</b>
Zanchi, Chiara	Brigada Villa et al.	D2	W14.2	Universal Dependency for Historical Languages (UD4HL): Towards Standardized Syntactic Data for Historical Languages
Zehentner, Eva	Zehentner & De Cesare	D2	W04	Ambiguity (avoidance) as a factor in language change
Zeng, Xiuwei	Zeng	D5	164	Ditransitive GIVE-construction in three Hainan Min-Chinese: Interaction between inherited structures and contact-induced changes