

BOOK OF ABSTRACTS

POSTERS

58th Annual Meeting of the *Societas Linguistica Europaea*

26 – 29 August 2025

Université Bordeaux Montaigne

Anna Kisiel

Phrasal compounds and the Syntax-Morphology interaction

Adolfo Ausín
(Michigan State University)

Keywords: Phrasal compounds, Syntax-Morphology interaction, Lexicalist Hypothesis, binding, sluicing

A long-standing debate in generative grammar has been whether syntax can access the internal structure of words. Proponents of the Lexicalist Hypothesis (Chomsky 1970) claim that there is a clear distinction between syntax and morphology and that syntax cannot access the internal structure of words. Bruening (2018) summarizes and presents several arguments against the Lexicalist Hypothesis. One of those arguments is based on phrasal compounds such as *God-is-dead theology* (Lieber 1988). In this paper, I revisit the evidence based on phrasal compounds and conclude (contrary to Bruening) that the phrasal compounds present a strong argument in favor of the Lexicalist Hypothesis.

The first element of a phrasal compound is a sentence that behaves like a word. This could be interpreted as an argument against the Lexicalist Hypothesis since it seems to show that it is possible to have a full sentence (a syntactic element) within a word. However, this argument is valid only if it is also shown that the elements outside the phrasal compound can establish syntactic relations with elements within.

The evidence that I will discuss (see (1)-(4)) shows that the elements that are inside a phrasal compound cannot be accessed by syntactic elements that appear outside. In (1)-(4) the a-sentence contains a phrasal compound whereas the b-sentence is a quasi-minimal pair where the main difference is the absence of the phrasal compound.

- (1) a. He_i never paid attention to those Obama_i-is-a-Muslim rumors.
b. *He_i never paid attention to the rumor that Obama_i is a Muslim.
- (2) a. *No woman_i was carrying lock-her_i-up signs.
b. No woman_i wants the president to lock her_i up.
- (3) a. *She always has a someone-will-fix-my-problems attitude, but she doesn't know who.
b. She is convinced that someone will fix her problems, but she doesn't know who.
- (4) a. *This is the woman_i my neighbor was carrying Lock-her_i-up signs.
b. This is the woman_i my neighbor was carrying signs asking the president to lock her_i up.

The grammaticality contrasts in (1)-(4) show that when a constituent appears inside of a phrasal compound (that is, inside of a word), such a constituent cannot enter into syntactic relations (Condition C, pronoun binding, sluicing, or resumptive pronouns) with elements in the main sentence. The lack of a Condition C violation in (1a) can only be explained by assuming that Obama inside the constituent Obama-is-a-Muslim cannot be bound by He in the main sentence. The contrast in (2) shows that the pronoun within the phrasal compound cannot be bound from outside by No woman. The contrast in (3) shows that an indefinite within

a phrasal compound cannot create the necessary conditions for sluicing. The contrast in (4) shows that a pronoun within the phasal compound cannot act as a resumptive pronoun.

The overall conclusion to be drawn from (1)-(4) is that elements with a phrasal compound cannot interact syntactically with elements outside the phrasal compound, supporting some version of the Lexicalist Hypothesis.

References:

- Bruening, Benjamin (2018), The lexicalist hypothesis: Both wrong and superfluous, *Language* 94(1), 1–42.
- Chomsky, Noam (1970), Remarks on Nominalization, in R. A. Jacobs, and P. S. Rosenbaum (eds.), (1970) *Readings in English Transformational Grammar*, Waltham: Ginn, 184–221.
- Lieber, Rochelle (1988), Phrasal compounds in English and the morphology-syntax interface, *Chicago Linguistic Society (Parasession on Agreement in Grammatical Theory)*, 24(2), 202–224.

Investigating correlations between subject pro-drop and indexing

A Corpus-Based Study of Communicative Efficiency

Recent advancements in linguistic typology and corpus-based studies enable the study of linguistic phenomena as continua rather than discrete categories, reducing information loss (Levshina 2021; Wälchli 2009). This study builds on this approach by examining two phenomena increasingly viewed as continuous: person indexing and pro-drop.

To refer to a subject, languages employ numerous strategies such as the use of a pronoun, a person index or both, making it functionally impossible to distinguish between a pronoun and a person index. Thus, many argue that there is a continuum of bondedness of these referring expressions, but that they are, for all intents and purposes, the same (Haspelmath 2013; Croft 2013; Siewierska 2004; Corbett 2006). Similarly, pro-drop, the omission of subject pronouns, ranges from rarely possible, as in English, to common, as in Japanese, with or without indexing on the verb (Ackema & Neeleman 2007; Kramer & Liu 2021). This study aims to find correlations between likelihood of subject prodrop and the presence of person indexing on verbs at the clause level, exploring possible trade-offs and indications of communicative efficiency.

Research indicates a correlation between pro-drop and overt argument indexing, potentially as a form of communicative efficiency (Haig 2018; Berdičevskis *et al.* 2020). However, cross-linguistic findings highlight the variability of informativeness in lexical arguments (Bickel 2003; Stoll & Bickel 2009). Addition-

ally, in SOV languages, the verb-final position can increase processing costs by separating the verb from its arguments (Ferrer-i Cancho 2004; Liu 2008). Ueno & Polinsky (2009) suggest that pro-drop can reduce these costs, with zero anaphora influenced by referent accessibility, in line with Ariel's accessibility theory (Ariel 1990). In addition, recent corpus-based work shows that subject omission positively correlates with the frequency of verb-final order (Levshina 2025).

Data and Methodology The data comprise 1000 sentences from oral corpora in the SUD collection (Gerdes *et al.* 2018), representing 4 languages: Spanish, Slovene, Hausa and Beja. A Python script was used to extract the following information for every clause: use and omission of subject pronoun (response variable), the presence of person indexing, argument structure of the clause, person and TAM (predictors). In addition, we operationalized accessibility of the subject referent as being mentioned in the previous context, using a semiautomatic approach. Correlations between these variables are examined using logistic regression models.

Expected Outcomes By analyzing the correlation between pro-drop rate and person indexing rate, this study tests theories on the positive correlation of argument indexing and pro-drop (Haig 2018; Berdičevskis *et al.* 2020). The findings will contribute to the broader discourse on communicative efficiency and the role of indexing in argument structure.

References

- Ackema, Peter, & Neeleman, Ad. 2007. Restricted pro drop in early modern Dutch. *The Journal of Comparative Germanic Linguistics*, **10**, 81–107.
- Ariel, Mira. 1990. Accessing Noun-Phrase Antecedents. Vol. 96. Londres: Routledge. *Linguistics*, 113–118.
- Berdičevskis, Aleksandrs, Schmidtke-Bode, Karsten, & Seržant, Ilja. 2020. Subjects tend to be coded only once: Corpus-based and grammar-based evidence for an efficiency-driven trade-off. *Pages 79–92 of: Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*.
- Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language*, **79**(4), 708–736.
- Corbett, G.G. 2006. *Agreement*. Agreement. Cambridge University Press.
- Croft, William. 2013. *Agreement as anaphora, anaphora as coreference*. Berlin, Boston: De Gruyter Mouton. Pages 95–118.
- Ferrer-i Cancho, Ramon. 2004. Euclidean distance between syntactically linked words. *Physical review. E, Statistical, nonlinear, and soft matter physics*, **70**(12), 056135.
- Gerdes, Kim, Guillaume, Bruno, Kahane, Sylvain, & Perrier, Guy. 2018 (Nov.). SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In: Lynn, Teresa, & Schuster, Sebastian (eds), *Universal Dependencies Workshop 2018*.
- Haig, Geoffrey. 2018. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics*, **56**(4), 781–818.
- Haspelmath, Martin. 2013. *Argument indexing: A conceptual framework for the syntactic status of bound person forms*. Pages 197–226.
- Kramer, Alex, & Liu, Zoey. 2021. A Data-driven Approach to Crosslinguistic Structural Biases. *Pages 344–346 of: Proceedings of the Society for Computation in Linguistics 2021*.
- Levshina, Natalia. 2021. Corpus-based typology: applications, challenges and some solutions. *Linguistic Typology*, **26**(03).
- Levshina, Natalia. 2025. The Paradox of SOV: A Case for Token-based Typology.”. *PsyArXiv*, 01.
- Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, **9**(2), 159–191.
- Siewierska, Anna. 2004. *Person*. Cambridge Textbooks in Linguistics. Cambridge University Press.
- Stoll, Sabine, & Bickel, Balthasar. 2009. *How Deep Are Differences in Referential Density?* Pages 543–555.
- Ueno, Mieko, & Polinsky, Maria. 2009. Does headedness affect processing? A new look at the VO–OV contrast1. *Journal of Linguistics*, **45**(3), 675–710.
- Wälchli, Bernhard. 2009. Data reduction typology and the bimodal distribution bias. *Linguistic Typology*, **13**(01).

Negative Existential Cycle in Turkish Sign Language (TİD)

Bahtiyar Makaroğlu
(Ankara University)

Keywords: negative existential cycle, Croft cycle, sign language, typology, Turkish Sign Language

The evolution of negation is typically characterized as a cyclical process; whereby new constructions emerge to encode existing functions. The Negative Existential Cycle (NEC) was originally proposed by Croft (1991) and subsequently developed by Veselinova (2013, 2014). An examination of the characteristics of the Croft's Cycle has been conducted in predominantly spoken languages. Consequently, the inquiries posed in relation to the NEC will facilitate a more profound comprehension of the sources and pathways of verbal negation in the domain of signed modality.

Despite the numerous studies on standard negation in Turkish Sign Language (TİD) (see Zeshan 2006; Gökgöz 2009, 2011; Pfau 2016; Makaroğlu 2021), there is a paucity of research on existential negation. The present study undertakes an investigation into the evolution of negative existential negator YOK 'not.exist' in TİD and it seeks to investigate the applicability of the Negative Existential Cycle as proposed by Croft (1991) and examines the potential of it as an explanatory framework in understanding the sources and pathways of standard negation. A hypothesis can be advanced that TİD may be considered as a language situated at Stage B of the NEC. This is due to the fact that it has discrete negative markers that are used for verbal negation and for the negation of existential clauses. However, as evidenced by the corpus data, TİD exhibits negative existential marker YOK which developed into the encoding of standard/verbal negation as illustrated in (1) and (2).

- (1) ₃SENT.LETTER₁ YOK PT1:PRO1
'They didn't send me a letter.' [01.010 S:00:07 E:00:09]
- (2) PT1:PRO1 ONE ₃INFORM₁ YOK
'They didn't give me any news.' [16.005 S:01:32 E:01:34]

The aim of testing the model is to delineate the progression of a negative existential into a standard negator, along with the mechanisms that govern the transition between stages. The present study is based on data from the TİD corpus (for details, see Dikyuva et al. 2017), which comprises a diverse group of 116 native signers from 26 cities across Turkey. The signers, who ranged in age from 12 to 65, were either congenitally deaf or became deaf before the age of three and had early exposure to TİD. This study examined 790 minutes of dyadic video conversations involving 56 native signers of TİD from 14 different cities in Turkey. The dataset comprised 93,151 sign tokens, representing approximately 11.64% of the TİD Corpus.).

In the context of the theory and modeling of language change, it is crucial to emphasize that the NEC is a variationist model. Its stages do not progress sequentially; rather, they tend to co-exist. The synchronic data from the TİD provide substantial support for this assertion, demonstrating that the standard negativity can be encoded in alternative ways. The current study also argues that TİD presents characteristics of stage B~C, wherein the YOK is progressively substituting for the verbal negator in the grammatical system. However, it has not yet been extended to the whole grammatical system and is currently limited to specific contexts. By applying NEC, the negative existential sign YOK would progress along a proposed pathway as outlined below:

negative existential marker > negative perfective marker > standard negator

Acknowledgments: This work is supported by the TÜBİTAK 2224A program.

References

- Croft, William (1991), The evolution of negation, *Journal of Linguistics* 27(1), 1–27.
- Dikyuva, Hasan, Bahtiyar Makaroğlu, and Engin Arık (2017), *Turkish Sign Language grammar*, Ankara: Ministry of Family and Social Policies Press.
- Gökgöz, Kadir (2009). *Topics in Turkish Sign Language (Türk İşaret Dili – TİD) Syntax: Verb movement, negation and clausal architecture*, MA thesis, İstanbul: Boğaziçi University.
- Gökgöz, Kadir (2011), Negation in Turkish Sign Language: The syntax of nonmanual markers, *Sign Language & Linguistics* 14(1), 49–75.
- Makaroğlu, Bahtiyar (2021), A corpus-based typology of negation strategies in Turkish Sign Language, *Dilbilim Araştırmaları Dergisi* 32(2), 111–147.
- Pfau, Roland (2016), A featural approach to sign language negation, in P. Larrivé & C. Lee (eds), (2016), *Negation and Polarity: Experimental Perspectives*, Dordrecht: Springer, 45–74.
- Veselinova, Ljuba (2013), Negative existentials: A cross-linguistic study, *Italian Journal of Linguistics* 25(1), 107–145.
- Veselinova, Ljuba (2014), The negative existential cycle revisited, *Linguistics* 52(6), 1327–1369.
- Zeshan, Ulrike (2006), Negative and interrogative structures in Turkish Sign Language (TİD), in U. Zeshan (ed), *Interrogative and negative constructions in Sign Languages*, Nijmegen: Ishara Press, 128–164.

Exceptive conditionals: A cross-linguistic survey

Bastian Persohn, Tom Bossuyt
(FSU Jena, Ghent University)

Keywords: unless, typology, conditionals, clause combining, negation

In this talk, we examine exceptive conditionals (‘except if, unless’; henceforth: ECs) in the world’s languages. Example (1) is an illustration. Here, Mandarin Chinese *chúfēi* signals that ‘we are busy’ constitutes a possible exception to the validity of the state-of-affairs depicted in the apodosis.

(1) Mandarin Chinese (Li & Thompson 1981: 636)

Chúfēi	wómen	hěn	máng wómen	yíding	lái	kàn	nǐ.
unless	1PL	very	busy 1PL	certainly	come	see	2SG
‘Unless we’re busy, we’ll certainly come to see you.’							

On the face of it, the meaning of ECs appears quite similar to that of canonical conditionals with a negated antecedent. However, there are important ontological and semantic differences between the two (von Fintel 1991, Geis 1973, Declerck & Reed 2000, among others).

Though ECs have received considerable attention in semantics and psycholinguistics (e.g. Declerck & Reed 2000, Espinosa et al. 2015, von Fintel 1991, Traugott 1997), most research on them has focused on individual European languages. On the other hand, “[i]t is not clear to what extent these exceptive C[onditional]C[lauses]s resemble or differ from one another in one language and across languages” (Liu 2019: 4).

In this talk, we present some results of the first-ever cross-linguistic survey of ECs, based on a global convenience sample of 85 languages. After laying out the definition of our subject matter, we examine the distribution of ECs. We show that such constructions are very common in some parts of the world (e.g. large parts of Eurasia), but rare to non-existent in others (e.g. Australia, Papunesia, South America), where they are usually not formally distinguished from regular negative conditionals. We then turn to the marking of exceptive antecedents, where we observe a strong preference for constructions that are (from a strictly synchronic point of view) transparently built around “free exceptives” (von Fintel 1993), such as English *except if*, or Levantine Arabic *ʔalla ʔiza* in (2). Specialized, opaque connectives like English *unless* or Mandarin Chinese *chúfēi*, on the other hand, are less common.

(2) Levantine Arabic (Cowell 1964: 331)

<i>Raḥ-na-ži</i>	<i>ʔalla</i>	<i>ʔiza</i>	<i>nəzl-et</i>	<i>maṭar.</i>
FUT-SUBJ.1PL-come.IPFV	except	if	descend.PFV-SUBJ.3SG.F	rain(F)
‘We’ll come unless (lit. except if) it rains.’				

On the syntactic plane, we show that ECs across languages, phyla and areas display a strong preference for a postposed protasis, the very opposite of what is known from canonical conditionals (Diessel 2019, Greenberg 1978). We argue that this is a function of discursive iconicity and, ultimately, of the ontological differences between the two types of conditionals. Whereas in canonical conditionals the protasis establishes the basis for the following argumentation, in ECs the apodosis lays the foundation, to which an exception can subsequently be addressed in the protasis (cf. Dancygier & Sweetser 2015: 183–187; Montolío 1999).

By surveying the distribution and structure of ECs, as well as aspects of their syntactic behaviour, we provide a first step towards closing the gap in the typological knowledge of these devices and contribute to a better cross-linguistic understanding of non-canonical conditionals in general.

References

- Cowell, Mark W. 1964. *A reference grammar of Syrian Arabic (based on the dialect of Damascus)*. Washington, D.C.: Georgetown University Press.
- Dancygier, Barbara & Eve Sweetser. 2015. *Mental spaces in grammar. Conditional constructions*. Cambridge: Cambridge University Press.
- Declerck, Renaat & Susan Reed. 2000. The semantic and pragmatics of *unless*. *English Language and Linguistics* 4(2). 205–241.
- Diessel, Holger. 2019. Preposed adverbial clauses: Functional adaptation and diachronic inheritance. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 97–122. Berlin: Language Science Press.
- Espino, Orlando, Isana Sánchez-Curbelo & Gustavo M. Rodríguez. 2015. Reasoning with exceptive conditionals: the case of ‘except if’. *Psicológica* 36. 99–122.
- von Fintel, Kai. 1991. Exceptive conditionals: The meaning of *unless*. *Proceedings of NELS* 22. 131–151.
- von Fintel, Kai. 1993. Exceptive constructions. *Natural Language Semantics* 1. 123–148.
- Geis, Michael L. 1973. *If and unless*. In Braj B. Kachru, Robert B. Lees, Yakov Malkiel, Angelina Pietrangeli & Sol Saporta (eds.), *Issues in linguistics: Papers in honor of Henry and Renée Kahane*. 231–253. Chicago: University of Illinois Press.
- Greenberg, Joseph H. 1978. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of human language*. 73–113. Cambridge: MIT Press.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Liu, Mingya. 2019. Current issues in conditionals. *Linguistics Vanguard* 5(3). 1–8.
- Montolío, Estrella. 1999. Las construcciones condicionales. In Ignacio Bosque & Violeta Demonte (Eds.), *Gramática descriptiva de la lengua Española*. vol. 3. *Entre la oración y el discurso. Morfología*. 3643–3737. Madrid: Espasa.
- Traugott, Elizabeth Closs. 1997. *Unless and but* conditionals. In Angeliki Athanasiadou & René Dirven (eds.), *On conditionals again*. 145–167. Amsterdam: Benjamins.

Writing does not impact the evolutionary dynamics of syntax

Carlo Y. Meloni, Chundra A. Cathcart, Jessica K. Ivani, Taras Zakharko, Guanghao You & Balthasar Bickel

(University of Zurich, University of Zurich and NCCR Evolving Language; University of Tübingen, University of Zurich, University of Zurich and LiRI – Linguistic Research Infrastructure; & University of Zurich and NCCR Evolving Language)

Keywords: Phylogenetic modeling, Clause linkage, Writing, Syntax, Language Evolution

Writing and reading have been shown to influence how the brain processes language (Dehaene et al. 2010; Cilibiasi, Adani, and Tsimpli 2019). This raises the question of whether these cognitive effects have left an imprint on the evolution of languages as they transitioned to written forms and as literacy spread across various regions of the world. It has been hypothesized that the advent of writing facilitated expressive complexity, specifically by promoting hierarchical structures (subordination, hypotaxis) and reducing reliance on concatenation (coordination, parataxis) (Small 1924; Mitchell 1985; Dąbrowska 2015). Although the development of writing systems was gradual and their adoption among populations occurred over a prolonged time, it is possible that written expression influenced spoken language, particularly through socially prestigious innovators. Such changes could have initiated linguistic evolution toward more complex syntactic constructions (Karlsson 2009).

To test these hypotheses, we conducted two complementary analyses. The first examined the immediate effect of writing on language use, focusing on sentence-level syntactic patterns in written and spoken contexts. Using Universal Dependencies data (de Marneffe et al. 2021), we analyzed about 100,000 sentences from 30 languages across ten families and three genres: Spoken, Fiction, and Wikipedia. We applied Poisson and negative binomial hierarchical regression models to test whether writing influenced (i) the number of clauses per sentence and (ii) the depth of hierarchical clause embeddings. Our models adjusted for phylogenetic and areal relatedness and for dataset-specific idiosyncrasies.

The second analysis investigated the effect of written traditions on the evolution of grammatical constructions. We compiled a dataset of 763 clause-combining constructions from 59 languages in the Indo-European, Sino-Tibetan, and Tupi-

Guaraní families, coding them for 18 syntactic features reflecting hierarchy in the form of structural asymmetries between clauses (Foley and Jr. 1984; Cristofaro 2003; Bickel 2010; van Gijn, Galucio, and Nogueira 2015). To model grammatical evolution, we employed an Ornstein-Uhlenbeck process (Butler and King 2004) with two regimes ('writing' and 'non-writing') for specific branches and time spans, incorporating random effects for feature type and language to account for cross-linguistic variability.

Our findings reveal no significant influence of writing on syntax in either analysis. For language use, we found no evidence that written genres increased the number of clauses or embedding depth in sentences compared to spoken genres. Similarly, phylogenetic and areal effects showed no measurable impact. For grammatical evolution, the probability of syntactic asymmetry did not differ substantially between the 'writing' and 'non-writing' regimes. However, we also found marginal evidence for reduced variance and increased selective pressure under the 'writing' regime (particularly within Indo-European), tentatively suggesting normativizing effects of writing on grammar evolution.

While the scope of our analysis is limited by its coverage of global linguistic diversity, we tentatively conclude that the advent of writing had little impact on the structural evolution of syntax. However, the subtle normativizing effects observed in some lineages, particularly Indo-European, warrant further investigation into the role of writing traditions in shaping grammatical variation.

References

- Bickel, Balthasar (2010), Capturing Particulars and Universals in Clause Linkage: A Multivariate Analysis, in Bril, Isabelle (eds), (2010), *Clause Linking and Clause Hierarchy: Syntax and Pragmatics*, Amsterdam: John Benjamins, 51–102.
- Butler, Marguerite A. and Aaron A. King (2004), Phylogenetic Comparative Analysis: A Modeling Approach for Adaptive Evolution, *The American Naturalist* 164(6), 683–695.
- Cilibrasi, Luca, Flavia Adani, and Ianthi Tsimpli (2019), Reading as a Predictor of Complex Syntax. The Case of Relative Clauses, *Frontiers in Psychology* 10.
- Cristofaro, Sonia (2003), *Subordination*, Oxford: University Press.

- Dąbrowska, Ewa (2015), Language in the mind and in the community, in Daems, Jocelyne, Eline Zenner, Kris Heylen, Dirk Speelman, and Hubert Cuyckens (eds), (2015), *Change of Paradigms – New Paradoxes: Recontextualizing Language and Linguistics*, Berlin / München / Boston: De Gruyter Mouton, 221–236.
- de Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, and Daniel Zeman (2021), Universal Dependencies, *Computational Linguistics* 47(2), 255–308.
- Dehaene, Stanislas, Felipe Pegado, Lucia W Braga, Paulo Ventura, Gilberto Nunes Filho, Antoinette Jobert, Ghislaine Dehaene-Lambertz, Régine Kolinsky, José Morais, and Laurent Cohen (2010), How Learning to Read Changes the Cortical Networks for Vision and Language, *Science* 330(6009), 1359–1364.
- Foley, William A. and Robert D. Van Valin Jr. (1984), *Functional Syntax and Universal Grammar*, Cambridge: University Press.
- Karlsson, Fred (2009), Origin and maintenance of clausal embedding complexity, in (eds), (2009), *Language Complexity as an Evolving Variable*, Oxford: Oxford University Press.
- Mitchell, Bruce (1985), *Old English Syntax: Concord, The Parts of Speech, and the Sentence*, Oxford: University Press.
- Small, George W. (1924), *The Comparison of Inequality*, Baltimore: University Press.
- van Gijn, Rik, Ana Vilacy Galucio, and Antonia Fernanda Nogueira (2015), Subordination strategies in Tupian languages, *Boletim do Museu Paraense Emílio Goeldi: Ciências Humanas* 10(2), 297–324.

Iconicity and upvoted graphicons. An exploration into the correspondence between upvoting on *Reddit* and use of graphicons.

Elisio Renouf

(University of Caen – Normandy)

Keywords: visual iconicity, graphicons, memes, pragmatics, computer-mediated communication

This paper deals with **visual** iconicity (Klinkenberg) as opposed to **sound** iconicity, i.e. the resemblance or similarity between a visual sign - here graphicons - and its referent. Herring and Dainas (2017, p.1) define graphicons as types of image icons, a set of 'graphical devices found on contemporary social media platforms: emoticons, stickers, GIFs, images, and videos.' Graphicons are shaped by cultural codes and interpretive communities, hence a specific use among Internet groups and forums. Graphicons have been shown to have pragmatic functions, namely 'mention, reaction, tone modification, riffing, action, and narrative sequence.' (Herring & Dainas 2017, p.1). This study investigates the role of graphicons (graphical icons such as emojis, GIFs, and images) in English online interactions on *Reddit* by examining their relationship with image iconicity and post popularity. Specifically, two research questions are tested: (1) whether a link exists between the use of graphicons, expressivity – how speakers express their opinions, beliefs and feelings, and post popularity, and (2) whether specific types of graphicons correlate with higher levels of popularity. To explore the first one, we analyse the 'top' posts across several subreddits, including *r/memes*, which features daily meme submissions characterised by extensive graphicon usage. A comparative analysis is conducted by measuring the number of positive votes ('upvotes') on top posts from *r/memes* and other popular subreddits, such as *r/AskReddit*. Posts are sampled by sorting content using the 'Top' filter for the 'All Time' period.

For the second research question, we examine the relationship between graphicon type and post popularity. The top 30 *Reddit* posts are ranked and indexed to identify those containing at least one graphicon, as categorised by Herring and Dainas (2017). Metrics such as graphicon type, upvote counts, and expressivity are analysed to determine potential correlations with post popularity. Results indicate that a majority of the most popular posts incorporate at least one form of graphicon, such as GIFs or images, which engage users visually while aligning with cultural codes. Within the dataset, the use of graphicons appears to have a direct correlation with upvote frequency.

To extend the analysis, we propose an examination of the number of comments on the most upvoted posts, particularly from *r/memes*, where all posts contain graphicons, compared with the most upvoted posts across *Reddit* as a whole. These findings contribute to understanding the interplay between visual expressivity and user engagement in digital spaces.

References

- Crystal, D. (2001) *Language and the Internet*. Cambridge: Cambridge University Press.
- Herring, S. C. (2004). 'Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior.' In S. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (1st ed., pp. 338–376). Cambridge University Press. <https://doi.org/10.1017/CBO9780511805080.016>

Herring, Susan C., & Ashley Dainas (2017), “‘Nice Picture Comment!’ Graphicons in Facebook Comment Threads’ in *Proceedings of the 50th Hawaii International Conference on System Sciences*. DOI.org (Crossref), <https://doi.org/10.24251/HICSS.2017.264>.

Klinkenberg, Jean-Marie (2012) [2001], ‘Pour une sémiotique cognitive’ in *Linx* 44, Presses universitaires de Paris Nanterre. Put online July 5, 2012. <http://linx.revues.org/1056> ; DOI : 10.4000/linx.1056

Is there a formula for intonation in poetry? Phonetic analysis of intonational features in ten Italian poetry readings

Federico Lo Iacono, Valentina Colonna, Antonio Romano

(University of Turin, Italy and LFSAG – Laboratorio di Fonetica Sperimentale ‘Arturo Genre’, Italy)

Keywords: Intonation, Phonetics, Prosody, Poetry readings, Voices of Italian Poets

In recent years, there has been a growing focus in experimental phonetics on the oral reading of poetry [1, 2, 3]. This is reflected in the development of models for analyzing this specific speech register and in the increasing interest of international research communities, as demonstrated by projects like the *Voices of Italian Poets* (VIP) [4, 5]. Despite advancements in data collection and analysis, many aspects of poetic speech remain under-explored. These include the relationships among metric and rhythmic patterns, and prosodic and melodic elements. Among these, intonation plays a central but underexamined role in defining the distinctive nature of poetic interpretation. The significance of intonation in poetry has long been acknowledged by linguists and literary critics, who highlighted its influence on composition and performance [6].

However, experimental investigations into poetic intonation remain scarce. For example, [7] provided an initial analysis of English poetic intonation, later expanded by [8]. Regarding Italian poetry, [4] and [9] underscored the importance of intonation, identifying patterns such as the *poetic declarative*, characterized by a medium-low, gently descending tonal profile. Operating on the premise that poetry reading is marked by unique rhythmic patterns and distinct approaches to pauses and intonation, this study investigates the intonational characteristics of Italian poetic speech. Ten poetry readings read by the authors themselves (5 men, 5 women), representing diverse generations, were selected from the *VIP Platform* [5]. Despite stylistic diversity, the corpus demonstrates that poetic speech is characterized by diverse intonational forms and functions shaped by its distinctive pragmatics. Indeed, due to its unique features, intonation in poetry cannot be adequately described using linear theoretical models. Instead, it necessitates the application of superpositional models, which are also sensitive to subtle yet distinctive variations. The annotations were carried out using *Praat* software and an updated *VIP protocol* [10]. Moreover, intonation patterns were analyzed using segmentation and alignment methods developed within the AMPER project [11]. Intonation groups were also described qualitatively using the classification framework introduced by [12] and adapted for poetic speech by [4].

This study confirms that poetry readings have distinctive intonational structures. Unlike prose readings, often dominated by descending declarative contours, poetry readings exhibit a higher proportion of less steeply descending poetic declaratives [4], frequently linked to explanatory or emphatic purposes. Continuative patterns tend to reflect greater

suspension, while interrogatives often display flat or non-rising contours, diverging from Italian read speech [13]. Additional features were also identified. Given the intrinsic link between music and poetry, the analysis also highlighted the complex rhetorical structure of the prosodic level of poetry readings. Beyond intonation, the study identified specific strategies for rhythmic features, including longer pauses, shorter inter-pause units, and extended vowel durations compared to other speech registers [14]. Finally, the acoustic parameters of jitter, shimmer, and HNR were also extracted with *Praat*. A preliminary comparison with those observed in prose readings or interviews of the poets themselves reveals significant differences between the speech registers. These differences indicate a tendency toward a more softened, whispered, and breathy voice in poetry readings [15].

References

- [1] Barbosa, P. A. (2023). The dance of pauses in poetry declamation. *Languages*, 8(1), 76. <https://doi.org/10.3390/languages8010076>
- [2] Wagner, P. (2013). Meter specific timing and prominence in German poetry and prose. In O. Niebuhr (Ed.), *Understanding Prosody: The Role of Context, Function and Communication* (pp. 219-236). De Gruyter. <https://doi.org/10.1515/9783110301465.219>
- [3] Meyer-Sickendiek, B., Hussein, H., & Baumann, T. (2017). Rhythmicalizer. In *INFORMATIK 2017: Deep Learning in heterogenen Datenbeständen* (pp. 2189-2200). Gesellschaft für Informatik. https://doi.org/10.18420/in2017_218
- [4] Colonna, V. (2022). *Voices of Italian Poets. Storia e analisi fonetica della lettura della poesia italiana del Novecento* (Vol. 5, pp. 1-272). Edizioni dell'Orso.
- [5] Lo Iacono, F., Colonna, V., & Romano, A. (2024). Preservation, conservation and phonetic study of the voices of Italian poets: A study on the seven years of the VIP archive. In *Proceedings of INTERSPEECH 2024 – Speech and Beyond* (pp. 3664-3668). International Speech Communication Association (ISCA).
- [6] Fabb, N., & Halle, M. (2008). *Meter in poetry: A new theory*. Cambridge University Press.
- [7] Byers, P. P. (1979). A formula for poetic intonation. *Poetics*, 8(4), 367-380.
- [8] Barney, T. (1999). Readers as text processors and performers: A new formula for poetic intonation. *Discourse Processes*, 28(2), 155–167. <https://doi.org/10.1080/01638539909545078>
- [9] Lo Iacono, F. (2024). A whirlwind of neo-avant-gardist voices: Unveiling sound experimentalism in the poetic performances of six members of the Gruppo 63. In *Libro de resúmenes I Congreso Internacional de Teoría de la Lírica y Poéticas Comparadas* (Vol. 1, pp. 89-90). Universidad de Salamanca.
- [10] Colonna, V., & Romano, A. (2023). VIP-RADAR: A model for the phonetic study of poetry reading. In *Proceedings of the 20th ICPHS* (pp. 594-598). Guarant International.
- [11] Contini, M., Lai, J.-P., Romano, A., Rouillet, S., de Castro Moutinho, L., Coimbra, R. L., Pereira Bendiha, U., & Secca Ruivo, S. (2002). Un projet d'atlas multimédia prosodique de l'espace roman. In *Proceedings of the International Conference Speech Prosody 2002* (pp. 227-230). Aix-en-Provence.

- [12] Romano, A. (2014–2018). *Etichette per l'analisi prosodica di file di parlato*. [www.lfsag.unito.it/ricerca/Etichette prosodiche 2014-18.pdf](http://www.lfsag.unito.it/ricerca/Etichette_prosodiche_2014-18.pdf)
- [13] Colonna, V., & Romano, A. (2023). Prose or poetry? A perceptive phonetic study. In *Proceedings of the 20th ICPHS*(pp. 347-351). Guarant International.
- [14] Lo Iacono F., Colonna V., Romano A. (Forthcoming) Stylistic vowel lengthening in Italian poetic speech. A comparative analysis of 14 contemporary Italian poetry readings. *Exploring the relation between duration and length. Italo-Romance, Romance and beyond. Linguistik Aktuell / Linguistics Today*, John Benjamins, Amsterdam.
- [15] Gafni C. & Tsur R. (2017). “Softened” Voice Quality in Poetry Reading: An Acoustic Study. *Style*. 51. 456. 10.5325/style.51.4.0456.

On the Case Licensing of Genitive Subjects in Japanese

Hideki Kishimoto
(Kobe University)

In Japanese, nominative case on the subject of a relative or a noun-complement clause can alternate with genitive case via nominative-genitive case conversion (in the presence of a higher nominal above the clause), as in (1).

- (1) [gakusei-{ga/no} yon-da] hon
 student- NOM/GEN read-PAST book
 ‘the book which the students read’

The two kinds of subjects are often claimed to be located in distinct syntactic positions. One perennial issue is how genitive subjects are Case-licensed. The most prominent hypothesis (advanced by Miyagawa 2011) is that due to the smaller clause size, the genitive subject is Case-licensed by D (located in an upper nominal domain) while residing in vP. Nevertheless, the issue on the position of genitive subjects has not settled yet, since no definitive evidence has been provided in the literature.

In this paper, I suggest that genitive subjects are moved to CP whose head is assigned the formal feature triggering an EPP effect, while nominative subjects are moved to TP owing to the inheritance of the formal feature by T (Chomsky 2008). It is argued then that the genitive subject is Case-licensed by D after its movement to CP, whereas the nominative subject is Case-licensed by T while residing in TP. I will offer new empirical evidence for this analysis, making use of the focus construction in which the focus particle *dake* ‘only’ is placed to the right of the tensed predicate.

In the type of focus construction at issue, *dake* takes TP as its focusing domain, and can be associated with elements located below TP but not elements above TP. In (2), *dake* can be associated with the nominative subject, obtaining the subject-focus interpretation that it is only the child that touched the vase, but this interpretation is missing for the genitive subject.

- (2) [kodomo-{ga/no} sawat-ta]-**dake** no] tubo
 child-NOM/GEN touched-PAST-only COP pot
 ‘the pot which the child touched only’

The interpretative difference shows that the genitive subjects appear in a position higher than the nominative subject. The interpretive fact of (2) further raises the question of where genitive objects of stative predicates, derived from nominative objects via nominative-genitive case conversion, are Case-licensed. It can be shown that genitive objects as well appear in CP with the same diagnostics.

- (3) [[_{TP} kyandii-{ga/no} sukina]-**dake** no] kodomo
 candy-NOM/GEN like-only COP child
 'the child who only likes candies'

(3) can have the object-focus interpretation if marked with nominative case but this interpretation is absent when it is marked with genitive case. Objects are not raised by the formal features for deriving EPP effects in syntax, and given that the focus construction diagnoses the LF position of arguments (Kishimoto 2009), it falls out that the genitive objects are moved into CP at LF and Case-licensed by D.

The new empirical evidence argues against the currently prominent hypothesis on genitive subjects, showing that genitively-marked arguments (derived via nominative-genitive case conversion) are moved into the edge of CP, where an upper D can have access to them for Case-licensing.

References

- Chomsky, Noam (2008), On Phases, in Robert Freidin, Carlos P. Otero and Maria L. Zubizarreta, (eds), (2008), *Foundational Issues in Linguistic Theory: Essay in Honor of Jean-Roger Vergnaud*, Cambridge, MA: MIT Press, 133-166.
- Kishimoto, Hideki (2009), Topic prominency in Japanese, *The Linguistic Review* 26, 465–513.
- Miyagawa, Shigeru (2011), Genitive subjects in Altaic and specification of phase, *Lingua* 121, 1265–1282.

Affix rivalry reconsidered: The role of non-canonical linguistic features in statistical modeling of evaluative morphology

Ivan Lacić
(University of Bologna)

Keywords: affix rivalry, classifier approach, evaluative morphology, Italian, non-canonical linguistic features

Affix rivalry refers to the interaction between two or more affixes that, in at least some instances, can generate words of identical or similar semantic types (Huyghe and Varvara 2023). A particularly apt domain for exploring affix rivalry is evaluative morphology, where constraints of formal economy frequently do not hold (Grandi 2015).

This study investigates rivalry among six Italian evaluative prefixes – *arci*, *extra*, *iper*, *stra*, *super*, and *ultra* – in their usage as adjectival intensifiers (e.g., *strabello* ‘super nice’) in the *itWaC* corpus (Baroni et al. 2009).

Given the polyfunctional nature of evaluative prefixes, derivatives were annotated by five annotators according to two semantic macro-values: intensification and non-intensification. Only intensified derivatives were further analyzed, resulting in a dataset comprising 48,069 tokens distributed across 3,686 types.

To address the constraints potentially relevant in shaping the rivalry landscape, a random sample of 450 occurrences per prefix was annotated for ten (extra)linguistic features. While prior research typically emphasizes semantic differences (Lieber 2023), phonological constraints (Plag 1999) or morphological factors (Lindsay and Aronoff 2013) as discriminating factors between rival affixes, this study underscores the significant role of less explored features. The features considered include: (i) unidirectional base-to-prefix association, (ii) base age, (iii) base frequency, (iv) base semantic class, (v) base emotionality, (vi) derivative syntactic function, (vii) base length, (viii) base stress placement, (ix) base initial sound, and (x) the manner of articulation of the initial consonant of the base, where applicable.

To assess the influence of these features on prefix selection, a Random Forest (RF) classifier (Breiman 2001) was fitted, with feature importance determined using SHAP values (Lundberg 2017). The classifier achieved an accuracy of 0.794, with the highest misclassification rates observed for *iper* (0.319) and *ultra* (0.389). However, given the substantial number of predictors in the most complex model and the potential for certain features to confuse the classifier (Bonami and Pellegrini 2022), further analysis was conducted to pinpoint the optimal feature set. To this end, a series of 39 RF models, each subjected to independent hyperparameter tuning, was constructed.

Assuming that classifier accuracy reflects the predictive power of the explanatory variables, the optimal models, achieving an accuracy of 0.798/0.799, all rely on two-member combinations of three non-canonical features, namely unidirectional base-to-prefix association (operationalized via ΔP values (Ellis and Ferreira-Junior 2009)), base emotionality (derived leveraging Neuraly’s BERT sentiment model), and base frequency. Among these, base-to-prefix association stands out as the most significant predictor of prefix selection, accounting for the distributional niches (Aronoff 2023) the rival prefixes occupy within the morphological system.

In addition to providing localized insights into the variables influencing the selection of one prefix over another, this study challenges the received wisdom by presenting evidence indicating that affix rivalry is governed not only by canonical linguistic factors but also by more nuanced features that have frequently been overlooked in previous research, suggesting a need of more context-sensitive approaches.

References

- Aronoff, Mark (2023), Three ways of looking at morphological rivalry, *Word Structure* 16(1), 49–62.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta (2009), The wacky wide web: A collection of very large linguistically processed web-crawled corpora, *Language resources and evaluation* 43, 209–226.
- Bonami, Olivier and Matteo Pellegrini (2022), Derivation predicting inflection: A quantitative study of the relation between derivational history and inflectional behavior in Latin, *Studies in Language* 46(4), 753–792.
- Breiman, Leo (2001), Random forests, *Machine learning* 45(1), 5–32.
- Ellis, Nick C. and Fernando Ferreira-Junior (2009), Constructions and their acquisition: Islands and the distinctiveness of their occupancy, *Annual Review of Cognitive Linguistics* 7(1), 188–221.
- Grandi, Nicola (2015), The place of evaluation within morphology, in N. Grandi and L. Körtvélyessy (eds), *Edinburgh handbook of evaluative morphology*, Edinburgh: Edinburgh University Press, 74–90.
- Huyghe, Richard and Rossella Varvara (2023), Affix rivalry: Theoretical and methodological challenges, *Word Structure* 16(1), 1–23.
- Lieber, Rochelle (2023), Ghost aspect and double plurality: On the aspectual semantics of eventive conversion and *-ing* nominalizations in English, in S. Kotowski and I. Plag (eds), *The Semantics of Derivational Morphology: Theory, Methods, Evidence*, Berlin: de Gruyter, 15–36.
- Lindsay, Mark and Mark Aronoff (2013), Natural selection in self-organizing morphological systems, in N. Hathout, F. Montermini, and J. Tseng (eds), *Morphology in Toulouse: Selected proceedings of Décembrettes 7*, München: Lincom Europa, 133–153.
- Lundberg, Scott and Su-In Lee (2017), A unified approach to interpreting model predictions, *arXiv preprint arXiv:1705.07874*.
- Plag, Ingo (1999), *Morphological productivity: Structural constraints in English derivation*, Berlin: de Gruyter.

Variation in the realisation of a glottal stop in Leivu South Estonian

Janek Vaab

South Estonian is a Finnic language, or a language group spoken by over 100 000 people, mainly in Estonia. Historically, South Estonian had three exclaves – Leivu, Lutsi, Kraasna – outside Estonia and the main body of the South Estonian area. Leivu was spoken near Alüksne in northeastern Latvia, and it is said to be the earliest offspring of South Estonian (Kallio 2021). According to the typological database of the Uralic languages UraTyp (Norvik *et al.*), South Estonian is the only Finnic language to have a glottal stop /ʔ/ as a phoneme, solely carrying several grammatical and lexical functions, such as marking imperatives and nominative plurals (Norvik *et al.*, UraTyp). Prior research has observed a tendency for the glottal stop to be unrealised in spoken South Estonian, particularly in three-syllable words and in pre-vocalic positions (Iva 2003, 2005). Examples 1 and 2 show realised (1) and unrealised (2) glottal stops in nominative plural.

- (1) *bat `rüü-ʔ `kiezvü-ʔ `kasv-i-va `äste* (Mets *et al.* 2014: 68)
but rye-PL barley-PL grow-IPF-3PL well
'But the rye and the barley grew well.'
- (2) *`sa-ije rüü rauttu `ussõ* (Mets *et al.* 2014: 61)
get.IPF.3PL rye.PL cut-PPP out
'Got the rye cut off.'

The aim of this study is to find the conditions and frequency of pronunciation of a glottal stop in Leivu. There are no more Leivu speakers left; thus, no more material could be recorded. In addition to the 2 hours and 40 minutes of previously transcribed recordings (Mets *et al.* 2014), 21 more were transcribed and annotated, creating a corpus of 4,591 potential glottal stop occurrences based on descriptions of Võro South Estonian (Nigol 1994; Keem & Käsi 2002; Iva 2003, 2005).

The glottal stop was realised in only 21,5% of the cases. To explain the variation between [ʔ] and Ø in Leivu South Estonian, a dataset was coded for morphophonological variables: the syllable count (Iva 2003, 2005), word quantity (*ibid.*), following segment (*ibid.*), morphological status (e.g., /ʔ/ as an independent suffix, as in *küriidä-ʔ* 'write-IMP', or part of a suffix, as in *kiele-geʔ* 'language-COM'), and the speaker. By applying conditional inference

trees, random forests, and regression modelling, it was determined that while realising /ʔ/ as Ø is the default pattern in Leivu South Estonian, the realisation of [ʔ] is most likely when it functions as an independent suffix and is followed by a pause. In contrast, word quantity (e.g., Q2 [ier:neʔ] ‘pea’ vs. Q3 [ier::neʔ] ‘peas’) was not a significant predictor, distinguishing Leivu from Lutsi, another South Estonian variety in Latvia.

These findings help us understand the historical morphophonological changes that contributed to the complete loss of the glottal stop in Mulgi, the westernmost South Estonian variety, while also highlighting the unique features of Leivu as a language island and the earliest offspring of South Estonian.

References

Iva, Sulev 2003. Lõunaeesti häälikud II. TÜ eesti keele õppetooli toimetised. Tartu: Tartu Ülikool. 68–91.

Iva, Sulev 2005. Glottal stop in Võro South Estonian. *Linguistica Uralica*, 2, 123–133.

Kallio, Petri 2021. The Position of Leivu. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 12(2), 123–143. <https://doi.org/10.12697/jeful.2021.12.2.05>

Keem, Hella, Inge Käsi 2002. *Eesti murded VI*. Võru murde tekstid. Aldi Sepp (eds.). Tallinn: Eesti Keele Instituut.

Mets, Mari, Anu Haak, Triin Iva, Grethe Juhkason, Mervi Kalmus, Miina Norvik, Karl Pajusalu, Pire Teras, Tuuli Tuisk, Lembit Vaba 2014. *Eesti murded IX. Lõunaeesti keelesaarte tekstid*. Ellen Niit (eds.). Tallinn: Eesti Keele Instituut, Tartu Ülikool.

Nigol, Salme 1994. Hargla murraku konsontantism. Tallinn: Eesti Keele Instituut.

Norvik, Miina, Yingqi Jing, Michael Dunn, Robert Forkel, Terhi Honkola, Gerson Klumpp, Richard Kowalik, Helle Metslang, Karl Pajusalu, Minerva Piha, Eva Saar,

Sirkka Saarinen and Outi Vesakoski: Uralic typology in the light of new comprehensive data sets (submitted ms to Journal of Uralic Linguistics). DOI 10.5281/zenodo.5236365.

INTEREST in Old and Middle English texts: Historizing the curious emotion

Javier E. Díaz-Vera
(University of Castilla-La Mancha)

Keywords: Interest, Conceptual Metaphor, Old and Middle English, Emotion Conceptualization, Figurative Language.

Interest, as a curious and engaging emotion, arises when a stimulus is both novel or complex and simultaneously comprehensible, thus triggering curiosity, learning, and active engagement (Silvia 2008). From a cognitive linguistic perspective, interest emerges when the balance between challenge and comprehensibility captures attention and fosters action. This study investigates how the emotion of interest was conceptualized in Old English (OE) and Middle English (ME) texts, focusing on three research questions:

1. What are the dominant metaphorical and metonymic expressions of interest in OE and ME?
2. How does the conceptualization of interest shift from a largely negative framing in OE to a more positive one in ME?
3. What role do figurative expressions play in this diachronic evolution?

Drawing on Conceptual Metaphor Theory (Lakoff and Johnson 1980; Geeraerts and Gevaert 2008), the analysis uses a corpus of representative OE and ME texts, including homilies, medical treatises, religious prose, and early philosophical and scientific writings. These texts provide insight into both everyday and intellectual uses of emotion terms.

The study identifies dominant figurative models in OE that construe interest through negative metaphors and metonymies such as LUST, ANXIETY, and LOSS OF CONTROL. For instance, metaphors of FIRE within the body represent interest as an uncontrollable and dangerous desire, suggesting a communal wariness of excessive emotional or cognitive involvement (Díaz-Vera 2025). This resonates with a medieval cultural climate that prioritized spiritual detachment and collective stability (Harrison 2001). In contrast, ME texts show a marked shift toward positive metaphorical framings of interest. By the 13th century, as scholastic and philosophical thinking flourished, interest becomes associated with metaphors such as NOURISHMENT and LIGHT—for example, *foode of the mynde* ('food of the mind')—which present curiosity as a vital resource for intellectual and personal development. These expressions reflect a growing cultural appreciation for individual inquiry, suggesting that emotional engagement and curiosity were being reframed as assets rather than liabilities. This evolution from danger to nourishment, from inner fire to enlightening light, illustrates a cognitive and cultural transformation in the understanding of intellectual emotion. Figurative language not only reflects but actively shapes this shift, highlighting how emotional conceptualizations evolve with changing socio-intellectual norms.

Ultimately, this study reveals how figurative expressions of interest in OE and ME texts offer a window into broader medieval attitudes toward learning, selfhood, and emotional regulation. By

mapping these diachronic changes through metaphor and metonymy, the paper contributes to historical cognitive linguistics and the study of emotion language across time.

Acknowledgments: Spanish Ministry of Science and Innovation (PID2022-138508NB-I00)

References

- Díaz-Vera, Javier E. (2025), *Positive Emotions in Old English Language and Thought: An Emotion Family Approach*. Berlin: Mouton.
- Geeraerts, Dirk and Caroline Gevaert (2008), Hearts and (angry) minds in Old English. In Farzad Sharifian, Rene Dirven and Ning Yu (eds), (2008), *Culture, Body, and Language: Conceptualizations of Internal Body Organs across Cultures and Languages*, Berlin: Mouton, 319–347.
- Harrison, Peter (2001), Curiosity, forbidden knowledge, and the reformation of natural philosophy in Early Modern England. *Isis* 92(2), 265–90.
- Lakoff, George and Mark Johnson (1980), *Metaphors We Live By*. Chicago: University of Chicago Press.
- Mortillaro, Marcello, Marc Mehu, and Klaus Scherer (2011), Subtly different positive emotions can be distinguished by their facial expressions. *Social Psychological and Personality Science* 2(3), 262–271.
- Silvia, Paul J. (2008), Interest: The curious emotion. *Current Directions in Psychological Science* 17. 57–60.

Living on the edge: The Balkan vs. non-Balkan nature of Torlak through the lens of parametric hierarchies

Jelena Živojinović

(University of Graz)

Keywords: Torlak, clitic doubling, parameter hierarchies, Balkan Sprachbund

Torlak, also referred to as Prizren-Timok or Torlakian, is a South Slavic variety spoken primarily in southeastern Serbia, with extensions into neighboring regions of Bulgaria and North Macedonia. Despite its linguistic distinctiveness, the classification of Torlak remains a matter of ongoing debate. Some scholars classify it as a Serbian (or Serbo-Croatian) dialect (cf. Rešetar 1907, Matasović 2011, a.o.), while others situate it as part of the Bulgarian linguistic continuum (cf. Stojkov & Bernštejn 1975). Its position within the Balkan Sprachbund is equally ambiguous, as it is often considered to occupy a midway position between Balkan and non-Balkan linguistic areas without a precise determination of its areal characteristics.

The genealogical uncertainty and areal positioning of Torlak present significant challenges for linguistic classification. While it is rooted in the Old Shtokavian tradition, as evidenced by its retention of pre-Neo-Shtokavian accentual features (cf. Vuković *et al.* 2022), it also exhibits features typical of Balkan Sprachbund, such as the postposed definite article, clitic doubling, and the productive use of synthetic past tenses. However, the extent to which these features are endogenous or contact-induced remains unclear. For example, while clitic doubling has been argued to be a Balkan Sprachbund feature developed through contact (cf. Živojinović 2021), its diachronic origins and mechanisms of emergence require further exploration.

To address these questions, I propose a formalized approach to analyzing Torlak's areality using the framework of Parametric Hierarchies (Roberts & Holmberg 2010; Roberts 2012, 2019). This model organizes the parameters of Universal Grammar into hierarchies, where macroparametric effects emerge as combinations of microparameters. Within this framework, Ledgeway *et al.* (in press) introduce the concept of the "catapult effect," whereby contact-induced changes result in abrupt movements across the parametric spectrum. Regular, endogenous changes typically proceed gradually, with incremental shifts between macro, meso-, and microparameters. By contrast, the catapult effect describes scenarios in which a variety undergoes rapid reconfiguration due to intense language contact, aligning its parametric values with those of the contact variety.

In this study, I apply the framework of Parametric Hierarchies to Torlak data to identify Balkan and non-Balkan features. I focus specifically on the phenomenon of clitic doubling, which has been described as a Balkan Sprachbund feature (cf. Živojinović 2021)

and explore its development in Torlak. Most existing studies suggest that Old Church Slavonic and Old Slavic lacked true clitic doubling in the sense described by Krapova & Cinque (2008), though some instances of clitic left dislocation are attested in Old Bulgarian (cf. Dimitrova-Vulchanova & Vulchanov 2008). This supports the hypothesis that true clitic doubling is not a diachronic inheritance, but rather a contact-induced phenomenon.

To explore this hypothesis, I focus on the recent diachrony of Torlak, assuming its closest genealogical relation to be with Old and Neo-Shtokavian varieties. BCMS (Bosnian, Croatian, Montenegrin, and Serbian)

does not exhibit clitic doubling, providing a baseline for comparison. According to Runić's (2014) analysis, certain Torlak subvarieties lack sensitivity to specificity, representing an earlier stage in the parametric hierarchy. In contrast, southern Torlak varieties, along with Macedonian, exhibit sensitivity to specificity/definiteness but not to topicality, diverging from Bulgarian, which is sensitive to both specificity and topicality. This pattern suggests that Torlak has been catapulted from its assumed genealogical baseline in BCMS and has converged with Macedonian as shown in Fig.1.

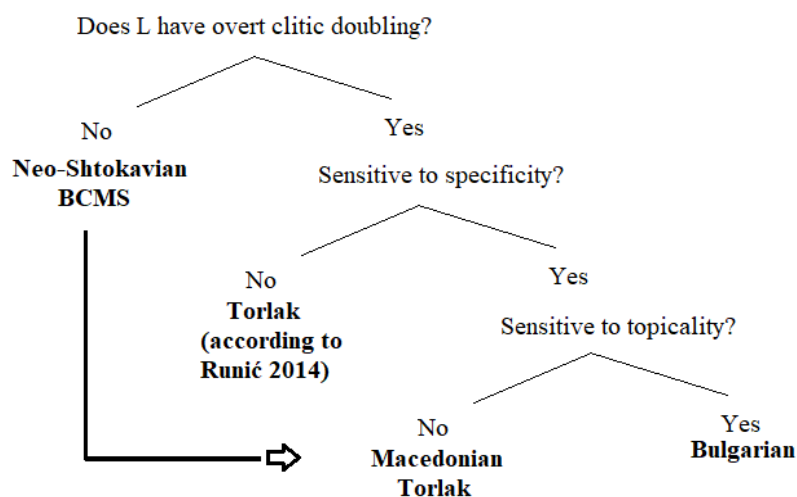


Figure 1: Parametric hierarchy for clitic doubling

Aknowledgments: This research has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement n. 101153454.

References

- Dimitrova-Vulchanova, Mila and Vulchanov, Valentin (2008), Clitic doubling and Old Bulgarian, in D. Kallulli and L. Tasmowski (eds), *Clitic Doubling in the Balkan Languages*, Amsterdam / Philadelphia: John Benjamins, 257–287.
- Krapova, Iliyana and Cinque, Guglielmo (2008), Clitic reduplication constructions in Bulgarian, in D. Kallulli and L. Tasmowski (eds), *Clitic Doubling in the Balkan Languages*, Amsterdam / Philadelphia: John Benjamins, 257–287.
- Ledgeway, Adam, Schifano, Norma and Silvestri, Giuseppina (in press), *Language Contact and Variation in Magna Graecia: Convergence, Layering and Hybridism*.
- Matasović, Ranko (2011), Što je srednjojužnoslavenski?, *Jezik* LVIII(3), 81–120.
- Rešetar, Milan (1907), *Der Štokavische Dialekt*, Wien: Alfred Hölder.
- Roberts, Ian (2012), Macroparameters and Minimalism. A programme for comparative research, in C. Galves, S. Cyrino, R. Lopez and J. Avelar (eds), *Parameter Theory and Linguistic Change*, Oxford: Oxford University Press, 320–354.
- Roberts, Ian (2019), *Parameter Hierarchies and Universal Grammar*, Oxford: Oxford University Press.

- Roberts, Ian and Holmberg, Anders (2010), Introduction: Parameters in minimalist theory, in T. Biberauer, A. Holmberg, I. Roberts and M. Sheehan (eds), *Parametric Variation: Null Subjects in Minimalist Theory*, Cambridge: Cambridge University Press, 1–57.
- Runić, Jelena (2014), *A New Look at Clitics, Clitic Doubling, and Argument Ellipsis: Evidence from Slavic*, PhD dissertation, University of Connecticut. <https://opencommons.uconn.edu/dissertations/457>
- Stojkov, Stojan and Bernštejn, S. B. (1975), *Bŭlgarski dialekten atlas: Juzozapadna Bŭlgarija*, Sofia: Bŭlgarska akademija na naukite.
- Vuković, Teodora, Escher, Anastasia and Sonnenhauser, Barbara (2022), Degrees of non-standardness. Feature-based analysis of variation in a Torlak dialect corpus, *International Journal of Corpus Linguistics* 27(2), 220–247.
- Živojinović, Jelena (2021), Torlak clitic doubling. A cross-linguistic comparison, in A. Blümel, J. Gajić, U. Junghanns and H. Pitsch (eds), *Advances in Formal Slavic Linguistics 2018*, Berlin: Language Science Press, 423–441.

Singular and plural possessive pronouns in relation to Danish collective nouns

Jørgen Schack & Eva Skafte Jensen (The Danish Language Council)

Keywords: collective nouns, possessive pronouns, agreement, member-level accessibility, animacy

The topic of our presentation is the use of singular and plural possessive pronouns in relation to Danish collective singular nouns, i.e. nouns such as *befolkning* ('population'), *familie* ('family'), *hold* ('team'), *klub* ('club'), *par* ('couple'), *parti* ('party'), *regering* ('government'). Collective nouns have a two-sided semantic structure, as one can perceive the referent of a collective noun both as 'a unit' and as 'a plurality of individuals' (Corbett 2022, 194). Due to the unique semantics of collective nouns, they can control elements, such as possessive pronouns, that imply they are either semantically singular or semantically plural.

We present the results of two empirical studies on the use of the possessive pronouns *sin/sit/sine*, *dens/dets* (all meaning 'its'), and *deres* ('their') in connection with a range of different collective nouns. We have examined language use in a contemporary newspaper corpus and in a corpus consisting of novels from the period 1870–1899. When we began our study of collective nouns, we assumed that the choice between singular and plural possessive pronouns was primarily a contextually determined choice. However, the results of our empirical studies indicate that the choice between singular and plural forms is not only contextually determined but also largely lexically determined. This is evident in the fact that some collective nouns, such as *parti* and *regering*, predominantly control singular forms, while others, such as *familie* and *par*, predominantly control plural forms. This applies to both late 19th-century Danish and contemporary Danish. The collective nouns that the two corpora have in common, e.g. *par*, *familie*, *regering*, occupy the same relative position on a "singular-plural scale", but the use of plural forms in such nouns has increased from the 19th century to the present day.

Based on a cognitive concept of "low" versus "high member-level accessibility", Joostens et al. 2007 classify Dutch collective nouns into three main groups in the gradient of member level accessibility: "from predominantly plural possessive and personal pronouns (e.g., *echtpaar* 'married couple') over singular possessive, but plural personal pronouns (e.g., *team* 'team') to singular possessive and personal pronouns (e.g., *regering* 'government')" (Joostens et al. 2007, 99). As for the use of possessive pronouns, we found corresponding results in our studies of equivalent Danish collective nouns. A cognitively based classification of collective nouns like the aforementioned goes hand in hand with a classification grounded in the concept of animacy (e.g. Corbett 2000, 56). Nouns that denote 'organisations' (i.e. entities with a formal organisational structure, such as *klub* and *parti*), are not animate *per se*, as they can be viewed both as a plurality of individuals and as an abstract entity that is not necessarily associated with particular individuals (Yamamoto 1999, 18-19). Collective nouns such as *par* and *familie*, on the other hand, profile the individuals that are part of such units in an immediate manner and are therefore animate *per se*.

References

- Corbett, Greville G. (2000), *Number*, Cambridge: Cambridge University Press.
- Corbett, Greville G. (2022), The Agreement Hierarchy revisited: The typology of controllers, *Word Structure* 15(3), 181–225.

Joosten, Frank, Gert De Sutter, Denis Drieghe, Stef Grondelaers, Robert J. Hartsuiker & Dirk Speelman (2007), Dutch collective nouns and conceptual profiling, *Linguistics* 1, 85-132.

Yamamoto, Mutsumi (1999), *Animacy and reference: a cognitive approach to corpus linguistics*, Stanford: Stanford University Press.

How do inflectional classes diverge? A case study in the Finnic languages

Jules Bouton & Sacha Beniamine

(Université Paris-Cité, LLF, CNRS & Surrey Morphology Group, University of Surrey)

The existence and persistence of inflectional classes is a recurring issue in studies of morphological complexity (Stump and Finkel, 2013; Maiden, 2018; Parker et al., 2022). A standing question is the extent to which inflection classes of closely related languages can provide phylogenetic signal. In this study, we hypothesize that although individual lexemes often change classes, the macrostructure of a class system tends to remain stable over time. We test this hypothesis by quantitatively comparing inflection class hierarchies in four closely related Finnic languages.

Inflection classes are groups of lexemes which inflect similarly within a morphological system. They are notably difficult to identify (see among others Corbett, 1982; Blevins, 2004) within a single system, as they depend on complex analytical choices (Carstairs, 1987). Across languages, this problem is compounded by the low comparability of language descriptions, which are often idiosyncratic and depend on varied philological traditions. The absence of large diachronic inflected lexicons makes it hard to study the evolution of inflectional classes. A possible workaround consists in using agent-based models (ABM) to simulate the emergence of inflectional systems (Parker et al., 2022) on synthetic data. Such results need to be confronted with quantitative evidence on real languages. Recent developments in quantitative morphology provide both tools and resources to approach this issue.

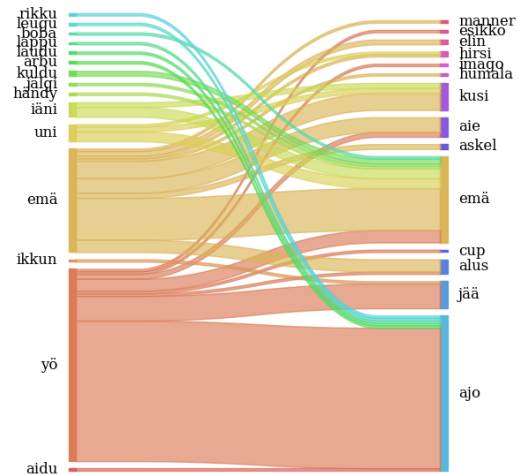
Recent work in computational morphology had made it possible to distribute lexemes in inflectional classes in a reproducible manner (Beniamine, 2018). This approach proceeds from large inflected lexicons, and calculates analogical patterns between pairs of forms. Lexemes that share identical sets of patterns are clustered in microclasses. Similarities across microclasses are represented as a hierarchy of inflectional classes (Beniamine, 2021). The toolkit *Qumin* (Beniamine and Bouton, 2025) makes these computations straightforward and easily reproducible for large inflected lexicons. Furthermore, the publication of four inflected lexicons of closely related Finnic languages (Estonian; Beniamine, Aigro, et al., 2024; Finnish, Livonian, Olonets; Bouton, 2024) offers an opportunity to study the distribution of inherited lexemes in inflectional classes.

We set out to compare inflectional classes across these four Finnic varieties. First, we annotate the inflected lexicons with cognacy information, extracted from the etymological data of the Wiktionary (Ylonen, 2022) and from the etymological project Sanat (Junttila et al., 2020). We derive a minimal lexicon of frequent inherited lexemes in the Finnic languages, linked to existing datasets and available on Zenodo¹. We then calculate alternation patterns, microclasses and inflectional lattices using *Qumin* on the four languages. We keep only inherited cognates that appear in the etymological data. We obtain formal hierarchies over the same cognate sets, which we compare systematically using distance and similarity metrics. Figure shows a visual comparison between Finnish and Olonets microclasses, in which the main classes remain large and constitute attractors, despite lexeme-level variations.

Our study confirms in all languages the stability of closed and highly frequent inflectional classes. We show that the hierarchical structure of inflection classes captures more similarities than the raw distribution in microclasses. This observation strengthens the claim that a dynamic representation of inflectional classes provides stronger evidence of inherited structures.

¹10.5281/zenodo.15133174

Figure 1: Distribution of lexemes in microclasses for Finnish (left) and Olonets (right). The flows on the chart show which classes the cognates belong. The labels are examples from each class.



Bibliography

- Beniamine, Sacha (2018). "Classifications flexionnelles. Étude quantitative des structures de paradigmes". PhD thesis. Université Sorbonne Paris Cité - Université Paris Diderot.
- (2021). "One lexeme, many classes: Inflection class systems as lattices". In: *One-to-many-relations in morphology, syntax, and semantics*. Ed. by Berthold Crysmann and Manfred Sailer, pp. 23–51.
- Beniamine, Sacha, Mari Aigro, et al. (2024). "Eesthetic: A Paralex Lexicon of Estonian Paradigms". In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*. Ed. by Nicoletta Calzolari et al. ELRA and ICCL, pp. 5526–5537.
- Beniamine, Sacha and Jules Bouton (2025). *Qumin 3.0.10 [Python package]*. Zenodo. DOI: 10.5281/zenodo.15008373.
- Blevins, Jim (2004). "Inflection classes and economy". In: *Explorations in Nominal Inflection*. Ed. by L. Gunkel, G. Müller, and G. Zifonun. De Gruyter, pp. 41–85. DOI: 10.1515/9783110197501.51.
- Bouton, Jules (2024). "Towards standardized inflected lexicons for the Finnic languages". In: *Proc. of the 9th International Workshop on Computational Linguistics for Uralic Languages*. ACL, pp. 59–66.
- Carstairs, A. D. (1987). *Allomorphy in Inflection*. Croom Helm linguistics series. Croom Helm.
- Corbett, Greville G. (1982). "Gender in Russian: An account of gender specification and its relationship to declension". In: *Russian linguistics* 6 (2), pp. 197–232.
- Junttila, Santeri, Sampsa Holopainen, and Juho Pystynen (2020). "Digital Etymological Dictionary of the Oldest Vocabulary of Finnish". In: *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 46.2, pp. 733–747.
- Maiden, Martin (2018). *The Romance verb: Morphomic structure and diachrony*. Oxford University Press.
- Parker, Jeff, Robert Reynolds, and Andrea D. Sims (2022). "Network Structure and Inflection Class Predictability: Modeling the Emergence of Marginal Detraction". In: *Morphological Diversity and Linguistic Cognition*, pp. 247–281.
- Stump, Gregory T. and Raphael Finkel (2013). *Morphological Typology: From Word to Paradigm*. Cambridge University Press.
- Ylonen, Tatu (2022). "Wiktextextract: Wiktionary as Machine-Readable Structured Data". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. ELRA, pp. 1317–1325.

The limits of Degree Modification in Basque spatial PPs: Variation across Projective and Region adpositions

Jürgen Etxeberria Gonzalez
(University of the Basque Country, UPV/EHU)

Keywords: adpositions, spatial PPs, degree modification, measure phrases, Basque

Spatial adpositions exhibit degree modification in various ways, but not all adpositions are subject to modification. Previous studies (Svenonius 2010, Franco 2016, Ursini & Wu 2024) observed that adpositions with certain semantic properties, such as Region adpositions, block measure phrases, whereas Projective adpositions permit them. In this study, I argue that, in Basque, gradability (Kennedy & McNally 2005) and licensing an overt ground are delimiting factors in these constructions. Consequently, neither Region adpositions nor bare NPs are candidates for degree modification.

The Projective/Region alternation traces back to Talmy's typology (2000). Spatial adpositions are lexical items (*in, at, under*) that form PPs, establishing a spatial relation between the *Figure*, a moving entity, and the *Ground*, which denotes a referential point in space (Talmy 2000), as in (1):

- (1) Katu-a ispilu(-aren) aurre-an dago.
cat-DET mirror-GEN front-INE be.3SG
Figure Ground Region
'The cat is in front of the mirror.'

This sentence pairs with *Region* adpositions (Svenonius 2006), which license an overt ground and denote undirected points within the ground. In contrast, *Projective* adpositions denote directed points, allowing measure phrases. In Italian (Ursini & Wu 2024), it has been observed that [+ degree] features are only present in Projective prepositions (*Mario va 10 metri dietro alla macchina* 'Mario goes 10 meters behind the car'). Region prepositions lack these features, blocking measure phrases (**Mario va 10 metri in cima alla montagna* 'Mario goes 10 meters on top of the mountain'). Basque data in (2) corroborates this:

- (2) a. Ander (*10 metro) etxe(-aren) **aurre-ra** etorri da.
Ander 10 meters house-GEN front-ALL come be.3SG
'Ander came 10 meters in front of the house.'
- b. Ander etxe-tik 10 metro **aurre-ra** doa.
Ander house-ABL 10 meters front-ALL go.3SG
'Ander goes 10 meters forward from the house.'

Two uses of *aurre* 'front' appear in (2). First, *etxe aurre* 'front of the house' (2a) denotes a Region of the ground, as a non-gradable construction, the measure phrase is blocked. In contrast, a

Projective adposition, *aurre-ra* ‘forward’ (2b), which does not denote a Region, is gradable and allows measure phrases.

Minimal pairs also illustrate this alternation. Adpositions as *goi* ‘up’ cannot combine with overt grounds, instead, a relational noun, *gain* ‘upperside’, is required (Etxepare 2013). In (4b), I present that the latter one blocks measure phrases:

- | | |
|--|--|
| <p>(3) a. *Mendi(-aren) goi-an.
 mountain-GEN up-INE
 ‘Somewhere up the mountain.’</p> | <p>b. Mendi-an 100 metro go-ra joan.
 mountain-ALL 100 meters up-ALL go
 ‘Go 100 meters up the mountain.’</p> |
| <p>(4) a. Mendi(-aren) gain-(e)an.
 mountain-GEN upside-INE
 ‘At the top of the mountain.’</p> | <p>b. *Mendi-an 100 metro gain-era joan.
 mountain-ALL 100 meters upside-ALL go
 ‘Go 100 meters in top of the mountain.’</p> |

Beyond measure phrases, aforementioned limitations extend to comparative adpositions, which are productive with gradable Projective adpositions (*atze-ra-go*/**atze-an-ago* ‘further back’). Since most NPs are non-gradable (Vela-Plo 2020) they cannot form comparative adpositions (**aulki-ra-go* ‘further to the chair’).

In sum, Basque measure phrases are limited to Projective adpositions, as they are gradable and scalar, while Region adpositions block them due to their non-gradable nature. Taken together, these findings emphasize the importance of gradability and overt ground licensing in measure phrase compatibility.

Acknowledgments

I am grateful to the three anonymous abstract reviewers for their insightful comments. This research has been supported by Basque Government (Grants for the Predoctoral Training Program PRE-2023-3-0177) and by the Consolidated Research Group Project (PID2023-148030NB-I00).

References

- Etxepare, Ricardo (2013), Basque Primary Adpositions from a Clausal Perspective, *Catalan Journal of Linguistics* 12, 1–42.
- Franco, Ludovico (2016), Axial parts, phi-features and degrammaticalization: the case of Italian *presso*/*pressi* in diachrony, *Transactions of the Philological Society* 114(2), 149–170.
- Kennedy, Christopher and Louise McNally (2005), Scale Structure, Degree Modification and the Semantics of Gradable Predicates, *Language* 81, 345–381.
- Svenonius, Peter A. (2006), The emergence of axial parts, in P. Svenonius and M. Pantcheva (eds.), *Nordlyd, Tromsø Working Papers in Language & Linguistics: 33.1, Special Issue on Adpositions*, Tromsø: University of Tromsø, 49–77.
- (2010), Spatial P in English, in G. Cinque and L. Rizzi (eds.), *The cartography of syntactic structures*, vol 6, Oxford: Oxford University Press, 127–160.

- Talmy, Leonard (2000), *Towards a cognitive semantics*, Cambridge, MA: The MIT Press.
- Ursini, Francesco-Alessio and Wu, Tong (2024), On Italian spatial prepositions and measure phrases: reconciling the data with theoretical accounts, *Linguistics* 62(2), 491–530.
- Vela-Plo, Laura (2020), *Drawing comparisons: A syntactic and semantic approach to Basque, Spanish and English comparative structures*, PhD dissertation, UPV/EHU.

User-oriented research data management infrastructure development for collaborative research

Katja Politt

(Bielefeld University, CRC 1646)

Keywords: data management, open science, metadata, consent forms, research data

In large collaborative research centers (CRCs), managing a broad variety of linguistic data – such as texts, videos, audio recordings, questionnaires, or psycholinguistic measurements – presents several challenges, including handling personal and sensitive information (e.g., gender, personality traits, or place of birth) and ensuring proper anonymization of audio and video data. These issues, together with legal and ethical requirements, complicate data sharing and reuse, which are essential components of adherence to FAIR (findable, accessible, interoperable, reusable) data principles. Different legal and ethical requirements must be adhered to when it comes to study planning, data management plans, and good research practices, even before the data have been collected (Jorschick et al. 2024). Structured research data management (RDM) can improve project documentation and collaboration efficiency (Mittal et al. 2023), but RDM strategies need to be tailored to the needs of the individual project in order to allow for effective collaboration and reduced time spent on data management (Kanza & Knight 2022, Pasquetto et al. 2017).

The poster presents the development of a user-oriented solution to these problems developed within a large CRC in Germany. First, a "wizard" designed to help researchers during the planning phase by automatically generating customized, legally compliant consent forms for participants. Second, a data management platform where collected (meta)data is stored, linked to participant consent, that allows researchers to track the extent of this consent when it comes to data sharing and reuse both between projects and with the wider scientific community. By addressing technical, legal, and ethical aspects of linguistic data management, the project developing these solutions aims to enhance the efficiency and sustainability of collaborative research within the CRC. The poster will present the key benefits of this approach and will outline the metadata requirements for the platform and the wizard, as well as their component structure.

References

- Jorschick, Annett, Paul Schrader & Hendrik Buschmeier (2024), What can I do with this data point? Towards modeling legal and ethical aspects of linguistic data collection and (re-)use. *Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies at LREC-COLING 2024*. ELRA and ICCL, 47–51. <https://aclanthology.org/2024.legal-1.8>
- Kanza, Samantha & Nicola J. Knight (2022), Behind every great research project is great data management. In: *BMC Res Notes* 15(20). <https://doi.org/10.1186/s13104-022-05908-5>
- Mittal, Deepti, Rebecca Mease, Thomas Kuner, Herta Flor, Rohini Kuner & Jamila Andoh (2023), Data management strategy for a collaborative research center. In: *GigaScience* 12, <https://doi.org/10.1093/gigascience/giad049>
- Pasquetto, Irene V., Bernadette M. Randles & Christine L. Borgman (2017), On the reuse of scientific data. *Data Science Journal* 16(8), 1-9. <https://doi.org/10.5334/dsj-2017-008>

Numeral classifiers in Indo-Aryan. Some new diachronic and typological insights

Krzysztof Stroński, Riddhi Nirkhe, Husendra Sethiya (Adam Mickiewicz University in Poznań)

The reanalysis of grammatical gender-based system leading to a classifier-based one has been well documented in eastern Indo-Aryan (IA) languages such as Assamese, Bengali and Oriya where classifiers have been in use from their early stages (Chatterji 1926: 779-780; Aikhenvald 2000: 379). Other eastern IA languages (e.g. Maithili) that have lost grammatical gender quite early, have introduced classifiers only recently (Barz and Diller 1985). Those IA languages in which both classifiers and grammatical gender coexist (e.g. Nepali) are rather uncommon (Tang 2019: 15).

The aim of the present paper is to show that concurrent systems are much more common than previously expected and that the diachrony of classifiers in IA points not only towards generally accepted Sino-Tibetan source (Allasonnière-Tang and Kilarski 2020) but also towards internal developments which may be driven by the register selection.

Contemporary eastern IA show variation as regards the number of classifiers and their distributional patterns, for example: their position (as given in (1) and (2)), as well as the possibility to encode grammatical gender (3) or definiteness (4):

1. Awadhi (Sohail 1986: 132)

pəʈʃas tʰo goiʈna
50 CLF cowdung cake

2. Bhatari (Sethiya 2023: 19)

goʈ-ək bokəɖa-ke kaʈ-laje
CLF-one goat-ACC cut-PFV
'One goat was chopped.'

3. Nepali

jəw-t-i bahini
one-CLF-F sister

4. Kortha (Paudyal: 116)

khāci=tə
bucket=CLF
'the bucket'

Our preliminary research also shows that Nepali classifier system can be traced back to the 14th c. (cf. Chalise 2015) and there are no Sino-Tibetan elements in it. Moreover it is rather the Sino-Tibetan language Newari which borrowed the classifiers from Nepali. Among eastern languages Maithili shows single examples of classifiers at the very early stage but their regular use has been reported only by the beginning of the 20th century.

IA languages spoken to the west of Magadhan tongues either have developed classifiers quite recently (e.g. Awadhi) or they do not show them at all. However, in several western historical dialects (e.g. Braj), as well as in contemporary Gujarati and Marathi we find elements which could serve as classifiers despite the fact of maintaining fully-fledged grammatical gender system (see ex. (5); cf. also Shibatani 2023).

5. Marathi

dəha dzəŋ a-l-e/ dzəŋi a-l-ja
 ten CLF come-PFV-PL.M / CLF.F come-PFV-PL.F
 'Ten people / women came.'

Outside of the main IA area Romani may show borrowed elements (ex. (6) with a CLF element borrowed from Turkish) or vernacular elements (ex. (7) showing *žen* < OIA *jana* 'man') that co-exist with 2 grammatical genders.

6. Sepečides Romani (Cech and Heinschink 1996: 40)

a) *jek tan-o Rom*
 one CLF-M Rom

b) *jekh tan-i romni*
 one CLF-F woman

7. Kelderashi Romani (Oslo 2018: 243)

dùj-žen-è kətàn-i
 two-CLF-PL soldier[M]-PL

The data analysed by us show that the early development of classifier system in eastern IA as well as in Nepali can be a result of pattern replication (possibly from Sino-Tibetan or Austroasiatic) but certainly the lexical elements are vernacular and those elements have been often borrowed from IA to neighbouring non-IA languages.

REFERENCES

- Aikhenvald, Alexandra. 2000. *Classifiers*. Cambridge: CUP.
- Allasonnière-Tang Marc and Marcin Kilarski. 2020. Functions of gender and numeral classifiers in Nepali. *Poznań Studies in Contemporary Linguistics* 56(1), 113–168.
- Barz, Richard K. and Anthony V.N. Diller 1985. Classifiers and standardization: some South and South-East Asian comparisons. In D. Bradley (ed.), *Papers in South-East Asian linguistics* No. 9: Language policy, language planning and sociolinguistics in South-East Asia, 155-184. Pacific Linguistics, A-67, 1985.
- Cech, Petra and Mozes F. Heinschink. 1996. *Sepečides-Romani*. (Languages of the World / Materials, 106.) München: Lincom Europa.
- Chalise, Bidurkumar 2015. *Nepāli avyaya*, [Nepali indeclinables]. Kathmandu: Unique Educational Publishers.
- Chatterji, Suniti Kumar. 1926. *The Origin and Development of the Bengali Language*. Part II: Morphology. London: George Allen & Unwin Ltd.
- Paudyal, Netra P. (forthcoming). *A grammar of Khortha*. Leiden: Brill.
- Peterson, John and Sunil Baraik. 2022. *A Grammar of Chotanagpuri Sadri. A lingua franca of eastern central India*. Mysore: Central Institute of Indian Languages (CIIL)
- Sethiya, Husendra. 2023. *Classifiers, Gender Marking and Relativization in Halbi and Bhatari: A Nominalization Perspective*. Deccan College PGRI, Pune. MA thesis.

- Shibatani, Masayoshi. 2023. Towards Dynamic Functional Typology — Classifiers, genders, and nominalization. *Historical Linguistics in Japan* 12, 1-52.
- Sohail, Ahmad. 1986. *A descriptive analysis of eastern Awadhi dialect of Jaunpur*. Aligarh Muslim University. PhD thesis.
- Tang, Marc. 2019. *A typology of classifiers and gender. From description to computation*. Studia Linguistica Upsaliensia 23. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0507-3.

“Upwards towards the forest and away from the river”: The Upper Negidal riverine landmark-based orientation system and the role of rivers in the expression of direction

Laurène Barbier

(Laboratoire Dynamique Du Language CNRS UMR 5596; Université Lumière Lyon 2)

Keywords: Orientation system, Frame of reference, Riverine landmark-based, Spatial events, Tungusic

Spatial frames of reference, as theorized by (Levinson 1996; Levinson 2003), have been applied to describe the many types of orientation systems attested in the world's languages, either relative, absolute or intrinsic. As shown in the literature on other language families such as, for example, Austronesian (Blust 2013; Nagaya 2020) and Sino-Tibetan (Post 2019; Shirai 2020), the description of spatial orientation has often proved useful in understanding the encoding of direction across languages (Honkasalo 2019: 533–544; Nagaya 2022). In this context, Upper Negidal, one of two varieties of Negidal, a now moribund North-Tungusic language (Whaley & Oskolskaya 2020) spoken in the Russian Far East by four speakers, remains underdescribed. Despite its critical linguistic situation, its orientation system has not yet been the object of an in-depth and systematic research.

The Negidals are a traditionally semi-sedentary people, established along the middle reaches of the Amgun River, a tributary of the Amur River (Pevnov & Khasanova 2006: 450). The North of the Lower Amur basin is indeed a region of rivers surrounded by taiga forests, which is reflected in the Negidal's traditional economy based on fishery. Given this close relationship with rivers, the aim of this study is to examine the characteristics of Upper Negidal's orientation system, in relation to riverine motion events. The purpose is to describe how orientation in Upper Negidal expresses information on direction and how closely related it is to the riverine surroundings. Hence, to what extent does riverine vocabulary play a particular role in Upper Negidal's spatial grammar? How is it embedded in the orientation system? And what is Upper Negidal's spatial frame of reference?

For the purpose of this study, I examined a corpus of narratives of various genres told by 8 Upper Negidal speakers (Pakendorf & Aralova 2017). From this corpus, I extracted 3729 utterances expressing spontaneous translational motion (i.e. not caused by an event or a person). The data was then coded following Slobin's Berkeley coding manual (2005).

The analysis of these narratives indeed reveals frequent references to rivers, highlighting the Negidals' profound connection to their riverine environment and how deeply intertwined it is with the social and linguistic practices of the community. This connection is reflected in Upper Negidal spatial grammar, since riverine verbs, adverbs and relational nouns are attested in the language system, as described in other Asian languages (Sims & Genetti 2017: 121–123; Honkasalo 2019: 533). Additionally, results reveal that Upper Negidal displays an absolute orientation system based on the river flow and topographical features linked to the rivers, such as riverbanks and forests rising from the riverbed, that characterizes Upper Negidal's area of practice.

This study shows that Upper Negidal orientation system can be represented as a biaxial system composed of a forest-river vertical axis, perpendicular to the river, and an upriver-downriver horizontal axis, therefore parallel to the river. This research sheds light on an unknown area of Upper Negidal spatial grammar, therefore contributing to a better understanding of this now endangered language system.

Acknowledgements

This research was funded by l'Agence Nationale de la Recherche (ANR) project *Spatial Asymmetries Across Languages: a Typological approach* (SALTA, ANR-20-CE27-0015) and by the 2024 JSPS Postdoctoral Fellowship for Research in Japan (short-term).

References

- Blust, Robert. 2013. *The Austronesian languages*. Asia-Pacific Linguistics, School of Culture, History and Language, College of Asia and the Pacific, The Australian National University.
- Honkasalo, Sami. 2019. "Chapter eight: Orientation, tense-aspect-mood, and modality". *A Grammar of Eastern Geshiza: A Culturally Anchored Description*. Helsinki: Helsingin yliopisto. 533–580.
- Levinson, Stephen C. 1996. Language and space. *Annual Review of Anthropology*. US: Annual Reviews 25. 353–382.
- Levinson, Stephen C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. 1st edn. Cambridge University Press.
- Nagaya, Naonori. 2022. Directionals, topography, and cultural construals of landscape in Lamaholot. *Linguistics Vanguard* 8(s1). 25–37.
- Pakendorf, Brigitte & Natalia Aralova. 2017. *Documentation of Negidal, a nearly extinct Northern Tungusic language of the Lower Amur*. <http://hdl.handle.net/2196/b644db81-725c-4031-935c-f33c763df152>. (Last accessed 27/05/2025)
- Pevnov, Alexander & Marina Khasanova. 2006. Negidal'cy i negidal'skij jazyk [Negidals and Negidal language]. In N. Kazansky (ed.), *Acta Linguistica Petropolitana*, vol. 2, 447–542. Nauka. Saint Petersburg.
- Post, Mark W. 2019. Topographical Deixis in Trans-Himalayan (Sino-Tibetan) Languages. *Transactions of the Philological Society* 117(2). 234–255.
- Shirai, Satoko. 2020. A geolinguistic study of directional prefixes in the Qiangic language area. *Himalayan Linguistics* 19. 365–392.
- Sims, Nathaniel & Carol Genetti. 2017. The Grammatical Encoding of Space in Yonghe Qiang. *Himalayan Linguistics* 16. 99–140.
- Slobin, Dan I. 2005. *Coding of Motion Events in Texts*.
- Whaley, Lindsay & Sofia Oskolskaya. 2020. The classification of the Tungusic languages. In *The Oxford Guide to the Transeurasian Languages*. Oxford University Press. 80–91.

The functional load of lenition in some Oceanic and Papuan languages of Island Melanesia

Lidia Federica Mazzitelli

(University of Naples L'Orientale)

Keywords: morphology; phonology; Oceanic; Austronesian; transitivity

Different morphophonological processes of consonant mutation occur in the Oceanic and Papuan languages spoken in Island Melanesia, a region that extends from the Bismarck Archipelago to New Caledonia, excluding the island of New Guinea. Here, I focus on one such process, the lenition of voiceless plosive consonants in intervocalic contexts (CVC), and I examine it in six languages of the region: Lakurumau, Kara, and Lamasong (Oceanic; New Ireland, Papua New Guinea); Kuot (Papuan, isolate; New Ireland, Papua New Guinea); Qaqet and Mali (Papuan, Baining; New Britain, Papua New Guinea). The data come from my own fieldwork as well as from descriptive grammars.

In all these languages, lenition of voiceless plosives (/p/ > /v/; /β/; /k/ > /ɣ/, /x/; /t/ > /r/) is a regular process, which also occurs beyond word boundaries; see (1a-b).

- (1) a. Lakurumau
 puraa ['pura] ~ *a=vuraa* [ə'vura]
 chicken DEF=chicken
 b. Qaqet
 "dap=ian=tit=a ["daβiandira]
 but=3DU.SBJ=go.CONT=DIST
 'They go now' (Hellwing 2019: 25)

Lenition is not reconstructed for Proto-Oceanic and is assumed to have developed independently in different daughter language groups. In the Oceanic languages of New Ireland, its emergence is attributed to contact with Kuot, the island's only non-Austronesian language (Ross 1994). Within the Baining group, lenition occurs only in Qaqet and Mali (Hellwig 2019: 26).

Lenition has a functional role in all these languages except Kuot. In Lakurumau and Kara, the absence of lenition in expected contexts marks syntactic intransitivity (2). In Lamasong, the blocking of lenition in intervocalic contexts indicates that the predicate is continuous and non-punctual (3). In Qaqet and Mali, the lenited form of certain verbal stems marks the verb as non-continuous (Qaqet) or non-present (Mali) (5).

- (2) Lakurumau
 kaabang ['kabəŋ] 'help.INTR' *xabong* [ɣə'bɔŋ] 'help.TR'
(3) Lamasong
 E=kis [e'kis] *e=xis* [e'xis]
 1SG.S=sit.CONT 1SG.S=sit.PUNCT
 'I am sitting' 'I sit down'
(4) Qaqet
 kut [kut] 'dig.CONT' *qut* [ɣut] 'dig.NCONT' (Hellwig 2019: 307)

In at least four languages, there appears to be a correlation between lenition and transitivity: in Lakurumau and Kara, the blocking of lenition signals syntactic intransitivity, while in Qaqet and Lamasong, it marks the verb as non-punctual and thus semantically less transitive. In Qaqet (and possibly Mali), the absence of lenition in continuous stems can be traced back to a now-lost consonantal durative prefix, which originally blocked the lenition process (Hellwig 2019: 308). In

Lakurumau, Kara, and Lamasong, conversely, there is no clear evidence of a diachronic, language-internal process preventing lenition. It is tempting to suppose that the Oceanic languages of New Ireland have borrowed not only the phonological phenomenon of lenition from their Papuan neighbors but also its functional correlation with (in)transitivity. However appealing this hypothesis may be, the absence of a grammatical function for lenition in Kuot—the closest Papuan neighbor of Lakurumau, Kara, and Lamasong—and the lack of evidence for direct contact between the languages of North-Central New Ireland and the Baining languages make it difficult to sustain.

References

- Hellwig, Birgit (2019) *A grammar of Qaet*. Berlin: Mouton de Gruyter
- Lindström, Eva (2002) *Topics in the grammar of Kuot: A non-Austronesian language of New Ireland, Papua New Guinea*. Stockholm: Stockholm University dissertation.
- Ross, Malcom (1994) Areal phonological features in north central New Ireland. In Tom Dutton & Darrell T. Tryon (eds.), *Language contact and change in the Austronesian world* (Trends in Linguistics. Studies and Monographs, 77), 551–572. Berlin: Mouton de Gruyter
- Schlie, Perry, and Virginia Schlie (1993) *A phonology of Kara*. Typescript. Ukarumpa: Summer Institute of Linguistics.
- Stebbins, Tonya N. (2011) *Mali (Baining) grammar: A language of the East New Britain Province, Papua New Guinea* (Pacific Linguistics, 623). Canberra: Pacific Linguistics.

Accounting for mass nouns in Asante-Twi

Madeline Ladore

Georg-August Universität Göttingen

Keywords: Mass-count distinction, Niger-Congo, Asante-Twi, Semantics, Morphology

Most nouns in Asante-Twi (Niger-Congo; Kwa) have a singular-plural number distinction, expressed morphologically as a noun-class prefix (1), while others lack a morphological number distinction (MND)(2). The nouns in (2) fall into two camps: All mass nouns (2a), which, like most languages, have no MND, and seemingly-count nouns that have lost their MND through the degradation of the noun class system (2b) (cf. Osam 1993).

(1) MND nouns

a-konya	n-konya
SG-chair	PL-chair

(2) non-MND nouns

a. a-ɲwa	m-frama	b. a-kutuo	n-homa
NC-oil	NC-air	NC-orange	NC-book

This change is not only morphological, but also semantic in nature, as (2a) and (2b) nouns fit in the same grammatical contexts: They combine with mass quantifiers *dodo* ‘much’ and *kakra* ‘a bit of’, which is not possible for MND nouns.

(3) a-kutuo dodo / kakra	m-frama dodo / kakra	a-/n-konya *dodo / *kakra
NC-orange much / a.bit.of	NC-air much / a.bit.of	SG-/PL-chair much / a.bit.of

Non-MND nouns like *ɛwɔɔ* ‘honey’, *nyadoa* ‘garden eggs’ can also be quantitatively compared with one another, while MND nouns like *nkonya* ‘chairs’ would make the comparison infelicitous.

(4) Kofi wɔ ɛ-wɔɔ bebre sen Akosua a ɔ-wɔ n-yadoa / #n-konya
Kofi has NC-honey more COMP Akosua REL she-has NC-garden.eggs / #PL-chair
Kofi has more honey than Akosua has garden eggs / #chairs

However, non-MND nouns can fit in grammatically count contexts in that they combine with numerals without classifiers or container phrases for a unit reading, not a kind reading (5). In this sense, they behave similarly to mass nouns in Yudja (Lima 2014) and Nez Perce (Deal 2017).

(5) me.hia n-homa mmiensa	me.tɔɔ a-ɲwa mmienu.
I.need NC-book three	I.buy-PST NC-oil two
‘I need three books.’	‘I bought two containers / #kinds of oil.’

Non-MND nouns also combine with count adjectives like *kesee* ‘big’, *ketewa* ‘small’, which can be interpreted idiomatically.

(6) a-kutuo kesee	m-frama kesee
NC-orange big	NC-air big
‘big orange(s)’	‘strong wind’

As nouns like *akutuo* ‘orange’ and *aɲwa* ‘oil’ pattern together, it is likely the case that they have the same underlying denotation that is different from that of MND nouns. I propose that non-MND nouns are inherently homogenous (both cumulative and divisive (Bunt 1985)) and that MND nouns are

inherently quantized (anti-divisive; no proper part of p is also p (Krifka 1989)). While true that homogenous denotations cannot fit in count contexts like (5) & (6), this poses no issue, as semantic operators (such as Deal's (2017) α_n) are able to shift homogenized denotations to quantized ones. This α_n -operator allows for non-MND nouns to be felicitous in count contexts, while it does not apply to non-MND nouns in mass contexts.

This research presents a novel description of the mass-count distinction in Asante-Twi, which has, to this point, gone uninvestigated at the semantic level. The behaviour of mass nouns in Asante-Twi is not typologically unique, as many other languages allow for mass nouns in count contexts. However, certain morphological changes in the noun class system have resulted in semantic changes to the mass-count divide in the language.

References

- Bunt, Harry C. 1985. *Mass terms and model-theoretic semantics*. Cambridge University Press.
- Deal, Amy Rose. 'Countability Distinctions and Semantic Variation'. *Natural Language Semantics* 25, no. 2 (June 2017): 125–71. <https://doi.org/10.1007/s11050-017-9132-0>.
- Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In *Semantics and Contextual Expression*, eds. R. Bartsch, J.F.A.K. van Benthem, and P. von Emde Boas, 75–115. Dordrecht: Foris Publications.
- Lima, S. (2014, August). All notional mass nouns are count nouns in Yudja. In *Semantics and Linguistic Theory* (pp. 534-554).
- Osam, Emmanuel Kweku. 'The Loss of the Noun Glass System in Akan'. *Acta Linguistica Hafniensia* 26, no. 1 (1993): 81–106.

Interpreting vague quantifiers in survey scales: The effects of respondents' socio-demographic background and scale type

Maria Reile & Rainer Reile

(University of Tartu & National Institute for Health Development)

Keywords: vague quantifiers, survey scales, semantics, experimental approach, random sampling

Vague quantifiers is an overarching term for measurement expressions used in surveys, such as quantity expressions, e.g., *many*, frequency expressions, e.g., *sometimes*, and intensity expressions, e.g., *somewhat* (agree) (e.g., Al Baghal 2014). In linguistics, vague quantifiers have been studied mostly from the viewpoint of scalar implicatures (e.g., De Carvalho et al. 2016) or focusing on the comparison of two (opposite) terms (e.g., Katzir & Ariel 2023). However, in surveys, the vague quantifiers are often presented on an ordered Likert-type scale. To better understand how to interpret such responses, the meaning of vague quantifiers should also be studied within a similar ordered set.

Previous research has shown cross-linguistic differences in the meaning of vague quantity expressions (Stateva et al. 2019), e.g., *some* in English and *einige* in German, denote a different proportion of quantity within the same context. Rocconi and colleagues (2020) have also found gender difference in the interpretation of frequency expressions, indicating an influence of socio-demographic factors. Consequently, in the context of localising cross-national survey instruments, linguistic variation may introduce measurement bias and affect the validity of the results. Given that previous research has predominantly focused on Indo-European languages (e.g., English, Dutch, German), studies on vague quantifiers in other languages are warranted.

In this study, we focused on Estonian vague quantifiers, more specifically, frequency and intensity expressions, e.g., *mõnikord* 'sometimes' and *väga halb* 'very bad', and studied whether their interpretation varied by survey scale type and socio-demographic characteristics of the participants. To that end, we conducted a web-based experiment using LimeSurvey (an online survey tool). In the first task of the experiment, participants were presented four vignettes (short stories) describing situations: characters' health situation (for intensity expressions) and shopping habits (for frequency expressions). The participants were asked to choose the best match from five vague quantifiers (presented on three different scales: symmetric wide, symmetric narrow, and biased to positive end-point) to best describe each situation. After vignettes, the participants had to position (using slider on a scale) vague quantifiers on a numeric scale with both end-points defined qualitatively.

Population-based random sample of 2400 Estonian native speakers aged 18–69 years was drawn from Population registry. In total, 300 invited individuals (95 men and 205 women) participated in the experiment (response rate 12.5%). Participant's mean age was 48.4 years and 63% of participants had higher education.

The preliminary results of the vignette task showed that scale type did not have statistically significant influence on the choice of vague quantifiers. Regardless of the scale type, the participants chose the same quantifier to match with the situation description in the vignette. However, positioning of the vague quantifiers on a suitable location of the numeric scale (the second task in the experiment) revealed a statistically significant effect of gender and education level, but not of age. These findings align with previous research on the interpretation of English frequency words (Schaeffer 1991, Rocconi

et al. 2020) and suggest a possible inter-linguistic variation in the interpretation of vague quantifiers among participants with different socio-demographic backgrounds.

Acknowledgements

This study was funded by the Kadri, Nikolai, and Gerda Rõuk's research grant (MHVEE7).

References

- Al Baghal, Tarek (2014), Numeric Estimation and Response Options: An Examination of the Accuracy of Numeric and Vague Quantifier Responses. *Journal of Methods and Measurement in the Social Sciences* 5(2), 58–75. <https://doi.org/10.2458/v5i2.18476>.
- De Carvalho, Alex, Anne C. Reboul, Jean-Baptiste Van Der Henst, Anne Cheylus & Tatjana Nazir (2016), Scalar Implicatures: The Psychological Reality of Scales. *Frontiers in Psychology* 7. <https://doi.org/10.3389/fpsyg.2016.01500>.
- Katzir, Nicole & Mira Ariel (2024), *A few or several ?* Construal, quantity, and argumentativity. *Language and Cognition* 16(1), 148–175. <https://doi.org/10.1017/langcog.2023.25>.
- Rocconi, Louis M., Amber D. Dumford & Brenna Butler. 2020. Examining the Meaning of Vague Quantifiers in Higher Education: How Often is “Often”? *Research in Higher Education* 61(2), 229–247. <https://doi.org/10.1007/s11162-020-09587-8>.
- Schaeffer, Nora Cate (1991), Hardly Ever or Constantly? Group Comparisons Using Vague Quantifier. *Public Opinion Quarterly* 55(3), 395–423. <https://doi.org/10.1086/269270>.
- Stateva, Penka, Arthur Stepanov, Viviane Déprez, Ludivine Emma Dupuy & Anne Colette Reboul (2019), Cross-Linguistic Variation in the Meaning of Quantifiers: Implications for Pragmatic Enrichment. *Frontiers in Psychology* 10. <https://doi.org/10.3389/fpsyg.2019.00957>.

Source-Goal Asymmetry and Construction Alternation: Evidence from Latin Preverbed Verbs of Caused Motion

Martina Giuliani
University of Pavia

Preverbs have received much attention in Indo-European and Latin Linguistics. McGillivray (2013) highlights the interplay of preverbation and argument structure, studying Latin preverbed verbs (PVs) of motion and rest. Iacobini et al. (2017) compare the expression of motion events in Ancient Greek and Latin, focusing on Source and Goal encoding. In the linguistics of space, Source–Goal asymmetry is well-studied, though mainly from a synchronic perspective (Stefanowitch/Rohde 2004; Lakusta/Landau 2005).

This work aims at putting together these research fields by examining Source–Goal asymmetries involving Latin PVs both synchronically and diachronically. To do so, I investigate the argument structure constructions of PVs of *fero* ‘bring’ in Early and Classical Latin texts from the LASLA corpus, within the framework of Cognitive Grammar (Langacker 1987, 1999, 2008) and adopting a constructionist approach (Goldberg 1996). Early and Classical Latin corpora are balanced to allow for diachronic comparison. I select PVs of *fero* ‘bring’ because caused motion was less investigated than intransitive motion in previous literature. The analysis focuses on PVs having a transfer meaning compounded with *ad*- ‘to’, *ab*- ‘away’, *circum*- ‘around’, *de*- ‘from’, *prae*- ‘in front’, *per*- ‘through’, and *trans*- ‘across’. I annotate their argument structure constructions, comparing them with those of the base verb and other caused motion verbs (e.g. *mitto* ‘send’). The main goals of this study are: (i) to provide a frequency-informed account of the argument structure constructions of the PVs of *fero* ‘bring’; and (ii) to investigate the Source–Goal asymmetry in the construal of caused motion events in Latin.

My data point out that the construction of PVs has implications on Source-Goal asymmetry. Latin PVs, on one hand, exhibit semantic congruence (Borillo 1998) between the portion of Path expressed in the verbal locus and the one expressed in the adnominal locus (Iacobini et al. 2017). Consider *adfero* ‘bring to(ward)’ and *aufero* ‘bring away’ taking Goal and Source preverbs, respectively: the former selects Direction/Recipient arguments, whereas the latter selects Source/Maleficiary arguments. On the other hand, Latin PVs exhibit the Goal-over-Source principle (Kopecka/Vuillermet 2021) according to which Goals are more frequently expressed than Sources cross-linguistically. Consider the Source PV *defero* ‘bring (away) from’: it selects Goals more often than Sources, contradicting the principle of semantic congruence described for the Source PV *aufero* ‘bring away’. Zanchi (2017: 171) observes similar cases with Ancient Greek preverbs, arguing that “if there is a mismatch between the spatial role expressed by the preverb and that played by the verbal participant, compounds containing a source-preverb can take a goal-participant whereas the opposite is not attested.”

The PVs examined feature construction alternation manifesting in different encodings of Recipient-like (R) participants. The latter take either the dative or the Direction prepositional phrase: *ad* ‘to’ + accusative. This alternation, attested with transfer and communication verbs, is explained through the semantics of verb, construction, and participants (Fedriani/Prandi 2014; Napoli 2018). I analyze the distribution of PVs with different constructions in Early and Classical Latin. In this period *defero* ‘bring (away) from’ is more often found with Rs prepositionally constructed. I show that

Source PVs tend to express Rs with *ad* + accusative for the sake of semantic transparency, given that the Latin dative case encodes both Goal-like and Source-like participants.

References

- Borillo, A. 1998. *L'espace et son expression en français*. Paris: Ophrys.
- Fedriani, C. & M. Prandi. 2014. Exploring a diachronic (re)cycle of roles: The Dative complex from Latin to Romance. *Studies in Language* 38(3): 566–604.
- Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago & London: The University of Chicago Press.
- Iacobini, C., L. Corona, N. De Pasquale & A. Buoniconto. 2017. How should a “classical” Satellite-Framed Language behave?: Path encoding asymmetries in Ancient Greek and Latin. In S. Luraghi, T. Nikitina & C. Zanchi (eds.), *Space in Diachrony*, 95–118. Amsterdam: John Benjamins.
- Kopecka, A. & M. Vuillemet. 2021. Source-Goal (a)symmetries across languages. *Co-edited special issue of Studies in Language* 45: 1–35.
- Lakusta, L. & B. Landau. 2005. Starting at the end: The importance of goals in spatial language. *Cognition* 96. 1–33. <https://doi.org/10.1016/j.cognition.2004.03.009>.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar (Vol. I). Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1999. *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.
- McGillivray, B. 2013. Latin preverbs and verb argument structure: New insights from new methods. In E. van Gelderen, M. Cennamo & J. Barðdal (eds.), *Argument Structure in Flux*, 119–133. Amsterdam: John Benjamins.
- Napoli, M. 2018. Ditransitive Verbs in Latin: A Typological Approach. *Journal of Latin Linguistics* 17(1): 51–91.
- Stefanowitsch, A. & A. Rodhe. 2004. The goal bias in the encoding of motion events. In G. Radden & K. Panther (eds.), *Studies in linguistic motivation*, 249–267. Berlin: De Gruyter Mouton.
- Zanchi, C. 2017. New evidence for the Source-Goal asymmetry. In S. Luraghi, T. Nikitina & C. Zanchi (eds.), 147–178.

Web resources

<https://www.lasla.uliege.be/>.

<https://hyperbase.unice.fr/>.

Swearing Beauty: The Sound Symbolism of Slavic Profanity

Martina Rybová

(Ludwig-Maximilians-Universität Munich)

Keywords: sound symbolism, profanity, sonority, phonology, West Slavic languages

As part of the lexical inventory of language, swear words can be approached from many perspectives: e.g. neurocognitively, they exhibit specific behaviour since they are not processed in the language centre, but in the limbic system, which is evolutionarily linked to emotions (Finkelstein 2018). Thus, phylogenetically, they have more in common with physiological reactions to threat and danger (the so-called fight-or-flight response; Stapleton 2022) than with verbal utterances themselves. Both these viewpoints are recently put in context with a potentially unique sound structure of swear words, mainly known as sound symbolism, making them clearly recognisable as means of verbal aggression (e.g. Reilly et al. 2020, and Lev-Ari & McKay 2023). Previous research on this topic has revealed both language-universal patterns, such as the absence of approximants (Lev-Ari & McKay 2023), and language-specific tendencies, i.e. prevalence of short, closed syllables and plosives in English and German (Bergen 2016, and Aryani et al. 2018).

This paper presents first findings from an analysis of the sound structure and possible sound symbolism of swear words in Czech, Slovak, and Polish, three geographically and socio-culturally closely related languages from the West branch of Slavic languages. This investigation aims to provide novel insights into the degree of association between sound and profanity, namely whether the spatial and/or socio-cultural proximity of languages entails similar perceptual tendencies—both regarding sound symbolism or iconicity in general and swear words in particular.

The initial results are based on sixty swear words and sixty neutral words, each with an equal total number of phonemes (e.g., in Polish, each sample contains 340 phonemes). These words were extracted from available lexicons and Swadesh lists of the respective languages and automatically transcribed into the International Phonetic Alphabet. Both samples were controlled for morphological structure (i.e., to avoid the influence of affixes such as *-ka* in Polish *sraczka* [sɾatʃka]) and analysed for phonological patterns as well as for sonority—a measure proposed by Aryani et al. (2018) as indicative of a swear word's emotional content and (un)pleasantness depending on what phonemes it consists of, though not yet verified (cf. Blohm et al. 2021, and Jacobs 2023).

Preliminary findings suggest that West Slavic swear words significantly favour trilled [r] and exhibit slightly lower sonority than neutral words. In the Czech data, we observed a mean sonority difference of -0.2, with some highly sonorous outliers such as *idiot* [ɪdɪjot]. This partially aligns with the 'unpleasant' sound qualities noted by Jacobs (2023), supporting the hypothesis that sonority may be one of the distinguishing features of profane language.

The main focus is to discuss the relevance of used measures as proxies for identifying swear words and for generating suitable pseudowords with similar perceptual characteristics, which could be used for further experimental research on the interplay between form and meaning of profane language means.

References

Aryani, Arash, Conrad, Markus, Schmidtke, David, & Jacobs, Arthur (2018), Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making, *PLoS ONE* 13(6).

- Bergen, Benjamin K. (2016), *What the f: What swearing reveals about our language, our brains, and ourselves*, New York: Basic Books.
- Blohm, Stefan, Kraxenberger, Maria, Knoop, Christine A., & Scharinger, Mathias (2021), Sound Shape and Sound Effects of Literary Texts. In D. Kuiken, and A. M. Jacobs (eds), (2021), *Handbook of Empirical Literary Studies*, De Gruyter, 7–38.
- Finkelstein, Shlomit R. (2018), Swearing and the brain. In K. Allan (ed), (2018), *The Oxford Handbook of Taboo Words and Language*, Oxford: Oxford Handbooks.
- Jacobs, Arthur M. (2023), *Neurocomputational poetics: how the brain processes verbal art*, London/New York: Anthem Press.
- Lev-Ari, Shiri, & McKay, Ryan (2023), The sound of swearing: Are there universal patterns in profanity?, *Psychon. Bull. Review* 30(3), 1103–1114.
- Reilly, Jamie, Kelly, Alexandra, Zuckerman, Bonnie M., Twigg, Peter P., Wells, Melissa, Jobson, Katie R., & Flurie, Maurice (2020), Building the perfect curse word: A psycholinguistic investigation of the form and meaning of taboo words, *Psychon. Bull. Review* 27, 139–148.
- Stapleton, Karyn, Beers Fägersten, Kristy, Stephens, Richard, & Loveday, Catherine (2022), The power of swearing: What we know and what we don't, *Lingua* 277, 1–16.

Underlying clauses of German contrastive fragments: Evidence for a Q-based approach

Miriam L. Schiele
(University of Tübingen)

Keywords: clausal ellipsis, German, acceptability judgments

In German fragments (1B), retaining the preposition (P) is favored (Molimpakis 2019, Lemke 2021). Merchant (2001) argues that P-omission is unacceptable because fragments derived by P(reposition)-stranding (2) (where grey shading = ellipsis) are ungrammatical in German.

- (1) A: Peter hat mit jemandem gesprochen. B: (Mit) wem?
 ‘Peter was talking to someone.’ ‘(To) who?’
- (2) * Wem₁ hat Peter mit t₁ gesprochen?
 ‘Who was Peter talking to?’

Alternatively, Nykiel & Hawkins (N&H) (2020) argue that case-assigning Ps are required in fragments with case-marked DPs. In the German example in (3), the masculine DP is unambiguously case-marked, whereas feminine DP exhibit case syncretism. While the P in (3B') forces an ACC-reading, a default reading as NOM is available in (3B'').

Furthermore, (3B'') could be derived from an isomorphic clause (4i) or a copular clause (4ii), as there is no consensus regarding the underlying structure of a fragment (Nykiel 2013, Ronai & Stigliano 2020, Rodrigues et al. 2009, Craenenbroeck 2010, and Szczegielniak 2008). Crucially, (4ii) circumvents P-stranding, rendering (3B'') acceptable. I hypothesize (H1) that P-omission is degraded if the fragment is unambiguously case-marked and derived by P-stranding.

- (3) A: Der Chef hat sich über jemanden aufgeregt.
 ‘The boss got upset about an employee.’
- B: Ja, (über) einen Mitarbeiter.
 ‘Yes, (about) an employee.M.’
- B': Ja, über eine Mitarbeiterin.
 ‘Yes, about an employee.F.’
- B'': Ja, eine Mitarbeiterin.
 ‘Yes, an employee.F.’
- (4) i. [CP [eine.ACC Mitarbeiterin.F.ACC]₁ hat [TP sich der Chef über t₁ aufgeregt]]
 ‘Yes, an employee the boss got upset about.’
- ii. [CP [eine.NOM Mitarbeiterin.F.NOM]₁ war [TP das t₁]]
 ‘Yes, an employee that was.’

Barros (2012) proposes that contrastive focus on the remnant forces one to interpret the ellipsis site as isomorphic, i.e., syntactically parallel to the antecedent clause, making copular clauses unavailable as underlying clause. Thus, (5B) is unacceptable due to P-stranding, despite being ambiguously case-marked. I hypothesize (H2) that P-omission is generally degraded in contrastive focus.

- (5) A: Der Chef hat sich über seinen Vorgänger aufgeregt.
 ‘The boss got upset about his predecessor.’

B: **Nein**, eine Mitarbeiterin.
 ‘No, an employee.F.’

48 German monolinguals were sourced via Prolific (online, unsupervised). The study included dialogues such as in (3) and (5), testing the factors PREPOSITION (P-retention/P-omission), CLAUSE (isomorphic/ambiguous), and FOCUS (contrastive/presentational), with 4 repetitions per condition selected from 16 lexical sets, with a 2:1 ratio of fillers to test items. Participants rated the naturalness of speaker B’s answer on a 7-point Likert-type scale, where 1 = fully natural and 7 = fully natural. Figure 1 shows the mean ratings. Participants’ raw ratings were z-scored and analyzed using linear mixed models in R’s *lmer*.

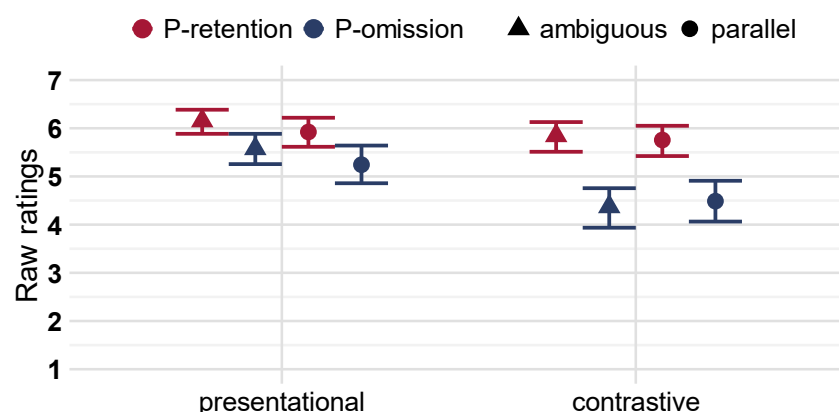


Figure 1

While PREPOSITION is significant ($t = 6.39$, $p < 0.01$), CLAUSE is not ($t = -0.25$, $p = 0.80$), contradicting with H1. When incorporating contrastive stimuli, an interaction between FOCUS and PREPOSITION emerges ($t = -3.56$, $p < 0.01$), supporting H2.

The data challenge N&H’s approach, as case marking is not significant, and contradict Merchant’s generalization, showing no effect of P-stranding. Instead, the findings support a Q-based approach (Ginzburg 1994; Griffiths 2019), where the implicit question (Q) eliciting a fragment must be a syntactically well-formed linguistic object. Crucially, in German, Q must retain the preposition. Under contrastive focus, a ‘corrective Q’ (Griffiths 2019), requiring identical structured meanings of the Q and the fragment. The F-marking on the P reinforces the need to retain the P, degrading P-omission further. Not only do my experimental results provide strong support for the Q-based approach to clausal ellipsis, but they undermine its competitors.

Acknowledgments

This research was supported by the German Research Foundation, project number 522896622.

References

- Barros, Matthew (2012), Short sources and pseudosluicing: A non-repair approach to island sensitivity in contrastive TP ellipsis, in A. Beltrama, T. Chatzikonstantinou, J. L. Lee, M. Pham, and D. Rak (eds), (2012), *Proceedings of the Annual Meeting of the Chicago Linguistic Society 48*, Chicago: Chicago Linguistic Society, 61-75.
- Ginzburg, Jonathan (1994), An update semantics for dialogue, in H. Bunt, R. Muskens, and G. Rentier (eds), (1994), *Proceeding of the 1st International Workshop on Computational Semantics*, Tilburg: Tilburg University, 111-120.

- Griffiths, James (2019), A Q-based approach to clausal ellipsis: Deriving the preposition stranding and island sensitivity generalisations without movement, *Glossa: a journal of general linguistics* 4(1), 1-41.
- Lemke, Robin (2021), *Experimental investigations on the syntax and usage of fragments*, Berlin: Language Science Press.
- Merchant, Jason (2001), *The syntax of silence: Sluicing, islands, and the theory of ellipsis*, Oxford: Oxford University Press.
- Molimpakis, Emilia (2019), *Accepting preposition-stranding under sluicing cross-linguistically: A noisy-channel approach*, doctoral thesis, London: University College London.
- Nykiel, Joanna (2013), Clefts and preposition omission in sluicing, *Lingua* 123, 74-117.
- Nykiel, Joanna and John A. Hawkins (2020), English fragments, minimize domains, and minimize forms, *Language and Cognition* 12(3), 411-443.
- Rodrigues, Cilene, Andrew Nevins, and Luis Vicente (2006), Cleaving the interactions between sluicing and preposition stranding, in D. Torck and W. L. Wetzels. (eds), *Romance Languages and Linguistic Theory 2006*, Amsterdam: John Benjamins, 175-198.
- Ronai, Eszter and Laura Stigliano (2020), Adjectival sluices in Hungarian: An argument for isomorphic sources, *Natural Language & Linguistic Theory* 38, 1197-1210.
- Szczegelnia, Adam (2008), Islands in sluicing in Polish, in N. Abner and J. Bishop (eds), *Proceedings of the 27th West Coast Conference on Formal Linguistics*, Somerville: Cascadilla Proceedings Project, 404-412.
- Van Craenenbroeck, Jeroen (2010), Invisible last resort: A note on clefts as the underlying source for sluicing, *Lingua* 120, 1714-1726.

The Diachronic Grammaticalization of Japanese Adverbs:

A Study of *Kamahete* and *Kesshite*

Mitsuko Takahashi

(Nagaoka University of Technology)

This study investigates the diachronic grammaticalization of the Japanese adverbs *kamahete* and *kesshite*, tracing their transition from intensifier adverbs with rich lexical meanings to negative polarity adverbs characterized by reduced semantic content and restricted contextual distribution. Both adverbs underwent similar semantic and functional shifts, moving from use in both affirmative and negative contexts to exclusive use in negative constructions. Their competitive interaction is also examined, revealing that by the end of the 19th century, *kamahete* had disappeared, while *kesshite* had assumed its role as the primary negative polarity adverb.

Until the 17th century, *kamahete* was widely used across genres such as war tales, essays, and *ukiyo zōshi*, functioning as an intensifier adverb in both affirmative and negative contexts and conveying a range of lexical meanings, including “truly,” “certainly,” and “absolutely.” It also played a pragmatic role in speech acts such as requests, commands, and invitations. From the 18th century onward, however, its affirmative usage declined, becoming restricted to negative contexts—particularly negative commands and prohibitions—most commonly appearing in the fixed expression *kamahete* – *na*, where it conveyed the negative meaning of “never” or “absolutely not.”

In contrast, *kesshite* first appeared in the early 18th century, initially conveying affirmative meanings such as “certainly” and “definitely.” For example:

Kesshite tonosama-no gyoi-ni iri meshikakae-raruru-wa hitsujō

(She would have **definitely** been favored and taken by the lord: this is certain.)

Ejima Kiseki, *Chūshin ryaku Taiheiki*, ca. 1712

Unlike *kamahete*, early *kesshite* did not co-occur with the negative particle *-na*. However, from the late 18th century to the 19th century, it increasingly appeared with various negative particles, including *-mai*, *-nu*, *-nai*, and *-na*, eventually forming the fixed expression *kesshite* – *nai* by the 20th century, where it came to express strong negation in the sense of “absolutely not” or “truly not.”

The data indicate that *kamahete* fell out of use by the end of the 19th century as *kesshite* solidified its function as a negative polarity adverb. The final recorded instance of *kamahete* appears in a prohibitive construction in a literary work from 1898, where it functions as a negative polarity adverb conveying strong prohibition.

This study contributes to our understanding of the historical evolution of negative polarity adverb in Japanese by offering insights into the mechanisms underlying grammaticalization and the influence of synchronic linguistic structures on diachronic change.

The shifts observed in *kamahete* and *kesshite* may exemplify broader patterns shared by other Japanese negative polarity adverbs (e.g., *taishite* “not particularly,” *ikkō* “not at all,” and *chittomo* “not in the least”). Moreover, comparable developments in other languages — for instance, the English adverb *hardly*, which in Old and Middle English conveyed meanings quite distinct from its modern usage, including “severely,” “harshly,” and “hard” — suggest the existence of cross-linguistic tendencies in diachronic grammaticalization.

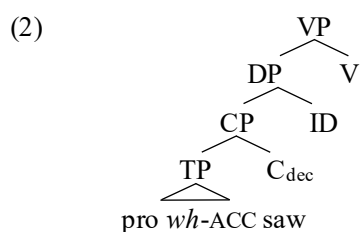
References

- Hachimonjiya-bon kenkyūkai (ed.) (1993) *Hachimonjiya-bon zenshū* Vol. 3. Tokyo: Kyūko Shoin.
- Nihon kokugo daijiten (JapanKnowledge) (<http://japanknowledge.com>)
- Takahashi, Mitsuko (2016) “Historical Changes of *Hardly*: Analyses from the Viewpoint of Diachronic Grammaticalization” (in Japanese) *The Bulletin of Saitama Prefectural University* 18: 1-17.
- Takahashi, Mitsuko (2018) “Grammaticalization of the Japanese Adverb *Kesshite*.” *Shakai Gakubu Ronsō* 29 (1): 97-115, Ryūtsū Keizai University.
- Takahashi, Mitsuko (2023) “Historical Changes of the Japanese Adverb *Kamahete*.” *Japanese/Korean Linguistics* 30: 683-692.

It is well-known that *wh* operators require licensors (Nishigauchi 1990, Chomsky 2015). However, Saito (2024) observes Japanese allows the construction involving *wh* operators without the licensors (the bare *wh* construction) as in (1).

- (1) Taro-wa [pro dare-o mita] to tutaeta.
 T.-TOP who-ACC saw C_{declarative} told
 ‘Taro told (me) who he saw.’

As shown in the translation, (1) has an interpretation like concealed questions, which means the head of this construction is DP. We adopt Nishigauchi’s (2020) analysis of concealed questions and assume the identity head ID (see also Barker 2016). (2) is the VP structure of (1).

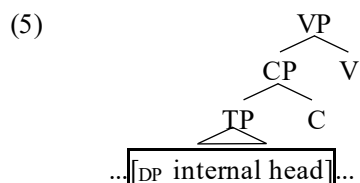


Next, let us consider (3) and (4).

- (3) Taro-wa Hanako-ga katta mono-ga wakaranai.
 T.-top H.-nom bought thing-nom know.neg
 ‘Taro doesn’t know the thing Hanako bought.’
- (4) Taro-wa [Hanako-ga nani-o katta] *to] wakaranai.
 T.-TOP H.-NOM what-ACC bought Cdec know.neg
 ‘Taro doesn’t know what Hanako bought.’

(3) shows the predicate *wakaru* ‘know’ takes the DP complement. However, the predicate cannot take the declarative CP complement, and we assume the ill-formedness of the bare *wh* construction in (4) is due to the selectional failure. The contrast in (3) and (4) indicates the CP in (2) is also long-distantly selected.

(2) shows the mirror image of the schematic structure of the Japanese head-internal relative clause in (5). In (5), the predicate semantically selects the internal head inside the relative clause in a long-distance way.



Thus, both constructions involve long-distance selection.

Regarding the interpretation of the head-internal relative clause, Kuroda (1976–7 [1992]: 147) observes the event denoted by the matrix clause and the one by the relative clause must form a single event. We assume this restriction is due to long-distance selection. In the head-external relatives like English, the relative clause is incorporated into the matrix event as an adjunct of the argument of the matrix clause.

However, the head-internal relative clause cannot adjoin the head inside the relative clause. Thus, the head-internal relative clause and the matrix clause are not syntactically related. To express a coherent meaning in a sentence, the two clauses must be related pragmatically, as the Relevancy Condition requires.

If our proposal that the bare *wh* construction involves long-distance selection, the Relevancy Condition also applies to this construction. (6) shows partitive expressions are banned in the bare *wh* construction. If declarative C is replaced by interrogative C, the sentence becomes well-formed. Thus, the bare *wh* construction, but not the interrogative CP complement, requires exhaustivity.

- (6) Taro-wa nani-o katta {ka/*to} ikutuka hookoku-o sinakatta.
 T.-TOP what-ACC bought C_{int}/C_{dec} some report-ACC didn't.do
 '(He bought a lot of things, but) Taro didn't report some of what he bought.'

To formulate a single event from the two events (the buying event and the reporting event), the members of the two events must be similar. In (6), the members are different because of the partitive expression. Thus, it fails to express the consistent event and becomes infelicitous.

References

- Barker, Chris (2016) "Why relational nominals make good concealed questions," *Lingua* 182, 12–29.
- Chomsky, Noam (2015) "Problems of Projection: Extensions," In *Structures, Strategies and Beyond: Studies in Honour of Adriana Belletti*, ed. by E. di Domenico, C. Hamann and S. Matteini, 3–16, John Benjamins, Amsterdam/Philadelphia.
- Kuroda, S.-Y. (1976–7) "Pivot-Independent Relative Clauses in Japanese," *Papers in Japanese Linguistics* 5, 157–179 (Reprinted in Kuroda (1992)).
- Kuroda, S.-Y. (1992) *Japanese Syntax and Semantics: Collected Papers*, Kluwer, Dordrecht/Boston.
- Nishigauchi, Taisuke (1990) *Quantification in the Theory of Grammar*. Kluwer, Dordrecht.
- Nishigauchi, Taisuke (2020) "Structure and Derivation of the Concealed Question," *Gengo Kenkyu* 157: 37–69.
- Saito, Mamoru (2024) "Wh-phrases as genuine focus," In *Rich Descriptions and Simple Explanations in Morphosyntax and Language Acquisition*, ed. by Giuliano Bocci, Daniele Botteri, Claudia Manetti and Vincenzo Moscati, 173–199, Oxford University Press, Oxford.

Ideophones in Haitian Creole: Typology of a Word Class

Patrice Kanndèl Edouard

Despite the growing academic interest in ideophones across languages (Akita & Pardeshi, 2019; Dingemanse, 2019; Hinton, Nichols, & Ohala, 1994; Lahaussois, Marsault, & Treis, 2024; Voeltz & Kilian-Hatz, 2001), studies focusing on their typological features in Caribbean Creoles remain limited. While languages like Jamaican Patois, Martinican Creole, Guadeloupean Creole, and Haitian Creole are recognized for their rich ideophonic inventories (Rougé & Véronique, 2024; Bartens, 1999, 2013; Edouard, 2024a-b; Prou, 1999), comprehensive analyses of this lexical class in Haitian Creole remain sparse.

This study examines the typological characteristics of ideophones in Haitian Creole, with a focus on their phonological, morphological, syntactic, and semantic properties. The research aims to address the gap in the literature by providing a detailed analysis of this lexical class within the context of Haitian Creole. The central research questions guiding this study are:

1. What phonological features distinguish ideophones in Haitian Creole, and how do they contribute to the iconicity and expressivity of these words?
2. How do morphological strategies such as reduplication and partial repetition enhance the meaning and function of ideophones in Haitian Creole?
3. In what ways do ideophones exhibit syntactic flexibility in Haitian Creole, and how do they function holophrastically or in collocation?
4. How do the semantic properties of ideophones in Haitian Creole mirror patterns observed in African languages, and what role do they play in preserving oral traditions?

The methodological approach involves the collection and analysis of a corpus of 400 ideophones gathered through fieldwork and written sources. This corpus serves as the basis for identifying ideophones as a distinct lexical class with unique structural and functional attributes. The analysis is framed within theoretical perspectives from linguistic typology and pragmatics, providing a multifaceted understanding of ideophones in Haitian Creole.

Phonologically, Haitian Creole ideophones exhibit features such as reduplication, lengthening, and tonal modulation (e.g., "tchaw-tchaw" for a spongy texture or "bim" for a small impact sound). These characteristics reinforce the iconicity of ideophones and enhance their expressivity, making them a vivid and dynamic component of the language.

Morphologically, ideophones in Haitian Creole lack inflectional markers but employ strategies like reduplication and partial repetition to amplify meaning and emphasize certain aspects of sensory experiences (e.g., "wounouwounou" for murmuring or "plaf-plaf" for fluttering tails).

Syntactically, ideophones in Haitian Creole demonstrate remarkable flexibility. They can function holophrastically, used as verbs, nouns, adjective or adverbs. Semantically, Haitian Creole ideophones vividly represent sensory experiences, often mirroring patterns observed in African languages (e.g., "dap" for a quick grasping motion).

The findings from this study contribute to a broader understanding of the linguistic and cultural significance of ideophones in Haitian Creole. By highlighting the unique phonological, morphological, syntactic, and semantic properties of ideophones, this research underscores their importance in enhancing expressivity and capturing the dynamics of human experience in Haitian

Creole. Additionally, the study addresses a gap in Creole linguistics and provides insights into the ways ideophones function as cultural and linguistic markers.

References

- Akita, K., & Pardeshi, P. (Eds.). (2019). Ideophones, mimetics and expressives. *Iconicity in Language and Literature*, Vol. 16. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/ill.16>
- Bartens, A. (1999). Ideophones and sound symbolism in Atlantic Creoles. *Suomalaisen Tiedeakatemia Toimituksia. Sarja Humaniora*, 304.
- Bartens, A. (2013). Creole languages. In Bakker, P. & Matras, Y. (Eds.), *Contact languages: A comprehensive guide* (pp. 65–158). Berlin: De Gruyter Mouton.
- Dingemanse, M. (2019). Ideophone as a comparative concept. In Akita, K., & Pardeshi, P. (Eds.), *Ideophones, mimetics, and expressives* (pp. 13–33). Amsterdam: John Benjamins.
- Edouard, P. K. (2024a). *Idéophonie et catégorisation en créole haïtien : une analyse linguistique des idéophones* (Bachelor's thesis). Université d'État d'Haïti, Faculté de Linguistique Appliquée.
- Edouard, P. K. (2024b). *Haitian Creole Ideophones: Typology of a word class*. Manuscript in preparation.
- Lahaussois, A., Marsault, J., & Treis, Y. (2024). Ideophones: honing in on a descriptive and typological concept. *Linguistic Typology at the Crossroads*, 4(1).
- Prou, M. E. (1999). Haitian Creole ideophones: An exploratory analysis. *Journal of Haitian Studies*, 5/6, 96–112.
- Rougé, J.-L., & Véronique, G. D. (2024). Word classes in Romance-related Creoles. In Ledgeway, A., & Maiden, M. (Eds.), *The Oxford guide to Romance languages* (pp. 689–706). Oxford: Oxford University Press.
- Voeltz, E. F. K., & Kilian-Hatz, C. (Eds.). (2001). *Ideophones*. Amsterdam: John Benjamins.

The effect of context lengths and linear bias on how surprisal predicts syllable duration in Taiwanese

Sheng-Fu Wang
(Academia Sinica, Taiwan)

Keywords: surprisal, prosody, spontaneous speech, language modelling, linear bias

The phonetic profile of a linguistic unit is known to correlate with its contextual probability, i.e., its likelihood to occur in a certain context (Seyfarth, 2014; Van Son & Van Santen, 2005; a.o.). This observation is often tied to listener-oriented speech planning to "smooth" the amount of information throughout an utterance, i.e., more likely units can be presented with weaker phonetic cues (shorter, more neutralized, etc), and vice versa (e.g., Aylet & Turk, 2004). In phonetic research, contextual probability is often estimated using bigram models, which assume a highly localized scope of planning. Given observations on nonlocal speech planning (e.g., Fuchs et al., 2013; Krivokapic, 2012) and neural language models' capacity to make probabilistic estimates from larger contexts, this study uses syllable duration in Taiwanese as a case study to assess this locality assumption by examining to what extent surprisal estimated from longer contexts improves prediction of phonetic variability, and whether local bias plays a role.

Syllable duration was extracted from a spontaneous speech corpus with 70K words (89K syllables). A set of masked language models were pretrained in the RoBERTa architecture (Liu et al., 2019) on a Taiwanese written corpus with 4.2M words (5.96M syllables) for 64 epochs. Models differ in size (Table 1) and whether the model was trained with RoBERTa's default positional embeddings and unconstrained attention mechanism or an attention mechanism with linear bias (Press et al., 2022), which exhibits strong linear decay to models' ability to pay long-distance attention. These models then generated word surprisal in the speech corpora with different bidirectional context lengths (1, 2, 4, 8, 16, 32, 64, 128 words). The effectiveness of surprisal in predicting syllable duration was evaluated on the gain in loglikelihood of a regression model with it over the base model (with speech rate, prosodic position, lexical tones, and baseline syllable duration, lexical frequency, bigram surprisal and informativity from on simple trigram models). Following previous studies on predictability, analyses focused on content words.

Results (Figure 1) show that with the default attention mechanism, the biggest improvements over the base model with shorter contexts. Model size also does not help with the improvements on utilizing longer contexts with the default attention mechanism. On the other hand, while models with linear bias underperform with smaller model sizes, at the two larger sizes, especially with "small", they outperform the default model while also utilizing longer contexts better. Crucially, we see lower surprisal (i.e., better at language modelling) with longer contexts and with the default attention mechanism, which does not completely reflect the results in gains in log-likelihood. While these findings show a scope of speech planning is not immediately local, they also show two sides of how local bias is important: Local contexts are more crucial, and it is through biasing towards local contexts that a RoBERTa model can better utilize longer contexts. Furthermore, it also shows that better language models in surprisal measurements do not guarantee better abilities to predict behavioral measurements.

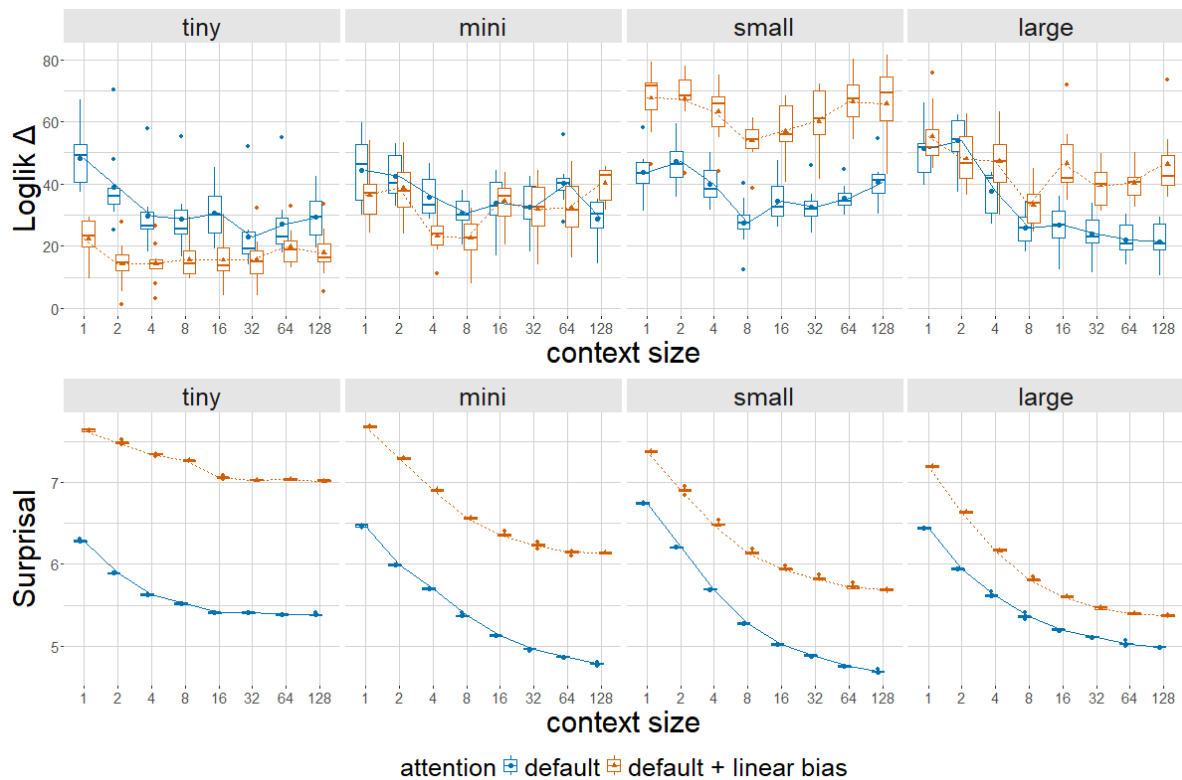


Figure 1: Statistical model's improvement in loglikelihood with surprisal added (Upper) and mean surprisal (Lower) as a function of context span size (x axis, in words) and model size (panels). Lower – surprisal

Table 1: Model sizes and corresponding parameters

Model Size	Parameters	Model Size	Parameters
Large	12 layers, 12 att. heads, 768d	Mini	4 layers, 4 att. heads, 256d
Small	8 layers, 4 att. heads, 512d	Tiny	2 layers, 2 att. heads, 128d

References

- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, 47(1), 31-56.
- Fuchs, S., Petrone, C., Krivokapić, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41(1), 29-47.
- Krivokapic, J. (2012). Prosodic planning in speech production. *Speech planning and dynamics*, 157-190.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Press, O., Smith, N., & Lewis, M. (2022) Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation. In *International Conference on Learning Representations (ICLR 2022)*.
- Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140-155.
- Van Son, R. J., & Van Santen, J. P. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech communication*, 47(1-2), 100-123.

The production of various focus types in *c'est* 'it is' clefts by L1 French children

Tess Wensink, Karen Lahousse & Cécile De Cat
(KU Leuven & University of Leeds)

Keywords: L1 acquisition, cleft sentences, Information Structure, syntax, French

INTRODUCTION. French adults produce *c'est* 'it is' clefts with various Information Structure (IS) articulations, including exhaustive focus-background, corrective focus-background (1), contrastive focus-background, new information focus-background, and all-focus (Dufter 2009, Karssenbergh and Lahousse 2018, Lahousse and Borremans 2014). While considerable research has been conducted on the use of clefts in adult French, there is limited understanding of their L1 acquisition. The present study seeks to address this gap by examining how L1 French children employ *c'est* clefts to encode these distinct focus types and how this relates to their syntactic development.

- 1) *Non c' est [moi]_{corrective focus} [qui commence]_{background}.*
no it is me who start
'No, it's me who is starting.'

(Palasis, Lisa, 5;6)

METHOD. The analysis draws on 1,087 *c'est* clefts produced by L1 French children aged 1-6, extracted from the Palasis (2009), York (De Cat and Plunkett 2002), and Lyon (Demuth and Tremblay 2008) corpora.

RESULTS. Our findings show that children's *c'est* clefts reflect adult IS patterns: corrective focus-background (28%), contrastive focus-background (37%), new information focus-background (17%) and all-focus clefts (5%). Importantly, the prevalence of focus types varies by age, with an increasing proportion of all-focus clefts, indicating that this IS type takes the longest to develop. Furthermore, younger children (up to 30 months) predominantly produce corrective focus clefts (39%), whereas older children favor contrastive focus clefts (39%). Since corrective focus exhibits a higher degree of contrastivity (Cruschina 2021), this suggests that younger children produce more clefts with higher contrastivity.

Children also produce various syntactic forms of clefts involving *c'est* X (see also (Lahousse and Jourdain 2023)):

1. without CRC (2),
2. with a CRC-attempt [without complementizer, without resumptive pronoun] (non-adultlike) (3),
3. with a CRC-attempt [without complementizer, with resumptive pronoun] (adultlike in informal French) (4),
4. with a full CRC [with complementizer, with resumptive pronoun] (adultlike in some varieties) (5),
5. with a full CRC [with complementizer, without resumptive pronoun] (1) (6), see Table 1.

Notably, even the syntactically non-adultlike clefts convey adultlike IS, supporting the view that IS precedes syntax in acquisition (Lambrecht 1994).

Surprisingly, children continue producing syntactically less complex CRC-attempts (10% from all clefts produced) even after mastering full CRCs. The CRC-attempts produced by two children older than 3 years and 4 months who have previously produced full CRCs, predominantly feature an all-focus articulation (57%). This reinforces the idea that all-focus clefts pose particular challenges.

CONCLUSION. Overall, these results indicate an early acquisition of various focus types and an early development of the interaction between IS and syntax.

Table 1. Syntactic types of *c'est*-clefts

Reduced clefts [- CRC]	Clefts with a CRC-attempt [- C]		Full clefts [+ C]	
	[- resumptive]	[+ resumptive]	[+ resumptive]	[- resumptive]
(2) <i>C'est moi.</i> it's me	(3) <i>C'est moi</i> it's me <i>lis.</i> read	(4) <i>C'est moi_i je_i</i> it's me I <i>lis.</i> read	(5) <i>C'est moi_i que</i> it's me that <i>je_i lis.</i> I read	(6) <i>C'est moi_i qui_i</i> it's me that <i>lis.</i> read
adultlike	non-adultlike	adultlike in informal French	adultlike in some varieties	adultlike

References

- Cruschina, Silvio (2021), The greater the contrast, the greater the potential: On the effects of focus in syntax, *Glossa: a journal of general linguistics* 6(1).
- De Cat, Cécile, and Plunkett, Bernadette (2002), QU'est ce qu'i (l) dit, celui+ Là?: notes méthodologiques sur la transcription d'un corpus francophone, *Romanistische Korpuslinguistik: Korpora und gesprochene Sprache/Romance Corpus Linguistics: Corpora and Spoken Language*, Narr.
- Demuth, Katherine, and Tremblay, Annie (2008), Prosodically-conditioned variability in children's production of French determiners, *Journal of Child Language* 35, 99–127.
- Dufter, Andreas (2009), Clefting and discourse organization: Comparing Germanic and Romance, in A. Dufter, and D. Jacob (eds), (2009), *Focus and background in Romance languages*, Amsterdam / Philadelphia: John Benjamins Publishing Company, 83–121.
- Karssenber, Lena, and Lahousse, Karen (2018), The information structure of French *il y a* clefts and *c'est* clefts: A corpus-based analysis, *Linguistics* 56(3), 513–548.
- Lahousse, Karen, and Borremans, Marijke (2014), The distribution of functional-pragmatic types of clefts in adverbial clauses, *Linguistics* 52(3), 793–836.
- Lahousse, Karen, and Jourdain, Morgane (2023), The emergence and early development of *c'est* 'it is' clefts in French L1, in C. Bonan, and A. Ledgeway (eds), (2023), *It-Clefts: Empirical and Theoretical Surveys and Advances*, Berlin: Walter de Gruyter, 135–156.
- Lambrecht, Knud (1994), *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*, Cambridge: Cambridge University Press.
- Palasis, Katerina (2009), *Syntaxe générative et acquisition: Le sujet dans le développement du système linguistique du jeune enfant* [Doctoral Dissertation, Université de Nice-Sophia Antipolis].

Why can some prefixes in Japanese change category?: Structural borrowing, constructionalization and an economy principle in word formation

Yupeng Yue & Yoshiki Ogawa
(Tohoku University)

Keywords: <Negative Prefix, Adjectival Noun, Structural Borrowing, Sino-Japanese, Economy>

According to Williams' (1981) Right-hand Head Rule on word formation, suffixes but not prefixes can determine the category of a derived word. In Japanese too, the prefix/suffix asymmetry holds for almost all the native affixes. However, the Japanese lexicon includes many affixes borrowed from Chinese, called *here as* Sino-Japanese (SJ) affixes. Among the SJ suffixes are *-teki/的* (adjectivizer) and *-sei/性* (nominalizer), and among the SJ prefixes are the negative prefixes *hu-/不*, *hi-/非*, and *mu-/無*.

A peculiar property of the SJ negative prefixes is that they can adjectivize the stems:

- (1) a. dootoku(-no/*-na) 'morality_N-Gen/*AGR_{AN}'
→hu-dootoku-na/*-no 'immoral-AGR_{AN}/*Gen'
b. kooritu(-no/*-na) 'efficiency_N-Gen/*AGR_{AN}'
→hi-kooritu-na/*-no 'inefficient-AGR_{AN}/*Gen'

In (1a,b), even when the stems to which *hu-* or *hi-* attach are nouns (N), the derived words are not N but adjectival nouns (AN), a kind of adjective, as shown by the *-na/-no* asymmetry (Yumoto et al. 2023).

Another peculiar property of the SJ negative prefixes is their compatibility with the AN-forming suffix *-teki/的* and the incompatibility of *hi-* with AN. *N-na*, *N-teki-na*, and *AN-na* are all possible, whereas *AN-teki-na* is impossible (e.g. (2a,b)). *Hu-AN-na* and *hu-/hi-N-na* are possible, whereas *hi-AN-na* is almost impossible. There are few lexicalized cases (e.g. (2c),(3b)).

- (2) a. kenkoo(-teki)-na 'health_N-adjectivizer-AGR_{AN}' (healthy)
b. kirei(*-teki)-na 'beautiful_{AN}-adjectivizer-AGR_{AN}' (beautiful)
c. enkyoku(-teki)-na (roundabout_{AN}-adjectivizer-AGR_{AN}) (roundabout)
(3) a. hu-kanou-na 'Neg-possible_{AN}-AGR_{AN}' (impossible)
b. ??hi-genmitu-na 'Neg-strict_{AN}-AGR_{AN}' (non-strict)

There are further restrictions when the negative prefixes co-occur with the adjectivizer: First, *hi-N-teki-na* is possible but *hu-N-teki-na* is not, although *hu-N-na* and *hi-N-na* are both possible (= (4a)). Second, *hu-N-teki-na*, *hu-AN-teki-na* and *hi-AN-teki-na* are (almost) impossible (= (4a), (5a,b)), although *hi-N-teki-na*, *hu-N-na*, and *hi-N-na* are fully possible (= (1a,b), (4c)). There are few lexicalized cases (e.g. (4b)).

- (4) a. *hu-kisoku(*-teki)-na* ‘Neg-rule_N-Adjectivizer-AGR_{AN}’ (irregular)
- b. *hu-ka-gyaku-teki-na* ‘Neg-can-reverse-Adjectivizer-AGR_{AN} (irreversible)’
- c. *hi-kooritu(-teki)-na* ‘Neg-efficiency_N-Adjectivizer-AGR_{AN}’ (inefficient)
- (5) a. **hu-tokusyu-teki-na* ‘Neg-special_{AN}-Adjectivizer-AGR_{AN}’ (cf. *vtokusyu(-teki)-na*)
- b. **hi-genmitu-teki-na* ‘Neg-strict_{AN}-Adjectivizer-AGR_{AN}’ (cf. ??*hi-genmitu-na*)

In this presentation, we will first claim that synchronically, *hu-* and *-teki* are Level I affixes, whereas *hi-* is either a Level I or II affix (Allen 1978). Second, we assume two different principles of economy of derivation (cf. Kageyama 1982, Chomsky 1992). Third, we claim that the adjectivhood of *hu-N* and *hi-N* is determined by the head-initial nature of the Negative Phrase in Chinese that has undergone structural borrowing into Japanese after which *hu-/hi-* were grammaticalized to prefixes (cf. Thomason and Kaufman 1988:74, Traugott 2015). Fourth, the borrowed structure [_{NegP} *hu/hi* (*Neg*) NP/AP] was renumerated as the root (ν AN) and then adjectivized by a zero-adjectivizer, as in [_{ANP} [_{ν AN} *hu/hi-N*]- ϕ _{AN}]] (Harley 2009). On the basis of these claims and assumptions, we will explain the ill-formedness of (5a,b) and the unacceptability of (2b) with *-teki* by an economy principle prohibiting adjectivization of an already adjectival category and another economy principle that prohibits overt adjectivization when zero-adjectivization is possible. These economy principles will also explain why the two Level I affixes *hu-* and *-teki* are incompatible, whereas the Level II affix *hi-* and the Level I affix *-teki* are compatible ((4a) vs. (4c)). Lexical borrowings from Chinese (e.g. (2c), (4b)) are exempt from the economy principles.

Acknowledgments:

This work was financially supported by JST SPRING, Grant Number JPMJSP2114 for the first author, and by the Grant-in-Aid for Scientific Research (C) from JSPS (Grant Number: 24K03939) and by the Grant-in-Aid for the Frontier Research Duo (FRiD) of Tohoku University, titled ‘Understanding the brain mechanism of language computation from analyzing the communication signals of birds’ for the second author.

References

- Allen, Margaret (1978), *Morphological Investigations*, PhD dissertation, University of Connecticut, Storrs.
- Chomsky, Noam (1992), Some Notes on the Economy of Derivation and Representation, in R. Friedin (ed), (1992), *Principles and Parameters in Comparative Grammar*, Cambridge, MA: MIT Press, 53-82.
- Harley, Heidi (2009), Chapter 7: compounding in distributed morphology, in By R. Lieber, and P. Štekauer (eds.), (2009), *The Oxford Handbook of Compounding*. Oxford University Press, New York, pp. 129-144.
- Kageyama, Taro (1982), Word formation in Japanese, *Lingua* 57, 215-258.
- Thomason, Sarah Grey and Terrene Kaufman (1988), *Language Contact, Creolization, and Genetics Linguistics*, Berkeley: University of California Press.
- Traugott, Elizabeth Closs (2015), Toward a coherent account of grammatical constructionalization, in J. Barðdal, E. Smirnova, L. Sommerer, and S. Gildea (eds), (2015), *Diachronic Construction Grammar*, Amsterdam: John Benjamins, 51-79.
- Williams, Edwin (1981), On the notions 'lexically related' and 'head of a word', *Linguistic Inquiry* 12, 245-274.
- Yumoto, Yoko, Yoko Sugioka and Takane Ito (2023) *Go-no Bunpou-e-no Izanai* [Invitation to the Grammar of Words], Tokyo: Hituzi Syobo.

Used Corpus

Balanced Corpus of Contemporary Written Japanese (BCCWJ), National Institute for Japanese Language and Linguistics.

<https://clrd.ninjal.ac.jp/bccwj/>

Corpus of Historical Japanese (CHJ) , National Institute for Japanese Language and Linguistics.

<https://clrd.ninjal.ac.jp/chj/>