

# **BOOK OF ABSTRACTS**

## **WORKSHOP TALKS**

**58th Annual Meeting of the *Societas Linguistica Europaea***

**26 – 29 August 2025**

**Université Bordeaux Montaigne**

**Anna Kisiel**

## **WS1 Compounding in word-formation paradigms**

**Nabil Hathout & Fabio Montermini**

# Paradigmatic competition between adjectival derivation and compounding

Amanda Pounder  
(University of Calgary)

Keywords: paradigm, derivation, compounding, adjectives, competition

How to choose between German *silberner Ring* (Adj N construction) and *Silberring* (compound), both meaning 'silver ring'? In (1), the two are clearly interchangeable:

(1) **Silberner Ring** mit Kringeln geschmiedet ... Erleben Sie die zeitlose Eleganz unseres handgefertigten **Silberrings** ..."  
<https://www.goldschmiede-von-gruenberg.com/freude/960-Ring-Kringel.html>, last accessed 15/1/2025

'**silver ring** forged with curls... experience the timeless elegance of our handmade **silver-ring**...'

The fact that we have both compounding and affixal derivation available and competing with each other points to a model of grammar which includes a word-formation paradigm, and indeed, one in which any morphological process a language may have at its disposal, concatenative or non-concatenative, may figure (cf. e.g. Pounder 2008, Winters 2017, Kunduract 2018). We will focus here on the competition between German compounding and suffixal derivation in pairings with the semantic function FROM('X'), where a noun X denoting a substance ('silver') corresponds to the source of some entity (Pounder 2000). Synchronically, we observe throughout modern history that both processes are available, although at all times derivation is less productive than compounding in this domain. We can model this by invoking the dynamic morphological paradigm, which provides the set of possible paths leading from a lexeme base of a given type to grammatical word-formation products. There will be as many paths for compounding operations as there are combinations of lexeme base types and semantic functions in compounds. Where different paths correspond to the same semantic function, as in the case of *silbern/Silber-*, we speak of "word-formation synonymy". This is where we see the potential for competition between operations and between processes. In this case, the relevant operations will have differing productivity indices.

Diachronically, we observe changes in the relations between derivation and compounding in the relevant German denominal subparadigm, for example an increasing preference for compounding with bases denoting substances and the semantic function FROM('X'). Where in the system are these changes reflected? As long as no other properties of the operations are changing, we will merely see adjustments of the productivity indices each manifests. The consequences of the shift will be most obvious in the lexicon, which we assume contains at least fragments of paradigm structure ("microparadigms") for any lexicalized path containing non-predictable information. In modern German, some derived adjectives of this type continue to be used (e.g. *golden*, *wächsern* 'waxen', *ledern* 'leathern' etc.) and will thus be represented in the lexical

paradigms of their bases. Their place in individual speakers' lexical paradigms may be tenuous, and low-frequency items are likely to disappear over time. Given the existence of the high-productivity compounding operations in the morphological component, the lexical paradigms need not explicitly contain the compound competitor; a compound will be created as a default. However, this does not prevent a speaker from choosing the derived form in a given instance. As we see in (1), there may be pragmatic motivations driving the choice, such as desire for stylistic variation.

An inclusive model of word-formation paradigm with dynamic macrostructure and lexical counterparts allows us to describe relations between processes including compounding and changes within these relations over time.

## References

- Kunduracı, Aysun (2018), The paradigmatic aspect of compounding and derivation, *Journal of Linguistics* 55, 563-609.
- Pounder, Amanda (2000), *Processes and Paradigms in Word-Formation Morphology: The Development of the Denominal Adjective System in German*, Berlin: Mouton.
- Pounder, Amanda. (2008), Adverb-marking patterns in Earlier Modern English Coordinate Constructions, in M. Dossena, R. Drury, and M. Gotti (eds), (2008), *English Historical Linguistics 2006, Series: Current Issues in Linguistic Theory*, Amsterdam: Benjamins 183-201..
- Winters, Svitlana (2017), *Lexical blending: System or sport?* University of Calgary dissertation.



## Blends as a special category of compounds in word-formation paradigms

Camiel Hamans  
(Adam Mickiewicz University, Poznań)  
(University of Amsterdam)

This contribution focusses on blends as a special form of compounds and highlights their position in paradigmatic word formation. The research question here is what the role of productivity is for successful paradigmatic word formation. The focus is mainly on blend formation and compounding in Germanic languages such as English, German and Dutch.

Before addressing this question, however, it must be determined in what way blends are a special category of compounds, and the category of blends must be demarcated.

Lexical blends can be defined as the result of a word-formation process that combines two or more source words into a single form, losing some phonological material in the process (Moreton et al 2017: 349). Word forms that are the result of a blending process are characterized by several properties, the most essential of which is the fact that the right-hand part is the formal, and often semantic, head of the new word form (Hamans 2021c):

- |                           |                             |                            |
|---------------------------|-----------------------------|----------------------------|
| (1) simultaneous (Adj)    | + broadcast (Noun)          | → simulcast (Noun)         |
| (2) die Kur 'cure' (F)    | + der Urlaub 'vacation' (M) | → der Kurlaub (M) [German] |
| (3) die Daten 'data' (pl) | + die Kartei 'file' (sg)    | → die Datei (sg) [German]  |

The head also determines the syllabic and prosodic structure of blends.

A difference with standard compounds is that blends do not follow the compound stress rule, which says that compounds have stress on the first syllable. This implies that one must distinguish between blends as the ones presented in (1)-(3) in which the leftmost part of the first source word is combined with the rightmost part of the second source words and forms such as:

- |  |             |          |
|--|-------------|----------|
| (4) situation                            | + comedy    | → sitcom |
| (5) California                           | + Texas     | → Caltex |
| (6) Semtin (Czech village of production) | + explosive | → Semtex |

In these forms the leftmost parts of both source words are combined. The stress pattern is that of compounds and therefore one better considers the word formation process operating here as a form of compounding of clipped forms. Therefore, they are called clipped compounds (Mattiello 2013, Hamans 2021a) or stub compounds (Spencer 1998:128-129, Hamans 2021b 199-210).

According to Fradin (2000:52), blends are unique and do not allow any analogically productive word formation and therefore are 'antithetical to paradigms.' 'Blends are type hapaxes. Unlike derived or compound units, they cannot form series. Each one is a (lexeme) type which is the only one to instantiate the morphological pattern it belongs to (Fradin 2015:389). Hamans (2011 and 2021b: 210-213) argues against this claim of uniqueness. Successful blends just form series:

- |                      |  |
|----------------------|--|
| (7) motel < motorist | + hotel  |
| boatel < boat        | + hotel  |
| rotel < rollendes    | + hotel (German) 'rolling hotel'                                 |
| zotel < ziekenhuis   | + hotel (Dutch) 'combination of hospital (ziekenhuis) and hotel' |

potel	< Polen	+ hotel (Dutch)	'hotel for Polish foreign workers)
(8) froogle	< frugal	+ google	
smoogle	< smart	+ google	
foogle	< false	+ google	
broogle	< brain	+ google	
poogle	< poop	+ google	(taking a poop when using google)
(9) Spanglish	< Spanish	+ English	
Dunglish	< Dutch	+ English	
Hunglish	< Hungarian	+ English	
Po(l)glish	< Polish	+ English	
Frenghish	< French	+ English	

Here, one must distinguish between two processes: the first being lexical blending (resulting in *motel* and *froogle*) and subsequent paradigmatic word formation. Paradigmatic extension is essential here for successful, productive word formation. This paradigmatic process resembles affixal derivation, however, it is better not to lump them together, since the rightmost part remains the head of the new word form in all respects. But it does make clear how a special category such as blends marks the transition area between derivation and compounding. In addition, truncation of the left wordform remains compulsory. However, this process makes clear how a special category such as blends marks the transition area between derivation and compounding.

## References

- [1] Fradin, Bernard (2000) Combining Forms, blends and related phenomena. Ursula Doleschal and Anna M. Thornton (eds.) *Exagrammatical and Marginal Morphology*. Munich: Lincom: 11-59.
- [2] Fradin, Bernard (2015). Blending. Peter O Müller, Ingeborg Ohnheiser, Susan Olsen and Franz Rainer (eds.). *Handbook on Word-Formation*. Berlin/New York: De Gruyter Mouton: 386-412.
- [3] Hamans, Camiel (2011). About Uniqueness and Productivity of Blends. In: *Folia Linguistica et Litteraria: Časopis za nauku jeziku i književnosti* 5: 69-79.
- [4] Hamans, Camiel (2021a). The difference between Blends and Clipped Compounds. Gabi Danon (ed.). *Proceedings of IATL 34-35. Papers from the 34<sup>th</sup> and 35<sup>th</sup> Conference of the Israel Association for Theoretical Linguistics. MIT Working Papers in Linguistics XXX*: 89-101.
- [5] Hamans, Camiel (2021b). *Borderline Cases in Morphology. A study in language change*. Lewedorp: Van Kemenade. PhD thesis University of Amsterdam.
- [6] Hamans, Camiel (2021c). Blends: an intermediate category at the crossroads of morphology and phonology. *Yearbook of the Poznań Linguistic Meeting*, 7(1): 99-128.
- [7] Mattiello, Elisa (2013) *Extra-Grammatical Morphology in English. Abbreviations, Blends, Reduplicatives and related Phenomena*. Berlin/Boston: De Gruyter Mouton.
- [8] Moreton, Elliott, Jennifer L. Smith, Katya Pertsova, Rachel Broad and Brandon Prickett (2017). Emergent Positional Privilege in Novel English Blends. *Language* 93,2: 347-380.
- [9] Spencer, Andrew (1998). Morphophonological Operations. Andrew Spencer and Arnold. M. Zwicky (eds.). *The Handbook of Morphology*. Oxford: Blackwell: 123-143.

# Transitional word-formation paradigms: The case of parasynthetic adjectival compounds

Elisa Mattiello  
(University of Pisa)

Keywords: parasynthetic adjectival compound, paradigm, analogy, body part, compound family

This paper investigates the concept of word-formation paradigms, focusing specifically on parasynthetic adjectival compounds (PACs), such as *blue-eyed* and *baby-faced*. A word-formation paradigm refers to a network of words that share formal and semantic patterns, where analogy plays a key role in the creation of similar new words. Paradigms, therefore, rely on (i) form and meaning similarity and (ii) the predictability of new words using a model (Bauer 1997; Antoniova and Štekauer 2015; Bonami and Strnadová 2019; Hathout and Namer 2019). Recent studies have extended the application of paradigms beyond derivation to compounding, emphasising the predictability and regularity of lexical networks (Fernández-Domínguez et al. 2020; Ruz et al. 2022). PACs that combine a body part with an adjectival suffix to make a complex second constituent (e.g., *eye* + *-ed*) straddle the boundaries between compounding and derivation, creating distinct paradigms based on both formal and semantic criteria.

The study is corpus-driven and employs the *Corpus of Contemporary American English* (COCA) to identify four of the most frequent complex elements in PACs: *-eyed*, *-faced*, *-headed*, and *-minded*. These elements were selected from 23 different body parts based on their high frequency in the data, enabling a systematic analysis of the patterns they generate. The qualitative analysis follows a paradigmatic approach, examining how all four elements create lexical networks based on both local and extended analogy. Local analogy occurs when a single exemplar serves as a model for new formations (e.g., *broad-minded* after *narrow-minded*), while extended analogy is evident when multiple compounds share at least one constituent leading to compound families (e.g., *angry-eyed*, *fierce-eyed*, *lazy-eyed*, *merry-eyed*, or *big-eyed*, *big-faced*, *big-headed*, *big-minded*). In PACs, the first constituent (which can be an adjective, noun, or, more rarely, an adverb, numeral or name) is crucial in determining the compound's meaning, which may be literal (e.g., *small-eyed*) or figurative (e.g., *demon-eyed*), and often involves connotative value such as pejoration (e.g., *empty-headed*) or amelioration (e.g., *doll-faced*).

The paper explores how these compounds form an overarching paradigm  $[[X_{A/N} \text{ [BODY PART]}] -ed]_A$ , with all four complex constituents generating distinct but analogous paradigmatic structures.

The research questions that the paper intends to address include:

- (i) Should PACs be integrated into word-formation paradigms? If so, would they belong to the same paradigms as derivatives?
- (ii) How does the syntagmatic combination of compounding with other morphological phenomena (e.g., *-ed* suffixation) interact with their integration into derivational paradigms?
- (iii) What is the role played by analogy in the creation, modelling and prediction of PACs?

The findings contribute to a broader understanding of how lexical patterns evolve and how new compounds are formed. The study highlights the value of paradigms in understanding the structure and productivity of PACs involving a body part, offering insights into the interplay of syntax, semantics, and analogy in compounding. It also extends the paradigm approach from derivational morphology to compounding,

shedding light on how new compounds may develop into compound families and showing regularities in terms of syntactic patterns and semantic realisations.

## References

- Antoniova, Vesna and Pavol Štekauer (2015), Derivational paradigms within selected conceptual fields - Contrastive research, *Facta Universitatis, Series: Linguistics and Literature* 13(2), 61–75.
- Bauer, Laurie (1997), Derivational paradigms, in G. Booij, and J. van Marle (eds), (1997), *Yearbook of Morphology 1996*, Dordrecht: Kluwer, 243–256.
- Bonami, Olivier and Jana Strnadová (2019), Paradigm structure and predictability in derivational morphology, *Morphology* 29, 167–197. <https://doi.org/10.1007/s11525-018-9322-6>
- Fernández-Domínguez, Jesús, Alexandra Bagasheva and Cristina Lara-Clares (eds), (2020), *Paradigmatic Relations in Word Formation*, Leiden / Boston: Brill.
- Hathout, Nabil and Fiammetta Namer (2019), Paradigms in word formation: What are we up to?, *Morphology* 29, 153–165. <https://doi.org/10.1007/s11525-019-09344-3>
- Ruz, Alba E., Cristina Fernández-Alcaina and Cristina Lara-Clares (eds), (2022), *Paradigms in Word Formation: Theory and Applications*, Amsterdam / Philadelphia: Benjamins.

# Modelling the paradigm of partiality/totality adjectival compounds in Italian

Flavio Pisciotta  
(University of Salerno)

Keywords: totality modifiers; approximation; evaluative morphology; adjectival compounds; word-formation paradigms.

The present study investigates the relationship between *mezzo* Adj ‘half Adj’ and *tutto* Adj ‘all Adj’, two Italian compound-like evaluative constructions including partial/total quantifiers. Such constructions are employed respectively as attenuation/approximation, and intensification strategies (Grandi 2017; Piuino 2023; Benigni 2023). Formally, they have been defined in different ways: as analytic constructions with a “high degree of internal cohesion” (*tutto*, Grandi 2017: 69), as well as affixoids (*mezzo*, Masini & Micheli 2020), but ultimately, their behaviour is quite consistent with the criteria for compoundhood (cf. Grandi 2006).

Previous research has always studied totality/partiality markers separately, even though they seem to be transparently related. In fact, when combined with adjectives, *mezzo* and *tutto* intuitively give rise to a number of semantically opposed doublets, in which *mezzo* reduces the degree of the property described by the adjective, while *tutto* can both intensify such property (1a) or mark that some boundary on the property scale has been fully reached (1b):

- (1)    a.    ***tutto bagnato***       ↔    ***mezzo bagnato***  
              ‘completely wet’                ‘partially wet’
- b.    ***tutto pieno***       ↔    ***mezzo pieno***  
              ‘completely full’                ‘partially full’

We hypothesize that this situation could be fruitfully modelled as a paradigm: word-formation paradigms are primarily structured around semantic contrasts (Štekauer 2014; Bonami & Strnadová 2019) and this is the case with completeness and halfness in the evaluative domain (Masini 2024). Moreover, the transparent relation between *tutto* and *mezzo* suggests a good degree of interpredictability (Ackerman et al. 2009; Bonami & Strnadová 2019; Sanacore et al. 2024) and mutual motivation between the elements of the two series.

Nonetheless, it is not clear if their output regularly results in semantic opposition: at least for *mezzo*, we find a variety of different outputs, encompassing quantitative and qualitative meanings (2a), while sometimes the only interpretation possible when *tutto* combined with the same filler is a more “compositional” (quantity-related) one (2b), at least in the case of adjectives denoting physical properties:

- (2)    a.    *un ruscello mezzo ghiacciato* ‘a partially frozen stream’ but also ‘a frozen-ish stream’  
          b.    *un ruscello tutto ghiacciato* ‘a completely frozen stream’

As a matter of fact, semantic differences could be expected, as *tutto* and *mezzo* appear to be at different stages of decategorialization from their original quantifier status. For instance, *mezzo*’s

formal behaviour seems to be closer to prefixoids, as it does not necessarily agree with the adjectival filler (3a), while *tutto* obligatorily agrees with it (3b):

(3)	a.	dei	calzini	<i>mezzi</i>	/	<i>mezzo</i>	<i>bucati</i>
		some.M.PL	sock.M.PL	half.M.PL		half.M.SG	holed.M.PL
		'some half-holed socks'					
	b.	<i>dei</i>	<i>calzini</i>	<i>tutti</i>	/	<i>*tutto</i>	<i>bucati</i>
		some.M.PL	sock.M.PL	all.M.PL		all.M.SG	holed.M.PL
		'some completely holed socks'					

Thus, we aim to understand to what extent totality and partiality constructions in Italian can be modeled as a paradigm, by assessing:

- i) how often *tutto* and *mezzo* combine with the same adjectives;
- ii) whether this always results in a semantic contrast, or if the semantic opposition is restricted to some specific sub-meaning(s) of the two constructions.

We do this by contrasting *tutto* Adj with an existing dataset of *mezzo* Adj constructions (n = 338) from the written Italian corpus CORIS (Rossini Favretti et al. 2002) (Pisciotta 2024). We extracted all the bases found with *mezzo* Adj (n = 141) and combined them with *tutto*, thus obtaining 1426 occurrences of *tutto* Adj. Currently, we are annotating them by presence of evaluative semantics, and by their adjectival base. In a second stage, we will compare the occurrences of the constructions combining with the same adjective to assess if they show some form of semantic contrast, distinguishing if the adjective is modified quantitatively or qualitatively.

Hopefully, our analysis will highlight the advantages of studying analytic constructions by considering paradigmatic relations among them for a deeper understanding of their behaviour in terms of productivity and semantic interpretation.

## References

- Ackerman, Farrell, James P. Blevins & Robert Malouf (2009), Parts and Wholes: Implicative Patterns in Complex Morphological Systems, in J. P. Blevins & J. Blevins (eds.), *Analogy in grammar: Form and acquisition*, Oxford: Oxford University Press, 54–81.
- Benigni, Valentina (2023), *È stato un mezzo disastro...* Fenomeni di convergenza tra approssimazione e intensificazione: uno studio contrastivo su italiano e russo, *Italica Wratislaviensia*, 14(1), 9–34.
- Bonami, Olivier & Jana Strnadová (2019), Paradigm structure and predictability in derivational morphology, *Morphology*, 29(2), 167–197.
- Grandi, Nicola (2006), Considerazioni sulla definizione e la classificazione dei composti, *Annali Online Lettere – Ferrara*, 1(1), 31–52.
- Grandi, Nicola (2017), Intensification processes in Italian: A survey, in M. Napoli & M. Ravetto (eds.), *Exploring Intensification: Synchronic, diachronic and cross-linguistic perspectives*, Amsterdam/Philadelphia: John Benjamins, 55–78.
- Masini, Francesca (2024), La prototipicità è una funzione semantico-valutativa? Presented at *LVII Congresso internazionale della Società di Linguistica Italiana*, Catania (Italy), 19–21 September 2024.

- Masini, Francesca & M. Silvia Micheli. (2020), The morphological expression of approximation: The emerging *simil*-construction in Italian, *Word Structure*, 13(3), 371–402.
- Pisciotta, Flavio (2024), Evaluative (half-)prefixes in Italian: the expression of approximation through morphemes of half quantity. Presented at the workshop *Unlocking evaluative morphology: conceptual and methodological challenges*, Berlin (Germany), 13-15 March 2024.
- Piunno, Valentina (2023), Intensifying constructions in Italian: Types, function and representation, in J.-P. Colson (ed.), *Phraseology, constructions and translation: corpus-based, computational and cultural aspects*, Louvain-la-Neuve: Presses Universitaires de Louvain, 399–408.
- Rossini Favretti, Rema, Fabio Tamburini & Cristiana De Santis (2002), CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model, in A. Wilson, P. Rayson & T. McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, Munich: LINCOM-Europa, 27–38.
- Sanacore, Daniele, Nabil Hathout & Fiammetta Namer (2024), A scenario-based approach to predictability in derivation, *Lingue e linguaggio*, 23(2), 337–359.
- Štekauer, Pavol (2014), Derivational paradigms, in R. Lieber & P. Štekauer (eds.), *The Oxford handbook of Derivational Morphology*, Oxford: Oxford University Press, 354–369.

# On the Explanatory Potential of Paradigmatic Relations in Compounding

Jan Radimský

(University of South Bohemia in České Budějovice)

**Keywords:** Paradigmatic relations; Compounding; Construction Morphology; Relational Morphology; Italian

This paper explores the explanatory potential of paradigmatic relations in the domain of compounding, understood as systematic co-variation between form and meaning across various linguistic items (Boyé & Schalchli, 2016; Bauer, 2019). Within the theoretical frameworks of Construction Morphology (Booij, 2010, 2016) and Relational Morphology (Jackendoff & Audring, 2020), paradigmatic relations (denoted as "≈") are conceptualized as links between constructions at different levels of abstraction. These include fully specified constructions (e.g. Italian *città modello* ≈ *lettore modello*), semi-schematic constructions (e.g., *N-modello* ≈ *N-simbolo*), and fully schematic templates (e.g., *N-N* ≈ *N-P-N*).

While paradigmatic relations between sister constructions often give rise to higher-order mother constructions (e.g., the semi-schematic [*N-modello*] for *città modello* ≈ *lettore modello*), there are cases where such relations reflect competition rather than unification, as observed in patterns like [*N-N*] ≈ [*N-P-N*]. This paper argues that paradigmatic relations serve as the foundational basis for constructions but extend beyond them to offer a broader explanatory framework for linguistic analysis.

The first section of the paper categorizes paradigmatic relations in compounding following Bauer (2019, 2022), distinguishing between paradigms of forms, paradigms of relationships, and hybrid paradigms. Empirical data from native and neoclassical compounding in Italian and French provide the primary basis for this analysis.

The second section illustrates the explanatory role of paradigmatic relations in specific linguistic phenomena:

1. **Semantic determination of compound elements:** The meaning of certain compound elements is defined exclusively by paradigms of forms, as in bound elements in neoclassical compounds or Italian native compounds like [*N-principe*].
2. **Determination of the relationship between elements:** It seems quite rarely the case that the nature of the relationship between compound elements is defined by paradigms of relationships alone (e.g., "made of" in *steel wire* ≈ *plastic bag*). It is determined more often jointly with a lexically fixed component (e.g., "homemade" in French [*N-maison*] constructions) and in certain contexts, these relationships remain determined only very broadly (e.g., French [*rayon-N*]).
3. **Productivity conditioned by paradigms of forms:** The productivity of semi-schematic constructions, as evidenced in the diachronic evolution of Italian **NN** compounds, highlights the critical role of pure paradigms of forms (i.e. those which do not involve specific semantic relationship between components) in shaping linguistic innovation.
4. **Development of higher-order constructions:** The emergence and evolution of new abstract templates rely on the interplay of paradigms of forms. For instance, the diachronic development of verbal-nexus NN compounds in Italian demonstrates this process.
5. **Linking elements driven by paradigmatic relations:** The presence and variation of linking elements in compounds (e.g., [*N-C*] constructions with nouns and neoclassical confixes) can be effectively explained through paradigmatic relations.

Further, paradigmatic relations illuminate phenomena such as inflectional behavior and contextual preferences of specific compound families (e.g., the Italian [*N-fantasma*] family, commonly appearing in artistic titles), but also competition between constructions.



By analysing these aspects, the paper demonstrates how paradigmatic relations enrich our understanding of the structure, variation, and diachronic trajectories of compounding patterns, underscoring their central role in morphological theory.

### Acknowledgements

This research was supported by the research project No. 25-15350S, entitled How do word-formation patterns emerge? An empirical diachronic analysis of Italian N+N compounds, funded by the Grant Agency of the Czech Republic (GAČR) and conducted at the Faculty of Arts, University of South Bohemia in České Budějovice.

### References

- Bauer, L. (2019). Notions of paradigm and their value in word-formation. *Word Structure*, 12(2), 153–175. <https://doi.org/10.3366/word.2019.0144>
- Bauer, L. (2022). Chapter 3. Interlocking paradigms in English compounds. In A. E. Ruz, C. Fernández-Alcaina, & C. Lara-Clares (Eds.), *Studies in Language Companion Series* (Roč. 225, p. 59–68). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.225.03bau>
- Bonami, O. & Strnadová, J. (2018). Paradigm structure and predictability in derivational morphology. *Morphology*, 1–31. Springer. DOI: [10.1007/s11525-018-9322-6](https://doi.org/10.1007/s11525-018-9322-6)
- Booij, G. E. (2010). *Construction morphology*. Oxford: Oxford University Press.
- Booij, G. E. (2016). Construction morphology. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 424–448.). Cambridge: Cambridge University Press.
- Boyé, G. & Schalchli G. (2016). The status of paradigms. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 206–234). Cambridge: Cambridge University Press.
- Goldberg, A. E. (2019) *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton, New Jersey: Princeton University Press.
- Hilpert, M. (2015) From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26/1, 113–147. doi:[10.1515/cog-2014-0001](https://doi.org/10.1515/cog-2014-0001).
- Jackendoff, R. and J. Audring (2020) *The texture of the lexicon: relational morphology and the parallel architecture*. First edition. Oxford / New York: Oxford University Press.
- Radimský, J. (2013), Position of the head in the Italian N-N compounds: the case of “mirror compounds”. *Linguistica Pragmatis*, 23/1, p. 41-52.
- Radimský, J. (2015). *Noun+Noun compounds in Italian. A corpus-based study*. České Budějovice: University of South Bohemia.
- Radimský J. (2018) Inflection of binominal ATAP compounds in French and Italian: a paradigmatic account. *Lingue e linguaggio* 2, 261–272. doi:[10.1418/91868](https://doi.org/10.1418/91868).
- Radimský, J. (2020). A paradigmatic account of lexical innovation: The role of repeated components in French N+N compounds. *Mediterranean Morphology Meetings; Vol 12 (2019): Rules, patterns, schemas and analogy*, 12, 77–91. <https://doi.org/10.26220/mmm.3251>
- Radimský, J. (2023). Tracing back the history of Italian Attributive-Appositive Noun+Noun compounds: First outcomes. *Linguistica Pragmatis*, 33(2), 136–155. <https://doi.org/10.14712/18059635.2023.2.3>
- Radimský, J. (2023). Where did the Italian Verbal-Nexus N+N compounds come from? *Mediterranean Morphology Meetings*, 13(0), Article 0. <https://doi.org/10.26220/mmm.4407>
- Štekauer, P. (2014). Derivational paradigms. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Derivational Morphology* (pp. 354–369). Oxford: Oxford University Press.

## Compounding as a trigger for derivational paradigmatic uniformity: Evidence from Hebrew

This study examines morphological regularization in Hebrew derived nominals as a result of compounding. Semitic morphology relies highly on non-concatenative morphology, where words are formed in patterns. The pattern consists of a vocalic pattern, consonantal slots and in some cases affixes (Berman 1978, Bolozky 1978, Schwarzwald 1981, Ravid 1990, Bat-El 1994, among others). Non-concatenative formation is obligatory for verbs. Nouns can be formed in various strategies, but verb-derived nominals only are formed in patterns. While there is a small set of verbal patterns, there are many nominal patterns, and there is some extent of irregularity manifested in pattern selection. Each Hebrew verbal pattern usually has one nominal pattern counterpart, that is typical for its derived action nouns (ANs). For example, ANs of *CaCaC* verbs are typically formed in *CCiCa* (*katav* 'wrote' - *ktiv* 'writing'), while ANs of *CiCeC* verbs are formed in *CiCuC* (*xipes* 'searched' - *xipus* 'searching').

However, there is some irregularity (Berman 1988, Ravid & Avidor 1998, Ravid 2000). For example, the AN of *rakad* 'danced', would be expected to be formed in *CCiCa* (*\*rkida*/*\*rekida*)<sup>1</sup>, but is formed in *CiCuC* (*rikud* 'dancing'). This irregularity is partially avoided in compound formation, where the derived nominal is the head, and the annex is its argument. For example, *ʔasar* 'prohibited' has an unexpected AN in the *CiCuC* pattern (*ʔisur* 'prohibiting') (1). This AN has a doublet in the *CCiCa* pattern - *ʔasira*. The latter surfaces only in compounds like *ʔasira(t)-išun* 'prohibiting smoking' (2) (see Berman 1988, Ravid & Avidor 1998), while *ʔisur* surfaces both with the complement and without it (3).

- (1) *kaše meod lehitgaber al ha-isur / \*ha-asira*  
'it is very difficult to stand against the prohibition' (<https://lib.cet.ac.il/pages/printitem.asp?item=16142>)
- (2) *ʔasirat ʔišun* nefoca hayom meʔod  
'prohibiting smoking is very common' (<https://www.fxp.co.il/showthread.php?t=10318548&page=3>)
- (3) *medabrim al mixtav legabey ʔisur ʔišun*  
'(they) talk about a letter regarding prohibiting smoking'  
(<http://www.kipa.co.il/community/show/4668093/>)

The formation in another pattern is motivated by semantic transparency, which is highly affected by the relations between the thematic roles that verbs assign and derived nominals. Verb-derived nominals are assumed to inherit properties of the verbs' argument structure (see for example, Grimshaw 1990, Rappaport Hovav & Levin 1992, Kratzer 1996, Alexiadou & Schafer 2006, Borer 2013). Studies have pointed out the thematic relations between elements of deverbal compounds, whose head is derived from verbs (e.g. *task assignment*) (Di Sciullo & Williams 1987, Di Sciullo 1992, Ralli 1992, Vogel & Napoli 1995, Di Sciullo & Ralli 1999, Ackema & Neeleman 2004, Leiber & Scalise 2006, Harley 2008, among others).

Compound formation brings about the formation of more regularized derived nominals, which gradually also surface out of compounds. This results in greater uniformity in the derivational paradigms of verbs and action nouns. I will demonstrate this in the talk for other types of derived nominals as well.

---

<sup>1</sup> In some cases there is an epenthetic vowel (*e/a*) between the first two consonants of the *CCiCa* pattern.

# Possible emergence of a paradigmatic organization of compounds: A case study of a Swedish child

Maria Rosenberg & Linda Sandström  
(Umeå University)

Keywords: paradigms, word-formation, compounding, language acquisition, diary data

Psycholinguistic evidence suggests that the lexicon has a paradigmatic organization, also in the domain of word formation (cf. Melloni & Dal Maso 2022). The present study addresses compounding in relation to paradigms in the early language development of Swedish, a highly compound-prone language. The study situates within a paradigmatic, word-based account, where derivational processes are assumed to manipulate form and meaning in predictable networks (e.g., Štekauer 2014, Bonami & Strnadová 2019, Hathout & Namer 2019, and Copot & Bonami 2024). Hence, “word systems are about multidirectional relationships between words and paradigm cells, in which no word has a privileged status” (Copot & Bonami 2024:221) and with the morphological family and the paradigmatic cell being two organizing dimensions (Copot & Bonami 2024:228).

In this case study, we assume that the child’s novel compounds would be paradigmatically motivated by their analogy to existing words – novel formations do not emerge in a void (cf. Gaeta 2022 and Bagasheva & Fernández-Domínguez 2022). The aim is thus to trace the emergence of compounds as one path into a paradigmatic organization of the lexicon, possibly originating via formal connections between position-bound constituents but rapidly spreading to a creative yet paradigmatic mixture of compound constituents based on form and/or semantics at different levels of abstraction (cf. Bauer 2022).

The data consist of around 500 compounds, with more than half being novel, drawn from diary notes of a Swedish-speaking child, (ages 1;9–4;0). A particular focus will be given to NN compounding, which predominates in the data as well as in Swedish in general. Most research on compound semantics tends to agree that a limited number of basic relations (around 10–15) can account for the majority of NN compounds (cf. Bauer & Tarasova 2013 and Bourque 2014). In the analysis, these basic relations can serve as a ground for arranging NN compounds into paradigms that consist of both compound families and compound series. Furthermore, adjectival compounds with affixoid-like first constituents are grasped early and lend themselves to a paradigmatic analysis, based on shared semantics. How the child gradually master liaison forms, which occur in some Swedish compounds, might provide another window into a possible paradigmatic organization. In addition, the analysis will consider Gaeta’s (2022:6) remark that, from a diachronic perspective, the transition from compound constituent into affix could depend on increase in paradigmaticity. We argue that a similar case can be observed in our data, where the distinction between position-bound compound constituents and derivational affixes is sometimes blurred. Overall, the data suggest a view of compounds emerging in smaller paradigmatic sets, or compound (sub-)families with position-bound constituents, in first or second position, sharing the same basic semantic relations (cf. Cetnarowska 2022; Bauer 2022).

In conclusion, by providing independent evidence from language development, we opt for a paradigmatic model encompassing morphological connections whether inflectional or derivational, originating from form and meaning relationships to end up in a self-organized system (cf. Bybee & Beckner 2010).

## References

- Bagasheva, Alexandra & Fernández-Domínguez, Jesús (2022), Fact-checking on compound verbs in English, in A.E. Ruz, C. Fernández-Alcaina & C. Lara-Clares (eds), *Paradigms in Word-Formation: Theory and Applications*, Amsterdam / Philadelphia: John Benjamins, 69–98.
- Bauer, Laurie (2022), Interlocking paradigms in English compounds, in A.E. Ruz, C. Fernández-Alcaina & C. Lara-Clares (eds), *Paradigms in Word-Formation: Theory and Applications*, Amsterdam / Philadelphia: John Benjamins, 59–68.
- Bauer, Laurie & Tarasova, Elizaveta (2013), The meaning link in nominal compounds, *SKASE Journal of Theoretical Linguistics* 10(13), 2–18.
- Bonami, Olivier & Strnadová, Jana (2019), Paradigm structure and predictability in derivational morphology, *Morphology* 29, 167–197.
- Bourque, Stephen Y. (2014), *Toward a Typology of Semantic Transparency: The Case of French Compounds*, Toronto: University of Toronto dissertation.
- Bybee, Joan L. & Beckner, Clay (2010), Usage-based theory, in B. Heine & H. Narrog (eds), *The Oxford Handbook of Linguistic Analysis*, Oxford: Oxford University Press, 827–855.
- Cetnarowska, Bożena (2022), Paradigm families in compounding: The case of English compound nouns headed by *-er* deverbal nouns, in A.E. Ruz, C. Fernández-Alcaina & C. Lara-Clares (eds), *Paradigms in Word-Formation: Theory and Applications*, Amsterdam / Philadelphia: John Benjamins, 99–128.
- Copot, Maria & Bonami, Olivier (2024), Baseless derivation: The behavioural reality of derivational paradigms, *Cognitive Linguistics* 35(2), 221–250.
- Gaeta, Livio (2022), Dangerous liaisons: An introduction to derivational paradigms, in A.E. Ruz, C. Fernández-Alcaina & C. Lara-Clares (eds), *Paradigms in Word-Formation: Theory and Applications*, Amsterdam / Philadelphia: John Benjamins, 3–18.
- Hathout, Nabil & Namer, Fiammetta (2019), Paradigms in word formation: what are we up to?, *Morphology* 29, 153–165.
- Jackendoff, Ray (2010), *Meaning and the Lexicon: The Parallel Architecture*, Oxford: Oxford University Press.
- Melloni, Chiara & Dal Maso, Serena (2022), For a topology of derivational paradigms, in A.E. Ruz, C. Fernández-Alcaina & C. Lara-Clares (eds), *Paradigms in Word-Formation: Theory and Applications*, Amsterdam / Philadelphia: John Benjamins, 21–56.
- Štekauer, Pavol (2014), Derivational paradigms, in R. Lieber & P. Štekauer (eds), *The Oxford Handbook of Derivational Morphology*, Oxford: Oxford University Press, 354–369.

# Pejorative compounding and paradigmatic relations: A cross-linguistic analysis of the concept HEAD

Muriel Norde, Elisa Mattiello, José Sanchez Fajardo, Francesca Masini, Kristel Van Goethem & Tijana Vesić Pavlović

(Humboldt-University Berlin, University of Pisa, University of Alicante, University of Bologna, F.R.S.-FNRS & Université catholique de Louvain & University of Belgrade)

Keywords: metonymy, pejoration, compounding, paradigms, cross-linguistic corpus studies

The human body is a rich source of productive patterns of pejoration cross-linguistically, often with racist or sexist connotation. In our talk, we focus on body parts that are connected to the category of [+upper body part] (henceforth HEAD), found in metonymical expressions featuring nouns meaning ‘head’, ‘face’, ‘brain’, or ‘skull’, such as English *-head* in *airhead* (Sánchez Fajardo 2022: 167–174) or *-face* in *buttface* (Mattiello 2024: 28–31). Similar examples are Dutch *stomkop* ‘stupid head’, Swedish *pappskalle* ‘cardboard skull’ or Serbian *pileći mozak* ‘chicken brain’, all meaning ‘fool’. In the Romance languages, by contrast, HEAD expressions are phrasal lexemes, e.g. Italian *testa di rapa* ‘head of turnip; fool’, or Spanish *cabeza de ladrillo* ‘brickhead; fool’. A Swedish and a Spanish example are given in (1-2):

- (1) *Den där **pappskallen** Harlekin har jag sett på andra bloggar också.* [svTenTen20]  
‘I have seen that fool Harlekin in other blogs as well’
- (2) *un trabajo de ficción para ver si el **cabeza de ladrillo** [...] entraba en razón* [esTenTen18]  
‘a fictional essay that is meant to see if the brickhead comes into his senses’

The aim of our talk is to present a comparative analysis of pejorative HEAD constructions in five languages: Dutch, Swedish, Italian, Spanish and Serbian, drawing data from web corpora at SketchEngine (MaCoCu for Serbian, TenTen for the other languages). For each language, we extract all HEAD collocations, removing irrelevant hits (like English *forehead*). The remaining types will be annotated for the collocate of HEAD (i.e. the first compounding member in the Germanic languages and in Serbian; the prepositional complement of the OF-phrase in the Romance languages), as well as for the part of speech of the collocate. In order to obtain detailed semantic profiles for each HEAD construction, we carry out a componential analysis of the 20 most frequent lemmas. This then allows us to identify clusters and regularities of HEAD expressions. The following two RQs will be addressed:

**RQ1:** Should compounds and phrasal lexemes be integrated into word-formation paradigms?

**RQ2:** Do cross-linguistic differences in compounding affect word-formation paradigms?

We adopt a constructionist approach to morphology (Booij 2010), which sees complex words not as a concatenation of morphemes, but as meaningful units in their own right that are paradigmatically related to other complex words. Where RQ1 is concerned, we argue that the notion of paradigm can be extended to compounding (e.g. Gaeta & Angster 2019; Hathout & Namer 2019; Bagasheva 2021), but also to phrasal lexemes. This implies that, like derivational constructions (Norde & Morris 2018), compounding constructions can be linked both interparadigmatically (e.g. all Swedish compounds ending in *-skalle* ‘skull’) and intraparadigmatically (e.g. all Swedish compounds made up of the same base and a different HEAD unit, e.g. *dumskalle* ‘stupid skull’ and *dumhuvud* ‘stupid head’). As to RQ2, we show that in both Modifier-Head and Head-Modifier languages, the evaluative unit undergoes a metonymization process (HEAD > PERSON), but that semantic secretion is even more prominent in Modifier-Head constructions, where the HEAD unit is used in the rightmost position, such that its paradigms are closer to affixation.

## References

- Bagasheva, Alexandra (2021), Paradigmaticity in Compounding, in J. Fernández-Domínguez, A. and C. L. Clares (eds.), *Paradigmatic Relations in Word Formation*, Leiden: BRILL, 21–48.
- Booij, Geert (2010), *Construction Morphology*. Oxford: Oxford University Press.
- Gaeta, Livio & Marco Angster (2019), Stripping paradigmatic relations out of the syntax, *Morphology* 29(2), 249–270.
- Hathout, Nabil & Fiammetta Namer (2019), Paradigms in word formation: what are we up to? *Morphology* 29(2), 153–165.
- Mattiello, Elisa (2024), A morphosyntactic and morphosemantic analysis of English slang suffixoids, *Studi e Saggi Linguistici* 62(1), 9–41.
- Norde, Muriel & Caroline Morris (2018), Derivation without category change. A network-based analysis of diminutive prefixoids in Dutch, in K. Van Goethem, M. Norde, E. Coussé and G. Vanderbauwhede (eds.), *Category Change from a Constructional Perspective* (Constructional Approaches to Language), Amsterdam / Philadelphia: John Benjamins, 47–90.
- Sánchez Fajardo, José A. (2022), *Pejorative Suffixes and Combining Forms in English* (Studies in Language Companion Series). Amsterdam, Philadelphia: John Benjamins Publishing Company.

# Compounding in a probabilistic paradigmatic model of word formation

Olivier Bonami, Louise McNally & Denis Paperno  
(Université Paris Cité, Universitat Pompeu Fabra & Universiteit Utrecht)

keywords : compounding, word formation, paradigm, French, quantitative morphology

Existing approaches to word formation embracing paradigmatic organization do not fully address how it influences the interpretation and production of novel lexemes. Baayen et al.'s (2019) LDL partially captures paradigmatic effects without providing a theory of morphological knowledge; representational approaches, from Construction Grammar (Booij 2010) and Relational Morphology (Jackendoff & Audring 2020) to radically paradigmatic models (Bonami & Strnadová 2019; Hathout & Namer 2022) provide such a theory but eschew explicitly modeling interpretation, production, and the way they are partially shaped by nonlinguistic knowledge. We argue for a novel, probabilistic approach, where paradigmatic knowledge mediates the relationship between form and meaning, together with context and world knowledge. We specifically assume that world knowledge comes into play in the form of probabilistic associations between lexeme families and scenarios, representations of situations and their participants (Sanford & Garrod 1998), which can be retrieved, e.g. by using LLMs (Sanacore et al. 2024). For instance French *descendeur* names agents or instruments in a variety of specific descending scenarios (downhill skiing, cycling, rappelling).

Here we adopt for compounding a model initially designed to deal with derivation. We take our exemplification from argumental compounds, but conjecture that the proposed architecture has broader applicability. We defend two key ideas. First, the notion of a derivational series needs to be extended to a more general notion of morphological series, which is essentially a word formation template with open slots, following e.g. Bagasheva (2011); Radimský (2020); Bagasheva & Fernández-Domínguez (2022). Second, compounds involve two lexeme families, but these evoke a single scenario. So for instance, the correct interpretation of French *tire-lait* 'breast pump' relies on a human breastfeeding scenario where milk, an action of drawing, and a pump all participate.

We will complement the theoretical elaboration with an empirical investigation of rivalry between verb-noun compounding and suffixation in the formation of French agent and instrument nouns, compare e.g. *tire-lait* above or *tire-bouchon* 'corkscrew' to *tireuse* 'female shooter', 'beer pump', 'optical printer'. We will build on available lexical data (Villoing 2002; Wauquier et al. 2020) and leverage distributional methods to show how compounding affords lower reliance on nonlinguistic cues by building on more lexical resources.

Specifically, given the very low granularity of the scenario information contributed by a verb such as *tirer* alone, and the number and variety of scenarios with which a noun like *lait* can be associated, the combination of the two in even a coarsely defined argumental relation vastly increases the specificity of the scenario information available to a hearer and, correspondingly, reduces the possible interpretations of the compound, when compared to a potential rival formed by simple derivation from one of the compound parts (e.g., *tireuse*). We argue that paradigmatic organization, together with independently available facts about scenarios (e.g. their salience), offers the prospect of facilitating probabilistic modeling of the preference for compounding as opposed to derivation in word formation.

## References

- Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, James P. Blevins, (2019), The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning, *Complexity* 2019.
- Bagasheva, Alexandra (2011), Beyond dichotomies : On the nature and classification of compound verbs in English, *Linguistica* 51(1), 39–63.
- Bagasheva, Alexandra & Jesús Fernández-Domínguez (2022), Fact-checking on compound verbs in English, In A. E. Ruz, C. Fernández-Alcaína, and C. Lara-Clares (eds.), *Paradigms in Word Formation: Theory and applications*, Amsterdam : John Benjamins, 69–98.
- Bonami, Olivier and Jana Strnadová (2019), Paradigm structure and predictability in derivational morphology, *Morphology* 29(2), 167–197.
- Booij, Geert (2010), *Construction Morphology*, Oxford : Oxford University Press.
- Hathout, Nabil and Fiammetta Namer (2022), ParaDis: A family and paradigm model, *Morphology* 32(2), 153–195.
- Jackendoff, Ray and Jenny Audring (2020), *The texture of the lexicon*, Oxford : Oxford University Press.
- Radimský, Jan (2020), A paradigmatic approach to compounding, in J. Fernández-Domínguez, A. Bagasheva and C. Lara Claes (eds.), *Paradigmatic relations in word formation*, Leiden: Brill, 164–185.
- Sanacore, Daniele, Nabil Hathout and Fiammetta Namer (2024), A scenario-based approach to predictability in derivation, *Lingue e linguaggio* (2), 337–359.
- Sanford, Anthony J. and Simon C. Garrod (1998), The role of scenario mapping in text comprehension, *Discourse Processes* 26(2–3), 159–190.
- Villoing, Florence (2002), *Les mots composés VN du français: réflexions épistémologiques et propositions d'analyse*, Université Paris X-Nanterre dissertation.
- Wauquier, Marine, Nabil Hathout, and Cécile Fabre (2020), Semantic discrimination of technicality in French nominalizations, *Zeitschrift für Wortbildung / Journal of Word Formation* 4, 100–119.



# Paradigmatic tendencies in NN compounds in the terminology of corporate communication in French: the example of sustainable development

Sophie Delforge, Anne Condamines, Fabio Montermini  
(CLLE, CNRS & Université de Toulouse Jean-Jaurès)

Keywords: <binominal compounds, sustainability, subordinate compounds, paradigmatic structures, textual terminology>

In this talk we present a study of NN French compounds in a corpus of financial reports from five companies over the last 20 years. The global size of each subcorpus and the number of NN compounds are given in Table 1.

Year	Number of words	Number of NN compounds
2022	1 581 597	3093
2017	1 521 646	2042
2012	1 149 804	1538
2007	930 257	1272
2003	725 570	722

Table 1: Global size and number of NN compounds in the corpus

Globally, we observe an increase in the use of binominal constructions such as *contrôle qualité* ('quality control'), *processus métier* ('business process') or *empreinte carbone* ('carbon footprint') in the fields of economics and corporate communication. In particular, we focus on the lexicon related to sustainable development, which expanded considerably due to the need to name potentially new concepts. As companies increasingly adopt CSR (Corporate Social Responsibility) practices, new terms and expressions emerge to reflect their commitment to sustainability, ethical governance, and social engagement.

Following the typology proposed by Bisetto & Scalise (2005), three major types of NN compounds may be identified in French, namely attributive, coordinate and subordinate compounds (see also Fradin, 2009; Villoing, 2012). In this talk, we focus particularly on subordinate NN compounds that are the most frequent in the database. In particular, we observe the emergence of patterns where specific lexemes are recurrently encountered as the right element (N2). These include words as *carbone* ('carbon'), *environnement* ('environment') *biodiversité* ('biodiversity') as in the following examples:

- (1) *empreinte carbone* ('carbon footprint'), *taxe carbone* ('carbon tax'), *bilan carbone* ('carbon footprint assessment')  
*stratégie environnement* ('environmental strategy'), *risque environnement* ('environmental risk'),  
*réseau environnement* ('environmental network')  
*norme biodiversité* ('biodiversity standard'), *enjeu biodiversité* ('biodiversity issue'), *démarche biodiversité* ('biodiversity programme')

The observation of corpus data highlights the tendency, for compounds, to evolve into paradigmatic structures and consequently, for the elements they contain, to adopt a more affix-like behaviour.

The noun *carbone* ('carbon'), when used as N2, is particularly productive. In our data, it combines with 25 different nouns (N1). Table 2 illustrates the growing centrality of climate-related concerns in economic and environmental discourse, by presenting the ten most frequent N1 encountered in the construction [N carbone] in the corpus:

(2)

[N carbone]	English translation	Number of occurrences
<i>empreinte carbone</i>	'carbon footprint'	470
<i>intensité carbone</i>	'carbon intensity'	124
<i>bilan carbone</i>	'carbon footprint assessment'	105
<i>neutralité carbone</i>	'carbon neutrality'	96
<i>émission carbone</i>	'carbon emission'	78
<i>fonds carbone</i>	'carbon fund'	75
<i>base carbone</i>	'carbon database'	69
<i>impact carbone</i>	'carbon impact'	58
<i>crédit carbone</i>	'carbon credit'	49
<i>réduction carbone</i>	'carbon reduction'	35

Table 2: Ten most frequent N1 encountered in [N carbone] compounds

The analysis we propose is based on two different but complementary frameworks. First, Construction Morphology (Booij, 2010) provides a suitable framework to explain the structure and evolution of these compounds, which can be described in terms of constructionalization (Traugott & Trousdale, 2013). Second, we adopt a textual terminology approach (Condamines & Picton, 2022), in order to show that the creation and diffusion of these compounds is strengthened in the terminology of a specific domain, especially one, like that of sustainability, which is constantly and rapidly changing.

## References

- Bisetto, A. & Scalise, S. (2005). The classification of compounds. *Lingue e linguaggio* 4(2), 319-332.
- Booij, G. (2010). *Construction Morphology*. Oxford University Press
- Condamines, A., & Picton, A. (2022). Chapter 10. Textual Terminology. Origins, principles and new challenges. *Terminology And Lexicography Research And Practice*, 219-236
- Fradin, B. (2009). IE, Romance: French. In R. Lieber, & P. Štekauer (Eds.), *The Oxford Handbook of Compounding*. Oxford University Press (pp. 417–435).
- Traugott, E. C. & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford University Press.
- Villoing, F. (2012). French compounds. *Probus* 24(1): 29-60

## **WS2 Connective negation: synchrony and diachrony**

**Mena B. Lafkioui & Johan van der Auwera**

# The pragmatic values of connective negation in contemporary Italian

Chiara Gianollo & Caterina Mauri  
(University of Bologna)

Keywords: connective negation, discourse-pragmatic function, emphatic negation, spoken corpus, semantic weakening

Italian has a dedicated marker, *né*, for the correlation of negative constituents (1): *né* can correlate phrases of various types, comprising full clauses, and takes part in the language's non-strict negative concord system (Scorretti 2001).

(1) KIParla, BOA1007

<i>non</i>	<i>ha</i>	<i>né</i>	<i>fatto</i>	<i>la</i>	<i>tesina</i>	<i>né</i>	<i>analizzato</i>
NEG	AUX.2SG	nor	do:PTCP	DEF:F.SG	essay	nor	analyze:PTCP
<i>una</i>	<i>sequenza</i>	<i>trascritta</i>					
INDEF:F.SG	sequence	transcribed					

'She has neither done the essay nor analyzed a transcribed sequence.'

Correlation represents a core functional domain for connective negation cross-linguistically (Haspelmath 2007, Sánchez López 2017, van der Auwera & Kookhan 2022, Salaberri 2022). At the same time, diachronic and typological research shows that markers used for negative correlation can also have further uses as discourse connectors and focus particles (Orlandini & Poccetti 2007, Gianollo 2020).

We aim to explore the main functional domains of *né* in a discourse-pragmatic perspective and its sociolinguistic correlates, by using data from spoken contemporary Italian (KIParla corpus; Mauri et al. 2019). We coded all the occurrences of *né* constructions in the Corpus (tot: 172) for the number of *né*(s), the semantic and morphosyntactic properties of the linked items, the cooccurrence with further negative expressions and further connectives, and their discourse functions.

A preliminary analysis of the occurrences clearly shows that they are employed as emphatic negators in discourse, through

- (i) *domain widening* (ex. *né polizia né carabinieri né niente* 'neither police nor gendarmerie nor anything'), in which *né* negates one or more exemplar(s) of a larger category, thus conveying the absence of potential exceptions;
- (ii) *domain exhaustivity* (ex. *né alto né basso* 'neither tall nor short'), in which *né* negates the opposite or complementary values of a category, entailing that all the (intermediate) category values are also negated;
- (iii) *enumeration* (ex. *né a me né a mia sorella* 'neither to me nor to my sister'), in which each relevant member of a given set, introduced in discourse, is separately negated.

The corpus also shows the competition with other strategies combining an additive interpretational component with a negative one (*e non* 'and not', *neanche*, *nemmeno* 'neither, not even'). In such contexts, *né* may be felt as insufficient to express the necessary discourse relation (parallelism, contrast, etc.), as shown by example (2), where the coordinating function of *né* is reinforced with an explicit conjunction *e* 'and'.

(2) KIParla, KPN015

*non ho voluto calcare né di olio e né di farina anche*  
NEG AUX.1SG want:PTCP press:INF nor of oil and nor of flour also  
'I didn't want to insist neither with oil (and) nor with flour (also).'

Though excluded by normative grammar, such constructions occur in spontaneous speech, pointing to a change in progress, possibly advancing the process of semantic weakening observed in the long trajectory from Latin to Italian (Gianollo 2020).

## References

- Gianollo, Chiara (2020), Grammaticalization Parameters and the Retrieval of Alternatives: Latin *nec* from Discourse Connector to Uninterpretable Feature, in R. Gergel & J. Watkins (eds), *Quantification and Scales in Change*, Berlin: Language Science Press, 33-65.
- Haspelmath, Martin (2007), Coordination, in T. Shopen (ed.), *Language Typology and Linguistic Description*, 2. ed., Cambridge: Cambridge University Press, 1-51.
- KIParla = Caterina Mauri, Eugenio Gorla, Silvia Ballarè, Massimo Cerruti, *KIParla. L'italiano parlato e chi parla italiano*, Università di Bologna-Università di Torino, <http://kiparla.it>.
- Orlandini, Anna & Paolo Poccetti (2007), Il y a *nec* et *nec*: Trois valeurs de la négation en latin et dans les langues de l'Italie ancienne, in F. Floricic (ed.), *La négation dans les langues romanes*, Amsterdam: Benjamins, 29-47.
- Salaberri, Iker (2022), A cross-linguistic study of emphatic negative coordination, *Studies in Language* 46, 647-717.
- Sánchez López, Cristina (2017), *Coordination and Correlatives*, in A. Dufter & E. Stark (eds), *Manual of Romance Morphosyntax and Syntax*, Berlin: De Gruyter, 647-688.
- Scorretti, Mario (2001), *Le strutture coordinate*, in L. Renzi, G. Salvi, A. Cardinaletti (eds), *Grande grammatica italiana di consultazione. Nuova edizione*, vol 1, Bologna: Il Mulino, 241-284.
- van der Auwera, Johan & Sepideh Kookhan (2022), Extending the typology: negative concord and connective negation in Persian, *Linguistic Typology at the Crossroads* 2, 1-36.

# When is Dutch *noch* ‘neither/nor’ (not) correlative? A corpus study

Daniel Van Olmen  
(Lancaster University)

Keywords: additive/connective negation, corpus, correlative, Dutch, non-correlative

One way in which Dutch can convey what van der Auwera and Van Olmen (forthc.) have called additive negation involves *noch* ‘neither/nor’. A feature of this expression is that it may be correlative or not, as in (1a) and (1b) respectively.

- (1) a. *Ik eet noch pizza noch pita.*  
I eat ADD.NEG pizza ADD.NEG pita  
b. *Ik eet pizza noch pita.*  
I eat pizza ADD.NEG pita  
‘I eat neither pizza nor pita.’

Research on other languages (e.g. van der Auwera 2021) has shown that this type of variation can be determined by various factors – such as additive negation’s position vis-à-vis the verb, the syntactic nature of what it connects and the presence/absence of clausal negation. For Dutch too, a range of potential determinants have been discussed in the literature (e.g. Moeyaert et al. 1986, Haeseryn et al. 1997, Broekhuis and Corver 2019). However, they have either not been tested against data or been considered only in isolation. This paper aims to address this gap with a multifactorial corpus analysis of *noch* in present-day Dutch.

The data comes from SoNaR (Oostdijk et al. 2013), a reference corpus of written Dutch, and particularly an arbitrary selection of texts from these increasingly formal types: chats; discussion lists; Wikipedia; reports and newspapers; and legal documents. Extra texts will be chosen until the target of 500 instances of *noch* per type is reached or somewhat exceeded, resulting in approximately 2,500 hits in total. In addition to the dependent variable of “correlative or not”, all cases will be coded for these literature-based independent variables: (i) text type (see the above distinctions); (ii) position before or after the finite verb; (iii) fixed expression, i.e. with non-compositional semantics, or not (e.g. *kind noch kraai hebben* ‘to have no next of kin’ but literally ‘to have neither child nor crow’); (iv) length, in terms of the number of both words and characters, of the first conjunct (i.e. *pizza* in 1a and 1b); (v) syntactic status of the conjuncts (e.g. phrases, finite verbs, clauses); (vi) indications of emphasis or not (e.g. capitalized *noch*); and (vii) clausal negation absent or not. On this data, a logistic regression analysis will be performed to assess the impact as well as the possible interactions of the different variables (Brezina 2018).

Preliminary results, based on a partially coded subset of the data, suggest that the syntactic status of the conjuncts plays no role whatsoever, despite (convincing) claims that correlative *noch* is hardly compatible with clausal conjuncts. The variable’s lack of statistical significance can, however, be attributed to the fact that *noch* seems to be used almost exclusively for phrasal conjuncts, like the noun phrases in (1). Another initial observation concerns text type: correlative *noch* is favored more in the chats than the more formal reports and newspapers. An explanation may lie in the negative-first principle (Jespersen 1917) and its potentially stronger impact in more informal language. A final

provisional finding has to do with fixed expressions, which prefer a single *noch* (e.g. regardless of text type).

## References

- Brezina, Vaclav (2018), *Statistics in Corpus Linguistics: A Practical Guide*, Cambridge: Cambridge University Press.
- Broekhuis, Hans and Norbert Corver, (2019), *Syntax of Dutch: Coordination and Ellipsis*, Amsterdam: Amsterdam University Press.
- Haeseryn, Walter, Kirsten Romijn, Guido Geerts, Jaap de Rooij and Maarten van den Toorn (1997), *Algemene Nederlandse Spraakkunst*, Groningen: Nijhoff.
- Jespersen, Otto (1917), *Negation in English and Other Languages*, Copenhagen: Høst and Søn.
- Moeyaert, Cyriel., Piet C. Paardekooper and Julien Peperstraete (1986), *Beknopte A.B.N.-Spraakunst*, Kapellen: Pelckmans.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman (2013), The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch, in Peter Spyns and Jan Odijs (eds), (2013), *Essential Speech and Language Technology for Dutch: Results by the STEVIN Programme*, Dordrecht: Springer, 219–247.
- van der Auwera, Johan (2021), Quirky Negative Concord: Croation, Spanish and French *Ni*'s. *Jezikoslovlje* 22(2), 195–225.
- van der Auwera, Johan and Daniel Van Olmen (forthc.), Additive Negation in Dutch, from Synchrony to Diachrony, Cyclical and Noncyclical, in Maj-Britt Mosegaard Hansen and Richard Waltereit (eds), (forthc.), *Cyclical Change in Grammar and Discourse*, Oxford: Oxford University Press.

## Three strategies of connective negation in Udmurt

Erika Asztalos

(Hungarian Research Centre for Linguistics, Budapest)

Keywords: connective negation, contact-induced variation, negation, scope, Udmurt

Despite a recent interest in the study of negation in Udmurt (Uralic; Russia) (cf. Kozmács 2017, Georgieva et al. 2021, Asztalos 2023), connective negation still remains an understudied topic in Udmurt linguistics. In this talk, I claim that Udmurt has two markers and three strategies of connective negation:

- 1) a native strategy involving *no ... no* (where *no* is a conjunction meaning ‘and; also’) + a negated predicate (1) (negation in Udmurt happens by means of an inflected negative auxiliary);

and two contact-induced strategies, both involving the doubling of the negative particle *ńe* ‘not’ borrowed from Russian:

- 2) *ńe ... ńe* + a negated predicate (2);

- 3) *ńe ... ńe* + an affirmative predicate (3):

- (1) *Busj-os-jn kenem no, mak no ug budj.*  
field-PL-INE hempseed also poppy\_seed also NEG.PRS.3SG grow.CNG.SG  
‘Neither hempseeds nor poppy seeds grow on the fields.’ (Udmurt corpus)

- (2) *ńe pęś-ez, ńe keźjt-ez ug lįktj.*  
not hot-DET not cold-DET NEG.PRS.3SG come.CNG.SG  
‘Neither hot nor cold weather is coming.’ (Udmurt corpus)

- (3) *Oźj bjdes pokoleńi-mj ńe źuč, ńe udmurt bud-i-z.*  
that\_way whole generation-1PL not Russian not Udmurt grow-PST-3SG  
‘Thus, our generation grew up neither as Russians nor as Udmurts.’ (Udmurt corpus)

The talk addresses the following questions:

- i) Is the use of the “native” and the Russian-induced strategies subject to interspeaker variation?
- ii) What factors condition whether *ńe ... ńe* stands with a negative or an affirmative predicate?
- iii) Does the scope of negation differ for the three strategies?
- iv) Why does Udmurt resort to *ńe*-doubling instead of borrowing *ńi ... ńi*, which is the standard marker of connective negation in Russian?

The research was carried out on contemporary data retrieved from the Udmurt corpus and was complemented by consultations with native speakers. The scope of the markers was tested by sentences with indefinite subjects (of the type “A guest me also, you also not visited” / A guest neither me nor you (not) visited”: if the subject can be interpreted as a non-specific guest, i.e., not necessarily the same person, that means negation scopes over the subject *a guest*; by contrast if *a guest* has the interpretation of a specific subject, that indicates that the subject stays outside the scope of negation).

Preliminary results suggest that the use of the “native” (1) vs. the “Russian induced” (2)–(3) strategies is subject to interspeaker variation, while the choice between 2) and 3) is motivated by



different presuppositions on the truth value of the predicate (i.e., in (3) it is presupposed that “our generation grew up in some way”, whereas in (2) there is no presupposition concerning whether anything is coming). Tests indicate that in (1) and in (2), negation has clausal scope, thus it scopes over the markers of connective negation. The difference between the two strategies consists in whether the marker is a negative concord item (2) or not (1). In (3), however, negation is expressed by the marker of connective negation itself. Structures like (3) are analysed as constituent negation constructions.

## References

- Asztalos, Erika (2023), The expression of constituent negation in Udmurt: From scope ambiguous to scope-transparent constructions, *Acta Linguistica Academica* 70 (2), 248–283.
- Georgieva, Ekaterina, Martin Salzmann & Philipp Weisser (2021), Negative verb clusters in Mari and Udmurt and why they require postsyntactic top-down word-formation, *Natural Language and Linguistic Theory* 39, 457–503.
- Kozmács, István (2017), A tagadó szerkezet változása az udmurtban, in A. Benő, and N. Fazakas (eds), (2017), *Élőnyelvi kutatások és a dialektológia*, Kolozsvár: Erdélyi Múzeum-Egyesület, 277–284.
- Udmurt corpus. [http://udmurt.web-corpora.net/udmurt\\_corpus/search](http://udmurt.web-corpora.net/udmurt_corpus/search) (accessed on 15.01.2025).

# Connective negation and areality in the Hindu Kush region and beyond

Henrik Liljegren

(Stockholm University)

Keywords: addition, areality, connective negation, emphasis, Hindu Kush

Languages in the linguistically diverse Hindu Kush region in High Asia display an intriguing inventory of connective negation constructions. The major strategy at use in various Hindu Kush languages, as in (1)–(3), tends to line up relatively closely with what is observed in a large macroarea stretching from the Eastern Mediterranean to the Bay of Bengal, equal to the extension of the historic Islamic Gunpowder Empires (Streusand 2010), exemplified here with Turkish (4). Strikingly similar lexical material is found in languages geographically far apart and phylogenetically distinct. Typically, both constituents are identically marked, and the markers almost invariably occur in pre-posed position: **NEG x NEG y**. Where-ever it occurs, whether in Hindu Kush or in the larger macroarea, it bears the undeniable mark of Persian influence.

- (1) Wakhi, Iranian (Obrtelová 2019: 583)

[Na	wuz=əm	əstar	vit-i=ət]	[nə	mol-iš]
CONEG	I=1SG	nothing	become.PST=and	CONEG	livestock-PL
'Neither I, nor the flock came to any harm.'					

- (2) Palula, Indo-Aryan (Liljegren 2019)

[Na	zinaawur-á	tas	the	ga	asár	th-íil-i	de]	[na
CONEG	beast-OBL	3SG.ACC	to	what	effect	do-PFV-F	PST	CONEG
ghrast-á	th-íil-i	de]						
wolf-OBL	do-PFV-F	PST						
'Neither wild animals nor wolves had touched him.'								

- (3) Purik, Sino-Tibetan (Zemp 2018: 360)

k <sup>h</sup> o	[na	mi	men]	[na	ane	men]
s/he	CONEG	man	N:EQ	CONEG	woman	N:EQ
'S/he's neither man nor woman.'						

- (4) Turkish, Turkic (van Schaaik 2020: 332)

Böyle	bir	balığ-ı	[ne	gör-dü-m]	[ne	de	işit-ti-m]
such	one	fish-ACC	CONEG	see-PST-1SG	CONEG	too	hear-PST-1SG
'Such a fish I have neither seen nor heard of.'							

A number of subtypes occur in individual languages of the Hindu Kush, some of them with a significant subareal distribution, while others appear to have emerged as the result of language-internal processes. Common to most of those is the combination of an overall strategy, similar to the one outlined above, and the use of language-specific discourse particles. One type that occurs in Gawarbat, exemplified in (5), is structurally the negative counterpart of an additive construction, **x ADD ... y ADD** 'both x and y'. Another type occurs in Nuristani Kalasha, exemplified in (6), as well as in neighbouring Indo-Aryan languages, in which two different particles are used in combination with the repeated connective negator, **NEG PRT<sub>1</sub> x ... NEG PRT<sub>2</sub> y**. A corresponding affirmative construction, **x PRT<sub>1</sub> ... y PRT<sub>2</sub>**, has an emphatic contrasting function, described by Liljegren & Svärd (2017). While it is too

early to identify a default strategy, or indeed to determine the exact distribution and semantic range of each subtype, the data suggest that negated emphasis and addition are senses each related to a particular subtype rather than to connective negation in general, cf. Haspelmath (2007: 17–19) and Salaberri (2022).

(5) Gawarbat, Indo-Aryan (unpublished corpus data)

[ <i>Nai=bi</i>	maramat-ã:	paisa	ti-met]	au	[ <i>nai=bi</i>	k <sup>h</sup> ud
CONEG=ADD	repair-DAT	money	give-PRS.3PL	and	CONEG=ADD	question
ke-met]						
do-PRS.3PL						

‘They neither give nor ask for repair money.’

(6) Nuristani Kalasha, Nuristani (Degener 1998: 169)

[ <i>Na=da</i>	war-gay	boy]	[ <i>na=ri</i>	ber-e	boy]
CONEG=FOC	in-go.ABS	could	CONEG=CNJ	out-come.ABS	could

‘He could neither go back nor come forward.’

## References

- Degener, Almuth (1998), *Die Sprache von Nisheygram im afghanischen Hindukusch*, Wiesbaden: Harrassowitz Verlag.
- Haspelmath, Martin (2007), Coordination. In Timothy Shopen (ed), (2007), *Language typology and syntactic description: Volume 2, Complex constructions*, Cambridge, UK / New York: Cambridge University Press, 1–51.
- Liljegren, Henrik (2019), Palula dictionary, *Dictionaria* (3), 1–2700.
- Liljegren, Henrik and Erik Svärd (2017), Bisyndetic contrast marking in the Hindukush: Additional evidence of a historical contact zone, *Journal of Language Contact* 10(3), 450–484.
- Obrtelová, Jaroslava (2019), *From Oral to Written: A Text-linguistic Study of Wakhi Narratives*, Uppsala: Uppsala University.
- Salaberri, Iker (2022), A cross-linguistic study of emphatic negative coordination, *Studies in Language* 46(3), 647–717.
- Schaaik, Gerjan van (2020), *The Oxford Turkish Grammar*, Oxford: Oxford University Press.
- Streusand, Douglas E (2010), *Islamic Gunpowder Empires: Ottomans, Safavids, and Mughals*, Boulder, Colo: Routledge.
- Zemp, Marius (2018), *A grammar of Purik Tibetan* (Brill’s Tibetan Studies Library volume 21), Leiden / Boston: Brill.

# Exploring connective negation in Basque: A diachronic corpus study

Iker Salaberri

(University of the Basque Country (UPV/EHU))

Keywords: negation, clause linkage, Basque, reconstruction, corpus linguistics

Despite recent advances, connective negation (CoNeg), the simultaneous negation and coordination of same-category coordinands, is an underresearched topic, particularly in comparison with other kinds of negation. Accordingly, many open questions remain, including when a CoNeg construction should be considered dedicated, whether CoNeg always involves emphasis, and how it emerges in languages that formerly did not have this clause-linkage function.

In view of this data situation, this paper analyzes the emergence, evolution and variability of CoNeg in Basque, none of which have been described in detail. This language can contribute to further delimiting CoNeg in cross-linguistic terms, for various reasons: standard negation (SN) and CoNeg are expressed by the same marker *ez*, yet SN is asymmetric in this language (Miestamo 2005: 153), whereas not all forms of CoNeg are; when CoNeg connects phrases, negative concord is obligatory (1a), whereas it is banned when whole (finite) clauses are linked (1b); in parallel to languages like Croatian (van der Auwera 2021: 205), when two or more phrases are coordinated, negation of the first is optional (1c).

- (1) a. *Ez=togu ez ollanda-ric, ez eperr-ic jan*  
NEG=AUX CONEG chick-PART CONEG partridge-PART eat.PFV  
'We have eaten neither chick nor partridge'  
(Mogel 1904 [1802]: 67)
- b. *Ez gal-du da ez gal-du-ren çu-re*  
CONEG lose-PFV AUX CONEG lose-PFV-FUT 2SG-GEN  
*gomendu-co-ric*  
mandate-RS-PART  
'Neither has anyone mandated by you been lost, nor will they be lost'  
(Etxepare 1980 [1545]: 104)
- c. *Çaudé presto, eta vigilant, ceren*  
stand.2SG ready and watchful because  
*ez=pai-t-a-qui-çu (ez) egun-a, ez oren-a*  
NEG=SUB-3SG-PRS-know-2SG CONEG day-DEF CONEG hour-DEF  
'Stand ready and watchful, because you know neither the day nor the hour'  
(Tartas 1666: 176)

CoNeg is analyzed drawing on a database of over 6000 negated clauses from 16<sup>th</sup>-to-20<sup>th</sup>-century Basque texts. The data are analyzed using Cramer's V and Theil's U statistical tests in order to search for correlations between four variables: (i) number of coordinands; (ii) syntactic complexity of the

coordinands, i.e., whether these are phrases, clauses or sentences; (iii) the phonological weight of coordinands, measured in number of syllables; (iv) the extent to which CoNeg markers are accompanied by reinforcing elements.

Functional factors are adduced in order to account for the tendencies observed: the strong propensity to have two (and no more) coordinands is related with speakers' processing and memory restrictions, and the preponderance of CoNeg markers reinforced by additive particles such as *ere* 'also' is accounted for as a result of the emphatic nature of this clause linkage function. Finally, it is argued that a dedicated CoNeg construction emerged via elision of the coordinating conjunction *eta* 'and' in syndetically coordinated sentences under the scope of negation, whereby the negator *ez* took over the coordinating function and was extended to each coordinand, thus being used correlatively.

### Abbreviations

1/2/3	1 <sup>st</sup> /2 <sup>nd</sup> /3 <sup>rd</sup> person
AUX	auxiliary verb
CONEG	connective negation
FUT	future
NEG	negator
PART	partitive
PFV	perfective
PRS	present
RS	relational suffix
SUB	subordinator

### Acknowledgements

This research has been partially supported by the research project TransGram PID2021-124769NB-I00, funded by the Spanish Ministry of Science and Innovation / Spanish Research Agency (AEI) /10.13039/501100011033, and the European Regional Development Fund (ERDF) "A way of making Europe"), as well as by the research project *Diachronic Linguistics, Typology and the History of Basque* (DLTB) (IT1534-22), funded by the Government of the Basque Autonomous Community.

### References

- Etxepare, Bernard (1980) [1545], *Linguae vasconum primitiae*, Bilbao: Euskaltzaindia, Ediciones Mensajero.
- Miestamo, Matti (2005), *Standard negation: The negation of declarative verbal main clauses in a typological perspective*, Berlin / New York: De Gruyter.
- Mogel, Juan Antonio (1904) [1802], *El doctor Peru Abarca*, Durango: Florentino Elosu.
- Tartas, Ioan (1666), *Onsa hilceco bidia*, Orthez: Jacques Rovyer.
- van der Auwera, Johan (2021), Quirky negative concord: Croatian, Spanish and French *ni*'s, *Jezikoslovlje* 22(2), 195–225.

# Connective negation in Slavic languages: synchrony and diachrony

Ljuba Veselinova  
(Stockholm University)

Keywords: connective negation, Slavic, historical-comparative method, areality, stability

This study, which is both synchronically and diachronically oriented, offers an overview of connective negation in Slavic languages. Following (van der Auwera et al. 2021), (Salaberri 2022) and (Haspelmath 2007), I study expressions that can be demonstrated to negate functionally equivalent constituents coordinated in a larger unit.

For the synchronic part, I review data from 14 modern languages spread across the traditional East, West and South groups using grammars and corpora. Building on van der Auwera et al. (2021), this study offers further details on the encoding, functions and frequency of Slavic connective negators. In the languages of the East group, we observe the particle *ni*, in most of West group, *ani*. Typically, these particles are used in complex constructions, *ni...ni* or *ani...ani*. In the South group, connective negation is encoded by longer or shorter constructions, *nito/niti..nito/niti* or *ni..ni*.

The most common use of connective negators is the encoding of negation in coordinated structures. In Croatian, the *niti* construction is used with clausal scope while the *ni* construction scopes over constituents only. Such a specialization appears valid for South West Slavic only. Connective negators are observed with non-negative functions as well. For instance, in Belarusian, the particle *ni* collocates with several interrogative words to yield expressions that indicate indetermination as in *kudy ni gljan'* 'wherever you look' (Byrila 1985: 175).

Connective negators show low frequency, between 0.02% to 0.12% in different corpora. Frequency correlates with generality of meaning. For instance, *ani* occurs in 0.03% of PolishWeb2019 but three times as much, 0.09% in SlovakWeb2021. In Polish, the uses of *ani* are equally distributed between connective negative constructions, scalar ones 'not even', and sometimes 'not'. In SlovakWeb2021, *ani* is used much less in the construction *ani...ani* and much more often as a single expression functioning as a general constituent negator, not always emphatic.

The diachronic part of the study includes a historical-comparative analysis of the modern dataset and discussing relevant material from Old Church Slavonic. The particle *ni* occurs in all languages; the observed variants are clearly fusions between *ni* and another element, typically a reinforcer. They correlate with specific groups which suggests they are newer creations. In Old Church Slavonic, *ni* had a variety of uses: a conjunction between equivalent constituents, negative answer, reinforcer of negation. Generally, the expression of connective negation appears to be a diachronically stable feature of Slavic languages that has been further consolidated by augmenting an older construction.

## REFERENCES

- Byrila, M. V. (1985) *Beloruskaja gramatyka*. Minsk: Nauka i Téxnika.
- Haspelmath, Martin. 2007. Coordination. In Timothy Shopen (ed.), *Language Typology and Syntactic Description: Volume 2: Complex Constructions*, vol. 2, 1–51. 2nd edn. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511619434.001>.
- Salaberri, Iker. 2022. A cross-linguistic study of emphatic negative coordination. *Studies in Language* 46(3). 647–717. <https://doi.org/10.1075/sl.20047.sal>.
- van der Auwera, Johan, Motoki Nomachi & Olga Krasnoukhova (2021) Connective negation and negative concord in Balto-Slavic. In Peter Arkadiev, J. Parkerys, I. Šeškauskienė & V. Žeimantienė (eds.), *Studies in Baltic and Other Languages. A Festschrift for Axel Holvoet on the occasion of his 65th birthday* (Open Series 16), 45–66, Vilnius: Vilnius University. doi: 10.15388/SBOL.2021.2.

# Changing forms of connective negation in French: The relevance and limits of cyclicity

Maj-Britt Mosegaard [Hansen](#)

(The University of Manchester)

**Keywords:** Negation, Connective negators, Diachrony, Cyclicity, French

I discuss the role of cyclicity in the evolution of connective negators (CONEGs) in French.

I first look briefly at the evolution of the negative coordinating conjunction (*ni...*)*ni* ('neither'/'nor') from Latin to Modern French, which has taken an almost perfectly circular path (Hansen 2021), from occurring exclusively in strong negative polarity (NP) contexts, to becoming compatible with weak NP contexts, back to strong NP contexts again, (1)-(3).

(1) *quia non viderunt nec sciunt* (1st c. BCE)

'because they have not seen nor do they know'

(2) *Quidez vus dunc k'il surrexist, ne qu'il vus puisse garantir ?* (c. 1130)

'So do you think that he was resurrected and that he can protect you?'

(3) *Je ne l'aime ni ne le déteste.*

'I neither love him nor hate him.'

*Ni* is, however, declining and increasingly confined to formal registers. In Contemporary French, it competes with the adverbial CONEGs *non plus/pas plus/pas davantage*, (4), which all originate in analytic comparatives ('not more'), (5):

(4) *Pierre n'a pas mangé et Marie non plus/pas plus/pas davantage.*

'Pierre hasn't eaten and neither has Marie.

(5) *... et que la Vierge Marie n'a pas plus de valeur qu'une autre femme.* (1542)

'... and that the Virgin Mary does not have greater value than another woman.'

The successive evolution of (i) *non plus*, (ii) *pas plus*, and (iii) *pas davantage* into CONEGs is charted using the written database *Frantext* (9<sup>th</sup>-21<sup>st</sup> c.). I focus, on the one hand, on the kinds of bridging contexts (Heine 2002) that may plausibly have triggered the changes (0, where the notion that Fatmé did not like the sentiment is merely a possible inference), and on the other hand, on the relationship between the three forms.

(6) *Fatmé [...] fut la première à s'ennuyer du sentiment. Le bramine, à qui il ne plaisoit pas plus qu'à elle, le quitta bientôt aussi.* (1742)

'Fatmé [...] was the first to become bored by the sentiment. The Brahmin, who didn't like it any more than she did/who didn't like it, either, soon abandoned it, too.'

Importantly, it is not clear that *ni* is actually being replaced by any of the newer forms, all of which occur even less frequently than *ni* in the contemporary part of *Frantext*, while only one, *non plus*, is

more frequent than *ni* in the contemporary spoken *Corpus CLAPI*. Nor can the more recent of the three latter forms meaningfully be said to renew the older ones, insofar as *non plus* is the most frequent and *pas davantage* the least frequent of these three CONEGs. For the time being, we thus cannot affirm that a cyclic development is involved. What looks superficially like a(n incipient) cycle may rather be a case of an onomasiologically “central” construction experiencing varying degrees of competition from a range of “satellite” constructions, none of which ultimately succeed in replacing it (Detges 2020). This type of scenario may in fact be more common across languages and construction types than cycles as such.

## References

- Corpus CLAPI*, Online database of contemporary spoken French, <http://clapi.ish-lyon.cnrs.fr>
- Detges, Ulrich (2020), Future markers in Western Romance. Cyclic change, synchronic variation and diachronic competition, *Journal of Historical Pragmatics* 21(2), 289-314.
- Frantext*, Online historical corpus of written French (950-present day), <https://www.frantext.fr/>
- Hansen, Maj-Britt Mosegaard (2021), Cyclic changes to the negative coordinating conjunction from Latin to Modern French, *Folia Linguistica Historica* 42(2), 223-254.
- Heine, Bernd (2002), On the role of context in grammaticalization, in Ilse Wischer, and Gabriele Diewald (eds), (2009), *New Reflections on Grammaticalization*, Amsterdam: John Benjamins, 83-101.



## Polish *ani* – an uncontroversial marker of negation connective, but for much longer?

Marzena Stępień  
(University of Warsaw)

Keywords: negative concord, connective, negation, Polish

The paper focuses on the Polish dedicated negation connective marker *ani* both in its standard connective usage and in other constructions.

Polish *ani* (Greszczuk 1993, Kallas 1994, Kisiel & Stępień 2016, van der Auwera J., Nomachi M. & Krasnoukhova 2021) has the capacity to connect both clauses (1) and nominal (noun) phrases (2), including prepositional phrases (3). It is claimed that the additional negation marker *nie* preceding the main verb is obligatory in most negation connective constructions, whereas *ani* can also be reduplicated in the sentence, introducing both clauses, phrases or clauses and phrases, the most typical context being the clausal-initial connective *ani*: (1a), (3a). It is also possible to extend the chain of clauses by putting the coordinating conjunction *albo* (or) into the work.

(1)

Pożyczek	nie	bierz	ani	nie	udzielaj.
loan.GEN.PL	NEG	take.IMP.2SG	nor	NEG	give.IMP.2SG

(Shakespeare, Hamlet, transl. Stanisław Barańczak, 1990)

"Neither take loans nor give them."

(1a)

Ani	nie	bierz	pożyczek,	ani	ich	nie	udzielaj
Nor	NEG	take.IMP.2SG	loan.GEN.PL,	nor	them	NEG	give.IMP.2SG

"Neither take loans nor give them."

(2)

Nie	pożyczaj	drugim	ani	od	drugich
NEG	lend.IMP.2SG	other.DAT.PL	nor	from	other.LOC.PL

(transl. Józef Paszkowski, 1862)

"Do not lend to others nor borrow from others."

(3)

Nie	bądź	dłużnikiem	ani	wierzycielem
NEG	be.IMP.2SG	debtor.INS.SG	nor	creditor.INS.SG

(transl. Maciej Słomczyński, 1978)

"Be neither a debtor nor a creditor."

(3a)

Ani nie bądź dłużnikiem, ani wierzycielem.

Ani	nie	bądź	dłużnikiem,	ani	wierzycielem
-----	-----	------	-------------	-----	--------------

nor NEG be.IMP.2SG debtor.INS.SG, nor creditor.INS.SG  
"Be neither a debtor nor a creditor."

(4)

Nie słyshałem ani jednego słowa. (KWJP)  
NEG hear.PST.1SG even one.GEN.SG.N word.GEN.SG  
"I didn't hear a single word."

In both cases (single vs reduplicated usage), the expression can be reduced to *ni*; however, in contemporary Polish this variant is considered obsolete. In all cases – connective negation with clausal scope and nominal phrase scope goes with the negative concord, similarly to the non-connective negation. Nonetheless, it was not always the case, as evidenced by the fact that in the 17th century both types of clauses with *ani* – with and without an additional negation marker *nie* – were possible (Piotrowska 2014) and are still observed in contemporary speech corpora. It can be argued that absence of explicit negation exponent *nie* is one of the indicatives of non-connective use of *ani*, although this is not universally applicable (4) (Kisiel, Stępień 2016).

The study adopts a functional approach incorporating corpus-based qualitative and quantitative analysis, including the most recent Polish text corpora: The Corpus of Contemporary Polish (Korpus Współczesnego Języka Polskiego – KWJP). It addresses the research questions formulated in the workshop proposal, namely: does the variety of *ani* usage call for plurality of terms, including pragmatic markers? Furthermore, the study explores whether emphasis, obligatoriness or exclusion of reduplication and/or additional negation exponent *nie* can serve as a criterion for distinguishing separate phenomena.

## References

- van der Auwera, Johan, Nomachi, Motoki and Krasnoukhova, Olga, (2021), Connective negation and negative concord in Balto-Slavic, in Arkadiev, P., Pakerys, J., Šeškauskienė, I. & Žeimantienė, V. (eds), (2021), *Studies in Baltic and Other Languages. A Festschrift for Axel Holvoet on the occasion of his 65th birthday. Vilnius University Open Series* 16, 45–66. doi: 10.15388/SBOL.2021.2
- Greszczuk, Barbara, (1993), Przeczenie przy „ani” – obligatoryjne czy fakultatywne? *Prace Filologiczne* XXXVIII, 107–137.
- Kallas, Krystyna, (1994), Syntaktyczne cechy spójnika i partykuły „ani”, *Polonica* XVI, 103–125.
- Kisiel, Anna and Stępień, Marzena, (2016), O granicach jednostek języka. Jeszcze o klasie otwartej na przykładzie polskich jednostek z elementem „ani”, *Linguistica Copernicana* 13, 87–105. <https://doi.org/10.12775/LinCop.2016.004>
- Kubicka, Emilia, (2006), Jednostki języka o kształcie „ni”, *Polonica* XXVI–XXVII, 147–160.
- Kieraś, Witold i in., (2024), Korpus Współczesnego Języka Polskiego. Dekada 2011-2020, *Język Polski* listopad 2024, <https://doi.org/10.31286/JP.001055>.
- Piotrowska, Agnieszka Ewa, (2014), Próba charakterystyki semantycznej wyrazów „ni” i „ani” oraz zaimków nieokreślonych z częstką „ni-” w dobie staropolskiej i średniopolskiej (na materiale słownikowym, *Poradnik Językowy* 7, 21–40.

## Coordinating negative clauses in Uralic languages

Matti Miestamo, Ksenia Shagal, Olli O. Silvennoinen & Chingduang Yurayong  
(University of Helsinki, LMU München, Åbo Akademi University & University of Helsinki)

Keywords: negation, coordination, typology, Uralic languages

This paper will examine the coordination of negative clauses in Uralic languages from a typological perspective. Some cross-linguistic observations on negative coordination were made already by Jespersen (1917), but it has not been the object of much large-scale typological work; a notable exception to this is Salaberri's (2022) work on emphatic negative coordination (ENC), exemplified by English *neither... nor*. In our ongoing typological work, we look at the coordination of negative clauses in a broader perspective, paying attention to clausal coordination structures in which at least one of the coordinated clauses is negative. We ask whether and how the marking of negation in coordinated clauses differs from the marking of negation in corresponding simple clauses and whether there are any structural differences between negative vs. non-negative clauses in coordination. The most salient differences vis-à-vis affirmatives are found in the form and behaviour of negative coordinators.

In this talk, the focus will be on Uralic languages, which are discussed against the larger typological background. Uralic languages are especially interesting in the typological perspective for at least two reasons: firstly, as most of them use a negative auxiliary for standard negation, it is interesting to see how these relate to coordination, and secondly, with their varied contact situations, they provide an interesting laboratory of how language contact may affect negative coordination. We will present a comprehensive typological overview of the coordination of negatives in Uralic languages and pay special attention to contact effects. Sampling will be exhaustive in the sense that all Uralic languages for which sufficient descriptive data sources can be found will be included in the study. Sources will include reference grammars and other relevant descriptive materials, such as the chapters in Miestamo et al. (2015). The picture given by these sources will be complemented by corpus data from spoken language and folkloric texts in three Uralic languages representing different branches of the family: Finnish (Finnic), Mari (Mari) and Mansi (Ugric). In the corpora, we will observe differences in frequency, distribution and function between alternative coordination constructions within these languages.

Preliminary results show interesting variation across the family, as regards for example the devices used for coordinating negatives. These range from a negatively polar coordinating clitic appearing on the negative auxiliary in some Finnic and Saamic languages, e.g. Finnish =*kä*, to the fused special negative coordinator *ega* 'and not, nor' in Estonian and the negative coordinators *ni...ni* 'neither...nor' borrowed from Russian in many Uralic languages spoken in Russia, e.g. Selkup, Komi and Votic. In the broader typological context, it is interesting to observe that similar negative coordinators are widely borrowed, e.g. from Spanish in the indigenous languages of Latin America. At the same time, some Uralic languages, e.g. Eastern Khanty, Erzya and South Saami, do not have special coordinators in negatives and use the same elements in negatives as in positive coordinated clauses.

### References

- Jespersen, Otto. 1917. *Negation in English and other languages*. Copenhagen: Høst.
- Miestamo, Matti, Anne Tamm & Beáta Wagner-Nagy. 2015. *Negation in Uralic Languages* (Typological Studies in Language 108). Amsterdam: Benjamins.
- Salaberri, Iker. 2022. A cross-linguistic study of emphatic negative coordination. *Studies in Language* 46 (3). 647-717.

## Bantu connective negation in the shadow of main clause negation

Maud Devos, Rasmus Bernander & Jacky Maniacky

(Royal Museum for Central Africa; and Ghent University, University of Gothenburg, Royal Museum for Central Africa)

Keywords: Bantu, Swahili, connective negation, standard negation, borrowing

If connective negation has received little attention in the languages of the world, it has been an ignored type of negation in Bantu languages. The reason could be that Bantu languages typically express anything negative on the main verb. Indeed, in a (growing) sample of 50 Bantu languages, connective negation is overwhelmingly expressed through a combination of standard negation marked on the verb and a conjunction. The disjunctive ‘or’ is used, but we more often find the coordinator ‘and’ of which additive and scalar meaning extensions (Andrason 2019) feed into connective negation. In Mongo, a Bantu language spoken in the DRC, the coordinator **la** ‘and, also, even’ (Hulstaert 1957:1097) combines with standard negation to express ‘neither...nor’.

(1) Mongo (Ruskin n.d.:512)

m- <b>pao</b> -kel-a	<b>la</b>	mbil'ene	<b>la</b>	lobi
SM1SG-NEG.FUT-do-FV	even	today	even	tomorrow
'I will <b>not</b> do it <b>neither</b> today <b>nor</b> tomorrow'				

Ambiguity with conditional and epistemic possibility is attested as well. In the Tanzanian language Rimi, for example, the conjunction **mbésá** used in connective negation is primarily used as a conditional (Olson 1964:239-241). In Nyakyusa, another Tanzanian Bantu language, the connective negation conjunction **pamo** can also be used as an epistemic possibility adverb expressing ‘maybe’ (Felberg 1996, Persohn 2017).

Borrowed conjunctions occur too. They either follow the prevailing non-dedicated pattern, like **àtá** in Kituba, a creole spoken in the the DRC and Congo Brazzaville (Fedherau 1992:2), derived from Arabic **hattá** ‘until’ but used in connective negation through its meaning extension ‘even’, or they borrow an inherently negative element like **né** in Mozambican Changana (Sitoe 2017:532), from Portuguese **nem** ‘nor’, or **wala** in Swahili and other East African Bantu languages, from Arabic **wala** ‘nor’ (Alruwali 2018). The inherently negative conjunctions, just like the non-negative ones, combine with a negative main verb (2a).

An investigation of the use of **wala**, both in Swahili varieties (like the Mozambican Swahili language Shangaji) and a historic Swahili corpus (de Schryver 2024) shows that the inherent negative meaning of **wala** (2a) tends to fade out allowing it to be used in affirmative utterances as well, either as a contrastive coordinator (‘but’) or as a disjunctive (‘or’) (2b). We hypothesize that the loss of negative meaning is facilitated by a strong tendency in Bantu languages for the expression of polarity to be assigned to the verb. This hypothesis is tested against a remarkably similar loss of negative meaning of **wala** in Sango, a lingua franca derived from the Ubangian language Ngbandi. In Sango **wala** is exclusively used as ‘or’ and mainly occurs in affirmative contexts (Samarin 2014).

(2) Shangaji

a. nakháraanga	<b>kha-ńttó</b> -zaliíwa	<b>waalá kha-ń</b> -khéyííwa
----------------	--------------------------	------------------------------

1a.sweet.potato      NEG.SM1-PRS.FOC-deliver.PASS (n)or    NEG.SM1-PRS-cut.PASS  
 'A sweet potato **neither** comes out of a seed **nor** out of a cutting.'

b. khúlá mwiiwa      o-ń-vúluwelaá      vaáthi    **waalá** likhóngóolo  
 each    3.fishbone      SM3-PRS-fall.APPL.REL    16.ground (n)or    5.bone  
 li-ń-vúluwelaá      vaáthi      yeéne mpákhá      a-thuúl-e  
 SM5-PRS-fall.APPL.REL    16.ground      s/he      until      SM1-take-SBJV  
 'Every fishbone **or** animal bone that has fallen on the ground, he has to take it'

#### References:

- Alruwali, Shatha. 2018. *Negation in Turaif Arabic: Not the last word*. University of Essex PhD dissertation.
- Andrason, Alexander. 2019. The conjunctive coordinator NA in Xhosa. Its categorial status and map of polyfunctionality. *Italian Journal of Linguistics* 31(1). 3-60.
- de Schryver, Gilles-Maurice. 2024. A Quarter Century of Swahili Corpus Building (1999-2024). Paper presented at the Start-up Meeting of the Modality in Swahili project, University of Gothenborg, 07/05/2024.
- Fehderau, Harold Werner. 1992. *Dictionnaire kituba (kikongo ya leta)-anglais-français et vocabulaire français-kituba*. Kinshasa: Editions CEDI.
- Felberg, Knut. 1996. *Nyakyusa-English-Swahili and English-Nyakyusa Dictionary*. Cape Town: Mkuki na Nyota Publishers.
- Hulstaert, Gustaaf. 1957. *Dictionnaire lomongo-français*. Tervuren: Musée Royal du Congo Belge.
- Olson, Howard S. 1964. *The Phonology and Morphology of Rjimi*. Hartford Seminary Foundation PhD dissertation.
- Persohn, Bastian. 2017. *The verb in Nyakyusa: A focus on tense, aspect and modality*. (Contemporary African Linguistics, 2.) Berlin: Language Science Press.
- Ruskin, Edward Algernon & Lily Adèle Ruskin. 1928. *Dictionary of the Lomongo language: Lomongo-English-French, English-Lomongo*. London: Christian Literature Soc.
- Samarin, William J. 2014. Swahili in Central African contact and colonization. In *In and Out of Africa. Languages in Question. In Honour of Robert Nicolai: Volume 2. Language Contact and Language Change in Africa*, 209-249. Louvain: Peeters.
- Sitoe, Bento. 2017. *Dicionário Português-Changana*. Maputo: Texto Editores.

# Connective negation in Romanian: a diachronic corpus study

Mihaela Iliaia  
(University of Ghent)

Keywords: negation, connective negation, Romanian, negative spread, negative concord

The present paper investigates structures containing connective negation (CoNeg) in Romanian, as illustrated in the examples in (1a–b). Recall that Romanian is a Romance language geographically surrounded from all sides by languages from different language families such as Finno-Ugric (Hungarian) and Slavic (Ukrainian, Bulgarian, and Serbian), with which it has continuously been in contact.

- (1) a. *Nici Maria nici Ștefan nu au venit.* (Phrasal use)  
 CONEG Maria CONEG Ștefan NEG.not have.3PL come  
 ‘Neither Maria nor Ștefan came.’
- b. *Nici Maria nu mi-a scris, nici eu nu am sunat-o.* (Clausal use)  
 CONEG Maria NEG.not me.DAT-has written CONEG I NEG.not have.1SG  
 called-her.ACC  
 ‘Maria neither has written to me nor did I call her.’

While Romanian negation and negative concord have been studied extensively, connective negation remains underexplored. This paper aims to analyze its evolution from the 16th to the 21st century, assessing whether Giannakidou's (1998, 2006) approach of negation applies to Romanian. Giannakidou's framework involves factors such as negative concord types, the use of N-words (negative polarity items, negative concord items, negative quantifiers), and their possible non-negative readings, and, based on these factors, Romanian shares certain features with both South-Slavic languages (cf. strict negative concord, Cristescu 2020: 54) and Romance languages (does not allow non-negative readings or negative spread).

This study aims at investigating structures with connective negation in Romanian by means of a corpus study both from a diachronic and a synchronic perspective. For the diachronic study, the *Pre-21st c. Romanian Corpus* (Iliaia 2020, 2023) is used, while for the synchronic study, I work with the *roTenTen21 Web Corpus*. Using an advanced CQL query, 200 examples per century are extracted, aiming to retain 100–150 relevant ones after noise removal.

This corpus study addresses research questions such as: (i) how does CoNeg interact with negative polarity and concord in Romanian, knowing that clausal CoNeg does not always align with non-connective clausal negation (e.g., Croatian, cf. van der Auwera 2021: 206)?; (ii) is there variation in how negators are realized in Romanian?; (iii) are Romanian CoNeg markers semantically ambiguous?; and (iv) are there significant historical changes in Romanian based on the collected data?

The data reveal changes over time: in the 16th c., there are very few cases of non-strict negative concord, which disappeared in the 21st c.; similarly, clausal CoNeg was more frequent and flexible in the 16th c., while in the 21st c., phrasal CoNeg is more common and dropping one negator is less acceptable.

## References

- Cristescu, Mihaela. 2020. Negation in the Aromanian dialect of the Romanian language: a typological analysis, *Fonetica și Dialectologie* XXXIX, București. 47–58.
- Giannakidou, Anastasia. 1998. Polarity Sensitivity as (Non)Veridical Dependency. *Linguistik Aktuell* Vol. 23, Amsterdam & Philadelphia: John Benjamins. xiv–281.
- Giannakidou, Anastasia. 2006. N-words and Negative Concord, in Martin Everaert et al. (eds.). *The Blackwell Companion to Syntax* vol. III. 327–391.
- Ilioiaia, Mihaela. 2020. Productivity of the Romanian mihi est pattern. *Revue Roumaine de Linguistique* 65(1), 49–67.
- Ilioiaia, Mihaela. 2023. *The MIHI EST construction: An instance of non-canonical subject marking in Romanian*. Berlin and Boston: De Gruyter, 317 p.
- van der Auwera, Johan. 2021. Quirky negative concord: Croatian, Spanish and French *ní*'s. *Jezikoslovlje* 22.2. 195–225.

# Connective negation in Baoding (Sinitic): forms, functions, and typological insights

Na Song

(INALCO and CRLAO-CNRS)

**Keywords:** Baoding Mandarin, Sinitic language, connective negation, correlative constructions, negation typology

This study investigates connective negation in Baoding Mandarin. It offers original data from an under-described Sinitic variety and contributing to typological debates on the form and function of negative coordination. Baoding employs two major strategies for connective negation: (i) a strategy shared with Standard Mandarin, combining negation with the additive adverb *iɛ*<sup>213</sup> ‘also’ (e.g., *pu*<sup>45</sup> + V, *iɛ*<sup>213</sup> + *pu*<sup>45</sup> + V); and (ii) a distinct bisyndetic strategy, in which identical negators are used before both conjuncts. This includes the general negator *pu*<sup>45</sup>, the negative existential *me*<sup>22</sup>, and the prohibitive *piɛ*<sup>51</sup>, as in:

- (1) *pu*<sup>45</sup> *tɕ<sup>h</sup>ɿ*<sup>45</sup> *pu*<sup>45</sup> *xɿ*<sup>45</sup> = *iã* *xa*<sup>51</sup> *ɕi*<sup>22</sup> = *iã*  
 NEG<sub>GEN</sub> eat NEG<sub>GEN</sub> drink = INT<sub>P.MIR</sub> DEM OK = INT<sub>P.MIR</sub>  
 ‘(You) neither eat nor drink? Is that okay?’
- (2) *to*<sup>45</sup> *sɿ*<sup>51</sup> *ɕɿ*<sup>22</sup> = *lɛ* *me*<sup>22</sup> *tɕ<sup>h</sup>ɿ*<sup>45</sup> *me*<sup>22</sup> *fã*<sup>22</sup> *me*<sup>22</sup> *ɕi*<sup>213-21</sup> *fɿ*  
 Already forty = COS NEG<sub>EXI</sub> car NEG<sub>EXI</sub> house NEG<sub>EXI</sub> wife  
 ‘(You are) already 40 years old, you have neither a car, nor a house, nor a wife.’
- (3) *ni*<sup>213</sup> *piɛ*<sup>51</sup> *tsɔ*<sup>45</sup> *piɛ*<sup>51</sup> *nɔ*<sup>51</sup>  
 2SG NEG<sub>PROH</sub> make.trouble NEG<sub>PROH</sub> cause a fuss  
 ‘Don’t make trouble or cause a fuss!’

Example (1) illustrates correlative negation with the general negator *pu*<sup>45</sup>, applied in verbal domains with emphatic force; Example (2) employs the negative existential marker *me*<sup>22</sup>, forming additive negation in the nominal domain—highlighting absence of multiple items. Example (3) is particularly noteworthy for showing that Baoding allows connective negation in prohibitive contexts—a domain typically excluded from correlative negation in languages like English *neither...nor....* The prohibitive use represents a typologically rare extension of connective negation into directive speech acts.

The use of bisyndetic negators in Baoding differs from Standard Mandarin, which typically uses asymmetric constructions like *bù...yě+bù...* or *méi(yǒu)...yě+méi(yǒu)...*, where *yě* ‘also’ functions as an additive adverb. By contrast, Baoding’s constructions — *pu*<sup>45</sup>... *pu*<sup>45</sup> ..., *me*<sup>22</sup>...*me*<sup>22</sup> ..., and *piɛ*<sup>51</sup>...*piɛ*<sup>51</sup> ... —show more syntactic symmetry and emphatic marking.

The findings of this study are based on personal fieldwork. These data were collected through recordings with native speakers in Baoding (Hebei province, China) as well as elicitation. This study addresses typological issues central to connective negation, including (i) scopal variation in Baoding’s negation markers, (ii) functional overlap with standard clause negation, and (iii) distributional flexibility within correlative and additive contexts (Haspelmath 2007, Jespersen 1917), including clause types such as imperatives and rhetorical questions.



Baoding also demonstrates complex combinations of negators, layering pragmatic meaning through emphatic and correlative connective negation forms (Déprez & Espinal 2020). Drawing on a broader typological framework, this study examines how these constructions interact with negative polarity and negative concord, contrasting them with systems of connective negation in other languages (Bond 2011, Gianollo 2018).

## References

- Bond, Oliver (2011), Negation in clause linkages. In *Language documentation and description*, Vol. 9, edited by Julia Sallabank, 77-120. London: School of Oriental and African Studies.
- Déprez, Viviane, and M. Teresa Espinal, eds. (2020), *The Oxford Handbook of Negation*. Oxford: Oxford University Press.
- Gianollo, Chiara (2018), *Indefinites between Latin and Romance*. Oxford: Oxford University Press.
- Haspelmath, Martin (2007), Coordination. In *Language typology and syntactic description*, Vol. 2, edited by Timothy Shopen, 1-51. Cambridge: Cambridge University Press.
- Jespersen, Otto (1917), *Negation in English and other languages*. København: Høst & Søn.
- Salaberri, Iker (2022), A cross-linguistic study of emphatic negative coordination. *Studies in Language* 46: 647-717.
- van der Auwera, Johan, and Sepideh Kookhan (2022), Extending the typology: Negative concord and connective negation in Persian. *Linguistic typology at the crossroads* 2: 1-36.

## Tracing connective negation in Balkan Languages: Areal convergence and divergent developments

Nikos Liosis  
(Aristotle University of Thessaloniki)

Keywords: connective negation, balkanism, bisyndesis, negative concord, (scalar) additivity

This typologically oriented talk examines connective negation (CN) in the Balkan languages. The geographical clustering of languages with dedicated connective negation markers (CNMs) indicates that contact is generally a strong trigger for their development (Salaberri 2022:677). Moreover, there is significant matter and pattern replication concerning the forms and types of standard negation and negative concord within the Balkan Sprachbund in particular (cf. Xherija 2015, Joseph 2020, 2022, a.o.). Therefore, the working hypothesis in this talk is that similar convergence phenomena are observed in the Balkans regarding CN. This hypothesis is tested with respect to three variables:

(a) CNMs optionally conform to the Balkan strict negative concord type (cf. van der Auwera et al. 2021:56–7). The term “optionality” is of key importance here, as the systematic asymmetries in the occurrence of CNMs and standard negation that arise are not merely distributive in nature (e.g., between the pre-verbal and post-verbal parts of the sentence) but also seem to be influenced by pragmatic or cognitive factors, such as, for instance, register variations, the linear distance between CNM and negation, etc.;

(b) the bisyndetic (‘neither...nor’), additive (‘neither’), and scalar additive (‘not even’) functions of CNMs (in line with Forker 2016) are co-expressed: For example, Arom. *neca* (Papahagi 1974: 871), Alb. *as* (Kostallari 1980: 60), Gr. *oute* (Barouni 2018) etc., encompass all three functions;

(c) CNMs participate in the univerbation of negative indefinites with shared morphological patterning, e.g., the collocation ‘not even one’ results in Alb. *as-një* ‘no one’, Med.Gr. *oud-enas*, Arom. *niṭi-un* etc.

Preliminary findings seem to confirm the hypothesis and suggest that the most likely origin of the diffusion of the observed similarities is Greek, where all these features have developed long ago (Liosis 2024). On the other hand, there are some significant idiosyncrasies, e.g. in Albanian *as...as që* the second negator is sometimes reinforced by an element which is homonymous with the complementizer *që* ‘that,’ in Vlax Romani the indefinite *či* evolved into a CNM and eventually into preverbal standard negation under the influence of Romanian CNM *nici* (Matras 2005: 189), in Greek and Balkan Turkish the verb can intervene between two coordinands, as in Persian (van der Auwera and Koohkan 2022) etc. Therefore, micro-variation analysis shows that there is indeed a common core regarding the morphological, syntactic, and pragmatic characteristics of connective negation in the Balkan languages, but there are also particularities of theoretical and typological interest that highlight the need for continued research in this direction.

### Select References

- Barouni, Maria (2018), *Topics in the Syntax–Semantics of Greek particles*, PhD dissertation, Rethimno: Crete University.
- Forker, Diana (2016), Toward a typology for additive markers, *Lingua* 180, 69–100.
- Joesph, Brian (2020), Contact in the Balkans, in R. Hickey (ed.), *The handbook of language contact*, New Jersey: John Wiley and Sons Ltd, 537–550.

- Joseph, Brian (2022), Some Positive thoughts on Albanian Negation, in D. Kyriazis (ed.), *Albanologu i arvanitëve. Festschrift kushtuar 80-vjetorit të lindjes së Akad. Titos Jochalas*, Tirana: Akademia e Shkencave e Shqipërisë, 196–200.
- Kostallari, Androkli (1980), *Fjalor i gjuhës së sotme shqipe*, Tirana: Instituti i Gjuhesisë dhe i Letërsisë.
- Liosis, Nikos (2024), Negative concord in Medieval Greek: lexical asymmetry and the role of correlative negation, in J. van der Auwera and Ch. Gianollo (eds.), *Negative Concord: A Hundred Years On*, Berlin: de Gruyter, 175–206.
- Matras, Yaron (2005), *Romani: A linguistic introduction*, Cambridge: Cambridge University Press.
- Papahagi, Tache (1974), *Dictionarul dialectului aroman general si etimologic*, Bucuresti: Editura Academiei Republicii Socialiste Romania.
- Salaberri, Iker (2022), A cross-linguistic study of emphatic negative coordination, *Studies in Language* 46(3), 647–717.
- van der Auwera, Johan, et al. (2021), Connective negation and negative concord in Balto-Slavic, in P. Arkadiev, et al. (eds.), *Studies in Baltic and other languages. A Festschrift for Axel Holvoet on his 65th birthday*, Vilnius: Vilnius University Press, 44–64.
- van der Auwera, Johan and Sepideh Koohkan (2022), Extending the typology: negative concord and connective negation in Persian, *Linguistic Typology at the Crossroads* 2, 1–36.
- Xherija, Orest (2015), Weak and strong NPIs: nobody and anybody in Albanian and Modern Greek, in C. Brown, et al. (eds.), *Proceedings of the 15th Texas Linguistic Society*, Austin: University of Texas, 184–201.

**WS3 Diachronic Studies on Minoritised and Under-researched Romance Varieties**

**Marc Olivier & Afra Pujol i Campeny**

## Old Romanian DPs' head-finality: fact or fiction?

The typological shift from head-finality, whereby complements (e.g., adjectives/genitives, direct objects) systematically precede heads (e.g., nouns, verbs) (1), to head-initiality, where heads are followed by their respective complements (2), is well documented for various stages of Latin and Romance varieties (cf. Ledgeway 2012).

- (1) *si seruus*                      *[[furtum]<sub>NP</sub>]*                      *faxit<sub>VP</sub>*  
 if slave.NOM                      robbery.ACC                      commit.SBJ.AOR.3SG  
 'if the slave were to commit a robbery' (Latin, *Leg XII Tab* (Vlp.Dig.9.4.2.1))
- (2) *et si*                      *[<sub>VP</sub> scrijbes*                      [...] [<sub>NP</sub> epistulam]]  
 and if                      write.IND.FUT.2SG                      letter.ACC  
 'and if you were to write [me] a letter' (Latin, Ter.467.25)

Nevertheless, the developments encountered in early stages of Romanian are less straightforward, inasmuch as there are no extant texts before the beginning of the sixteenth century. While the syntax seen in the earliest texts is overwhelmingly head-initial, traces of head-finality have been previously reported in the literature (Brăescu, Dragomirescu, and Nicolae 2015; Nicolae 2019). Now, if Romanian were to still show these relics at such a late date, it would significantly represent a unique case within Romance.

In our presentation we aim to zoom in on old Romanian determiner phrases (DPs), focusing on how they can be captured from a typological point of view (i.e., head-final vs head-initial constructions) and assessing which types of texts (e.g., translations from Old Church Slavonic/Greek, non-translated texts), if any, show traces of head-finality. Crucially, our in-depth analysis departs from previous studies in that we put forward a holistic approach, considering not only the position of nouns with respect to adjectives (cf. 3-4), possessives, and genitives, but also the syntax of various determiners, such as definite articles. Deliberately glossing over details, head-initial varieties are taken to overtly lexicalized functional categories such as, but by no means limited to, definite articles and auxiliaries (Dryer 2009: 205).

- (3) *legiei*                      *creştinească*  
 law=the.GEN                      Christian.F.SG.NOM  
 'the Christian religion' (old Romanian, CC<sup>2</sup>.1581: 6)
- (4) *Dumnezeiasca*                      *slujba*  
 Holy.F.SG=the.NOM                      service=the.NOM  
 'the Holy service' (old Romanian, CL.1570: 7<sup>v</sup>)

Among others, a vanishing head-final syntax would imply a dwindling use and incomplete grammaticalization of definite articles, which would be, for example, dropped in DPs expressing inalienable possession, as well as in those containing abstract or generic nouns, as seen in other early Romance varieties, such as fifteenth century French (5) and fourteenth century Laziale variety (6) (cf. also Lombardi 2007).

- (5) *que*                      *vos*                      *portoiz*                      *mon*                      *cors*  
 that                      you                      carry.SBJ.PRS.2PL                      my                      body  
 'that you carry my body' (French, *Le Mort d'Arthur* 29-30)

(6) *Lassai*                      *mea*                      *cappa*  
 leave.IND.PRT.1SG    my                      cape  
 ‘I left my cloak’ (Laziale, *Cronica dell’Anonimo Romano* 183.14)

Thus, the aims of our presentations are as follows: (i) to analyze old Romanian data from a typological perspective, focusing on the head directionality parameter; (ii) to offer a thorough description of the syntax of old Romanian definite articles, which are to be taken as ultimately instantiating a strong argument in favour of the already generalized head-initiality; (iii) to provide a principled explanation for different word orders encountered in old Romanian DPs, taking into account both internal factors, e.g., the head-initiality, and external factors, e.g., non-native structures produced by translators.

#### (SELECTED) REFERENCES

- Brăescu, R., A. Dragomirescu, and A. Nicolae (2015). ‘(Non-)configurationality and the internal syntax of adjectives in old Romanian’. *Bucharest Working Papers in Linguistics* 17: 55–74.
- Dryer, M. (2009). ‘The Branching Direction Theory of Word Order Correlations Revisited’. In: Scalise, Magni, and Bisetto (eds), *Universals of Language Today*. Berlin: Springer, 185–207.
- Ledgeway, A. (2012). *From Latin to Romance. Morphosyntactic Typology and Change*. Oxford: Oxford University Press.
- Lombardi, A. (2007). ‘Definiteness and possessive constructions in medieval Italo-Romance’. In: Lepschy and Tosi (eds), *Histories and Dictionaries of the Languages of Italy*. Ravenna: Longo, 99–118.
- Nicolae, A. (2019). *Word Order and Parameter Change in Romanian. A Comparative Romance Perspective*. Oxford: Oxford University Press.

# The Diachrony of Progressives in Venetan: A Syntactic Analysis

Alberto Frasson  
(University of Wrocław)

Keywords: progressive, aspect, prepositional infinitive, Italo-Romance, diachrony

This paper examines the evolution of the progressive construction in Venetan, a north-eastern Italo-Romance variety with a long literary tradition, spoken in a diglossic context alongside Italian, with no official recognition (Berruto 1987).

Venetan uses progressive constructions made up of a *be* form followed by an infinitive introduced by the preposition *drio* ('behind').

- (1) Son                    **drio**    **lavor-ar**.  
be.PRS.1SG        **behind** **work-INF**  
'I am working.'

This structure is independently attested with progressive readings in the diachrony of Venetan:

- (2) (...)    et        **dreo**    **disnar**    elo        me        dis-e (...)  
              and       **behind** **eat-INF** he        me        say-PRS.3SG  
'(...) and while eating, he tells me (...)'

*Commemoriali* (14th century; in Tomasin 2013)

Until the 18th century, the structure allowed for the pronominalization of the infinitive (3) or its replacement with a DP (4), while retaining a progressive reading. Furthermore, *drio* also functioned as a regular locative preposition (5):

- (3) A-ve-u            feni-o            de        mett-er zo        la    scrittura? - **Ghe son**        **drio**.  
have-2PL-you    finish-PTCP    of        put-INF down    the writing    **it** **be.PRS.1SG** **behind**  
'Are you done writing? – I am doing it.' (*lit.*: 'I am behind it.')

*Sior Todero Brontolon* (Goldoni, 18th century)

- (4) Xe        debotto        do        mesi    che        **sè**                **drio**    **sta**        **fabbrica**.  
be-PRS.3 almost        two        months that    **be-PRS.2PL**        **behind** **this**        **work**  
'You have been working on this for almost 2 months.' (*lit.*: 'You are behind this work.')

*La Casa nova* (Goldoni, 18th century)

- (5) Chi        xe                    **drio**    quella    portiera?  
who    be.PRS.3                **behind** that    door  
'Who is there behind that door?'

*Sior Todero Brontolon* (Goldoni, 18th century)

In Modern Venetan, pronominalization (6) and replacement (7) are not possible. Moreover, *drio* no longer behaves as a preposition: in locative contexts, the complex preposition *da-drio de* ('from-behind of') is used, which cannot receive a progressive reading:

(6) (\*Ghe) son drio \*(lavor-ar).  
 It be.PRS.1SG behind work-INF  
 'I am working (*lit.* I am behind work).'

(7) Sè \*drio / da-drio de sta fabrica.  
 be.PRS.2PL behind from-behind of this factory.  
 'You are behind this factory.'

I propose that the Old Venetan infinitive acted as a secondary predicate in a small clause (SC) introduced by the preposition *drio* (Raposo 1989). This structure can be represented as follows:

(8) [TP Son [PP drio [SC [VP lavorar]]]]

The PP headed by *drio* introduces a small clause, which contains a VP headed by the infinitive, predicating an action that the subject of the matrix clause is performing. Following Safir (1983), I analyze small clauses as constituents, which accounts for the availability of pronominalization and replacement.

In Modern Venetan, *drio* has grammaticalized into a progressive aspectual marker (Casalicchio 2019), losing its prepositional use. The infinitive is directly integrated into the matrix clause:

(9) [TP Son [AspP drio [Asp lavorar [VP [V ~~lavorar~~]]]]]

In this configuration, *drio* occupies Spec-AspP, while the infinitive undergoes head movement to Asp, which blocks pronominalization and replacement.

This grammaticalization aligns with Roberts and Roussou (2003): grammaticalization involves the reanalysis of functional heads in the clausal spine and entails loss of syntactic independence, structural simplification and upward reanalysis of prepositions as functional heads.

## References

- Berruto, G. 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.
- Casalicchio, J. 2019. Gerunds become prepositional infinitives in Romance Small Clauses: The effects of later Merge to the syntactic spine. *Probus* 31-1: 75-117.
- Raposo, E. 1989. *Prepositional infinitival constructions in European Portuguese*. In Jaeggli, O. & K. Safir (eds.), *The null subject parameter*. Dordrecht: Kluwer, 277–305.
- Roberts, I. and A. Roussou. 2003. *Syntactic change*. Cambridge University Press.
- Safir, K. 1983. On Small Clauses As Constituents. *Linguistic Inquiry* 14-4: 730-735.
- Tomasin, L. 2013. Quindici testi veneziani 1300-1310. *Lingua e stile* 48: 3-48.



# Object clitics and zero anaphora in Caribbean French: Changes of the referential paradigm through language contact with Creole

Amalia Canes-Nápoles      Sophie Repp  
UNIVERSITY OF COLOGNE

## KEYWORDS:

objects, morphosyntactic variation, Caribbean French  
corpus linguistics, language change

The use of object clitics for reference to given objects in spoken French in the Caribbean (CarFr) reflects ongoing processes of language change. For Guadeloupe and Martinique, it has been observed that given objects that in Hexagonal French (HxFr) would be referred to by a clitic are not expressed (Kriegel & Ludwig 2018, 81):

- (1) nous allons remettre **un sac** et nous allons  $le_{HxFr}/\emptyset_{GuFr}$  mettre à la presse  
we will place INDF bag and we will PRO:CL-ACC.3SG place in the press  
We are going to place a bag and we are going to press it

This departure from HxFr has been attributed to interference from the local creole languages (Bellonie 2008; Pustka 2007; Damoiseau 2003). Another feature observed for CarFr is the reduction of the clitic paradigm in terms of case and gender distinctions. (2) illustrates the elimination of the accusative/dative distinction at the pronominal level (Damoiseau 2003, 52-53). In a way, the omission of given objects as in (1) can be interpreted as an extreme reduction of the pronominal paradigm: neither case nor gender are expressed. The elimination has also been related to changes in verb transitivity, i.e., argument structure (Manessy 1994; Damoiseau 2003). However, these hypotheses face critical challenges. In Guadeloupean and Martinican French, object clitics—though underspecified for gender—do exist (see (3) taken from Ludwig et al. 2002, 22), countering claims of creole-induced simplification. Furthermore, the local creole systems themselves differentiate direct from indirect NP objects, e.g., through prepositions, which speaks against changes in argument structure.

- |  |                                  |
|--|----------------------------------|
| (2) Il $lui_{HxFr}/\emptyset_{GuiaFr}$ écrit | (3) An ka ba'w <b>li</b> .       |
| 3SG PRO:CL-DAT.3SG write                     | 1sg PROG give.2SG PRO:CL-ACC.3SG |
| He writes him                                | 'I am giving it to you.'         |

In general, the absence of pronominal anaphora in CarFr has been sparsely discussed, apart from isolated studies addressing the challenges posed by pronouns like *y* and *en* for Creole-speaking learners (Bellonie 2008; Pustka 2007). Moreover, the existing claims about object clitics (or lack thereof) await support from robust empirical support, particularly regarding characteristics of the referents denoted by the clitics, including their cognitive status (referent activation), semantic characteristics (competing referents with overlapping semantic features), and discourse status (topicality /focus), which likely shape variation in their use.

This study contributes to filling these gaps through a corpus-based analysis of unscripted radio programmes from Guadeloupe and Martinique: five programmes, each lasting approximately 26–30 minutes, featuring one moderator and 12 speakers with a balanced gender distribution, each contributing an estimated 2–5 minutes to share their opinions on the programme’s topic. Our work examines how language-internal and language-external factors influence the varied expression of object arguments and proposes that Caribbean French varieties exhibit an expanded system of referential forms incorporating both clitics and zero anaphora. Additionally, it argues for a reorganization of the functional mapping of anaphors, confirming that pronominal systems in Romance languages are particularly susceptible to instability in contexts of language contact (Léglise 2013). Crucially, this analysis challenges deterministic views of paradigm reduction as an inherent tendency within French, instead highlighting the dynamic interaction of contact-induced and context-dependent factors.

## REFERENCES

- Bellonie, J.-D. (2008). De l'intérêt de corpus diversifiés pour la réflexion sociolinguistique et la didactique du flm en martinique... et ailleurs. *Verbum*, 30(4), 287–298.
- Damoiseau, R. (2003). *Eléments de grammaire comparée français-créole*. Cayenne: Ibis Rouge Editions.
- Kriegel, S., & Ludwig, R. (2018). Le français en espace créolophone–guadeloupe et seychelles. *Romanistisches Jahrbuch*, 69(1), 56–95.
- Léglise, I. (2013). The interplay of inherent tendencies and language contact on french object clitics. *The Interplay of Variation and Change in Contact Settings*, 137–163.
- Ludwig, Ralph, Montbrand, Daniele, Pouillet, Hector, & Telchid, S. (2002). *Dictionnaire créole-français: Avec un abrégé de grammaire créole et un lexique français-créole*. Paris: SERVEDIT – Maisonneuve & Larose – Éditions Jasor.
- Manessy, G. (1994). *Le français en afrique noire. mythe, stratégies pratiques*. Paris: L'Harmattan.
- Pustka, E. (2007). Le mythe du créole l1. *Romanistisches Jahrbuch*, 57, 60–83.

## *Verb position and clitic placement in Old Catalan*

**Aim:** To offer a description of the evolution of sentential word order and clitic placement in Old Catalan (11<sup>th</sup>-14<sup>th</sup>c) by demonstrating the connection between the two phenomena, thus updating preceding studies (Fischer 2002; Pujol i Campeny 2018), and widening the datapool used in them by relying on CICA (Torruella et al. 2009).

**Verb position:** We follow Pujol i Campeny (under review) in distinguishing two stages within Old Catalan: **(a)** Early Old Catalan (10-12<sup>th</sup>c), when the language exhibited a relaxed V2 system (akin to that of contemporary neighbouring varieties), with verb movement to the low left periphery, the systematic appearance of a preverbal constituent, and expletive *si* (Poletto 2005; Wolfe 2018). **(b)** Later Old Catalan (13<sup>th</sup>-14<sup>th</sup> c), when the language starts losing V-to-C in favour of V-to-T due to the loss of **(i)** expletives and **(ii)** unmarked XVS strings (now used in informationally specific contexts connected to contrast and veridicality), a key cue for the acquisition of V2 (Pujol i Campeny 2024). This innovation sets Catalan aside from neighbouring Romance varieties, which do not start losing V2 until the 15<sup>th</sup>c.

**Clitic placement:** With respect to clitic placement, the data analysed indicate that, despite the application of the Tobler Mussafia Law (TML), many exceptions are found in Early Old Catalan (11<sup>th</sup>c): unlike Ibero-Romance varieties, we find cases involving absolute first position by the 13<sup>th</sup>c; when we also observe that clitic placement became sensitive to information structure, leading to a reanalysis of the TML in favour of an informationally driven system (as described by Batllori et al. 2005), drifting from the Romance continuum.

**Bringing the two phenomena together:** We claim that the loss of V2 is connected to the loss of TML. As shown in Table 1, Old Catalan develops informationally sensitive clitic placement as it loses unmarked V-to-C and unmarked XVS sequences in the 13<sup>th</sup>c, after which proclisis becomes associated to XV(S) contexts where the preverbal constituent is focalised. Interestingly, in the 14<sup>th</sup>c, the type of foci encoded by XV(S) strings diminishes, while proclisis starts becoming the default option. These results render Catalan an innovative variety sitting in the centre of the Romania. In this sense, we hypothesise that the presence of a clitic placement system sensitive to information structure attested in Western Ibero-Romance (Batllori et al. 2005) could be an east-to-west spread of the innovation.

	11 <sup>th</sup> century	12 <sup>th</sup> century	13 <sup>th</sup> century	14 <sup>th</sup> century
<b>Verb position</b>	<ul style="list-style-type: none"> <li>- Systematic V-to-Fin</li> <li>- All foci types in the left periphery</li> <li>- All topic types in the left periphery</li> <li>- Expletive <i>si</i> à la Oil</li> </ul>	<ul style="list-style-type: none"> <li>- Systematic V-to-Fin</li> <li>- All foci types in the left periphery</li> <li>- All topic types in the left periphery</li> <li>- No expletive <i>si</i> but cases of expletive <i>e</i></li> </ul>	<ul style="list-style-type: none"> <li>- Unmarked V-to-T</li> <li>- Informationally and stylistically marked V-to-Fin and XVS</li> <li>- No expletives</li> </ul>	<ul style="list-style-type: none"> <li>- Unmarked V-to-T</li> <li>- Informationally marked V-to-Fin and XVS (reduction of contexts vs 13<sup>th</sup> century)</li> <li>- No expletives</li> </ul>
<b>Clitic placement</b>	<ul style="list-style-type: none"> <li>- Tobler Mussafia Law clitic placement</li> </ul>	<ul style="list-style-type: none"> <li>- Tobler Mussafia Law clitic placement</li> <li>→ but proclisis possible after conjunctions <i>e</i> ‘and’ and <i>o</i> ‘or’</li> </ul>	<ul style="list-style-type: none"> <li>- Information structure-sensitive clitic placement</li> <li>- First cases of clitic placement in absolute first position</li> </ul>	<ul style="list-style-type: none"> <li>- Information structure-sensitive clitic placement</li> <li>→ but incipient generalised proclisis in discourse topic continuity contexts</li> </ul>

*Table 1* - Summary of changes in verb position and clitic placement in Catalan, 11<sup>th</sup>-14<sup>th</sup>

## References:

- Batlloori, M., Iglesias, N., & Martins, A.M. 2005. Morfologia i sintaxi dels clítics pronominals en català medieval. *Caplletra*, 38, 137–177.
- Fischer, S. 2002. The Catalan Clitic System. A Diachronic Perspective on its Syntax and Phonology. Berlin: Mouton de Gruyter.
- Poletto, Cecilia. 2005. Sì and e as Expletives in Old Italian. In Batllori, Montserrat & Hernanz, Maria Lluïsa & Picallo, Carme & Roca, Francesc (eds.), *Grammaticalization and Parametric Variation*, 206–235. Oxford: Oxford University Press.
- Pujol i Campeny, A. 2018. Word Order in Old Catalan. PhD thesis. University of Cambridge.
- Pujol i Campeny, A. 2024. ‘Fronting in Old Catalan: Narrative text and reported speech’. *Transactions of the Philological Society*. DOI: <https://doi.org/10.1111/1467-968X.12299>.
- Pujol i Campeny, A. Under review. ‘Word order in Early Old Catalan’. *Journal: Probus*.
- Torruella, J. Saldanya, M. Martines, M. 2009. *Corpus Informatitzat del Català Antic*.
- Wolfe, S. 2018. Probing the Syntax of a Problematic Particle: Old French ‘si’ Revisited. *Transactions of the Philological Society* 116(3). 332–362. (doi:10.1111/1467-968X.12123)

# The use of participles in an Old Gascon *corpus*: Some reflections

Barbara Francioni & Natalia Romanova  
(University of Caen Normandie)

Keywords: Old Gascon, Participles, Syntactic Annotation, Medieval Legal Texts, Universal Dependencies

The creation of *corpora* of annotated texts is fundamental for any historical linguistics research. Nevertheless, we still see a lack of resources for some varieties, especially in diachronic perspective. The use of deep-learning AI-technologies (Dozat & Manning 2017) offers a possible solution. Focusing on one concrete example, this paper aims at presenting the advantages of semi-automatic syntactic annotation by bootstrapping (Peng et al. 2022) of an Old Gascon *corpus* for linguistic research. To do so, we will show the annotation choices for past, present, and future participles with the help of some contextual examples.

Our *corpus* of medieval legal texts written in a variety of Old Occitan used in Gascony has been annotated using the Arborator platform (Guibon et al. 2020) and the integrated BertForDeprel (Guiller 2020) parser in the Universal Dependencies (UD) syntactic framework (de Marneffe et al. 2021). To date, it comprises six fully annotated and validated texts: *Coutumes et Privilèges de l'Entre-Deux-Mers* (Lépicier 1861), *Coutumes and Règlement municipal de Bagnères* (Soutras & Dejeanne 1882; Luchaire 1881), *Charte des Boucheries d'Orthez* (Tucóo-Chala 1957), *Fors et Charte d'Herrère* (Cheronnet & Dumonteil 1996), and *Les Fors Anciens de Béarn* (Ourliac & Gilles 1990).

The analysis of the use of different types of participles in the *corpus* will, firstly, allow us to present the annotation choices made to apply UD annotation strategies to a minoritised variety such as Old Gascon, which in itself represents a challenge, given the differences not only between modern and old Romance syntax, but also among old linguistic varieties such as Old French (for which UD guidelines already exist, see Prévost et al. 2024) and Old Gascon. Secondly, and most importantly, the study of documentary rather than literary texts implies dealing with the ambiguous nature of these sources, situated between the two poles of the conservative Latin model taught to notaries and men of letters and the completely innovative use of vernacular syntactic structures for which no graphic or syntactic standard existed at the time. Participles – either past, present, or future – are a striking instance of this kind of ambiguity, often presenting *ablativus absolutus*-like structures whose syntax and morphology has been adapted to the new vernacular style. This is especially true of forms classifiable as “future participles” that are generally assumed to have completely disappeared in both late medieval Latin and Romance languages by grammarians and linguists (Barbato 2017: 150). Their morphology (-*eyr(a)*, -*eir(a)* etc.), even if very similar to that of adjectives or nouns of the -ARIUS-suffix type (among others, Anglade 1921: 48-49), is clearly linked to their function as heads of Verb-Phrases (e.g. *talhadas donadeyras au Rey*, “taxes that have to/will be paid to the King”) and is a very interesting example of this conservative-innovative linguistic ambiguity. The possible occurrence of forms of the future participle will also be verified in other existing documentary *corpora* of Old Occitan, such as *Documents linguistiques galloromans* and *Chartes d'Occitanie*, in order to trace the limits of the circulation of these forms.

## References

- Anglade, J. (1921), *Grammaire de l'ancien provençal ou ancienne langue d'oc*, Paris: Librairie C. Klincksieck.  
Barbato, M. (2017), *Le lingue romanze. Profilo storico-comparativo*, Bari: Laterza.

*Chartes d'Occitanie*, URL: <https://chartoc.huma-num.fr/>.

- Cheronnet, B. & Dumonteil, J. (1996), Carta de poblacion de Herrera (21 de novémer de 1278), *Ligam/DiGaM*, 7, 5-8.
- De Marneffe, M.-C., Manning, C.D., Nivre & Zeman, D. (2021), Universal Dependencies, *Computational Linguistics*, 47(2), 255-308.
- Documents linguistiques galloromans (DocLing)*, dir. by Martin Glessgen. URL: <https://www.rose.uzh.ch/docling/>.
- Dozat T.C. & Manning D. (2017), Deep Biaffine Attention for Neural Dependency Parsing, in: *International Conference on Learning Representations (ICLR)*, URL: <https://arxiv.org/abs/1611.01734>.
- Guiller Kirian (2020), *Analyse syntaxique automatique du pidgin-créole du Nigeria à l'aide d'un transformer (BERT): Méthodes et Résultats*, Mémoire de Master, Paris, Sorbonne Nouvelle.
- Guibon G., Courtin M., Gerdes K. & Guillaume B. (2020), When Collaborative Treebank Curation Meets Graph Grammars, in: *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille: European Language Resources Association, 5291–5300, URL: <https://aclanthology.org/2020.lrec-1.651/>.
- Lépicié, J. (1861), Coutumes et privilèges de l'Entre-Deux-Mers, in: *Archives historiques du département de la Gironde*, 101-130.
- Luchaire, A. (1881), *Recueil de Textes de l'ancien dialecte gascon d'après des documents antérieurs au XVe siècle suivi d'un glossaire*, Paris: Maisonneuve.
- Ourliac, P. & Gilles, M. (1990), *Les Fors anciens de Béarn*, Paris: Éditions du Centre National de la Recherche Scientifique.
- Peng Z., Gerdes K., & Guiller K. (2022), Pull your treebank up by its own bootstraps, in: L. Becerra, B. Favre, C. Gardent, & Y. Parmentier (eds.) (2022), *Journées Jointes des Groupements de Recherche Linguistique Informatique, Formelle et de Terrain (LIFT) et Traitement Automatique des Langues (TAL)*, 139-153.
- Prévost S., Grobol L., Dehouck M., Lavrentiev A., & Heiden S. (2024), Profiterole : un corpus morpho-syntaxique et syntaxique de français médiéval, *Corpus*, 25, URL: <https://journals.openedition.org/corpus/8538>.
- Soutras, F. & Dejeanne, J.-M.-L. (1882), Fors et coutumes de Bagnères-de-Bigorre, *Bulletin de la Société Ramond*, 4: 1-16.
- Tucó-Chala, P. (1957), La charte des boucheries d'Orthez (2 novembre 1270), *Annales du Midi*, 69/40: 323-332.

## Adapting the AIS for geographic and phylogenetic research on language change dynamics

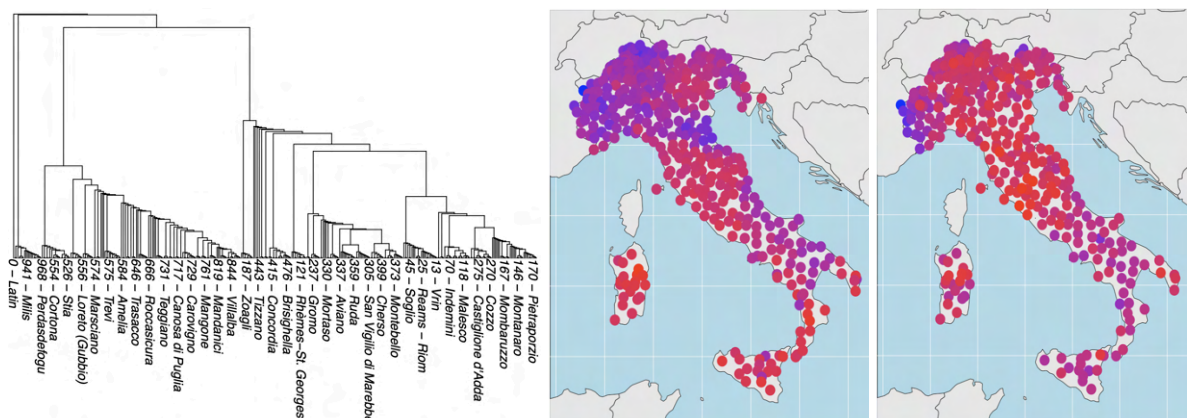
Borja Herce  
University of Zurich

Keywords: Romance, language change, lexicon, phonology, geography

Traditional dialectal surveys (e.g. AIS (Atlante Italo-Svizzero), ALF (*Atlas Linguistique de la France*), ALI (Atlante Linguistico Italiano), ALPI (Atlas Lingüístico de la Península Ibérica), etc.) are monumental works that document an enormous amount of information on the linguistic varieties spoken in hundreds of geographic locations. Despite the wealth of data (phonological, morphosyntactic, and lexical) that they have to offer, and despite our increasing ability to work with and process effectively large data volumes, these resources are underused in contemporary quantitative linguistic research. The main reason for this is that the methods and formats of most of these surveys predate the computer era and are not suited for automated manipulation. Inconsistent transcription practices, incomparable data points, etc. represent a persistent problem. Alongside digitalization, effort has to be put on increasing the computational usability of these resources to address questions related to language change and its predictors.

In this paper, I present an upgrade to the AIS data as digitized by Loporcaro et al. (2021). It consists of a clean-up and selection of variables, and the addition of Classical Latin and cognacy information. In total, 132,080 data points (349 meanings across 408 varieties) have emerged from this coding, each consisting of a phonologically transcribed word form, its meaning, variety/location, and cognate set. Specifically, the newly added information allows us to separate innovations from retentions and identify different types of change (i.e. lexical vs phonological). Thanks to this information, we are able to explore the division between languages and dialects (see Wichmann 2020), and ask if the amount/speed of change in different levels is correlated (see Greenhill et al. 2017, Heeringa & Hinskens 2011), and whether extralinguistic aspects like geographic isolation or population sizes can predict the rates of change in different varieties (Bromham et al. 2015, Trudgill 2020). Through the cognate set data, we also build a phylogenetic tree of the AIS varieties, which allows us to estimate (and discount where appropriate) the lack of independence brought about by shared ancestry.

Preliminary results suggest that the amount of change since Latin at the phonological and lexical levels is correlated. Conservativeness (in red in the maps below) reaches its highest levels in Nuorese and Calabrese varieties when it comes to phonology, and Nuorese and Tuscan when it comes to the lexicon. I will argue that the correlation of innovativeness at different levels is due to phonological change being causally associated with innovation beyond the phonological domain, as the loss of expressivity brought about by sound change (i.e. homophony, syncretism) precipitates lexical replacements (e.g. derivatives like *agnellus* instead of *agnus* ‘lamb’ or *rānucula* instead of *rāna* ‘frog’) and morphosyntactic innovations (e.g. obligatory subject pronouns, periphrastic past, etc.).



## References

AIS = Jaberg, Karl, & Jud, Jakob. (1928–1940). *Sprach- und Sachatlas Italiens und der Südschweiz (AIS)*. Zofingen: Ringier.

ALF = Gilliéron, Jules, & Edmont, Edmond. (1902–1910). *Atlas Linguistique de la France*. Paris: Champion.

ALI = M.G. Bartoli, G. Vidossi, B.A. Terracini, G. Bonfante, C. Grassi, A. Genre, L. Massobrio. (1995). *Atlante linguistico italiano*. Roma: Istituto Poligrafico e Zecca dello Stato/Libreria dello Stato.

ALPI = Navarro Tomás, Tomás, Aurelio M. Espinosa hijo, Luís F. Lindley Cintra, Francesc de Borja Moll, Armando Nobre de Gusmão, Aníbal Otero, Lorenzo Rodríguez Castellano, Manuel Sanchis Guarner. (1962). *Atlas Lingüístico de la Península Ibérica, I, Fonética, I*, Madrid, CSIC.

Bromham, Lindell, Xia Hua, Thomas G. Fitzpatrick, and Simon J. Greenhill. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences* 112, no. 7: 2097–2102.

Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, and Russell D. Gray. (2017). Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences* 114, no. 42: E8822–E8829.

Heeringa, Wilbert, and Frans Hinskens. (2011). The measurement of Dutch dialect change: Lexicon versus morphology versus sound components. *Taal en tongval* 63, no. 1: 79–98.

Loporcaro, Michele, Stephan Schmid, Chiara Zanini, Diego Pescarini, Giulia Donzelli, Stefano Negrinelli and Graziano Tisato. (2021). AIS, reloaded: A digital dialect atlas of Italy and southern Switzerland. In: A. Thibaut, M. Avanzi, N. Lo Vecchio and A. Millour (eds), *Nouveaux regards sur la variation dialectale*. Strasbourg: Editions de Linguistique et de Philologie: 111–136.

Trudgill, Peter. (2020). Sociolinguistic typology and the speed of linguistic change. *Journal of Historical Sociolinguistics* 6, no. 2: 20190015.

Wichmann, Søren. (2020). How to distinguish languages and dialects. *Computational Linguistics* 45, no. 4: 823–831.



## Inter-paradigm leveling in present-day Judeo-Spanish

Branka Arrivé  
(INALCO, Paris)

Keywords: agency, identity, language ideology, nominal morphology, paradigm leveling

Judeo-Spanish is one of the six recognized non-territorial minority languages in France (DGLFLF 2016). It currently has no intergenerational transmission and is classified as “severely endangered” on the UNESCO endangerment scale (Moseley 2010).

Revitalization efforts began to bear fruit with the creation of online communities, especially during the Covid-19 pandemic, when speakers and learners increasingly joined together in “digital homelands” (Held 2010), in which conversations are held entirely in Judeo-Spanish.

Speaker agency is essential in constructing a “Judeo-Spanish identity”, which positions itself in contrast with Spanish. The policy practiced by the online groups is based on the selection of attested forms that diverge from Spanish and on the adoption of unattested forms from languages that influenced Judeo-Spanish in the past or based on the analogy of non-standard or stigmatized varieties of Spanish.

This paper chooses to analyze language change in present-day Judeo-Spanish by focusing on the phenomenon of inter-paradigm leveling or paradigm migration from non-prototypical to prototypical noun classes (Zadok & Bat-El 2014). There are two patterns that can be observed starting from the late 19th century: 1) nouns ending in *-e* or a consonant tend to join either the noun class ending in *-o* (e.g. *el arabe* > *el arabo*, ‘the Arab’) or the noun class ending in *-a* (*kriza*, cf. Sp. *crisis*, ‘crisis’) and 2) masculine nouns ending in *-a* either become feminine (e.g. *la problema*, ‘the problem’) or change the ending, either by dropping the *-a* (*el problem*) or by replacing it with *-o* (*el problema*). While these patterns were inconsistent in the 19th and 20th century, they have become reinforced and encouraged today as a way of marking Judeo-Spanish as distinct from Spanish. The second pattern (phonetic correspondence between the noun ending and its gender) has even affected some very common nouns in certain documents although frequency remains relevant in resistance to change. There have been calls for normativization and certification of Judeo-Spanish and if they are adopted it is expected that the current heterogeneity of the nominal forms will be reduced.

Inter-paradigm leveling is motivated by the concurrent factors of the prevalence of the prototypical nominal paradigms ending in *-o* or *-a* and by language contact. The form *el arabo* was probably influenced by the Italian *l’arabo*, whereas the nouns of Greek or Latin origin ending in *-is* or *-e* in Spanish are not attested in classical (18th century) Judeo-Spanish and were borrowed from French in the late 19th century, using the pattern *-se* > *-za*, *-sse* > *-sa* (*kriza*, *baza*, *klasa*, ‘crisis’, ‘base’, ‘class’ etc.).

The study relies on four collections of written and spoken material in Judeo-Spanish, two of which are multi-genre diachronic corpora: CORHIJE (Corpus Histórico Judeoespañol of the CSIC, Madrid) and CoDiAJe (Corpus diacrónico anotado del judeoespañol of the Hebrew University in Jerusalem) and the other two reflect the present-day language: the publication *Aki Yerushalayim* (Jerusalem, founded in 1979) and the weekly show *Enkontros de alhad* (Buenos Aires, started in August 2020), used as speech corpus.

## References

*Aki Yerushalayim*. Revista kulturala Djudeo-Espanyola (2008-2024), Jerusalem.

<https://yerushalayimaki.wixsite.com/ladino>

CoDiAJe - The Annotated Diachronic Corpus of Judeo-Spanish. Director: Aldina Quintana.

<http://recursos-glif.upf.edu/teitok/codiaje/>

Délégation générale à la langue française et aux langues de France (2016), *Les langues de France*, 2nd edition

*Enkontros de alhad. Enkontros avlados enteramente en Ladino* (2020-2024), Buenos Aires: eSefarad.

<https://esefarad.com/tag/enkontros-de-alhad/>

García Moreno, Aitor and Javier Pueyo Mena (2013-2024), Corpus Histórico Judeoespañol CORHIJE.

<http://esefardic.es/corhije>

Held, Michal (2010), The people who almost forgot: Judeo-Spanish online communities as a digital home-land, *El Prezente* 4, 83-101.

Moseley, Christopher (ed), (2010), *Atlas of the World's Languages in Danger*, 3rd ed. Paris: UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000187026>

Zadok, Gila and Outi Bat-El (2015), Inter-paradigm leveling in Hebrew verbal system, *Morphology* 25, 271-297.

# Sociolinguistic Dynamics and Diachronic Change in Royasque: Insights from a Minoritized Romance Variety at the Franco-Italian Border

Charlène Chaupré-Berki

UR4299 Centre Interdisciplinaire de Recherches sur les langues et la pensée (CIRLEP) Université de Reims  
Champagne Ardenne

## Abstract

This study investigates the diachronic evolution of Royasque, a minoritized Romance variety spoken in the alpine region straddling the Franco-Italian border. Situated at the crossroads of Occitan and northern Italian linguistic traditions, Royasque presents a unique case of language contact, preservation, and shift. The research explores the resilience of this variety in retaining archaic morphosyntactic features—such as subject clitic doubling (e.g., *mi i vèn*), synthetic perfect forms, and verb-second word order—while also being shaped by sustained contact with dominant national languages, namely French and Italian.

Drawing from a mixed-methods approach, the analysis combines diachronic textual data from early 20th-century regional documents and literary sources with contemporary fieldwork. Interviews conducted with speakers across generations in both formal and informal settings offer insights into patterns of usage, linguistic attitudes, and perceptions of linguistic identity. The study triangulates historical and ethnographic data to provide a comprehensive account of both structural change and sociolinguistic dynamics.

The findings indicate that despite the absence of institutional support and ongoing sociolinguistic pressure toward assimilation, Royasque has maintained a number of conservative Romance features. This is consistent with other studies of Romance varieties exhibiting syntactic archaisms under conditions of marginalization (Dragomirescu & Nicolae 2018; Ledgeway 2000, 2012). At the same time, instances of lexical borrowing, syntactic calques (e.g., prepositional infinitives mirroring French structures), and convergence phenomena with Italian and French align with observed trends in other non-standardized Romance varieties (Kasstan 2015; Remberger 2010).

Generational variation further reveals a shift in linguistic ideologies, with younger speakers displaying both reduced competence and more ambivalent or utilitarian stances toward the variety. These findings highlight the importance of sociolinguistic context in understanding language change and maintenance, as emphasized by Kirilenko et al. (2024) and D'Alessandro (2017).

By documenting the structural and social transformations of Royasque, this research contributes to the broader discussion of Romance language evolution in borderland and minoritized contexts. It also underscores the need to better integrate community-based perspectives and non-standard varieties into models of language change.

**Keywords:** Royasque, morphosyntactic archaisms, language contact, Romance linguistics, minoritized varieties, sociolinguistic variation

# The encoding of Unaccusativity in Old and Middle *Champenois*

Daniela Baldassarre

Roma - Università degli studi Guglielmo Marconi

## Abstract

**Keywords:** Unaccusativity, Old and Middle *Champenois*, Auxiliary Selection, Diachrony, Language Change

The topic of split intransitivity in Romance varieties represents a field of study of great complexity. Regardless of the theoretical perspective adopted (semantic or syntactic), many studies have shown how the auxiliary ‘have’ has progressively spread at the expense of ‘be’ in many Romance varieties (for example, Aranovich 2007, Cennamo 2001, 2002, 2008, Ledgeway 2003a, 2009a, Legendre, Sorace 2003, Mithun 1991, and Sorace 2000, 2011). Several works have focused on split intransitivity in the earliest phase of the langue d’oil (Buridant 2000, 2019, Dufresne, Dupuis 2010, Dupuis, Dufresne 2012, and Cassarà, Stein 2023).

The present contribution fits into a space that has not yet been explored, namely the study of Unaccusativity in the different domains of the early phase of the Langue d’*Oil*.

This contribution aims at: 1) investigating the encoding of Unaccusativity in Old and Middle *Champenois* (a neglected variety of the Langue d’*Oil* domain) and discussing the paths of development of the reflexes of Latin *esse* and *habere* as perfective auxiliaries with one-argument verbs; 2) casting light on the process leading to the use of *avoir* as the only perfective auxiliary in contemporary *Champenois*; 3) investigating the connection between the post-verbal and null subjects in unaccusative constructions and the semantic classes of predicates involved.

Data analysis presented in this study is based on an extensive analysis of eight works in verse/prose from the *Champenois* literary tradition (*Cligés – Chansons – Œuvres - Mémoires ou Vie de saint Louis - Le dit dou lyon - L’art de dictier- Récit du voyage en Terre Sainte - Proposition ... par devant... le comte d’Eu*). From each text all occurrences of unaccusative predicates is gathered and classified in terms of types and tokens. The same analytical approach is taken for all subjects. The diachronic segment of the reference texts ranges from the mid-12th to the mid-15th century. In this study, the classification system of monovalent predicates used is based on Sorace’s (2000) model.

Preliminary results show that: 1) the auxiliary *aveir*, already from a very early phase of the language (around the 12th-13th century), penetrates into the functional domains of *estre*, by appearing with verbs of change of place and change of state, which represent the nucleus of prototypical unaccusatives; 2) the spread of *aveir* is sensitive to semantic factors, 3-4) which affect also the post-verbal position of the subject and null subject in unaccusative constructions.

## References

- Aranovich, Raul (2007), Split auxiliary selection from a cross-linguistic perspective, in R. Aranovich (ed.), (2007) *Split auxiliary systems*, Amsterdam: Benjamins, 1-24.
- Buridant, Claude (2019), *Grammaire du français médiéval*. Strasbourg: Éditions de linguistique et de philologie (Bibliothèque de Linguistique Romane, 16).
- Cassarà, A. and Stein, A., (2023), Variation des diagnostics inaccusatifs dans l’histoire du français, in *SIDF (Colloque de la Société Internationale de Diachronie du Français)*, LMU Munich, Germany, March 22-24, 2023.
- Caudal, P., Burnett, H. and Troberg, M. (2017), Les facteurs de choix de l’auxiliaire en ancien français: étude quantitative, in Prevost, S. and Fagard, B. (eds.) (2017), *Le français en diachronie. Dépendances syntaxiques, Morphosyntaxe verbale, Grammaticalisation*, Berna: Peter Lang, 237-265.
- Cennamo, Michela (2001), L’inaccusatività in alcune varietà campane: teorie e dati a confronto, in A. Leoni, F., Krosbakken, E.S., Sornicola, R. and Stromboli, C. (eds.) (2001), *Dati empirici e teorie linguistiche*.

- Atti del XXXIII Congresso Internazionale delle Società di Linguistica Italiana (Napoli, 28-30 ottobre 1999)*, Roma: Bulzoni, 427-452.
- (2002), La selezione degli ausiliari perfettivi in napoletano antico: fenomeno sintattico o sintattico-semantic?, *Archivio glottologico italiano*, 87, (2), 175-222.
- (2008), The rise and development of analytic perfects in Italo-Romance, in T. Eythórsson (ed.) (2008), *Grammatical Change and Linguistic Theory: The Rosendal Papers (Linguistik Aktuell/Linguistics Today 113)*, Amsterdam: Benjamins, 115-142.
- Dufresne, M. and Dupuis, F. (2010), Les structures inaccusatives en français médiéval, in Neveu, F. and others (eds.) (2010), *Congrès Mondial de Linguistique Française - CMLF 2010*. Paris: Institut de Linguistique Française, 159-176.
- Ledgeway, Adam (2003a), L'estensione dell'ausiliare perfettivo «avere» nell'antico napoletano: Intransitività scissa condizionata da fattori modali, in *Archivio glottologico italiano* 88, 27-71.
- (2009a), *La grammatica diacronica del napoletano*, Tübingen: Max Niemeyer Verlag.
- Legendre, G., and Sorace, A., (2003), Auxiliaires et intransitivité en français et dans les langues romanes, in Godard, D. (Ed.) (2003), *Les Langues Romanes: Problèmes de la Phrase Simple*. Paris: CNRS editions. 185-234.
- Marchello-Nizia, C., Combettes, B., Prevost S. and Scheer T. (2020), *Grande grammaire historique du français*, Boston: De Gruyter Mouton.
- Mithun, Marianne (1991), Active/Agentive Case Marking and Its Motivation, in *Language* 67(3), 510-546.
- Perlmutter, David (1978). Impersonal Passives and the Unaccusative Hypothesis, in *Proceedings of the 4th Annual Meeting of the Berkeley Linguistic Society*, 157-189.
- Sorace, Antonella (2000), Gradients in auxiliary selection with intransitive verbs, in *Language* 76(4), 859-890.
- Smith, John Charles (2016), French and northern Gallo-Romance, in Ledgeway, A. and Maiden, M. (eds.) (2016), *The Oxford guide to the Romance languages*, Oxford: Oxford University Press, 292-318.

# ***Gone With the Aoristic Drift: The Case of the Ancient Venetan Vernacular Varieties***

Francesca Valcamonico  
(University of Bergamo)

Keywords: <*aoristic drift*, ancient Venetan vernacular varieties, Compound Past, Simple Past, grammaticalisation>

The phenomenon of the *aoristic drift* (Squartini and Bertinetto 2000) describes the gradual process by which the Compound Past (or Present Perfect) has taken over the functions previously fulfilled by the Simple Past (or Past Simple), leading to its decline. Since Harris's seminal study (1982), this phenomenon has been divided into four *stages*, defined according to the degree of grammaticalisation achieved by the compound form. Over the last forty years, research on this topic has mainly focused on Spanish, which, according to Harris's work, belongs to the third stage, with particular attention to the varieties spoken in Central and South America, which show a different pattern in the use of Simple and Compound Past in relation to Peninsular Spanish (Schwenter 1994a, 1994b; Escobar 1997; Schwenter and Torres Cacoulllos 2008; Copple 2009, 2011; Jara 2009; Rodríguez Louro and Howe 2010; Torres Cacoulllos 2011; Howe 2013; Hernandez 2014; Soto 2014; Montoro del Arco 2017; Kleinherenbrink 2019). However, other Romance languages, especially those that are not considered standard, such as dialect varieties, have not received the same attention. In this context, the Venetan dialects are an important example of linguistic varieties that have reached the end of the aoristicisation process, having completely abandoned the use of the Simple Past in favour of the Compound Past (Skubic 1971).

The aim of this paper is to present the results of the study on *aoristic drift* carried out on a *corpus* of ancient Venetan texts, covering the period from the beginning of the 14<sup>th</sup> century to the end of the 17<sup>th</sup> century. After giving an overview of the current state of research on the subject, I will discuss the philological, linguistic, and textual criteria used for the selection of the texts included into the *corpus*. I will present the textual analysis that made it possible to compare data from different textual genres (judicial acts, vulgarised religious texts, letters, plays), adapting Fido's (1988) model used for Boccaccio's *Decameron*. Finally, I will describe the linguistic categories (Comrie 1976, Bertinetto 1986) used for the quantitative and qualitative analysis of the verb forms, together with the results obtained, which show how certain peripheral linguistic structures of these varieties (specifically, the relative clause expressing anteriority and the periphrasis *cominciare a + infinito*) facilitated the change under study during the 16<sup>th</sup> and the 17<sup>th</sup> century. The synchronic and diachronic analysis of the collected data made it possible to reconstruct the course of the linguistic shift and place it into the wider context of the Romance languages. It was thus possible to describe more closely the transition from the third to the fourth stage of the *aoristic drift* in the ancient Venetan vernacular varieties.

## **References**

- Bertinetto, Pier Marco (1986), *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*, Firenze: Accademia della Crusca.
- Comrie, Bernard (1976), *Aspect*, Cambridge: Cambridge University Press.
- Copple, Mary T. (2009), *A diachronic study of the Spanish perfect(ive): Tracking the constraints on a grammaticalizing construction*, Ph.D. dissertation, University of New Mexico.
- Copple, Mary T. (2011), Tracking the constraints on a grammaticalizing perfect(ive), *Language Variation and Change*, 23, 163–191.
- Escobar, Anna María (1997), Contrastive and innovative uses of the present perfect and the preterite in Spanish in contact with Quechua, *Hispania*, 80, 859-870.

- Fido, Franco (1988), *Il regime delle simmetrie imperfette. Studi sul «Decameron»*, Milano: Franco Angeli.
- Harris, Martin (1982), The 'Past Simple' and the 'Present Perfect' in Romance, in Vincent, Nigel, Harris, Martin (eds.) (1982), *Studies in the Romance Verb*, Londra: Croom Helm, 42-70.
- Hernández, José Esteban (2004), *Present Perfect Variation and Grammaticization in Salvadoran Spanish*, Ph.D. dissertation, University of New Mexico.
- Howe, Chad (2013), *The Spanish Perfects. Pathways of Emergent Meaning*, London: Palgrave Macmillan.
- Jara, Margarita (2009), El pretérito perfecto simple y el pretérito perfecto compuesto en las variedades del español peninsular y americano, *Signo&Seña*, 20, 263-291.
- Kleinherenbrink, Carmen (2019), *Today is Perfect, isn't it? Another Look at the Functions of the Present Perfect in the Spanish Variety of Alicante*, MA Dissertation, Leiden University.
- Montoro del Arco, Esteban T. (2017), El pretérito perfecto compuesto con valor aorístico en el habla urbana de Granada, *Orillas. Revista d'Ispanística*, 6, 455-470.
- Rodríguez Louro, Celeste, Howe, Chad (2010), Perfect Potential: Semantic Change in Narrative Contexts across Spanish, *Revista Internacional de Lingüística Iberoamericana*, 8/2(16), 157-174.
- Schwenter, Scott (1994a), The Grammaticalization of an Anterior in Progress: Evidence from a Peninsular Spanish Dialect, *Studies in Language*, 18, 71-111.
- Schwenter, Scott (1994b), "Hot news" and the Grammaticalization of Perfects, *Linguistics*, 32(6), 995-1028.
- Schwenter, Scott, Torres Cacoullos, Rena (2008), Defaults and Indeterminacy in Temporal Grammaticalization: The 'Perfect' Road to Perfective, *Language Variation and Change*, 20, 1-39.
- Skubic, Mitja (1971), Contributo alla conoscenza delle sorti del preterito in area veneta, *Studi di grammatica italiana*, 1, 117-178.
- Soto, Guillermo (2014), El pretérito perfecto compuesto en el español estándar de nueve capitales americanas: frecuencia, subjetivización y deriva aorística, in Azpiazu Torres, Susana (al cuidado de) (2014), *Formas simples y compuestas de pasado en el verbo español*, Lugo: Editorial Axac C.B., 131-146.
- Squartini, Mario, Bertinetto, Pier Marco (2000), The Simple and Compound Past in Romance Languages, in Östen, Dahl (ed.) (2000), *Tense and Aspect in the languages of Europe*, Berlin: Mouton De Gruyter, 403-439.

**Vowel length in an Occitan dialect.**  
**Comparing grammatical description with production data**  
Judith Meinschaefer (Freie Universität Berlin)

**Background:** This study focuses on the distribution of vowel length in the Limousin variety of Occitan as spoken in Nontron (Dordogne). In previous studies of this variety, located in the border region between French and Occitan, contrasting vowel length has been reported, which is noteworthy given that contemporary French and many synchronic Occitan varieties lack long vowels. Contrastive vowel length is found in very few underived lexemes, though; examples include words such as *pan* ['po] 'bread' and *pòst* ['po:] 'board' (Javanaud 1981: 17).

**Research question:** The aim of the study is, first, to determine whether and in which environments Nontron Limousin shows vowel length, and second, how diachronic descriptions of long vowels in this variety fit in with newer data on this or other varieties of Limousin.

**Data:** The diachronic description of vowel length in Nontron Limousin reported here is based primarily on the *Grammaire limousine* published in 1876 (Chabaneau 1876), written by a native speakers of the variety. In addition, the study draws on data from the *Atlas linguistique de la France* (Gilliéron & Edmont 1902), which systematically recorded vowel length and stress. According to Chabaneau's detailed study, in Nontron Limousin vowels are long in two cases. First, length is prosodically conditioned, depending on stress and vowel quality and occurs only in a penult. Lengthening following the diachronic deletion of a coda consonant is the second source of long vowels, variably occurring after deletion of a liquid or sibilant – where the deleted consonant sometimes serves as a morphological exponent –, but not after any other consonant. As a result, verb forms contrasting only with respect to vowel length can be observed, such as *chan't[a]*, from Latin CANTĀTUM, 'sing.PTCP.PRF' and *chan't[a:]*, from Latin CANTĀRE 'sing.INF' (Chabaneau 1876: 12). As to synchronic data, it is clear that Occitan and (even more so) Limousin Occitan are moribund with very few speakers left. Short newspaper articles read by (anonymous) native Limousin speakers (<https://ieo-lemosin.org/actualites/ieo-lemosin/las-cronicas-lo-pais-vos-parla-en-linha?lang=fr>) were segmented and subject to acoustic analysis with Praat. The focus was on comparing vowel durations in lexical contexts relevant to the two factors conditioning length according to Chabaneau. The results show that, while the data contain a great deal of variation, they support various aspects of Chabaneau's description.



**Analysis:** With respect to an analysis in terms of the underlying phonological system, Chabaneau’s description raises two questions: First, how should the system be represented? Second, how could it have developed from a proto-Romance system and how does it differ in terms of underlying representations from the two neighboring systems, “Standard” French and Lengadocian Occitan, given that both lack long vowels? The analysis proposed here for the Limousin is cast in a standard version of Optimality Theory (Prince & Smolensky 1993, Krämer 2017), assuming a mechanism of compensatory lengthening (de Chene & Anderson 1979) triggered by consonant deletion, as well as a preference for an uneven trochee (Bullock 1995, Repetti 2000).

## References

- Bullock, Barbara E. (1995): The uneven trochee in French. *Rivista di Linguistica* 7(2). 273–292.
- Chabaneau, Camille. 1876. *Grammaire Limousine: Phonétique. Parties du discours*. Paris: Maisonneuve.
- Chene, Brent E. de & Stephen R. Anderson (1979): Compensatory Lengthening. *Language*. 55(3). 505–535. doi:10.2307/413316.
- Gilliéron, Jules & Edmont, Edmond. 1902. *Atlas linguistique de la France*. Paris: Champion.
- Javanaud, Pierre G. (1981): *The vowel system of Lemosin: A phonological study*. Doctoral dissertation, Göteborg University.
- Krämer, Martin (2017): Current issues and directions in Optimality Theory: Constraints and their interaction. In S. J. Hannahs & A. R. K. Bosch (Hrsg.), *The Routledge handbook of phonological theory*, 37–67. London: Routledge. doi:10.4324/9781315675428-3.
- Prince, Alan S. & Paul Smolensky (1993): *Optimality Theory: Constraint interaction in Generative Grammar*. Technical Report Rutgers University.
- Repetti, Lori (2000): Uneven or Moraic Trochees? Evidence from Emilian and Romagnol Dialects. In Lori Repetti (ed.), *Phonological theory and the dialects of Italy*, 273–288. Amsterdam: Benjamins.

# OV orders in Old Venetian as a non-unitary phenomenon

Maddalena Bonadio & Cecilia Poletto

(Università di Padova; & Università di Padova and Goethe-Universität Frankfurt am Main)

Keywords: <Old Venetian, Syntax, OV orders, Deictic elements, Deictic adverbs>

In this work we intend to analyze the distribution of OV orders in Old Venetian, which is nowadays considered as a dialect but has a considerable written tradition, in comparison with the Old Florentine system described in Poletto (2014). We will take into account a single text written by a single author so as to minimize the possible different grammars, which is long enough to be a sufficient source of data and at the same time reliable in philological terms, i.e. the *Venetian Tristano* (*Tristano Veneto*). We investigated the amount of OV orders with different types of objects: bare quantifiers and complex quantified expressions, definite objects, indefinite objects. Indefinite DPs invariably occur in VO order, which provides a strong argument that Old Venetian, on a par with Old Florentine and most probably other Old Romance varieties, was underlyingly VO. OV orders are derived through movement of the object to positions located higher than the past participle movement span. Actually, the past participle in Old Venetian is located in a rather low position, which renders OV orders, and consequently movements of different types of objects, visible. Old Venetian confirms the tendency of quantifiers to occur in OV order, an observation that has already been made for both Germanic (see Svenonius 2000) and Romance (see the original finding by Kayne 1975 in French). Furthermore, bare quantifiers are more prone to OV than complex quantified expressions, which is also a tendency that confirms data from Old Italian and Old French. Two further classes of elements display OV rather massively, which have never been identified so far in the literature. The first class are deictic pronouns (59 cases, 48% in OV order) or expressions containing a deictic determiner (125 cases, 23% in OV order). The second class is represented by some locative and tense adverbs and complex PPs, which are also massively OV. There are 140 cases of deictic locative adverbs, 69% of which occur in OV order; there are 25 cases of deictic temporal adverbs, 59% of which occur in OV order. On a par with quantifiers, which are moved to a dedicated position, we can surmise that deictic elements move out of the vP area to reach the higher phase, in which they can be bound by the speaker's coordinates (see Giorgi 2010). Interestingly, when we have more than one element in OV order, or look at their distribution with respect to adverbs located in the low IP area, we can extrapolate the following hierarchy:

bare quantifiers → definite DPs/ Deictic elements

More generally, this means that what we call "OV" is not a unitary phenomenon, but rather a set of different movements triggered by distinct features that emerge only in those languages where the past participle is low enough to render this movement visible. Time permitting, we will provide an explanation for the new cases of deictic elements (pronouns and PPs) as occurring in multiple Topic positions of the vP periphery.

## References

- Donadello, Aulo (1994), *Il libro di messer Tristano ("Tristano veneto")*, Venezia: Marsilio.
- Giorgi, Alessandra (2010), *About the speaker: Towards a syntax of indexicality*, Oxford: Oxford University Press.
- Kayne, Richard S. (1975), *French syntax: The transformation cycle*, Cambridge, MA: MIT press.

Poletto, Cecilia (2014), *Word order in Old Italian*, Oxford: Oxford University Press.

Svenonius, Peter (2000), *The derivation of VO and OV*, Amsterdam / Philadelphia: John Benjamins.

## Computer Simulation: Grammatical Gender from Latin to Old Occitan

Matthias Schöffel (Bayerische Akademie der Wissenschaften), Marinus Wiedner (Universität Freiburg)

Keywords: Language shift, simulation, Old Occitan, Latin, grammatical gender.

This communication is based on the study by Polinsky/Everbreeck (2003), who simulated the development of grammatical gender from Latin to Old French with a connectionist model. Building on this, we want to simulate the development of gender from Latin to Old Occitan. Therefore, we use a *long-short-term memory* (LSTM) architecture combined with an attention mechanism in opposition to heuristic models (cf. Marr/Mortensen 2020).

As a next step, we also focus on CVAE (Conditional Variational Autoencoder) models, which learn to generate word forms conditioned on linguistic features. It enables controlled and diverse generation by combining a latent variable space with observed inputs. This approach is particularly relevant for modeling complex morphological changes, such as the grammatical gender shift from Latin to Old Occitan. The character-based decoder now produces morphologically structured word forms. Moving forward, we aim to refine the generation further by addressing deviations from stem words and enhancing semantic consistency through targeted model adjustments.

A gender reduction from three to two genders took place in the transition from Latin to Old Occitan, during which the neuter disappeared. The neuter nouns (e.g. Latin third declension MARE) had to be reattributed to either masculine or feminine (cf. it. *il mare*<sub>masc</sub> vs. fr. *la mer*<sub>fem</sub> vs. both genders in Old Occitan), and we aim at simulating this development, starting on the character level.

For this simulation, we use nouns from the *Dictionnaire de l'occitan médiéval* (DOM), the largest work of Old Occitan lexicology. We also included variants digitized beforehand via a tailored OCR model (cf. Garcés Arias/Pai/Schöffel/Heumann/Aßenmacher 2023). As a starting point for the model training we take the linked etyma from the *Französischen Etymologischen Wörterbuch* (FEW).

In addition to the lexicographic (and therefore normalised) data we use nouns extracted from original manuscripts from the 13<sup>th</sup> and 14<sup>th</sup> century, semi-automatically transcribed with a Transkribus-model for Old Occitan Handwriting (cf. Wiedner 2023). We then annotated the texts using available PoS tagger, and manually correcting the results. Afterwards, we manually combined these nouns with their respective etyma, including information from the FEW and the *Thesaurus Linguae Latinae* (TLL) on gender (and possible variation) and the accusative

forms. We want to see if there are differences in the simulations' outcome with this non-normalised, 'authentic' data in comparison to the data taken from the DOM.

We will present and discuss the basic idea as well as preliminary results.

## References:

DOM = *Dictionnaire de l'occitan médiéval*. <<http://www.dom-en-ligne.de/>>.

FEW = Wartburg, Walther von, et al. (1922–2022): *Französisches Etymologisches Wörterbuch (FEW). Eine Darstellung des galloromanischen Sprachschatzes*. 25 Bände, Bonn/ Heidelberg/ Leipzig/ Berlin/ Basel, Klopp/ Winter/ Teubner/ Zbinden. <<https://apps.atilf.fr/lecteurFEW/>>

Garcés Arias, Esteban, Pai, Vallari, Schöffel, Matthias, Heumann, Christian, Aßenmacher, Matthias. 2023. Automatic Transcription of Handwritten Old Occitan Language. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, Singapore. Association for Computational Linguistics, 15416–15439.

Marr, Clayton, Mortensen David (2020): „Computerized Forward Reconstruction for Analysis in Diachronic Phonology, and Latin to French Reflex Prediction“, *Proceedings of LT4HALA 2020 - 1<sup>st</sup> Workshop on Language Technologies for Historical and Ancient Languages*, 28–36.

Polinsky, Maria, Everbroeck, Ezra van (2003): „Development of Gender Classifications: Modeling the Historical Change from Latin to French“, *Language* 79, 356–390.

TLL = *Thesaurus linguae Latinae (TLL) open access*. München. <<https://publikationen.badw.de/de/thesaurus>>

Wiedner, Marinus (2023). OldOccitanHandwriting, (Modell-Nr. 52822, CER=3,51 %), PyLaia-model for handwritten Occitan of the 13th and 14th century. <<https://readcoop.eu/de/modelle/old-occitan-handwriting/>>

The purpose of this paper is to report on subject expression in Gascon. Gascon is a distinct variety in the Occitan continuum which has largely been understudied, especially where syntax is concerned (an opinion shared by Sibille and Bach 2024). The received wisdom is that Gascon is a V2 language with null subjects of the consistent type, and that it was always thus (Ledgeway 2020, Vance et al 2010). It is not clear from the literature however whether the null subject is of the V2 type, with a preponderant distribution in main clauses as in Medieval French (Larrivée *et al* 2024), or of the consistent type, as in Iberian languages. An empirical issue arises as to the data to be used, in its representativeness, both in terms of the texts fully representing the variety (see Glessgen 2021) and in terms of ordinary language (see, for French, Pinzin and Goux 2024). In this exploratory study, we refer to the 15<sup>th</sup> century *Fors du Béarn*: on the one hand, the custumal text type has been shown to be less conservative for Medieval French, and it might be so for Gascon as well; on the other, the later date allows any development from expectations to be pinned down. We use a treebank approach to a fully UD-annotated version of the text to be part of the SOLIDGRAM corpus under development. The dimension that is explored in this quantitative approach is specifically the distribution of null subjects in root as compared to embedded environments. The preliminary results appear to confirm the consistent nature of null subject language status, and interestingly not a V2 status, given the absence of a quantitative asymmetry in the distribution of null subjects between root and embedded. The results we compare to available quantitative studies for historical Spanish (Dufter 2011) and Portuguese (Galves in preparation), and find a closely comparable lack of asymmetry. This absence is all the more significant that the rate of expressed subject in post-verbal position is indicative of a V2 distribution for Gascon. We hope that these results provide a baseline for comparisons to other stages of Gascon, surrounding Occitan varieties and other Romance varieties.

- Dufter, A. 2011. Sujetos pronominales y estructura informativa en el español tardomedieval. Andreas Dufter and Daniel Jacob (Eds), *Syntaxe, structure informationnelle et organisation du discours dans les langues romanes*. Berlin : Peter Lang. 49-78.
- Frascarelli, M. 2018. The interpretation of pro in consistent and partial null-subject languages. Federica Cognola and Jan Casalicchio (Eds), *Null subjects in Generative Grammar: A synchronic and diachronic perspective*. Oxford: Oxford University Press. 211-239.
- Glessgen, M. 2021. Pour une histoire textuelle du gascon médiéval. *Revue de linguistique romane*, 85(339-340), 325-384.
- Larrivée, P., C.Poletto, F. Pinzin and M. Goux Asymmetry as a general cue for V2 (loss). *Isogloss* 10,7.
- Ledgeway, A. 2020. Variation in the Gallo-Romance left-periphery: V2, complementizers, and the Gascon enunciative system. Sam Wolfe and Martin Maiden (Eds), *Variation and Change in Gallo-Romance Grammar*. Oxford: Oxford University Press.
- Pinzin, F. and M. Goux. 2024. *How genre affects word order: a diachronic analysis of French*. MS, Frankfurt and Caen.
- Sibille, J. and X. Bach. 2024. Syntax. L. Esher and J. Sibille (Eds). *Manuel de linguistique occitane*. Berlin: De Gruyter. 351-398.
- Vance, B., B. Donaldson and B. D. Steiner. 2010. V2 loss in Old French and Old Occitan. In *Romance Linguistics 2009: Selected Papers from the 39th Linguistic Symposium on Romance Languages (LSRL)*, Tucson, Arizona, March 2009 (Vol. 315, p. 301). Philadelphia : Benjamins.

## **New elements for the history of the Croissant (northernmost Occitan area)**

This contribution suggests a slightly revised scenario for the history of the northernmost occitan area, known as the Croissant (Brun-Trigaud 1989, Esher et al. 2021). This area, heavily influenced by Oïl language(s), has long been under-studied because it was not considered as authentic Occitan. Moreover, the limited existing research has often been framed by the search for a “border” between Oc and Oïl languages, rather than examining the area in its own right.

Over the past decade, significant advances have been made in the study of this region (Esher et al. 2021; Guérin; Author 2021 ; amongst others), driven by an ANR project that has facilitated extensive sociolinguistic and dialectal surveys. These efforts have resulted in numerous descriptive grammars and lexical inventories, providing a foundation for new syntheses. This updated documentation allows us to reassess the region's history, particularly the processes through which French influenced Croissant varieties.

Our hypothesis is that the Croissant area did not result from egalitarian contact of Oc dialects with Oïl dialects as it is still most often hypothesized. Instead, we argue that the Croissant primarily reflects unilateral borrowing from French. In this respect, the region aligns with broader Occitan trends (Chambon 2009), albeit over a longer timescale, involving borrowings from earlier stages of French. This influence was further amplified by the region's proximity to Oïl-speaking areas and the interplay of regional French, Oïl dialects, and the diffusion of French through urban-network. Rather than being as a “transition zone” or “contact zone” between Oïl and Oc dialects, the Croissant's specific characteristics stem largely from the unidirectional influence of French on local Oc dialects. Notably, there is no historical or sociolinguistic evidence of significant influence from Oïl varieties, nor of influence of Oc varieties on Oïl ones. Instead, French proper is attested as the primary influence, as evidenced by its exclusive use for writing documents, alongside Latin, as early as the 14th century. Consequently, the Croissant's varieties are rarely attested in written sources (Brun 1923: 51).

To illustrate this hypothesis, we analyze the lexicon of two close doculects from the Croissant that have been recently documented (Dompierre-Les-Églises (Guérin 2019, henceforth DLE) and Saint-Agnant-de-Versillat (Author 2021, henceforth SAV). Numerous cases implies borrowing from French at very various stages:

- For instance lexems with /we/ (fr. /wa/) illustrate borrowing before contemporary French (SAV : /mwεzir/ ‘rotten’, /ʃwεzir/ ‘choose’ /pwε/ ‘pea’, /pwεvr/ ‘pepper’, etc.
- forms shared with other Oïl dialects often trace back to Middle French. For instance DLE /ʃom(ə)nir/ ‘to go mouldy’ is attested in many varieties in Poitou (Oïl) and Limousin (Oc) area but also in Middle French (FEW 14:245b) and the diffusion from French account for this scattered distribution ;
- Often, similarity with a form in another Oïl variety cannot be supported by a contact scenario : DLE /bija/ ‘studboar’ is attested only in the distant Bourgogne area (FEW 22/2:5b).

This analysis shows that the very high lexical variation between the data points is mostly due to the differential effect of the influence from French. A systematic comparison of the 230 different forms (for two lexicons of 2500 forms each) show taht the differences are mostly due to difference in the adoption or integration of French borrowings. For instance DLE has /nɔr/ vs SAV /bry/ ‘daughter-in-law’, a clear north-south difference (FEW 7:246a).

The complexity of borrowings from different stages of French, combined with interference from regional French and Oïl varieties, helps explain why the patterns identified by Chambon (2004) around urban centers in the Occitan region are less distinct in the Croissant. This discrepancy may arise from two concurrent pathways of French influence: one following the urban network, as in

other parts of Occitania, and another stemming from rural diffusion of French (or presumed French) forms from adjacent Oil-speaking areas.

Though historically marginalized within the Occitan linguistic landscape, the Croissant may serve as a laboratory for understanding how French influenced the restructuring of Oc languages. This region exemplifies how the influence of a dominant standard language can simultaneously drive convergence and amplify divergence within varieties.

## References

Author 2021.

Brun, A. (1923) *Recherches historiques sur l'introduction du français dans les provinces du midi*, Honoré Champion, Paris

Chambon, Jean-Pierre. 2003. « L'histoire linguistique de l'Aquitaine : de la romanisation à la fragmentation (1er–6e siècles). Éléments pour un modèle », in : *Regards croisés, Recherches en Lettres et en Histoire, France et Hongrie*, textes publiés sous la responsabilité de Jean-Luc Fray et Tivadar Gorilovics, 35-52. Debrecen : Gossuth Egyetemi Kiado-Debreceni Egyetem & Clermont-Ferrand : Presses de l'Université Blaise Pascal.

Chambon J.-P. (2004) « Les centres urbains directeurs du midi dans la francisation de l'espace occitan et leurs zones d'influence : esquisse d'une synthèse cartographique », *Revue de Linguistique Romane*, 68/269-270 : 5-13.

Chambon J.-P. (2009) « Prolégomènes à l'étude historique des emprunts de l'occitan au français » in André Thibault (ed.) *Gallicismes et théorie de l'emprunt linguistique*, Paris : L'Harmattan, 45-69.

Esher, L., Russo, M., Quint, N., & Guérin, M. (2021) *Le croissant linguistique: entre oc, oil et francoprovençal: des mots à la grammaire, des parlers aux aires*, Paris : L'Harmattan.

FEW=Wartburg, Walther von (1922-) *Französisches etymologisches Wörterbuch : eine Darstellung des galloromanischen Sprachschatzes*

Greub, Yan & Jean-Pierre Chambon (2009) « Histoire des dialectes dans la Romania : Galloromania ». In E. Gerhard, M-D. Glessgen, C. Schmitt & W. Schweickard (dir.), *Histoire linguistique de la Romania, Manuel international d'histoire linguistique de la Romania*, vol. 3, 2499-2520. Berlin & New York : Walter de Gruyter.

Guérin M. (2019) *Grammaire du parler marchois de Dompierre-les-Églises (Haute-Vienne)*, Paris : L'Harmattan.

Guylaine Brun-Trigaud (1989) *Formation du concept Croissant : contribution à l'histoire de la dialectologie française au XIXe siècle*. Université de Paris 13.



## **WS4 Disentangling contact and inheritance in lexical semantics**

**John Mansfield & Paul Widmer**

# Rigorous testing of deep-time hypotheses using semantic divergence

Erich Round & James Elhindi

University of Surrey and University of Queensland; & University of Melbourne

**Keywords:** deep-time relatedness, polysemy networks, semantic divergence, phonological divergence, languages of Australia

**Introduction:** A deeper understanding of polysemy and colexification may open the way to significant new progress on long-standing problems in linguistics. We show how polysemy networks based on large-scale cross-linguistic data breathe new life into the rigorous testing of hypotheses of deep-time relatedness (Campbell 1973, Harrison 2003, Campbell & Poser 2008).

**Background:** Over time, true cognates a) diverge phonologically; b) diverge semantically; c) undergo lexical replacement. Hypotheses of deep-time relatedness (DTR) can be contentious, but consensus has emerged (Kesler & Lehtonen 2006, Ringe & Eska 2013, Creolin 2019) that any level-headed evaluation of DTR must compare degrees of similarity in word pairs: a) phonologically, b) semantically and c) against chance. For task c) the agreed, best approach is a Monte Carlo test, where word pairs of interest are compared against a large, baseline sample of randomly paired words. Because this entails thousands of pairwise comparisons, methods for tasks a) and b) must be scalable and automated. Computational methods for task a) are now sufficient (List & Forkel 2024). The weak point has been task b), semantic comparison. The current standard is to compare only word pairs that satisfy semantic equivalence, effectively adopting a hypothesis according to which deep-time cognates never diverge semantically. This *Stasis Hypothesis* is false, but has been adopted as a heuristic, ‘least-bad’ option, explicitly due to the unavailability of any viable, sufficiently non-subjective, scalable alternative (Campbell 1973).

**Contribution:** Based on polysemy networks, we present a scalable, non-subjective measure of semantic distance and demonstrate its application to DTR hypotheses among language families of Australia. Our method appears to be more informative than reliance on the incorrect Stasis Hypothesis alone.

**Semantic divergence:** Paths of semantic change can be inferred empirically from the typology of polysemy (Evans & Wilkins 2000), but expertly hand-curated polysemy databases are limited primarily to high-resource languages (Rzymiski et al. 2020) and fragments of the lexicon. Working with low-resource languages, we took entire wordlists and dictionaries, applied parsing tools to extract succinct lexical glosses, and built an empirical network of recurrent polysemies (15k edges, 3k nodes).

**Results:** Improving on Round & Elhindi (2022), we bin word pairs according to their distance in the semantic network, and examine the probability distribution of phonological distances in each bin. We do this for language pairs of known time depth in Pama-Nyungan (Bowern & Atkinson 2021, Bouckaert et al. 2018). This produces ‘distribution spectra’ of divergences at various time depths (Figure 1), creating a multi-dimensional baseline for measuring divergence. For pairs of non-Pama-Nyungan families, we find distinct spectra and strikingly different signatures in frequent vs rarer vocabulary (Figures 2, 3).

**Conclusions:** Even for low-resource languages, it now appears possible to cease reliance on the false Stasis Hypothesis when evaluating DTR hypotheses. However, these are just first steps. Our baseline (in Figure 1) currently incorporates only divergence over time. One future direction is to distinguish

presence/absence of recent contact at each time depth, to account for convergence. This may shed additional light on the deep-time spectra found in Figures 2, 3.

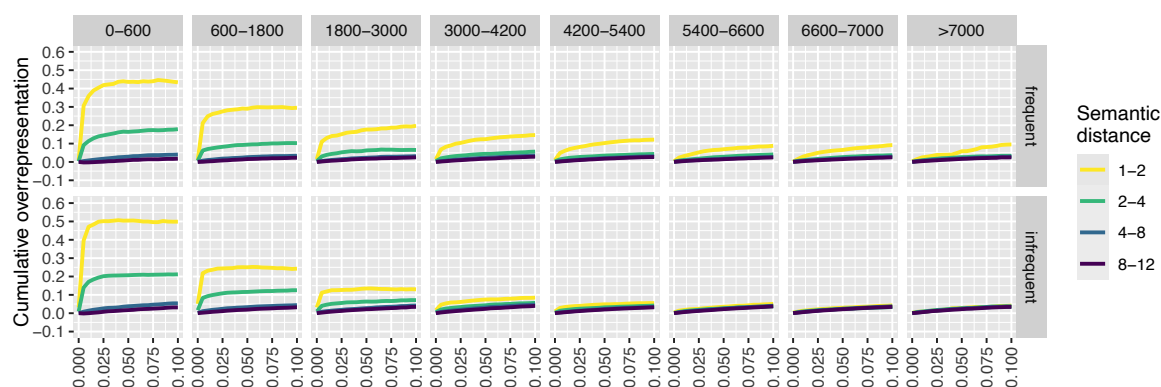


Figure 1: Distribution spectra for 4 semantic distance bins and 8 historical depth bins, split into frequent vs infrequent lexical meanings. Lines that arc strongly upwards are semantic distances at which high phonological similarity most strongly outstrips random chance. At any time depth, the 4 curves provide a depth-specific, multi-dimensional signature of expected phonological divergences for word-pairs with various semantic distances.

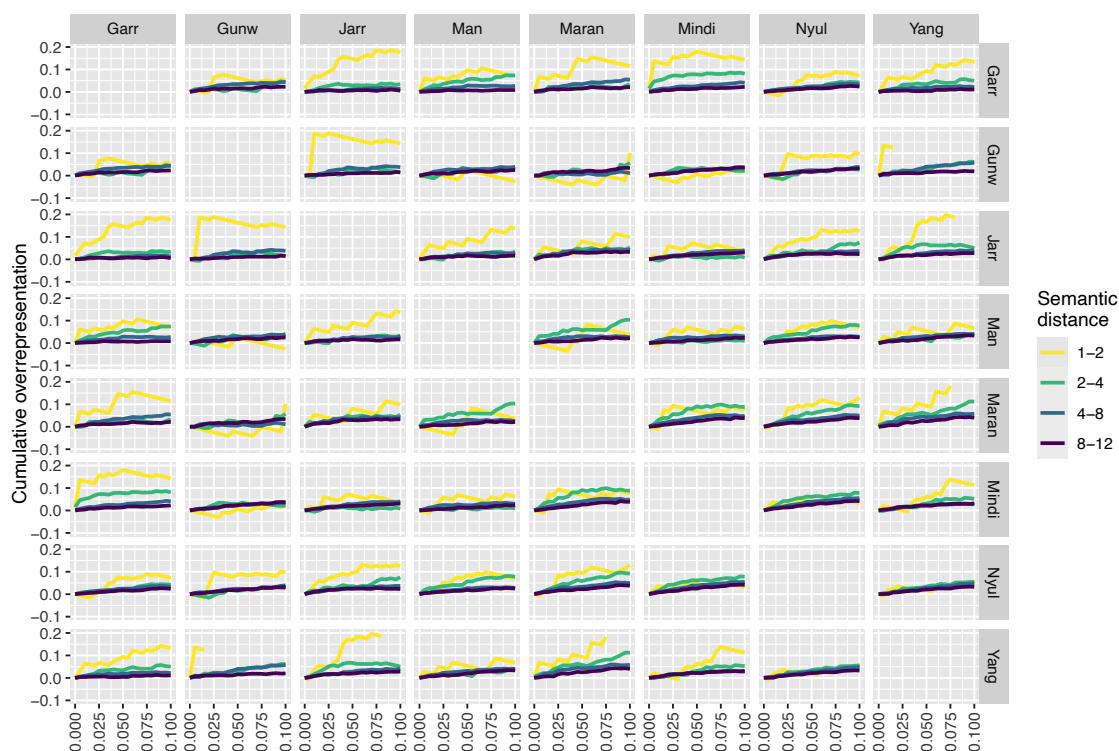


Figure 2: Distribution spectra for paired non-Pama-Nyungan families (frequent lexical meanings). Note: scale of vertical axis is not the same as Figure 1.

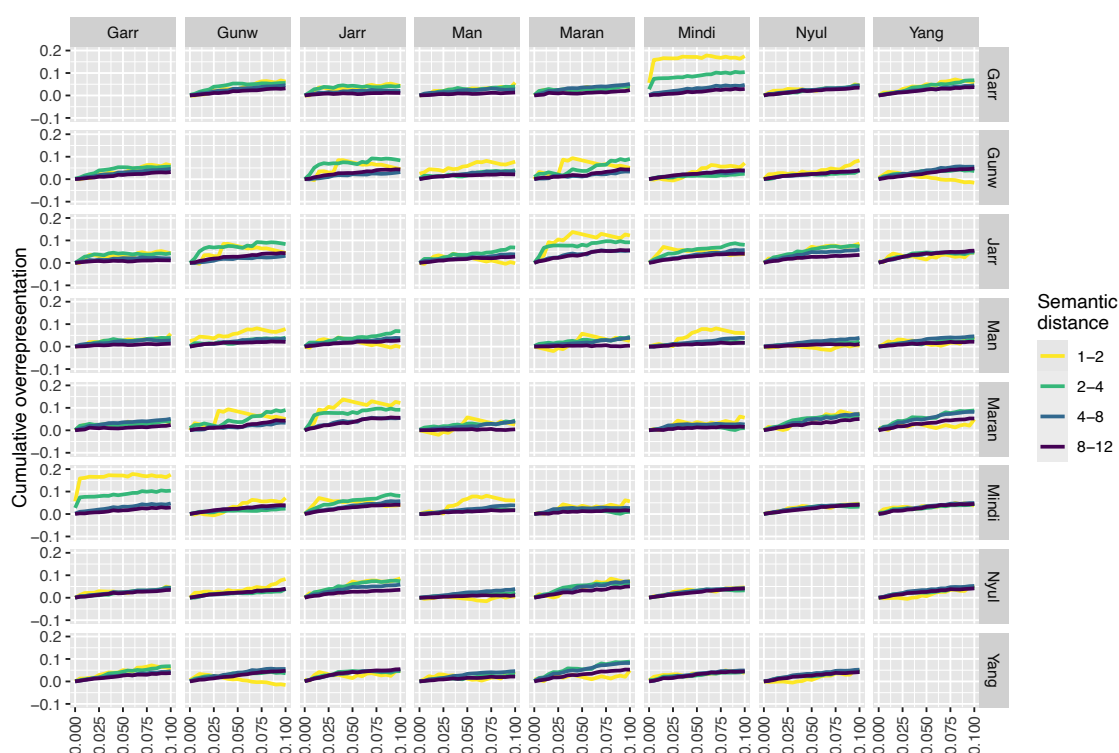


Figure 3: Distribution spectra for paired non-Pama-Nyungan families (infrequent lexical meanings). Note: scale of vertical axis is not the same as Figure 1.

**Acknowledgments** We are grateful for support to ER: Global Professorship GP300169 funded by the British Academy, and Consolidator Grant EP/Y02429X/1 selected by the ERC and funded by the UKRI.

## References

- Bouckaert, Remco R., Claire Bower, and Quentin D. Atkinson (2018) The origin and expansion of Pama–Nyungan languages across Australia. *Nature ecology & evolution* 2:741–649.
- Bower, Claire and Quentin Atkinson (2012) Computational phylogenetics and the internal structure of Pama–Nyungan. *Language* 88.4:817–845.
- Campbell, Lyle (1973) Distant genetic relationship and the Maya–Chipaya hypothesis. *Anthropological Linguistics* 15.3:113–135.
- Campbell, Lyle and William J. Poser (2008) *Language classification: History and method*. Cambridge: Cambridge University Press.
- Ceolin, Andrea (2019) Significance testing of the Altaic family. *Diachronica* 36.3:299–336.
- Evans, Nicholas and David Wilkins (2000) In the mind’s ear: the semantic extensions of perception verbs in Australian languages. *Language* 76.3:546–592.
- Harrison, Sheldon P. (2003) “On the limits of the comparative method”. In: *The handbook of historical linguistics*. Ed. by Brian D. Joseph and Richard Janda. Chicago: Blackwell, 213–243.
- Kessler, Brett and Annukka Lehtonen (2006) “Multilateral comparison and significance testing of the Indo– Uralic question”. In: *Phylogenetic methods and the prehistory of languages*. Ed. by Peter Forster and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, 33–42.
- List, Johann-Mattis and Robert Forkel (2024) LingPy. Python library for historical linguistics. <https://lingpy.org/>
- Ringe, Don and Joseph F Eska. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press, 2013.

- Round, Erich R. and James Elhindi (2022) Sharper evaluation of deep-time hypotheses using both phonological and semantic divergence. Presented at ICHL, Oxford, 2022. URL: <https://ichl.ling-phil.ox.ac.uk/abstracts/373>.
- Rzymiski, Christoph, et al. (2020) The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific data* 7.1:1–12.

## Semantic evolution and lexical gender evolution – in phylogeny and geography

Gerd Carling, Professor, Goethe University/Frankfurt am Main

Nour Efrat-Kowalsky, Wissenschaftliche Mitarbeiter, Goethe University/Frankfurt am Main

The current presentation will look at lexical semantic evolution and how this process relates to the evolution of gender assignment of lexemes – in phylogeny and geography. In gender languages, gender is normally marked at agreeing components, such as numeral, article, adjective, or verb (Corbett 1991, 2013; Corbett and Fraser 2000). This agreement targets either the complete vocabulary, or alternatively parts of it. As languages evolve, gender as a typological feature may either emerge (from nothing), disappear (move from gender to no gender), or alternatively switch categories (e.g., from masculine – feminine to masculine – feminine – neuter). In this process, the fluctuation between categories take place both in the vocabulary as well as at a system level. In the about 20% of the world's languages that have gender (Allasonnière-Tang et al 2021) it is possible to observe the tendency of change between various gender systems: sexus systems (masculine – feminine) are more frequent and stable, as well as more likely to occur emerge and occur at proto-language levels. This is a global trend, based on a comparative phylogenetic study of 3,200 languages [Carling et al to appear]. However, what still remains obscure is how fluctuations of gender systems operate in the vocabulary and how this is related to stability, semantic property of lexemes, or language contact. In the presentation, we will make use of two gender-coded lexical data sets for the Indo-European and Arawak language families, and address this question with respect to semantic stability, semantic property and propensity for gender fluctuation, in phylogeny and geography. Adapting a phylogenetic comparative model (Carling – Cathcart 2021), we will look at the difference in gender change in the lexicon in relation to change in gender systems. Preliminary results indicate that a change in the gender system, i.e., a loss or gain of categories (masculine, feminine, neuter, animate, inanimate) or systems at such (gender, no gender) starts with a fluctuation of gender in the vocabulary. Here, we hypothesize that the trajectory is either from less stable categories (inanimate, feminine, neuter) to more stable categories (animate, masculine) – in case of a gender loss, or vice versa – in case of gender gain. However, the semantic properties of nouns (such as animacy, humanness, concreteness) as well as the general substitution rates of concepts may also impact this process. Towards this end, we contrast the gender gain and loss of concepts with the semantic properties as well as the substitution rates of concept. The other option is that gender fluctuations are impacted by areal contact. To test this, we will set up a phylogeographic model (Cathcart et al 2018), testing to what level the gender fluctuations of languages depend on fluctuations in geographically adjacent languages. This model will be set up for both gender systems, as well as gendered lexemes.

Allasonnière-Tang, Marc, Olof Lundgren, Maja Robbers, Sandra Cronhamn, Filip Larsson, One-Soon Her, Harald Hammarström, and Gerd Carling. 2021. "Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems." *Nature Humanities & Social Science - Communications* 8:331.

Carling, Gerd, and Chundra Cathcart. 2021b. "Reconstructing the evolution of Indo-European grammar." *Language* 97(3):561-598.

Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson, and Erich Round. 2018. "Areal pressure in grammatical evolution: An Indo-European case study." *Diachronica* 35 (1):1-34.

Corbett, Greville G. 1991. *Gender*, Cambridge textbooks in linguistics, 99-0104661-0. Cambridge: Cambridge Univ. Press.

Corbett, Greville G. 2013. "Gender typology." In *The Expression of Gender*, edited by Greville G. Corbett, 87-130. Berlin - New York: Mouton de Gruyter.

Corbett, Greville G., and Norman M. Fraser. 2000. "Gender assignment: a typology and a model." In *Systems of Nominal Classification*, edited by Gunter Senft, 293-325. Cambridge: Cambridge University Press.

## Areal semantic patterns in nominal classification

John Mansfield, Ruth Singer and Frank Seifart

University of Zurich, University of Melbourne, CNRS Structure and Dynamics of Languages

Noun classes and classifiers partition the world into semantic classes, though the semantics may become increasingly incoherent due to grammaticalisation effects (Dixon 1986; Mel'cuk 2006). Following recent work on areal semantics (Koptjevskaja-Tamm & Liljégren 2017; Schapper & Koptjevskaja-Tamm 2022; François 2022) we investigate whether the allocation of nominal concepts into noun classes exhibit areal characteristics (Allasonnière-Tang et al. 2021). In Indo-European, where noun classes (genders) are semantically bleached, class partitions have a strongly phylogenetic distribution (McCarthy et al. 2020). But for semantically richer classifier systems, we might expect that more areal spread could occur (Seifart 2018). The study by McCarthy et al provides a useful measure of semantic similarity between nominal classification systems, treating them as partitions of nominal concepts into sets. For example, Spanish allocates the concepts {FRUIT, MOON, FLOWER} to the same set, since the relevant nouns are all of the same gender. Mutual-information (MI) measurements can be used to capture the similarity of partitioning between languages. For example, German partitions {FRUIT, MOON, FLOWER} into three different genders, which would produce low MI for Spanish:German over this set of concepts.

To test for areal semantic effects in nominal classification, we target three geographic regions, namely northern Australia, the northwest Amazon and northern Papua New Guinea, for which we coded grammatical classification of a standard list of 395 nominal concepts, with 23 languages coded so far (see Table 1 for examples). We then apply McCarthy et al.'s statistical method, using adjusted mutual information (AMI) as the basis for similarity measurements between pairs of languages. Figure 1 illustrates a pairwise comparison between northern Australian languages, Murrinhpatha and Anindilyakwa, with coloured bands representing the number of concepts that are co-classified by pairs of classes. In this instance, there is a moderate degree of similarity ( $AMI = 0.35$ ), largely due to alignment between the inanimate and animate classes. We then analyse all such pairwise similarity metrics, as illustrated in Figure 2, where deeper colours represent greater partition similarity. Preliminary results indicate that there is not a strong similarity-clustering with regions: for example, many Australia:Amazon language pairs are just as similar as pairs within Australia. These results also suggest an effect that we had not anticipated: higher similarity scores generally occur between languages with richer classification into largish number of semantically motivated classes (e.g. Murrinhpatha:Anindilyakwa, in Figure 1), while languages with fewer, more semantically bleached classes (such as the masc/fem classification in Alamblak, and other Papuan languages) appear to have generally low similarity to other languages.

Further work will include coding of additional Amazonian and Papuan languages, and testing for whether any areal effect can be detected in a regression model. This will use areal location as a predictor of classification similarity, while controlling for semantic richness of the classification system, and phylogenetic distance between languages. While several limitations can be identified in this method (e.g. lack of cognacy coding), it nonetheless provides a novel test of areal semantics in nominal classification.



Table 1. Examples of noun classification in Australian and Amazonian languages

	Daly		Arnhem		NW Amazon	
	Murrinhpatha	M. Tjevin	Wubuy	Mawng	Bora	Muinane
BIRD	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ngarra-</i> (IV)	<i>na-</i> (I)	<i>-co</i> (POINT)	<i>-i</i> (ROUND)
FISH	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ngarra-</i> (IV)	<i>na-</i> (I)	<i>-be</i> (MASC)	<i>-bo</i> (MASC)
WASP	<i>ku</i> (ANIM)	<i>awu</i> (ANIM)	<i>ana-</i> (II)	<i>niny-</i> (II)	<i>-co</i> (POINT)	<i>-cu</i> (ROUND)
FIREWOOD	<i>thungku</i> (FIRE)	<i>tjendji</i> (FIRE)	<i>ngarra-</i> (IV)	<i>ma-</i> (III)	<i>-wa</i> (PLANK)	<i>-gai</i> (PLANK)
FIRESTICK	<i>thungku</i> (FIRE)	<i>tjendji</i> (FIRE)	<i>ngarra-</i> (IV)	<i>niny-</i> (II)	<i>-ho</i> (OBLONG)	<i>-hu</i> (OBLONG)
ALCOHOL	<i>kura</i> (LIQ)	<i>wudi</i> (LIQ)	<i>ngarra-</i> (IV)	<i>nung-</i> (IV)	<i>-pacyo</i> (LIQ)	<i>faihu</i> (LIQ)

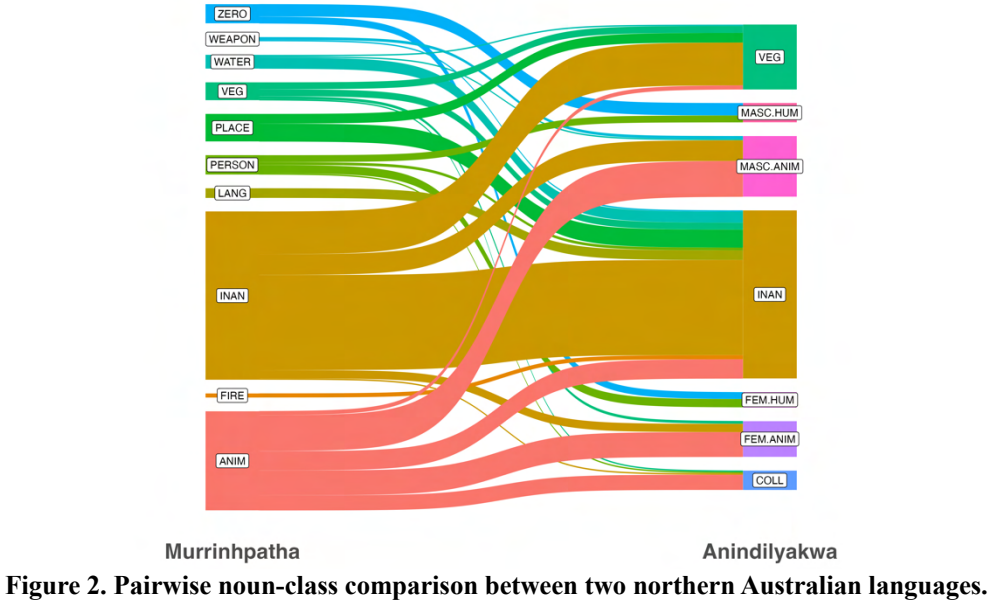


Figure 2. Pairwise noun-class comparison between two northern Australian languages.

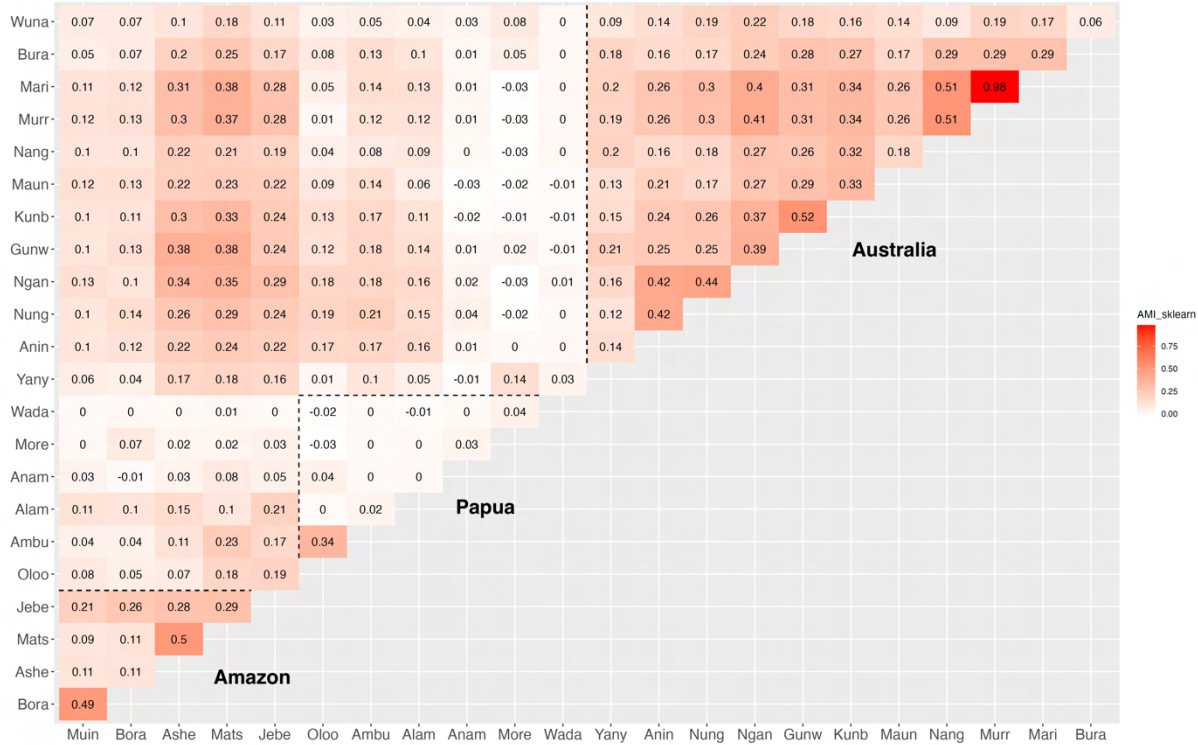


Figure 3. Heatmap of similarities between all language pairs.

## References

- Allasonnière-Tang, Marc & Lundgren, Olof & Robbers, Maja & Cronhamn, Sandra & Larsson, Filip & Her, One-Soon & Hammarström, Harald & Carling, Gerd. 2021. Expansion by migration and diffusion by contact is a source to the global diversity of linguistic nominal categorization systems. *Humanities and Social Sciences Communications*. Palgrave 8(1). 1–6. (doi:10.1057/s41599-021-01003-5)
- Dixon, R. M. W. 1986. Noun classes and noun classification in typological perspective. In Craig, Colette G. (ed.), *Noun Classes and Categorization: Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983* (Typological Studies in Language), 105–112. John Benjamins Publishing Company. (doi:10.1075/tsl.7.09dix)
- François, Alexandre. 2022. Lexical tectonics: Mapping structural change in patterns of lexification. *Zeitschrift für Sprachwissenschaft*. De Gruyter 41(1). 89–123. (doi:10.1515/zfs-2021-2041)
- Green, Ian. 2003. The genetic status of Murrinh-patha. In Evans, Nicholas (ed.), *The non-Pama-Nyungan Languages of Northern Australia*, 125–158. Canberra: Pacific Linguistics.
- Kapitonov, Ivan. 2021. *A Grammar of Kunbarlang* (Mouton Grammar Library). Vol. 89. Berlin: De Gruyter Mouton. (<https://doi.org/10.1515/9783110747058>)
- Koptjevskaja-Tamm, Maria & Liljegren, Henrik. 2017. Semantic patterns from an areal perspective. In Hickey, Raymond (ed.), *The Cambridge Handbook of Areal Linguistics* (Cambridge Handbooks in Language and Linguistics), 204–236. Cambridge: Cambridge University Press. (doi:10.1017/9781107279872.009)
- List, Johann-Mattis & Rzymiski, Christoph & Tresoldi, Tiago & Greenhill, Simon & Forkel, Robert (eds.). 2019. *CLICS<sup>3</sup>*. Jena: Max Planck Institute for the Science of Human History. (<https://clics.clld.org/>) (Accessed August 28, 2023.)
- McCarthy, Arya D. & Williams, Adina & Liu, Shijia & Yarowsky, David & Cotterell, Ryan. 2020. Measuring the similarity of grammatical gender systems by comparing partitions. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 5664–5675. Online: Association for Computational Linguistics. (doi:10.18653/v1/2020.emnlp-main.456)
- Mel’cuk, Igor. 2006. Gender and noun class. *Aspects of the theory of morphology*, 322–383. De Gruyter Mouton. (doi:10.1515/9783110199864.2.322)
- Schapper, Antoinette & Koptjevskaja-Tamm, Maria. 2022. Introduction to special issue on areal typology of lexico-semantics. *Linguistic Typology*. De Gruyter Mouton 26(2). 199–209.
- Seifart, Frank. 2018. The semantic reduction of the noun universe and the diachrony of nominal classification. In McGregor, William B. & Wichmann, Søren (eds.), *The Diachrony of Classification Systems* (Current Issues in Linguistic Theory), 9–32. John Benjamins Publishing Company. (doi:10.1075/cilt.342.02sei)

# Investigating drivers of lexical change by following colexification patterns

Kim Gfeller & Paul Widmer  
(Universität Zürich & Universität Zürich)

Keywords: Colexification, lexical change, phylogenetic methods, change rates, polysemy

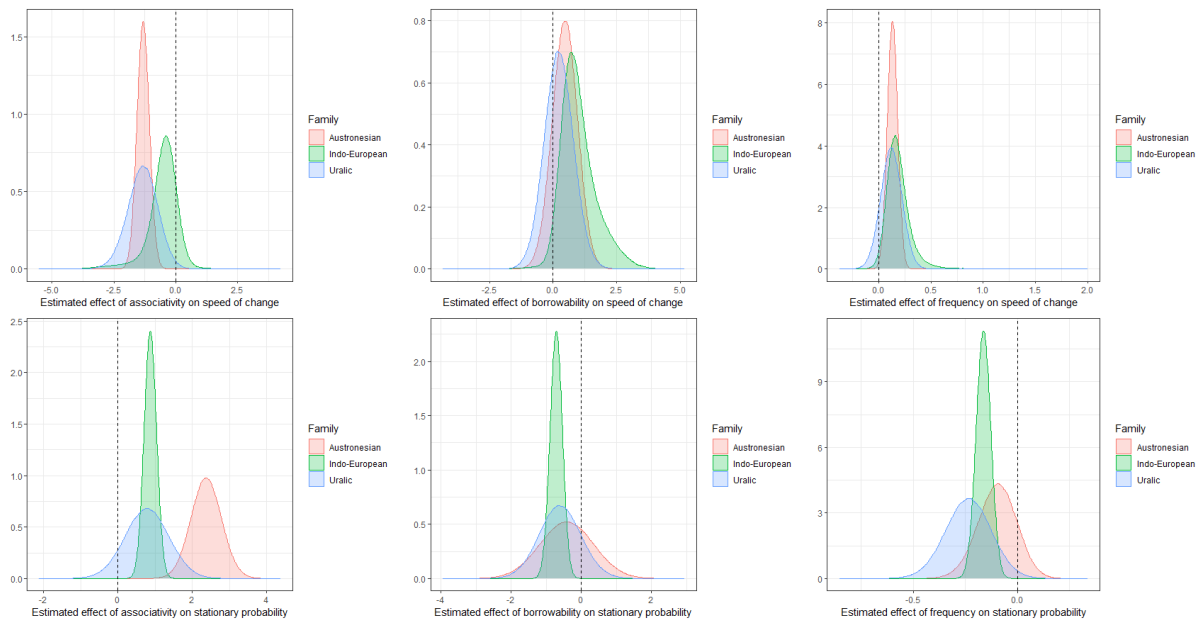
Lexical semantic change involves changes in the meaning of lexemes and in the structure of the lexicon. Compared to other areas of language change, semantic change is less easily reconstructed or predicted, partially because there is a large number of possible changes that the meaning of a lexeme can undergo (Urban 2014). Many studies have investigated samples of attested changes (e.g., Zalizniak et al. 2024, Traugott & Dasher 2002), which may show an incomplete picture, since many languages have only been researched for a short time. In this study, we aim to measure the speed and directionality of changes of specific form-meaning relations and identify correlates of lexical semantic change.

We propose to investigate lexical semantic change using colexification patterns, colexification being the expression of two distinct concepts by a single lexeme (François 2008). It has been shown that related or similar concepts are more likely to colexify cross-linguistically (Xu et al. 2020, Brochhagen & Boleda 2022). Colexification patterns are more likely to be borrowed compared to phonological material (Gast & Koptjevskaja-Tamm 2022).

Colexifications are connected to lexical semantic change because when a lexeme shifts its meaning from A to B, there is an intermediate stage during which the lexeme expresses both meanings A and B simultaneously. At this stage, meanings A and B are colexified. Colexifications are also useful for large-scale studies since they can be derived from dictionary data, which is readily available for many languages.

Due to the connection between colexification and lexical semantic change we can follow the presence or absence of colexification patterns along phylogenies and identify factors that influence changes. In this study, we use dictionary data of three language families – Indo-European, Austronesian, and Uralic – from the CLICS database (Rzyski et al. 2020) and use Bayesian phylogenetic models to estimate change rates and stationary probabilities of common colexifications. We then use a Bayesian regression model to investigate the influence of three lexical properties – associativity, frequency of use, and borrowability – on these rates and probabilities. Associativity reflects the degree of semantic relatedness between words, based on experimental data (e.g., De Deyne et al. 2019 for data from English speakers). Frequency of use is derived from subtitle corpora and indicates how often a concept appears in everyday speech. Borrowability measures the likelihood of a concept being borrowed across languages, using data from the World Loanword Database (Haspelmath & Tadmor 2009).

We find that associativity has a positive effect on the long-term preference for two concepts being colexified but negatively impacts the speed of change. By contrast, frequency of use and borrowability have a positive effect on the rate of change, but a negative effect on the long-term preference. We also observe differences between the three language families, suggesting that cultural variation may play a role.



- Brochhagen, Thomas, and Gemma Boleda (2022), When do languages use the same word for different meanings? The Goldilocks principle in colexification, *Cognition* 226: 105179.
- De Deyne, S., Navarro, D.J., Perfors, A. et al. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research* 51, 987–1006.
- François, Alexandre (2008), Semantic maps and the typology of colexification, *From polysemy to semantic change: Towards a typology of lexical semantic associations* 106: 163-215.
- Gast, Volker, and Maria Koptjevskaja-Tamm (2022), Patterns of persistence and diffusibility in the European lexicon, *Linguistic Typology* 26: 403-438.
- Haspelmath, Martin, and Uri Tadmor (2009), World Loanword Database, <https://wold.clld.org>. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Rzyski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast et al. (2020), The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies, *Scientific data* 7: 13.
- Traugott, E. C., and Dasher, R. B. (2002), *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Urban, Matthias (2014), *Lexical Semantic Change and Semantic Reconstruction*. Routledge Handbook of Historical Linguistics, London - New York: Routledge.
- Xu, Yang, Khang Duong, Barbara C. Malt, Serena Jiang, and Mahesh Srinivasan (2020), Conceptual relations predict colexification across languages, *Cognition* 201: 104280.
- Zalizniak A. et al (2024), Database of Semantic Shifts, <https://datsemshift.ru>. Moscow: Institute of Linguistics, Russian Academy of Sciences.

# Dialexification: A tool for detecting areal and hereditary trends in lexical change

Mathieu Dehouck, Alexandre François & Tom Poulton  
(Lattice, CNRS, ENS–PSL)

*Keywords:* semantic change, etymology, colexification, areality, lexical database

In addition to colexification (François 2008) – wherein two meanings A and B are encoded by the same word in a language – a different sort of link exists when A and B are distributed across two words that share the same etymology. For example, Spanish *esperar* ‘wait’ and French *espérer* ‘hope’, which are cognate, indicate a historical connection between the two senses {WAIT–HOPE}. That connection has been called *dialexification* (François & Kalyan 2023): “two senses are dialexified if they are the meanings of cognate forms”. Dialexification often captures ancient cases of colexification {A–B}, which split up as A was kept in one language and B in another. Such former links can be reconstructed through the work of historical linguists, as they group words into cognate sets.

Just like colexification, dialexification can help assess the relatedness of meanings. Connections can be universal, due to common human experience: eg. the {WATER–RIVER} pair is attested globally. Conversely, some links are restricted to certain areas or families: thus, Ding & Dong (2016) showed how {THUNDER–DRAGON} are prominently colexified in Sino-Tibetan languages.

Patterns of lexification can be inherited along phylogenetic lineages; but they can equally spread across neighbouring languages through contact (Gast & Koptjevskaja-Tamm 2018). While these two types of diffusion can be hard to disentangle (as contact also occurs within phyla), co-/dialexification is a powerful tool for detecting, at least, contact across different families.

“EvoSem” is a database project designed to catalogue patterns of dialexification across the world’s families (Dehouck et al. 2023). As of May 2025, *EvoSem* features approximately 30k concepts, expressed by 215k words from 3,300 languages; these are organised into 25k cognate sets, representing 135 protolanguages. This wealth of data can be exploited to propose some statistical observations.

For every pair of concepts, *EvoSem* lists the etyma that dialexify them. For instance, *EvoSem* finds the pair {WATER–RIVER} in 33 different cognate sets, belonging to numerous families: Arawak, Austronesian, Dravidian, Indo-European, Semitic, Tibeto-Burman, Turkic...; this dialexification is globally well distributed. Conversely, among the 14 etyma dialexifying {FIRE–FIREWOOD}, six come from Pama-Nyungan alone [Figure]: this proves that this pattern is overrepresented in Australian languages – confirming earlier studies (Schapper et al. 2016).

14 etyma dialexifying **firewood** — **fire**

FAMILY	ETYMON	REFLEXES
Algonquian	*mehše	Algonquin misan
		Miami mihsi
		Wiyot bás
		Yurok mech
Austronesian	*CuNuh	
Austronesian	*Sapuy	
Indo-European	*h <sub>2</sub> éydʰ-o-s	Sanskrit रुध् + édha
		Old Irish áed
Malayo-Polynesian	*aluten	
Malayo-Polynesian	*suluq	
Mon-Khmer	*()ʔus	
Pama-Nyungan	*cuma	
Pama-Nyungan	*puri	
Pama-Nyungan	*waca	
Pama-Nyungan	*wi:	
Pama-Nyungan	*yaʔu	
Pama-Nyungan	*waru	Manjilyjara waru
		Warlpiri warlu
Uto-Aztecan	*kut	

Figure – Table from *EvoSem*, listing all the etyma that dialexify {FIRE–FIREWOOD}

In this talk, we present new statistical techniques to measure dialexifications according to geographic or genetic proximity. For every language family or area, we can calculate the degree to which semantic pairs deviate from their expected co-occurrence frequency, as measured globally. This way, we can empirically identify which dialexifications are widespread around the world, vs. which are regionally or genetically specific – based on their deviation from global tendencies. Thus, a pattern overrepresented in a certain area, especially if it involves different families, is likely to reflect areal diffusion.

In sum, the concept of dialexification, and the *EvoSem* database built on it, provide a heuristic tool for diagnosing areal signals, and for outlining the lexical profiles of different regions around the world.

## References

- Dehouck, Mathieu, Alexandre François, Siva Kalyan, Martial Pastor & David Kletz. 2023. *EvoSem: A database of polysemous cognate sets*. In *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, 66–75. Singapore: Association for Computational Linguistics.
- Ding, Hongdi; Sicong Dong. 2024. Colexification of "thunder" and "dragon" in Sino-Tibetan languages. *Asia Pacific Translation and Intercultural Studies* 11 (2): 77–99.
- François, Alexandre. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In Martine Vanhove (ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, 163–215. Amsterdam: Benjamins.
- François, Alexandre & Siva Kalyan. 2023. Dialexification: A tool for studying cross-linguistic patterns of semantic change. Paper presented at ICLC16 (16th International Cognitive Linguistics Conference), Heinrich-Heine- Universität, Düsseldorf.
- Gast, Volker & Maria Koptjevskaja-Tamm. 2018. The areal factor in lexical typology. In Daniël Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of Linguistic Variation*, 43–82. Berlin, Boston: De Gruyter.
- Schapper, Antoinette; San Roque, Lila; Hendery, Rachel (2016). Tree, firewood and fire in the languages of Sahul. In Päivi Juvonen; Maria Koptjevskaja-Tamm (eds.). *The Lexical Typology of Semantic Shifts*, 355–422. Berlin, Boston: De Gruyter.

# Number-driven dislexification in the expression of size properties: The case of East Caucasian

Rita Popova & Michael Daniel

(Saarland University & University of Tübingen; Laboratoire Dynamique du Langage CNRS)

Keywords: East Caucasian, dislexification, number, suppletion, property words

East Caucasian languages is a family of somewhat 50 languages indigenous to the Caucasus. Some languages of the family demonstrate a peculiar (dis-)lexification pattern (in terms of François 2022, Haspelmath 2023) in the semantic domain of size properties. Namely, the property of being small in size is expressed by two unrelated stems depending on the number of objects it is ascribed to. Referring to one small entity, one stem is used, while the other is used when referring to multiple small objects. The same may hold for the property of being big in size, as (1) demonstrates.

(1) Tsakhur:

- |    |                         |                  |                   |               |
|----|-------------------------|------------------|-------------------|---------------|
| a. | <i>ma-na</i>            | <i>xe-b-na</i>   | <i>iš</i>         | <i>diš-ē?</i> |
|    | this-ATR.NPL            | big.SG-3-ATR.NPL | matter            | COP.NEG-Q1    |
|    | 'Isn't it a big deal?'  |                  |                   |               |
| b. | <i>č'ek-in</i>          | <i>umud-bi</i>   | <i>wo-d-im-mi</i> |               |
|    | big.PL-ATR              | hope-PL          | be-NPL-ATR-PL     |               |
|    | 'There are high hopes.' |                  |                   |               |

(Kibrik 1999: 773, 859)

While number-driven stem alternations in the expression of size properties are reported for diverse languages elsewhere in the world, the pattern is overall cross-linguistically rare (Nurmio 2017, Popova & Daniel 2023). The latter study also observes that the alternation of stems is not characteristic of property words from other semantic fields, such as 'long', 'old', 'red', etc.

In our talk, we present data on lexification patterns observed in this domain across East Caucasian languages. First, we demonstrate that within East Caucasian, it is attested in the areally separated (south, central, northwest) languages belonging to different branches (Lezgic, Lak, Andic and Nakh) which are not in direct contact with each other. Such a broad distribution of a cross-linguistically uncommon phenomenon is intriguing. It is not immediately clear what type of evolutionary scenario can account for such a distribution.

Second, we demonstrate that in all languages in our sample, this stem alternation can be attributed to a single pathway of historical semantic change, also observed in the languages discussed by Maiden (2014) and Nurmio (2017). In languages such as Megleno-Romanian or Swedish, the present-day suppletive plural allomorph of 'small' or 'big' at the ancestral stage was a separate lexeme that was used to describe structure of substances or entities composed of numerous parts, e.g. modern Swedish *små* 'small.PL' developed from Old Norse *smár* 'composed of small parts' as in *smár skógr* 'forest consisting of small trees' (Nurmio 2017).

Different East Caucasian languages are at various stages of a similar process, whereby the use of the 'component structure' word evolves from a lexical choice into a grammatically conditioned stem alternation. We explore the possibility of an explanation that accounts for the distribution of number-governed adjectives of size, incorporating both genealogical inheritance and contact-induced change.

## References

- François, Alexandre. 2022. Lexical tectonics: mapping structural change in patterns of lexification. *Zeitschrift für Sprachwissenschaft* 41(1). 89–123.
- Haspelmath, Martin. 2023. Coexpression and synexpression patterns across languages: comparative concepts and possible explanations. *Frontiers in Psychology* 14.
- Kibrik, Alexander E. 1999. *Elementy tsaxurskogo jazyka v tipologičeskom osveščanii* [Studies in Tsakhur: a typological perspective ]. Moskva [Moscow]: Nasledie. 943.
- Maiden, Martin. 2014. Two suppletive adjectives in megleno-romanian. *Revue Romane* 49.
- Nurmio, Silva. 2017. The development and typology of number suppletion in adjectives. *Diachronica* 34(2). 127–174.
- Popova, Rita & Michael Daniel. 2023. Small is big: number-governed dislexification in property words. Manuscript.



## Areal spread and genealogical inheritance of discourse formulas in Daghestan

Nina Dobrushina

(Laboratoire Dynamique du Langage (CNRS, Lyon))

Formulas, conversational routines, language contact, inheritance, Daghestan

There is a common belief that pragmatically motivated, interactional discourse phenomena are easily borrowed (Beier & Sherzer 2002; Aikhenvald 2007; Andersen 2014; Daniels & Brooks 2019). Although the diffusion of discourse formulas (aka conversational routines) may indicate intensive communication across language borders and reveal large linguistic areas, it is severely understudied. Even less is known about the inheritance of discourse formulas: to what extent are discourse formulas stable across generations? as easy targets for borrowing, can they resist the influence of contact?

In this talk, I will analyze the interplay of contact and inheritance factors in several discourse formulas in the languages of Daghestan and their neighbors (about 45 languages), using elicitation, grammars and dictionaries as sources.

The Republic of Daghestan is an area of high linguistic density. Most of the languages spoken in Daghestan belong to the Nakh-Daghestanian (East Caucasian) language family. Despite this linguistic diversity, the peoples of Daghestan are culturally very homogeneous. They are almost exclusively Muslim, share many similarities in their occupations and customs, and are in close contact with each other. There was never a single lingua franca common to all of Daghestan before the arrival of the Russian language in the mid-20th century, but in some parts of North and South Daghestan Avar and Azerbaijani were used as lingua francas (Chirikba 2008: 30).

Among the formulas studied in this paper are commemorative formulas. These are expressions used as modifiers to accompany a reference to a deceased person. The study shows that there is a great variation across the region, and that both matter and structural copying are abundant. There are three most common patterns that roughly correspond to the genealogical grouping of the languages: the speakers of Avar-Ando-Tsezic languages use the structural pattern that can be literally translated as ‘May the sins be washed away’ (ex. 1), the speakers of Lezgian languages use the formulas that include the Arabic word *rahmat* (most probably borrowed via Azerbaijani) (ex. 2), while the speakers of Lak and Dargwa languages use the formula with another matter borrowing - *ʔaʔpa* (‘forgive’ in Arabic) (ex. 3).

(1) Anchiq dialect of Karata

<i>cʼaqʼo-w</i>	<i>hekʷa</i>	<i>w-ukʷa</i>	<i>munah</i>	<i>m-uč̣a-w</i>
good-M	men	M-be.PST	sin	N-wash.PTCP-M

‘He was a good man, **may his sins be washed away**’ (Gadzhimagomedov 2019)

(2) Lezgian

<b>rähmet xa-ji</b>	Mehamd-a-n	rikʼ	ala-j	zatʼ
mercy be-AOR.PTCP	Mehammad-OBL-GEN	heart	be.on-PTCP	thing
tir	ima			
COP.PST	this			

‘The late Mehammad loved this thing.’ (Tsnal dialect, Fazir Dzhaliev, p.c.)

(3) Standard Dargwa

<b>ʔa'pa</b>	<b>b-arq'</b>	ʔa'h-si	adam	w-iri
<b>forgiveness</b>	<b>n-do.pfv.OPT</b>	good-ATTR	men	M-be.PST

‘The deceased was a good man.’ (Yusupov 2014: 350)

The distribution of some other previously studied formulas is largely similar (Naccarato & Verhees 2021). In all cases, we see a very strong areal signal; no other phonological or grammatical phenomena of contact origin spans such vast areas in North Caucasus.

Since the areal distribution of formulas strongly overlaps with the genealogical distribution, it is difficult to distinguish inheritance from the result of contact. In order to do this, I consider several outliers: speech communities located outside their genealogical area. They clearly show that the formulas are inherited, but are quickly substituted by a borrowing under certain conditions. I conclude that the areal signal is by far stronger in the diffusion of discourse formulas than the genealogical signal.

#### References

- Aikhenvald, Alexandra (2007). Semantics and pragmatics of grammatical relations in the Vaupés linguistic area. Oxford University Press. 237-266.
- Andersen, Gizle (2014). Pragmatic borrowing. *Journal of Pragmatics*, 67, 17-33.
- Beier, Christine, Michael, Lev, & Sherzer, Joel (2002). Discourse forms and processes in indigenous lowland South America: An areal-typological perspective. *Annual Review of Anthropology*, 31(1), 121-145.
- Chirikba, Viacheslav A. (2008). The problem of the Caucasian Sprachbund. In: Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 25–94. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Daniels, Don, and Joseph Brooks (2019). The History of \*= a: Contact and Reconstruction in Northeast New Guinea. *Journal of Language Contact* 12.3: 533-568.
- Gadžimagomedov, Magomed Z. (2019). *Ančixskij jazyk [Anchiq]*.
- Naccarato, Chiara and Samira Verhees (2021). Morning greetings (Maps & Data). In: *Typological Atlas of the Languages of Daghestan (TALD)*. <http://lingconlab.ru/dagatlas>

# A Language Model perspective on cross-linguistic colexification patterns

Roksana Goworek, Maria Koptjevskaja-Tamm & Haim Dubossarsky

(Queen Mary University of London; Stockholm University; & Queen Mary University of London, University of Cambridge and The Alan Turing Institute)

Keywords: colexification, language model, genetic relations, language contact, cross-linguistic

Cross-linguistic colexification research has gained significant interest in recent years, with new datasets curated and innovative analysis methods developed (Vanhove ed. 2008; Zalizniak et al. 2012; Juvonen & Koptjevskaja-Tamm eds. 2016; Rzymiski et al. 2020). Here we propose Multilingual Language Models (mLMs) as a powerful framework to investigate cross-linguistic colexification on both large- and fine-scales analyses. mLMs can be thought of as polyglots: trained on dozens of languages, they acquire implicit knowledge about linguistic patterns. Here, we teach them to disambiguate polysemous words in one language, and ask a simple but revealing question: *how useful is this training when the model is tested on a different language?*

The underlying assumption is intuitive: more similar languages should support ('transfer to' in NLP terms) each other more effectively (Lauscher et al. 2020; He et al. 2022). This similarity is usually attributed to shared genetic inheritance, areal contact, and their interaction, along with possible universal patterns, remains a core issue in typological research (Gast & Koptjevskaja-Tamm 2018; Jackson et al. 2019; Georgakopoulos et al. 2022; Liljegren 2022; Norcliffe & Majid 2024). Our experimental setup offers a novel lens through which to analyze cross-linguistic colexification. We present a large-scale cross-linguistic analysis that leverages the polyglot nature of mLMs to systematically evaluate this question at both macro and micro levels. Our factorial design compares transfer performance across language pairs that differ in genetic relatedness and contact history. For instance, we examine transfer from Polish to Russian and from German to Hindi, pairs with varying degrees of genetic and areal affinity, and contrast these with more distant combinations like Azerbaijani–German. In each case, we assess how well colexification knowledge in language X supports polysemy disambiguation in language Y. We observe that both genealogical closeness and areal contact can enhance transfer.

We complement this investigation with a micro-level analysis of dozens of colexified words across 10 languages that were collected by native speakers (Goworek et al. 2025), providing a detailed account of specific colexification examples and identifying which trained languages contribute most to their disambiguation. Findings like this can prompt in-depth investigations into previously unnoticed colexification patterns across languages, and more broadly, into the types of semantic overlaps that enabled them, and their possible universality. These overlaps are crucial for understanding semantic change given that polysemy and colexification often mark transitional stages in the evolution of word meaning (François 2008). By highlighting such patterns at scale, our approach provides empirical support for hypothesized semantic trajectories, revealing latent relationships that may foreshadow future shifts in meaning.

The results raise meta-theoretical questions about how mLMs perceive polysemy and homonymy. Do they grasp semantic ambiguity like humans, or detect different patterns? As AI plays a growing role in linguistic analysis, understanding where human and machine language abilities align or differ is crucial—for computational linguistics and for the future of typology, semantics, and language change.

Our method also tackles key methodological issues in cross-linguistic research. Prior work shows that choices like word lists or language comparisons can bias results (Schapper & Koptjevskaja-Tamm 2022). By using transfer-based evaluation, our approach—applicable to 100+ mLM-supported languages—avoids such subjective choices. As studying lexico-semantic patterns is vital to understanding cross-linguistic colexification, scalable and reproducible methods like ours offer a promising path for more robust and diverse future research.

#### Acknowledgements

This research was partly supported by the research program “Change is Key!”, supported by Riksbankens Jubileumsfond (under reference number M21-0021).

#### References

- François, Alexandre (2008), Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (ed), (2008), *From polysemy to semantic change: Towards a typology of lexical semantic associations*, Amsterdam / Philadelphia: John Benjamins Publishing Company, 163–215.
- Gast, Volker, and Maria Koptjevskaja Tamm (2018), The Areal Factor in Lexical Typology: Some Evidence from Lexical Databases. In D. Van Olmen, T. Mortelmans, and F. Brisard (eds), (2018), *Aspects of Linguistic Variation*. Berlin: De Gruyter Mouton, 43–81. <https://www.degruyter.com/viewbooktoc/product/506391>.
- Georgakopoulos, Thanasis, Grossman, Eitan, Nikolaev, Dmitry and Polis, Stéphane (2022), Universal and macro-areal patterns in the lexicon: A case-study in the perception-cognition domain, *Linguistic Typology*, 26 (2), 439–487. <https://doi.org/10.1515/lingty-2021-2088>.
- Goworek, Roksana, Harpal Karicut, Muhammad Shezad, Nijaguna Darshana, Abhishek Mane, Syam Bondada, Raghav Sikka, Ulvi Mammadov, Rauf Allahverdiyev, Sriram Purighella, Paridhi Gupta, Muhinyia Ndegwa and Haim Dubossarsky (2025). SenWiCh: Sense-Annotation of Low-Resource Languages for WiC using Hybrid Methods. arXiv preprint arXiv: 2505.23714.
- He, Junxian, Zhou, Chunting, Ma, Xuezhe, Berg-Kirkpatrick, Taylor and Neubig, Graham (2022), Towards a Unified View of Parameter-Efficient Transfer Learning, *International Conference on Learning Representations*. <https://openreview.net/forum?id=ORDcd5Axok>
- Jackson, Joshua Conrad, Joseph Watts, Teague R. Henry, Johann-Mattis List, Robert Forkel, Peter J. Mucha, Simon J. Greenhill, Russell D. Gray, and Kristen A. Lindquist (2019), Emotion Semantics Show Both Cultural Variation and Universal Structure, *Science* 366 (6472) (December 20, 2019): 1517–22. <https://doi.org/10.1126/science.aaw8160>.

- Juvonen, Päivi, and Maria Koptjevskaja-Tamm (eds) (2016), *The Lexical Typology of Semantic Shifts*. Cognitive Linguistics Research 58. Berlin/New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110377675>.
- Lauscher, Anne, Vinit Ravishankar, Ivan Vulić, and Goran Glavaš (2020), From Zero to Hero: On the Limitations of Zero-Shot Language Transfer with Multilingual Transformers, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4483-99. <https://doi.org/10.18653/v1/2020.emnlp-main.363>
- Liljegren, Henrik (2022), Kinship Terminologies Reveal Ancient Contact Zone in the Hindu Kush, *Linguistic Typology* 26 (2), 211–45. <https://doi.org/doi:10.1515/lingty-2021-2080>.
- Norcliffe, Elisabeth, and Asifa Majid (2024), Verbs of Perception: A Quantitative Typological Study, *Language* 100 (1), 81–123.
- Rzymiski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast, et al. (2020), The Database of Cross-Linguistic Colexifications, Reproducible Analysis of Cross-Linguistic Polysemies, *Scientific Data* 7 (1) (January 2020). <https://doi.org/10.1038/s41597-019-0341-x>.
- Schapper, Antoinette & Koptjevskaja-Tamm, Maria (2022), Introduction to special issue on areal typology of lexico-semantics, *Linguistic Typology* 26(2), 199–209. <https://doi:10.1515/lingty2021-2087>
- Vanhove, Martine (ed) (2008), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*. Studies in Language Companion Series 106. Amsterdam: John Benjamins.
- Zalizniak, Anna A., Maria Bulakh, Dmitrij Ganenkov, Ilya Gruntov, Timur Maisak, and Maxim Russo (2012), The Catalogue of Semantic Shifts as a Database for Lexical Semantic Typology, *Linguistics* 50 (3), 633–69. <https://doi.org/doi:10.1515/ling-2012-0020>

## **WS5 Empirical studies on syntactic alternation across languages and theoretical frameworks**

**Yanis Da Cunha, Pegah Faghiri & Juliette Thuilier**

# A cross-linguistic experiment on the role of emotion intensity in the psych alternation

Addison J. Blanchard-Rooney & Niklas Wiskandt

(CNRS UMR 5565 Dynamique du Langage, Université Lumière Lyon 2 & Heinrich-Heine-Universität Düsseldorf)

Keywords: emotions, psych verbs, psych alternation, experiment, Romance languages

We report on a novel perspective on the psych alternation (cf. e.g. Rott et al. 2020), employing experimental methodology to investigate a yet unexplored semantic effect within a theoretically-dominated domain.

In Romance languages, speakers can opt to realize the experiencer of a psych verb in one of three ways, as demonstrated here in French (Legendre 1993: 373): As the object in an active voice construction (1a), or as the subject in a pronominal voice construction using a detransitivizing reflexive clitic (1b). These two constructions form syntactic alternation pairs. Stative passives, as shown in (1c), represent yet a further possibility.

- (1) a. *Le sort de-s personne-s âgé-e-s préoccupe Pierre.*  
DEF.M fate of-PL person-PL aged-F-PL worry.3SG Pierre  
'The fate of old People worries Pierre.'
- b. *Pierre se préoccupe du sort de-s personne-s âgé-e-s.*  
Pierre REFL worry.3SG of.DEF.M fate of-PL person-PL aged-F-PL  
'Pierre worries about the fate of old people.'
- c. *Pierre est préoccup-é du sort de-s personne-s âgé-e-s.*  
Pierre be.3SG worry.PTCP.M of.DEF.M fate of-PL person-PL aged-F-PL  
'Pierre is worried by the fate of old people.'

While previous research on this topic has been primarily theoretical (e.g., Belletti & Rizzi 1988 for Italian; Royo 2018 for Catalan), this study takes an empirical perspective, investigating the role of emotional intensity on speakers' choice of voice construction. Our cross-Romance syntactic experiment employs 16 piloted visual stimuli, which portray transitive actions between two people, of which one is the experiencer of an emotion. Actions and facial expressions were devised in accordance with the six basic universal emotions (Ekman 1999), and with two levels of emotion intensity (high vs. low) for each basic emotion.

In an online experiment, 49 native speakers of Catalan, Spanish, French, Italian, Portuguese, and Romanian are shown 10 stimuli, each accompanied by a psych verb. Each emotion is shown exactly once, and both intensities at least once. Additional controlled filler stimuli were created. Each stimulus is presented individually. Participants are tasked to generate a written description of the stimulus and to use the verb from the caption in their description. These descriptions are analyzed from a qualitative morphosyntactic and semantic perspective, and from a quantitative perspective, using multinomial

logistic regression to model the effect of emotion intensity within and across emotion classes and languages. Moderation by the emotion itself is also calculated and sociodemographic covariates considered.

Regarding an effect of emotional intensity on the voice alternation, we expect an increased perception of causality with active voice to induce a correlation between active voice and higher emotional intensity. We also assume that the emotional category itself moderates this effect in an interaction term, i.e., the strength of the expected voice tendencies to not be equally pronounced across all verbs. When comparing the languages of investigation, we predict similar tendencies across the language family as a whole, however with language-dependent effect strengths. Strikingly, other considered covariates, e.g., participant bias and emotion class, show stronger effects in the data.

The mixed-methods nature of the experiment, as well as the novel consideration of emotional intensity as an influencing factor, offer a new perspective into this syntactic alternation. Furthermore, consideration of additional covariates provides significant results. Our findings aid in creating a more wholistic understanding of speakers' motivation in choosing a morphosyntactic construction when presented with multiple options.

## Glossary

3	third person	M	masculine	REFL	reflexive
DEF	definite	PL	plural	SG	singular
F	feminine	PTCP	participle		

## References

- Belletti, Adriana and Luigi Rizzi (1988), Psych-verbs and  $\theta$ -theory, *Natural Language and Linguistic Theory* 6(3), 291–352. doi:10.1007/BF00133902.
- Ekman, Paul (1999), Basic emotions, in T. Dalgleisch, and M. Power (eds.), (2009), *Handbook of Cognition and Emotion*, Chapter 3, Chichester: John Wiley & Sons Ltd., 45–60.
- Legendre, Géraldine (1993), Antipassive with French psych-verbs, in *Proceedings of the twelfth west coast conference on formal linguistics*, Stanford, CA: CSLI Publications, 373–388.
- Rott, Julian A., Elisabeth Verhoeven and Paola Fritz-Huechante (2020), Valence orientation and psych properties: Toward a typology of the psych alternation, *Open Linguistics* 6(1), 401–423. doi:10.1515/opli-2020-0020.
- Royo, Carles (2018), Els verbs psicològics pronominals catalans i l'alternança emotiva/volitive, *Études romanes de Brno* 39(1), 115–138. doi:10.5817/ERB2018-1-8.



## Escaping (from) the dative alternation. Alternating indirect objects without a direct object.

Alexander Van Herpe & Dirk Pijpops  
(Universiteit Antwerpen)

Keywords: dative alternation, indirect object, language processing, constructional grammar, collocation analysis.

The Dutch dative alternation between the ditransitive and the prepositional dative construction as in (1)-(2), is for a large part determined by processing-related factors, such as the givenness and complexity of theme and recipient, that have a stable influence among various alternating verbs – although lexical effects do appear stronger than in English (Bernolet & Coleman 2016, Coleman 2020 and Engel et al. 2024). Meanwhile, for the Dutch transitive-prepositional alternation, where the direct object alternates with a prepositional realization, as in (3)-(4), construal-related differences that are tied to specific verbs appear comparatively stronger predictors. For example, for *verlangen* ‘desire’ in (3), the prepositional variant is much more likely when the desire is construed as a longing, whereas the transitive variant is preferred if the desire is construed as a demand (Pijpops et al. 2021). Much like the dative alternation, the choice between both variants is multifactorially determined and processing-related factors do come into play, but unlike the dative alternation, they play second fiddle and usually only overrule the difference in construal in extreme cases.

- (1) *Alexander had (naar) Parmenion reeds versterkingen gezonden.*  
‘Alexander had already sent Parmenion reinforcements.’
- (2) *Voor de veldslag had hij (aan) zijn generaals duidelijke instructies gegeven.*  
‘Before the battle, he had given his generals clear instructions.’
- (3) *We zijn gewoon twee mensen die (naar) een kind verlangen.*  
‘We are just two people that long for a child.’
- (4) *Samen bouwen we (aan) een mooie toekomst.*  
‘Together, we are building a beautiful future.’

Between both alternations, a group of verbs such as *ontsnappen* ‘escape’ in (5) or *ontlopen* ‘avoid’ in (6), can be found. These verbs express some movement by the subject away from the object; other such alternating verbs include *ontvallen* ‘pass away’, *ontgroeien* ‘outgrow’ and *ontkomen* ‘evade’. Functionally, the single object of these two-place verbs is an indirect one (Vandeweghe 2013: 86), and it could therefore be hypothesized that they alternate much like other indirect objects in the dative alternation. That is, we would expect the choice between both variants to be primarily driven by processing-related variables. Formally, however, in taking only a single object, these verbs are at least superficially more alike the verbs of the transitive-prepositional alternation. As such, one could hypothesize that their alternation is primarily determined by construal-related factors tied to individual verbs.

- (5) *Darius wist (aan) zijn achtervolgers te ontsnappen.*

‘Darius managed to escape his pursuers.’

- (6) *De kans om (aan) een controle te ontlopen, was quasi nihil.*  
‘The probability to avoid a check-point, was almost nil.’

We aim to find out if the *aan*-alternation for this group of *ont*-verbs behaves more like the dative- or the transitive prepositional alternation. To address this issue, we will conduct a quantitative corpus study, based on the SoNaR corpus of written Dutch (Oostdijk et al. 2013), following the methodology laid out by Pijpops et al. (2021), making use of collostructional analysis (Gries 2004) and mixed effects logistic regression modelling.

## References

- Bernolet, Sarah & Timothy Coleman. 2016. Sense-based and lexeme-based alternation biases in the Dutch dative alternation. In Jiyoung Yoon & Stefan Thomas Gries (eds.), *Corpus-based Approaches to Construction Grammar*, 165–198. Amsterdam/Philadelphia: John Benjamins.
- Coleman, Timothy. 2020. The emergence of the dative alternation in Dutch: towards the establishment of a horizontal link. In Maria Fedriani, Chiara and Napoli (ed.), *The diachrony of ditransitives*, vol. 351, 137–168. Berlin: De Gruyter. <http://dx.doi.org/10.1515/9783110701371-005>.
- Gries, Stefan Th., Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International journal of corpus linguistics*, 9(1), 97-129.
- Engel, Alexandra, Elsy Andries, Laura Rosseel, Benedikt Szmrecsanyi & Freek Van de Velde. 2024. Constraints and lexical conditioning in the dative alternation. A cross-linguistic analysis of English and Dutch. *Languages in Contrast* Published.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste & Ineke Schuurman. 2013. The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In Peter Spyns & Jan Oostdijk (eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*, 219–247. Heidelberg: Springer.
- Pijpops, Dirk, Dirk Speelman, Stefan Grondelaers & Freek Van de Velde. 2021. Incorporating the multi-level nature of the construction into hypothesis testing. *Cognitive Linguistics* 32(3). 487–528.
- Vandeweghe, Willy. 2013. *Grammatica van de Nederlandse zin*. 8e herzien. Antwerpen: Garant.

# The material/product alternation in French and Italian: A view on production data and within-type variation in two verb framed languages

Barbara Schirakowski, Manuela Treffer & Judith Meinschaefer  
(Freie Universität Berlin)

**Keywords:** material/product alternation, created result, French, Italian, Lexical Mapping Theory

**Background:** Most Romance languages [RL] are considered verb-framed languages (Talmy 1985), as they typically encode the event result within the verb root. A few recent studies have discussed whether certain syntactic alternations that involve non-verbal result expressions are restricted in this language type (Lewandowski 2014).

**Phenomenon:** This study addresses the material/product (M/P) alternation – as a lesser studied type of argument alternation – in French and Italian. In the M/P alternation, the material and the product (result) of a creation event surface as either object DP or PP, cf. (1) and (2-3) for data from this study.

- |     |         |    |   |                        |                        |                           |      |      |
|-----|---------|----|---|------------------------|------------------------|---------------------------|------|------|
| (1) | a.      | DP | V   | DP <sub>PRODUCT</sub>  | PP <sub>MATERIAL</sub> | “effected-object variant” |      |      |
|     | b.      | DP | V   | DP <sub>MATERIAL</sub> | PP <sub>PRODUCT</sub>  | “resultative-PP variant”  |      |      |
| (2) | French  | a. | <i>La personne a sculpté une poupée avec du bois.</i>     |                        |                        |                           |      |      |
|     |         |    | the person has carved a doll with                         | PART                   | wood                   |                           |      |      |
|     |         | b. | <i>La personne a sculpté le bois en (une) poupée.</i>     |                        |                        |                           |      |      |
|     |         |    | the person has carved the wood in a                       |                        | doll                   |                           |      |      |
| (3) | Italian | a. | <i>L'uomo ha intagliato una bambola di legno.</i>         |                        |                        |                           |      |      |
|     |         |    | the man has carved a doll of                              |                        | wood                   |                           |      |      |
|     |         | b. | <i>L'uomo ha intagliato del legno a forma di bambola.</i> |                        |                        |                           |      |      |
|     |         |    | the man has carved  | PART                   | wood                   | to                        | form | of   |
|     |         |    |   |                        |                        |                           |      | doll |

In previous research the resultative-PP alternant (cf. the b-examples) has been deemed largely unacceptable in RL (Folli and Harley 2020). Furthermore, effected objects (cf. the a-examples) are not as widely available as they are, e.g., in English (Mateu 2003).

## Research questions:

- 1) (Under what conditions) does the material/product alternation occur in French and Italian?
- 2) (How) do the languages differ from one another?

**Data:** In a production study, a total of 1347 creation event descriptions was elicited from 96 participants, providing as stimuli both pictures and lexemes (deemed likely to allow the alternation). The data was coded for syntactic means of result lexicalization and for the structure of the resultative PP (if present).

**Results:** The M/P alternation was attested in French and Italian for > 50 % of the lexemes. In both languages, the effected-object alternant rather occurs with obligatorily transitive and agentive verbs (cf. 2-3) than with ergative verbs. Verbs vary with respect to the frequency of the alternants. Differences between the languages are related to prepositional inventories. While French *en* ‘in’ (contrasting with French *dans* ‘in’) yields unambiguous resultative semantics with the attested PPs, Italian PPs headed by *in* are often ambiguous between locative, directional and resultative readings.

Overall, the results show that the created-result frame is available in RL, and that lexical factors play a role for the availability and frequency of usage of both alternants.

**Analysis:** A formal analysis of the M/P alternation must take into account not only lexico-semantic contrasts between verb classes and the semantic fine structure of individual lexemes, but also the semantics of prepositions. This study presents such a formal model of the alternative mappings seen in the M/P alternation within the framework of Lexical-Functional Grammar (Bresnan 1982), drawing on the architectural assumptions and feature-based mapping principles of Lexical Mapping Theory (Bresnan and Kanerva 1989, Findlay, Taylor and Kibort 2023).

## References

- Bresnan, Joan W. (1982), The passive in lexical theory, in J. W. Bresnan (ed.), *The mental representation of grammatical relations*, Cambridge: MIT Press, 3–86.
- Bresnan, Joan W. and Jonni M. Kanerva (1989), Locative Intension in Chichewa: A Case Study of Factorization in Grammar, *Linguistic Inquiry* 20(1), 1–50.
- Findlay, Jamie Y., Roxanne Taylor and Anna Kibort (2023), Argument structure and mapping theory, in M. Dalrymple (ed.), *The Handbook of Lexical Functional Grammar*, Berlin: Language Science Press, 699–778. doi:10.5281/zenodo.10037797.
- Folli, Raffaella and Heidi Harley (2020), A Head Movement Approach to Talmy’s Typology, *Linguistic Inquiry* 51(3), 425–470. [https://doi.org/10.1162/ling\\_a\\_00351](https://doi.org/10.1162/ling_a_00351).
- Lewandowski, Wojciech (2014), The locative alternation in verb-framed vs. satellite-framed languages: A corpus study of Spanish and Polish, *Studies in Language* 38(4), 864–895. <https://doi.org/10.1075/sl.38.4.08lew>.
- Mateu, Jaume (2003), Parametrizing Verbal Polysemy: The Case of *bake* Revisited, *Proceedings of the 39th Annual Meeting of the Chicago Linguistic Society*, 256–269. Chicago, IL: Chicago Linguistic Society.
- Talmy, Leonard (1985), Lexicalization patterns: Semantic structure in lexical forms, in T. Shopen (ed.), *Language typology and syntactic description. Vol. 1 Clause structure*, Cambridge: Cambridge University Press, 57–149.

# Classifier constructions in English and German: Compounds vs. head-classifier constructions

Elnora ten Wolde & Thomas Schwaiger  
(University of Graz & University of Vienna)

Keywords: Noun phrase, Classification, Premodification, Compounding, Functional Discourse Grammar

Classifier constructions serve the purpose of specifying the type of entity denoted, whereby the classifying element conceptually restricts the property, not the referent set (Gunkel and Zifonun 2009). English has three classifying forms: adjective-head (*black tea*), compounds (*fire arms*) and head-classifier constructions (HCCs; *man of faith*). What is interesting about the latter two constructions is that, despite the difference in form, they can often be used without a clearly discernible difference in meaning: e.g. *date of birth* and *birth date*. Nevertheless, corpus data show that there often is a clear preference for one form: e.g. *stone wall* (73,977)/*wall of stone* (2,464) (enTenTen). On the other hand, some meaning relations do not allow for alternatives: e.g. *man of substance*/\**substance man*.

German has four classifying constructions: adjective-head (*schwarzer Tee* 'black tea'), compounds (*Steinmauer* 'stone wall'), post-head genitives (*Zeit der Trauer* 'time of sadness') and post-head prepositional phrases (*Mann von Mut* 'man of courage') (Zifonun 2010). The language seems to prefer compounds such as *Geburtsdatum* 'birth date' (78,402) to *Datum der Geburt* 'date of birth' (367) (deTenTen), although post-head genitives in general are also frequently used. In some cases, however, here too only one form seems to be available: e.g. *Frau von Verstand* 'woman of sense' (Zifonun 2010).

These alternations, or lack thereof, have been little discussed for German (Gunkel and Zifonun 2009, Zifonun 2010, Kopf 2021) and even less for English (Keizer & ten Wolde 2024). Using data from contemporary English and German from the TenTen corpora, the present study explores the differences between the various noun-classifier constructions in these languages, addressing the following questions:

- 1) What alternation preferences can be observed?
- 2) Is there a correlation between the choice of construction and the type of relation between the classifier and the head noun?
- 3) When alternation is possible, what factors determine the choice of a particular form (semantics; syntactic considerations; genre; degree of conventionalization)?

This study focuses on head nouns commonly used in these constructions as has been discussed in previous research, including abstract, concrete, relational and collective first nouns. Up to 100 of the most frequent types per head noun have been extracted from the corpora (including frequency numbers and genre information). The classifying constructions have been coded for type and whether alternation is attested in corpora. The classifying first nouns have been coded for count or non-count, singular or plural and modification restrictions. Preliminary results show that the semantic relation between the two nouns is the primary predictor for the variation selected in both languages, and there are strong parallels between both languages. Finally, working within Functional Discourse Grammar (Hengeveld and Mackenzie 2008), this study explores how respective alternations can be modelled across closely related, yet different, languages.

## References

- Gunkel, Lutz and Gisela Zifonun (2009), Classifying modifiers in common names, *Word Structure* 2(2), 205–218.
- Hengeveld, Kees and J. Lachlan Mackenzie (2008), *Functional Discourse Grammar: A Typologically-Based Theory of Language Structure*, Oxford: Oxford University Press.

- Keizer, Evelien and Elnora ten Wolde (2024), *Of birds of prey and men of honour: Head-classifier constructions in English*, in L. Gardelle, E. Mignot, and J. Neveux (eds), (2024), *Nouns and the Morphosyntax / Semantics Interface*, Cham: Palgrave Macmillan, 181–208.
- Kopf, Kristin (2021), Genitiv- und von-Attribute: Bestimmung des Variationsbereichs, in M. Konopka, A. Wöllstein, and E. Felder (eds), *Determination, syntaktische Funktionen der Nominalphrase und Attribution*, Heidelberg: Heidelberg University Publishing, 135–172.
- Zifonun, Gisela (2010), Possessive Attribute im Deutschen, *Deutsche Sprache* 38(2), 124–153.

### **Corpora**

TenTen Corpus Family. Available via Sketch Engine (<http://www.sketchengine.eu>).  
<https://www.sketchengine.eu/documentation/tenten-corpora/>.

# Alternation in Abui P-indexing

George Saad, Katherine Walker, Eva van Lier

University of Amsterdam, Universität zu Köln, Palacký University Olomouc

## Background and research aim

Abui is a minority language of Alor (Indonesia), in intense contact with Alor Malay, the regional lingua franca. Abui P-indexing on verbs is a very complex system, with alternations conditioned by both lexical factors and semantic factors, such as animacy, definiteness, and affectedness (Kratochvíl 2014). When indexing P arguments, speakers choose between six prefix paradigms, several proclitics, and zero-marking (Kratochvíl, Saad, and Delpada under review). In this talk, we focus on the influence of animacy. In contrast to Abui, Alor Malay has no P-indexing (nor any other verbal morphology).

The contact situation results in an under-documented pattern of bilingualism, recently termed Late/Delayed/Adult Vernacular Production (Saad, Arnold, and Peddie in prep.; Anderbeck 2015; Peddie 2021): children are raised in Alor Malay, but acquire passive competence in Abui through overheard speech (Kratochvíl 2007; Saad 2020). They become active speakers of Abui only in late adolescence/early adulthood as they enter the community of adults, who use Abui in daily life. Under these circumstances, we may expect younger speakers with less input to reconfigure the complex Abui P-indexing system. In this paper we investigate the nature of this potential contact-induced change in Abui animacy-conditioned P-indexing.

## Method/Approach

We use data from a production experiment, in which speakers describe short videoclips showing transitive events, involving animate and inanimate P arguments in combination with specific verbs. Using a cross-sectional approach, we compare younger speakers (under 25) to older speakers (above 40). We combine this with a longitudinal approach, comparing data from a single group of 18 younger speakers, gathered in two waves: when they were (*pre*)*adolescents* (aged 9-16) in 2015 (*T1*) and when they had become *young adults* (aged 17-25) in 2023 (*T2*).

## Results

The data show that i) There is no trend towards overall loss of overt P-indexing in younger speakers ii) In general, younger speakers behave more like older speakers as they age from *T1* to *T2*. (iii) Certain prefixes are used more often by younger speakers to index animate Ps, while other prefixes are used less (compared to older speakers). (iv) Younger speakers innovate: they use many Malay verb stems, which index animate P with an Abui prefix. Thus, compared to older speakers, there is both a spread and a loss of animacy effects on younger speakers' indexing systems, depending both on the indexing strategy and on the lexical verb (class). We compare our results to data from other studies on acquisition of differential argument marking and on contact-induced change in complex argument marking morphology across languages.

## References

- Anderbeck, Karl. 2015. "Portraits of Language Vitality in the Languages of Indonesia." In *Language Documentation and Cultural Practices in the Austronesian World: Papers from ICAL 12 (A-PL 019)*, edited by I Wayan Arka, L. N. Seri Malini, and I.A.M Puspani, 19–47. Canberra: Pacific Linguistics.
- Kratochvíl, František. 2007. "A Grammar of Abui: A Papuan Language of Alor." PhD dissertation, Leiden: Leiden University.
- . 2014. "Differential Argument Realization in Abui." *Linguistics* 52 (2): 543–602. <https://doi.org/10.1515/ling-2013-0072>.
- Kratochvíl, František, George Saad, and Benediktus Delpada. under review. "Abui Language Cameo." In *The Oxford Guide to the Papuan Languages*, edited by Nicholas Evans and Sebastian Fedden. Oxford: Oxford University Press.
- Peddie, Emma. 2021. "Delayed Vernacular Production in Melanesia: A Case Study-Based Analysis." MA (Hons.) thesis, Edinburgh: University of Edinburgh.
- Saad, George. 2020. "Variation and Change in Abui: The Impact of Alor Malay on an Indigenous Language of Indonesia." PhD dissertation, Leiden: Leiden University.
- Saad, George, Laura Arnold, and Emma Peddie. in prep. "Late Vernacular Production in Island South East Asia and the Pacific." Manuscript.



The effect of discourse accessibility on syntactic alternations in English

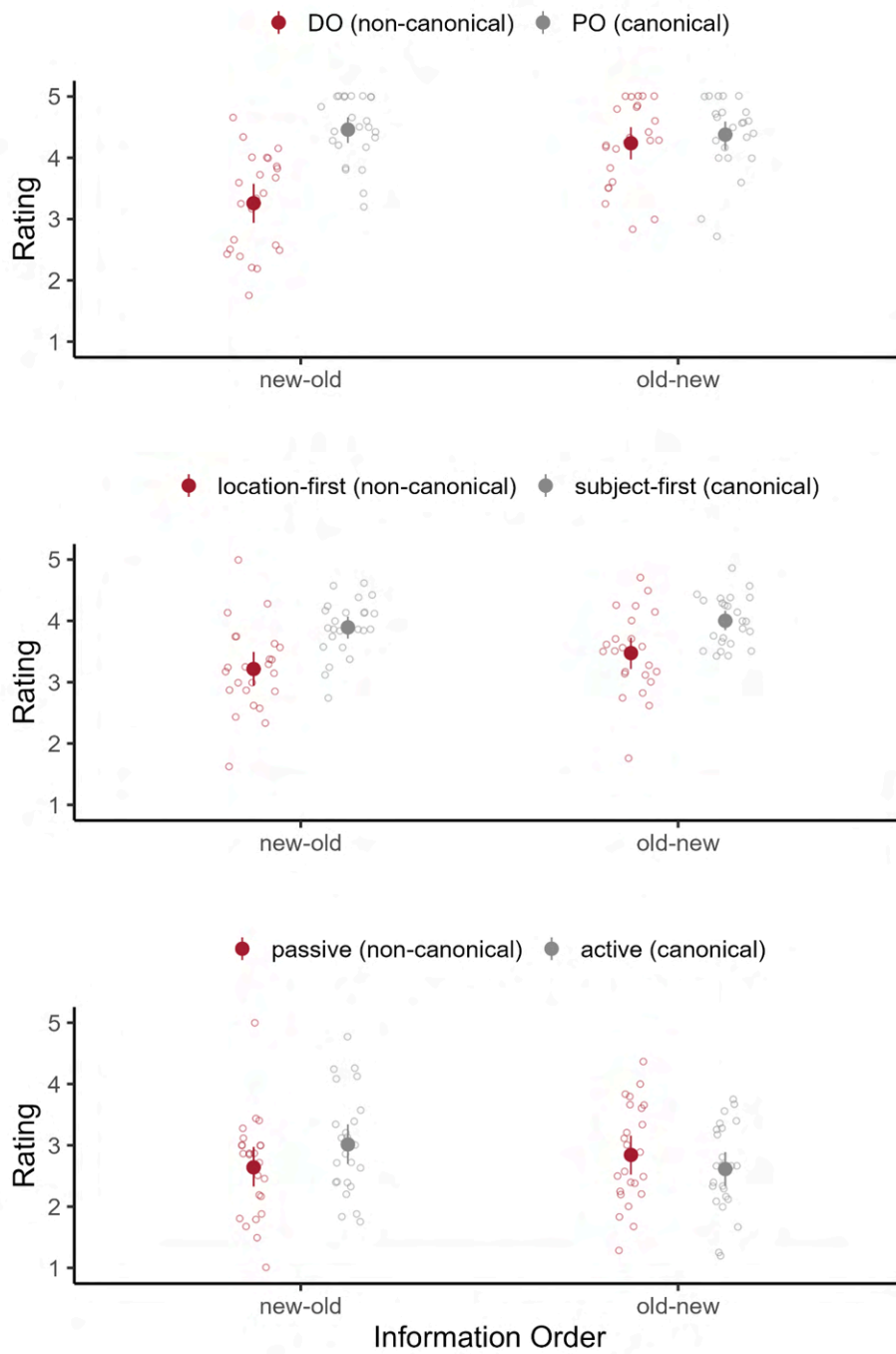
**Introduction.** There are many ways to put the same idea into words. A widely-cited case study is the dative alternation, where the direct object construction (DO) is almost always interchangeable with the prepositional object construction (PO) (see **Sample Items**). But what causes a speaker to choose one construction over the other? In the case of the dative construction, a key idea has been *information structure*, and, specifically, that DO has a preference for old information to precede new information<sup>1</sup>. Given that PO is more frequent (canonical) than DO, perhaps more generally, non-canonical constructions have a preference for old-new word order<sup>2-11</sup>, but evidence for this proposal is limited<sup>13-14</sup>. In contrast, it is possible that the preference for old-new order is an idiosyncratic property of DO. We tested whether non-canonical constructions have a preference for old-new word order, across three English alternations.

**Method.** We conducted three experiments, one for each alternation: dative, locative, and active/passive. On each trial, we presented participants with a context sentence, followed by a target sentence that varied in canonicity (see **Sample Items**). The context sentence mentioned one, two, or none of the arguments in the target sentence (so old arguments were preceded by *the* and new arguments by *a/an*). Thus, each experiment had a 2x2x2 design: the information status of the 2 arguments and the construction (canonical/non-canonical).

**Results & Discussion.** For brevity, we consider only sentences with new-old and old-new information status (**Figure 1**). For each experiment, we fit a Bayesian cumulative regressions with the full random effects structure. The term of interest was the interaction between information order and canonicity. We found that, in the dative and active/passive experiments, old-new was preferable to new-old in the non-canonical construction relative to the canonical construction. No such interaction was found for the locative experiment. However, unlike in the dative alternation, a post-hoc comparison revealed that the active (canonical) construction is more sensitive to information structure than the passive (non-canonical construction). Thus, in each alternation that we tested, we found a different effect of information structure on acceptability. This suggests that information structure is a critical tool for explaining why speakers may choose one construction over the other, but that there is no clear connection between the canonicity of a construction and its preferred information structure.

**Sample Items** (the manipulated arguments are in boldface)

1. Double Object:  
PO (canonical): The professor sent **the grant** to **the administrator**.  
DO (non-canonical): The professor sent **the administrator** **the grant**.
2. Locative:  
Subject-first (canonical): **The weapon** lay behind **the box**.  
Location-first (non-canonical): Behind **the box** lay **the weapon**.
3. Active-Passive:  
Active (canonical): **The detective** took **the weapon**.  
Passive (non-canonical): **The weapon** was taken by **the detective**.



**Figure 1.** The results of the experiments for each alternation. Hollow circles are item means, and error bars are 95% bootstrapped confidence intervals over item means.

## References

1. Goldberg, A. E. (2019). *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton University Press.
2. Halliday, M. A. (1967). *Notes on transitivity and theme in English: Part 2*. *Journal of linguistics*, 3(2), 199-244.
3. Chafe, W. (1987). *Cognitive constraints on information flow. Coherence and grounding in discourse*, 11, 21-51.
4. Clark, H. H., & Clark, E. V. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace Jovanovich.
5. Gundel, J. K. (1988). *Universals of topic-comment structure*. *Studies in syntactic typology*. John Benjamins.
6. Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). *Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering*. *Language*, 76(1), 28-55.
7. Birner, B. J., & Ward, G. (1996). *A crosslinguistic study of postposing in discourse*. *Language and Speech*, 39(2-3), 113-142.
8. Birner, B. J., & Ward, G. (1998). *Information Status and Noncanonical Word Order in English* (Vol. 40). John Benjamins Publishing.
9. Birner, B. J., & Ward, G. (2009). *Information structure and syntactic structure*. *Language and linguistics compass*, 3(4), 1167-1187.
10. Clifton, C., & Frazier, L. (2004). *Should given information come before new? Yes and no*. *Memory & cognition*, 32(6), 886-895.
11. Brown, M., Savova, V., & Gibson, E. (2012). *Syntax encodes information structure: Evidence from on-line reading comprehension*. *Journal of memory and language*, 66(1), 194-209.

12. Kaiser, E., & Trueswell, J. C. (2004). *The role of discourse context in the processing of a flexible word-order language*. *Cognition*, 94(2), 113-147.
13. Fedorenko, E., & Levy, R. (2007). *Information structure and word order in Russian sentence comprehension*. In *Poster presented at the 20th CUNY conference on human sentence processing*, La Jolla, CA.

# Auxiliary selection in French and Italian: a corpus-driven analysis

Greta Viale, Andrea Briglia & Anne Carlier

University of Verona/Sorbonne University (Paris), Sorbonne nouvelle/UMR LLP (Paris), Sorbonne University/STIH (Paris)

**Keywords:** Auxiliary selection, agency, aspect, French, Italian

In French and Italian, intransitive verbs can select either ‘have’ or ‘be’ as auxiliaries in the perfect tense. The complexity of auxiliary selection (AS) is highlighted by verbs that allow both auxiliaries in the same language (i), and others that differ in auxiliary choice across languages (ii).

- (i) Fr. *Les prix ont monté spectaculairement*  
DEF.PL price.PL have.PRS.3PL rise.PRF.PTC spectacularly  
‘the prices have risen dramatically’  
*Le taux d’inflation est monté à 3 %*  
DEF.SG rate of inflation be.PRS3sg rise.PRF.PTCP to 3%.  
‘the inflation rate has risen to 3%’
- (ii) Fr. *Il a disparu.* It. *È scomparso.*  
PRO.NOM.3SG have.PRS.3SG disappear.PST.PTCP be.PRS.3SG disappear.PST.PTCP  
‘he has disappeared’

Extensive research has been devoted to the syntactic and semantic verb features that influence AS. A major hypothesis is Sorace’s (2000, 2015) “Auxiliary Selection Hierarchy”, proposing a scale with telicity and agentivity as the two extreme poles: telicity aligns with ‘be’ selection, while agentivity favors ‘have’.

The state of the art, however, presents several gaps that this study aims to address: a lack of large-scale corpus studies, insufficient attention to verbs allowing both auxiliaries, and an unclear understanding of why ‘have’ is more common in French.

This study focuses on verbs that allow auxiliary alternation in Italian and French, using both quantitative and qualitative methods to identify the syntactic and semantic parameters driving auxiliary selection (AS). It analyzes approximately 1,000 sentences from Sketch Engine (Jakubíček et al. 2013) — composed of 100-sentence samples for a selection of Italian and French verbs with the highest statistical rates of auxiliary alternation — manually annotated for 16 parameters related to the semantic features of the subject, aspect markers and syntactic construction (see Fig. 1 & 2 below). To evaluate the influence of these parameters on AS, the study employs  $\chi^2$  clustering (Kass 1980) and random forest analysis (Kuhn 2008). These two statistical methods are complimentary because of their output: CHAID progressively creates smaller sub-groups of sentences by maximizing differences between sub-groups and minimizing differences between each one. Hence, it groups similar sentences and selects, among the given parameters, the strongest predictor to split each node (Figure 1). In contrast, RF provides a ranked list of all the splitting parameters, sorted in decreasing order of their importance in predicting AS (Figure 2). Findings are corroborated by a usage-based questionnaire using Likert-scale acceptability judgments, structured as a Latin square (Bross 2019). This multifaceted approach ensures robust and reproducible results.

Key findings include:

- **Refinement of agentivity parameter:** Agentivity alone cannot fully predict ‘have’ selection; we introduce a statistically validated hierarchical agentivity scale incorporating the *human* feature and *internal cause* (Levin & Rappaport Hovav 2005; Mateu 2009).
- **Aspectual semi-auxiliaries:** In Italian, these are partially “transparent” with respect to the AS properties of the infinitive; that is, the choice between ‘be’ and ‘have’ as an auxiliary depends to a certain extent on the AS of the verb in the infinitive form, unlike in French (Viale et al. 2024).
- **Cross-linguistic differences:** Contrary to the view that French simply shifted more verbs to ‘have’ selection, this research highlights fundamental differences in auxiliary selection principles: Italian AS is primarily driven by semantic factors (Sorace 2000, 2015; Viale & al. 2024), while French AS is more construction-dependent.

References

Bross, F. (2019). *Acceptability Ratings in Linguistics: A Practical Guide to Grammaticality Judgments, Data Collection, and Statistical Analysis*. Version 1.02. Mimeo.

Jakubíček, M, A. Kilgarriřf, V. Kovář, P. Rychlý, & V. Suchomel (2013). TheTenTen corpus family. *7<sup>th</sup> International Corpus Linguistics Conference CL*, 125–127.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *App. Statist.* (29(2): 119–127. <https://doi.org/10.2307/2986296>

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28(5): 1–26. <https://doi.org/10.18637/jss.v028.i05>

Levin, B., & M. Rappaport Hovav (2005). *Argument Realization*. Cambridge : CUP/

Mateu, J. (2009). Gradience and auxiliary selection in Old Catalan and Old Spanish. In P. Crisma, & G. Longobardi Eds. *Historical Syntax and Linguistic Theory*, 176–193. Oxford: Oxford University Press.

Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language* 76(4) : 859-890.

Sorace, A. (2015). The cognitive complexity of auxiliary selection: from processing to grammaticality judgements. In M. Rosemeyer, & R. Kailuweit Eds, *Auxiliary Selection Revisited. Gradience and Gradualness*, 23–43. Berlin: De Gruyter. <https://doi.org/10.1515/9783110348866-002>

Viale, G., A. Briglia, M. Mucciardi, & A. Carlier. (2024). To be or to have? That is the question: Auxiliary Selection in Italian. *Isogloss*, 10(3), 1-30. <https://doi.org/10.5565/rev/isogloss.346>

Fig 1. CHAID applied on the annotated Italian sentences

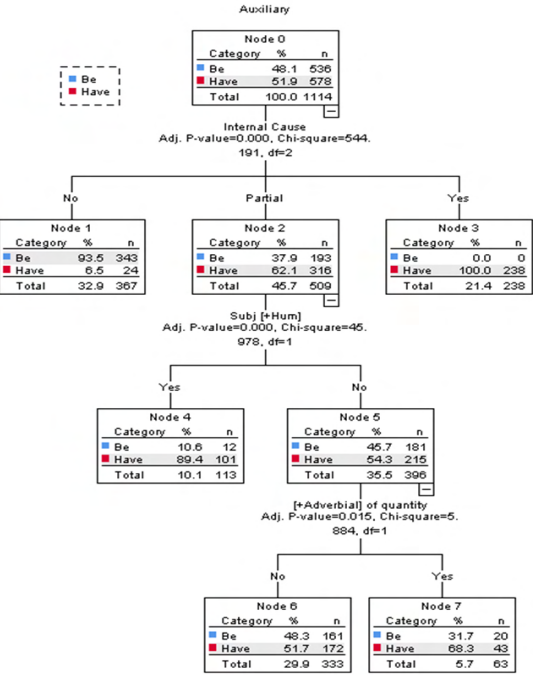
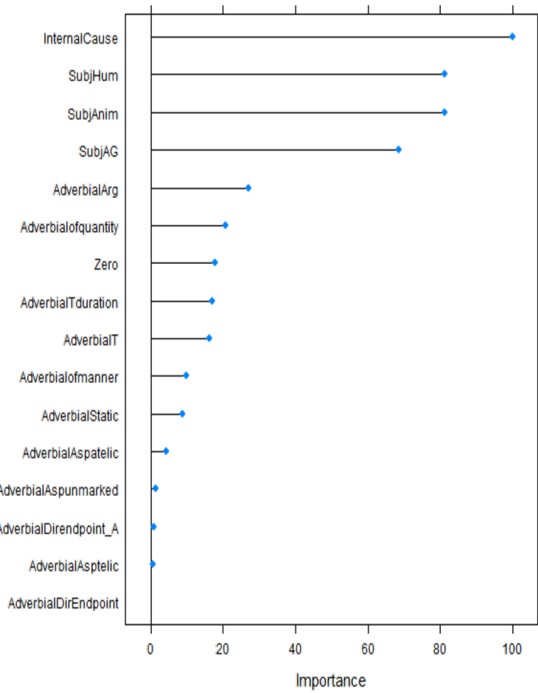


Fig 2. Random Forest applied on the annotated Italian sentences



# Clausal complementation patterns with *remember*, *forget* and *expect*: Combining individual and population level analysis of syntactic variation

Hubert Cuyckens, Nicola Swinburne & Peter Petré  
KU Leuven, University of Antwerp & University of Antwerp

Keywords: syntactic variation, verb complementation, population-level change vs individual change, historical syntax, English

Long-term variation is often viewed on the population level, with the role of individuals considered ‘reduced below the level of linguistic significance’ (Labov 2012: 265). Such views narrow the scope of research by tightly focusing on social identity, minimizing the influence of individual differences. By studying individuals, we may uncover how they accommodate change in their understanding/use of language over their lifetime, and how individual differences in cognitive processing impact the spread of variants (Petré and Anthonissen 2020, Blas Arroyo 2023).

This paper provides a study of 20 renowned male white authors across three 100-year periods of late Modern English (from c. 1650, 1700, 1800, respectively). It investigates changes in their use of competing variants of finite vs. nonfinite complement clauses (CCs) with the complement-taking predicates (CTPs) *remember*, *forget* and *expect*. In this variation, the finite CCs (*that*-clauses and zero-complementation clauses; the older forms) compete with the nonfinite CCs (the *to*-infinitive, the newer form, although first attested late 14th century; see Los 2005: 254-255) and the *ing*-form (emergent). A preference for a certain form could be due to processing considerations, a (personal) determination of a semantic/pragmatic difference, or even a (conscious) social association (as suggested by Cheshire, Kerswill and Williams 2005; and contrary to Labov 2001: 28). While the nature of their competition appears to change, all patterns continue to coexist.

Data consist of >500,000 words per individual, annotated for CCs with *remember*, *forget*, and *expect*, and coded on seven variables. Multifactorial classification models (cf. Fonteyn and Nini 2020) are employed to determine which language-internal factors condition individuals’ variation, and how these constraints vary between individuals, between periods, and during the lifespan of an individual. Preliminary results from >4,000 instances show individuals tend to organise their own behaviour using partly idiosyncratic systems; we argue this accounts for the persistence of long-term variation. A drop in the degree of idiosyncrasy between the earlier and later periods shows potential standardisation at play. In addition, lifespan analysis reveals (almost) all authors increase their use of *that*-clauses (previously regarded as an older form) across their lifespan. We propose this tendency relates to all of them being professional writers who are increasingly placing more emphasis on the clarity of their prose.

Generally, the results suggest that individuals may have different cognitive representations of certain constructions, and that population-level change is simply a change in the average semantic-pragmatic or social understanding of the different forms, and/or the relative processing advantages.

## References

- Blas Arroyo, José Luis (2023), Individual variation and frequency change in Early Modern Spanish: Alignment and intra-speaker (in)stability in a corpus of 18th century ego-documents. *Journal of Historical Linguistics*, 1–43, Published online: 17 August, <https://doi.org/10.1075/jhl.22059.bla>



- Cheshire, Jenny, Paul Kerswill, and Ann Williams (2005), Phonology, grammar and discourse in dialect convergence, in P. Auer, F. Hinskens, and P. Kerswill (eds.), (2009), *Dialect Change: Convergence and Divergence of Dialects in Contemporary Societies*. Cambridge: Cambridge University Press, 135–167.
- Fonteyn, Lauren, and Andrea Nini (2020), Individuality in syntactic variation. *Cognitive Linguistics* 31(2), 279–308.
- Labov, William (2001), *Principles of Linguistic Change*, vol. 2, Oxford: Blackwell.
- Labov, William (2012), What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics* 10, 263–293.
- Los, Bettelou (2005), *The Rise of the To-Infinitive*. Oxford: Oxford University Press.
- Petré, Peter, and Lynn Anthonissen (2020), Individuality in complex systems: A constructionist approach, *Cognitive Linguistics* 31(2), 185–212.

# Voice syncretism in the psych alternation

Jessica Schamne & Niklas Wiskandt

(Heinrich-Heine-Universität Düsseldorf)

Keywords: psych verbs, psych alternation, causative alternation, grammatical voice, syncretism

We present a novel typological perspective on the psych alternation (cf. Rott et al. 2024), i.e. the process of deriving intransitive subject experiencer verbs from transitive object experiencer verbs like in (1) and vice versa like in (2), based on patterns of voice syncretism (cf. Bahrt 2021).

(1) Spanish (Rott et al. 2024: 6)

- a. *El concierto alegra a Sofia.*  
the concert make.happy.3SG to Sophie  
'The concert makes Sophie happy.'
- b. *Sofia se alegra (con el/por el/del concierto).*  
Sophie REFL make.happy.3SG (with the/for the/of the concert)  
'Sophie gets happy about the concert.'

(2) Turkish (Rott et al. 2024: 34)

- a. *Adam kurtçuk-lar-dan endişe-len-ir.*  
Man maggot-PL-ABL worry-INCH-PRS  
'The man is worried due to the maggots.'
- b. *Kurtçuk-lar adam-ı endişe-len-dir-ir*  
maggot-PL man-ACC worry-INCH-CAUS-PRS  
'The maggots worry the man.'

Previous research on the psych alternation has addressed a limited range of mostly Indo-European languages, and has almost exclusively dealt with the detransitivizing direction. It has been debated, usually on a language-specific basis, whether the psych alternation is a subtype of the causative alternation (e.g. Alexiadou & Iordăchioaia 2014) or something different (e.g. Rozwadowska & Bondaruk 2019). The goal of this talk is to contribute to this debate from a typological perspective.

Recent typological studies (Rott et al. 2020, 2024) have shown the morphosyntactic diversity of psych alternation marking, but have not clarified the theoretical question of the syntactic classification and its cross-linguistic significance. This talk takes a new typological approach inspired by the study on voice syncretism by Bahrt (2021):

The proposed method targets the grammatical markers used to form psych alternation pairs, and identifies all occurrences of syncretisms with other grammatical markers in the given language. The focus lies on syncretisms with other voice functions. The method is applied to a genealogically and geographically diverse sample of 20 languages, encompassing both transitivity and detransitivizing alternations, to overcome the prevailing mostly euro-centric perspective on the domain. Based on the

patterns we identify in the sample, we show that the verbal marking in the psych alternation regularly syncretizes with causative and anticausative marking.

Causative syncretism can be attested for almost all languages featuring a transitivizing psych alternation in the sample. Some languages like Warrongo (waru1264, Northern Pama-Nyungan) feature complex syncretism, i.e. a psych alternation marker syncretizing with more than one voice, with causative and applicative voice. When there is complex syncretism with two voices, the psych alternation syncretizes with causative and applicative voice. The only instance for a complex syncretism with three voices is also found in Warrongo, but only marginally documented. It features a syncretism involving causative, applicative and anticausative voice functions.

In the detransitivizing psych alternation type, all markers syncretize with the anticausative. Complex syncretism patterns (e.g. also including the passive and reflexive) are significantly more frequent than in the transitivizing type.

The results provide a typological substantiation to the assumption of a close relation between the psych alternation and the causative alternation. This outcome allows for two possible interpretations that the talk discusses: Either psych transitivization and detransitivization are independent voice functions, but there is a cross-linguistic tendency for them to syncretize with causative and anticausative marking; or the psych alternation is simply a subtype of the causative alternation.

## Glossary

3	third person	CAUS	causative	PRS	present
ABL	ablative	INCH	inchoative	REFL	reflexive
ACC	accusative	PL	plural	SG	singular

## References

- Alexiadou, Artemis and Gianina Iordăchioaia (2014), The psych causative alternation, *Lingua* 148. 53–79. doi:10.1016/j.lingua.2014.05.010.
- Bahrt, Nicklas N. (2021), *Voice Syncretism* (Research on Comparative Grammar 1), Berlin: Language Science Press. doi:10.5281/zenodo.5948872.
- Rott, Julian A., Elisabeth Verhoeven and Paola Fritz-Huechante (2020), Valence orientation and psych properties: Toward a typology of the psych alternation, *Open Linguistics* 6(1), 401–423. doi:10.1515/opli-2020-0020.
- Rott, Julian Andrej, Elisabeth Verhoeven and Paola Fritz-Huechante (2024), Directionality in the psych alternation: a quantitative cross-linguistic study, *Linguistic Typology* 28(1), 147–191. doi:10.1515/lingty-2021-0060.
- Rozwadowska, Bożena and Anna Bondaruk (2019), Against the Psych Causative Alternation in Polish, *Studies in Polish Linguistics Special* (1), 77–97. doi:10.4467/23005920SPL.19.007.10987.

## Argument inversion with psychological predicates in Spanish

José M. García-Miguel & Victoria Vázquez-Rozas  
(Universidade de Vigo & Universidade de Santiago de Compostela)

Keywords: psychological predicates, syntactic alternations, argument inversion, conversive diathesis, collostructional analysis

In Spanish, like in other languages, clauses with psychological predicates deviate from canonical transitivity. Neither of their participants, the Experiencer and the Stimulus, can be classified as a prototypical Agent-subject or a prototypical Patient-object, as both have properties of both subject and object. This ambivalent nature of the participants in experiential events provides a rationale for the existence of alternating structures with psychological predicates and for the non-canonical marking of their arguments in many languages (Haspelmath 2001; Rott et al. 2024).

This paper presents a usage-based study of the diathesis alternation illustrated in (1)-(2)

- |     |                            |             |                    |           |                 |
|-----|----------------------------|-------------|--------------------|-----------|-----------------|
| (1) | (A ella)                   | le/la       | alegr-ó            |           | ver=me          |
|     | (to her)                   | 3SG.DAT/ACC | make.happy-PST.3SG |           | see.INF=1SG.OBJ |
|     | Exp Obj                    | Exp         | V                  |           | Stim Subj       |
|     | 'It pleased her to see me' |             |                    |           |                 |
|     | 'She was happy to see me'  |             |                    |           |                 |
|     |                            |             |                    |           |                 |
| (2) | (Ella)                     | se          | alegr-ó            | al/de     | ver=me          |
|     | (she)                      | REFL.3      | make.happy-PST.3SG | at.the/of | see.INF=1SG.OBJ |
|     | Exp Subj                   | Vmiddle     |                    | Stim Obl  |                 |
|     | 'She was happy to see me'  |             |                    |           |                 |

(1) and (2) convey the same referential meaning and have the same information structure, with the Experiencer as the topic and the Stimulus as the focus. However, there are changes in the syntactic coding of the participants and the predicate: the argument selected as the subject in (1) appears as an oblique adjunct in (2), while the object in (1) is cast as the subject in (2); and the verb undergoes a change in voice from active in (1) to middle in (2). This diathetic relationship can be seen as an instance of inversion or conversive diathesis (Kulikov 2011; De Benito 2022), motivated by the fact that the semantic distance between the two main arguments is smaller than in canonical transitive clauses.

The objective of this study is to identify the factors that influence the choice between the alternates illustrated by (1)-(2). To narrow down the scope of the analysis, the investigation is centered on those constructions in which the Stimulus is expressed by a complement clause (finite or infinitive), either subject, as in (1), or adjunct, as in (2).

We present a multifactorial analysis of corpus data extracted from CORPES XXI. Our dataset comprises 10010 concordance lines, obtained by searching patterns such as **ProClit + Vpsych + (al/de) + que/Vinf**, where ProClit stands for 1sg and 3sg pronominal clitics (*me, se, le, lo, la*) and Vpsych stands for a set of 13 selected psychological verb lemmas.

The CORPES data specify external features such as geolectal variation and register variation. The example set has been annotated for linguistic features such as Experiencer person, TAM of Vpsych and of V2-stimulus, Preposition, V2 lemma.

The statistical analysis shows the significance of linguistic factors (person of the Experiencer, Aspect and the finiteness of the Stimulus) in the selection of the active voice vs. middle voice as the dependent variable, and the comparatively lower relevance of geolectal and register factors.

Furthermore, a collostructional analysis (Stefanowitsch & Gries 2003) was conducted to observe distinctive collexemes in V2 for each constructional variant (separating Active-Middle, Prep and V2 finite vs. infinitive). The results provide insights into the broader tendencies underlying the subtle aspects of meaning and usage that differentiate the alternants. However, each verb displays idiosyncratic characteristics. In addition to the relevance of individual psychological verbs as predictors of syntactic variation, the observation of co-varying collexemes Vpsych - V2 also shows lexical preferences related to the semantic properties of the individual verbs and the constructions.

In sum, the view presented here adheres to the Construction Grammar tenet that there is always some semantic distinction between alternating constructions, and that meaning differences are reflected in the relative frequency of lexical and grammatical features of each construction.

## References

- De Benito Moreno, Carlota (2022), *The Middle Voice and connected constructions in Ibero-Romance: A variationist and dialectal account*. Amsterdam / Philadelphia: John Benjamins.
- Haspelmath, Martin (2001), Non-canonical marking of core arguments in European languages, in A. Y. Aikhenvald, R. M. W. Dixon, and M. Onishi (eds.), *Non-canonical marking of subjects and objects*. Amsterdam / Philadelphia: John Benjamins, 53–83.
- Kulikov, Leonid. 2011. Voice Typology, in Jae Jung Song (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press.
- Rott, Julian Andrej, Elisabeth Verhoeven & Paola Fritz-Huechante (2024), Directionality in the psych alternation: A quantitative cross-linguistic study, *Linguistic Typology* 28(1), 147–191.
- Stefanowitsch, Anatol & Stefan Th. Gries (2003), Collostructions: Investigating the interaction of words and constructions, *International Journal of Corpus Linguistics* 8(2), 209–243.

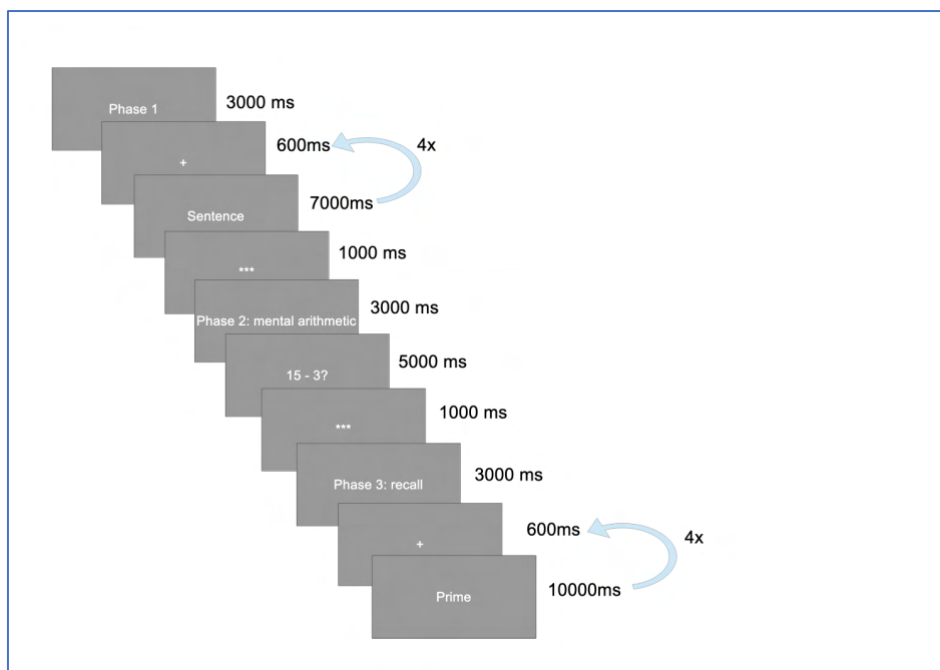
# Syntactic alternation of French psych-verbs: Experiencer first!

Juliette Thuilier, Pegah Faghiri & Léa Pineau  
(Université Toulouse Jean Jaurès/CLLE/IUF, LISN & CLLE)

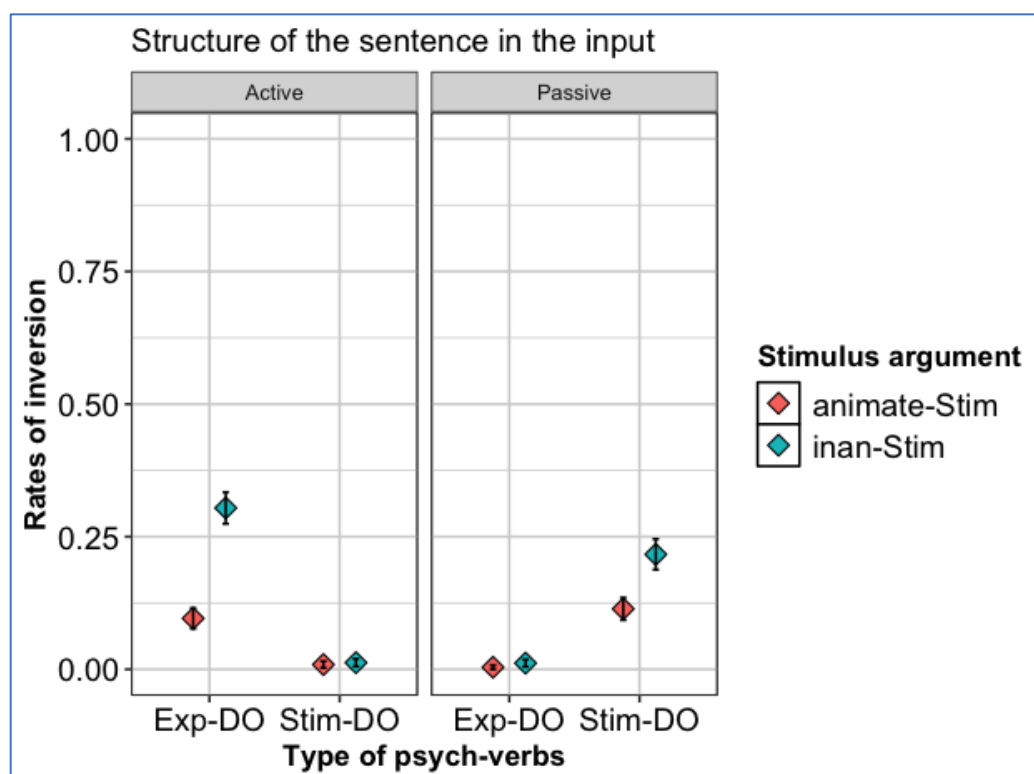
French transitive psych-verbs with an experiencer object (such as *étonner* ‘to surprise’) occur more frequently in the passive voice than the active voice in oral corpora (Blanche-Benveniste 2000, Hamma *et al.* 2017). This observation is in line with Ferreira’s (1994) hypothesis that the experiencer, being more prominent than the stimulus, tends to align with subject function. Using sentence recall (Tanaka *et al.* 2011), we collected experimental data to test Ferreira’s hypothesis in French, and to evaluate a competing hypothesis: given that the experiencer is always animate, using the passive voice for experiencer-object psych-verbs reflects the tendency to align subjects with animate referents (da Cunha & Abeillé 2020 2022, Thuilier *et al.* 2021). We manipulated two intra-item variables –stimulus animacy and voice – and one inter-item variable – semantic role of the direct object. The experiment included 12 items with exp-DO verbs (1) and 12 items with stim-DO verbs (2). Other variables (length, definiteness) were controlled to favor the passive structure, considered to be the rare alternate.

<b>(1) EXP-DO</b>
<b>Pendant la nuit</b> , un tir de carabine a épouvané le villageois. [active/inan-stim] un voleur de tracteur a épouvané le villageois. [active/anim-stim] le villageois a été épouvané par un tir de carabine. [passive/inan-stim] le villageois a été épouvané par un voleur de tracteur. [passive/anim-stim]
<i>During the night, {a bullet/a thief} scared the farmer.</i>
<b>(2) STIM-DO</b>
<b>Tout au long du procès</b> , un témoin de la défense a apprécié l'atmosphère. [active/inan-stim] un témoin de la défense a apprécié la magistrate. [active/anim-stim] l'atmosphère a été appréciée par un témoin de la défense. [passive/inan-stim] la magistrate a été appréciée par un témoin de la défense. [passive/anim-stim]
<i>All along the trial, a defense witness appreciated {the atmosphere/the magistrate}.</i>

The experiment was organized into 24 blocks of 4 sentences, each containing one experimental item. For each block, participants were exposed to 4 sentences successively with the instruction to memorize them. After a distraction task, the primes (in bold in the examples) from the 4 sentences were presented sequentially, and participants were asked to recall aloud the target sentence from the prime (see Fig. 1). The responses of the 141 participants were recorded, transcribed and sorted according to whether they were correctly recalled and whether they showed voice inversion. Inversion rates are presented in Fig. 2: participants tend to change the voice to align the experiencer with the subject function. When there is an asymmetry of animacy (animate experiencer / inanimate stimulus), the probability of having the experiencer as subject significantly increases. The two competing hypotheses are, in fact, complementary: the effect of animacy is added to the effect of argument prominence. The statistical analysis (mixed-effects logistic regression with participants and items as random effects) confirms the observations: an animate stimulus disfavors inversion (coef. -0.65,  $p < .0001$ ) and the interaction between the active structure and the Exp-DO verbs strongly favors inversion (coef. 1.77,  $p < .0001$ ). Our results highlight the value of experimental approaches for the study of syntactic alternations and call for cross-linguistic comparisons of the argument structure effects of psych-verbs in voice alternations.



**Fig. 1: Course of the experiment**



**Fig. 2: Inversion rates (active to passive on the left panel; passive to active on the right panel)**

Keywords: Language variation, Methodology of linguistic research, Romance languages, Syntax

This research is supported by the *Institut Universitaire de France* (IUF) and by the PULCO research project, funded by the French National Research Agency (ANR-22-CE28-0017-01).

## References

- Blanche-Benveniste, C. (2000). Analyse de deux types de passifs dans les productions de français parlé. L. Schøsler (éd.). *Le passif, Études romanes*, 45, 303-319.
- da Cunha, Y. & Abeillé, A. (2020). L'alternance actif/passif en français : une étude statistique sur corpus écrit. *Discours : revue de linguistique, psycholinguistique et informatique*, (27), 25.
- da Cunha, Y. & Abeillé, A. (2022). L'alternance actif/passif en français parlé : un modèle statistique. *Langue française*, 216, 63-80.
- Ferreira, F. (1994) Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language* 33(6). 715–736.
- Hamma, B. & Tardif, A. & Badin, F. (2017). Le passif à l'oral. P. Larrivée & F. Lefevre, (éds). *Français contemporain vernaculaire* (FRACOV).
- Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65:318–330.
- Thuilier, J., Grant, M., Crabbé, B. & Abeillé, A., (2021) Word order in French: the role of animacy, *Glossa: a journal of general linguistics* 6(1): 55.



# Development and validation of a targeted elicitation method for combinations of voice markers

Niklas Wiskandt  
(Heinrich-Heine-Universität Düsseldorf)

Keywords: grammatical voice, valency operations, voice stacking, elicitation, alternations

This talk proposes a new elicitation method in the domain of grammatical voice alternations, specifically targeting combinations of voice markers.

Within the generally well-studied area of voice alternations, combinations of voice markers, also known as voice marker stacking, are an underdescribed behavioral subdomain (for recent studies see e.g. Alsina 2023, and Creissels 2024: 349–355), even though they occur in many languages and have the potential for significant insights into the nature of voice markers and voice alternations. Voice marker stacking appears as the concatenation of morphological voice markers in agglutinating languages like Turkish (1), as nesting of periphrastic voice constructions in more analytic languages such as Portuguese (2), or as a combination of both.

- (1) *Baba-ları kardeş-ler-i öp-üş-tür-dü.*  
Father-3PL.POSS sibling-PL-ACC kiss-RECP-CAUS-PST  
'Their father made the siblings kiss each other.'

(Turkish < Turkic; nucl1301)

- (2) *O pai fez o-s irmão-s se beija-r-em.*  
DEF father make.PST.3SG DEF-PL brother-PL REFL kiss-INF-3PL  
'The father made the siblings kiss each other.'

(Brazilian Portuguese < Romance < Indo-European; braz1246)

The limited previous studies describe only few languages and focus on affix order. Empirical and cross-linguistic studies on the phenomenon face the problem that adequate amounts of data on voice marker stacking are difficult to find, e.g. in typical corpus data. In particular, these data are usually not sufficient to analyze less frequent stacking combinations and the alternation relations between different stacking options. We address this problem by developing an elicitation questionnaire for cross-linguistic use. The questionnaire includes elements that target seven voice functions, their possible combinations, and the order of those combinations (e.g., whether only RECP-CAUS is possible, only CAUS-RECP, or both). In each element of the questionnaire, the consultant is presented with a brief scenario, followed by several questions that target different voice constructions. For each question, the consultant is asked to give a description of one sentence in their language. For the example scenario (3) below, question (3a) produces a simple transitive construction, (3b) elicits reciprocal marking, and (3c) targets a causative stacked onto the reciprocal like in (1) and (2).

- (3) *You are in a playground. You see two kids, brother and sister, who are playing there, accompanied by their father. The brother gives his sister a kiss. The sister kisses her brother back. Later, the kids tell you that they kissed because their father wanted them to do it.*
- How can you describe what the brother does?
  - How can you describe what the two kids do?
  - How can you describe what the father did with the kids?

Subsequently, we validate the questionnaire by applying it to two languages as test candidates, one with agglutinating voice marking, and one with analytic voice marking. The complex syntactic alternation patterns in the data elicited from native speakers of these test candidates confirm the cross-linguistic applicability of the method, but the need for slight adaptations, depending on the type of voice marking that the target language features, and on available alternative strategies to express the target meaning. Beyond artifacts of the method, restrictions operating on alternations do not depend on the morphosyntactic type of the marker, which motivates a hypothesis for a subsequent typological study. Furthermore, not all grammatical combinations are actually speakers' preferred constructions in elicitation responses, e.g. while speakers can produce combinations of three voice markers, they frequently revert to biclausal constructions instead. But even though the realization of alternative strategies instead of voice stacking is a challenge for the elicitation method, it is still successful in recording details of the voice stacking inventory of a language that are hard to detect in corpus data, and capturing alternation relations between strategies used to describe equivalent scenarios.

## Glossary

3	third person	INF	infinitive	RECP	reciprocal
ACC	accusative	PL	plural	REFL	reflexive
CAUS	causative	POSS	possessive	SG	singular
DEF	definite	PST	past		

## References

- Alsina, Alex (2023) Order of Valency-Changing Morphemes, in P. Ackema, S. Bendjaballah, E. Bonet, and A. Fábregas (eds.), *The Wiley Blackwell Companion to Morphology*, Hoboken: Wiley, 1–35. doi:10.1002/9781119693604.morphcom057.
- Creissels, Denis (2024), *Transitivity, Valency, and Voice*, Oxford: Oxford University Press. doi:10.1093/9780198899594.001.0001.

# The German Object Order from the Information Theoretic Perspective

Sophia Voigtmann  
(University of Kassel)

Keywords: Historical linguistics, Information structure, Language variation, Syntax

The relative order of dative (Dat) and accusative (Acc) objects in German is variable. Dat>Acc (“dative before accusative object”; 1a) is the canonical order, but Acc>Dat (“accusative before dative object”, 1b) can be found when the Acc is e.g. given (Lenerz 1977; Rauth 2020; Speyer 2011; 2016).

- 1) a) Ich gebe [einem Athleten]<sub>Dat</sub> [den Ball]<sub>Acc</sub>.  
I give an athlete the ball.  
‘I give an athlete the ball.’  
b) Ich gebe [den Ball]<sub>Acc</sub> einem Athleten]<sub>Dat</sub>.  
I give the ball an athlete.  
‘I give the ball to an athlete.’

We propose two previously unconsidered factors to influence the object order: the position of the full verb (FVP) and the clauses’ information profile, testing two hypotheses using logistic regression (R Core Team 2023):

(H1) Acc>Dat is more likely when the lexical verb precedes the objects.

(H2) Acc>Dat is less likely when the clause’s lexical information profile is uneven.

When the lexical verb containing the valency information follows the objects (FV-VL), Dat>Acc is preferable because recipients discard sentence continuations with a transitive verb earlier (Levy 2008). If the full verb is presented first (FV-V2), a certain object order is less crucial as the objects’ necessity is already known.

(H2) refers to the Uniform Information Density (UID) (Levy & Jaeger, 2007): In lexically uneven clauses, the more common Dat>Acc is preferred as familiarity with a certain construction (e.g. Futrell et al. 2020) facilitates processing even in bumpy information profiles.

We test the stability of these assumptions over time by conducting a corpus study using the Anselm (~16<sup>th</sup> century), RIDGES (16<sup>th</sup>/17<sup>th</sup> century), GerManC (17<sup>th</sup>/18<sup>th</sup> century), the Tiger and TüBa-D/Z corpus (modern German). The objects were annotated automatically using among others morphologic or dependency information given in the corpora. We analyze 1733 clauses here (76% modern data). Acc>Dat occurs overall in <16%.

Each clause was annotated for the FVP, DORM (Cuskley et al. 2021), a measure for UID based on unigram-lemma-surprisal of each word in each corpus – because it neutralizes any grammatical information and is preferable for DORM –, the object’s length difference, their givenness status following the RefLex-Scheme by Riester and Baumann (2017) and the publication century of each text. In the historical data, Acc>Dat occurs in lexically uneven clauses only under the FV2 condition, otherwise, Dat>Acc is more likely in less smooth clauses, showing an effect of the FVP (Figure 1a). Nowadays, Dat>Acc slightly increases with uneven clauses independent of the FVP: Lexically harder-to-process sentences have the more common order and vice versa. As more variation was possible in the past, speakers were more sensitive to means of facilitating processing like the FVP then.

Furthermore, given accusative objects are linked to the Acc>Dat-order. Still, the likelihood of Dat>Acc increases when the full verb is in the right sentence bracket even for given accusative objects (Figure

1b). Thus, we find further (and period-stable) evidence for the influence of the position of the full verb in line with our prediction.

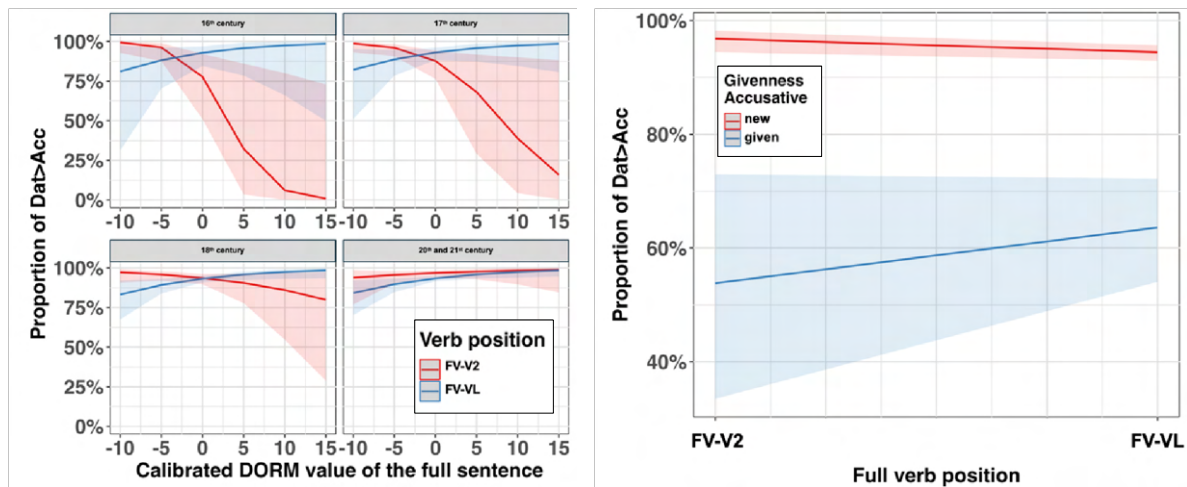


Figure 1a (left) Interaction plot (Lüdecke 2024) of DORM, FVP and period. Figure 1b (right) interaction of FVP with givenness of accusative object.

	Est.	Std. Error	z-value	p-value		Est.	Std. Error	z-value	p-value
Intercept	1.26	0.25	4.95	<0.001 ***	DORM: Period	0.07	0.04	1.73	0.08 .
DORM	-0.06	0.07	-0.86	0.39	Length ratio: FVP	0.18	0.09	2.03	<0.05 *
Length ratio	-0.83	0.04	-1.88	0.06 .	AccInfo-Status: FVP	1.18	0.55	2.15	<0.05 *
AccInfo-Status	-2.11	0.44	-4.78	<0.001 ***	AccInfo-Status: Period	-0.47	0.25	-1.87	0.06 .
DatInfo-Status	0.08	0.41	0.19	0.84	DatInfo-Status: Period	0.65	0.24	2.66	<0.01 **
FVP	0.96	0.49	1.98	0.047 *	FVP: Period	-0.63	0.29	-2.19	<0.05 *
Period	0.38	0.16	2.42	<0.05 *	DORM: FVP	0.35	0.13	2.67	<0.01 **
DORM: Length ratio	0.02	0.01	2.09	<0.05 *	DORM: FVP: Period	-0.16	0.08	-2.05	<0.05 *

## References

- Cuskley, C., Bailes, R., & Wallenberg, J. (2021). Noise resistance in communication: Quantifying uniformity and optimality. *Cognition*, 214, 104754. <https://doi.org/10.1016/j.cognition.2021.104754>
- Futrell, R., Gibson, E., & Levy, R. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, 44(3), e12814. <https://doi.org/10.1111/cogs.12814>
- Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Narr.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems 19*:

- Proceedings of the 2006 Conference* (pp. 849–856). The MIT Press.  
<https://doi.org/10.7551/mitpress/7503.003.0111>
- Lüdecke, D. (2024). *sjPlot: Data Visualization for Statistics in Social Science* [Computer software].  
<https://CRAN.Rproject.org/package=sjPlot>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rauth, P. (2020). *Ditransitive Konstruktionen im Deutschen. Geschichte und Steuerung der Objektfolge im Mittelfeld*. Stauffenburg.
- Riester, Arndt, und Stefan Baumann. 2017. The RefLex Scheme – Annotation Guidelines. *SinSpec* 14.
- Speyer, A. (2011). Die Freiheit der Mittelfeldabfolge im Deutschen – ein modernes Phänomen. *Beiträge Zur Geschichte Der Deutschen Sprache Und Literatur*, 133, 14–31.
- Speyer, A. (2016). The relative object order in High and Low German. In S. Featherston & Y. Vearsley (Eds.), *Quantitative Approaches to Grammar and Grammatical Change. Perspectives from Germanic*. (Vol. 290, pp. 143– 164). De Gruyter.

# Word Order Variation in Three-place Predicates from a Cross-domain Perspective

Xinyi Gao & Milad Shariatmadari

(École des Hautes Études en Sciences Sociales and Laboratoire CRLAO; Université Sorbonne Nouvelle and Laboratoire LATTICE)

Keywords: animacy, definiteness, relative length, Kurmanji Kurdish, Mandarin Chinese

The relative order of direct and indirect objects has been the subject of many usage-based studies cross-linguistically (e.g., Wasow 2002; Bresnan et al. 2007; Bresnan and Ford 2010; Stallings and MacDonald 2011; Faghiri 2016; Faghiri et al. 2018; Faghiri and Samvelian 2020; Zhang and Xu 2023). However, these studies focused solely on either the preverbal or postverbal domain, as most languages position arguments on one side of the verb depending on their type (OV or VO).

According to WALS data (Dryer and Gensler 2013), of the 213 studied VO languages, only three varieties of Chinese allow arguments to be placed on both sides of the verb, including the XVO order. A number of OV languages, including Kurmanji, use a mirror-image placement of arguments on both sides of the verb (OVX) (Haig 2022; Shariatmadari et al. to appear). These fairly rare word orders offer an opportunity to study the effect of previously investigated determinants of word order using data that involves cross-verb domains (pre and postverb).

In the present study, taking a quantitative approach and using the data of The Lancaster Corpus of Mandarin Chinese (McEnery & Xiao 2004) and the Pewan Corpus for Kurmanji Kurdish (Esmaili & Salavati 2013; Esmaili et al. 2013), we investigated the effect of several cognitive-functional factors, including definiteness (definite, indefinite, generic), animacy (animate, inanimate), relative length (number of words for Kurmanji, number of characters for Mandarin), and semantic role, on the position of the O and the X in relation to the verb (preverbal vs. postverbal).

In our corpus study of Kurmanji, we found statistically significant effects of definiteness, animacy, and relative length (Table 1), corresponding to definite-first (O), animate-first (X), and long-first preferences. Accordingly, the postverbal Xs were generally inanimate (goal) and short. For Mandarin data, we found statistically significant effects for the same factors (Table 2), with definite-first (O) and animate-first (X) preferences. However, in terms of length, we observed a long-first preference in the preverbal domain and an end-weight preference in the postverbal domain (when both O and X are placed after the verb). Consequently, the preverbal Xs were typically animate (recipient) and long.

Predictor	$\chi^2$	df	p
definiteness-O	49.44	2.00	< .001
animacy-X	42.29	1.00	< .001
rel_length	24.71	1.00	< .001

Table 1. The results of the omnibus test for Kurmanji data

Predictor	$\chi^2$	df	p
definiteness-O	21.44	2.00	< .001
animacy-X	10.27	1.00	0.001
rel_length	5.77	1.00	0.016

Table2. The results of the omnibus test for Chinese data

By comparing our findings to those of previous studies, we cautiously argue that the definite-first and animate-first preferences are consistent in these two languages regardless of the preverbal or postverbal domains. However, in terms of length, there are two possible scenarios: a language like Kurmanji has a consistent long-first preference across domains, whereas a language like Mandarin has a long-first preference in the preverbal domain and an end-weight preference in the postverbal domain.

## References

- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In G. Bouma, I. Kraemer & J. Zwarts (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.
- Dryer, Matthew S. & Orin D. Gensler. 2013. Order of Object, Oblique, and Verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/84> (accessed 10 January 2024).
- Esmaili, Kyumar S., Donya Eliassi, Shahin Salavati, Purya Aliabadi, Asrin Mohammadi, Somayeh Yosefi & Shownem Hakimi. 2013. Building a test collection for Sorani Kurdish. In Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications (AICCSA). IEEE. 1–7.
- Esmaili, Kyumars S., & Shahin Salavati. 2013. Sorani Kurdish versus Kurmanji Kurdish: an empirical comparison. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 300–305.
- Faghiri, Pegah. 2016. *La variation de l'ordre des constituants dans le domaine préverbal en persan: approche empirique. Linguistique*. Université Sorbonne Paris Cité; Université Sorbonne Nouvelle Paris 3.
- Faghiri, Pegah & Pollet Samvelian. 2020. Word order preferences and the effect of phrasal length in SOV languages: evidence from sentence production in Persian. *Glossa: a journal of general linguistics*. <https://www.glossa-journal.org/article/id/5334/> (accessed 10 January 2024).
- Faghiri, Pegah, Pollet Samvelian & Barbara Hemforth. 2018. “Is there a canonical order in Persian ditransitive constructions? Corpus based and experimental studies”. In Agnes Korn & Andrej Malchukov (eds.), *Ditransitive constructions in a cross-linguistic perspective*, 165–185. Wiesbaden: Reichert Verlag.

Haig, Geoffrey. 2022. Post-predicate constituents in Kurdish. In Yaron Matras, Ergin Öpengin & Geoffrey Haig (eds.), *Structural and Typological Variation in the Dialects of Kurdish*, 335–378. London: Palgrave MacMillan.

McEnery, Anthony & Zhonghua Xiao. 2004. The Lancaster Corpus of Mandarin Chinese: A Corpus for Monolingual and Contrastive Language Study. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Stallings, Lynne M. & Maryellen C. MacDonald. 2011. It's not just the "heavy NP": relative phrase length modulates the production of heavy-NP shift. *Journal of psycholinguistic research* 40(3). 177–187.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.

Zhang, Dong & Jiajin Xu. 2023. Dative alternation in Chinese: A mixed-effects logistic regression analysis. *International Journal of Corpus Linguistics*, 28(4), pp.559-585.



## **WS6 From explicit to implicit: constraints and explanations**

**Liesbeth Degand & Hongling Xiao**

# The explicitness and implicitness of causal relations in Portuguese

Amália Mendes & Pierre Lejeune

(University of Lisbon - School of Arts and Humanities and Centre of Linguistics)

Keywords: discourse relations; explicitness; implicitness; causal relations; Portuguese

Discourse relations (DRs) ensure coherence between discourse units and may be explicitly marked with connectives or left implicit and inferred through linguistic context. Corpus-based studies (Asr & Demberg 2012; Zufferey & Gygas 2016; Hoek et al. 2017) have explored whether the explicit or implicit rendering of DRs may be explained by the continuity (Murray 1997), causality-by-default (Sanders 2005) or complexity (Hoek et al. 2017) hypotheses. A multilingual study of the TED-MDB corpus reveals that some of these factors do not hold cross-linguistically (Mendes et al. 2023). Other approaches have analysed the role played by non-connective cues in the inference of a sense in implicit discourse relations: a corpus-based study showed that only a small set of discourse relations are exclusively signalled by connectives (Das & Taboada, 2018), and an experimental study highlighted the importance of contextual cues, such as antonyms in implicit contrast relations (Crible & Demberg 2020).

We propose to explore some of these hypotheses by focusing on Causal relations (Reason and Result) in Portuguese by exploring a manually annotated Portuguese corpus, in order to answer the following questions:

- Are Causal relations more frequently implicit than explicit, and are they the most frequently implicit relation, as expected from the hypotheses suggested in the literature?
- Does linear temporal order of the arguments (as in Result relations) affect the implicit realization of the DRs, compared to the non-linear temporal order (Reason)?
- Is subjective (epistemic) cause a factor that favours explicitness, compared with objective cause? And are causal relations with a speech act value more frequently implicit?
- Are DRs involving different Attribution sources more frequently explicit?
- What contextual cues play a role in signalling a Causal interpretation in implicit relations?

We take the CRPC-DB, available online at the PORTULAN CLARIN platform ([portulanclarin.net](http://portulanclarin.net)), as the source of data for this analysis (Mendes & Lejeune 2022). The CRPC-DB contains originally-written Portuguese texts (newspaper, fiction and technical texts) annotated for DRs (14,224 DRs) following the Penn Discourse Treebank annotation scheme (Prasad et al. 2008), i.e., it includes relations of the type explicit, implicit, alternative lexicalizations, entity relations (and also No Relation). To analyse the data extracted from the corpus, we use a quantitative (descriptive statistics, chi-squared test) and a qualitative approach.

Preliminary results indicate that (i) causal relations are slightly more frequently explicit than implicit, contrary to what the continuity and cognitive complexity hypotheses would predict; (ii) temporal linear order has an impact, as predicted by continuity (the forward relation Result is more frequently implicit, while the backward Reason is more frequently explicit); (iii) there seems to be no effect of subjective vs. objective cause. We will analyse the contexts to explore relevant contextual patterns related to the explicit/implicit relation type and compare our results with the findings in the literature for other languages and with the results presented for Portuguese on a corpus of translated transcripts of TED Talks (Mendes et al. 2023).

## Acknowledgments

This work was partially supported by Fundação para a Ciência e a Tecnologia as part of the project of Centro de Linguística da Universidade de Lisboa (UID/00214).

## References

- Asr, Fatemeh Torabi, and Vera Demberg (2012), Implicitness of discourse relations, in *Proceedings of COLING 2012*, 2669-2684.
- Crible, Ludivine, and Vera Demberg (2020), The role of non-connective discourse cues and their interaction with connectives, *Pragmatics & Cognition* 27(2), 313–338.
- Das, Debopam, and Maite Taboada (2018), Signalling of Coherence Relations in Discourse, Beyond Discourse Markers, *Discourse Processes* 55(8), 743–770.
- Hoek, Jet, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted J. M. Sanders (2017), Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study, *Journal of Pragmatics* 121, 113–31.
- Mendes, Amália, Deniz Zeyrek, and Giedre Oleškevičienė (2023), Explicitness and implicitness of discourse relations in a multilingual discourse bank, *Functions of Language* 30(1), 67-91.
- Murray, John D. (1997), Connectives and narrative text: The role of continuity, *Memory and Cognition* 25(2), 227-236.
- Sanders, Ted J. M. (2005), Coherence, causality and cognitive complexity in discourse, in *Proceedings/Actes SEM-05: First International Symposium on the Exploration and Modelling of Meaning*, 105–114.
- Prasad, Rashmi, Nikil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber (2008), The Penn Discourse Treebank 2.0, in *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*, 2961–2968.
- Zufferey, Sandrine, and Pascal Gygax (2016), The role of perspective shifts for processing and translating discourse relations, *Discourse Processes* 53(7), 532-555.

## Entropy & Repetitions

### Towards observing the unfolding of task-based conversations

Anaïs C. Murat & Carl Vogel  
(Trinity College Dublin & Trinity College Dublin)

**Keywords:** task-based dialogue; conversation flow; information structure; repetition; entropy

Conversations are dynamic. Utterances are performed; meaning is negotiated; and, eventually, the interactions end. In an ideal scenario, this end occurs when both parties are satisfied with the content that has been shared, communicative goals have been sufficiently fulfilled, and nothing else is deemed necessary to be added. Our hypothesis is that these conversation phase types and their links may be inferred through indicators of information density, apart from explicit connectives. Hence, we present a method to investigate the turn-by-turn unfolding of task-based dialogues using the HCRC Map-Task Corpus (Thompson et al. 1993) as a representative source.

The method relies on the interpretation of two metrics sampled at each dialogue turn: (1) a measure of **linguistic repetitions** (self and other-repetitions, closed and open-class word categories) from one turn to another, and (2) **entropy** ( $H$ ) (Shannon 1948) as measurements of the lexical spread throughout the dialogue both for the speaker ( $H_S$ ) and the conversation ( $H_C$ ). Repetitions are calculated using a Jaccard index which counts whether the tokens of a turn were types in the preceding one, similarly to Murat et al. (2022). Our method for entropy is adapted from Gnjatović et al. (2018): every turn updates the speaker and the conversation's dictionaries, and  $H_S$  and  $H_C$  are computed using the token distribution in each dictionary. More specifically, to gauge the impact of the turns over the course of the interaction, we look at  $\Delta H$ , the difference in entropy from one turn to another. The ratio  $\Delta H_S / \Delta H_C$  thus provides information on how a given turn benefits the speaker as compared to the conversation. For instance, it can be expected that saying something completely new will increase both  $H_S$  and  $H_C$ , while referring to something that the other participant had already mentioned will likely increase  $H_S$  further than  $H_C$ .

Using statistical data mining in light of deeper qualitative analyses of turns (such as comparisons with speech acts (marked with "\*" in this text), turn index and lexical content), we show that turns that start dialogue units (e.g. "ready" turns\*, initial questions, or instructions\*) are more likely to display an entropy behaviour such that  $\Delta H_S / \Delta H_C \simeq 1$ ; those that negotiate the common ground (e.g. open questions\*, explanations\* or rejections\*) rather fall into  $\Delta H_S / \Delta H_C > 1$ ; and those that close dialogue units (acknowledgements\*, confirmation\*, etc.) lead to a decrease of both  $H_C$  and  $H_S$ . Repetitions allow further distinctions between speech acts, such as those that require more planning (e.g. instructions\* display more self-repetitions) and those that build upon others' productions (e.g. clarifications\* use more other-repetitions). We thus argue that examination of lexical spread in dialogues through repetitions and entropy (indicators associated with information density) supports quantitative and qualitative characterization of interaction flow.

*This publication has emanated from research conducted with the financial support of Taighde Éireann-Research Ireland under Grant. No. 18/CRT/6223.*

## References

- Gnjatović, Milan, Jovica Tasevski, Branislav Borovac & Nemanja Maček (2018), An Entropy-Based Approach to Automatic Detection of Critical Changes in Human-Machine Interaction, in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 000175–000178. doi:10.1109/CogInfoCom.2018.8639869. ISSN: 2380-7350.
- Murat, Anais, Maria Koutsombogera & Carl Vogel (2022), Mutual Gaze and Linguistic Repetition in a Multimodal Corpus, in *Proceedings of the Language Resources and Evaluation Conference*, 2771–2780. Marseille, France: European Language Resources Association (ELRA).
- Shannon, C. E (1948), A Mathematical Theory of Communication, *Bell System Technical Journal* 27(3). 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- Thompson, Henry S., Anne Anderson, Ellen Gurman Bard, Gwyneth Doherty-Sneddon, Alison Newlands & Cathy Sotillo (1993), The HCRC Map Task corpus: natural dialogue for speech recognition, in *Proceedings of the workshop on Human Language Technology - HLT '93*, 25. Princeton, New Jersey: Association for Computational Linguistics. doi:10.3115/1075671.1075677.

# Analyzing the impact of (dis)continuity during the reading of explicit and implicit discourse relations

Isabel Falé, Paula Luegi, Amália Mendes & Jéssica Gomes

(Universidade Aberta and Centre of Linguistics of University of Lisbon; University of Lisbon - School of Arts and Humanities and Centre of Linguistics of University of Lisbon)

Keywords: discourse relations; discourse processing; (dis)continuity; explicitness; implicitness

Discourse relations (DR) may be explicitly or implicitly conveyed, i.e., a discourse marker with a connective function linking two arguments of a DR may be present or absent. Also, the relation between the two arguments can have different senses (e.g., Cause or Concession). The ‘continuity’ hypothesis (Murray 1997) suggests that readers/listeners expect events in discourse to follow a linear temporal sequence, being continuity described through several dimensions, such as time, space, action continuity and perspective (Segal *et al.* 1991; Das & Egg 2023). Corpus-based studies show that discontinuous relations (e.g., Contrast) are conveyed explicitly much more often than continuous relations (e.g., Cause) (Asr & Demberg 2012; Hoek *et al.* 2017; Mendes *et al.* 2023), and experimental studies suggest that the processing of implicit-discontinuous relations is harder than implicit-continuous DRs (Zufferey & Gyax 2016).

Recent results for European Portuguese (EP), contrasting cause-continuous and confirmation-discontinuous DRs, show that reading times are slower when (i) the connective is implicit, (ii) the DR is confirmation-discontinuous, (iii) the connective is implicit in confirmation-discontinuous relations, compared with implicit cause-continuous DRs (Falé *et al.* 2024).

Adding to recent studies in EP, our goal is to test implicit and explicit DRs of three different types: (i) a causative reading of *de facto* ‘indeed’, where the second argument provides a subjective cause for the content of the first argument (continuous relation); (ii) a confirmation reading of *de facto* ‘indeed’, where the first argument conveys the perspective of an external source and the second argument conveys the speaker’s own perspective (discontinuous - perspective shift), and (iii) a contrastive reading of *na verdade* ‘in fact’ (discontinuous - negative polarity). In a self-paced reading experiment in EP, the participants will read sentences similar to the examples presented in (1), (2), and (3), with and without the connective, followed by a comprehension question (4).

## (1) CAUSE

*O Pedro deve estar a desenvolver-se de forma saudável. (De facto,) o rapaz cresceu muito nos últimos meses deste verão.*

‘Pedro must be developing healthily. (CONNECTIVE) the boy has grown a lot in the last few months of this summer.’

## (2) CONFIRMATION

*A Sofia esperava que o Pedro estivesse mais alto. (E de facto,) o rapaz cresceu muito nos últimos meses deste verão.*

‘Sofia expected Pedro to be taller. (CONNECTIVE) the boy has grown a lot in the last few months of this summer.’

(3) CONTRASTIVE

*A Sofia pensava que o Pedro estava da mesma altura. (Na verdade,) o rapaz cresceu muito nos últimos meses deste verão.*

‘Sofia thought Pedro was the same height. (CONNECTIVE) the boy has grown a lot in the last few months of this summer.’

(4) Comprehension question:

*O Pedro engordou no final deste verão?*

‘Did Pedro gain weight at the end of this summer?’

Based on the literature, we put forward the following hypotheses: H1 – implicit DRs are harder to process than explicit DRs; H2 – discontinuous DRs are harder to process than continuous DRs; H3 – implicit-discontinuous DRs are harder to process than implicit-continuous DRs; H4 – contrastive DRs, due to their negative polarity, are harder to process than cause and confirmation DRs.

#### Acknowledgments

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) as part of the project of Centro de Linguística da Universidade de Lisboa (UID/00214) and project number UI/BD/154500/2022.

#### References

- Asr, Fatemeh Torabi, and Vera Demberg (2012), Implicitness of discourse relations, in *Proceedings of COLING 2012*, 2669-2684.
- Das, Debopam, and Marcus Egg (2023), Continuity in discourse relations, *Functions of Language* 30(1), 41–46.
- Falé, Isabel, Paula Luegi, Jéssica Gomes, and Amália Mendes (October, 2024), Relações causais e confirmativas com “de facto”: processamento de relações discursivas explícitas e implícitas em Português Europeu. Talk presented at *ENAPL*, Azores, Portugal.
- Hoek, Jet, Sandrine Zufferey, Jacqueline Evers-Vermeul, and Ted J. M. Sanders (2017), Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study, *Journal of Pragmatics* 121, 113–31.
- Mendes, Amália, Deniz Zeyrek, and Giedre Oleškevičienė (2023), Explicitness and implicitness of discourse relations in a multilingual discourse bank, *Functions of Language* 30(1), 67-91.
- Murray, John D. (1997), Connectives and narrative text: The role of continuity, *Memory and Cognition* 25(2), 227-236.
- Segal, Erwin, Judith Duchan, and Paula Scott (1991), The role of interclausal connectives in narrative structuring, *Discourse Processes* 14(1), 27–54.
- Zufferey, Sandrine, and Pascal Gyga (2016), The role of perspective shifts for processing and translating discourse relations, *Discourse Processes* 53(7), 532–555.

## ***Decoding discourse with punctuation***

Jet Hoek & Merel C.J. Scholman  
(Radboud University & Utrecht University)

Keywords: discourse relations, language processing, corpus, punctuation, psycholinguistics

Prior research has shown that readers tend to understand texts better and integrate information faster when the discourse relations are signalled (e.g., Millis & Just, 1994; Sanders & Noordman, 2000). However, the scope of this work largely remained on lexical elements such as connectives. The current study focuses on a non-lexical cue, namely punctuation markers such as commas, colons, and parentheses. To illustrate the discourse signaling function of punctuation, consider Example (1), taken Nunberg (1990):

(1) [He reported the decision]<sub>S1</sub> {:/;} [we were forbidden to speak with the chairman directly.]<sub>S2</sub>

In this example, S2 can be interpreted as an elaboration of the decision when it contains a colon, whereas it more likely expresses the reason for why “he” reported the decision and not the chairman directly when the sentence contains a semi-colon. Hence, even though punctuation does not carry a lexical meaning, it can play a role in discourse relation interpretation. The mapping is not straightforwardly one-to-one, however: the same punctuation markers may be used to realize different relations.

In this project, we will explore the co-occurrences between punctuation and specific discourse relations in a corpus study using the discourse-annotated PDTB corpus (Webber et al., 2019). Should punctuation serve as a cue for specific discourse relations, the distribution of discourse relations should be different for the various types of punctuation markers. Indeed, preliminary results suggest several interesting correlations between punctuation and discourse relations. For example, semicolons (;) typically signal contrast or parallel relations, dashes (–) can signal specification or equivalence relations, and colons (:) are frequently used to introduce explanations or lists.

These insights will be used in a story continuation study (to be completed before the SLE conference), which investigates to what extent comprehenders are guided by punctuation in their construction of discourse relations. Participants will be given the first segment of a relation and one of a selection of punctuation signals and will be asked to provide the second segment. These continuations will be annotated for continuation. Annotators will be blind to what punctuation signal was presented to the participants. If readers have learned the mappings between the punctuation and relation types from their exposure to these mappings in natural language, then the distribution of relation types per punctuation type should resemble that found in the corpus study.

This line of research will contribute to our understanding of how non-lexical cues, specifically punctuation, function in signaling discourse relations. We aim to uncover patterns that highlight the role of punctuation in text coherence and reader comprehension and offer insights into the cognitive processes involved in reading and text interpretation.

## **References**

- Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128–147.
- Nunberg, G. (1990). *The linguistics of punctuation*. 18. Center for the Study of Language (CSLI).
- Sanders, T. J. M., & Noordman, L. G. M. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*(1), 29, 37–60.
- Webber, B., Prasad, R., Lee, A., & Joshi, A. (2019). *The Penn Discourse Treebank 3.0 annotation manual*. Philadelphia, University of Pennsylvania.



# The Processing of Peripheral Gerundive Clauses: A Psycholinguistic Study

Juliana Gomes, Purificação Silvano, António Leal & Violeta Magalhães

(Faculty of Arts and Humanities of the University of Porto, Centre for Linguistics of the University of Porto; Faculty of Arts of the University of Porto, Centre for Linguistics of the University of Porto, INESC TEC; Faculty of Arts and Humanities of the University of Macau, Centre for Linguistics of the University of Porto; & Faculty of Arts and Humanities of the University of Porto, Centre for Linguistics of the University of Porto)

**Keywords:** Discourse relations; peripheral gerundive clauses; self-paced reading; semantic processing; compound gerund

Discourse relations (DRs) can be implicit, not linguistically signaled, or explicit, marked by words (e.g., "because" for *Explanation*) or by syntactic structures (e.g., relative clauses for *Elaboration*) (Taboada & Das, 2013). However, some constructions pose challenges in the study of DRs, as is the case of peripheral gerundive clauses (GCs) (Lobo, 2003), which are typically not introduced by connectors. Moreover, the syntactic structure of GCs cannot be consistently associated with a specific DR (cf., e.g., Lobo, 2003; Móia & Viotti, 2004; Silvano, Leal & Cordeiro, 2019). Therefore, interpreting GCs, particularly their connection to the main clause, relies on other grammatical elements.

This study aims to investigate the factors determining the inference of two DRs (Explanation and Result) in European Portuguese (EP). To achieve this, we propose a self-paced reading (SPR) coupled with a 5-point *Likert* scale study that measures reading times (RTs) and acceptability rates (ARs) in relation to the roles of clause ordering and lexical cues in the processing of GCs, compared to finite clauses. We designed a study with six conditions, crossing three experimental factors: linear order (finite clause-final position, GC-final position); presence or absence of explicit lexical marking (GC without explicit lexical marking, GC with a causative verb, and finite clause with lexical marker); and DR (Result and Explanation). Linear order was chosen because it influences the classification of DRs, particularly the pair *Result–Explanation*. These two DRs were selected due to their distinguishable lexico-semantic encoding. In addition, causality has been extensively studied for its central role in discourse processing. Thus, we aimed to contribute to this ongoing line of research. The selected conditions allowed us to examine varying degrees of causal markedness—from implicit (GC without lexical marking) to explicit forms (GC with a causative verb, finite clauses with lexical markers). Example (1) with finite and gerundive *Result* clauses in final position illustrates the data.

We remotely tested participants using PCIBex software and analyzed the RTs and ARs (*Linear Regression, Jamovi*) of 62 native undergraduate EP speakers (~22.8 y.o.) of UPorto. Results show that: (i) GCs in the second position are easier to process (lower total RTs,  $F(5, 37205) = 35.3, p < 0.001$ ); (ii) finite sentences with discourse markers (DMs) are less costly (cf. (1a),  $F(5, 37205) = 35.3, p < 0.001$ ) and (iii) GCs with causative verbs are costlier than those without marking (cf. (1b-1c),  $F(5, 37205) = 35.3, p < 0.034$ ). Furthermore, the integration of the DM appears to be local since sentences with DMs show longer integration times in the critical segment (DM), but the final processing (*wrap-up effect*) is shorter.

Additionally, *Result* sentences (finite or gerundive) appear to be less costly locally (cf. critical segment and spill-over) than *Explanation* sentences, except for GCs without lexical marking.

This study represents the initial phase of a comprehensive investigation into the processing of DRs in EP. Future research will expand upon this first study by incorporating additional variables and exploring a broader range of DRs within the analytical framework.

(1) As declarações do presidente criaram um ambiente de mal-estar na empresa,

a. **de maneira que** os funcionários realizaram uma greve de 2 dias (finite clause with lexical marking (*de maneira que*))

b. tendo **levado** os funcionários a realizar uma greve de 2 dias (GC with lexical marking (*levar a*))

c. tendo os funcionários realizado uma greve de 2 dias (GC without lexical marking).

*The president's statements created an atmosphere of unease in the company, so much so that employees went on a 2-day strike.*

### Acknowledgments

National funds have funded this research through FCT – Fundação para a Ciência e a Tecnologia, I.P., within the UIDB/00022/2020 project.

The last author is funded by FCT – Fundação para a Ciência e a Tecnologia, I.P., ref. 2024.02966.BD.

### References

- Lobo, Maria (2003), *Aspectos da sintaxe das orações subordinadas adverbiais*. PhD Dissertation: Universidade Nova de Lisboa.
- Móia, Telmo & Viotti, Evani (2004), Sobre a semântica das orações gerundivas adverbiais, in I. Duarte , and I. Faria (eds), (2004), *Actas do XX Encontro da Associação Portuguesa de Linguística*, 715-729, Lisboa: Associação Portuguesa de Linguística.
- Taboada, Maite & Das, Debopam (2013), Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue Discourse* 4, 249-281.
- Silvano, Purificação; Leal, António & Cordeiro, João. 2019. Algumas reflexões sobre a classificação de orações gerundivas em Português Europeu. *Revista da Associação Portuguesa de Linguística*, 5, 325-347.

# ***Salient signals: The role of quantifiers during discourse processing***

Merel C.J. Scholman & Ted J.M. Sanders  
(Utrecht University)

Keywords: quantifiers, discourse relations, language processing, reading, listening

Various linguistic markers play a role as processing instructions when readers are building a cognitive representation of discourse. Connectives and cue phrases such as *because* and *however* help readers process coherence relations (e.g., Cozijn et al., 2011; Sanders & Noordman, 2000). However, little is known regarding the effect of signals other than connectives on the processing of coherence relations. This work focuses on quantifiers such as “several”, which can function as a signal of an upcoming list relation. Indeed, an offline continuation study has shown that such quantifiers do elicit list expectations (Scholman et al., 2020).

Experiment 1 (n=80, crowdsourced via Prolific) focused on whether the presence of a list signal caused the subsequent list to be read faster during *online* processing. The target items (n=16) were presented in two conditions: one containing a list signal (bolded in Example 1 for illustrative purposes) and one without a list signal (Example 2).

Example 1. [List signal condition:] The singer had written **several** songs during a summer holiday. He wrote a song about his childhood. {He also}<sub>target</sub> {composed one}<sub>spill</sub> about his girlfriend. The songs were released a month ago and have already reached the charts.

Example 2. [Control condition:] The singer went on a songwriting retreat over the summer. He wrote a song about his childhood. {He also}<sub>target</sub> {composed one}<sub>spill</sub> about his girlfriend. The songs were released a month ago and have already reached the charts.

Log-transformed reading times were modeled using a linear mixed-effects regression model with condition as predictor variable. The results showed no evidence of an effect of the list signal on reading times (target:  $\beta = -0.02$ , SE = 0.02,  $t = -0.97$ ,  $p = .34$ ; spill:  $\beta = -0.01$ , SE = 0.02,  $t = -0.51$ ,  $p = .61$ ).

These results suggest that readers are not sensitive to the list signal, which is not in line with the offline story continuation results reported by Scholman et al. (2020). One possible explanation for this discrepancy is that readers can only exploit alternative signals during online processing when these signals are salient enough. To address this, we conducted a follow-up study (Experiment 2), which employed an identical paradigm as Experiment 1, with the exception of the presentation of the quantifier: quantifiers were presented in bold font, thereby making the signal more salient. The results showed an effect of the bolded list signal on reading times on the spillover region (target:  $\beta = -0.01$ , SE = 0.01,  $t = -0.46$ ,  $p = .65$ ; spill:  $\beta = -0.02$ , SE = 0.01,  $t = -2.14$ ,  $p = .03$ ). To further explore the role of signal salience on the extent to which that signal can facilitate processing, we will conduct a self-paced *listening* study (Experiment 3), in which the salience of signals is manipulated in a more natural manner using prosodic stress. Our work will provide more insight into the cognitive processes that come into play during discourse comprehension.

## References

- Cozijn, R., Noordman, L. G., & Vonk, W. (2011). Propositional integration and world-knowledge inference: Processes in understanding because sentences. *Discourse Processes*, 48(7), 475–500. <https://doi.org/10.1080/0163853X.2011.594421>
- Sanders, T. J., & Noordman, L. G. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse processes*, 29(1), 37-60. [https://doi.org/10.1207/S15326950dp2901\\_3](https://doi.org/10.1207/S15326950dp2901_3)
- Scholman, M. C., Demberg, V., & Sanders, T. J. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes*, 57(10), 844–861. <https://doi.org/10.1080/0163853X.2020.1813492>

# From German to Japanese? Corpus-based analysis of implicitness in events across languages

Natalia Levshina, Marco Centola, Alice Gazzinelli, Laura Mello, Maria Morhy, Nivedha Narayanaswamy & Yuan Quin

(Radboud University, Radboud University, Federal University of Minas Gerais, Federal University of Minas Gerais, Federal University of Minas Gerais, Radboud University & Radboud University)

**Keywords:** high- and low-context cultures, implicitness, agent, temporal reference, corpora

## Aims and background

There have been attempts in cross-cultural and cross-linguistic studies to classify cultures and languages with regard to their general preference for explicit or implicit way of communicating information. The most famous probably is the classification of cultures into low-context and high-context ones (Hall 1976). According to this classification, cultures like Swiss Germans, Germans and Scandinavians prefer highly explicit communication with low reliance on context, whereas the Japanese and Chinese leave a lot of information implicit because they strongly rely on context. In linguistics, a related classification has been proposed by Bisang's (2009), who argues that languages differ in their overt and hidden complexity. However, systematic empirical support for such classifications has been largely missing. Our project fills in this gap by encoding different aspects of event semantics in a sample of diverse languages from different countries: American English, Brazilian Portuguese, German (as spoken in Germany), Italian (Italy), Japanese (Japan), Mandarin Chinese (Mainland China), Russian (Russia) and Tamil (India).

## Data and method

We create corpora from online newspapers (30 articles per language/country). Our investigation of implicitness focuses on implicit and explicit expression of different aspects of events: agent and time reference. We have created and tested an annotation schema for identification of events based on the approach in TimeBank (Pustejovsky et al. 2003). Events and their aspects are annotated manually, but we are also training Large Language Models on the annotated data, with the purpose of identification of implicit information in corpora in different languages (cf. Roth and Anthonio 2021). We test if the differences between the languages/cultures are significant with the help of mixed-effect logistic models predicting whether a specific type of information is implicit or explicit based on the language. The individual articles, newspapers and topics are tested as random effects.

## Preliminary results

Our preliminary results of the analysis of agents and time references in eight languages and 1700 events indicate that the existing cultural classifications are supported only partly. With regard to the expression of agents, German displays the highest proportion of overt agents, and Japanese the lowest, supporting the existing classifications. As for time reference, Russian and Brazilian Portuguese display low-context behaviour, often using overt ways of encoding time, which is against the expectations. Therefore, we conclude that reliance on context strongly depends on which type of information language users want to express.

## References

Bisang, Walter (2009), On the evolution of complexity—Sometimes less is more in East and mainland Southeast Asia, in G. Sampson, D. Gil, and P. Trudgill (eds.), *Language complexity as an evolving variable*, Oxford: Oxford University Press, 34–49.

Hall, Edward T. (1976), *Beyond Culture*, Garden City, NY: Anchor Press/Doubleday.

Pustejovsky, James, José Castao, Robert Ingria, Roser Sauri, Robert Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir Radev (2003), TimeML: Robust Specification of Event and Temporal Expressions in Text, in M.T. Maybury (ed.), *New Directions in Question Answering*, AAAI Press, 2834.

Roth, Michael, and Talita R. Anthonio (2021), UnImplicit shared task report: Detecting clarification requirements in instructional text, in *Proceedings of the 1st Workshop on Understanding Implicit and Underspecified Language*, Online: ACL, 28–32.

## The interaction of register and null object instantiations across languages

Nico Lehmann, Jozina Vander Klok, Maryam Bahmani, Elisabeth Verhoeven & Aria Adli  
(Humboldt-Universität zu Berlin, Humboldt-Universität zu Berlin, Universität zu Köln, Humboldt-Universität zu Berlin, Universität zu Köln)

Keywords: null objects, typology, register, variation, syntax

Languages vary as to their null object properties (Huang 1984, and Cole 1987), yet across languages, communicative situations related to informal discourse seem to display higher rates of null referential objects (e.g. Huang 1984, Meyerhoff 2011, Culy 1996, and Schäfer 2021, on German). Language users thus specify referential objects either explicitly or implicitly depending on the communicative situation, which highlights the register sensitivity of the phenomenon. Our research question centres on how null object properties of typologically diverse languages interact with register. In **Yucatec Maya**, null objects (ec: empty categories) are the default (1), while they are common in **Javanese** (2) and **Persian** (3) (Sato 2015, and Sato & Karimi 2016); in **German**, however, null objects are restricted to topic drop in the prefield as in (4) (Huang 1984, Cardinaletti 1990, Schäfer 2021, and Huang & Yang 2024).

### (1) Yucatec Maya

t=u            y-il-aj-e'ex-o'ob            ec            in            suku'un-o'ob  
PRFV=A.3    0-see-CMPL-B.2.PL-3.PL    \_            A.1.SG    elder.brother-PL  
'Did my brothers see you?' [CoCoYum-YT-OCE-0.33]

### (2) Javanese

Aku            nge-rusa-ke            ec  
1.SG-FM.REL ACT-broke-CAUS \_  
'I (am the one who) broke (that vase)' (Widhyasmaramurti 2008)

### (3) Persian

vali            az            in            ke            maryam            ec            na-khund            tajib=kard-am  
but            from    this            that            Maryam            \_            NEG-read.3SG    surprise=did-1SG  
'But the fact that Maryam didn't read (the book) surprised me.' (Rasekhi 2016)

### (4) German

ec hab            ich            gestern            gekauft  
\_    have    I            yesterday            bought  
'I bought (it) yesterday.' (Cardinaletti 1990)

These properties correlate with further typological differences: whereas the Yucatec verb shows obligatory object agreement, the other three languages do not have object agreement. Persian attaches an object marker to specific objects (Sato & Karimi 2016) and has both clitic and full pronouns, whereas Javanese only has full pronouns. German is distinguished as the only sentence-oriented language, e.g. requiring a subject, in contrast to the other, discourse-oriented languages, e.g. allowing discourse-bounded anaphors (Huang 1984). Null objects are also affected by semantic/pragmatic factors, e.g. type of object, animacy, person, referent status or discourse topicality (Lambrecht & Lemoine 2005,

Schwenter 2006, Pérez-Leroux, Pirvulescu & Roberge 2008, Alamillo 2009, Schnell & Barth 2018, and Schäfer 2021). For Persian, German, Bislama and Romance languages, for instance, it has been observed that null objects occur primarily with inanimate entities, things or abstract concepts and less with human entities (Cardinaletti 1990, Meyerhoff 2002, Schwenter 2006, Pérez-Leroux, Pirvulescu & Roberge 2008, and Alamillo 2009).

Understudied aspects include which register parameters favour null objects and what role semantic/pragmatic factors play in this variation. We will present a corpus study on null objects in German, Persian, Javanese and Yucatec Maya across various registers using the open-access Lang\*Reg corpus, which includes language output of the same language users in six communicative situations, distinguished by the parameters social hierarchy, social distance, mode and interactivity. Our study annotates (null) object linguistic features such as person, number, NP-type, reference, saliency, animacy, and discourse topic status. We expect to find lower pronoun rates in more informal contexts (between familiar interlocutors) for Persian, Javanese, and German, though based on the diverging grammatical properties the actual drop-rates should be much lower in German than in Persian and Javanese. Persian and Javanese are expected to show more subtle differences, given that object marking and object clitics are found only in Persian. In contrast, independent pronouns are extremely rare in the core clause in Yucatec, which is why we anticipate pronouns to require additional motivations such as emphasis, the latter being expected to be more frequent in informal situations.

## References

- Alamillo, Asela Reig. 2009. "Cross-dialectal variation in propositional anaphora: Null objects and propositional *lo* in Mexican and Peninsular Spanish." *Language Variation and Change* 21 (3): 381–412. <https://doi.org/10.1017/s0954394509990111>.
- Cardinaletti, Anna. 1990. "Subject/object asymmetries in German null-topic constructions and the status of specCP." In *Grammar in Progress*, edited by Joan Mascaró and Marina Nespor, 75–84. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110867848.75>.
- Cole, Peter. 1987. "Null Objects in Universal Grammar." *Linguistic Inquiry* 18 (4): 597–612. <http://www.jstor.org/stable/4178562>.
- Culy, Christopher. 1996. "Null objects in English recipes." *Language Variation and Change* 8 (1): 91–124. <https://doi.org/10.1017/s0954394500001083>.
- Huang, C.-T. James. 1984. "On the Distribution and Reference of Empty Pronouns." *Linguistic Inquiry* 15 (4): 531–74. <https://doi.org/10.2307/4178404>.
- Huang, C.-T. James, and Barry C.-Y. Yang. 2024. "Topic drop and pro drop." *Language and Linguistics* 25 (1): 1–27. <https://doi.org/10.1075/lali.00147.yan>.
- Lambrecht, Knud, and Kevin Lemoine. 2005. "Definite null objects in (spoken) French: A Construction-Grammar account." In *Grammatical Constructions*, edited by Mirjam Fried and Hans C. Boas, 13–55. Constructional Approaches to Language. <https://doi.org/10.1075/cal.4.03lam>.
- Meyerhoff, Miriam. 2002. "Formal and cultural constraints on optional objects in Bislama." *Language Variation and Change* 14 (3): 323–46. <https://doi.org/10.1017/s0954394502143031>.
- Meyerhoff, Miriam. 2011. *Introducing Sociolinguistics*. London: Routledge.
- Pérez-Leroux, Ana Teresa, Mihaela Pirvulescu, and Yves Roberge. 2008. "Null objects in child language: Syntax and the lexicon." *Lingua* 118 (3): 370–98. <https://doi.org/10.1016/j.lingua.2007.07.002>.



- Rasekhi, Vahideh. 2016. "Missing Objects in Persian." *Cahier de Studia Iranica* 58: 157–74.
- Sato, Yosuke. 2015. "Argument ellipsis in Javanese and voice agreement." *Studia Linguistica* 69: 58–85.  
<https://doi.org/10.1111/stul.12029>.
- Sato, Yosuke, and Simin Karimi. 2016. "Subject-object asymmetries in Persian argument ellipsis and the anti-agreement theory." *Glossa: A Journal of General Linguistics* 1 (1).  
<https://doi.org/10.5334/gjgl.60>.
- Schäfer, Lisa. 2021. "Topic drop in German: Empirical support for an information-theoretic account to a long-known omission phenomenon." *Zeitschrift Für Sprachwissenschaft* 40 (2): 161–97.  
<https://doi.org/10.1515/zfs-2021-2024>.
- Schnell, Stefan, and Danielle Barth. 2018. "Discourse motivations for pronominal and zero objects across registers in Vera'a." *Language Variation and Change* 30 (1): 51–81.  
<https://doi.org/10.1017/s0954394518000054>.
- Schwenter, Scott A. 2006. "Null Objects across South America." In *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, edited by Timothy L. Face and Carol A. Klee, 23–36. Somerville, MA: Cascadilla Proceedings Project.
- Widhyasmaramurti, W. 2008. "'Tak' and 'kok' in Javanese Language." Master's thesis, Utrecht University.

## Inferring Implicit Discourse Relations in Polish and European Portuguese

Purificação Silvano (University of Porto), Aleksandra Tomaszewska (Institute of Computer Science, Polish Academy of Sciences), António Leal (University of Porto), Maciej Ogrodniczuk (Institute of Computer Science, Polish Academy of Sciences), Evelin Amorim (INESC TEC) & Joana Gomes (University of Porto)

**Keywords:** discourse relations, aspectual classes, cross-linguistic variation, temporal relations, annotation

Text coherence relies on semantic connections between its parts. These connections, often referred to as discourse relations (DRs), have been theorized in various frameworks, including Rhetorical Structure Theory (Mann & Thompson, 1988) and Segmented Discourse Representation Theory (Asher & Lascarides, 2003). DRs have proven effective in explaining diverse linguistic phenomena, such as temporal and nominal anaphora, and the overall structure of discourse. However, fully understanding these phenomena and discourse structures requires a deeper insight into the features of DRs, particularly those implicit and not signaled by any explicit linguistic marker. In previous work, we demonstrated that, in sentences containing adverbial perfect participial clauses, certain linguistic factors – e.g., linear order of discourse, temporal relations, and aspectual classes – can predict specific DRs, and that these factors vary across languages. In the present study, we aim to identify the linguistic factors that influence the inference of implicit DRs at the textual level, thereby advancing our understanding of their defining features and facilitating their identification. To achieve this, we compiled a corpus of 30 fables, chosen for their richness in discourse relations and the availability of materials in different languages without significant licensing complications, in European Portuguese and Polish – two languages with notable typological differences. Each fable was annotated by two annotators per language, focusing on DRs, the aspectual classes of the situations represented by their arguments, and the temporal relations between situations. We employed ISO 24617-8 (ISO, 2016) as it provides an interoperable schema for annotating local discourse relations, being both language- and genre-agnostic. For aspectual classes, we used the ontology proposed by Moens (1987), while temporal relations were categorized as anteriority, posteriority, and simultaneity.

Our preliminary findings suggest that, across the languages analyzed, the number of implicit discourse relations (DRs) slightly exceeds the number of explicit ones. Nevertheless, certain DRs exhibit a strong reliance on explicit connectors, such as Concession (1). In contrast, other DRs, such as Synchrony (2), can occur with or without explicit markers. Furthermore, some explicit DRs are realized through different connectors, as observed in the case of Cause (3).

The most frequently occurring implicit DRs in our dataset are Asynchrony (4) and Conjunction. Regarding the inference of implicit DRs, our preliminary analysis indicates that, in the absence of explicit connectors, other linguistic factors, such as temporal relations (cf. Examples 3b and 4), play a pivotal role. These temporal relations are often determined by specific combinations of aspectual classes.

(1) *Everyone loved the idea, **but** there was one major obstacle.*

(2) a. ***As** he crossed a river, he caught sight of his reflection in the river.*

b. *The hare eagerly accepted, laughing at the idea.*

(3) a. *O cão estava feliz **porque** tinha encontrado um bom naco de carne.*

*The dog was happy because it had found a nice piece of meat.*

b. *Um dia, a tartaruga ficou cansada dessas humilhações e resolveu desafiar a lebre para uma corrida.*

*One day, the tortoise got tired of these humiliations and decided to challenge the hare to a race.*

(4) *One day, the tortoise (...) decided to challenge the hare to a race. The hare eagerly accepted.*

Additionally, we identified certain language-specific differences in how DRs are constructed, highlighting cross-linguistic variation in the realization of discourse structure. These insights will contribute to enhancing the ISO 24617-8 standard for annotating discourse relations.

## References

Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Bunt, H. & Prasad, R. (2016). ISO DR-core (ISO 24617-8): Core concepts for the annotation of discourse relations. In *Proceedings of the 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-12)* (pp. 45–54).

ISO. (2016). *ISO 24617-8:2016. Language resource management, Part 8: Semantic relations in discourse (DR-Core)*. International Organization for Standardization.

Mann, M. & Thompson, S. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8, 243–281.

Moens, M. (1987). *Tense, Aspect and Temporal Reference* (Doctoral dissertation). University of Edinburgh.

# Argument drop and competition effects: An experimental study on French and Spanish PPs

Steffen Heidinger & Yanis da Cunha  
(University of Graz)

Keywords: pronouns; animacy; competition; acceptability judgments; Romance languages

**Background.** Compared to subjects and direct objects, complements of prepositions are a lesser studied context for the alternation between overt and null pronouns. One factor for the distribution of null pronouns is the preposition itself (Pavón 2003, Troberg 2020). While French *avec* 'with' allows null pronouns (1) (Zribi-Hertz 1984), other prepositions such as French *vers* 'towards' or Spanish *con* 'with' do not (2-3). In this talk we provide insights in this understudied domain by considering argument drop from the perspective of animacy-related competition effects (Zribi-Hertz 2000).

**Claim.** The availability of the null pronoun, as a competitor to the overt (strong) pronoun, has an impact on the animacy restrictions of the overt pronoun: the preference of strong pronouns for human referents holds with prepositions allowing null pronouns (e.g. French *avec* 'with'), but is absent with prepositions not allowing null pronouns (e.g., French *vers* 'towards' or Spanish *con* 'with').

**Data collection.** We conducted online acceptability judgment experiments for French and Spanish. We created 16 items in each language following a 2\*2 design (animacy: human vs. inanimate; pronoun: strong vs. null). Half of the items contained argument-drop prepositions (French *avec* 'with', *sans* 'without'; Spanish *delante* 'in front', *detrás* 'behind') and the other half contained no-drop prepositions (French *malgré* 'despite', *vers* 'towards'; Spanish *con* 'with', *sin* 'without'). For each language, we recruited 32 participants (from France and Spain).

**Results** are displayed in Figures 1 and 2. The data first show that the acceptability of argument drop depends on the language: while in Spanish it is at most mildly acceptable, independently of animacy, in French, it is fully acceptable with inanimate antecedents only. Despite this difference, in both languages, the availability of argument drop has the same competition effect on the animacy restrictions of strong pronouns. Whenever argument drop is not available, strong pronouns show no animacy restrictions: they are acceptable with both human and inanimate antecedents. When argument drop is available, strong pronouns show the following animacy restriction: they are more acceptable with human than with inanimate antecedents.

**Discussion.** The exact nature of the competition effect (Müller & Sternefeld 2001) depends on assumptions about the animacy features of strong and null pronouns in the two languages. Based on a Stochastic OT analysis (Boersma & Hayes 2001) we assume that neither strong nor null pronouns are specified wrt animacy. The observed competition effect then emerges from an interaction between two constraints. While an economy principle requires the use of a null pronoun whenever available, an outranking faithfulness constraint requires referents with semantic gender (= human) to be expressed by gender-marked forms (i.e., strong and not null pronouns). It follows that strong pronouns are restricted to human antecedents only when a null competitor (i.e., argument drop) is available.

(1) Cette valise, je voyage toujours avec  $\emptyset$ . 'This suitcase, I always travel with it'

(2) Cette dame/boutique, je me dirige vers \* $\emptyset$ /elle. 'This lady/shop, I go towards her/it'

(3) Esa señora/maleta, siempre viajo con \* $\emptyset$ /ella. 'This lady/suitcase, I always travel with her/it'

## Acknowledgements

This research was funded in whole or in part by the Austrian Science Fund (FWF) 10.55776/F1003.

## References

- Boersma, Paul and Hayes, Bruce (2001), Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32(1), 45-86.
- Müller, Gereon and Sternefeld, Wolfgang (2001), The Rise of Competition in Syntax : A Synopsis. In G. Müller & W. Sternefeld (eds.), *Competition in Syntax*. Berlin: De Gruyter Mouton, 1-68.
- Pavón, María Victoria (2003). *Sintaxis de las partículas*. Madrid: Visor.
- Troberg, Michelle (2020), Les prépositions orphelines : Un réexamen à la lumière du SP étendu. *Arborescences*, 10, 185-206.
- Zribi-Hertz, Anne (1984), Orphan Prepositions in French and the Concept of 'Null Pronoun'. *The Indiana University Linguistics Club*.
- Zribi-Hertz, Anne (2000), Les pronoms forts du français sont-ils +animés? In M. Coene, W. De Mulder, P. Dendale, and Y. d'Hulst (eds.), *Traiani Augusti vestigia pressa sequamur. Studia linguistica in honorem Lilianae Tasmowski*. Padova: Unipress, 663-679.

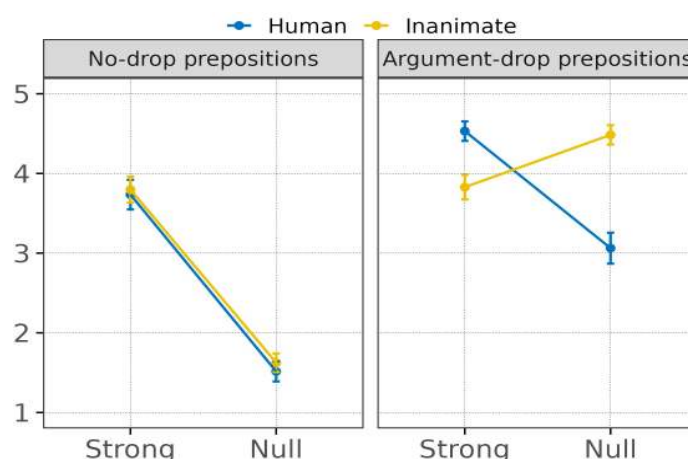


Figure 1: Mean acceptability in French vs. Animacy and Pronoun type

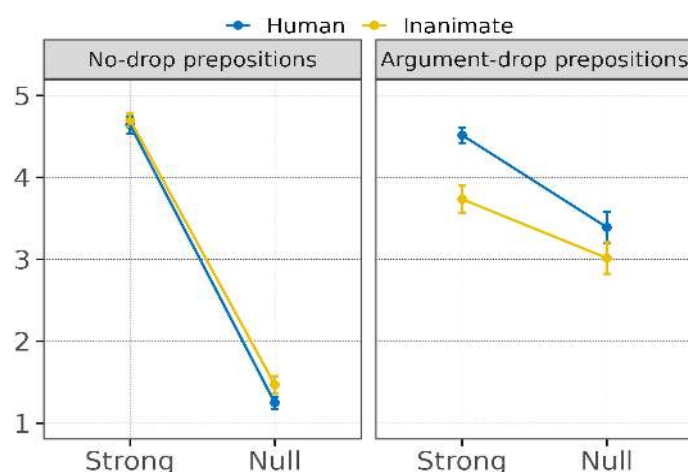


Figure 2: Mean acceptability in Spanish vs. Animacy and Pronoun type

# Causal and concessive connectives in English-German translation and interpreting

Volker Gast & Eberhard Gade

(Friedrich Schiller University Jena)

Keywords: translation, interpreting, implicitation, explicitation, connectives

Lapshinova-Koltunski et al. (2022) demonstrate that interpreting tends to favour implicitation in the expression of certain discourse relations, in contrast to written translation, which shows a tendency towards explicitation (see also Defrancq et al. 2015). They interpret this tendency as a strategy for managing high cognitive load (see also Defrancq 2015; Defrancq & Plevoets 2018 on cognitive load in interpreting). While this explanation is plausible, the hypothesis that implicitation goes hand in hand with high cognitive load warrants further investigation. For instance, interpreting is a spoken mode, which may itself account for the relative scarcity of connectives in this register (cf. Shlesinger & Ordan 2012; Przybyl et al. 2022; Gast & Borges 2023 on the spoken nature of interpreting). Furthermore, different connectives may be associated with varying degrees of implicitation.

In our presentation we investigate the relationship between implicitation and (hypothesized) cognitive load in interpreting and written translation. Based on a new translation and interpreting corpus (the MultiModalEuroParl corpus/MMEP), we analyze the distribution of causal and concessive connectives in translations and interpretations of speeches made in the European Parliament. Causal connectives are further sub-classified according to the conceptual domain of the relation (cf. Sanders & Sweetser 2009). As potential indicators of cognitive load during interpreting, we consider speech rate and *décalage*.

## General hypotheses

1. Implicitation correlates with mode of mediation (interpreting > written translation).
2. The correlation between implicitation and mode of mediation is mediated by ontological domain.
3. Within interpreting,
  - a) implicitation correlates positively with speech rate;
  - b) implicitation correlates positively with *décalage*.

More specifically, our hypotheses are:

- I.
  - a) The connectives *although* and *because* are weakened or omitted more often in interpreting than in written translation.
  - b) In the case of *because*, mode of mediation as a predictor of implicitation interacts with ontological domain.
- II.
  - a) In interpreting, the weakening of *although* or *because* correlates positively with speech rate in the preceding part of the interpretation.
  - b) In interpreting, the weakening of *although* or *because* correlates positively with *décalage* in the preceding part of the interpretation.

Hypothesis Ia will be tested by comparing the rates of weakening in written translations and the corresponding interpretations. Hypothesis Ib will be tested using semi-automatic annotations based on multilingual embeddings and translation correspondences. To test Hypothesis IIa, we will operationalize speech rate at the occurrence of a connective *c* as the number of words (normalized per second) in a 7 secs. window preceding *c* in the original speech. For Hypothesis IIb, we will calculate the average *décalage* per speech and use deviations from this average, measured within the same 7 secs. window preceding *c*, as a predictor variable for implicitation.

Preliminary results provide support for Hypothesis Ia. We have not so far observed a significant interaction between ontological domain and mode of mediation as predictors of implicitation (Hypothesis Ib). However, additional data will be processed to further explore this hypothesis. The findings related to the hypotheses in II are not conclusive. Exploratory analyses reveal an effect of *décalage* as hypothesized, but not of speech rate.

While the primary focus of this talk will be on the methodological challenges, we will also present preliminary results related to the hypotheses outlined above, and consider, as an outlook, potential theoretical implications relating to cognitive models of mediated language use, such as Gile's (2009) Effort Model.

## References

- Defrancq, Bart. 2015. Corpus-based research into the presumed effects of short EVS. *Interpreting* 17.1: 26–45.
- Defrancq, Bart and Koen Plevoets. 2018. Over-uh-load, filled pauses in compounds as a signal of cognitive load. In C. B. Mariachiaro Russo and B. Defrancq (eds.), *Making way in corpus-based interpreting studies, vol. 1*, 43–64. Singapore: Springer.
- Defrancq, Bart, Koen Plevoets and Cédric Magnifico. 2015. Connective markers in interpreting and translation: where do they come from? In Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2015. Current Approaches to Discourse and Translation Studies*, 195–222. Heidelberg: Springer.
- Gast, Volker and Robert Borges. 2023. Nouns, Verbs and Other Parts of Speech in Translation and Interpreting: Evidence from English Speeches Made in the European Parliament and Their German Translations and Interpretations. *Languages* 8.1, no 39.
- Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training*, 2nd ed. Amsterdam: Benjamins.
- Lapshinova-Koltunski, Ekaterina, Christina Pollkläsener and Heike Przybyl. 2022. Exploring explicitation and implicitation in parallel interpreting and translation corpora. *The Prague Bulletin of Mathematical Linguistics* 5: 5–22.
- Przybyl, Heike, Alina Karakanta, Katrin Menzel and Elke Teich. 2022. Exploring linguistic variation in mediated discourse: Translation vs. interpreting. In Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska and Silvia Bernardini (eds.), *Empirical Investigations into the Forms of Mediated Discourse at the European Parliament. Translation and Multilingual Natural Language Processing*, 191–218. Berlin: Language Science Press.
- Sanders, Ted and Eve Sweetser. 2009. Introduction: Causality in language and cognition – what causal connectives and causal verbs reveal about the way we think. In T. Sanders and E. Sweeters (eds), *Causal Categories in Discourse and Cognition*, 1–18. Berlin: Mouton De Gruyter.
- Shlesinger, Miriam and Noam Ordan. 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target. International Journal of Translation Studies* 24: 43–60.

## **WS7 From Linguistics to Animal Communication and Back**

**Jacob Ayers, Alexandra Bosshard, Maël Leroux, Remo Nitschke  
and Simon Townsend**



Title: Communicative efficiency in preverbal children and nonhuman ape gestural communication

Authors: [Anonymised Author List]

## **Abstract**

### **Introduction**

The world's languages are unfathomably diverse. At the same time, this complexity is not random: Statistical regularities called 'language laws' apply to most human written, signed, and spoken languages [1, 2], reflecting energetically efficient communication. However, it remains unclear if human non-linguistic communication (e.g., gestures) follows similar patterns and at which point in human development communicative efficiency emerges, as no studies have yet assessed the communicative system of preverbal children. Similarly, because language does not fossilise, we have limited understanding of the emergence of these properties in human evolution, and to address these questions we must rely on comparative studies with other species. One way to compare the degree of efficiency across gestural (human and nonhuman) communicative systems is to apply the recently developed measure of optimality (Psi) [3-7]. Language optimality measures how efficient a communicative system is as compared to its most efficient expression, considering the duration and frequency of the units assessed and present in the repertoire (e.g., letters in words, duration, and frequency of gesture types) [3]. For the first time, we combine linguistic advances and gestural methodologies to assess the efficiency of four ape species' gestural repertoires: chimpanzees, gorillas, bonobos, and preverbal children. We aim to assess the degree of efficiency of a shared gestural system on two duration measures: One that quantifies the minimum duration of a signal necessary to be

understood (Minimum Action Unit - MAU, [8]), which should be shaped by selective pressures over evolutionary time, and one that includes the full duration of a gestural performance (Performed Action Unit – PAU: MAU + persistence/redundancy), which is most likely affected by proximate factors within each individual interaction. Finally, we will perform a species comparison on optimality MAU values to address whether and how species affects communicative system efficiency.

### **Methods**

This analysis will be based on a gestural databased comprising more than 22.000 gestures from 22 groups of 5 species (5 eastern chimpanzee groups, 4 western chimpanzee groups, 4 gorilla groups, 2 bonobo groups, 7 human groups). For each group, bootstrapped optimality values (Psi) will be computed, with resampling to control for uneven samples. We will compare gestural repertoire optimization based on MAU and PAU durations and analyse optimality distributions between species using a linear model with bootstrapped optimality values as the response variable, species as a predictor, and community as a random factor. Analyses will be conducted in R.

### **Expected results**

We expect higher optimality values based on MAU durations across groups. Species differences are predicted to significantly impact optimality values, with human gestures being more optimised than apes’.

### **Discussion**

This first-of-its-kind analysis identifies communicative efficiency in ape and child gestural communication and compares their efficiency on the same scale. By assessing

preverbal human children and ape gestures we hope to draw more direct comparisons between species, and to provide a suitable comparison with the efficiency of spoken language studies. In doing so we provide a better understanding of the evolution and onset of the unique communicative means that is language.

## References

1. Bentz C, Ferrer-i-Cancho R. Zipf's law of abbreviation as a language universal. In: Bentz C, Jäger G, Yanovich I, editors. *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*. University of Tübingen, 2015.
2. Börstell C, Hörberg T, Östling R. Distribution and duration of signs and parts of speech in Swedish Sign Language. *SL&L*. 2016;19(2):143–96.
3. Petrini S, Casas-i-Muñoz A, Cluet-i-Martinell J, Wang M, Bentz C, Ferrer-i-Cancho R. The optimality of word lengths. Theoretical foundations and an empirical study. arXiv; 2023. Available from: <http://arxiv.org/abs/2208.10384>
4. Ferrer-i-Cancho R, Bentz C. The evolution of optimized language in the light of standard information theory. In: *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)* Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika; 2018.
5. Gibson E, Futrell R, Piantadosi SP, Dautriche I, Mahowald K, Bergen L, et al. How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*. 2019;23(5):389–407.
6. Levshina N. *Communicative efficiency: language structure and use*. Cambridge New York, NY Port Melbourne, VIC New Delhi Singapore: Cambridge University Press; 2022. p. 291.
7. Liu H, Xu C, Liang J. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*. 2017;21:171–93.
8. Grund C, Badihi G, Graham KE, Safryghin A, Hobaiter C. GesturalOrigins: A bottom-up framework for establishing systematic gesture data across ape species. *Behav Res Methods*. 2023.

## Exploring Sparrows' Semantics

Ambre Salis, Aria Zhang, Vicky Lin, Philippe Schlenker, Emmanuel Chemla & Julia Schroeder  
(Imperial College London; Imperial College London; Imperial College London; Institut Jean Nicod;  
Laboratoire de Sciences Cognitives et Psycholinguistique & Earth Species Project; Imperial College  
London)

**Keywords:** animal communication, animal linguistics, comparative linguistics, house sparrows, semantics, call

The field of animal linguistics has recently introduced a set of tools and specific methodologies to investigate animal calls using linguistic concepts. In addition to procedures on how to explore the core meaning of calls (Berthet et al. 2023), or call combinations in birds (Schlenker et al. 2024a), more precise hypotheses such as featural interpretation have been suggested in birds (Salis et al. 2024, Schlenker et al. 2024b). Featural interpretation pertains to the concept that meaning-bearing units are composed of acoustic features rather than discrete notes. This perspective moves away from the traditional view in animal linguistic studies that equates 'one call type' with 'one meaning.' This hypothesis serves as a compelling explanation for heterospecific communication (Schlenker et al. 2024b) but may also elucidate the apparent paradox of highly social species having seemingly limited vocal repertoires, as exemplified by the house sparrow, *Passer domesticus*.

House sparrows represent one of the most prevalent bird species globally and are central to ecological studies about their reproduction, population dynamics, and genetics (Anderson 2006). Despite their widespread presence, their vocal repertoire and coding mechanisms remain surprisingly under-researched. Given their apparently limited vocal repertoire (approximately 10 calls described, Nivison 1978) and their intricate social structure (gregarious flocks organized by social dominance, exhibiting social monogamy during the reproductive season), they present an excellent opportunity to test the featural interpretation hypothesis and explore how meaning is encoded in this type of species.

By uniting the expertise of linguists and ecologists, we are therefore undertaking a project that investigates the call system of house sparrows. Specifically, our objectives are (i) to precisely identify the core meaning of their most common calls, (ii) to identify the acoustic diversity inside each call type, notably by differentiating contextual variation from individual signatures and (iii) to explore the potential use of featural coding mechanisms, for example in food discovery contexts. To attain the first objective, we use 360° cameras coupled with high quality recorders to precisely describe the surrounding context when specific notes are produced. For the second objective, recordings of known individuals are carried out in both winter and spring to assess the within- and inter-individual variation in their most common call, the 'cheep' call. To do this, we are taking advantage of a monitored population of house sparrows on Lundy Island (UK). This population is a closed, small population (~50 individuals) that are all ringed and with RFID tags, allowing individual recognition while still recording natural behaviours. Finally, the third objective focuses on experimentally testing whether food discovery calls are produced in a different manner depending on the quality and the quantity of food,

with the hypothesis that call rate, peak frequency and/or non-linearities can inform about these different situations. The initial findings of these three questions will be presented, with the expectation that they will encourage new discussions on how methodologies in animal linguistics can be adapted to different species and will address potential future directions.

### Acknowledgments

This project passed the European Research Executive Agency's ("REA") assessment process (MSCA fellowship) and has subsequently been transferred to the UK. UKRI (UK Research and Innovation) is therefore the founder of this fellowship examining semantics of sparrows: SPANTICS (grant reference: EP/Z002508/1).

### References

- Anderson, Ted (2006), *Biology of the ubiquitous house sparrow: from genes to populations*. Oxford University Press: UK.
- Berthet, Mélissa, Coye, Camille, Dezechache, Guillaume, and Kuhn, Jeremy (2023), Animal linguistics: a primer. *Biological Reviews* 98(1), 81–98.
- Nivison, James Jay (1978) *The vocal behavior and displays of the house sparrow, Passer domesticus L., in the United States*. Thesis, Wayne State University.
- Schlenker, Philippe, Salis, Ambre, Leroux, Mael, Coye, Camille, Rizzi, Luigi, Steinert-Threlkeld, Shane, and Chemla, Emmanuel (2024a) Minimal Compositionality versus Bird Implicatures: two theories of ABC-D sequences in Japanese tits. *Biological Reviews* 99(4), 1278-1297.
- Schlenker, Philippe, Coye, Camille, Salis, Ambre, Steinert-Threlkeld, S, Ravaux, Lucie and Chemla, Emmanuel (2024b) Anti-Babel: Three degrees of interspecies comprehension. *Mind & Language* (in press, preprint available at : <https://ling.auf.net/lingbuzz/007168> )
- Salis, Ambre, Badaire, Flavien, Coye, Camille, Leroux, Mael, Lengagne, Thierry, Schlenker, Philippe, and Chemla, Emmanuel (2024) Mechanisms of mobbing call recognition: Exploring featural decoding in great tits. *Animal Behaviour* 216, 63-71.

## Rule-based sequences in sooty mangabey vocal communication

Auriane Le Floch<sup>1,2,3,4,11\*</sup>, Cédric Girard-Buttoz<sup>3,5,6</sup>, Tanit Souha Azaiez<sup>2,3,4</sup>, Natacha Bande<sup>3</sup>, Roman M. Wittig<sup>1,3,6</sup>, Steven Moran<sup>2,8</sup>, Klaus Zuberbühler<sup>2,4,7,9</sup>, Catherine Crockford<sup>1,3,6,9\*\*1</sup>  
The Ape Social Mind Lab, Institut des Sciences Cognitives Marc Jeannerod, CNRS, Bron, France

<sup>1</sup>The Ape Social Mind Lab, Institut des Sciences Cognitives Marc Jeannerod, CNRS, Bron, France

<sup>2</sup>Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

<sup>3</sup>Taï Chimpanzee Project, Centre Suisse de Recherches Scientifique en Côte d'Ivoire, Abidjan, Ivory Coast

<sup>4</sup>Taï Monkey Project, Centre Suisse de Recherches Scientifique en Côte d'Ivoire, Abidjan, Ivory Coast

<sup>5</sup>ENES Bioacoustics Research Laboratory, Centre de Recherche en Neurosciences de Lyon, CNRS, Inserm, University of Saint-Etienne, Saint-Etienne, France

<sup>6</sup>Department of Human Behaviour, Ecology and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>7</sup>School of Psychology & Neuroscience, University of St Andrews, Scotland

<sup>8</sup>Linguistic Research Infrastructure (LiRI), University of Zurich, Zurich Switzerland

<sup>9</sup>Joint senior authors

<sup>10</sup>Lead contact

\*Corresponding authors: Auriane Le Floch (aurianelefloch@me.com); Catherine Crockford (catherine.crockford@isc.cnrs.fr).

Investigating how non-human animals produce call sequences provides key insights into the evolutionary origins of meaning in vocal communication, including syntax. Many genetically distant species combine calls into structured sequences that follow specific rules. However, most studies focus on only a narrow portion of the vocal repertoire. We addressed this gap by documenting the vocal sequence repertoire and structural rules of wild sooty mangabeys (*Cercocebus atys*), a West African monkey species. Over 11 months, we followed two habituated groups in Taï National Park, Ivory Coast, and manually annotated 1,672 vocal utterances. Annotation involved visual inspection of spectrograms combined with auditory verification to identify distinct sound elements—continuous sounds with clear beginnings and endings. We defined a call as one or more elements of the same type produced in succession with inter-element intervals of less than two seconds, and a sequence as a series of different consecutive call types, each separated by less than one second. To detect rule-based structure, we used Bayesian models to predict the likelihood of a call occurring in

specific sequence positions and identified statistically significant patterns to formalise the sequence “grammar”.

Our analyses revealed that sooty mangabeys combine most of their calls into diverse sequences but rely heavily on a restricted set of combinations. Among the most frequent sequences, we identified:

(1) Ordering – we identified four call types that were more likely to occur in specific positions (e.g., first or second); (2) Affixation – one call, ‘hoo’, never occurred alone, did not involve element repetitions and was consistently attached to other stand-alone calls; (3) Recombination – independently emitted short sequences (e.g., the bigrams ‘grunt<sup>n</sup>\_twitter<sup>n</sup>’ and ‘twitter<sup>n</sup>\_grunt<sup>n</sup>’) were concatenated into longer ones; (4) Non-adjacent dependencies – a rule linking the first and last call of a sequence irrespective of intervening calls, such as iterative ‘grunt<sup>n</sup>\_twitter<sup>n</sup>’ sequences terminated with a ‘grunt<sup>n</sup>’; (5) Potential recursion – the self-embedding of call pairs within larger identical structures (e.g.,  $AB \rightarrow ABABCAB$ ), with ‘grunt<sup>n</sup>\_twitter<sup>n</sup>’ bigrams embedded multiple times in extended sequences. Although these patterns are not equivalent to human syntax, they formally resemble features such as positional constraints, morphological affixation, ordered dependencies, and recursive embedding. We note this parallel with caution and use it solely to frame evolutionary questions. Clarifying how these structural rules map onto meaning remains the next critical step.

<sup>n</sup>indicates that each call can have any number of vocal element repetitions.

## Reviews

----- REVIEW 1 -----

SUBMISSION: 9748

TITLE: Organisational rules in sooty mangabey vocal sequences

AUTHORS: Auriane Le Floch, Cédric Girard-Buttoz, Tanit Souha Azaiez, Natacha Bande, Roman M. Wittig, Steven Moran, Klaus Zuberbühler and Catherine Crockford

----- Overall evaluation -----

SCORE: 5 (excellent: strong accept)

----- Reviewer's confidence -----

SCORE: 4 ((high) I am fairly familiar with the area of this abstract)

----- Review -----

This is a well written, clear abstract. I wish more specific examples were discussed, ideally with reference to parallels in human language, but I hope that will be done in the actual paper.

We thank the reviewer for the positive feedback. In response to the suggestion, we now include specific examples of structural patterns in sooty mangabey sequences and carefully note formal parallels to features of human language, while explicitly emphasising our cautious approach to such comparisons.

----- REVIEW 2 -----

SUBMISSION: 9748

TITLE: Organisational rules in sooty mangabey vocal sequences

AUTHORS: Auriane Le Floch, Cédric Girard-Buttoz, Tanit Souha Azaiez, Natacha Bande, Roman M. Wittig, Steven Moran, Klaus Zuberbühler and Catherine Crockford

----- Overall evaluation -----

SCORE: 5 (excellent: strong accept)

----- Reviewer's confidence -----

SCORE: 4 ((high) I am fairly familiar with the area of this abstract)

----- Review -----

I rate this abstract a strong accept.

My main comment is this:

I would have liked to see short explanation or preview as to how (as in, with what methodology and definition for these terms) the authors found hierarchical and recursive structures in the sooty mangabey repertoire. I hope this will be explained in the presentation.

Regardless of this small critique, I believe this paper fits well into the workshop theme and will be of interest to linguists and animal communication researchers alike.

We thank the reviewer for their positive evaluation and insightful comments. In response, we have clarified in the abstract how we operationalised key concepts such as hierarchical structure, recursion, recombination, and non-adjacent dependencies, and how these were tested statistically

----- REVIEW 3 -----

SUBMISSION: 9748

TITLE: Organisational rules in sooty mangabey vocal sequences

AUTHORS: Auriane Le Floch, Cédric Girard-Buttoz, Tanit Souha Azaiez, Natacha Bande, Roman M. Wittig, Steven Moran, Klaus Zuberbühler and Catherine Crockford

----- Overall evaluation -----

SCORE: 4 (very good: accept)

----- Reviewer's confidence -----

SCORE: 5 ((expert) consider me an expert on this abstract, I know the field very well, I know that my judgment is correct)

----- Review -----

Very interesting, I would like to know more! The abstract runs a bit short in specifying the procedure followed for the characterization and decomposition of calls, if accepted for presentation the criteria followed should be made explicit.

We thank the reviewer for their positive feedback and insightful comments. We have now incorporated a clear description of our call annotation procedure in the abstract.



## **Versatile vocal combinations as a tool to convey a larger range of meanings in chimpanzees**

Human language is a highly versatile combinatorial communication system able to generate an infinite number of meanings via the combinations of phonemes into words and words into hierarchically structured sentences<sup>1</sup>. Non-human animals also use call combinations to modify meanings of single vocal units using different combinatorial mechanisms<sup>2–6</sup>. Combinatorial mechanisms describe the way the meanings of single calls are altered when strung into call combinations. To date, at most one combinatorial mechanism per species has been documented. Whether this apparent vast evolutionary gap between human language and other animal communication systems is real or an artefact of studies largely focusing on only a few call combinations per species remains unclear. Here we assessed the potential for chimpanzees to use vocal combinations to expand the range of meanings they can convey through versatile use of various combinatorial mechanisms. This species is of particular interest here since they utter hundreds of call combinations that demonstrate structural rules, in particular positional bias of certain call types and dependencies between calls in the combinations.

Using 4,323 utterances recorded in 53 Taï Forest chimpanzees, we first assess the overall potential for vocal combinations to convey information about several events within a single utterance. We found that the likelihood of uttering call combinations rather than single calls doubled when chimpanzees encountered more complex situations characterised by concomitant events such as when greeting individuals in a food patch during a fusion. The number of different situations individuals encountered also correlated with the number of different vocal combinations they used. This relationship was stronger in adults than in immatures, indicating that the capacity to match events to call combinations may develop throughout ontogeny.

Second, we assessed how the potential meanings of utterances containing two call types (bigrams) were derived from the meanings of their component calls, by assessing the events in which each utterance was emitted. Focusing on the 16 most common bigrams in the chimpanzee vocal repertoire, we found four combinatorial mechanisms that expand meanings: new meaning creation, meaning

combination, meaning disambiguation, and call order effect on meanings. Each combinatorial mechanism was evident in at least two bigrams. Also, most call types were utilised in several bigrams and in more than one combinatorial mechanism across a wide range of daily life events. Such a versatile combinatorial communication system can enable small repertoires to encode many more meanings than there are call types. Such a system has never been documented and may be transitional between more rudimentary systems previously described and open-ended systems such as human language.

## References

1. Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K. & Bickel, B. Compositionality in animals and humans. *PLoS Biol.* **16**, 1–7 (2018).
2. Suzuki, T. N., Wheatcroft, D. & Griesser, M. Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* **7**, 10986 (2016).
3. Leroux, M. *et al.* Call combinations and compositional processing in wild chimpanzees. *Nat. Commun.* **14**, 2225 (2023).
4. Engesser, S., Ridley, A. R. & Townsend, S. W. Meaningful call combinations and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci.* **113**, 5976–5981 (2016).
5. Arnold, K. & Zuberbühler, K. Semantic combinations in primate calls. *Nature* **441**, 303–303 (2006).
6. Coye, C., Zuberbühler, K. & Lemasson, A. Morphologically structured vocalizations in female Diana monkeys. *Anim. Behav.* **115**, 97–105 (2016).

# Learning new sounds: Understanding and overcoming phonological interference in songbirds and humans

Dina Lipkind, Virginia Valian, Tomas Tomka, Richard Hahnloser & Suzanne van der Feest  
(York College, City University of New York, New York,  
Hunter College, City University of New York, New York,  
ETH Zurich, ETH Zurich, & The Graduate Center, City University of New York)

Keywords: Second language learning, phonology, birdsong, vocal learning, zebra finch

Humans and songbirds share the ability to learn multiple languages or songs – an important adaptation to complex and dynamic social environments. However, this ability has a paradoxical downside: what we learn can get in the way of further learning. A prominent example is the notorious difficulty of learning the sounds of a second language (L2), especially in adulthood, due to interference from learned sound categories of the native language (L1). The precise mechanism that causes interference in L2 learning is still unknown, as is an effective intervention for reducing interference (Bundgaard-Nielsen 2011, Flege 1993, Kuhl 2008, Flege 1987, Sakai 2018 and Mokari 2017). Also unknown is whether phonological interference is a uniquely human phenomenon or whether it is shared with other species of vocal learners. Understanding the mechanism of interference and its generality across species is necessary for gaining deeper insights into learning and categorization of sounds, and for developing more efficient methods to facilitate second language learning.

We used an experimental paradigm of “second song” learning in zebra finches (Lipkind 2013, 2017 and Toutounji 2024). to test for the existence of phonological interference in birdsong learning and to attempt to eliminate interference by manipulating the birds’ auditory input. Artificial bird “tutors” induced juvenile males to learn first one synthetic song (S1) and then a second song (S2), differing from S1 in selected features. We tested if and how the acoustic difference between S1 and S2 affected birds’ learning success of S2 sounds.

Preliminary results show that, like humans, zebra finches encounter interference from a previously learned S1 sound when attempting to learn a new, similar, S2 sound (differing from S1 by a small pitch interval). Instead of fully learning the new sound, the birds sing a sound with an intermediate pitch between S1 and S2. However, we can fully eliminate this interference by adding a second new sound shifted by the same pitch interval but in the opposite direction from the old sound. While a single new target results in only partial learning, exposure to *two opposing new targets* leads to the successful learning of both.

These preliminary findings provide the first evidence for “phonological” interference in songbirds, and a new direction for understanding and resolving interference. In humans, native language sound categories remain somewhat plastic throughout life to adapt to diverse speakers (Olasagasti 2020 and Clayards 2008). This life-long L1 plasticity is hypothesized to be served by a Bayesian-inference-based mechanism (Kleinschmidt 2015 and Theodore 2019) that infers changes in the acoustic-phonetic characteristics of L1 categories from ongoing auditory experience. L1 plasticity might have the downside of interfering with the acquisition of L2 sound categories because hearing new L2 sounds may lead to partial adjustment of similar L1 categories rather than formation of new L2 categories. We hypothesize that input with two new targets at opposite acoustic directions from an old sound (i.e., an L2 target and an artificially generated opposing target) may reduce interference because it has a low probability of belonging to a common category and would therefore bias a Bayesian-based mechanism to reject the new sounds as evidence for a shift in an old category and form new sound categories instead.

## References:

- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, 32(1), 51–67. <https://doi.org/10.1017/S0142716410000287>
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809. <https://doi.org/10.1016/j.cognition.2008.04.004>
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, 93(3), 1589–1608. <https://doi.org/10.1121/1.406818>
- Flege JE. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47–65.
- Kleinschmidt, D. F., & Florian Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. <https://doi.org/10.1037/a0038695>
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000. <https://doi.org/10.1098/rstb.2007.2154>
- Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., & Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, 498(7452), 104–108. <https://doi.org/10.1038/nature12173>
- Lipkind, D., Zai, A. T., Hanuschkin, A., Marcus, G. F., Tchernichovski, O., & Hahnloser, R. H. R. (2017). Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-01436-0>
- Mokari, G. P., & Werner, S. (2017). Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners’ production and perception of Standard Southern British English vowels. *Lingua*, 185, 81–95. <https://doi.org/10.1016/j.lingua.2016.07.008>
- Olasagasti, I., & Giraud, A. L. (2020). Integrating prediction errors at two time scales permits rapid recalibration of speech sound categories. *ELife*, 9, 1–22. <https://doi.org/10.7554/eLife.44516>
- Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187–224. <https://doi.org/10.1017/S0142716417000418>
- Theodore Rachel M, & Monto Nicholas R. (2019). Distributional learning for speech reflects cumulative exposure to a talker’s phonetic distributions. *Psychonomic Bulletin & Review Bull Rev*, 26(3), 985–992.
- Toutounji, H., Zai, A. T., Tchernichovski, O., Hahnloser, R. H. R., & Lipkind, D. (2024). Learning the sound inventory of a complex vocal skill via an intrinsic reward. *Science Advances*, 10(13). <https://doi.org/10.1126/sciadv.adj3824>

# Investigating zebra finch responses to simulated vocal interactions in a virtual reality system

Maxine Guttman, Catherine Alexander, Gabriella Pilguy, Eitan Globerson, Ofer Tchernichovski & Erich D. Jarvis

(Department of Psychology, Hunter College, CUNY; Department of Psychology, Hunter College, CUNY; Department of Psychology, Hunter College, CUNY; Technion: Israel Institute of Technology; Department of Psychology, Hunter College, CUNY, CUNY Graduate Center, New York, NY, USA; The Rockefeller University.)

**Keywords:** birdsong, virtual reality, cultural transmission, vocal communication, vocal sequences, hierarchical structure

Vocal communication in songbirds includes both culturally transmitted (Fehér et al. 2009) and innate (Elie & Theunissen 2016) behaviors that have been studied intensively in both the field and lab. Songbirds often vocalize thousands of times a day. Their vocal exchanges of songs and calls can influence—attract or repel—other birds. We have a good understanding of the mechanisms through which songbirds learn and produce complex vocal sequences and rhythms (Hahnloser et al. 2002). But we know much less about how they perceive and interpret either these vocalizations or their hierarchical structure. In spoken language, semantics and hierarchy are strongly linked. But even in the most studied songbird, the zebra finch, it is not clear if any features of song or call exchanges have any semantic value. Rhythm, vocal coordination, and hierarchical-structure patterns may instead influence behavior through aesthetic qualities, perhaps as music affects human listeners (Rothenberg et al. 2014). Phonotaxis and operant-conditioning experiments are limited in their ability to reveal how high-level features of zebra finch vocal communication could affect their behavior.

We developed a virtual reality system—Avian Anime—to explore how vocal sounds that vary in hierarchy, rhythm, and acoustic structure may affect zebra finch behavior in different ecological contexts. In our system, a bird is housed in a cage where it hears sounds from speakers on each side and sees animations from video monitors that act as walls. The system collects data from microphones above and proximity sensors on each perch and responds to the bird's behavior in real time. It automatically estimates absolute preference levels for specific sounds. It also captures how each sound or combination of sounds and animations affects the bird's movement patterns. Avian Anime can simulate many different ecological contexts (e.g., courtship). For example, an animated male or female zebra finch can approach the live zebra finch and then move and vocalize in response to its calls and movements. This constitutes a closed feedback loop between the system and the bird, where parameters such as response delay and predictability can be experimentally manipulated.

Preliminary results demonstrate that our system can successfully engage with both male and female zebra finches over prolonged sessions. We found that interactions with different visual and auditory scenarios produce consistent results. We identified different behavioral responses with a much higher signal-to-noise ratio than standard methods such as phonotaxis. We found, for example, that interactive short calls attract the birds and increase their activity levels, whereas other types of sounds repel them or persistently decrease their activity levels. We are currently using Avian Anime to systematically explore birds' behavioral responses over time to different kinds of calls and songs—a large range of natural and synthetic sounds that vary in rhythm, order, symmetry, and hierarchy. This lets us document zebra finch responses and interaction dynamics in a scalable and generic manner. We hope that this approach can eventually form a suitable foundation for semantic analysis of vocal communication patterns across species.

## References

- Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246), 564-568.
- Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal cognition*, 19, 285-315.
- Hahnloser, R. H., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902), 65-70.
- Rothenberg, D., Roeske, T. C., Voss, H. U., Naguib, M., & Tchernichovski, O. (2014). Investigation of musicality in birdsong. *Hearing research*, 308, 71-83.

## Compositionality in Wild Bonobo Vocal Communication

A crucial hallmark of human language is the capacity to combine elements. For example, morphemes can be combined into words (e.g., “bio” + “logy” = “biology”) or words into sentences (“Biology is interesting”). This is possible due to compositionality, whereby meaningful units are combined into larger structures whose meaning is determined by the meanings of the parts and the way they are combined (Partee 2004). Compositionality can be trivial (the meaning of a combination is derived from adding the meaning of its parts, as in “blond dancer”) or non-trivial (the meaning of one element modifies the meaning of the other element, as in “bad dancer”) (Steinert-Threlkeld 2020; Martin 2022). Several studies in birds and primates have demonstrated that animals are capable of combining meaningful vocalisations into trivially compositional structures (e.g., Engesser et al. 2016; Suzuki and Matsumoto 2022; Leroux et al. 2023), but to date, unambiguous evidence of non-trivial compositionality in animals from systematically collected quantitative data is still lacking (Schlenker et al. 2016, 2023, 2024; Sauerland 2016; Kuhn et al. 2018; Steinert-Threlkeld 2020; Leroux et al. 2023; Trujillo and Holler 2024; Beckers et al. 2024).

Here, we provide robust empirical evidence for non-trivial compositionality in the vocal communication of wild bonobos (*Pan paniscus*). Firstly, we leveraged a framework that investigates meaning by considering all aspects of context that co-occur with the emission of the signal (Berthet et al. 2023). This approach defines the meaning of a signal as the set of features of circumstances (hereafter, FoCs) that appear at a rate greater than chance across the signal’s occurrences. We recorded 700 bonobo vocal utterances and collected a set of more than 300 FoCs for each utterance. Secondly, using a method adapted from Distributional Semantics, a linguistic approach that quantifies meaning similarities between words (Harris 1954), we used these FoCs to map bonobo utterance types within a semantic space and quantify meaning similarities between utterance types. Lastly, we investigated whether the bonobo call system is compositional. To this end, we used a multi-step process developed by Trujillo and Holler, previously developed to identify non-trivial compositionality in human multimodal communication (Trujillo and Holler 2024). This approach considers a combination AB as compositional if i) the meaning of A is distinct from that of B, ii) the meaning of AB is different from that of A and that of B, and iii) the meaning of AB is derived from the meaning of A and B. The fourth step determines whether a compositional combination is trivial or non-trivial: the combination AB represents a non-trivial compositional structure if it is compositional (i.e., it fulfils criteria i-iii) and if iv) the meaning of AB is different from the meaning of A+B.

We found that bonobos produce four two-call compositional structures involving all call types of their repertoire. Most strikingly, three of the compositional structures detected exhibit non-trivial compositionality. These findings indicate that compositionality is a prevalent feature of the bonobo vocal system, which may display more parallels with human language than previously thought.

## References

- Beckers GJL, Huybregts MAC, Everaert MBH, Bolhuis JJ (2024) No evidence for language syntax in songbird vocalizations. *Front Psychol* 15:. <https://doi.org/10.3389/fpsyg.2024.1393895>
- Berthet M, Coye C, Dezechache G, Kuhn J (2023) Animal linguistics: a primer. *Biological Reviews* brv.12897. <https://doi.org/10.1111/brv.12897>
- Engesser S, Ridley AR, Townsend SW (2016) Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences* 113:5976–5981

- Harris ZS (1954) Distributional structure. *WORD* 10:146–162.  
<https://doi.org/10.1080/00437956.1954.11659520>
- Kuhn J, Keenan S, Arnold K, Lemasson A (2018) On the -oo suffix of Campbell’s monkeys. *Linguistic Inquiry* 49:169–181. [https://doi.org/10.1162/LING\\_a\\_00270](https://doi.org/10.1162/LING_a_00270)
- Leroux M, Schel AM, Wilke C, et al (2023) Call combinations and compositional processing in wild chimpanzees. *Nat Commun* 14:2225. <https://doi.org/10.1038/s41467-023-37816-y>
- Martin JR (2022) Compositional routes to (non)intersectivity. PhD thesis, Harvard University
- Partee BH (2004) Compositionality. In: *Compositionality in formal semantics: selected papers*. Blackwell, Malden, Mass., pp 153–181
- Sauerland U (2016) On the definition of sentence. *Theoretical Linguistics* 42:..  
<https://doi.org/10.1515/tl-2016-0007>
- Schlenker P, Chemla E, Schel AM, et al (2016) Formal monkey linguistics. *Theoretical Linguistics* 42:1–90. <https://doi.org/10.1515/tl-2016-0001>
- Schlenker P, Coye C, Leroux M, Chemla E (2023) The ABC-D of animal linguistics: are syntax and compositionality for real? *Biological Reviews* brv.12944. <https://doi.org/10.1111/brv.12944>
- Schlenker P, Salis A, Leroux M, et al (2024) Minimal Compositionality versus Bird Implicatures: two theories of ABC-D sequences in Japanese tits. *Biological Reviews*.  
<https://doi.org/10.1111/brv.13068>
- Steinert-Threlkeld S (2020) Toward the emergence of nontrivial compositionality. *Philos of Sci* 87:897–909. <https://doi.org/10.1086/710628>
- Suzuki TN, Matsumoto YK (2022) Experimental evidence for core-Merge in the vocal communication system of a wild passerine. *Nat Commun* 13:5605. <https://doi.org/10.1038/s41467-022-33360-3>
- Trujillo JP, Holler J (2024) Conversational facial signals combine into compositional meanings that change the interpretation of speaker intentions. *Sci Rep* 14:2286.  
<https://doi.org/10.1038/s41598-024-52589-0>



# Where Do Hockett's 'Design Features of Language' Stand 65 Years Later?

Michael Pleyer & Elizabeth Qing Zhang  
(Nicolaus Copernicus University in Toruń & Jiangsu Normal University)

Keywords: Animal Communication, Language Evolution, Design Features, Hockett

In this talk, we aim to give an overview of two central developments regarding Hockett's design features of human language: 1. The current state of the presence of Hockett's design features in non-human animals 2. recent developments in more fundamental re-evaluations of Hockett's design features.

Hockett's design features (Hockett 1960, 1963) have remained highly influential in linguistics and the language sciences. They still feature prominently in many linguistics textbooks (Wacewicz et al. 2023) and popular press publications when dealing with the comparison of human language with other animal communication systems and the evolution of language (e.g. Johansson 2021).

However, in the roughly 65 years since they were first proposed, they have not remained without criticism (e.g. Bender 1990; Oller 2004; Wacewicz & Żywicznyński 2015). Many such criticisms have focussed on the necessity to add certain design features (Aitchison 2008) or on certain features being a direct consequence of more basic properties or not being characteristic of language as a whole. For example, Hockett's design features have rightly been criticised for disregarding signed languages (Johansson 2021). In addition, research on animal communication systems has also shown that many features are instead more widely shared with other animal communication systems, at least to some degree (see e.g. Pleyer & Hartmann 2024 for a review). For example, duality of patterning – the combination of single meaningless elements into meaningful composite units – has been documented in chestnut-crowned babblers (Engesser et al. 2019) and limited compositionality seems to present in the communication systems of other animals, such as some bird and monkey species (Townsend et al. 2018).

Going beyond this, there have been increasing calls for a more fundamental re-evaluation of Hockett's design features. For one, there has been criticism that some features, such as arbitrariness, are too language-centric to be fruitfully applied to comparisons with animal communication systems (Watson et al. 2022). In addition, it has been argued that more attention needs to be paid to the underlying cognitive capacities and social and ecological settings involved in the realisation of design features (e.g. Wacewicz & Żywicznyński 2015). For example, for a feature such as displacement – the ability of language to refer to things that are remote in both time and space – a species-comparative perspective should not only simply outline to what degree the feature is present in other animals. Instead, it should also investigate the cognitive capacities underlying displacement – such as the ability for mental time travel – and investigate to which these capacities are present in other animals (Pleyer & Zhang 2022). A more fundamental re-evaluation of Hockett's design features should also take into account properties that more recent research has identified as being central to human language on all levels of analysis, such as its inherently multimodal nature (Cohn & Schilperoord 2024) or the inherently dynamic nature of language as a complex adaptive system (Beckner et al. 2009) and investigate to what extent such properties also characterise non-human communication systems (Hebets et al. 2016; Zhang & Pleyer 2024).

## References

- Aitchison, J. (2008). *The articulate mammal: An introduction to psycholinguistics* (5th ed.). Routledge.
- Bender, M. L. (1990). [Review of *The Genesis of Language: A Different Judgment of Evidence*, by M. E. Landsberg]. *Anthropological Linguistics*, 32(1/2), 179–183. <http://www.jstor.org/stable/30028149>
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59, 1–26. <https://doi.org/10.1111/j.1467-9922.2009.00533.x>
- Cohn, N., & Schilperoord, J. (2024). *A Multimodal Language Faculty: A Cognitive Framework for Human Communication*. Bloomsbury Publishing.
- Engesser, S., Holub, J. L., O'Neill, L. G., Russell, A. F., & Townsend, S. W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proceedings of the National Academy of Sciences*, 116(39), 19579–19584.
- Hebets, E. A., Barron, A. B., Balakrishnan, C. N., Hauber, M. E., Mason, P. H., & Hoke, K. L. (2016). A systems approach to animal communication. *Proceedings of the Royal Society B: Biological Sciences*, 283(1826), 20152889.
- Hockett, C. F. (1960). The origins of speech. *Scientific American*, 203, 88–96.
- Hockett, C. F. (1963). The problem of universals in language (J. H. Greenberg, Ed.; pp. 1–22). MIT Press.
- Johansson, S. (2021). *The Dawn of Language: The story of how we came to talk* (F. Perry, Trans.). MacLehose Press.
- Oller, D. K. (2004). Underpinnings for a theory of communicative evolution. In D. K. Oller & U. Griebel (Eds.), *Evolution of communication systems: A comparative approach* (pp. 49–65).
- Pleyer, M. & Hartmann, S. 2024. *Cognitive Linguistics and Language Evolution* (Cambridge Elements in Cognitive Linguistics). Cambridge University Press.
- Pleyer, M., & Zhang, E. Q. (2022). Re-evaluating Hockett's design features from a cognitive and neuroscience perspective: The case of displacement. In A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, Y. Jadoul, M. Martins, Y. Oseki, E. D. Rodrigues, O. Vasileva, & S. Waciewicz (Eds.), *Proceedings of the Joint Conference on Language Evolution (JCoLE)*. <https://doi.org/10.17617/2.3398549>
- Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. *PLoS Biology*, 16(8), e2006425.
- Waciewicz, S., Pleyer, M., Szczepańska, A., Ewa Poniewierska, A., & Żywicznyński, P. (2023). The representation of animal communication and language evolution in introductory linguistics textbooks. *Journal of Language Evolution*, lzac010. <https://doi.org/10.1093/jole/lzac010>
- Watson, S. K., Filippi, P., Gasparri, L., Falk, N., Tamer, N., Widmer, P., Manser, M., & Glock, H.-J. (2022). Optionality in animal communication: A novel framework for examining the evolution of arbitrariness. *Biological Reviews*, 97(6), 2057–2075. <https://doi.org/10.1111/brv.12882>
- Zhang, E. Q., & Pleyer, M. (2024). Toward Interdisciplinary Integration in the Study of Comparative Cognition: Insights from Studying the Evolution of Multimodal Communication. *Comparative Cognition & Behavior Reviews*, 19, 85–90. <https://doi.org/10.3819/CCBR.2024.190017>

# Construing and identifying referentiality in interspecies languaging

Rea Peltola

(University of Caen Normandy, CRISCO UR4255)

Keywords: Human-Animal interaction, semantics, referentiality, indexicality, embodiment

This paper explores referentiality in interaction involving dogs and 6-12-year-old human children. The focus is on the vocal and bodily activities that the children treat as referring to something observable or otherwise shared in the situation. Ethological research has shown that animal vocalisations not only express emotional states but also convey context-specific indexical information. This functional referentiality has been observed across species also in situations involving a third party or an object (Faragó et al. 2010, Pongrácz et al. 2024).

In this study, the objective is not to measure the degree of referentiality in dog behaviour, as such, but to identify 1) what canine vocalisations and bodily cues are treated as referential, 2) which human language forms children use to respond to these, 3) in which types of interactional contexts referentiality is perceived, and 4) how human children tailor iconic and indexical referential relationships in their own speech to adapt to the dog's assumed communicative properties.

The data consist of 15 video recordings presenting child-dog dyads in ecological French- and Finnish-speaking context (Grandgeorge et al. 2020, Peltola et al. 2025). The post-anthropocentric (Braidotti 2024) theoretical framework combines biosemiotics (Uexküll 2011, Tønnessen 2022) and cognitive linguistic approaches to meaning construal, embodiment, and perspective-taking (Johnson 2018, Tomasello 2019).

The study shows that children highlight the perceived referential potential of dog sounds by incorporating imitated vocalisations into semantically complex constructions. In *non pas rhoo rhaa* 'no not rhoo rhaa', which is a response to the dog's repeated growl in a ball game, the child imitates the canine sound (Harjunpää 2022, Mondémé 2022), and treats it as carrying a local referential meaning recognisable for both playing partners, namely announcing the dog's taking possession of the ball. As the child simultaneously kicks the ball, so that the dog cannot reach it, the event to which the growl is referentially connected is cancelled both linguistically and physically. When orienting to objects at the focus of interaction, such as toys, the dog's non-vocal behaviour also serves as a resource for proform referentiality, e.g. *no ni, oliha se siellä* 'well, it was there after all' uttered after the dog has found the last hidden treat in an interactive toy. The dog's gaze is frequently incorporated into referential processes, e.g. *tu ne la vois même pas* 'you don't even see it' or *pas ça, ça* 'not that one, this one' are responses to dog's looking steadily at the hidden toy location or changing the gaze direction from one human hand to another. Simultaneously with the verbal utterances, the children use non-lexical sounds, (empty) hands and body posture as indexical signs, thereby embracing the multimodal dimension of dog-human interaction.

The paper brings to the fore meaning construction in interspecies interactions. Excluding semantics contributes to maintaining the status of human language as a system apart that cannot be analysed on the same grounds as others. The notion of languaging opens up the boundaries and places human languages among other modes of interaction, including those used by other species (Demuro & Gurney 2023)

## References

- Braidotti, R. (2024 [2019]), *Tieto ihmisen jälkeen*, transl. by K. Kortekallio, Tampere: niin&näin.
- Demuro, E., Gurney, L. (2023), Can nonhumans speak? Linguaging and worlds in posthumanist applied linguistics, *Linguistics Frontiers* 6, 92–105.
- Faragó, T., Pongrácz, P., Range, F., Virányi, Z., Miklósi, Á. (2010), ‘The bone is mine’: affective and referential aspects of dog growls, *Animal Behaviour* 79, 917–925.
- Grandgeorge, M., Gautier, Y., Bourreau, Y., Mossu, H., Hausberger, M. (2020), Visual attention patterns differ in dog vs. cat interactions with children with typical development or autism spectrum disorders, *Frontiers in Psychology* 11, 2047.
- Harjunpää, K. (2022), Repetition and prosodic matching in responding to pets’ vocalizations, *Langage et société* 176, 69–102.
- Johnson, M. (2018), The embodiment of language, in Newen, A. & De Bruin, L. & Gallagher, S. (eds.), *The Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press, 623–639.
- Mondémé, C. (2022), Quand les animaux participent à l’interaction sociale. Nouveaux regards sur l’analyse séquentielle, *Langage et société* 176, 9–24.
- Peltola, R., Wu, Y., Grandgeorge, M. (2025), Interpersonal distance, mouth sounds, and referentiality in child-dog play: A pluridisciplinary approach, *Language & Communication* 103, 126–148.
- Pongrácz, P., Dobos, P., Zsilák, B., Faragó, T., Ferdinandy, B. (2024), ‘Beware, I am large and dangerous’ – human listeners can be deceived by dynamic manipulation of the indexical content of agonistic dog growls, *Behavioral Ecology and Sociobiology* 78, 37.
- Tomasello, M. (2019), *Becoming Human: A Theory of Ontogeny*, Cambridge, Mass.: The Belknap Press of Harvard University Press.
- Tønnessen, M. (2022), The evolutionary origin(s) of the Umwelt. *Biosemiotics* 15, 451–455.
- Uexküll, J. von. (2011 [1982]). The Theory of Meaning, in Maran, T. & Martinelli, D. & Turovsk, A. (eds.), *Readings in Zoosemiotics*, Berlin: De Gruyter, 61–76.

# A Data-Driven Approach to Estimating Animal Vocal Repertoires and their Usage

Yuhang Wang & Richard Hahnloser

(Institute of Neuroinformatics UZH / ETH Zurich )

Keywords: < vocal repertoire, vocal structure, comparative bioacoustic, acoustic signal processing, animal vocalization>

Animal vocal sequences including human speech are organized into hierarchical sequences composed of voice primitives. In human speech, these primitives are known as phonemes or phones. By contrast, the corresponding vocal primitives in animals have been poorly studied.

In birdsong, the roughly hundred-millisecond long song syllables are often considered basic vocal units, since syllables display little acoustic variability within types but large differences across types (Sainburg, Thielk & Gentner, 2020). However, neurophysiology suggests that vocal units should be sought at a much finer 10-ms resolution imposed by the pace of neural dynamics (Leonardo & Fee, 2005). We propose a simple data-driven approach to estimate song repertoires.

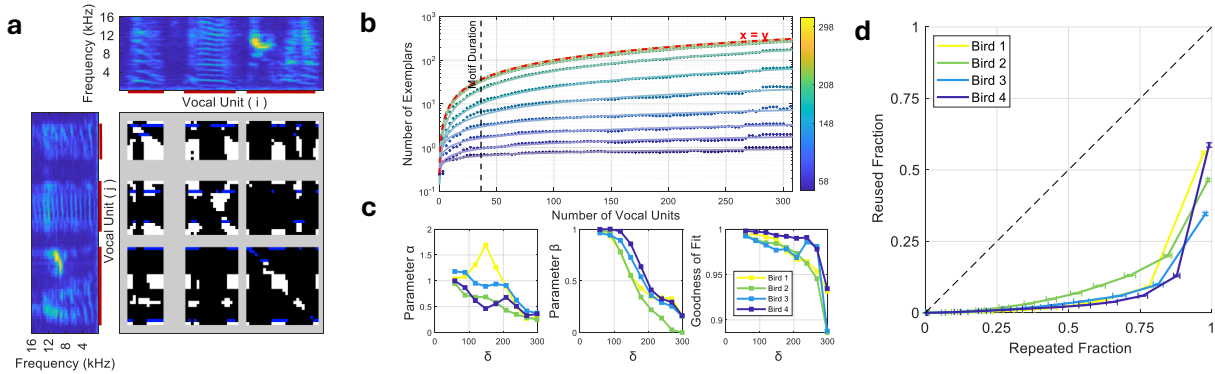


Figure 1: (a) Binarized distance matrix between vocal units  $i$  and  $j$  in a zebra finch song segment (black: distance  $d_{ij} > \delta$ ; white:  $d_{ij} \leq \delta$ ;  $\delta = 200$ ). The vocal units (song spectrogram) are shown on the left and on top (syllable extent is indicated by red lines). The blue pixels indicate the exemplars (left) and the vocal units they approximate (top). (b) Relationship between the number of vocal units in a song segment (x-axis), and the repertoire size (y-axis) for diverse thresholds  $\delta$  (dots) and the fits  $y = \alpha n^\beta$  (curves) for diverse  $\delta$  (color scale). (c) The dependence of parameters  $\alpha$  and  $\beta$  on  $\delta$  (left, middle) and the goodness of fit (right). (d) The fraction of repeated versus the fraction of reused vocal units in pairs of song motifs. The curves ( $N=4$  birds) are obtained by varying  $\delta$  (mean and error bars; 150 motifs per bird).

We segment birdsong into vocal units of 16-ms time window and identify dominating sets, which we refer to as exemplars, such that any vocal unit is no further away from an exemplar than some maximal threshold distance  $\delta$ . We use a greedy approximate algorithm to find the minimum set of exemplars — the repertoire, Figure. 1a. By increasing the song segment that we subject to this analysis (letting it contain  $n$  vocal units), we approximately need  $y = \alpha n^\beta$  exemplars, in reminiscence of Heap’s law, Figure 1b, c.

Our analysis allows us not only to identify repeated vocal units that birds consistently produce at a given time point across song motifs but also reused vocal units produced at diverse times within a motif, e.g., as part of different syllables. Surprisingly, we find that for the smallest  $\delta$  at which every vocal unit is repeated, on average 50% (range 35-65%,  $N=4$  birds) of all vocal units are excessively reused, Figure. 1d. Thus, zebra finches frequently reuse sounds across diverse time points in a motif which agrees with findings that some neural patterns repeat across song motifs not only at the same time, but also at other times when vocal output is similar (Vyssotski et al., 2016).

We propose a simple algorithm for identifying the repertoire of any vocal signal, revealing new insights into zebra finch song. The generality of our procedure enables its application to other birds’ songs and to mammalian vocalizations (including human speech), illustrating how easily communication signals can be hierarchically decomposed into comparable repertoires.

## References

- Sainburg, T., Thielk, M. & Gentner, T. Q. Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Comput. Biol.* 16, e1008228 (2020).
- Leonardo, A. & Fee, M. S. Ensemble coding of vocal control in birdsong. *J. Neurosci. Off. J. Soc. Neurosci.* 25, 652–661 (2005).
- Vyssotski, A. L., Stepien, A. E., Keller, G. B. & Hahnloser, R. H. R. A Neural Code That Is Isometric to Vocal Output and Correlates with Its Sensory Consequences. *PLOS Biol.* 14, e2000317 (2016).

**WS8 Insights into lexical iconicity: onomatopoeia,  
ideophones, and sound symbolism**

**Maria Flaksman & Chris Smith**

# Onomatopoeias and ideophones in Bulgarian: exploring criteria for differentiation

Alexandra Bagasheva

(Sofia University “St. Kliment Ohridski”)

Key words: onomatopoeias, ideophones, Bulgarian, differential properties, word formation

Ideophones are considered a rarity in Bulgarian and are classified either as verbal interjections (Hauge 1995, Maldjieva 2021) or as onomatopoeias (Tilkov, Stoyanov and Popov 1983). In the arch-class of interjections, Maldjieva (2021) distinguishes 18 classes with distinct syntactic properties, among which onomatopoeias (e.g., *myau* ‘meow’), predicative interjections (e.g., *drân-drân* ‘incessant blabbing’) and adverbs (e.g., *kucuk-kucuk* ‘describing limping movement’). The latter two seem to fall under the concept of *ideophone*. Starting from a very comprehensive definition of ideophones as “marked words that evoke vivid sensory scenes in imitative fashion” (Akita and Dingemanse 2019, 1), in the proposed talk, the possibilities for an informed demarcation between traditionally recognized onomatopoeias in Bulgarian (such as *drus-drus* ‘a series of vertical shakes’, *zââârrrr* ‘the sound of a bell or a traditional ring-tone’, *fraaasss* ‘a bang-like sound’), a possible class of ideophones (as suggested by Kovatcheva 2014, 1, e.g., *bliz* ‘lick’, *blâs* ‘push’, *brâk* ‘insert hand’) and verbal interjections (such as e.g., *bež* ‘sudden dashing away’, which can be used as an imperative) are explored.

Assuming there is a prototypical, if not a canonical, space that captures the class of “ideophones as conventionalized depictions” (Akita and Dingemanse 2019, 1), including the following properties: i) marked wordhood status – deviance from typical lexical items in the language in terms of phonology, length, heightened role of reduplication, etc.; ii) vague semantics since “[i]deophones are not like normal words to which meanings are readily assigned. They are simply sounds used in conveying a vivid impression” (Okpewho 1992, 92) and iii) a predicate function “qualificative or adverb in respect to manner, colour, sound, smell, action, state or intensity” (Doke 1935, 119), in the talk answers to the following research questions are sought: Is there a distinct class of ideophones in Bulgarian? Are both onomatopoeias and ideophones (if any) types of interjections and what are the possible differentiating properties among these?

The central dimension of differentiation hypothesized is their respective word formation status. Onomatopoeias are directly sound imitative and, apart from reduplication motivated euphonicly, no other word formation pattern can be associated with their origin, as they are simplex words. The second set of data comprises ideophones (identified as deverbal ones in Kovatcheva 2014), verb-based and necessarily involving back-formation or paradigmatic word formation (Haspelmath and Sims 2010; Plag 2018; Booij 2008). The two sets of data share imitativeness as a distinguishing property from all other interjections – sound imitatives (25 items) and other imitatives (25 items), extracted from the Corpus of Spoken Bulgarian, BED, Kovatcheva (2014) and Maldjieva (2021).

The preliminary conclusion is that *mimetics* is better suited as a cover term, wherein *onomatopoeias* as sound symbolic imitatives, resulting from imitative coinage (Flask 2022) and *ideophones* (derived by affix removal or reanalysis) as image-depictives are recognized as similar but distinct



subclasses, which coheres with Körtvélyessy's idea of onomatopoeia as a unique species (Körtvélyessy 2020). In view of a shared degree of iconicity (albeit monomodal vs. multimodal) and depictive properties, both should be set aside from the class of interjections in the language.

## References

- Akita, Kimi and Mark Dingemanse (2019), Ideophones (Mimetics, Expressives), *Oxford Research Encyclopaedia of Linguistics* 2019 | 10.1093/acrefore/9780199384655.013.477.
- Booij, Geert (2008), Paradigmatic morphology, in B. Fradin, (ed), (2008), *La Raison Morphologique: Hommage à la Mémoire de Danielle Corbin*, Amsterdam/Philadelphia: John Benjamins, 29–37.
- Doke, Clement (1935), *Bantu Linguistic Terminology*. London: Longmans, Green, and Co.
- Flask, Maria. 2022. Universality of onomatopoeic words decreases with de-iconization. Paper delivered at *Word-Formation Theories VI & Typology and Universals in Word-Formation V*, 23 - 26 June 2022, Košice, Slovakia.
- Haspelmath, Martin and Andrea Sims (2010), *Understanding Morphology*, London: Hodder and Stoughton.
- Hauge, Kjetil Rå (1995), Kâm klasifikacijata na tât narečenite glagolni meždumetiya v bâlgarski [Towards the classification of the so-called verbal interjections in Bulgarian], *Problems of Sociolinguistics* 4, 9–12.
- Körtvélyessy, Livia (2020), Onomatopoeia – A Unique Species?, *Studia Linguistica* 74(2), 506–551. <https://doi.org/10.1111/stul.12133>.
- Kovatcheva, Mira (2014), An Unrecognized class of words in Bulgarian and their word-formation, *Rivista di Linguistica* 26(2), 183–208.
- Maldjieva, Vyara (2021), Sintaktični strukturi realizirani ot mejdumetiya v bâlgarskiya ezik [Syntactic structures realized by interjections in Bulgarian], in J. L. Banasiak, A. Kiklewicz, J. Mazurkiewicz-Sułkowska (eds), (2021), *Języki słowiańskie dziś: w kręgu kategorii, struktur i procesów*, WUŁ–IS PAN, Łódź–Warszawa 2021, „Prace Slawistyczne. Slavica”, v. 150, 239–255.
- Okpewho, Isidore (1992), *African Oral Literature. Backgrounds, Character and Continuity*, Bloomington and Indianapolis: Indiana University Press.
- Plag, Ingo (2018), *Word-Formation in English*, Cambridge: Cambridge University Press.
- Tilkov, Dimitar, Stoyanov, Stoyan & Popov, Konstantin (eds) (1983), *Gramatika na sâvremenniya bâlgarski knižoven ezik: Vol. 2. Morfologiya* [Grammar of contemporary Bulgarian: Vol 2. Morphology], Izdatelstvo na Bâlgarskata akademiya na naukite [Publishing House of the Bulgarian Academy of Sciences].

## Primary sources

- BED: *Bâlgarski etimologičen rečnik* [Bulgarian etymological dictionary]. Vol. 1–7. Sofia: Bulgarian Academy of Sciences, 1971–2010.
- Corpus of Spoken Bulgarian. <http://metashare.ibl.bas.bg/repository/browse/corpus-of-spoken-bulgarian/5b64d54668b411e281b65cf3fcb88b70b06592a0147141109b0d88529dc6ec04/> (last accessed 5 November 2024).

# Phonesthemic sound symbolism in Formosan languages

Amy Pei-jung Lee

(National Dong Hwa University)

A phonestheme is defined as ‘a recurrent sound-meaning association that is not definable by contrast’. Structurally, a phonestheme is sub-morphemic which cannot occur independently. Semantically, a phonestheme has a representative meaning, yet its residue in the word is meaningless. Many phonesthemes in language are also onomatopoeic, which pertain to ideophones.

This paper is a follow-up study on the phonesthemes identified in Formosan languages based on Blust (1988, 2003a, 2003b) and Lee (2023), with both descriptive and cognitive approaches. Phonesthemes in Formosan languages can be categorized into two types. The first type is a monosyllabic root with a -CVC which only occurs at morpheme-final position. The second is a meaning-associated segment, either a single consonant or a pattern of consonants in successive syllables (cf. Blust 1988: 5).

The first type of phonesthemes can be further divided into those which are related to onomatopoeic ideophones and those which are not. This type exhibits the -CVC form, such as \*-tuk ‘knock, pound, beat’ or \*-pak ‘slap, clap’. To fulfil the requirement of a canonical word, a monosyllabic root of -CVC is reduplicated to form lexicalized reduplication, such as *tuktuk* ‘knock, pound’ (Kavalan) or *baqbaq* ‘slap, clap’ (Paiwan).

Those not related to onomatopoeic ideophones, on the other hand, exhibit associative sound symbolism (Johansson *et. al* 2020), of which the sound-meaning association can be accounted for from a cognitive perspective. For example, \*-pun ‘heap, pile, assemble, collect, gather’ is reflected in Formosan languages as ‘a place where rivers converge’, ‘a stomach’, ‘a well’, ‘a mound’, ‘a barn where grains are collected’, or ‘to meet up, to be together’, showing that a CONTAINMENT image-schema is at work. Similarly, \*-pit ‘press, squeeze together; narrow’ is reflected in Formosan languages with the following meanings: ‘chopsticks’, ‘eyelashes’, ‘to close eyes, to blink eyes, to wink’, ‘to heal a wound’, or ‘a button, to button one’s clothes, to sew a broken cloth’. The sound-meaning association not only demonstrates imitation of the vocal gestures, but also reveals a CONTACT image-schema.

The second type is manifested with the following phonesthemes: ŋ- ‘related to oral or nasal area’ (Blust 2003a), h-/f-/s- ‘air-related actions’, -r or -l ‘shaking, trembling’ (Lee 2023). The sound symbolism of these phonesthemes can be explicated by articulatory gestures as proposed by Thompson and Do (2019).

Keywords: phonestheme, sound symbolism, ideophone, articulatory gesture, Formosan language

## References

- Blust, Robert A. (1988), *Austronesian Root Theory: An Essay on the Limits of Morphology*. John Benjamins Publishing.
- Blust, Robert A. (2003a), The phonestheme ŋ- in Austronesian languages. *Oceanic Linguistics* 42.1: 187–212.
- Blust, Robert A. (2003b), A Note on Monosyllabic Roots in Kavalan. *Oceanic Linguistics* 42.1: 239–243.
- Johansson, Niklas Erben, Andrey Anikin, Gerd Carling and Arthur Holmer. (2020), The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 2020; 24(2), 253–310.
- Lee, Amy Pei-jung. (2023), Phonesthemes in Formosan languages. Paper presented at The 32nd Annual Meeting of the Southeast Asian Linguistics Society (SEALS 2023), 16–18 May, Chiang Mai University, Chiang Mai, Thailand.
- Thompson, Arthur Lewis and Youngah Do. (2019), Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: a journal of general linguistics* 4(1: 72), 1–40. DOI: <https://doi.org/10.5334/gigl.872>

## Iconicity in use -an onomasiological approach to words of enthusiasm in English

This paper investigates iconic (onomatopoeic, echoic, expressive, ideophonic) words from an onomasiological, or conceptual, viewpoint. It aims at providing a fine-grained usage-driven perspective on iconicity.

We introduce a mix-method approach to the study of semantic changes in iconic words. Thus, we combine (1) methods of research of lexicology and (2) usage-driven analysis. The obtained data provide a more accurate and fine-grained understanding of semantic changes in iconic vocabulary.

Methods of lexicology are the traditional methods of semantic analyses based on dictionary definitions of iconic words. There are currently several data-driven psycholinguistic hypotheses regarding iconicity and usage of iconic words. Thus, Dingemanse (2012: 663) shows that iconic words are most likely to describe conceptual categories of sound, movement, visual impressions, inner feelings, and cognitive states. Winter et al (2023) have recently proposed that certain semantic traits and certain categories of words correlate with higher iconicity ratings. According to their study, the following categories of words score highest in iconicity: (1) sensory words, (2) early acquired words, (3) words occupying spare semantic neighbourhoods, (4) words which are structurally marked and ‘playful’, and (5) (in English specifically) words belonging to the grammatical classes of interjections and verbs.

**Material and methods:** In this talk, following our preliminary study (Author 2024) on the diachronic evolution of English nouns depicting the idea of ‘enthusiasm, energy, or vitality’, **we extend the study to the class of verbs of enthusiasm with a view to comparing the results.** We use the *HTOED* and the *OED* dictionaries and combine these data with the analysis of their collocational behaviour in several diachronic and contemporary corpora (OEC, EHBC and COHA).

Initial **results** indicate that the iconicity cycle (Flaksman 2020) is tied to the general phenomenon of competition and regulation within the lexicogrammatical continuum: there is a balance between (1) innovation, creativity, and expressivity on the one hand, and (2) economy, stability, and convention on the other (cf. Goldberg 2019). We conclude that describing and modelling the processes of semantic change in onomatopoeia and related phenomena shall be based on the analysis of real-life interactions where iconic words occur (that is, an emphasis shall be put on exploring the usage-driven data).

Thus, what is perceived as iconicity (cf. Winter et al 2023) appears at the crossroads of (1) objective iconic form-meaning configurations (acoustic qualities and articulatory properties of speech sounds comprising iconic words), (2) conventionalised meanings of iconic words (which (can) form phraseological units and collocations), and (3) their contextualised interpretation (mapping).

## References

- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Davies Mark. (2010) *The Corpus of Historical American English* (COHA). <https://www.english-corpora.org/coha/>.
- Flaksman, Maria. 2020. Pathways of De-iconization: How Borrowing, Semantic Evolution, and Sound Change Obscure Iconicity. In Pamela Perniss, Olga Fischer & Christina Ljungberg (eds.), *Operationalizing Iconicity*. Amsterdam/Philadelphia: John Benjamins. 75-104.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Author 2024. “*Rah-rah!* Investigating the variation in phonosemantic motivation in a set of iconic nouns expressing the concept <enthusiasm energy vitality>. A diachronic semantic approach”, *Lexis* [Online],

23 | 2024, Online since 25 April 2024, connection on 11 November 2024. URL:  
<http://journals.openedition.org/lexis/7929>; DOI: <https://doi.org/10.4000/lexis.7929>

*Sketch Engine*, <http://www.sketchengine.eu>. <https://www.sketchengine.eu/>

*The Oxford English Dictionary*, OED3, <https://www.oed.com/>.

*The Historical Thesaurus of English* (2nd edition, version 5.0). 2024. University of Glasgow. <https://ht.ac.uk>.

*The Oxford English Corpus* OEC, via Sketch Engine. <https://www.sketchengine.eu/>

*The English Historical Book Collection*, via Sketch Engine. <https://www.sketchengine.eu/>

Winter, Bodo, Gary Lupyan, Lynn K. Perry, Mark Dingemanse, Marcus Perlman. 2023. Iconicity ratings for 14,000+ English words. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02112-6>

# Figurative verbs in Dolgan

Eugénie Stapert

(Kiel University)

Keywords: Figurative verbs, Dolgan language, Onomatopoeia, Turkic, Sound symbolism

This paper investigates figurative verbs in Dolgan, a Turkic language spoken in arctic Siberia (Russian Federation). Figurative verbs are known to form a special category of verbs in Sakha, Dolgan's closest genealogical relative, both in terms of their meaning and their phonological structure (Kharitonov 1954, Afanasyev 1993, Shamayeva & Prokopeva 2018, and Monastyrev et al. 2018). In this language, figurative verbs are characterised as verbs which may describe the shape of a human face, the shape, height and figure of a person, or the structure, form and appearance of a material object (Shamayeva & Prokopeva 2018). In addition, they may be onomatopoeic. In their study Shamayeva & Prokopeva (2018) show that in Sakha there is a correlation between the phonological structure of figurative verbs and their semantics. More specifically, figurative verbs with non-labialised vowels are associated with wide, flattened shapes, whereas figurative verbs with labialised vowels are associated with elongated or oval forms.

While figurative verbs have been investigated quite well for Sakha, a similar detailed investigation is still lacking for Dolgan (Däbritz 2022: 137). Figurative verbs have been mentioned for Dolgan by several scholars (c.f. Ubryatova 1985, Artemyev 2013, and Däbritz 2022), but an in-depth study of this category of verbs has not yet been possible due to a lack of data. The recent publication of a large corpus of Dolgan spontaneous texts and folklore stories (Däbritz et al. 2022), has improved this situation significantly. The Dolgan corpus consists of 14 hours of transcribed and glossed material and provides the necessary data to fill this gap in the research on Dolgan figurative verbs.

The current paper aims to establish whether the correlation between phonological structure and semantics, as proposed by Shamayeva & Prokopeva (2018) for Sakha, holds true for Dolgan as well. Taking the works by Shamayeva & Prokopeva (2018) and Monastyrev et al. (2018) as a basis, a list of onomatopoeic figurative verbs in Sakha will be compiled. These verbs will be compared with their equivalents in Dolgan for phonological structure as well as for semantics, using data from the corpus (Däbritz et al. 2022), as well as from dictionaries (Stachowski 1993, and Stachowski 1998). The combination of both oral and written data sources for Dolgan allows for a nuanced comparison, including speaker-internal phonetic variation as well as variation across speakers. Based on these data, it is investigated a) whether all figurative verbs in Sakha have an equivalent in Dolgan b) whether a correlation exists between phonological form and semantics, and c) whether the correlation matches the pattern found for Sakha, or that Dolgan displays a different pattern.

Finally, since a substantial part of figurative verbs in Sakha has been associated with Mongolian origins, a detailed analysis of this domain in Dolgan can provide additional insights into its history of contact.

References:

- Afanasyev, Lazar (1993), *Phonosemantics of figurative verbs of the Yakut language*. Dissertation. Yakutsk.
- Artemyev, Nikolay (2013), *Dolganskij jazyk. 10–11 klassy. Učebnoe posobie dlja obščeeobrazovatel'nyh učreždenij. Čast' 2. Morfologija*. [The Dolgan language. 10th to 11th grade. Textbook for secondary schools. Part 2. Morphology]. Saint Petersburg: Almaz-Graf.
- Däbritz, Chris Lasse (2022), *A grammar of Dolgan: A Northern Siberian Turkic language of the Taimyr Peninsula*. Leiden/Boston: Brill.
- Däbritz, Chris Lasse, Kudryakova, N., Stapert, E. (2022), *INEL Dolgan Corpus. Version 2.0*. Publication date 2022-11-30. <https://hdl.handle.net/11022/0000-0007-F9A7-4>. Archived at Universität Hamburg. In: The INEL corpora of indigenous Northern Eurasian languages. <https://hdl.handle.net/11022/0000-0007-F45A-1>.
- Kharitonov, Luka (1954), *Types of verbal stem in Yakut*, Moscow/Leningrad: Publishing House of the USSR Academy of Sciences.
- Monastyrev, Vladimir et al. (2018). On semanticization of phonemes in Yakut figurative verbs. *Turkish Online Journal of Design, Art and Communication*, special edition, 2641-2648.
- Shamayeva, Anastasia & Svetlana Prokopeva (2018), Concept analysis of Yakut figurative verbs with Mongolian parallels. *Turkish Online Journal of Design, Art & Communication* 8, 1778 – 1785.
- Stachowski, Marek (1993), *Dolganischer Wortschatz*, Zeszyty naukowe Uniwersytetu Jagiellonskiego. Kraków: Nakl. Uniewersytetu Jagiellonskiego.
- Stachowski, Marek (1998), *Dolganischer Wortschatz: supplementband*. Krakow: Nakl. Uniewersytetu Jagiellonskiego.
- Ubryatova, Elizaveta (1985), *Jazyk noril'skich dolgan [The language of the Norilsk Dolgans]*, Novosibirsk: Nauka.

# Iconic mappings facilitate spoken word production: Implications for lexical representation and processing

Giulio Massari, Fanny Meunier & Raphaël Fargier

Université Côte d’Azur, CNRS, BCL, France

**Keywords:** iconicity; sound symbolism; lexical processing; language production.

Despite standard assumptions of an arbitrary relationship between word forms and meanings, empirical research has shown the pervasive existence of iconic mappings in language. Such mappings, established through cross-modal (i.e., sound symbolic) or unimodal (i.e., visual) correspondences, have been systematically observed within the lexicon of oral and sign languages (Perniss & Vigliocco, 2014; Winter et al., 2024). For instance, in English, words referring to round objects were found to contain more round-articulated phonemes like /a, o, b, m/, while words referring to spiky objects to rather contain sharp-articulated phonemes like /e, i, k, t/ (Sidhu et al., 2021). Iconicity seems hence to imply a non-arbitrary connection between conceptual features of referents and formal features of lexemes. Therefore, processing consequences might be expected when words featuring iconic mappings are accessed for production and comprehension. However, while studies on sign language revealed the impact of iconicity upon lexical processing (Sehyr & Emmorey, 2021; Thompson et al., 2009), as well as its interaction with other lexical factors, such as word age of acquisition (Vinson et al., 2015), the implications of iconic mappings in the oral modality remain virtually unknown.

To clarify this, we first tested whether spoken production is facilitated when there is iconic congruence between conceptual and phonological representations of words. We designed a picture naming task, in which native French speakers (N=24) named images through words which were either iconically congruent to their image, i.e., their phonology symbolically matched the visual shape (round or spiky) of the depicted object (e.g., *ballon* – ‘balloon’), or iconically neutral, i.e., their phonology did not bias towards any shape (e.g., *dauphin* – ‘dolphin’). Words in each condition were further selected to be either early or late acquired (i.e., before/after 4.5 years: Fig. 1). Results showed that participants were faster to name pictures associated with iconic as compared to non-iconic words. However, this effect was only reliable among late acquired words, as consistent with observed interactions between iconicity and age of acquisition in sign naming (Fig. 2). By showing facilitation in the retrieval of words featuring

iconic phonological-conceptual links, our results support the hypothesis that iconicity impacts lexical processing in spoken language as well. Such results, in turn, carry insights about the cognitive processes underlying the retrieval of lexical representations, as accounted for by models of language production, and their potential differences for iconic and non-iconic mappings (Roelofs & Ferreira, 2019). To account for the observed facilitation, a non-modular conception of the different features in a word’s representational network (e.g., conceptual, lexical-semantic, phonological) seems to be required. Specifically, our results suggest that such features should be connected interactively, rather than accessed in a discrete sequence (Rapp & Goldrick, 2000). To better understand this mechanism, we explored whether facilitated retrieval of iconic mappings is due to faster activation of a word’s phonology based on its conceptual features. We designed a conceptual replication of the picture naming task described above, this time employing an EEG technique. Regression analyses on forthcoming data will reveal whether modulations of the brain’s electrical signal by psycholinguistic variables related to phonology (e.g., word length) or lexical semantics (e.g., animacy) display the same time course for iconic as compared to non-iconic words.









	Early acquired		Late acquired	
Iconic	 Ballon (‘balloon’)	 Carotte (‘carrot’)	 Bombe (‘bomb’)	 Équerre (‘set square’)
Non-iconic	 Tomate (‘tomato’)	 Renard (‘fox’)	 Dauphin (‘dolphin’)	 Antenne (‘antenna’)

Figure 1: experimental stimuli.

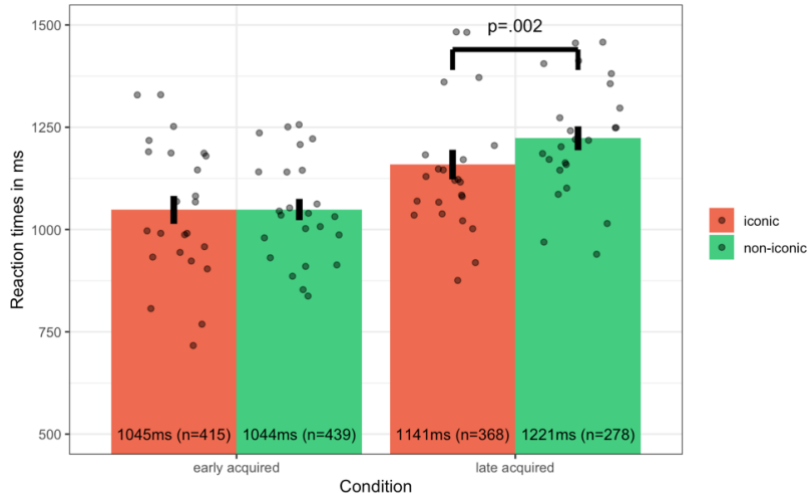


Figure 2: results from Exp. 1.



## References

- Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: To a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B*, 369: 20130300.
- Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human Language* (pp. 35–50). The MIT Press.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107(3), 460–499.
- Sehyr, Z.S., & Emmorey, K. (2021). The effects of multiple linguistic variables on picture naming in American Sign Language. *Behavior Research Methods*, 54, 2502–2521.
- Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, 28(4), 1390–1398.
- Thompson, R.L., Vinson, D.V., & Vigliocco, G. (2009). The link between form and meaning in American Sign Language: Lexical processing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 550–557.
- Vinson, D., Thompson, R.L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language*, 82, 56–85.
- Winter, B., Lupyan, G., Perry, L.K., Dingemanse, M., & Perlman, M. (2024). Iconicity ratings for 14,000+ words. *Behavior Research Method*, 56, 1640–1655.

# IDEOPHONIC WORDS IN THREE TORRICELLI LANGUAGES: SHARED AND DIFFERENTIAL FEATURES

Jose A. Jódar-Sánchez

Universitat de Barcelona / University at Buffalo

**Abstract:** The goal of this talk is to provide an overview of ideophonic words in three Torricelli languages, a family of severely endangered languages spoken in Papua New Guinea. The three languages in the study are Srenge (ISO: lsr), Walman (ISO: van), and Yeri (ISO: yev). Data comes from extensive fieldwork by various researchers at the University at Buffalo who have documented and described the languages in question. The methodology is based on elicitation of ideophonic words and as well as on analysis of text-based examples of ideophonic words and their contexts. First, I describe some of the common phonological features of ideophonic words in these three languages which set them apart from nonideophonic words. For instance, ideophonic words tend to begin with stops, have certain consonant clusters, and be reduplicated more often than nonideophonic words. Statistical evidence is offered for the special phonological status of ideophonic words. Second, I present some morphological and syntactic features of ideophonic words which are not shared by each of these languages. Third, I discuss the word class status of ideophonic words in each of these languages. Fourth, I survey the semantic fields covered by ideophonic words in these languages in relation to Dingemanse's (2012) implicational hierarchy of semantic areas expressed by ideophones crosslinguistically. Contextualized examples are provided throughout the presentation to illustrate all these features. A future goal of this research is to reconstruct and characterize ideophonic words in Proto-Torricelli and to create an elicitation tool to investigate them in other Torricelli (or Papuan) languages.

**Keywords:** ideophones, phonology, morphosyntax, semantics, Torricelli languages, Papua New Guinea.

## References

Dingemanse, M (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistic Compass*, 6(10), 654-672. DOI: 10.1002/lnc3.361.

# Phonemic Bigram Surprisal and the Decline of Expressivity in Iconic Lexicon

Maria Flaksman<sup>1</sup> Alexander Kilpatrick<sup>2</sup>

<sup>1</sup>University of Bamberg <sup>2</sup>University of Aizu

Imitative (iconic) words (ideophones, onomatopoeic words, etc.) directly convey impressions of sensory imagery (*ka-boom*, *crash*). Expressivity in imitative words manifests on a phonological level and is connected with *markedness* (use of rare speech sounds (*ugh* [əx]), phonotactic violations (*vroom*), expressive lengthening (*grrrr*)). This study explores the potential of *phonemic bigram Surprisal* as a markedness feature. Diachronically, expressivity declines as imitative words are integrated into the lexicon [4]. The *iconic treadmill hypothesis* [2] suggests that integrated imitative words *de-iconize* under the influence of sound changes (which reduce phonetic variability), which triggers new imitative coinage. The *aim* is to test ITH in the context of Surprisal.

**Procedure:** Previously [1], a large corpus (N=54 million lexemes) of spoken American English [3] was cross referenced with a pronouncing dictionary to obtain phonemic transcriptions. The number of phonemes in each word was totaled and used to calculate average Surprisal across bigrams. This was then cross-referenced with a dataset containing morpheme counts [5] and an iconicity dataset [6] where participants assigned Likert scores to 14,776 words according to how much each word “sounds like” its meaning. 13,294 words in the master dataset matched to the iconicity dataset, other samples were discarded (<https://tinyurl.com/m7nzjn3s>).

**Results:** A multiple linear regression model ( $F(3:13291) = 277.8, p < .001$ ) conducted to assess the impact of average Surprisal, phonemic length, and morpheme count on iconicity revealed that average surprisal ( $\beta = 0.0215, t = 4.211, p < .001$ ) and morpheme count ( $\beta = 0.0758, t = 5.386, p < .001$ ) have significant positive relationships with iconicity while phonemic length has a significant negative effect on iconicity ( $\beta = -0.1102, t = -24.706, p < .001$ ).

**Discussion:** These findings speak to the relationship between expressivity and communicative efficiency, see [7]. As words de-iconize, in accordance with the ITH, they become less expressive, manifesting as more predictable sound patterns and fewer morphemes. In particular, the data reveal that higher average Surprisal and the expression of a greater number of morphemes per phoneme are associated with higher iconicity, while greater phonemic length is associated with reduced iconicity.

## References

1. Kilpatrick, A., & Bundgaard-Nielsen, R. L. (2025). Exploring the dynamics of Shannon’s information and iconicity in language processing and lexeme evolution. *PLoS One*, 20(4), e0321294.
2. Flaksman, M. 2017. Iconic treadmill hypothesis: The reasons behind continuous onomatopoeic coinage. In A. Zirker, M. Bauer, O. Fischer, & C. Ljungberg (Eds.), *Dimensions of Iconicity* (pp. 15–38). John Benjamins Publishing Company.
3. Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior research methods*, 41(4), 977-990.
4. Dingemanse, Mark. 2017. “Expressiveness and System Integration: On the Typology of Ideophones, with Special Reference to Siwu”. *STUF. Language Typology and Universals* 70(2): 363–384.
5. Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior research methods*, 50, 1568-1580.
6. Winter, B., Lupyan, G., Perry, L. K., Dingemanse, M., & Perlman, M. 2023. Iconicity ratings for 14,000+ English words. *Behavior Research Methods*, 56(3), 1640-1655.
7. Levshina, N. (2022). *Communicative efficiency*. Cambridge: Cambridge University Press.

## Iconicity in the language of comics: A Swedish case study

Maria Wiktorsson  
Malmö University  
School of Arts and Communication

**Keywords:** iconicity, ideophones, comics language, sensory imagery, Swedish

Comics are multimodal forms of storytelling, where images and linguistic items combine to create lively narratives. They also offer rich material for investigating lexical iconicity. Expressive items in both dialogue bubbles and panels commonly add dimensions of sound (*bang*), motion (*swish*), sensory experiences (*yuk*), and inner feelings or cognitive states (*argh*, *gak*). Even though comics are silent and visual, readers' understanding of the sound dimension of written words affords iconic mappings that bring sound and emotion to the stories (Oyola 2011). Petersen (2007, 2009) describes how readers experience the sounds through a form of sub-vocalisation, triggering iconic mappings and creating a form of narrative erotics. Iconicity in comics also extends beyond properties of words and includes iconic aspects of font sizes and styles (Forster et al 2012) or shapes of dialogue bubbles (Forceville et al 2010).

Expressive Items in comics are often discussed as onomatopoeic expressions or interjections (Forster et al 2012; Guynes 2014). However, the full scope of relevant expressive items in comics cannot be covered by these categories. As the examples in the first paragraph illustrate, iconic mappings do not relate only to sound. Some recent studies have acknowledged this limitation and are instead treating the items as ideophones (e.g. Pischedda 2024). Ideophones display iconic mappings between form and meaning, including, but extending beyond, the sound dimension (Akita & Dingemanse 2019). Like ideophones, expressive items in comics depict sensory imagery not only of sound but also of movement, other sensory experiences, inner feelings and cognitive states (Dingemanse 2012).

This study aligns with recent work (Pischedda 2024) in treating expressive items in comics as ideophones but extends the research into new territory. To date, no detailed linguistic studies of expressive items in comics originally written in Swedish exist. This study addresses the gap by analyzing ideophones in the Swedish comic *Rocky* by Martin Kellerman: known for its realistic representation of spoken Swedish interaction (Fägersten 2019). Through manual analysis of the first *Rocky* album (Kellerman 1999), ideophones are identified and categorized by lexical status. English influence on both spoken Swedish and the language of comics (Valero Garces 2014; Pischedda 2017), necessitates checking items against both Swedish and English dictionary sources. Items not found in the sources are categorized as non-lexical. These are further sub-categorized as ‘comics items’ (if found in comics items’ compilations like Gasca & Gubern 2008 and Taylor 2018) or ‘creative constructions’ if not.

Findings show that Rocky contains a high number of items depicting sensory imagery. While most items do have iconic sound mappings, they often represent feelings, reactions, or cognitive states through depicting human vocalizations. Many are fully lexicalized interjections in Swedish, often written in larger fonts or elongated (*ååååååååååååh*). Swedish lexical words dominate, but the material also includes English lexical words (*crunch*, *yikes*), comics items (*fwoom*, *blam*), and creative constructions (*frodobolf*, *uööözzz*). Creative constructions with related meanings can show form similarity, *guäck* ('disgust'), *guläck* ('negative taste experience'), and *uuääck* ('vomiting'), with possible influence from related lexicalized words *äckel* ('disgust') and *äcklig* ('disgusting').

## References

- Akita, Kimi and Mark Dingemanse (2019), Ideophones (Mimetics, Expressives), in *Oxford Research Encyclopedia of Linguistics*, Oxford: Oxford University Press.
- Dingemanse, Mark (2012), Advances in the cross-linguistic study of ideophones, *Language and Linguistics Compass* 6/10, 654–672.
- Forceville, Charles, Tony Veale and Kurt Feyaerts (2010), Balloonics: The visuals of balloons in comics, in J. Goggin, and D. Hassler-Forest (eds), (2010), *The rise and reason of comics and graphic literature: Critical essays on the form*, Jefferson NC: McFarland & Co, 56–73.
- Forster, Iris, Susanne R. Borgwaldt, and Martin Neef, (2012), Form follows function: Interjections and onomatopoeia in comics, *Writing systems research* 4.2, 122–139.
- Fägersten, Kristy B. (2019), Linguistics: Comics conversations as data in Swedish comic strips, in M. J. Smith, M. Brown and R. Duncan (eds), (2019), *More critical approaches to comics: Theories and methods*, New York: Routledge, 145–159.
- Gasca, Luis and Román Gubern (2008), *Diccionario de onomatopeyas del comic*, Cátedra.
- Guynes, Sean A. (2014), Four-color sound: A Peircean semiotics of comic book onomatopoeia, *Public Journal of Semiotics* 6.1, 58–72.
- Kellerman, Martin (1999), *Rocky*, vol. 1, Stockholm: Ordfront Galago.
- Oyola, Osvaldo (2011), This is not a sound: The treachery of sound in comic books, in *Sounding Out: The Sound Studies Blog*, Retrieved from <http://soundstudiesblog.com/2011/06/>
- Petersen, Robert S. (2007), The acoustics of manga: Narrative erotics and the usual presence of sounds, *International Journal of Comic Art* 9(1), 578–90.
- Petersen, Robert S. (2009), The acoustics of manga, in J. Heer and K. Worcester (eds), (2009), *A Comics Studies Reader*, Jackson, MS: University Press of Mississippi, 163–171.
- Pischedda, Pier Simone (2017), Anglophonic influence in the use of sound symbolism in Italian Disney comics: A corpus-based analysis, *Open Linguistics* 3.1, 591–612.
- Pischedda, Pier Simone (2024), The sound of the Italian comic book: Representing noises, senses, and emotions across 80 years, *Open Linguistics*, 10(1).
- Taylor, Kevin (2018), *Ka-Boom!: A Dictionary of Comic Book Words, Symbols & Onomatopoeia*, Kevin J Taylor.
- Valero Garces, Carmen (2014), Onomatopoeia and Unarticulated Language in the translation and production of comic books, in F. Zanettin (ed), (2014), *Comics in Translation*, 237–255.

# Beautiful and hideous words: The effect of cross-linguistic phonesthetics and iconicity on word formation

Niklas Erben Johansson & Andrey Anikin  
(Lund University)

Keywords: iconicity, phonesthetics, typology, tone, word formation

Speech sounds enable us to produce spoken language through double articulation, but they can also carry intrinsic iconic associations to specific meanings, thereby speeding up linguistic transmission. For instance, high and/or rising intonation is often perceived as smaller and more positive than low and/or falling intonation. Similar iconic effects are found across a wide range of semantic domains, influencing words throughout the lexicons of individual languages (Blasi et al. 2016; Erben Johansson et al. 2020; Monaghan and Fletcher 2019; Sidhu et al. 2021; Winter and Perlman 2021). This enables iconicity to be used as a crucial strategy for meaning-making and word formation. At the same time, language users frequently perceive certain sounds or languages as more beautiful than others, and specific speech sounds can elicit strong affective arousal (Aryani et al. 2018, 2020), suggesting a potential link between phonesthetic and iconic associations. How, then, might phonesthetics shape the words we use?

We investigated whether there are universal phonesthetic judgments regarding the sound of languages. This study included 2,125 recordings from 228 languages across 43 language families, evaluated by 820 native speakers of English, Chinese, or Semitic languages, who rated how much they liked the sound of each language. The results showed that recordings of languages perceived as familiar, even when misidentified, and breathy female voices were rated as more pleasant. However, there was little consensus among raters about which languages sounded most beautiful, as personal preferences and perceived resemblance to culturally branded “beautiful” or “ugly” languages significantly influenced judgments. Despite this variability, some population-level phonesthetic preferences were observed. The clearest preference was for non-tonal languages, a trend most pronounced among Chinese-speaking participants—the only tonal participant language in the study. This raises an intriguing question: While lexical tones may be perceived as less pleasing, they are integral to the phonological systems of many languages. How, then, are tones used in tonal languages to convey words with positive or pleasing meanings versus negative or displeasing ones?

In a work-in-progress follow-up study, we collected basic vocabulary items with various positive and negative meanings (e.g., beautiful-ugly, good-bad, correct-wrong, sweet-bitter, happy-sad) from tonal languages representing over 50 language families. The study aims to investigate whether: (a) words with positive valence tend to feature higher and/or more rising tones compared to words with negative valence, aligning with iconic patterns; and (b) words with negative valence exhibit greater tonal complexity, such as more variable pitch contours and tonal contrasts, than words with positive valence. We also explore how tone types can be categorized and quantified to assess how underlying factors influence word formation.

Taken together, this presentation examines how our perception of speech, at both the segmental and language levels, facilitates the creation of shortcuts in communication. It also seeks to define the similarities and differences between phonesthetic and iconic associations on a cross-linguistic scale.

## References

- Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why “piss” is ruder than “pee”? The role of sound in affective meaning making. *PLOS ONE*, 13(6), e0198430. <https://doi.org/10.1371/journal.pone.0198430>
- Aryani, A., Isbilen, E. S., & Christiansen, M. H. (2020). Affective Arousal Links Sound to Meaning. *Psychological Science*, 31(8), 978–986. <https://doi.org/10.1177/0956797620927967>
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, 113(39), 10818–10823. <https://doi.org/10.1073/pnas.1605782113>
- Erben Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, 24(2), 253–310. <https://doi.org/10.1515/lingty-2020-2034>
- Monaghan, P., & Fletcher, M. (2019). Do sound symbolism effects for written words relate to individual phonemes or to phoneme features? *Language and Cognition*, 11(2), 235–255. <https://doi.org/10.1017/langcog.2019.20>
- Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (2021). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*, 28(4), 1390–1398. <https://doi.org/10.3758/s13423-021-01883-3>
- Winter, B., & Perlman, M. (2021). Size sound symbolism in the English lexicon. *Glossa: A Journal of General Linguistics*, 6(1), Article 1. <https://doi.org/10.5334/gjgl.1646>

# Ideophones in signed languages? Mapping ‘depicting constructions’ onto the canonical definition of ideophones

Sébastien Vandenitte  
(University of Jyväskylä)

Keywords: iconicity, depiction, signed languages, typology, ideophones

Language and communication researchers have become increasingly interested in depictive meaning-making and its manifestations across diverse signed and spoken languages. Recent endeavours have relied on the operationalisation of modality-agnostic comparative concepts and canons for such comparisons (Haspelmath, 2010; Brown & Chumakina, 2013), i.e., concepts designed for comparative purposes that are not tied to a specific channel or modality (e.g., Hodge et al., 2023). Dingemanse (2019, p. 16) proposes the concept of ideophone as ‘*a member of an open lexical class of marked words that depict sensory imagery*’. Following this definition, Dingemanse qualifies previous ideophonic analyses of several phenomena in signed languages, e.g., Bergman & Dahl (1994), Ajello et al. (2001), Hogue (2011), and Lu & Goldin-Meadow (2018). In this presentation, I discuss research on the use of specific depicting constructions in signed languages to re-examine Dingemanse’s claim that there are no ideophones in signed languages. Illustrated with entries of LSFB (French Belgian Sign Language) lexical databases and utterances retrieved from the LSFB Corpus (Meurant, 2015), the discussion will start from criteria that easily fit depicting constructions and gradually progresses towards criteria that are less obvious fits. As a starting point, I show how depicting constructions are used to express sensory experience. In particular, the depicting constructions under scrutiny prototypically consist in using the space in front of the signer to depict a referent’s position and motion on a reduced scale (Perniss, 2012; Ferrara & Halvorsen, 2017). I then discuss how this strategy can diachronically lead to the emergence of conventionalised form-meaning pairings that may still leave space for a strong depictive potential (Cormier et al., 2012; Johnston, 2012; Wilcox & Xavier, 2013; Ferrara & Halvorsen, 2017). Finally, I ask whether these conventionalised form-meaning pairings can be described as ‘marked’ in signed languages and whether they can be said to form an open lexical class (Schwager & Zeshan, 2008; Kimmelman & Börstell, 2023). Using the criteria listed in the definition, it is argued that lexicalised depicting constructions are closer to the concept of ideophone than suggested earlier. Building on the similarities, the presentation points to questions and interests spanning both signed and spoken language research and suggests some grounds for cross-pollination between these research traditions.

## References

- Ajello, R., Mazzoni, L., & Nicolai, F. (2001). Linguistic gestures: Mouthing in Italian sign language (LIS). In P. Boyes Braem & R. Sutton-Spence (Eds.), *The Hands Are the Head of the Mouth: The Mouth as Articulator in Sign Languages* (pp. 231–246). Signum-Verlag.
- Bergman, B., & Dahl, Ö. (1994). Ideophones in Sign Language? The place of reduplication in the tense-aspect system of Swedish Sign Language. In C. Bache, H. Basbøll, & C. E. Lindberg (Eds.), *Tense, Aspect, and Action: Empirical and Theoretical Contributions to Language Typology* (Vol. 12, pp. 397–422). De Gruyter.
- Brown, D., & Chumakina, M. (2013). What there might be and what there is: An introduction to Canonical Typology. In D. Brown, M. Chumakina, & G. Corbett (Eds.), *Canonical Morphology and Syntax* (pp. 1–19). Oxford Academic.
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., & Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language & Communication*, 32(4), 329–348. <https://doi.org/10.1016/j.langcom.2012.09.004>



- Dingemanse, M. (2019). Chapter 1. 'Ideophone' as a comparative concept. In K. Akita & P. Pardeshi (Eds.), *Iconicity in Language and Literature* (Vol. 16, pp. 13–33). John Benjamins Publishing Company. <https://doi.org/10.1075/ill.16.02din>
- Ferrara, L., & Halvorsen, R. P. (2017). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture*, 16(3), 371–395. <https://doi.org/10.1075/gest.00001.fer>
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3), 663–687.
- Hodge, G., Barth, D. & Reed, L. (2023). Auslan and Matukar Panau: a modality-agnostic look at quotatives. *Language Documentation & Conservation*, 12, 85–125.
- Hogue, R. L. (2011). *OOs and AAs: Mouth Gestures as Ideophones in American Sign Language*. [Doctoral dissertation]. Washington D.C.: Gallaudet University.
- Johnston, T. (2012). Lexical Frequency in Sign Languages. *Journal of Deaf Studies and Deaf Education*, 17(2), 163–193. <https://doi.org/10.1093/deafed/enr036>
- Kimmelman, V., & Börstell, C. (2023). Word classes in sign languages. In E. van Lier (Ed.), *The Oxford Handbook of Word Classes* (pp. 809–829). Oxford University Press.
- Lu, J. C., & Goldin-Meadow, S. (2018). Creating Images With the Stroke of a Hand: Depiction of Size and Shape in Sign Language. *Frontiers in Psychology*, 9, 1276. <https://doi.org/10.3389/fpsyg.2018.01276>
- Meurant, L. 2015. Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB). LSFB-Lab, University of Namur. URL: <http://www.corpus-lsfb.be>
- Perniss, P. (2012). 19. Use of sign space. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign Language* (Vol. 37, pp. 412–431). De Gruyter Mouton. <https://doi.org/10.1515/9783110261325.412>
- Schwager, W., & Zeshan, U. (2008). Word classes in sign languages: Criteria and classifications. *Studies in Language*, 32(3), 509–545.
- Wilcox, S., & Xavier, A. N. (2013). A framework for unifying spoken language, signed language, and gesture. *Revista Todas as Letras*, 15(1), 88–110.

# Arbitrariness in iconicity: A contrastive study of degrees of iconicity in sign languages & pictorial language

Saghie Sharifzadeh & Thomas Sähn  
(Sorbonne Université Paris)

Keywords: iconicity, arbitrariness, semiotics, sign languages, pictorial language

Iconicity, as understood in current applied research, represents a specific connection between the “form of a sign” and its “meaning”, which is generally analyzed in contrast to Saussure's understanding of signs as arbitrary, whether that connection is described as “natural”, “resembling”, “similar”, or “motivated” (e.g. Taub 2001: 8; Perniss et al. 2010: 1–2; Dingemanse et al. 2015: 604; Perlman et al. 2018: 1; ...). Some authors have argued that iconicity should rather be defined in terms of degrees (e.g. Caselli and Pyers 2020; Ortega 2017); others even pointed out that there can be no iconicity in language without arbitrariness (e.g. Wilcox 2004: 140ff.), as was implicit in the Saussurean analysis (1971/1916: 100–101). Through different semiotic approaches to iconicity (cf. Groupe  $\mu$  1992; Hjelmslev 1971/1948; Peirce 1995/1931–1958), highlighting qualitative and quantitative degrees of iconicity in languages, this paper aims to demonstrate that ‘arbitrariness’ and ‘iconicity’ constitute the two ideal poles of a continuum on which the units of a specific language can be placed according to those degrees. Based on a categorization of the various connections between the signifying units of sign languages and what they denote, this study will compare sign languages, French Sign Language (FSL) in particular, with written/oral languages on the one hand and pictorial languages on the other.

We will first show how the codified units of FSL can be divided into smaller units, either ‘discrete’ or ‘distinctive’, and reveal the role that arbitrariness and iconicity play in articulating these units into signifying ones. In a second step, we will then highlight various types of iconic connections and show that signs can share common features with extralinguistic, intralinguistic or interlinguistic ‘objects’ (in the Peircian sense) and, through additional symbolic, indexical or iconic connections, even refer to other objects in ways comparable to Napoli (2019)'s “chains”. The fundamental distinctions brought to light in these first two sections will result in an initial classification of the lexicon of FSL but personal and situational transfers, which are characteristic of sign languages (e. g. Cuxac 2000, 31–95), also enable the formation of signifying units which are not codified by any “normative authority” (Ammon 1995: 80ff.). In a final section, devoted to a qualitative study of narratives in sign languages and pictorial languages, we will therefore demonstrate that signers consciously or unconsciously resort to the distinctive units which are (proto)typical of a given object in a way that is shared with pictorial languages.

## References

- Ammon, Ulrich (1995), *Die deutsche Sprache in Deutschland, Österreich und der Schweiz: Das Problem der nationalen Varietäten*, Berlin/New York: De Gruyter.
- Caselli, Naomi and Pyers, Jennie (2020), Degree and Not Type of Iconicity Affects Sign Language Vocabulary Acquisition, *Journal of Experimental Psychology: Learning, Memory and Cognition* 46(1), 127–139.

- Cuxac, Christian (2000), *La Langue des Signes Française: Les Voies de l'Iconicité*, *Faits de Langues* 15–16, Paris: Ophrys.
- Dingemanse, Mark; Blasi, Damián E.; Lupyan, Gary; Christiansen, Morten and Monaghan, Padraic (2015), Arbitrariness, Iconicity, and Systematicity in Language, *Trends in Cognitive Sciences* 19 (10), 603–614.
- Groupe µ. (1992), *Traité du signe visuel: Pour une rhétorique de l'image*, Paris: Le Seuil.
- Hjelmslev, Louis (1971), La Structure fondamentale du langage [transl. from Danish by Una Canger], in Hjelmslev, L. (1971), *Prolégomènes à une théorie du langage* [1948], Paris: Éditions de Minuit, 173–227.
- Napoli, Donna Jo (2017), Iconicity Chains in Sign Languages, *On Looking into Words (And Beyond): Structures, Relations, Analyses* 3, 517–546.
- Ortega, Gerardo (2017), Iconicity and Sign Lexical Acquisition: A Review, *Frontiers in Psychology* 8, 1–14.
- Peirce, Charles Sanders (1995), *Collected Papers of Charles Sanders Peirce* [1931–1958], Cambridge (Mass.): Harvard University Press.
- Perlman, Marcus; Little, Hannah; Thompson, Bill and Thompson, Robin (2018), Iconicity in Signed and Spoken Vocabulary: A Comparison between American Sign Language, British Sign Language, English, and Spanish, *Frontiers in Psychology* 9, 1–16.
- Perniss, Pamela; Thompson, Robin L. and Vigliocco, Gabriella (2010), Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages, *Frontiers in Psychology* 1, 1–15.
- Saussure, Ferdinand de, (1971), *Cours de linguistique générale* [1916], Paris: Édition Payot & Rivages.
- Taub, Sarah (2001), *Language from the Body: Iconicity and Metaphor in American Sign Language*, Cambridge, UK: Cambridge University Press.
- Wilcox, Sherman (2004), Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed language, *Cognitive Linguistics* 15(2), 119–147.

# Iconic lexicons: What’s actually depicted in them?

Thomas Van Hoey  
(FWO; KU Leuven, Leuven)

Keywords: iconicity, lexical typology, ideophone, depiction, language comparison

In the past fifteen years or so, the interest in lexical iconicity, i.e., ideophones, expressives, mimetics, onomatopoeias etc. has been on the rise. Examples include Japanese *pikapika* ‘flashing’, Siwu *ɖabodɖabo* ‘gait of a duck walking’ or Pastaza Kichwa *polang* ‘manatee jumping up from the water surface’. Such items can be cross-linguistically defined as “*marked words that depict sensory imagery and belong to an open lexical class*” (Dingemanse 2019). At an *abstract level*, these words often depict senses like: sound, motion, color, dimension, configuration, textures, temperatures, pain, smell, taste, inner feelings, cognitive states, time, evaluations and so on (Dingemanse 2012; Van Hoey 2023). It has been observed that the range of the iconic inventories differs across languages and language families.

But how similar are iconic inventories at a *more fine-grained level*? What kinds of *concepts* are expressible in an ideophonic manner? And to what degree do they resort to *iconic means* to express these different concepts, such as phonological structure mappings, morphological mechanisms?

To tackle such questions, this study introduces the Depicticon, a typological database containing the iconic lexicons of over 50 languages from 22 language families; see <https://osf.io/zrq56/> for the range of languages. Iconic items are semantically tagged for the concepts they depict, similar to the appendix of Kakehi, Tamori & Schourup’s (1996) dictionary for Japanese mimetics, and phonologically enriched with phonotactic information. Through the technique of onomasiological profiling (Speelman, Grondelaers & Geeraerts 2003) the similarity between inventories is then calculated and visualized in 2D and 3D plots.

This study makes use of the increasingly open and digitized data age we are living in and provides a first typological step towards large-scale comparison of iconic lexicons, while disambiguating universally depictable concepts and language-particular idiosyncratic depictions.

Funding: FWO grant 1209725N “Borrowing iconic words” is gratefully acknowledged.

## References

- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives* (Iconicity in Language and Literature, ILL 16), 13–33. Amsterdam: John Benjamins. 10.1075/ill.16.02din.
- Kakehi, Hisao, Ikuhiro Tamori & Lawrence C. (Lawrence Clifford) Schourup (eds.). 1996. *Dictionary of iconic expressions in Japanese* (Trends in Linguistics 12). Berlin ; New York: Mouton de Gruyter.
- Speelman, Dirk, Stefan Grondelaers & Dirk Geeraerts. 2003. Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37. 317–337.
- Van Hoey, Thomas. 2023. A semantic map for ideophones. In Thomas Fuyin Li (ed.), *Handbook of Cognitive Semantics: Vol 2*, 129–175. Leiden: Brill.

## Actual dynamics of Mongolian ideophonic vocabulary

This paper focuses on two interrelated phenomena of Khalkha Mongolian ideophones: (i) their semantic extension and shift and (ii) usage change conditioned by change in society and ways of living. We draw upon our database of Mongolian ideophones which currently contains ca. 700 ideophonic roots and over 3000 individual forms with examples from fieldwork, dictionaries, literature and internet sources. A large portion of our data comes from elicitation within fieldwork (43 speakers).

Mongolian ideophones consist of roots and suffixes deriving nominals and verbs (Ichinose 1991). The majority (76%) of roots pertain to the visual domain, describing shapes, visual effects, and their movements. Their original meanings are those describing animals, people, and shapes as they occur in the natural environment of Mongolian nomads (Oberfalzerová 2010).

Investigating (i), we show that ideophones are, further, used in extended meanings to describe (ridicule or praise) human exterior and behavior. Thus, the ideophone *danhalza*- ‘to move repeatedly of someone with big head’, is used to describe proud and pompous behavior. Specifically, in our data, several types of behavior (e.g. proud and arrogant behavior, conflict behavior, flirting) are conspicuously “crowded” with ideophones. We are interested in the relation of such groups of ideophones to their source domains.

Investigating (ii), we base our findings on fieldwork data gathered between 2022-2024 among rural and urban speakers (elicitation with 26 speakers born, and continually living in rural areas, and 17 speakers born and living in Ulaanbaatar). We observe several types of change connected to sociolinguistic factors: Loss of ideophones, loss of the original meanings connected to the rural life-style but preservation of derived meanings (e.g. *oli* ‘to but’ (rural language), ‘to snitch on sb.’ (urban language), and deideophonization. We present an analysis of sociolinguistic factors at play: (a) massive transfer of rural population into urban areas (Konagaya et al. 2013), (b) inclusion of ideophones into the standardized form of the language (Narantuya 2021) and into the curriculum of primary education, (c) weakening of the natural transmission. As an illustration, in a three-generational group of informants with rural background and children currently living in an urban center (July 2022, 8 members of one family in Övörkhangaï province), we observed strong command of ideophones by the oldest generation, while the middle generation has inherited a part by natural transmission. The third generation exhibited poor natural command of ideophones, but has acquired certain lexemes via the education system.

In a synthesis of (i) and (ii) we show that (a) the shifted meanings of ideophones describing social behavior are highly productive in certain spheres such as less standard journalism and comments to political events, (b) ideophones are used as part of scientific terminology e.g. *arzgar suhai* (literally ‘bristly tamarisk’) ‘*Tamarix hispida*’, (c) strong decline of the actual use of the ideophonic system is observed. We propose a general tendency for ideophonic systems to be stronger in less urbanised and more traditionally living communities.

## References:

- Ichinose, M. (1991). Symbolic Words describing shapes and states of things in Mongolian. *Gengo Kenkyu* (Journal of the Linguistic Society of Japan) 1991/ 99, 121– 132.
- Oberfalzerová, A. (2010). The use of iconopoeic words in spoken Mongolian. *Mongolo-Tibetica Pragensia* 3(1), 7–34.

Konagaya, Y., Maekawa, A., Characteristics and Transformation of the Pastoral System in Mongolia. 2013. In YAMAMURA, Norio; FUJITA, Noboru a MAEKAWA, Ai (eds.). *The Mongolian ecosystem network: Environmental issues under climate and social changes*. Ecological research monographs. Japan: Springer.

Narantuyaa, SH., 2021, *Dürsleh ügiin tailbar toli* (Dictionary of ideophones). Ulaanbaatar: Udam soyol.

## **WS9 Isomorphism and optionality in language**

**Benoît Leclercq & Cameron Morin**

## ***Allo- Allo-! Clearing up some confusion around a constructionist concept***

Bert Cappelle  
(Université de Lille)

Keywords: allostructions, allomorphs, isomorphism, optionality, Construction Grammar

For some time now, considerable weight has been given within Construction Grammar to the idea that two or more constructions at the same level of generality can be linked ‘horizontally’. These are then sometimes called *allostructions* – formally distinct but semantically similar daughters of a more schematic mother node or ‘constructeme’ (Cappelle, 2006; De Vaere et al., 2020; Minami, forthcoming; Perek, 2015; Ungerer, 2024; Zehentner, 2023). The view that constructions at the same level of grammatical abstraction may be seen as related to one another can be traced back to at least Lambrecht’s (1994, p. 6) use of *allosentence*, which was borrowed from Daneš (1966, p. 232). Despite being rooted in a long tradition, the term *allostruction*, just like the concept it refers to, is still not well delineated and understood. In this theoretically oriented talk, I will attempt to address two open questions. First, do allostructions have to share some aspect of form? Second, do allostructions support the principle of optionality (i.e., in a usage context, there are several alternatives to choose from) or, rather, the principle of isomorphism (i.e., one form corresponds to only one function and vice versa)?

As to whether two or more forms eligible for the status of ‘allostructionhood’ should have some portion of their form in common, the answer is yes. But this in itself is not a sufficient condition. Crucially, the allostruction candidates must also share a specific meaning not found in other forms. Thus, while the double object construction and the *to*-prepositional dative share little form except for abstract categories (a verb and two NPs, one of which is realized as the complement of a preposition in one of the alternants), they also share a crucial aspect of meaning (roughly, ‘transfer’). As this meaning is specific to these two constructions, they are allostructions. By contrast, *cat* and *rat* (to take a pair of lexical constructions) can hardly be considered allostructions. This is not because their shared form (*consonant+/æ/*) may be coincidental, as the two forms still share a more abstract form – they’re both nouns – but rather because their shared meaning (say, ‘small, furry animal’) is not exclusively tied to this pair of words. After all, we want to avoid treating all members of a broad lexical field, such as ‘common mammals found in or near one’s house’, as allostructions.

As for the seemingly competing principles of optionality and isomorphism, these play out differently for constructions differing in complexity. As suggested by Minami (forthcoming), who refers to Jakobson (1980, p. 74), speakers cannot choose freely between allomorphs in complementary distribution (e.g. *in-*, *il-*, *im-*, *ir-*), which are then functionally equivalent. Allostructions of phrasal/clausal constructions show more obvious isomorphism (subtle functional differences). Such more complex combinations also allow for greater *freedom* in Jakobson’s sense (namely, leeway in selecting lexical items for syntactic slots). Still, *optionality* – understood as free choice between allostructions – is restricted, though not excluded, by the lack of full functional equivalence among alternatives.



## References

- Cappelle, Bert (2006). Particle placement and the case for “allostructions.” *Constructions Online*, SV1-7, 1–28.
- Daneš, František (1966). A Three-Level Approach to Syntax. In J. Vachek (Ed.), *L'École de Prague d'aujourd'hui* (Vol. 1, pp. 225–240). University of Alabama Press.
- De Vaere, Hilde, Kolkman, Julia, & Bellig, Thomas (2020). Allostructions revisited. *Journal of Pragmatics*, 170, 96–111. <https://doi.org/10.1016/j.pragma.2020.08.016>
- Jakobson, Roman (1980). Part II: Two aspects of language and two types of aphasic disturbances. In *Fundamentals of Language* (pp. 67–96). De Gruyter Mouton.  
<https://www.degruyterbrill.com/document/doi/10.1515/9783110889611-003/html>
- Lambrecht, Knud (1994). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge University Press.
- Minami, Yusuke (forthcoming). Allostructions and constructemes in Construction Grammar. In *Elsevier Encyclopedia of Language and Linguistics* (3rd edition). Elsevier.
- Perek, Florent (2015). *Argument Structure in Usage-Based Construction Grammar: Experimental and Corpus-Based Perspectives*. John Benjamins.  
[https://books.google.be/books/about/Argument\\_Structure\\_in\\_Usage\\_Based\\_Constr.html?hl=nl&id=3x\\_IBwAAQBAJ](https://books.google.be/books/about/Argument_Structure_in_Usage_Based_Constr.html?hl=nl&id=3x_IBwAAQBAJ)
- Ungerer, Tobias (2024). Vertical and horizontal links in constructional networks: Two sides of the same coin? *Constructions and Frames*, 16(1), 30–63. <https://doi.org/10.1075/cf.22011.ung>
- Zehentner, Eva (2023). Allostructions re-revisited. *Constructions*, 15(1), Article 1.  
<https://doi.org/10.24338/cons-569>

# Future temporal reference and optionality in English: Distributional semantic insights across registers

Chiara Paolini, Ruixue Wang & Bingjie Zeng

(KU Leuven, Andover Shanghai)

The study of future temporal reference (FTR) in English (*will* versus *be going to*) has garnered significant attention in recent years, particularly in the context of its variation across different registers (Engel and Szmrecsanyi 2022), regional varieties (Bohmann 2024), and second-language learning (Dubois et al. 2024).

- (1) a. [The friends]<sub>subject</sub> *will* [meet]<sub>verb</sub> for a coffee  
b. [The friends]<sub>subject</sub> *are going to* [meet]<sub>verb</sub> for a coffee

Overall, these studies highlight the inherent complexities and variability of the future tenses usage, presenting a significant methodological challenge to predict how speakers will reference future events in different linguistic contexts. Focusing on register variation, Engel and Szmrecsanyi (2022), show how FTR usage of individual speakers is greatly conditioned by language-internal variables (e.g., negation favors *be going to* in written formal register, stative verbs favor *will* in spoken language, while *be going to* in written language), using variationist corpus-based techniques and psycholinguistic evidence. However, these studies do not account for register-based lexical variation as a language-internal factor in their prediction models, raising the question: to what extent does the semantics of the materials in the argument slots (subject, verb) play a role in determining speakers' choices in different situational contexts?

In the ongoing debate between no synonymy (Goldberg 1995; more recently, Leclercq and Morin 2023) and intra-speaker optionality (Ma, Hoey, and Szmrecsanyi 2025; Szmrecsanyi et al. In print;), this study aims to provide further empirical evidence on intra-speaker probabilistic variation (i.e., optionality) of FTR variants across registers by integrating the semantic properties of the lexical context into the classic variationist model to predict variant choice.

We re-examined two of the four register subsets of the FTR dataset (Engel and Szmrecsanyi 2022): the written-informal register based on the British section of the Global Web-based English (GloWbE; Davies 2013) and the spoken-informal register based on the British National Corpus 2014 (BNC 2014; Love et al. 2017). To assess the semantic features of the lexical context in a cognitive-solid and bottom-up way, we employed Vector-Space-Model (VSM) Semantic Predictors (Paolini et al. 2023, 2025). Using the *NephoSem* count-based, type-level vector space model (Lenci 2018; QLV 2021), the semantic predictors are automatically built by separately clustering the lemmas of the subject and verb argument slots (e.g., *friends* and *meet* in examples 1(a-b)), extracted from each FTR subset. Once evaluated, these are then combined with each dataset's original top-down and manually-annotated formal predictors and fitted into a binomial logistic regression model with mixed-effects to predict speakers' choice.

Analyses reveal a differential impact of lexical variation across the two registers. In the written-informal register, VSM-semantic predictors slightly enhance the overall predictive power of top-down predictors by mirroring and reinforcing significant variables such as sentence type, proximity, and subject person. In contrast, alternates in the spoken-informal register appear largely unaffected by lexical variation, as

evidenced by the absence of significant VSM-semantic predictors in the regression model. Overall, VSM-semantic predictors demonstrate strong potential in capturing usage-based lexical patterns, thus representing a valuable addition to the variationist toolkit.

From a more theoretical perspective, the present study demonstrates that the role of the lexical material in the FTR sentences can vary considerably depending on the register in which variant choice occurs. The results for the spoken register, in particular, seem to confirm the neutralization of subtle (lexical) semantic differences between variants in spoken contexts, as postulated by Sankoff (1988), ultimately supporting the probabilistic nature of intra-speaker variant choice – a dimension that linguistic isomorphism, in all its forms and definitions, fails to adequately capture.

## References

- Bohmann, Axel. 2024. "Future-Time Reference in World Englishes." *World Englishes* 43 (1): 2–22. <https://doi.org/10.1111/weng.12634>.
- Davies, Mark. 2013. "Corpus of Global Web-Based English: 1.9 Billion Words from Speakers in 20 Countries (GloWbE)." <https://www.english-corpora.org/glowbe/>.
- Dubois, Tanguy, Magali Paquot, and Benedikt Szmrecsanyi. 2024. "Future-Time Reference in Spoken EFL: More Complex than in ENL?," October. <https://doi.org/10.1075/jsls.00030.dub>.
- Engel, Alexandra, and Benedikt Szmrecsanyi. 2022. "Variable Grammars Are Variable across Registers: Future Temporal Reference in English." *Language Variation and Change* 34 (3): 355–78. <https://doi.org/10.1017/S0954394522000163>.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Leclercq, Benoît, and Cameron Morin. 2023. "No Equivalence: A New Principle of No Synonymy." *Constructions*, June, 2023: Special Issue "35 Years of Constructions" (Editors: Lotte Sommerer & Stefan Hartmann). <https://doi.org/10.24338/CONS-535>.
- Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4 (1): 151–71. <https://doi.org/10.1146/annurev-linguistics-030514-125254>.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina, and Tony McEnery. 2017. "The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations." *International Journal of Corpus Linguistics* 22 (3): 319–44. <https://doi.org/10.1075/ijcl.22.3.02lov>.
- Ma, Ruiming, Thomas Van Hoey, and Benedikt Szmrecsanyi. 2025. "Isomorphism-Inspired Theorising about Optionality and Variation: No Empirical Support from English Grammar." *English Language & Linguistics*, May, 1–21. <https://doi.org/10.1017/S1360674325000097>.
- Paolini, Chiara, Benedikt Szmrecsanyi, Mariana Montes, and Hubert Cuyckens. 2023. "Distributional Semantics and the English Dative Alternation: Recipient and Theme Slots Matter (to Some Extent)." *Paper of the Linguistic Society of Belgium* 17 (December). <https://doi.org/10.61430/CSGM7397>.
- QLVL. 2021. "Neposem." Zenodo. <https://doi.org/10.5281/ZENODO.5710426>.

Sankoff, David. 1988. "Sociolinguistics and Syntactic Variation." In *Linguistics: The Cambridge Survey: Volume 4: Language: The Socio-Cultural Context*, edited by Frederick J. Newmeyer, 4:140–61. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620577.009>.

Szmrecsanyi, Benedikt, Matt Hunt Gardner, Ruiming Ma, and Thomas Van Hoey. Accepted. "Empirical Accountability Meets Theorizing about Language Variation." In *Empirical Accountability in Variation Linguistics: Taking the next Step*, edited by Patrica Cukor-Avila, Sali A. Tagliamonte, and Guy Bailey. Cambridge: Cambridge University Press.

Paolini, Chiara, Dirk Speelman, Hubert Cuyckens and Benedikt Szmrecsanyi (2025, *in print*). The English dative alternation in vector space: how much does meaning matter?

Szmrecsanyi, Benedikt, Matt Hunt Gardner, and Thomas Van Hoey (in print). *Empirical accountability meets theorizing about language variation*.

# Isomorphism as a constraint on redundancy? Evidence from English and Dutch

Pijpops, Dirk & Eva Zehentner

(University of Antwerp; & University of Zurich)

Keywords: isomorphism, who-did-something-to-whom, redundancy, efficiency, corpus

Isomorphism holds that languages tend towards one-on-one correspondences between form and function (Haiman 1980). However, one of the main functions of grammar, viz. marking who did what to whom, is often performed through multiple formal strategies (Levshina 2021). For example, in Dutch sentences such as (1)-(6), the giver of a hug can be formally distinguished from its recipient through (i) word order, as in (1), (ii) pronominal case marking, as in (2), (iii) verbal agreement, as in (3), (iv) prepositions, as in (4), any combination of those strategies, as in (5), or in fact, none at all, as in (6), where the sentence is left ambiguous.

- (1) *Morgen kan je papa een dikke knuffel geven.* (word order)
- (2) *Papa kan jij morgen een dikke knuffel geven.* (pronominal case marking)
- (3) *Papa kun je morgen een dikke knuffel geven.* (verbal agreement)
- (4) *Aan papa kan je morgen een dikke knuffel geven.* (preposition)
- (5) *Morgen kun jij aan papa een dikke knuffel geven.* (word order, pronominal case marking, preposition)

All: 'You can give daddy a big hug tomorrow'.

- (6) *Papa kan je morgen een dikke knuffel geven.* (ambiguous)

'You can give daddy a big hug tomorrow' or 'Daddy can give you a big hug tomorrow'

Such systemic redundancy, i.e. the availability of various formal strategies in the language system to perform the same function, seems to fly in the face of isomorphism (Van de Velde 2014). However, it could be claimed that, while language does exhibit systemic redundancy, (extreme) syntagmatic redundancy in language is rare. That is, there is evidence suggesting that strategies tend to be used in complementary fashion in actual language use, such that it is typically only one or maximally two strategies that are drawn on in one and the same sentence (cf. Leufkens 2020: 83-84).

We test this claim by investigating both Dutch and English corpus data, zooming in on how the agents and recipients of the cognate verbs *geven* 'give' and *give* – which are both usually animate and volitional – are formally disambiguated. We choose Dutch and English as closely related languages that have the same four strategies at their disposal for marking who did what to whom; crucially, though, Dutch grammar leaves its language users more leeway in determining when to use which strategy, through variation in all four strategies, as exemplified in (1)-(6). By contrast, in English, only prepositional marking is truly variable.

Our results show that double marking is the default in both Dutch and English. Furthermore, there is tentative evidence that Dutch language users make use of the greater flexibility offered by the system to modulate the degree of redundancy in their sentences in an ad-hoc way, while English grammar exhibits a more consistent level of double marking throughout. In addition, redundancy in Dutch but not English seems to be impacted by processing complexity. We interpret these findings as supporting

isomorphism as a guiding factor in strategy use, but in a more gradient way, which is impacted, among other things, by language-specific systemic properties.

## References

- Haiman, J. 1980. The Iconicity of Grammar: Isomorphism and Motivation. *Language*. 56.3: 515–540.
- Levshina, N. 2021. Cross-Linguistic Trade-Offs and Causal Relationships Between Cues to Grammatical Subject and Object, and the Problem of Efficiency-Related Explanations. *Frontiers in psychology*. 12: 648200.
- Leufkens, S. 2020. A functionalist typology of redundancy. *Revista da Abralin* 19(3). 79–103..
- Van de Velde, F. 2014. Degeneracy: the maintenance of constructional networks. In R. Boogaart, T. Coleman & G. Rutten (eds.), *Extending the Scope of Construction Grammar*, vol. 1, 141–179. Berlin: Mouton de Gruyter.

# Didymophilia in language

Freek Van de Velde  
(KU Leuven)

Keywords: alternation, exaptation, degeneracy, free variation

Proponents of the no-synonymy hypothesis have argued that unconditioned ‘free’ variation between alternating variants in language is exceedingly rare or non-existent. Indeed, synchronic variationist linguists have built lofty multivariate models that are able to explain large swaths of the variance in near-synonymous constructions (see Pijpops 2019, Ch.2 for a discussion), especially when taking into account social indexation of the forms: even if the distribution of alternating variants cannot ‘residuelessly’ be explained by language-internal (e.g. pronominality, lexical biases) and cognitive-semantic factors (e.g. animacy, topicality, priming) alone, the added explanatory power of social variables (SES, gender, region), often suggests the variants are prone to second-order indexing (in the sense of Silverstein 2003).

Diachronic variationists, on the other hand, have for the most part focused on cases in which an incoming mutant gradually takes over an older form (Blythe & Croft 2012), ousting the latter form from the language. This scenario indicates that a neat distinction on functional or social grounds of the two forms is not very stable. The demarcation is not stable through time. Moreover, while it is not unheard of that near-synonymous forms develop new meanings in a process of exaptation (Van de Velde & Norde 2016; De Smet & Van de Velde 2020), competing constructions can also converge over time, rather than divide the functional space among them (De Smet et al. 2018). Doubt with regard to a neat division of labour between competing constructions also comes from studies pointing out that synonymy and non-isomorphic tendencies can have advantages for the system, which is protected against function loss in times of perturbation (Van de Velde 2014; Fonteyn & Maekelberghe 2018), and that competing constructions can co-exist – a phenomenon that could be called ‘didymophilia’: a predilection or fascination for twins.

Retaining a residu of unexplained variance is advantageous, and this may be the reason behind the pervasiveness of variation. I will discuss a number of case studies in which the distribution of two associated variants changes over time, but rather than a smooth replacement, the two variants co-exist for a remarkably long time, and small-scale semantic effects creep in. These case studies are: strong vs. weak preterites in Dutch, V1 conditionals vs. asyndetic conditionals in Dutch, internal vs. external possessors in Dutch, and ‘red’ (AUX-V) versus ‘green’ (V-AUX) word order in the Dutch verbal endgroup. For the preterites, the alternants become sensitive to Aktionsart over time (De Smet & Van de Velde 2020). For V1 conditions, the alternants become associated with different lexical-semantic regions (Nijs & Van de Velde 2023). For the red vs. green word order, the alternants are sensitive to valency considerations (Sevenants et al. 2025).

Blythe, Richard & William Croft. 2012. ‘S-curves and the mechanisms of propagation in language change’. *Language* 88(2): 269-304.

De Smet, Hendrik, Frauke D’hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. ‘The changing functions of competing forms: attraction and differentiation’. *Cognitive Linguistics* 29(2): 197-234.

- De Smet, Isabeau & Freek Van de Velde. 2020. 'Semantic differences between strong and weak verb forms in Dutch'. *Cognitive Linguistics* 31(3): 393-416.
- Fonteyn, Lauren, & Charlotte Maekelberghe. 2018. 'Competing motivations in the diachronic nominalisation of English gerunds'. *Diachronica* 35(4): 487-524.
- Nijs, Julie & Freek Van de Velde. 2023. 'Resemanticising 'free' variation. The case of V1 conditionals in Dutch'. In: Kristin Kopf & Thilo Weber (eds.), *Free variation in grammar. Empirical and theoretical approaches*. Amsterdam: John Benjamins. 229-257
- Pijpops, Dirk. 2019. *How, why and where does argument structure vary? A usage-based investigation into the Dutch transitive-prepositional alternation*. PhD, KU Leuven.
- Sevenants, Anthe, Freek Van de Velde & Dirk Speelman. 2025. 'Investigating lexical-semantic effects on morphosyntactic variation using elastic net regression'. *Corpus Linguistics and Linguistic Theory* (published ahead of print).
- Silverstein, Michael. 2003. 'Indexical order and the dialectics of sociolinguistic life'. *Language and Communication* 23: 193-229.
- Van de Velde, Freek. 2014. 'Degeneracy: the maintenance of constructional networks'. In: Ronny Boogaart, Timothy Coleman & Gijsbert Rutten (eds.), *The extending scope of construction grammar*. Berlin: De Gruyter Mouton. 141-179.
- Van de Velde, Freek & Muriel Norde. 2016. 'Exaptation. Taking stock of a controversial notion in linguistics'. In: Muriel Norde & Freek Van de Velde (eds.), *Exaptation and language change*. Amsterdam: John Benjamins. 1-35



# Variation and isomorphism in the English comparative alternation

Hikaru Hotta & Martin Hilpert

(University of Neuchâtel)

Keywords: English comparatives, frequency effects, isomorphism, schematicity, variation

This study presents a corpus-based analysis of the English comparative alternation (e.g., periphrastic *more proud* vs. morphological *prouder*). Previous studies have identified general factors underlying the comparative alternation (e.g., Mondorf 2009). For instance, adjectives with more syllables in predicative usage tend to favor the periphrastic variant. However, substantial empirical evidence suggests that both an abstract schema (e.g., ADJ + the suffix *-er*) and its specific instances (e.g., *prouder*) can be symbolic units that are cognitively represented in the mind and conventionalized within the community (e.g., Katz, Rexer, and Lukatela 1991). This may suggest that competition between variants operates not only at the schematic level, as previous studies have demonstrated, but also at the lexically specific level. The primary goal of this study is to investigate the potential impact of lexically specific representations in the comparative alternation.

We reanalyze Hilpert (2008)'s data, which comprises 247 alternating adjectives from the British National Corpus, using a mixed-effects logistic regression model. This approach accommodates nested structures within the dataset. Specifically, our model distinguishes between group-level predictors (e.g., number of syllables) and observation-level predictors (e.g., predicative or attributive usage). Individual adjectives are included as random effects.

The model shows that while most fixed predictors from Hilpert (2008) remain relevant factors in the alternation, the strength of their effects is substantially moderated. Moreover, a closer examination of the random effects reveals idiosyncratic patterns among adjectives, such as monosyllabic adjectives favoring the periphrastic form (e.g., *apt*, *real*) or trisyllabic adjectives taking the morphological form with considerable frequency (e.g., *untidy*, *unhappy*). The model also accounts for diverging tendencies within certain groups, such as disyllabic *-ly* adjectives showing contrasting tendencies, with *likely* and *costly* favoring the periphrastic form while *lovely* and *ugly* exhibit the opposite tendency.

One key factor underlying these idiosyncrasies is the skewed distribution of adjectives toward one variant over the other. Further analysis shows that the degree of skewness is largely proportional to the token frequency of adjectives in comparative forms. In other words, the more frequently an

adjective is used in comparative forms, the stronger the bias becomes toward either variant.

We discuss these findings in light of the overarching themes of the workshop, which concern the relationship between variation and isomorphism. In the case of the English comparative alternation, the two competing forces may operate at different levels of schematicity. At the schematic level, there is variation that can be predicted by multiple general factors. At the lexically specific level, there is isomorphism, where variants appear to substitute for one another. Importantly, the boundary between variation and isomorphism is gradient: for less frequent forms, a speaker may primarily rely on probabilistic knowledge at the schematic level to select a conventional variant, while directly accessing lexically specific representations for highly entrenched comparative forms.

## References

- Hilpert, Martin. (2008). The English Comparative – Language Structure and Language Use. *English Language and Linguistics* 12(3). 395–417.
- Katz, Leonard, Karl Rexer, and Georgije Lukatela. (1991). The processing of inflected words. *Psychological Research* 53(1). 25–32.
- Mondorf, Britta. (2009). *More support for more-support: the role of processing constraints on the choice between synthetic and analytic forms*. Amsterdam and Philadelphia: John Benjamins.

## What makes alternates contrast? Evidence from a (non-)chain shift

Juliette Kayenbergh  
(KU Leuven, FWO)

Keywords: contrast, isomorphism, progressive construction, simple form, semantic change

Acknowledgments: We gratefully acknowledge the financial support of the FWO (FWO PhD fellowship - 11PBL24N).

The isomorphic principle maintains that languages strive towards ideal form-meaning mappings whereby a single meaning corresponds to a single form and vice versa (Haiman 1980, and Goldberg 1995). This view has been challenged, however, by extensive evidence that language variation is very common and a natural – arguably even desirable – aspect of language (e.g. Poplack 2018, and Gardner et al. 2021). In response, several attempts have been made to reinterpret the isomorphic principle in line with the facts of variation and change (e.g. Leclercq & Morin 2023, and Cai & De Smet 2024).

This paper contributes to this effort by going back to the theoretical roots of the isomorphic principle in Saussurian structuralism. The isomorphic principle is essentially a continuation of the idea that signs define one another through relations of contrast. The mutual hold that signs are thought to have over each other is supported by evidence from semantic change. Specifically, in chain shifts, change in one expression triggers change in a semantically related contrasting expression, just as the structuralist view would predict (e.g. Buyle & De Smet 2018).

However, the present paper addresses a development that behaves as a chain shift only in part, resulting in a new contrast but also in variation. The development in question is the rise of English *BE Ving* (*she is sleeping*) and its impact on the meaning of its simple form counterparts (*she sleeps*). The rise of *BE Ving* and concomitant shifts in the interpretation of simple forms are investigated for seven verbs selected to represent the full range of situation types (see e.g. Vendler 1957, and Smith 1997). These are *stand*, *sleep*, *talk*, *eat*, *read*, *build*, and *catch*. Data have been collected from the EEBO, CLMET and BNC corpora, covering Early Modern, Late Modern and Present-day English and consist of random samples capped at 150 instances per verb per period. For each simple form token, the most plausible interpretation is established, i.e. progressive (1), perfective (2) or habitual (3). If the rise of *BE Ving* affected the use of simple forms, the latter are expected to shift away from contexts favouring a progressive interpretation, as in (1), as *BE Ving* gains in frequency.

(1) **Sleep** you, or wake you, Lady bright? (CLMET, 1798)

(2) [He] bit his lip, and **read** the letter twice, [...]; and then he asked what o'clock it was. (CLMET, 1837)

(3) though he **eats** very well, [...] he seems to have lost all care of his person (CLMET, 1841)

Results paint a more complex picture. Present tense contexts behave more or less as expected, with the rise of *BE Ving* leading to interpretational changes in the simple form. However, in past tense contexts, the rise of the new paradigmatic alternate *BE Ving* has little noticeable impact on the simple forms. In these contexts, variation persists.

The results indicate that isomorphic and non-isomorphic outcomes can coexist in the same development. A possible explanation is that variation emerges as a side-effect of contrasts that are functional elsewhere in the system.

## References

- Buyle, Anouk and Hendrik De Smet (2018), Meaning in a changing paradigm: the semantics of you and the pragmatics of thou, *Language Sciences* 68, 42-55.
- Cai, Yingying and Hendrik De Smet (2024), Are categories' cores more isomorphic than their peripheries? *Frontiers in Communication* 9, 1310234.
- Gardner, Matt H., Eva Uffing, Nicholas Van Vaeck and Benedikt Szmrecsanyi (2021), Variation isn't that hard: Morphosyntactic choice does not predict production difficulty, *PLoS ONE* 16(6), e0252602.
- Goldberg, Adele E. (1995), *Constructions: A Construction Grammar Approach to Argument Structure*, Chicago: University of Chicago Press.
- Haiman, John (1980), The iconicity of grammar: Isomorphism and motivation, *Language* 56(3), 515-540.
- Leclercq, Benoît, and Cameron Morin (2023), No equivalence: A new principle of no synonymy, *Constructions* 15(1), 1-16.
- Poplack, Shana (2018), Categories of grammar and categories of speech: When the quest for symmetry meets inherent variability, in N. Lapidus Shin, and D. Erker (eds), (2018), *Questioning Theoretical Primitives in Linguistic Inquiry*, Amsterdam: John Benjamins, 7-34.
- Smith, Carlota S. (1997), *The Parameter of Aspect* (Studies in Linguistics and Philosophy 43), Dordrecht: Springer Netherlands.
- Vendler, Zeno (1957), Verbs and Times, *The Philosophical Review* 66(2), 143-160.

### A three-layer analysis of the English dative alternation

The relationship between pervasive variability and the structuralist “one-form-one-meaning” principle, a postulate endorsed across various theoretical frameworks, continues to spark debate. In an earlier chapter of this debate, Newmeyer (1983: 116) astutely observed that much of the debate hinges on how the notion of “meaning” is conceptualized (cf. Weber & Kopf 2023: 3). Newmeyer’s observation remains pertinent, and we accordingly aim to advance the discussion by offering a more theoretical contribution.

Specifically, we advocate for a “Three-Layer Approach” to meaning, a framework rooted in Coseriu structural-functional linguistics and neo-Gricean pragmatics (Coseriu 1989, Levinson 2000, Belligh & Willems 2021). This approach differentiates between (i) encoded language-specific meanings, (ii) conventionalized senses, and (iii) specific readings in discourse. Furthermore, we draw on Coseriu’s (1989) distinction between *polyvalence* (a single expression with multiple senses) and *polymorphy* (“grammatical synonymy”, i.e. more than one expression for a single encoded meaning).

We elaborate on De Vaere’s (2023) analysis of the German dative alternation and explore its applicability to the English dative alternation. The German and English alternations are illustrated below:

- (1) a. *Er schickt seiner Mutter einen Brief.* (IOC)  
b. *Er schickt einen Brief an seine Mutter.* (POC)
- (2) a. *He sends his mother a letter.* (DOC)  
b. *He sends a letter to his mother.* (POC)

The German and English ditransitive alternation are similar. Both involve a RECIPIENT that can be expressed with or without a preposition. Corpus evidence further indicates that both alternations are also associated with similar motivating factors related to the THEME and RECIPIENT (pronominality, definiteness, length, etc.) (De Vaere 2023: 95). However, there is an important formal difference. The German alternants are “indirective constructions” (Malchukov, Haspelmath, & Comrie 2010): the German non-prepositional alternant is not a Double Object Construction (DOC) but an Indirect Object Construction (IOC), sharing its alignment pattern with the Prepositional Object Construction (POC).

The three-layer analysis conceptualizes the German ditransitive construction as an overarching AGENT-THEME-GOAL (ATG) *constructeme*, with IOC and POC as two “allostructions”. A distributional analysis indicates that neither IOC nor POC are confined to a CAUSED POSSESSION and CAUSED MOTION meaning, respectively. The difference between IOC and POC involves polymorphy of the GOAL argument within the constructeme. A similar analysis applies to the English dative alternation, but in English the polymorphy of the GOAL argument coincides with two alternating word order patterns to a much larger extent than in a case language such as German.

### References

- Belligh, T., & Willems, K. (2021). What's in a code? The code-inference distinction in Neo-Gricean Pragmatics, Relevance Theory, and Integral Linguistics. *Language Sciences*, 83, 101310.  
<https://doi.org/10.1016/j.langsci.2020.101310>
- Coseriu, E. (1989). Principes de syntaxe fonctionnelle. In G. Kleiber & G. Roques (Eds.), *Travaux de linguistique et de philologie* (Vol. XXVII, pp. 5–46). Klincksieck.
- De Vaere, H. (2023). *The ditransitive alternation in present-day German: A corpus-based analysis*. John Benjamins. <https://doi.org/10.1075/sigl.6>

- Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.
- Malchukov, A., Haspelmath, M., & Comrie, B. (2010). Ditransitive constructions: A typological overview. In A. Malchukov, M. Haspelmath, & B. Comrie (Eds.), *Studies in ditransitive constructions: A comparative handbook* (pp. 1–64). De Gruyter Mouton.
- Newmeyer, F. J. (1983). *Grammatical theory: Its limits and its possibilities*. University of Chicago Press.
- Weber, T., & Kopf, K. (2023). Chapter 1. Free variation, unexplained variation? In K. Kopf & T. Weber (Eds.), *Free variation in grammar: Empirical and theoretical approaches* (pp. 1–20). John Benjamins. <https://doi.org/10.1075/slcs.234.01web>

# Isomorphism and optionality at the level of clause-linking

Mathilde Pinson

(Sorbonne Nouvelle University – Paris)

Keywords: Isomorphism, optionality, clause-linking, language change, iconicity

The question whether the distinction between two alternative forms is semantically or sociolinguistically motivated can be tackled at all levels of the lexico-syntactic continuum, including clause-linking (complementation and adverbial subordination).

In line with the principles of isomorphism and iconicity, Givón's (1985) binding theory asserts that the degree of syntactic dependency between the main clause and the subordinate clause reflects the degree of semantic dependency between the processes encoded in the two clauses. This implies that the presence or absence of *to* after the verb *help* exhibits a semantic distinction (cf. e.g. Dixon 1991; Duffley 1992; Fischer 1995).

(1a) *I helped him  $\emptyset$  eat.*

(1b) *I helped him **to** eat.*

Researchers typically estimate that (1a) encodes a more manipulative, agentive and direct form of assistance on the part of the helper than (1b).

In the same vein, Kortmann (1997) argues that polymorphemic subordinators are more likely than monomorphemic subordinators to express complex semantic relations. This suggests that *as if* and *like* are unlikely to be synonymous.

(2a) *You look **like** you have lost some weight.*

(2b) *You look **as if** you have lost some weight.*

If one follows Kortmann, (2a) may convey an evidential/epistemic interpretation, while (2b) may imply that the proposition included in the subordinate clause is actually untrue, which is more cognitively complex than (2a) as it encodes a comparison with a fictitious reality (counterfactual interpretation).

However, these two examples of semantic contrast are at odds with the well-known observations that the distinction between *help  $\emptyset$*  and *help to* and between *like* and *as if* is largely due to diatopic and/or diaphasic variation (e.g. McEnery and Xiao 2005 on *help  $\emptyset$ /to*; Bryant 1962, Whitman 1974 on *like/as if*).

A corpus study, based on the Old Bailey Proceedings, the Times Digital Archives and the Corpus of Historical American English, shows that these two contradictory perspectives can be reconciled with the help of a diachronic approach, in line with De Smet (2019). Although the evolution of *like* and *help  $\emptyset$*  are different, they show that sociolinguistic and semantic distinctions have both been operational, but at different periods in their histories. Semantic and sociolinguistic distinctions have been unstable

over time, because of a combination of factors including *horror aequi*, semantic persistence, functional reallocation, expressivity and the colloquialisation of English.

It is hoped that these two case studies will show how semantic and sociolinguistic approaches to formal alternation can be reconciled when cognitive and social principles are combined with a diachronic dimension.

## References

- Bryant, M. M. (1962), *Current American Usage*, New York: Funk & Wagnalls Company Inc.
- De Smet, Hendrik (2019), The motivated unmotivated: Variation, function and context. In Bech, Kristin & Ruth Möhlig-Falke (eds.), *Grammar – Discourse – Context: Grammar and Usage in Language Variation and Change*. Berlin: De Gruyter, 305-332.
- Dixon, R. M. W. (1991), *A New Approach to English Grammar on Semantic Principles*, Oxford: Clarendon Press.
- Duffley, Patrick (1992), *The English Infinitive*, London: Longman.
- Fischer, Olga. 1995. The distinction between *to* and bare infinitival complements in Late Middle English, *Diachronica* 12, 1-30.
- Givón, Talmy (1985), Iconicity, isomorphism and non-arbitrary coding in syntax. In Haiman, John (ed.), *Typological Studies in Language, vol. 6: Iconicity in Syntax*, Amsterdam/Philadelphia: John Benjamins, 187-219.
- Kortmann, Bernd (1997), *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*, Berlin/New York: Mouton de Gruyter.
- McEnery, Anthony & Richard Xiao (2005), *Help or help to*: What do corpora have to say? *English Studies* 86.2, 161-187.
- Whitman, C. D. (1974), *Like* as a conjunction, *American Speech* 49(1), 156-158.



## Some remarks on no synonymy and individual variation

Peter Uhrig

(Friedrich-Alexander-Universität Erlangen-Nürnberg)

Keywords: Construction Grammar, Principle of No Synonymy, individual variation

Uhrig (2015) claims that “[t]he principle of no synonymy is overrated.” While Leclercq & Morin (e.g. 2023) disagree, I will attempt to defend that position and review the arguments in the light of recent updates both to the theoretical framework (e.g. Goldberg 2019) and to the body of literature discussing (near) synonymy, alternations, variation and optionality (e.g. Gardner et al. 2021, Laporte et al. 2021, Levshina & Lorenz 2022, Leclercq & Morin 2023).

One particular aspect I would like to focus on is individual variation by the same speaker, as in examples (1) and (2) taken from Uhrig (2022):

- (1) I don’t know if it’s ... if you call that **ideology** [aɪ], I mean, for President Obama, you know, ending the Iraq war and bringing the soldiers home was something he campaigned on ... so ... it, it was more of a ... you can call it **ideology** [ɪ], you can call it a policy priority.  
<http://go.redhenlab.org/pgu/0139>
- (2) argument for Rick Perry, which is an interesting one, it's a compelling one, is that he has the right **ideology** [aɪ], and he has experience implementing that **ideology** [aɪ]. I would be willing to wager, and I think anyone could ... would agree that there are lots of people who have that same **ideology** [ɪ] and who have some level of experience doing that, and they may be Cambodian, they may be women, they may be LGBT, there are a range of folk who have that **ideology** [aɪ], if he... <http://go.redhenlab.org/pgu/0140>

I would claim that in both cases it might not be straightforward to argue that for the pronunciation variants of *ideology* (with initial [aɪ] and with initial [ɪ]) speakers “flexibly use them as part of their dynamic construction of sociolinguistic identity” (Leclercq & Morin 2023: 10). Instead, I would prefer to maintain that there is free variation, leading to (obviously quite rare) pockets of synonymy, and that there is nothing wrong with permitting synonymy in a constructionist model of language. We might at the same time under- and overestimate speakers’ abilities if we claim that all variation is driven by semantic or pragmatic differences.

A second point I would like to make is that even where it holds strictly speaking, the principle of no synonymy is overrated in relation to actual communicative events. One case in point is priming effects (see e.g. Hoey 2005), which can influence the choice between alternatives (such as *for instance* and *for example*) without a discernible difference in meaning. While all types of sociolinguistic factors have been mentioned as potentially part of the meaning in a wider sense, I would argue that priming cannot be sensibly construed as such.

### Works Cited

- Gardner, Matt Hunt, Eva Uffing, Nicholas Van Vaeck & Benedikt Szmrecsanyi. 2021. Variation isn’t that hard: Morphosyntactic choice does not predict production difficulty. *PLoS ONE* 16(6), e0252602.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Hoey, Michael. 2005. *Lexical priming: A new theory of words and language*. Abingdon: Routledge.

Laporte, Samantha, Tove Larsson & Larissa Goulart. 2021. Testing the Principle of No Synonymy across levels of abstraction: A constructional account of subject extraposition. *Constructions and Frames* 13(2). 230–262.

Leclercq, Benoît & Cameron Morin. 2023. *No Equivalence: A new principle of no synonymy*. *Constructions* 15, 1–16.

Levshina, Natalia & David Lorenz. 2022. Communicative efficiency and the Principle of No Synonymy: predictability effects and the variation of *want to* and *wanna*. *Language and Cognition* 14:2, 249–74.

Uhrig, Peter. 2015. Why the Principle of No Synonymy is Overrated. *Zeitschrift für Anglistik und Amerikanistik* 63(3). 323–337.

Uhrig, Peter. 2022. *Large-Scale Multimodal Corpus Linguistics – The Big Data Turn*. Habilitation Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.

# One-to-one, many-to-many, tom[eɪ]to-tom[ɑ:]to ...

## On how cognitive linguistics and sociolinguistics are perhaps closer to each other than meets the eye

The cognitive-functional ideal of a one-to-one correspondence between a form and a meaning, explicable through economy and communicative efficiency (Goldberg 2019; Levshina & Lorenz 2022), comes with the responsibility of having to identify potential meaning differences between competing alternatives at various levels, namely semantics, pragmatics, discourse-function, and social meaning (Leclercq & Morin 2023). By contrast, the variationist ideal of ordered heterogeneity, witnessed by the considerable amount of structured variance in actual language data, comes at the price of having to dismiss (subtle) meaning differences a priori in favor of a strong focus on self-proclaimed extra-linguistic factors like AGE, GENDER, or SOCIO-ECONOMIC STATUS to account for that variance (Labov 1972, 2014; Tagliamonte 2012).

The present study proposes that this divide might be less dramatic than it seems. For one thing, isomorphism requires pure token repetition, which is merely a theoretical concept. In actual language use, speakers are always confronted with variable perceptual input which they must abstract away from, even if it is seemingly the same utterance type (Ambridge & Lieven 2015; Schmid 2018). And if schematization is treated as a crucial, domain-general process, cognitive linguistics should embrace the idea of (admittedly very different degrees of) variation. This does not run counter to the pursuit of identifying meaning differences between competing variants, only that these differences are probabilistic rather than categorical (Daugs & Lorenz 2024). And for another thing, the “built-in probabilistic indexicality” of many utterances, which, among other things, reflects social contingency, must be the result of both their degree of conventionality as well as their degree of entrenchment (Schmid 2020: 23). Therefore, a speaker’s unique social identity, which essentially emerges from meaning associated with recurring situational contexts (Ellis 1996), will very much impact their linguistic choices in a probabilistic fashion and is thus best not located outside the linguistic system. This does not run counter to the idea of the ‘envelope of variation’, only that meaning differences are also a function of socio-cultural aspects that need to be accounted for.

The testing ground for the claims above will be the case of intensifiers in US American English, specifically regionally stratified and socially indexed forms like *hella* and *wicked* and their competitors. Corpus and web-based data (e.g. COCA, NOW, TV Corpus, Twitter, Youglish) will be utilized to discuss how socio-pragmatic properties can be fruitfully integrated into cognitive-functional usage-based frameworks of language, most notably Schmid’s (2020) *Entrenchment-and-Conventionalization Model*, which unifies the communal and the individual side of the linguistic system.

### References

- Ambridge, Ben & Elena Lieven. 2015. A constructivist account of child language acquisition. In Brian MacWhinney & William O’Grady (eds.), *The handbook of language emergence*, 478–510. Chichester: John Wiley & Sons.

- Daug, Robert & David Lorenz. 2024. A radically usage-based, collostructional approach to assessing the differences between negative modal contractions and their parent forms. *Corpus Linguistics and Linguistic Theory* (aop).
- Ellis, Nick. 1996. Sequencing in SLA: phonological memory, chunking and points of order. *Studies in Second Language Acquisition* 18. 91–126.
- Goldberg, Adele. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 2014. What is to be learned. The community as the focus of social cognition. In Martin Pütz, Justyna Robinson & Monika Reif (eds.), *Cognitive sociolinguistics: Social and cultural variation and language use*, 23–52. Amsterdam: John Benjamins.
- Levshina, Natalia & David Lorenz. 2022. Communicative efficiency and the Principle of No Synonymy: predictability effects and the variation of *want to* and *wanna*. *Language and Cognition* 14(2). 249–74.
- Schmid, Hans-Jörg. 2018. Unifying entrenched tokens and schematized types as routinized commonalities of linguistic experience. *Yearbook of the German Cognitive Linguistics Association* 6(1). 167–82.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Tagliamonte, Sali A. 2012. *Variationist sociolinguistics: Change, observation, interpretation*. Oxford: Wiley-Blackwell.

## **Probing deeper into No Equivalence: A usage-based, radically dynamic perspective**

Stefan Hartmann & Tobias Ungerer  
(Heinrich Heine University Düsseldorf & University of Toronto)

Keywords: isomorphism, optionality, no equivalence, usage-based linguistics

Goldberg's (1995) "Principle of No Synonymy", reframed by Leclercq & Morin (2023) as the "Principle of No Equivalence", posits that two constructions which differ in form must be semantically, pragmatically, and/or socially distinct. In this talk, we outline a usage-based perspective on three aspects of No Synonymy/No Equivalence (henceforth NS/NE) that have previously remained underexplored: (1) its causal origins, (2) its scope within a dynamic theory of meaning, and (3) its relation to optionality, i.e., the existence of 'alternative ways of saying 'the same' thing' (Labov 1972: 188).

Regarding the first aspect, NS/NE are fundamentally descriptive and synchronic principles, pertaining to language at the level of the linguistic community. In contrast, the causal mechanisms that give rise to NS/NE are inherently diachronic developments, connected to cognitive factors at the level of individual language users. If we conceive of language as a complex adaptive system (Beckner et al. 2009), and if we accept the distributional hypothesis that the meaning of linguistic constructions is largely inferred from distributional cues, then it follows that, as no two items can have exactly the same distribution, they will develop semantic, pragmatic, or social differences.

Concerning the second aspect, the degree to which NS/NE apply depends on the researcher's view of how meaning is constructed in language (Cappelle 2024). Some versions of the isomorphic principle rest on a bilateral view of the linguistic sign (Saussure 1916), which also still influences the way we conceive of constructions in Construction Grammar as form-meaning pairs (Ungerer & Hartmann 2023). However, there is a growing consensus that meaning-making is much more dynamic and interactional (e.g., Scott-Phillips 2015; Goldberg 2019; Schmid 2020). If we follow this approach faithfully, and if we take the above-mentioned distributional hypothesis seriously, then we can postulate an even broader Principle of NS/NE according to which not even the meaning of the same form is ever fully stable.

Finally, we argue that isomorphism and optionality do not exclude each other but are rather closely connected. The general tendency of languages to avoid synonymy can have two complementary effects on linguistic innovation. On the one hand, it can lead to statistical preemption and the "blocking" of some theoretically possible forms (e.g., \*stealer – thief; Boyd & Goldberg 2011). On the other hand, the very fact that these "blocked" forms seem unconventional can lead language users to use them for, e.g., stylistic or humorous purposes in the spirit of linguistic extravagance (Haspelmath 1999). This can, at least temporarily, create the impression that multiple near-synonymous forms co-exist and can be optionally substituted for each other. However, over time, comprehenders may ascribe distinct pragmatic meanings to these apparent near-synonyms, based on Grice's (1975) maxim of Manner or Levinson's M-Heuristic ("Marked message indicates marked situation", Levinson 2000: 136), thus yielding optionality between forms that are nevertheless not fully equivalent.

In sum, we propose that a deeper understanding of NS/NE requires a stronger focus on the usage-based mechanisms underlying these principles as well as their implications vis-à-vis recent nuanced views of form-meaning isomorphism.

## References

- Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a Complex Adaptive System: Position Paper. *Language Learning* 59 Suppl. 1. 1–26.
- Boyd, Jeremy K. & Adele E Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language* 87(1). 55–83.
- Cappelle, Bert. 2024. *Can Construction Grammar be proven wrong?* Cambridge: Cambridge University Press.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Grice, H. P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, vol. III, 183–198. New York: Academic Press.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.
- Labov, William. 1972. *Sociolinguistic patterns*. Oxford: Blackwell.
- Leclercq, Benoît & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. *Constructions* 15(1).
- Levinson, Stephen C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: MIT Press.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Payot.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Scott-Phillips, Thom. 2015. *Speaking our minds: Why human communication is different, and how language evolved to make it special*. London: Pelgrave.
- Ungerer, Tobias & Stefan Hartmann. 2023. *Constructionist approaches: Past, present, future*. Cambridge: Cambridge University Press.

# Empirical accountability meets theorizing about language variation: The Principle of Optionality

Thomas Van Hoey, Ruiming Ma, Matt H. Gardner & Benedikt Szmrecsanyi  
(FWO & KU Leuven, KU Leuven, University of Oxford, KU Leuven)

Keywords: optionality, variation, dysfluencies, principle of isomorphism, spoken corpus

Functional and cognitive linguists in particular tend to consider language variation (specifically: **optionality** between different ways of saying the same thing) dysfunctional, abnormal, and inconvenient to language users, a theoretical stance that follows from adopting the Principle of Isomorphism (Haiman 1980), the Principle of No Synonymy (Goldberg 1995) or the Principle of No Equivalence (Leclercq & Morin 2023). Parallel to these principles, there exists the idea that absolute complexity (i.e. the length of the grammatical descriptions of language) is proportional to relative complexity, i.e. processing and production difficulty (Miestamo 2008). Together, then, these two ideas assume that optionality increases the relative complexity of language production. But is that actually so?

In this contribution, we use a corpus-based psycholinguistics research design with a variationist twist and analyze SWITCHBOARD, a corpus of conversational spoken American English. We ask if and how grammatical optionality correlates with two symptoms of production difficulty, namely filled pauses (um and uh) and unfilled pauses (speech planning time), building on a pilot study (Gardner et al. 2021). The investigated optionality contexts consist of 20 well-known types, ranging from the dative alternation (Bresnan et al. 2007), over *-body/-one* suffixation (D’Arcy et al. 2013), to gerund/infinitive clause complementation (Deshors & Gries 2016).

We further include three factors that may influence production difficulty: the number of variants at play in an optionality context, the constraints that govern choosing, and the probabilistic cueing of a choice (how much does the context predict variant choice?). Mixed-effects regression modeling shows that, overall, optionality contexts do not make speech production more dysfluent, regardless of how cued optionality contexts are, how many constraints are in operation, or how many variants there are to choose from.

The results thus call into question old dogmas like the Principle of Isomorphism. We instead propose an alternative—the Principle of Optionality (“Languages and language users favor the availability of different ways of saying the same thing”), a principle that can in fact be motivated by the existing literature.

## Funding

The C1 project C14/22/045 “Relatively complex or relatively simple? Toward new ways of analyzing language variation” is gratefully acknowledged.

## References

- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Boume, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- D’Arcy, Alexandra, Bill Haddican, Hazel Richards, Sali A. Tagliamonte & Ann Taylor. 2013. Asymmetrical trajectories: The past and present of *–body/–one*. *Language Variation and Change* 25(3). 287–310.
- Deshors, Sandra C. & Stefan Th. Gries. 2016. Profiling verb complementation constructions across New Englishes. *International Journal of Corpus Linguistics* 21(2). 192–218.
- Gardner, Matt Hunt, Eva Uffing, Nicholas Van Vaec & Benedikt Szmrecsanyi. 2021. Variation isn’t that hard: Morphosyntactic choice does not predict production difficulty. (Ed.) Stefan Th. Gries. *PLOS ONE* 16(6). e0252602.
- Goldberg, Adele E. 1995. *Constructions*. Chicago: University of Chicago Press.
- Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56(3). 515–540.
- Leclercq, Benoît & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. *Constructions* 1–16.
- Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 23–42. Amsterdam, Philadelphia: Benjamins.



# **WS10 Issues in the formal and functional typology of focus**

**Olivier Duplâtre & Pierre-Yves Modicom**

## Focus marking strategies in Gawarbati (Indo-Aryan)

Anastasia Panova, Shinichiro Ishihara & Henrik Liljegren  
(Stockholm University, Lund University & Stockholm University)

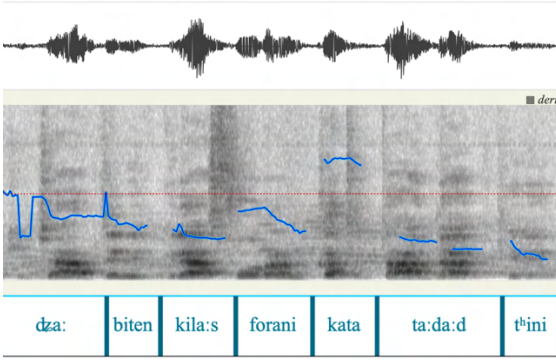
Keywords: focus, information structure, prosody, word order, Indo-Aryan.

Previous studies of focus in Indo-Aryan languages have dealt with the marking of focus by prosody, see, e.g., Féry et al. (2016), Urooj et al. (2019), Kügler (2020) on Hindi/Urdu, Hayes & Lahiri (1991) on Bengali. Patil et al. (2008) and Jabeen (2019, 2022, among others) have further examined the ways in which focus in Hindi/Urdu is marked by prosody and manipulation of word order, i.e., placing the focus constituent in the immediately preverbal position (as earlier claimed in Butt & King 1996).

This paper contributes to the discussion of focus marking in Indo-Aryan by describing narrow focus marking strategies and their interplay in Gawarbati, an underdocumented Indo-Aryan language spoken in the Hindukush region. Our preliminary results show that the marking of focus by pitch peaks, pitch compression and word order in Gawarbati form a complicated trade-off. Furthermore, we see that these particular focus marking strategies are tightly integrated into more general information-structural patterns observed in the natural language data. This study is based on a corpus of spontaneous conversations recorded in the speaker community in 2021-2024.

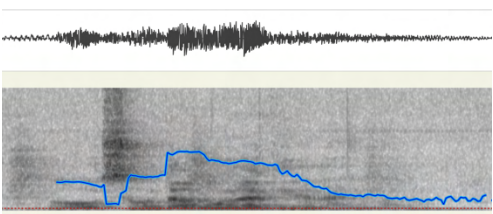
For the pilot study, we analyzed examples of focus in the sense of Krifka (2008) as manifested in wh-questions and answers to wh-questions in a corpus fragment. As expected from the previous experiment-based literature (Jabeen 2019), we found cases where the focused word has a F0-peak with a falling contour, and the focus constituent is located right before the verb, which is often its in-situ position, cf. the focus constituent *kata ta:da:d* in (1).

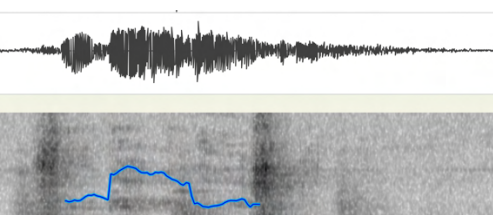
(1)



<i>d̪a:</i>	<i>biten</i>	<i>kila:s</i>	<i>forani</i>	<i>kata</i>	<i>ta:da:d</i>	<i>tʰini</i>
<i>d̪a-a</i>	<i>biten</i>	<i>kila:s</i>	<i>for-an-i</i>	<i>kata</i>	<i>ta:da:d</i>	<i>tʰini</i>
say-IMP.SG	please	class	fourth-GEN-F	how.many	amount	be.PRS.3F.SG
'tell (me) please, how many (students) are there in the fourth class?'						

Importantly, however, in many (if not most) examples, contexts with narrow focus also include topicalization or frame-setting. For example, in (2)-(3) the source argument *kinen-a/lawans-an-a* [where.GEN-M/Drosh-GEN-M] is the focus of the sentence, while the direct object *sauda:* [goods-OBL] is left-dislocated and, apparently, topicalized. In such cases, the left-dislocated topicalized constituent bears the highest F0-peak (with a rising contour), while the focus constituent gets a lower F0 with a falling pitch contour.

- (2)
- 
- |    |        |        |        |            |
|----|--------|--------|--------|------------|
| au | sauda: | mauda: | kinena | a:nima:niu |
|----|--------|--------|--------|------------|
- au*      *sauda-a*      *mauda-a*      *kinen-a*      *a:n-ima:niu*  
 and      goods-OBL      RED-OBL      where.GEN-M      bring-PRS.2PL  
 'and where do you bring the stuff from?'

- (3)
- 
- |        |     |           |         |      |
|--------|-----|-----------|---------|------|
| sauda: | ama | lawansana | a:nimek | nori |
|--------|-----|-----------|---------|------|
- sauda-a*      *ama*      *lawans-an-a*      *a:n-imek*      *nori*  
 goods-OBL      we      Drosh-GEN-M      bring-PRS.1PL      now  
 'We bring the stuff from Drosh now.'

Presumably, the focus constituent in examples like (2)-(3) is still perceived as focused because of other focus marking strategies. First, it appears in the preverbal position (although it can be argued that this is merely a consequence of the displacement of the direct object, which normally precedes the verb). Second, at least in (3), the source argument is followed by the post-focal pitch compression that has been claimed to indicate focus in Hindi (Kügler 2020).

In the talk, we will present a comprehensive corpus-driven account of narrow focus contexts in Gawarbatī, in the spirit of recent bottom-up approaches to information structure such as Matić & Wedgwood (2013) and Ozerov (2018, 2021).

## References

- Butt, Miriam & Tracy H. King. 1996. Structural Topic and Focus without Movement. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the first LFG conference*, 1–15. Stanford: CSLI Publications.
- Féry, Caroline, Pramod Pandey, and Gerrit Kentner. 2016. The prosody of focus and givenness in Hindi and Indian English. *Studies in Language* 40, 302–39.
- Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9, 47–96.
- Jabeen, Farhat. 2019. *Prosody and Word Order: Prominence Marking in Declaratives and Wh-Questions in Urdu/Hindi*. Ph.D. Thesis, University of Konstanz, Konstanz, Germany.
- Jabeen, Farhat. 2022. Word Order, Intonation, and Prosodic Phrasing: Individual Differences in the Production and Identification of Narrow and Wide Focus in Urdu. *Languages*, 7(2), 103.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55, 243–76.

- Kügler, Frank. 2020. Post-focal compression as a prosodic cue for focus perception in Hindi. *Journal of South Asian Linguistics* 10, 38–59.
- Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: the significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49 (1), 127–163.
- Ozerov, Pavel. 2018. Tracing the sources of information structure: Towards the study of interactional management of information. *Journal of Pragmatics* 138, 77–97.
- Ozerov, Pavel. 2021. Multifactorial Information Management (MIM): summing up the emerging alternative to Information Structure. *Linguistics Vanguard* 7(1), 20200039.
- Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry, and Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics* 1, 55–72.
- Urooj, Saba, Benazir Mumtaz, and Sarmad Hussain. 2019. Urdu intonation. *Journal of South Asian Linguistics* 10, 2–22.

## Prosodic realization of focus in Czech: A production experiment

Anna Staňková, Radek Šimík, Tomáš Bořil & Daniela Kořánová  
(Charles University)

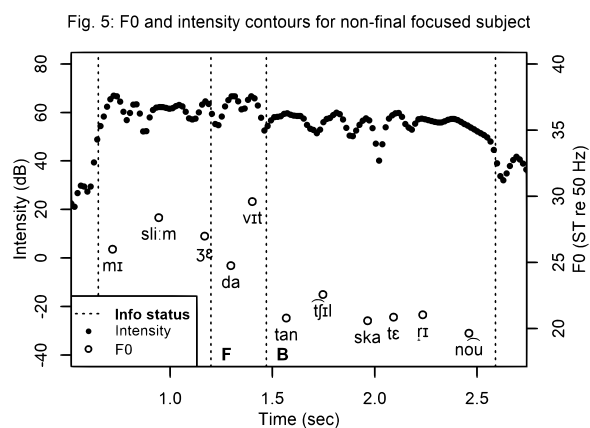
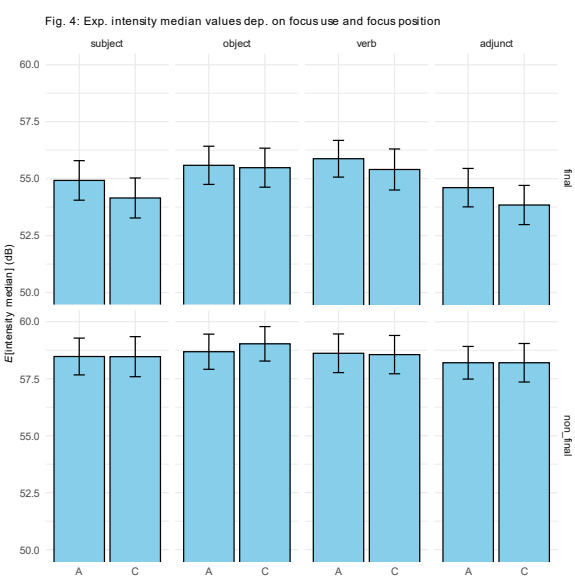
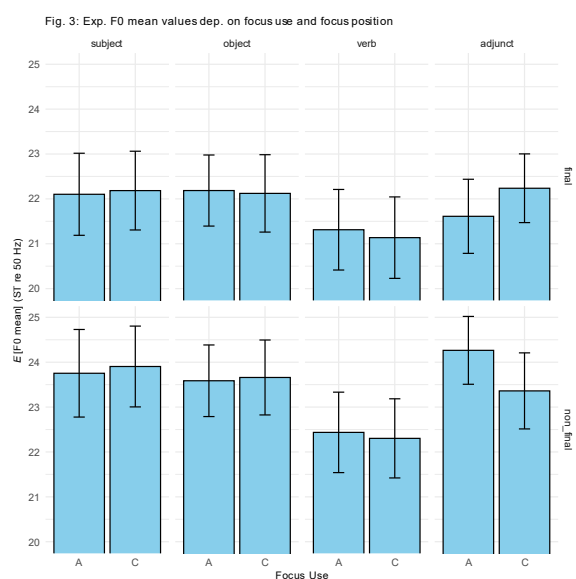
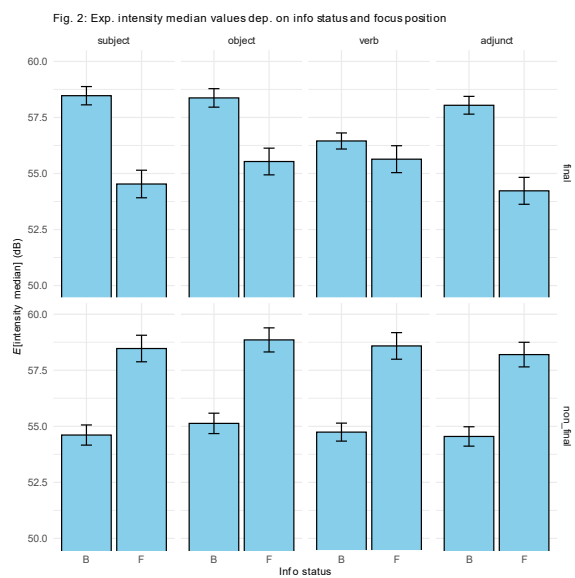
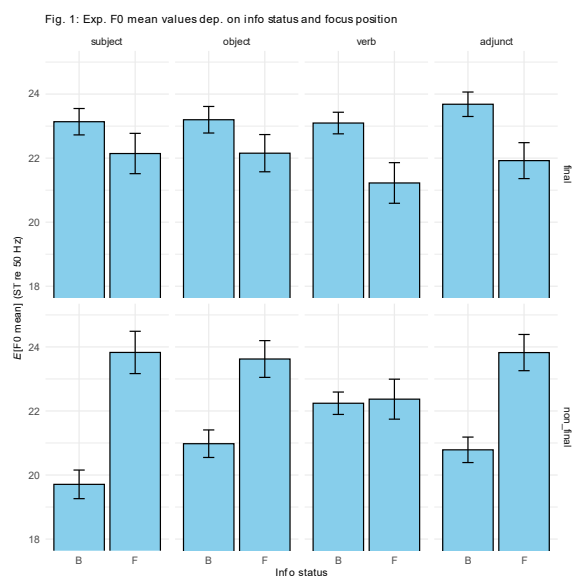
Keywords: focus, givenness, information structure, Czech, prosody

**Background** A focused constituent gives rise to alternative denotations relevant in the current discourse; what is not focused is called the background (Rooth 1992, Krifka 2008). Two prominent uses of focus are in answers to *wh*-questions (answerhood f.) and in corrections of contextually salient utterances (corrective f.); these uses have been observed to be realized by different prosody (Zybatow & Mehlfhorn 2000 for Russian; Katz & Selkirk 2011 for English). Focus encoding in Czech, a language with very flexible word order (Siewierska & Uhlířová 1998), has traditionally been looked at from a syntactic perspective. Focus has been observed to be realized towards the end of an utterance (Mathesius 1939, Sgall et al. 1980 and Firbas 1992). That focus is (also) encoded prosodically is well-known (Daneš 1957, Šimík & Wierzbica 2017, Hamlaoui et al. 2018 and Šimík 2024), but a deeper understanding of how focus and information structure affect prosody in this strongly discourse-configurational language, is still missing.

**Experiment** We ran a production experiment where participants read aloud responses to prerecorded contexts. We manipulated (i) info status of the utterance (focus vs. background), (ii) focus position (final vs. non-final), (iii) focus use (answerhood vs. corrective), all within subjects & items, and (iv) syntactic function of the focused element (within subjects, but between items). We constructed 36 items and mixed them with 44 fillers. The stimuli were distributed on lists using Latin Square and the order of presentation was pseudo-randomized. 68 participants took part in the study (49f, 19m). We measured F0 and intensity of the constituents.

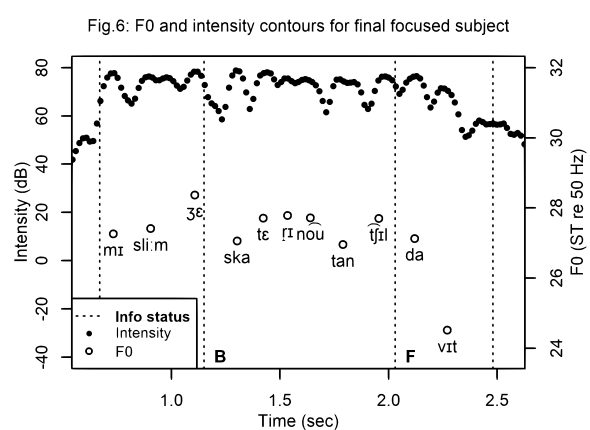
We fitted two linear mixed-effects models to analyze mean F0 and intensity median, using info status, focus position, and their interaction as fixed effects; and random intercepts and slopes for participants, along with random intercepts and interaction slopes for items as random effects. Focus in its default utterance-final position is not prosodically prominent relative to its background. Focus is prominent when occurring in a non-final position, followed by its (contextually given) background (interaction of info status and focus position;  $p < .001$  for F0 and intensity; see Figs. 1, 2, 5 & 6). The crossed interaction is due to two opposing simple effects: boosted and reduced prominence in focus and background, respectively, implying that a combination of focus prosodic prominence and givenness deaccentuation is at play in these structures, supporting the crucial role of relative prominence (Féry & Samek-Lodovici 2006 and Groeben et al. 2017). We fitted two other LMMs to predict mean F0 and intensity median, with focus type, focus position, and their interaction as fixed effects, and random intercepts for participants and items. There was no difference in answerhood vs. corrective focus encoding (no effect of focus type;  $p = 0.764$  for F0,  $p = 0.237$  for intensity). This did not change with word order (no interaction between focus type and focus position;  $p = 0.18$  for F0 and  $p = 0.827$  for intensity; see Figs. 3 & 4). We thus see that Czech speakers are sensitive to focus-as-alternatives (Rooth 1992), but fail to encode different focus uses.

**Conclusion** Czech possesses the capacity to encode focus (and givenness) prosodically, by modulating F0 and intensity, but systematically utilizes this option only for non-final foci. If final, focus remains prosodically non-prominent relative to its given background.



### Contour 1

(lit.) I think that [David]F [danced with Kateřina]B



### Contour 2

I think that [with Kateřina danced]B [David]F

## References

- Daneš, František (1957), *Intonace a věta ve spisovné češtině*, ČS AV.
- Féry, Caroline & Samek-Lodovici, Vieri (2006), Focus projection and prosodic prominence in nested foci, *Language* 82(1), 131–150.
- Firbas, Jan (1992), *Functional sentence perspective in written and spoken communication*, Cambridge University Press.
- Groebe, Lena & Šmík, Radek & Kügler, Frank (2017), Stress shift, focus, and givenness in Czech, in *FASL 24: The NYU Meeting 2015*, 180–199.
- Hamlaoui, Fatima & Żygis, Marzena & Engelmann, Jonas & Wagner, Michael (2018), Acoustic correlates of focus marking in Czech and Polish, *Language and Speech* 62(2), 358–377.
- Katz, Jonah & Selkirk, Elisabeth (2011), Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English, *Language* 87, 771–816.
- Krifka, Manfred (2008), Basic notions of information structure, *Acta Linguistica Hungarica* 55(3–4), 243–276.
- Mathesius, Vilém (1939), O tak zvaném aktuálním členění větném, *Slovo a slovesnost* 5(4), 171–174.
- Rooth, Mats (1992), A theory of focus interpretation, *Natural Language Semantics* 1(1), 75–116.
- Sgall, Petr & Hajičová, Eva & Buráňová, Eva (1980), *Aktuální členění věty v češtině*, Academia.
- Siewierska, Anna & Uhlířová, Ludmila (1998). An overview of word order in Slavic languages, in A. Siewierska (ed), (1998), *Constituent Order in the Languages of Europe*, De Gruyter Mouton, 105–150.
- Šmík, Radek (2024), Prosodic reflexes of information structure, in D. Šipka & W. Browne (eds), (2024), *The Cambridge Handbook of Slavic Linguistics*, Cambridge University Press, 104–126.
- Šmík, Radek & Wierzba, Marta (2017), Expression of information structure in West Slavic: Modeling the impact of prosodic and word-order factors, *Language* 93(3), 671–709.
- Zybatow, Gerhild & Mehlhorn, Grit (2000), Experimental evidence for focus structure in Russian, in T. H. King & I. A. Sekerina (eds), (2000), *Proceedings of the Annual Workshop on Formal Approaches to Slavic Linguistics*, 414–434.

# Asymmetries in grammatical marking of focus and content interrogatives: A cross-linguistic study

Ante Petrović

University of Zagreb

**Keywords:** information structure; syntax; typology; focus construction; content interrogative; construction asymmetry

This talk investigates asymmetries in grammatical marking of narrow focus and content interrogatives that are conditioned by the syntactic function of the narrowly focal or the interrogative phrase. The goals are twofold: 1) to provide a cross-linguistic view of the asymmetries, and 2) to bring forth an empirical argument to the discussion whether content interrogatives are a type of focus constructions by comparing the two construction types.

The study assumes a non-derivational view of syntax and adopts Lambrecht's (1994: 213) definition of focus as the non-presupposed semantic component of a proposition expressed by a clause.

Zimmermann (2011: 1203) and Velleman (2014: 186–188) have discussed asymmetries in obligatoriness of focus marking in a cross-linguistic perspective. Zimmermann hypothesises that such asymmetries commonly feature obligatory marking of narrow subject focus.

This investigation aims to shed light on asymmetries in grammatical focus marking based on a dataset spanning 55 languages from 22 families and five continents. The research questions are: 1) what kinds of asymmetries are attested, 2) which syntactic functions tend to trigger special treatment cross-linguistically, 3) whether asymmetries tend to involve both declarative narrow focus constructions and content interrogatives. Zimmermann's and Velleman's claims motivate the hypothesis that it is the grammatical relation associated with the most actor-like argument that is singled out for obligatory or additional marking. Furthermore, the assumption that content interrogatives are a type of focus construction (cf. Horvath 1986: 118–122; Lambrecht 1994: 282–286; Croft 2022: 375) bears out the hypothesis that asymmetries should involve both declarative focus and content interrogative constructions. Prosody is not considered.

In addition to splits in obligatoriness of marking, the data attest to two kinds of asymmetries: certain syntactic functions triggering additional marking, and distinct constructions used for different syntactic functions. Splits in obligatoriness of marking for narrow focus in declaratives are found in K'iche' and colloquial French, where only the most actor-like grammatical relation is obligatorily marked. Content interrogatives feature a split in obligatoriness of marking in Min Bei Chinese and Central Khmer, where only 'why' interrogative phrases must be marked. Six languages in the sample feature a special verb form in clauses where either the narrowly focal or the interrogative phrase is of a certain syntactic function. This function is the most actor-like relation in four languages: K'iche', Yucatec Mayan, Benchnon and Somali. On the other hand, it is the object in Amharic and adjuncts in Oluta Popoluca. Finally, Huehuetla Tepehua and Semelai use one construction to mark subject and certain adjunct interrogative phrases and another construction to mark object and other adjunct interrogative phrases.



In conclusion, the data show that the relation associated with the most actor-like argument triggering obligatory or additional marking is a frequent but not universal pattern. Furthermore, the assumption that content interrogatives are categorically a type of focus construction is not supported by the non-correspondences between these two construction types in obligatoriness splits and asymmetries involving different constructions for different syntactic functions.

## References

- Croft, William. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316145289>.
- Horvath, Julia. 1986. *FOCUS in the theory of grammar and the syntax of Hungarian*. Dordrecht: Foris.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Velleman, Leah Bridges. 2014. *Focus and movement in a variety of K'ichee'*. Austin: University of Texas, PhD thesis.
- Zimmermann, Malte. 2011. The grammatical expression of focus in West Chadic: Variation and uniformity in and across languages. *Linguistics* 49(5). 1163–1213. <https://doi.org/10.1515/ling.2011.032>.

## Contrast and *focus* sub-types across European languages: an experimental study

The relationship between *focus* and contrast has emerged as a key topic in studies on information structure. Building on the framework of Alternative Semantics [1], many scholars argue that contrast should not be treated as a distinct notion from *focus*, but as an inherent aspect of it, manifesting in various intensities. Under this perspective, contrast is not either absent or present (and *focus* either neutral or contrastive), but rather develops along a *continuum*, resulting in *focus* sub-types marked through distinct linguistic configurations [2, 3, 4, 5, 6].

This hypothesis has been tested in several languages, with a noteworthy argument presented by [7]. The study suggests that in languages such as Italian and French, which use both syntax and prosody to encode information structure, it is the additive combination of these two strategies that establishes the *continuum* of contrast.

The present study aims to extend this hypothesis by examining *focus*-marking strategies in German and Russian. Although *focus* in these languages has been largely studied [8, 3, 9, 10, 13, 14], the novelty of our analysis lies in the use of the same protocol employed in [7], ensuring full comparability of the results.

In addition to the existing *corpus* of 35 Italian and French speakers from [7], data were collected from 5 speakers of German and 10 speakers of Russian. Using a task adapted from [11], participants produced six target utterances, in which *focused* subjects represent three levels of contrastiveness: broad *focus*, identification *focus*, and correction *focus*.

The data are being evaluated with a two-factor linear regression model. Preliminary findings show that the four languages studied exhibit differences in how linguistic markers are mapped onto specific *focus* subtypes. In Italian, prosody primarily signals identification, with syntax (clefts) added specifically to mark correction. In French, clefts serve as the primary marker for identification, with prosody layered on top in corrective utterances. In Russian, word order (e.g., *focus* fronting, postverbal subject) is the main strategy, with prosody added on canonical word orders to indicate correction. In German, *focus* is mainly delivered by prosody, through phonetic cues such as *f0*, intensity and glottalization; the differentiation between identification and correction is achieved by a specific modulation of pre-nuclear and nuclear prominences [12].

These results highlight two key aspects: first, they confirm that *focus* and contrast, along with *focus* subtypes created by their interaction, are signalled through distinct combinations of linguistic markers. Second, they reinforce the concept of a *continuum* of contrast, rather than pointing to a binary distinction between contrastive and non-contrastive *foci*. As stated in [2], “the greater the contrast, the greater the potential,” and, as our findings suggest, the more linguistic means are employed to mark it. Crucially, Italian, French, German, and Russian demonstrate varying patterns of either additivity or trade-off between prosody, word order, and syntax, all of which, in some way, fit into this model. Our findings also underscore the need for future research to apply a repeatable and replicable methodology across a broader range of languages, fostering robust cross-linguistic comparisons.

## References

- [1] Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics*, 6:75–116.
- [2] 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):1–30.
- [3] Féry, Caroline. 2013. Focus as prosodic alignment. *Natural Language & Linguistic Theory*, 31(3):683–734.
- [4] Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4):243–276.
- [5] Molnár, Valeria. 2002. Contrast from a contrastive perspective. *Language and Computers*, 39:147–162.
- [6] Vallduví, Enric & Vilks, Maria. 1998. On rheme and kontrast. In Peter Culicover & Louise McNally, (eds.), *The Limits of Syntax*, Academic Press, New York, 79–108.
- [7] De Paolis, Bianca Maria. 2024. Focus-induced variations in prosody and word-order in native and non-native Italian and French. Doctoral dissertation, Università di Torino / Université Paris 8.
- [8] Büring, Daniel. 1997. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London, Routledge.
- [9] Féry, Caroline. 1993. *German intonational patterns*. Tübingen: Niemeyer.
- [10] King, Tracy Holloway. 1995. *Configuring topic and focus in Russian*. Stanford, CA: CSLI Publications.
- [11] Gabriel, Christoph. 2010. On Focus, Prosody, and Word Order in Argentinian Spanish. A Minimalist OT Account. In *Revista Virtual de Estudos da Linguagem*, Special issue 4:183–222.
- [12] Roessig, Simon. 2024. The inverse relation of pre-nuclear and nuclear prominences in German. *Laboratory Phonology*, 15(1): 1–43.
- [13] Bryzgunova, Elena Andreevna. 1983. *Zvuki i intonacija russkogo jazyka* (6th edition). Russkij jazyk, Moskva.
- [14] Janko, Tat'jana Evgen'evna. 2008. *Intonacionnye strategii russkoj reči v sopostavitel'nom aspekte*. Jazyki slavjanskichkul'tur, Moskva.

# **Focus marking. A typological overview from the *Catalogue of Information Structure in the World's Languages***

Damiana Porcellato & Luigi Talamo  
(Universität des Saarlandes)

Keywords: Focus, Information structure, Typology, Language variation

Among the main domains identified in information structure analysis, focus is one of the most widespread and relevant, as it refers to the expression of new information and is generally assumed to designate the element of the clause conveying the most prominent information (Matić 2015).

As is commonly known from the study of individual languages, a variety of specialized grammatical and lexical means signal focus, linking information-structural encoding with various formal features of language. Strategies associated with focus are, for example, prosody as in English, syntactic constructions like clefts as in French, word order variations as we find in Hungarian (Miller 2006), and morphemes such as the agreement suffix *-mə/ə* in Yukaghir (Matić 2015). The formal expression of focus thus relies on the use of available grammatical means.

Aiming to a sample of up to 25 languages belonging to 13 language families, we aim to present data on how information structure is conveyed across these languages, and, more specifically, which grammatical features are involved in the domain of focus. This typological study provides an overview of how strategies as well as categories related to information structure are expressed cross-linguistically, enabling a deeper analysis of the lexical and grammatical means involved in the domain of focus. Our study pays special attention to the polyfunctionality attested among focus-related markers, such as Hindi particles *bhi* and *hi*, which are generally used for emphasis in addition to marking focus on preceding constituents (Butt 2014, Choudury 2015), and Māori particle *ko*, which serves a range of functions, including topic as well as focus marking (Calhoun et al. 2016).

Our data is closely tied to the development of a new cross-linguistic database, the *Catalogue of Information Structure in the World's Languages*, which contains extensive information on information-packaging devices across the languages of our sample. The *Catalogue* is a literature-based collection of sources on individual languages, and it is meant to provide researchers with detailed information regarding the grammatical means used to encode the main domains identified in information structure analysis: focus, topic, givenness (as in Krifka & Musan 2012) and word order variation. The development of this tool is of broad interest due to its accessibility through a web application and capacity for real-time updates.

In light of the recent discussions regarding the notion of focus as a cross-linguistic valid category (Matić and Wedgwood 2013) and alternative proposals for a 'bottom-up study of Information Structure' (Ozerov 2021), our study contributes to the field and has the potential to shed new light on this research domain from a cross-linguistic perspective.

## References

- Butt, Miriam (2014), Questions and Information Structure in Urdu/Hindi, in M. Butt, and T. Holloway King (eds), *Proceedings of the LFG14 Conference*. CSLI Publications.
- Calhoun, Sasha, Yui, Naoko, and Karena Kelly (2016), Change in Māori focus/topic *ko*: the impact of language contact on prosody, in C. Carignan, and M. D. Tyler (eds), *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology*. ISSN 2207-1296.
- Choudury, Arunima (2015), *Interaction between Prosody and Information Structure: Experimental evidence from Hindi and Bangla*. Doctoral Dissertation. USC.
- Krifka, Manfred, and Renate Musan (2012), Information Structure: Overview and Linguistic Issues, in M. Krifka, and R. Musan (eds), *The Expression of Information Structure*, Berlin/Boston: De Gruyter, 1–44.
- Matić, Dejan, and Daniel Wedgwood (2013), The Meanings of Focus: The Significance of an Interpretation-Based Category in Cross-Linguistic Analysis, *Journal of Linguistics* 49 (1): 127–163.
- Matić, Dejan (2015), Information Structure in Linguistics, in J. D. Wright (eds), *International Encyclopedia of the Social and Behavioral Sciences. 2nd Edition*, Oxford: Elsevier, 95–99.
- Miller, Jim (2006), Focus in the Languages of Europe, in G. Bernini, and M. L. Schwartz (eds), *Pragmatic Organization of Discourse in the Languages of Europe*, Berlin/New York: Mouton de Gruyter, 121-214.
- Ozerov, Pavel (2021), Multifactorial Information Management (MIM): Summing up the emerging alternative to Information Structure, *Linguistics Vanguard* 7 (1), 20200039.

## Focus on Topics

Ekkehard Koenig & Letizia Vezzosi  
(FU Berlin & Universität Freiburg) (Università di Firenze)

Keywords: (multiple) focus, focus markers, contrastive topics, intensifiers, Germanic languages

Our point of departure are sentences like (1-4) found in Germanic languages.

Apart from matters of style and register, the sentence pairs in these examples are equivalent and have the following properties (cf. Jacobs, 1996; Koenig & Siemund, 1999):

- They both involve double focusing characterized by a fall-rise (tone) on the topic (indicated by √) and a falling tone (\) on the second focalized expression.
- The scope of the two focalized expressions is the reverse of that indicated by the linear order. The scope of the negation includes the rest of the sentence.
- In both the (a) and (b) members of these sentence pairs a person A is opposed to an alternative B (such that B is a function of A, e.g. A. Per vs. B. Per's wife in (3)).
- The two members of each pair differ formally in so far, as in (a) only the emphatic marker ('intensifier', 'emphatic') occurs in first position.
- Data of type (a) seem to be a specific feature of Germanic languages, but are also found in French, whereas constructions of type (b) are found in a wide variety of European languages

On the basis of relevant previous work, we will try to answer the question whether the so-called 'intensifiers' ('emphatics'), i. e. expressions like Germ. *selbst*, Engl. *X—self*, It. *stesso*, Sp. *mismo*, which are generally assigned to a class of their own (cf. 5), can be subsumed under the more general class of focus markers, analogously to the expressions discussed in Koenig 1991), Siemund 2000 and Koenig & Gast 2006.

The following observations support such an analysis:

- the so-called 'intensifiers' occur as topics in sentences with double focusing ((1-4).
- the relevant focalized expression ((1)a-(4)a) and the post-posed one ((1)b-(4)b) evoke alternatives to the denotation of the focus associate, just like their unaccented prenominal counterparts (Germ. *selbst*, Swed. *själv*, Fr. *lui-même* vs. *même*). The alternatives in question, however, are functions of the value given (6).
- In languages which combine reflexive pronouns with 'intensifiers' the alternatives evoked are more likely and potential, rather than real values (7).
- In adverbial positions, the relevant expressions may take wide scope and narrow scope over a negation, just like auch 'also', and nur ('only') (8).
- The so-called 'emphatics' allow multiple occurrences, unlike adverbs, but like focus markers (9).
- Just like focus markers, 'intensifiers' can also be used as topic markers (1-4).
- Both focus markers and the so-called 'intensifiers' may manifest sortal restrictions in combining with associates: le président lui-même vs. aujourd'hui même; \*Selbst ist das Buch nicht interessant.

A successful attempt to subsume intensifiers under a more comprehensive category allows us to provide more general and precise characterization of function words typically assigned to the 'dustbin' category adverb. Moreover, we will show that it allows us to raise new questions and provide new answers.

### Examples

- (1) a. Engl. a. (My wife thinks we can go ahead) √Myself, I am \not in favor of the project b. I myself am not in favor of the project.  
b. Germ. (Mein Bruder wird Dir helfen.) a. √Selbst habe ich \keine Zeit. b. Ich /selbst habe \keine Zeit.  
'My brother will help you. I have not got the time myself'.

- (3) Swed. (Pers fru bor i Stockholm). a. √Själrv bor Per \inte i Stockholm. b. Per /själrv bor \inte i Stockholm. 'Peters wife lives in Stockholm. Peter himself does not live in Stockholm.'
- (4) Fr. (a) Lui-même, Paul n'habite pas à Paris. (b) Paul lui-même n'habite pas à Paris.
- (5) a. German: *selbst, eigen, leibhaftig, persönlich, höchstpersönlich, von selbst, von sich aus, an sich, höchstselbst, allein*, etc.  
 b. Italian: *stesso, in persona, proprio, personalmente, da sè, da solo, di suo*
- (6) a. Paul's father speaks English and French. Paul himself does not speak any foreign language.  
 b. ?? The president of Albania also speaks Italian and English. Paul himself does not speak any foreign language.
- (7) Man muss diesen Politiker vor sich SELBST schützen.  
 'This politician needs to be protected against himself.'
- (8) a. Karl hat sein Auto nicht selbst gewaschen. 'K. has not washed his car himself.'  
 b. Wie kann Karl sich über andere beklagen, wenn er Normen selbst nicht beachtet.  
 'How can K. complain about others, if he does not respect norms himself.'
- (9) John himself admires Mary herself, (rather than her husband). Only Trump admires only Trump.

## References

- Jacobs, Joachim (1996), Bemerkungen zur I-Topikalisierung, in Inger Rosengren (ed.), *Sprache und Pragmatik 41*. Lund: Lunds Universitet, 1- 48.
- Koenig, Ekkehard (1991), *The meaning of focus particles: A comparative perspective*. London: Routledge.
- Koenig, Ekkehard & Peter Siemund (1999), Intensifikatoren und Topikalisierung. Kontrastive Beobachtungen zum Deutschen Englischen und anderen germanischen Sprachen. In H. Wegener (ed.), *Deutsch kontrastiv. Typologisch vergleichende Untersuchungen zur deutschen Grammatik*. Stauffenburg, 87-110.
- König, Ekkehard & Volker Gast (2006). Focused assertion of identity: A typology of intensifiers. *Linguistic Typology* 10.2: 223-276.
- Krifka, Manfred (2008), Basic notions of information structure. *Acta Linguistica Hungarica* 55. 243–276.  
<https://doi.org/10.1556/ALing.55.2008.3-4.2>
- Siemund, Peter (2000), *Intensifiers in English and German*. London: Routledge.

## Clause-medial focus marking – the perspective from West-African languages

In this talk, we discuss clause-medial focus marking in different West-African languages from the Mabilia and Chadic families. In these languages, focus is marked by morphological particles, which mark fronted foci (Mabilia and Chadic) and in situ foci (Mabilia). We are pursuing two main goals, to (i) develop a typology for the different morphological focus marking strategies; (ii) show that morphological focus marking leads to massive ambiguity.

Focus marking in many European languages mainly relies on prosody, i.e. the nuclear pitch accent of the clause being placed on the focused constituent. In addition to this, a dedicated syntactic position can be used as well, especially for contrastive or mirative focus types (Cruschina 2012), which is mainly left peripheral (Italian, English) or sentence final (Spanish, Italian).

The West-African languages under consideration in this talk do not employ phonology for focus marking, mainly because they are tone languages, but rely on other means, focus marking by morphological particles, which occur ex-situ as well as typologically exceptionally in-situ, cf. (1) from Dagbani (Mabilia) (Olawsky 1999, Issah 2020).

- (1) a. Noo **ka** Adam korigi.      b. Adam korigi      **la** noo.  
fowl FOC Adam slaughter      Adam slaughterer FOC fowl  
'Adam slaughtered a FOWL.'      'Adam slaughtered a FOWL.'

In the talk, we will discuss two classes of languages employing morphological focus markers. On the one hand, there are languages in which the marker attaches to the focused constituent itself (Type 1), (2). On the other hand, we also find languages in which the focus marker occurs in a fixed low position in the clause, following the verb (Type 2). This gives rise to the patterns in (3).

- (2) a. subject Tns V-Asp [object-1<sub>FOC</sub> **FM**] object-2  
b. subject Tns V-Asp object-1 [object-2<sub>FOC</sub> **FM**]
- (3) a. subject Tns V-Asp **FM** object-1<sub>FOC</sub> object-2  
b. subject Tns V-Asp **FM** object-1 object-2<sub>FOC</sub>

In both types of morphological focus marking languages, it seems to be the case that focus marking is subject to relatively strict constraints, as the focus marker, even in Type 1 languages, like Dagbani in (1b), must be on the main clausal projection line. Thus, direct marking of, for example, possessors or adjectives, is simply impossible. This restriction, as well as the fixed position of the marker in Type 2 languages, leads to massive focus ambiguity, as the same position of the focus marker is compatible with very different focus interpretations (e.g. focus on possessor, possessum, NP or VP), see (4) from Likpakpaanl (Mabilia) (Acheampong 2024), a type 2 language, illustrating the



possessor / possessum ambiguity. Note that (4) is also compatible with a VP-focus interpretation.

- (4) Mari nan kər        Piita    aa-kɔla    **la**.  
Mari PST slaughter Peter POSS-fowl FOC  
a. ‘Mary slaughtered PETER’s fowl.’  
b. ‘Mary slaughtered Peter’s FOWL.’

Such cases of focus ambiguity highlight the importance of context and pragmatic reasoning for focus interpretation in the West-African languages.

## References

- Acheampong, Samuel, O. 2024. The syntax of Likpakpaanl elliptical constructions. PhD dissertation, Frankfurt University.
- Cruschina, Silvio 2012. *Discourse-Related Features and Functional Projections*. Oxford: Oxford University Press.
- Issah, Samuel A. 2020. *On the structure of A-bar constructions in Dagbani: Perspectives of wh-questions and fragment answers*. Frankfurt am Main: Peter Lang.
- Olawsky, Knut. 1999. *Aspects of Dagbani grammar with special emphasis on phonology and morphology*. Munich: Lincom.

## From syntax to focus and focus to syntax?

Marianne Mithun

(University of California, Santa Barbara)

Keywords: clefts, contrast, diachrony, focus, prosody

As is well known, focus constructions are marked formally in various ways: prosodically, morphologically, and/or syntactically. Their functions have also been classified in various ways, i.e. in terms of scope (narrow/broad), alternatives (information focus/contrastive focus) and more (Lambrecht 1994, Krifke 2007, Büring 2010, Zimmermann & Onea 2011, Güldemann et al. 2015, van der Wal 2016, Zimmermann 2016, Aissen 2023). But formal devices marking focus sometimes serve other functions as well. The lack of strict form-function isomorphism has raised questions about the categoriality of the notion itself. Some of the apparent fuzziness might make sense if we can uncover trajectories of development linking the various functions.

This possibility is explored here with examples from Mohawk, a North American Iroquoian language. Here focus is marked prosodically, morphologically, and syntactically. The basic focus construction consists of an intonation unit beginning with a pitch reset and coherent descending pitch contour, with the focused constituent in initial position at the pitch peak, typically with extra-high pitch. Contrastive focus is marked with a particle after the initial focused constituent.

Developmental trajectories can be discerned for both constructions. There are no philological records sufficiently deep to document stages of development, nor does comparison of related languages allow reconstruction by the comparative method, since the same constructions occur in all of them. But internal reconstruction provides clues. The contrastive focus construction apparently developed from a bi-clausal cleft construction; the modern particle originated as anaphoric demonstrative plus copula *né*: 'that' + 'e is'. This particle now also marks contrast on its own. The basic focus construction has developed further into basic syntactic structure. Constituents are now ordered in descending order of importance to the message at that point rather than their syntactic role. The modern verb morphology apparently originated in basic SOV clause structure. When unstressed pronouns fused with following verbs as prefixes, the pragmatic force of earlier weaker focus constructions, presumably with less prosodic prominence, began to fade. Non-focal elements no longer need be presupposed: they are simply relatively backgrounded.

Ultimately, the apparent lack of sharp categoriality can be less surprising with the recognition that languages are constantly evolving.

### References

- Aissen, Judith (2023), Documenting topic and focus, in P. Jens and L. Michael (eds), *Language Documentation and Conservation Special Publication 26*: 11–57. <http://hdl.handle.net/10125/75032>
- Büring, Daniel (2010), Toward a typology of focus realization, in Malte Zimmermann and Caroline Féry (eds), *Information Structure*. Oxford: Oxford University Press, 177–205.
- Güldemann, Tom, Sabine Zerbian, and Malte Zimmermann (2015), Information Structure with Special Reference to Africa. *Annual Review of Linguistics* 1(1):155–178.
- Krifka, Manfred (2007), Basic notions of information structure, in C. Féry, G. Fanselow, and M. Krifka (eds), *Interdisciplinary Studies on Information Structure*, ed. Potsdam: Universitätsverlag.
- Lambrecht, Knud (1994), *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- van der Wal, Jenneke (2016), Diagnosing focus. *Studies in Language* 40.2:259–301.

Zimmermann, Malte and Edgar Onea (2011), Focus marking and focus interpretation. *Lingua* 121:1651–1670.

## Particle positions and focus effects: An Edoid case

Ronald P. Schaefer & Francis O. Egbokhare  
Southern Illinois University Edwardsville & University of Ibadan

Keywords: secondary auxiliary focus, paired focus particles, postposed and preposed focus particles, multi-element focus, Emai, Edoid

Focus effects continue to garner attention (Zimmermann and Féry 2010, Féry and Ishihara 2016, Assmann et al. 2023) despite reservations by some about the viability of focus as a definable linguistic category (Matić and Wedgwood 2013). Cross-linguistically, focus effects can be signaled prosodically, morphologically, and/or syntactically. We assess a sample of focus effects in Emai, an under-analyzed member of West Africa's Edoid group. Its unmarked default focus is shown by the question-answer pair below, with VP elements equivalent to 'sell yam' in focus.

<i>é<sup>↓</sup>mé</i>	<i>óje</i>	<i>↓ú-ì?</i>	<i>ɔ́</i>	<i>↓shén</i>	<i>émà.</i>
what	Oje:DST	PST:do-PFV	3SG:DST	PST:sell:PFV	yam
'WHAT did Oje do?'			'He SOLD YAM.'		

In addition to unmarked default focus, Emai has focus particles that postpose, prepose, or both. Some particles also distinguish polarity. Under narrow contrastive focus, particle *li/nì* (affirmative) or *kì* (negative) postpose to a clausal constituent that is fronted (*émà lí óje ló shén ákhò* [yam PF Oje:DST FUT sell:PFV tomorrow] 'It is YAM that Oje will sell tomorrow'). The CV particle in such constructions prosodically phrases with not-in-focus elements, since vowel (*i*) elides and the consonant cliticizes to clausal subject (i.e. *óje*). Under broad focus, a particle preposes to an entire clause, e.g. affirmative *é<sup>↓</sup>rí* and negative *kí*, as in *kí óje ↓fí jí ókpósó úkpóràn ò* [ SN Oje:DST PST:hit:PFV ART woman stick INT] 'It isn't that OJE HIT THE WOMAN WITH A STICK, you know.' With the entire clause in focus no particle prosodic phrasing occurs.

Polarity and the broad/narrow distinction aside, Emai secondary auxiliaries (SAs) mark focus with identical twin particles, either additive phasal aspect *gbò* 'also, too' or distal manner demonstrative *ì<sup>↓</sup>yó* 'like that.' Each particle pair occurs as preposed and postposed to a focused SA, e.g. *dégbè* 'carefully' in *óje ↓gbó dégbè gbó é jí émàè* [Oje:DST PST:ADD carefully ADD eat ART food] 'Oje REALLY CAREFULLY ate the food' or *dóbó òì* subject reflexive 'by himself' in *óje í<sup>↓</sup>yó dóbó òì í<sup>↓</sup>yó híán jí órán* [Oje:DST PST.DMD SREFL 3SG DMD cut the wood] 'Oje ALL BY HIMSELF cut the wood.' Particle pairs in these constructions do not prosodically phrase with a following element; their vowels do not elide.

Corresponding SA constructions without a particle pair mark do not engage focus. Their clause exhibits only default focus, as with *dégbè* (*óje ↓dégbè é jí émàè* [Oje:DST PST:carefully eat ART food] 'Oje carefully ATE THE FOOD') and *dóbó òì* (*óje ↓dóbó òì híán jí órán* [Oje:DST PST.SREFL 3SG cut the wood] 'Oje himself CUT THE WOOD').

A SA sequence can also manifest an extended or multi-element focus via multiple particle pairs where at least one particle is shared: *ójé ɖzá ìɖyɔ gbó dégbè gbó dóbó òì gbó é ɔlí émàè* [Oje:DST PST:RES DMD ADD carefully ADD SREFL 3SG ADD eat ART food] ‘As a result Oje that way REALLY CAREFULLY ALL BY HIMSELF ate the food.’ We note in conclusion that Emai primary auxiliary forms that convey TAMP do not accept focus particles, thereby contrasting with focus effects of auxiliary TAMP forms in other Benue-Congo languages (Hyman and Watters 1984, Hyman and Polinsky 2010, Downing and Hyman 2016).

## References

- Assmann, Muriel, Daniel Buring, Izabela Jordanoska, and Max Prüller (2023), Towards a Theory of Morphosyntactic Focus Marking, *Natural Language & Linguistic Theory* 41, 1349-1396.
- Downing, Laura J, and Larry M. Hyman (2016), Information structure in Bantu, in Caroline Féry, and Shinichiro Ishihara (eds), (2016), *The Oxford Handbook of Information Structure*, Oxford: Oxford University Press, 790-813.
- Féry, Caroline, and Shinichiro Ishihara (eds.) (2016), *The Oxford handbook of information structure*. Oxford: Oxford University Press.
- Hyman, Larry M, and John R. Watters (1984), Auxiliary Focus, *Studies in African Linguistics* 15, 233-273.
- Hyman, Larry M, and Maria Polinsky (2010), Focus in Aghem, in Malte Zimmermann, and Caroline Féry (eds), (2010), *Information Structure: Theoretical, Typological, and Experimental Perspectives*, Oxford: Oxford University Press, 206–233.
- Matić, Dejan, and Daniel Wedgwood (2013), The Meanings of Focus: The Significance of an Interpretation-based Category in Cross-Linguistic Analysis, *Journal of Linguistics* 49(1), 127-163.
- Zimmermann, Malte, and Caroline Féry (eds.) (2010), *Information structure: Theoretical, typological, and experimental perspectives*. Oxford: Oxford University Press.

## Pragmatics of *ta:n* Focus Particle in Tamil

Rajamathangi Shanmugasundaram  
(University of Lausanne)

**Keywords:** Focus, discourse particle, common ground, reactivation, ellipsis.

This study analyzes the particle *ta:n* in Tamil, a Dravidian language, focusing specifically on its pragmatic role as a discourse particle. The particle *ta:n* serves multiple functions in Tamil, many closely related to the grammatical expression of focus. Murugaiyan (2009) provides a detailed discussion on its origin and function as a focus marker in different contexts, while Velupillai (1981) identifies its use as an emphatic clitic in cleft constructions.

I illustrate the role of *ta:n* in fragment stripping (an elliptical phenomenon), as shown in (1), where it emphasizes previously mentioned entities and reactivates them in the common ground (CG). In such elliptical constructions (the utterance uttered by another speaker), the subject of the fragment obligatorily takes the additive marker *-um* and the focus particle *ta:n*.

- (1) a. **[Speaker A]:** *ʃɔ:n* [sku:l-kku po:-n-a:n].  
John[3SGM.NOM] school-DAT go-PST-3SGM  
John went to school.'
- b. **[Speaker B]:** *me:rij-um ta:n [e]*  
Mary[3SGF.NOM]-ADD TA:N EP  
Mary, too.'

In my understanding, *ta:n* is related to the emphasis effect observed by Sperber and Wilson (1996), who analyze repetition in discourse as creating emphasis. Similarly, Jackson (2016) discusses how repetition by the same speaker reinforces the truth of a proposition. While repetition in yes/no question–answer contexts is often analyzed as polarity focus (Goodhue 2018, Wilder 2013), I argue that *ta:n* cannot be solely analyzed as polarity focus because its use in dialogue not only emphasizes the truth of a proposition but also reactivates a referent in the common ground (CG), signaling its relevance in the current discourse.

This pragmatic use of *ta:n* observed in non-elliptical contexts, its presence is optional but impactful. Consider the examples in (2): in (2-a), *ta:n* reactivates an assurance made by the speaker, adding a “don’t worry” flavor. Without *ta:n*, as in (2-b), the assurance is not as emphatic.

- (2) **Context:** Somu had a maths examination today. He finds mathematics difficult, but he has worked very hard to secure a pass mark. In the course of a discussion about his performance, his father says as in (2-a), for which his mother reassures the other that he would do well, uttering (2-b).
- a. *avan exam epadi paṇṇi iru-pp-a:n-o.*  
3SGM.NOM exam how do be-WILL-3SGM-DISJ  
'How he would have done the exam!'
- b. *avan nalla ta:n paṇṇi iru-pp-a:n.*  
3SGM.NOM good TA:N do be-WILL-3SGM-COMP

- ‘He would have done good for sure (don’t worry).’
- c. avan        nalla paṇṇi iru-pp-a:n.  
 3SGM.NOM good do    be-WILL-3SGM-COMP  
 ‘He would have done good.’

I compare this role of *ta:n* to the German discourse particle *doch*, which reactivates facts already part of the CG but potentially forgotten or doubted by the listener (Zymla et al. 2015). However, unlike *doch*, which has been argued to be a modal particle, *ta:n* is emphatic in nature and historically derived from a reflexive/anaphoric marker widely attested across Dravidian languages.

According to Zimmermann (2011), *doch* signals that the speaker assumes the proposition *p* to be inactive in the current discourse because the addressee may have forgotten or doubted it. Similarly, in Tamil, *ta:n* in fragment stripping highlights entities introduced by one speaker and reactivated by another.

This distributional and pragmatic behavior of *ta:n* supports its analysis as a discourse particle consistent with its role in marking contrastive and reactivated focus. To address its syntactic status, I follow Rizzi (1997) in assuming that *ta:n* is merged in  $\text{Foc}^0$ , within the left periphery of the clause.

## References

- Jackson, Rebecca Claire (2016) “The Pragmatics of Repetition, Emphasis, and Intensification”. PhD thesis. University of Salford.
- Murugaiyan, Appasamy (2009) “Focus constructions in Modern Tamil”. In: *International Journal of Dravidian Linguistics* 38.2, pp. 43–69.
- Rizzi, Luigi (1997) “The Fine Structure of the Left Periphery”. In: *Elements of Grammar*. Ed. by Liliane Haegeman. Springer, 1997, pp. 281–337.
- Sperber, Dan and Deirdre Wilson (1996) *Relevance: Communication and Cognition*. 2nd ed. Wiley-Blackwell, 1996, p. 338.
- Velupillai, A. (1981) “Cleft sentences in Tamil”. In: *The Sri Lanka Journal of the Humanities* 7.
- Zimmermann, Malte (2011) “Discourse particles”. In: *Theoretical Linguistics* 37.1–2, pp. 39–74.
- Zymla, Mark-Matthias et al. (2015) “Modelling the Common Ground for Discourse Particles”. In: *Proceedings of the 20th International Lexical Functional Grammar Conference (LFG15)*. 2015, pp. 420–440.

## *Fare+Che+Infinitives* in Turinese Italian: Focus-Sensitivity and Information Structure

Tommaso Sgrizzi  
(IUSS Pavia)

KEYWORDS: low-*vp*, infinitives, focus, turinese Italian, cartography

The "*fare+che+infinitive*" (FCI) construction is a syntactically intriguing and pragmatically rich structure found in Turinese Italian.

- (1) Se vai al supermercato fai che comprare i pomodori.  
If go.2SG to-the supermarket do.2SG that buy.inf the tomatoes.  
"If you go to the supermarket, buy the tomatoes."

(1) is used when the subject can perform an action that would otherwise be postponed or is unnecessary at that moment. The combination of the complementizer "*che*" (that) and an infinitive verb raises crucial questions about how an infinitive can be licensed by a finite complementizer (Satik, 2022).

Different structural tests (e.g., availability of low topicalization and focalization; interaction with negation, tense, and modality) show conflicting results wrt a biclausal or monoclausal analysis. The FCI is not compatible with stative verbs; it resists clitic climbing, but allows other transparency effects (e.g., wrt tense and mood); and the whole FCI constitutes a unique binding domain.

Here we set aside the nature of *che* (we propose that it realizes a high functional head within the low-*vP* periphery, following Belletti, 2004 – mediating the requirements of the causative verb *fare* regarding the eventuality type of the embedded dynamic VP, much like the high Force head expresses clausal type, see Rizzi, 1997). To account for the context in which the FCI is typically uttered, we argue that the entire TP is focalized (main clause-focus) in line with (Greco, 2019). This yields the interpretation in (2), where other commonly performed alternatives are negated. Notice the incompatibility of the FCI in unmarked contexts (3).

- (2) Visto che mi sono svegliato presto, [<sub>FocP</sub> [<sub>TP</sub> ho fatto che fare una corsa] Foc<sup>o</sup> [... [<sub>TP</sub> ...  
Since I got up early, [I went for a run]<sub>Foc</sub> **¬(Having breakfast, Showering, etc..)**  
(3) \*Ogni giorno [faccio che andare a lavoro alle 8].  
\*Every day [do.1<sup>ST</sup>SING.PRES that go.INF to work at 8]

As predicted by our hypothesis, main clause focalization in the left periphery is not permitted, since the relevant structural position is already occupied by the TP.

- (4) \*Visto che avevo tempo, [IL LATTE]<sub>Foc</sub> ho fatto che prendere (*non il vino*)  
\*Since I had time [MILK]<sub>Foc</sub> have.1<sup>ST</sup>.SG do.PSTPRT that buy.INF (*not wine*)

The incompatibility of the FCI with negation naturally follows by our analysis, as the TP is internally merged into a position higher than that of negation (Foc).

- (5) Visto che sono di buon umore, (\*non) faccio che (non) rimproverare nessuno.  
Since I am in a good mood (\*not) do.1<sup>ST</sup>.SING.PRES that (not) scold.INF nobody

In (5) we see that negation is licensed only within the focalized TP, as the higher negation before *fare* is not able to license the NPI *nessuno*. Other aspects will be addressed, such as the role of the embedded subject and the theory of control, passivization, and the comparison with the structure in (6), present in Standard Italian but showing several differences with the FCI, although the apparent similarities.



- (6) Non faccio (altro) che mangiare dolci ultimamente.  
Not do.1<sup>ST</sup>.SING.PRES (else) that eat.INF sweets lately  
'I do nothing but eat sweets lately'

WORD COUNT WITHOUT TITLE, KEYWORDS, AND REFERENCES: 499

## References

- Belletti, A. (Ed.). (2004). *Structures and beyond: Volume 3: The cartography of syntactic structures*. Oxford University Press.
- Greco, M. (2019). *The syntax of surprise: Expletive negation and the left periphery*. Cambridge Scholars Publishing.
- Rizzi, L. (1997). The Fine Structure of the Left Periphery. In *Kluwer International Handbooks of Linguistics* (pp. 281–337). Springer Netherlands. [https://doi.org/10.1007/978-94-011-5420-8\\_7](https://doi.org/10.1007/978-94-011-5420-8_7)
- Satik, D. (2022). The maximal size of infinitives: A truncation theory of finiteness. *Lingbuzz*/005910. <https://doi.org/lingbuzz/005910>

## **Focus and Fronting in Guadeloupean Creole: Examining the role of complementation in focus**

William Kezerian, Vasily Tselioudis, Béatrice Jeannot-Fourcaud, Isabelle Barrière & Renauld Govain  
(Syracuse University, Columbia University, University of the Antilles,  
Saint Elizabeth University & State University of Haiti)

Keywords: focus, syntax, creole, subordination, fieldwork

Guadeloupean Creole is spoken in the French overseas department of Guadeloupe, which has been historically minoritized and whose general linguistic descriptions (Bérnabe 1983, Damoiseau 2012, and Bonan 2013) lack empirical evidence of its notably unconventional focus constructions. Considering its relatively recent recognition as a regional language of France, our team had the opportunity to conduct linguistic fieldwork on the syntax of its focus constructions, particularly whether speakers allow focusing on noun phrases and verbs that are situated within a complement clause. Focus in this language requires reduplication and fronting of focus-marked verbs, however, only fronting when focusing nouns (Byrne et al., 1993). Thus, the present study aims to examine factors that prohibit or allow the movement of a focused NP or verb across a complementizer phrase, which would appear to contrast with the notion that movement of a phrase to outside of a CP boundary is illicit (Haegeman 1991, Buring 2009).

We administered randomized Guadeloupean Creole (GC) acceptability judgment tasks to 34 bilingual native speakers of GC and French using PsychoPy2024.1.5, collecting 1,674 data points. Participants listened to GC audio prompts with and without contrastive focus, after which they ranked the focus-marked sentence on a three-point acceptability scale. We also collected ethnographic interviews and oral narratives elicited with storyboards in both languages to compare our acceptability judgment tasks against natural usage, and to gather data on educational background, demographics, and other sociolinguistic factors that might affect focus-marking. Our results reveal a preference for focusing elements external to a CP, though a variety of factors contribute to the acceptability of focusing on CP-internal elements. These include the inclusion of the focus-marking morpheme “sé”, how agentive the noun is, the use of complementizers (Dixon 2006, Lohninger et al., 2020, and Tramutoli 2021), and the transitivity of the focused verb. We also observed that language proficiency and attitude towards each language influenced focus-marking acceptability. These findings suggest that focus movement of a noun or verb from inside a CP is indeed licit in Guadeloupean Creole, but relies on particular morphemes and the aligning of certain syntactic conditions to be acceptable. These findings lack any previous linguistic documentation in Guadeloupean Creole. Our research suggests that this particular Antillean Creole language layers multiple mechanisms of focus in a given phrase, and may have more transparent CP boundaries for focus movements when compared to traditional frameworks.

This work will contribute to the linguistic understanding of Antillean Creole languages’ syntactic complexity by documenting focus constructions unique to GC, a language that differs significantly from Haitian Creole and other French-based Creoles. We aim not only to expand the descriptive grammar of GC, but to situate these novel findings within modern understandings of movement barriers, focus typology, the role of complementation, and the syntactic behavior of Antillean Creole syntactic phrases.

## Examples

### NP Focus on a CP-internal noun

#### 1. a. Without focus

Sara ka kwè frè a-y ké gannyé kous-la.  
Sara PROG believe brother DET-3SG FUT win race-DET  
"Sara believes her brother will win the race."

#### b. With focus movement

(Sé) **kous-la** Sara ka kwè frè a-y ké gannyé.  
FOC **race-DET** Sara PROG believe brother DET-3SG FUT win  
"Sara believes her brother will win **the race**."

### VP focus on a CP-internal verb

#### 2. a. Without focus

David ka anonsé pawk-la ouvè jòdla.  
David PROG announce park-DET open today  
"David is announcing that the park opens today."

#### b. With focus on the CP-internal verb

(Sé) **ouvè** David ka anonsé pawk-la ouvè jòdla.  
FOC **open** David PROG announce park-DET open today  
"David is announcing that the park **opens** today."

### PP focus disambiguating PP attachment by using functional morphemes *ki* and *la*

#### 3. a. An vwè tifi-la ki ka travay évè-w la an boutik-la.

1SG see.PFV girl-DET REL PROG work with-2SG DET in store-DET  
"I saw the girl who you work with at the store."

#### b. (Sé) **an boutik-la** an vwè tifi-la ka travay évè-w. FOC **in store-DET** 1SG see.PFV girl-DET PROG work with-2SG "I saw the girl **at the store** who you work with."

#### c. (Sé) **an boutik-la** an vwè tifi-la ki ka travay évè-w la. FOC **in store-DET** 1SG see.PFV girl-DET REL PROG work with-2SG DEIC "I saw the girl who you work with **at the store**."

## References

- Bernabé, Jean (1983), *Fondal-natal: Grammaire basilectale approchée des créoles guadeloupéen et martiniquais*, vol. 1–3, Paris: L'Harmattan.
- Bonan, Céline (2013), *The core grammar of Guadeloupean Creole: A descriptive and comparative approach*, Master's thesis, Università Ca' Foscari Venezia.
- Büiring, Daniel (2009), *Towards a typology of focus realization*, in M. Zimmermann & C. Féry (eds), *Information Structure: Theoretical, Typological, and Experimental Perspectives*, Oxford: Oxford University Press. [Online edition, Oxford Academic, 1 Feb. 2010]

- Byrne, Francis, & Winford, Donald (eds) (1993), *Focus and Grammatical Relations in Creole Languages*, vol. 12, Amsterdam/Philadelphia: John Benjamins Publishing Company.  
<https://doi.org/10.1075/cll.12>
- Damoiseau, Robert (2012), *Syntaxe créole comparée*. Martinique, Guadeloupe, Guyane, Haïti, Paris: Karthala.
- Dixon, R. M. W., & Aikhenvald, A. Y. (eds) (2006), *Complementation: A Cross-Linguistic Typology*, Oxford: Oxford University Press.
- Haegeman, Liliane (1991), *Introduction to Government and Binding Theory*, First edition, Oxford: Blackwell Publishers.
- Hazaël-Massieux, Marie-Christine (1996), *Du français, du créole et de quelques situations plurilingues: Données linguistiques et sociolinguistiques*, Paris: Éditions Publibook.
- Koopman, Hilda (1984), *The Syntax of Verbs*, New York: Academic Press.
- Lohninger, Maximilian, & Wurmbrand, Susi (2020), Typology of complement clauses, *Linguistic Inquiry* 51(2), 353–398. [https://doi.org/10.1162/ling\\_a\\_00353](https://doi.org/10.1162/ling_a_00353)
- Tramutoli, Laura (2021), ‘Fact type’ complementizer in Guadeloupean Creole, *Journal of Pidgin and Creole Languages* 36(2), 336–361. <https://doi.org/10.1075/jpcl.00078.tra>
- Vaillant, Pascal (2009), *Induction of French structures into Creole grammar*, paper presented at the 8th Creolistics Workshop, Giessen (Germany), April 2009.

# Re-evaluating Wh-Movement in Focus Structures: Implications for Syntax, Boolean Algebra, and Cross-Linguistic Focus Sensitivity

Yoshio Endo

(Kanda University of International Studies)

Key words: focus, cartography, syntax, wh-movement, Boolean algebra

This paper revisits Bach & Horn's (1976) and Chomsky's (1977) foundational work on wh-movement in focus-sensitive structures, particularly examining why sentences like (1) are grammatical while sentences like (2) are not in the framework of the cartography of syntactic structures.

(1) Who did you see pictures of?

(2) ?Who did you burn pictures of?

This asymmetry, often explained by constraints against movement out of DP in the structure in [see • (pictures • who)], could also arise from an alternative structural analysis: that *who* is understood as directly merged outside DP, as in [(see • pictures) • who)].

I propose a novel interpretation supporting this explanation, highlighting how syntactic arrangements in English can be extended by findings in Japanese, where similar wh-movement focus phenomena allow for wh-element with accusative Case marking of V as in (3), while this accusative Case marking of *who* is impossible with *burn pictures of who*, as in (4) .

(3) John-o • (syasin-de • miru )

John-Acc picture-of see

(4) ?John-o • (syasin-de • moyasu )

John-Acc picture-of burn

This modern Japanese accusative Case particle *o* was a focus marker in Old Japanese. Even in modern Japanese, the element immediately left of the verb is interpreted as the focus (cf Belletti 2004). In other words, [(see • pictures) • who)], which allows for Wh-movement in English, can be represented as [John• (pictures •see)] in Japanese, where the direct object *John* is positioned immediately to the left of the complex predicate *see*

*pictures.*

Incidentally, in English, *burn pictures of who* becomes grammatical when *who* is D-linked as in (5). Similarly, in Japanese, accusative Case also becomes possible when a wh-phrase is D-linked.

(5) A: I heard you were really angry with your ex-boyfriend.

B: Yeah, I was.

A: What did you do?

B: Well... I burned some of his stuff.

A: Like what?

B: Old letters, photos...

A: **Who did you burn pictures of?**

This paper further relates these observations to Boolean algebra, demonstrating that the structural flexibility in [see • (pictures • John)] versus [(see • pictures) • John] aligns with the associative law in Boolean terms  $a \bullet (b \bullet c) = (a \bullet b) \bullet c$ , suggesting syntactic operations might reflect principles akin to mathematical laws. The properties of Boolean algebra are generally found in natural languages in coordinate structures. For instance, the distributive law  $a \bullet (b \bullet c) = (a \bullet b) \ \& \ (a \bullet c)$  is seen in [*John • (held • Mary) and (kissed her)*] = [(*John held Mary*)] & [(*John kissed her*)]. Similarly, the structure following the associative law discussed above can be interpreted as having a concealed coordinate structure, as *see pictures of John* implies a coordinate structural relationship like *see pictures & see John*. Wh-movement is impossible in English when the verb is like *burn*, and concealed coordination cannot occur in Japanese for such cases. Neither can the wh-element be accompanied by an accusative Case particle. This offers a fresh perspective on syntactic computation, i.e., operations in the syntactic component can borrow mathematical laws, blurring syntax and mathematics boundaries.

The associative law also appears in English constructions involving consecutive motion verbs. For example, sentences like (6) can be analyzed as *come • (go • eat)*, while sentences like (7) can be analyzed as (*come • go*) • *eat*. Despite the conjunction *and*, such

structures defy Ross's (1967) Coordinate Structure Constraint. Wh-movement remains viable in sentences like (8).

(6) Come and go eat lunch with us!

(7) Come go and eat lunch with us!

(8) What meal did you ask him to come and go eat with us?

#### References:

- Bach, E., and G. Horn. 1976. Remarks on 'Conditions on Transformations'. *Linguistic Inquiry* 7:265-299.
- Belletti, A. 2004. Aspects of the Low IP area. In: *The Structure of IP and CP. The Cartography of Syntactic Structures*. Vol. 2, 16–51, Oxford: Oxford University Press.
- Chomsky, N. On wh-movement. In *Formal Syntax*. 71-132. San Diego, CA: Academic Press.
- Ross, J.R. 1967. Constraints on Variables in Syntax, Ph.D. dissertation, MIT.

# **WS11 Lexical ambiguity in the mind: theoretical challenges and interdisciplinary approaches**

**Lucie Barque & Richard Huyghe**



## **Short words are more likely to colexify multiple meanings than long words: empirical and experimental evidence**

Alexey Koshevoy  
(Aix-Marseille University)

This talk investigates the phenomenon of colexification, where a single wordform is associated with multiple meanings (François, 2008; Haspelmath, 2023). While prior research has primarily explored factors like conceptual similarity (Karjus et al., 2021) and frequency of use (Kanwal et al., 2017), our work examines the relatively understudied relationship between word length and colexification. If short words are preferred for colexifying two and more meanings, then this relationship alone should explain why short words usually have more meanings than their longer counterparts – a relationship that was established in Piantadosi et al. (2012). In this talk, I will discuss two studies aimed at testing the hypothesis of whether short words are more likely to be used to colexify more meanings than longer words, using both empirical (study 1) and experimental (study 2) methods.

*Study 1:* This preregistered study tested, for the first time on a diverse cross-linguistic dataset, a direct correlate of two well-known linguistic “laws”. Both the law of meaning (frequent words have more meanings) and the law of abbreviation (frequent words are shorter) are motivated by efficiency principles: communication is less costly (to speakers at least) when short words are re-used for multiple meanings, making them more frequent. Consequently, we should observe a correlation between the brevity of words in most languages and the number of meanings they are associated with. Studying the 633,308 wordforms in 1,952 languages representing 192 families from the Lexibank database (List et al., 2022), we evidence a robust correlation between a word’s brevity and the probability that a word colexifies two or more distinct meanings. Unlike the law of abbreviation, the strength of this relationship is quite variable and is stronger for better-documented languages.

*Study 2:* Our preregistered study examines whether word length influences word choice for colexification using a novel dyadic communication game and a computational model grounded in the Rational Speech Act (RSA) framework (Degen, 2023). In the dyadic communication game, the participants are presented with 3 meanings (in our experiment, the participants need to communicate about 3 aliens) that they need to communicate about using two words (one short and one long). The communication game is designed in such a way that the participants are also provided with context that they can use to disambiguate between different meanings, allowing them to colexify two meanings using one word. In our experiment, participants need to exchange messages about three aliens. Two of these aliens produce one distinct sound each, while the last one produces both of the sounds that the two other aliens produce. The participants can rely on the information about the sounds (the context) that the aliens make to disambiguate between them, provided that they refer to the two aliens that each produce a distinctive sound using one single word. Our RSA-based model predicts that speakers should colexify multiple meanings using short words if they are acting pragmatically by relying on both context and word length to make their communication successful. Data collection for this experiment is ongoing, and I will

report the experimental results during my talk.

Together, these studies contribute to our understanding of the processes that shape colexifications in natural languages.

## References

- Degen, J. (2023). The Rational Speech Act Framework. *Annual Review of Linguistics*, 9(1), 519–540. <https://doi.org/10.1146/annurev-linguistics-031220-010811>
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf’s Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52. <https://doi.org/10.1016/j.cognition.2017.05.001>
- Karjus, A., Blythe, R. A., Kirby, S., Wang, T., & Smith, K. (2021). Conceptual similarity and communicative need shape colexification: An experimental study. *arXiv:2103.11024 [Cs]*. <http://arxiv.org/abs/2103.11024>
- François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Ed.), *Studies in Language Companion Series* (Vol. 106, pp. 163–215). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.106.09fra>
- Haspelmath, M. (2023). Coexpression and synexpression patterns across languages: Comparative concepts and possible explanations. *Frontiers in Psychology*, 14, 1236853. <https://doi.org/10.3389/fpsyg.2023.1236853>
- List, J.-M., Forkel, R., Greenhill, S. J., Rzymiski, C., Englisch, J., & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, 9(1), Article 1. <https://doi.org/10.1038/s41597-022-01432-0>
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291. <https://doi.org/10.1016/j.cognition.2011.10.004>

## Lexical Ambiguity Processing in Alzheimer's Disease: Evidence of Alterations in Executive and Lexico-Semantic Processes

Camille Rostand\*, Marie Thébaut\*, Julien Dumurgier, Claire Paquet & Karin Heidlmayr  
(University Department of Speech and Language Therapy Teaching and Training, Sorbonne University; University Department of Speech and Language Therapy Teaching and Training, Sorbonne University; Paris Cité University, UMRS 1144, INSERM and INSERM U1153; Paris Cité University, UMRS 1144, INSERM and Centre of Cognitive Neurology, AP-HP Nord; & Paris Nanterre University, UMR 7114 CNRS Modyco)

\*These authors contributed to an equal amount.

Keywords: lexical ambiguity, homonyms, semantics, neurodegenerative dementia

Alzheimer's disease (AD) and other neurodegenerative dementias (ND) are the most frequent causes for mild or major neurocognitive disorders (DSM-V; American Psychiatric Association, 2013). AD is characterised by a gradual decline of cognitive, emotional and behavioural capacities. With regard to language processing, deficiencies in lexico-semantic processing can be observed from the early stages on. The present study aims to investigate how these alterations in language processing affect the comprehension of contextually embedded homonyms, which are a case of lexical ambiguity in which words can take completely different meanings in different contexts. Previous work on the processing of lexical ambiguity in dementia patients has provided evidence on altered semantic word-meaning priming in this population (e.g., Chenery et al., 1998; Faust et al., 1997; Piccirilli et al., 2015; Taler et al., 2009). In the present study, 13 early stage dementia patients (age:  $72.77 \pm 6.39$  years; MMSE:  $23.54 \pm 2.44$ ) and 16 healthy age-matched control participants (age:  $73.50 \pm 4.79$  years; MMSE:  $29.50 \pm 0.73$ ) were exposed to French homonyms (e.g., *palais*) that were placed at the end of a short discourse which biased interpretation towards one of the word's meanings (dominant vs non dominant meaning; e.g., *palais* 'palace' vs. *palais* 'palate'; see also, Dommes, 2006). Participants were asked to carry out a semantic similarity judgement between the meanings of the discourse context and a subsequent target word that was either semantically compatible with the discourse context (e.g., *monument* 'monument'), incompatible but related to the contextually irrelevant meaning of the homonym (e.g., *bouche* 'mouth'), or incompatible and fully unrelated to any of the meanings of the homonym (e.g., *radio* 'radio'). There was a main effect of semantic compatibility on response times (incompatible related > incompatible unrelated > compatible), suggesting a partial co-activation of both meanings of homonyms requiring the inhibition of the contextually irrelevant meaning. Moreover, there was a main effect of group with dementia patients showing overall slower response times and higher error rates compared to healthy controls. Importantly, there was also a group by context interaction on error rates. Post-hoc analyses revealed that dementia patients showed higher error rates when confronted to a meaning-dominant context for the embedded homonyms, especially when they had to reject semantically incompatible related target words. Finally, error rates were found to be negatively related to both, independent measures of executive control (BREF, *Batterie Rapide d'Efficiency Frontale*) and of lexico-semantic processing (TLE, *Test de Langage élaboré*; Rousseaux & Cas, 2012), suggesting a joint contribution of executive and lexico-semantic deficits in lexical ambiguity processing in Alzheimer's disease. Taken together, these findings suggest that

Alzheimer's patients suffer from an insufficient inhibition of contextually irrelevant meanings of ambiguous words placed in a discourse context, but also from a degradation of semantic representations as such, affecting more strongly the dominant meaning. Further research should shed light on the online influence of context during lexical access in cases of lexical ambiguity, e.g. using electroencephalographic measures, in dementia patients and healthy controls.

## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
- Chenery, H. J., Ingram, J. C. L., & Murdoch, B. E. (1998). The resolution of lexical ambiguity with reference to context in dementia of the Alzheimer's type. *International Journal of Language & Communication Disorders*, 33(4), 393–412. <https://doi.org/10.1080/136828298247703>
- Dommes, A. (2006). *La compréhension d'ambiguïtés lexicales présentées dans différents contextes phrastiques et discursifs chez des adultes jeunes et âgés: Effets des contraintes contextuelles introduites et de la familiarité des sens des mots ambigus* [PhD Thesis]. Université Paris Nanterre.
- Faust, M. E., Balota, D. A., Duchek, J. M., Gernsbacher, M. A., & Smith, S. (1997). Inhibitory Control during Sentence Comprehension in Individuals with Dementia of the Alzheimer Type. *Brain and Language*, 57(2), 225–253. <https://doi.org/10.1006/brln.1997.1747>
- Piccirilli, M., D'Alessandro, P., Micheletti, N., Macone, S., Scarponi, L., Arcelli, P., Petrillo, S. M., Silvestrini, M., & Luzzi, S. (2015). Impairment of homonymous processing in Alzheimer's disease. *Neurological Sciences*, 36(8), 1331–1336. <https://doi.org/10.1007/s10072-015-2085-5>
- Rousseaux, M., & Cas, P. D. (2012). *TLE: test de langage élaboré pour adultes*. Ortho éd.
- Taler, V., Klepousniotou, E., & Phillips, N. A. (2009). Comprehension of lexical ambiguity in healthy aging, mild cognitive impairment, and mild Alzheimer's disease. *Neuropsychologia*, 47(5), 1332–1343. <https://doi.org/10.1016/j.neuropsychologia.2009.01.028>

# Is polysemy more durable in English verbs than in adjectives?

Hilke Ceuppens

(KU Leuven)

Keywords: polysemy; word class; syntactic differentiation; semantic loss

Research has shown that polysemy is processed differently across word classes. Lopukhina et al. (2018), for instance, demonstrate that participants have a harder time differentiating between the senses of adjectives than between the senses of verbs. One possible explanation is that word classes are associated with different degrees of syntactic versatility. The greater a word's syntactic versatility, the more clues it offers that can be exploited in word sense disambiguation.

There is already some evidence to support the role of syntax in differentiating word senses in verbs. For example, Reinöhl & Ellison (2024) show that verbs are less open to metaphorical interpretation when used without overt argument. As another example, different frames have been shown to facilitate homonymy learning (Casenhiser 2005). Such findings suggest that, in verbs, different senses are more easily distinguished when they are linked to different syntactic frames. Adjectives, in contrast, cannot enter into different argument structures, meaning that beyond the attributive/predicative contrast, syntactic differentiation cannot aid sense disambiguation.

This paper hypothesizes that the difference between adjectives and verbs will be reflected in patterns of semantic loss. In earlier work, Ceuppens & De Smet (accepted) have shown that semantic loss is more likely where sense disambiguation is compromised. Therefore, it is expected that verbs, showing greater versatility and less sense overlap, are likelier to be able to sustain polysemy than adjectives. Conversely, adjectives, lacking syntactic cues to distinguish between senses, are more prone to semantic loss than verbs.

The hypothesis is tested by comparing rates of loss of source sense in 200 frequency-matched polysemous adjectives and verbs. Random samples with frequency information are compiled on the basis of the lemmatized version of the *Corpus of Historical American English*. Next, the samples are annotated for loss of source sense using the *Oxford English Dictionary*. Senses marked as obsolete, archaic, rare and historical are considered to be lost. Origin (native versus non-native words), network size (based on the number of senses in the OED) and compositionality (simplex versus complex) are annotated as control variables. Table 1 illustrates the possible combinations of dependent (loss / no loss) and independent variable (verb / adjective).

	Adjective	Verb
Loss	<i>Sad</i> († 'having had one's fill, sated' > 'sorrowful')	<i>Hinder</i> († 'to do harm to' > 'to keep back')
Polysemy sustenance	<i>Dull</i> ('slow of understanding' > 'cheerless, gloomy')	<i>Tie</i> ('to bind, fasten' > 'to be equal with in a contest')

Higher rates of loss are expected in adjectives than in verbs. By comparing two word classes, this paper explores the link between word class, syntactic properties and outcomes of ambiguity in the form of loss versus polysemy sustenance.

## Acknowledgements

Research funding: this research is funded by the Research Foundation – Flanders (FWO), grant 11PF524N.

## References

- Casenhiser, Devin (2005), Children's resistance to homonymy: an experimental study of pseudohomonyms, *Journal of Child Language* 32(2), 319-343.
- Ceuppens, Hilke & Hendrik De Smet (accepted), When does semantic change lead to semantic loss? Metaphor vs inference-driven metonymy, *Journal of Historical Linguistics*.
- Lopukhina, Anastasiya, Anna Laurinavichyute, Konstantin Lopukhin & Olga Dragoy (2018), The mental representation of polysemy across word classes, *Frontiers in Psychology* 9, 192.
- Reinöhl, Uta & Mark T. Ellison (2024), Metaphor forces argument overtiness, *Linguistics* 62(4), 795-847.

# Factors influencing the selection of translation equivalents for translation-ambiguous irregular polysemes

Jūratė Žukauskaitė  
(University of Agder)

Keywords: polysemy, translation, sense dominance, entrenchment, translation-ambiguity

Speakers constantly encounter lexical ambiguity – that is, words that have the same form but different meanings/senses depending on context. Understanding how this ambiguity is resolved is important for modelling language processing and the mental representation of word meaning. Although lexical ambiguity has been studied for decades, many questions remain – particularly regarding polysemy, despite its prevalence: 84% of English words have more than one sense (Rodd et al., 2004).

How polysemes are processed across languages is even less explored, though it offers valuable insights into bilingual language use. This study investigates the translation of irregular polysemes, whose sense relations are relatively idiosyncratic and typically based on metaphor (Apresjan, 1974; Vicente, 2018). Irregular polysemes are often translated into another language using different words, depending on the intended sense. This phenomenon – referred to as translation ambiguity – has been shown to affect word learning and processing in bilinguals (for a review, see Tokowicz & Degani (2010)). The complexity introduced by such one-to-many cross-linguistic mappings warrants greater attention in theories of bilingual lexical and translation models.

Building on prior research, this study proposes an account of irregular polyseme translation and identifies three key factors: (1) source word sense entrenchment, (2) cross-linguistic meaning overlap, and (3) translation equivalent (TE) frequency in the target language. Entrenchment, a concept stemming from cognitive linguistics, refers to the strength of a cognitive representation (Langacker, 1987). In polysemy research, its counterpart is sense dominance: the more frequently a sense is used, the more entrenched in the mental lexicon – and thus accessible – it becomes. In a decontextualized translation task, bilinguals are expected to select TEs corresponding to the more entrenched senses. The second factor, the degree of meaning overlap, assesses whether the number of shared senses between the source word and its TE influences the likelihood of its selection. This is operationalized as cumulative entrenchment: the sum of entrenchment scores of all senses the TE corresponds to. The third factor explores whether more frequent TEs are selected more often, building on findings that frequent words are processed more efficiently (Brysbaert et al., 2018).

These factors were tested using a decontextualized single-word L2-to-L1 translation task with 60 DE–EN bilinguals (including 30 translators), conducted online via the Gorilla platform (Anwyl-Irvine et al., 2020). Participants translated 26 irregular translation-ambiguous polysemes into German and wrote English sentences using the same words. Sentence data served to estimate sense entrenchment, assuming that participants would write sentences with the most accessible (i.e. entrenched) sense. Entrenchment was calculated as the proportion of times each sense appeared across all responses.

The paper presents results from individual- and group-level analyses and a mixed-effects beta regression model, with TE proportion as the dependent variable. As hypothesized, all three factors significantly influenced TE selection, in line with the *Gravitational pull (entrenchment) hypothesis* (Halverson, 2003, 2024). Results are discussed in light of their implications for research on polysemy, bilingualism, and translation.

## References

- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Apresjan, Ju. D. (1974). Regular polysemy. *Linguistics*, 12(142). <https://doi.org/10.1515/ling.1974.12.142.5>
- Brysbaert, M., Mander, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Halverson, S. L. (2003). The cognitive basis of translation universals. *Target. International Journal of Translation Studies*, 15(2), 197–241.
- Halverson, S. L. (2024). The gravitational pull hypothesis: Status and prospects. In J. Marco & I. Tello (Eds.), *Towards an empirical verification of the gravitational pull hypothesis: Evidence from the COVALT Corpus* (pp. 21–52). Peter Lang.
- Langacker, R. W. (1987). *Foundations of cognitive grammar* (Vol. 1). Stanford University Press.
- Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, 28(1), 89–104. [https://doi.org/10.1207/s15516709cog2801\\_4](https://doi.org/10.1207/s15516709cog2801_4)
- Tokowicz, N., & Degani, T. (2010). Translation ambiguity: Consequences for learning and processing. In B. VanPatten & J. Jegerski (Eds.), *Research in second language processing and parsing* (pp. 281–294). John Benjamins Publishing Company. <https://doi.org/10.1075/lald.53.12tok>
- Vicente, A. (2018). Polysemy and word meaning: An account of lexical meaning for different kinds of content words. *Philosophical Studies*, 175(4), 947–968. <https://doi.org/10.1007/s11098-017-0900-y>



# What can language models tell us about regularities in adjectival ambiguity?

Contextual word representations (Peters et al. (2018); Devlin et al. (2019)) brought significant improvements over a range of semantic tasks (Brown et al. (2020)) in natural language processing. One can hypothesize that this is due to their enhanced capacity to model meaning variations in context. We propose two experiments to study whether systematicity can be detected in the treatment of adjectival polysemy in Hungarian by language models.

Regular polysemy has been approached through the lens of cognitive processes (Lakoff (1987)), pragmatic inference (Falkum (2011)), or dynamic processes in the lexicon (Pustejovsky (1995); Briscoe and Copestake (1996); Asher (2011)). Dynamic lexicon theories often rely on meaning decomposition. They combine world knowledge with linguistic information in their account of systematic polysemy. Contextual word embeddings share some of these characteristics. Multi-layer neural nets can learn complex interactions between words and dynamically modify their representations. They encode similarities along specific dimensions, comparable to meaning decomposition. Moreover, there is no inherent separation between linguistic and extra-linguistic knowledge in the data or the model structure. However, despite recent advances (Gabriel Grand and Fedorenko (2022); Erk and Apidianaki (2024)), vector space representations are not easily interpretable, and the proxy tasks used to evaluate the treatment of ambiguity (Erk et al. (2009); Pilehvar and Camacho-Collados (2019)) struggle to reflect the lexical knowledge of humans. We therefore propose two experiments designed to study the treatment of systematic ambiguity by large language models.

Instead of analyzing word representations, we focus on the behavior of the model. In particular, we examine how the prediction about the noun following the adjective is affected when we artificially modify either the context or directly the representation of the adjective. To that effect, we use manually and semi-automatically created Hungarian datasets where adjectives are assigned to coarse semantic groups and fine-grained synonym classes. In the first experiment, we construct coordination patterns in the form of "*adjective noun<sub>1</sub> and noun<sub>2</sub>*" of two types: one where the nouns are frequent collocates with adjectives from the same synonym group, and one where they are collocates of distinct synonym groups. We then measure the surprisal (negative log probability) of the model (Nemeskey (2021)) for the second noun to verify if contexts eliciting different meanings increase the surprisal. In the second experiment, we examine broader semantic groups of adjectives, e.g. adjectives expressing states of mind, traits of character etc. Such groups of words share multiple common usage scenarios (Atkins and Rundell (2008)). The meaning shifts between these scenarios are heavily influenced by extra-linguistic factors. We examine whether traces of systematic generalization can be retrieved from the model's prediction above the immediate collocate level, which would indicate that the model abstracts away from extra-linguistic towards semantic information. Our findings contribute to the discussion on the role of dynamic lexical processes in language models and the interaction between world knowledge and linguistic knowledge in regular polysemy.

## References

- Nicholas Asher. 2011. *Lexical meaning in context*. Cambridge University Press.
- B. T. Sue Atkins and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press.

- Ted Briscoe and Ann Copestake. 1996. Controlling the application of lexical rules. In *SIGLEX Workshop on Breadth and Depth of Semantic Lexicons*.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Katrin Erk and Marianna Apidianaki. 2024. Adjusting interpretable dimensions in embedding space with human judgments. In *NAACL: Human Language Technologies (Volume 1: Long Papers)*, pages 2675–2686, Mexico City, Mexico. Association for Computational Linguistics.
- Katrin Erk, Diana McCarthy, and Nicholas Gaylord. 2009. Investigations on word senses and word usages. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 10–18, Suntec, Singapore. Association for Computational Linguistics.
- Ingrid Lossius Falkum. 2011. *The Semantics and Pragmatics of Polysemy: A Relevance--Theoretic Account*. Ph.D. thesis, University College London.
- Francisco Pereira Gabriel Grand, Idan Asher Blank and Evelina Fedorenko. 2022. Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behavior*.
- George Lakoff. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.
- Dávid Márk Nemeskey. 2021. Introducing huBERT. In *XVII. Magyar Számítógépes Nyelvészeti Konferencia (MSZNY2021)*, pages 3–14, Szeged.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Mohammad Taher Pilehvar and Jose Camacho-Collados. 2019. WiC: the word-in-context dataset for evaluating context-sensitive meaning representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1267–1273, Minneapolis, Minnesota. Association for Computational Linguistics.
- James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.

## The Role of Previous Experience and Speaker Community in Word-meaning: A Study of Word-Meaning Priming and Lexical Ambiguity Resolution

Every individual belongs to different communities, and some of them use language in different ways. For example, British English speakers use *flat* to refer to an apartment, while American English speakers may mean "two-dimensional." Studies indicate that listeners rely on social cues, like age or gender, to anticipate the speaker's meaning. Moreover, previous experience with word-meaning influences later interpretation of the word.

We conducted three online studies to explore how individuals use a speaker's identity and prior experience with word meanings to interpret ambiguous words. Participants (n=120) per condition completed an image-association task with ambiguous words, each paired with four images: one for the subordinate meaning, one for the dominant meaning, and two distractors. Participants answered the question "which picture goes best with [ambiguous word]?". We manipulated whether participants saw a priming-video of a speaker using one meaning of the ambiguous words and varied the speaker identities to compare the selection of subordinate images.

In Study 1 (Fig 1), we compared two conditions: The "video" condition had a priming-video of a speaker telling a story with the subordinate meaning of seven ambiguous words (i.e. *bat*, *bow*) before the image-association task, where the same speaker guided the task. The "no-video" condition only had the image-association-task. A linear mixed effect model found that subordinate images were chosen more often in "video" than in "no-video" ( $\beta = .69$ ,  $SE = .11$ ,  $p = 8.26e-11$ ). In Study 2 (Fig 2) we compared two conditions with "no-video" from Study 1: "One-speaker" condition was similar to "video" in Study 1, but we included an introduction-video of the speaker talking about themselves. In the "Two-speakers" condition, a different speaker appeared in the priming-video telling the same story. The introduction and the image-association task were identical to "One-speaker". We found no significant differences between these two conditions ( $\beta = -.14$ ,  $SE = .09$ ,  $p = 0.15$ ) and subordinate choices were significantly higher in both conditions in comparison with "no-video" ( $\beta_{\text{One-speaker}} = .63$ ,  $SE = .15$ ,  $p = 1.67e-05$ ;  $\beta_{\text{Two-speakers}} = .54$ ,  $SE = .11$ ,  $p = 7.41e-07$ ). This suggests that having two different speakers is not enough to reduce the effect of prior experience when we interpret ambiguous words.

---

1. Walker, A., & Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology*, 2, 219–237.

2. Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of cognitive neuroscience*, 20(4), 580-591.

3. Curtis, A. J., Mak, M. H., Chen, S., Rodd, J. M., & Gaskell, M. G. (2022). Word-meaning priming extends beyond homonyms. *Cognition*, 226, 105175.

4. Images belong to the Aragon Government and have been created by Sergio Palao for ARASAAC (<http://www.arasaac.org>), which distributes them under the Creative Commons BY-NC-SA License

In Study 3 (Fig. 3), we used job-specific word-meanings of six ambiguous words (i.e. *chicken*, *pot*) to test if priming decreases when different meanings are expected for each speaker. In the “Food” condition, a chef used the food-sense of the ambiguous words in the video and guided the image-association task, which included a chef’s avatar. In the “Farm” condition, the priming-video was identical to “Food”; but a farmer guided the image-association task, with a farmer’s avatar. In the “no-video” condition, participants only saw the image-association task and no avatar was included. We found significant differences between “Farm” and the other two conditions ( $\beta_{\text{Food}} = .25$ ,  $SE = .10$ ,  $p = .01$ ;  $\beta_{\text{No-video}} = -.37$ ,  $SE = .11$ ,  $p = .000689$ ). This suggests that when word-meanings are associated with different speaker-communities, having speakers from different communities can reduce the priming effect from prior experience.

Word count: 495

Fig 1:

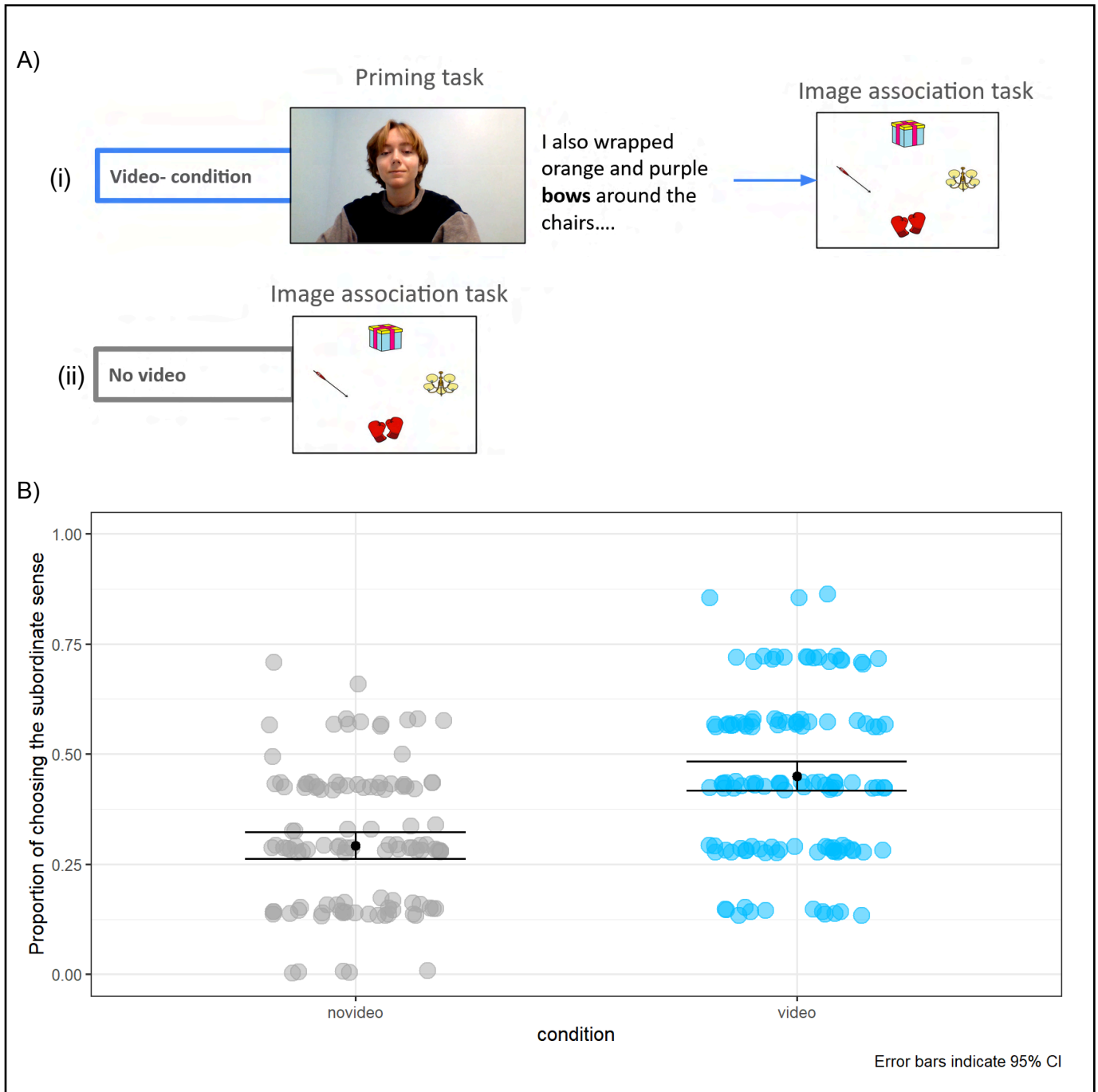


Fig. 1: Fig. 1: A) Design of Study 1 representing each condition of the study: (i) "video" (blue) and (ii) "no video" (gray). B) Results of Study 1, representing the proportion of subordinate choices in each condition

Fig 2:

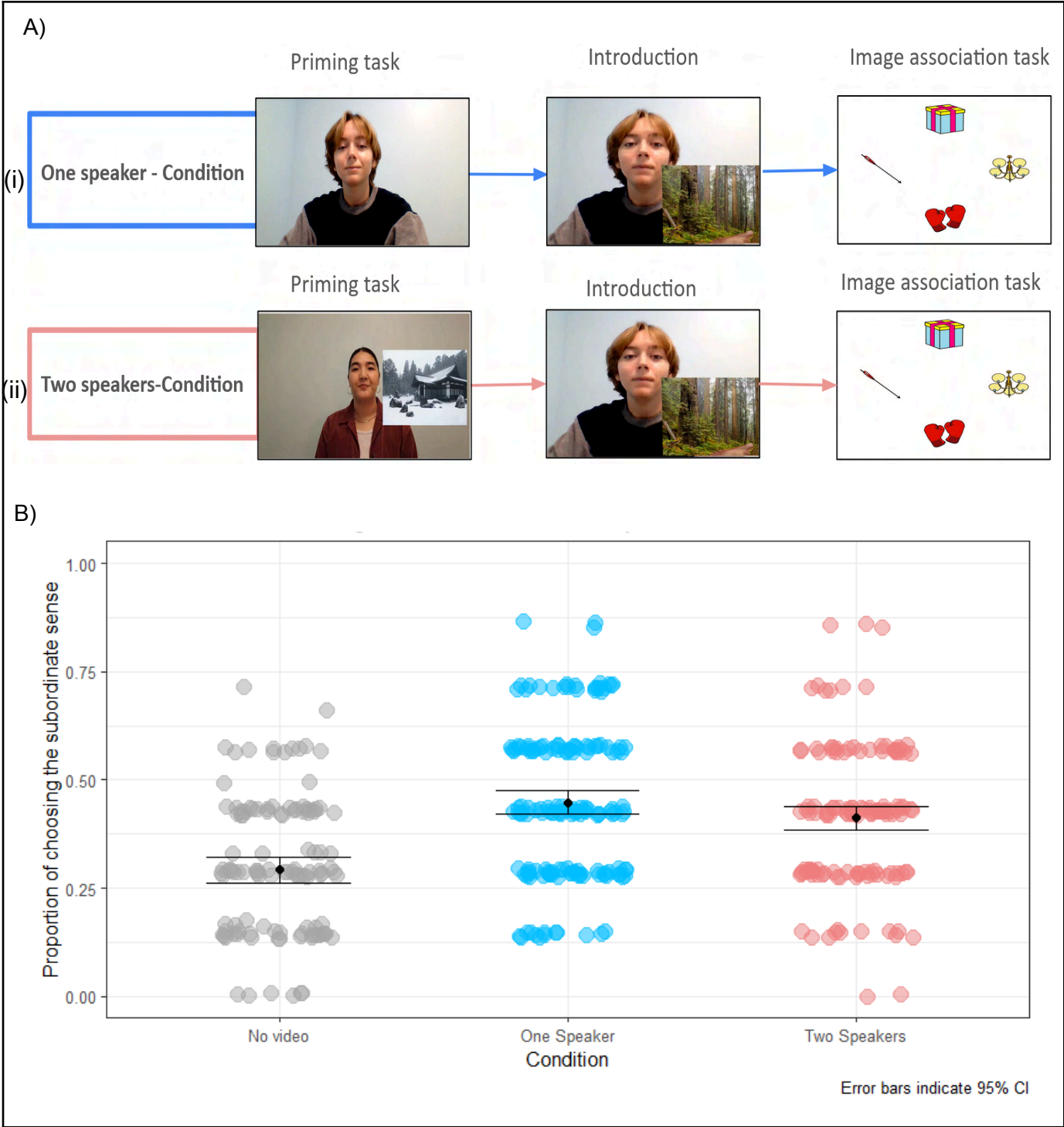


Fig 2: A) Design of study 2 representing each condition of the study: (i) “One- speaker” (blue) and (ii) “Two-speakers” (pink). B) Results of the second study: The plot shows the proportion of subordinate choices on each condition of Study 2 and “no-video” in study 1 (gray).

Fig 3:

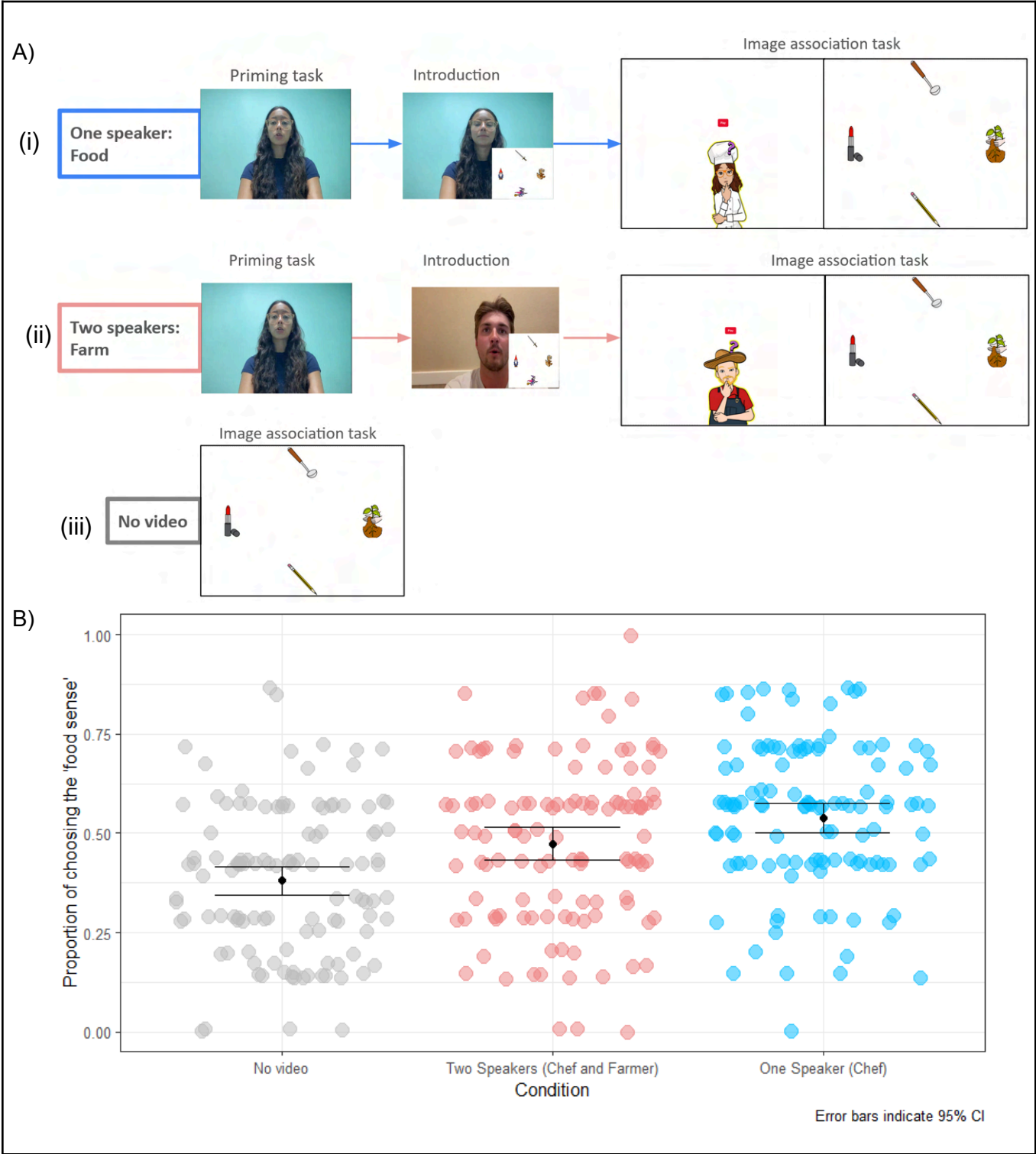


Fig 3: A) Design of Study 3, including a representation of each condition of the study: (i) “Food” (blue), where one speaker uses the “food” senses. (ii) “Farm” (pink) where there are two speakers (a chef and a farmer). The chef uses the “food” senses and the farmer guides the image association task (iii). “No video” (gray) where there is only an image association task. B) Results of study 3: The plot shows the proportion of subordinate choices on each condition of Study 3

# Ambiguity of abstraction: Semantic information affects words processing

Tommaso Lamarra, Caterina Villani & Marianna Marcella Bolognesi  
(University of Bologna)

Keywords: Generality, Linguistic ambiguity, Concreteness, Specificity, Abstraction

Linguistic ambiguity displays contrasting results across different processing tasks (lexical decision and semantic categorization). In fact, ambiguous words (“spring”, “board”) typically show a facilitatory effect, whereas an inhibitory effect arises when a competitive selection of relevant semantic information is required (Eddington & Tokowics 2015). This contrast is further supported by the evidence that words with more senses generally produce shorter responses in word recognition but less accurate semantic categorizations (Yap et al. 2011).

Hereby, we propose to extend this debate considering the influence on mental representation and processing of two correlated, yet distinct, variables involved in conceptual abstraction: concreteness, a measure of the degree of perceptibility of a referent (“democracy” vs “chair”); and categorical specificity, which describes the range of inclusiveness of a word’s denotation (“animal” vs “beagle”, “art” vs “impressionism”; Bolognesi & Caselli 2023). These semantic and psycholinguistic variables encode two distinct aspects of referentiality and are related to two relevant notions in the debate about ambiguity, namely indeterminacy and generality (Lyons 1977 and Gillon 1990). Indeed, the representation of more abstract words is typically more semantically sparse and less defined than concrete concepts (Borghi 2022). Similarly, the representation of generic words (hypernyms) is associated with a multitude of contexts and exemplars that share less salient information, whereas specific words (hyponyms) denote more precise referents (more specialized sense) (Barque & Chaumartin 2009) and are associated with less, but semantically related, contexts (Rambelli & Bolognesi 2024). Such differences might be reflected in variations in behavior-related measurements during online word processing.

To address this question, we investigated the influence of both concreteness and specificity on decision latencies in two behavioral experiments: a lexical decision task (words vs. non-words) and a semantic decision task (abstract vs. concrete).

Our stimuli were carefully constructed to include words systematically high and low in both specificity and concreteness and balanced for word length and frequency. Specifically, we used WordNet (Miller 1995) to generate a hypernym and a hyponym for each of 30 abstract and 30 concrete words selected from a previous study (Villani et al. 2022), resulting in a final set of 120 words. Moreover, we conducted a norming study to obtain concreteness and specificity ratings on our stimuli in order to operationalize the two variables on a continuous scale.

Results showed a non-significant effect of concreteness and specificity on reaction times in lexical decision task and a significant effect of both variables in semantic decision task. These findings align with previous research, which suggests that a shallow processing mechanism is involved in lexical decision task (Pexman et al. 2017) rather than a deeper semantic access to the meaning of the words. Indeed, when stimuli are controlled for relevant lexical variables, such as length and frequency, the semantic variables may have a less significant influence (Yap et al. 2011). Interestingly, while we replicated the well-known advantage of concrete words over abstract ones (Borghi et al. 2017), we also found first evidence of a facilitatory effect of specific words over generic ones, regardless of their level of concreteness. These results suggest that the instantiation of a precise referent and more



salient information improves the performance of participants in semantic categorization. On the contrary, a dense representation with more competitive information/exemplars requires more time to be processed, likely due to an inherent ambiguity.

### Acknowledgements

The study was funded by the European Research Council (Grant Agreement: ERC-2021-STG-101039777). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

### References

- Barque, Lucie, and Chaumartin, François-Régis (2009), Regular polysemy in WordNet, *Journal for language technology and computational linguistics*, 24(2), 5-18.
- Bolognesi, Marianna M., and Caselli, Tommaso (2023), Specificity ratings for Italian data, *Behavior Research Methods*, 55(7), 3531-3548.
- Borghi, Anna M. (2022), Concepts for which we need others more: The case of abstract concepts, *Current directions in psychological science*, 31(3), 238-246.
- Borghi, A. M., Binkofski, Ferdinand, Castelfranchi, Cristiano, Cimatti, Felice, Scorolli, Claudia, and Tummolini, Luca (2017), The challenge of abstract concepts, *Psychological Bulletin*, 143(3), 263.
- Eddington, Charles M., and Tokowicz, Natasha (2015), How meaning similarity influences ambiguous word processing: The current state of the literature, *Psychonomic bulletin & review*, 22, 13-37.
- Gillon, Brendan S. (1990), Ambiguity, generality, and indeterminacy: Tests and definitions, *Synthese*, 85, 391-416.
- Lyons, John. (1977), *Semantics*, vol. 1, Cambridge University Press.
- Miller, George A. (1995), WordNet: a lexical database for English, *Communications of the ACM*, 38(11), 39-41.
- Pexman, Penny M., Heard, Alison, Lloyd, Ellen, and Yap, Melvin J. (2017), The Calgary semantic decision project: concrete/abstract decision data for 10,000 English words, *Behavior research methods*, 49, 407-417.
- Rambelli, Giulia, and Bolognesi, Marianna M. (2024, May), The Contextual Variability of English Nouns: The Impact of Categorical Specificity beyond Conceptual Concreteness, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 15854-15860).
- Villani, Caterina, Orsoni, Matteo, Lugli, Luisa, Benassi, Mariagrazia, and Borghi, A. M. (2022), Abstract and concrete concepts in conversation, *Scientific Reports*, 12(1), 17572.
- Yap, Melvin J., Tan, Sarah E., Pexman, Penny M., and Hargreaves, Ian S. (2011), Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification, *Psychonomic bulletin & review*, 18, 742-750.

# From word recognition to translation: How polysemy and ambiguity shape language processing within and across languages

Victoria Tkacikova, Gabriela Terrazas Duarte & Natasha Tokowicz  
(University of Pittsburgh)

Keywords: translation ambiguity, polysemy, language processing, norming, word processing

Lexical ambiguity plays a significant role in language processing within and across languages, and the way we measure ambiguity can shape the accuracy and reliability of predictions about lexical processing. In the first part of this talk, we explore within-language ambiguity by comparing how lexical decision performance is predicted using a word's number of senses (NOS) as estimated by dictionaries versus human norms. Previous studies examining lexical processing have used lexicographer dictionaries to investigate the effect of polysemy (words with multiple related meanings) on word recognition but using lexical databases or dictionaries to quantify NOS may not accurately capture the way people process these words (Gernsbacher 1984). Rice et al. (2019) analyzed the interactions of polysemy with established predictors of lexical decision latencies, finding that NOS interacted with word features like context availability and orthographic neighborhood. Their study used the number of entries from WordNet to define NOS. Using subjectively-normed NOS, a reanalysis of the study showed that, when compared using Bayesian Information Criterion (BIC), subjective NOS was a better model fit than WordNet NOS ( $DBIC = 7$ ; a "strong" effect, based on standards in Kass & Raftery 1995). Word frequency and context availability still interacted, but orthographic neighborhood predictors did not. Our findings suggest that subjectively-normed NOS are better able to predict lexical processing, underlining the importance of how polysemy is operationalized when studying language ambiguity.

In the second part of this talk we discuss cross-linguistic ambiguity, focusing on "translation ambiguity", which occurs when words have multiple translations across languages (e.g., Prior et al. 2007). Prior research has demonstrated that translation ambiguity is at least partially related to within language ambiguity (e.g., Degani et al. 2016). Examining these issues cross linguistically provides an opportunity to examine how the nature, degree, and regularity of the relationship between senses impacts the processing and representations of ambiguous words as well as the salience and respective importance of these properties in the mental lexicon. We discuss translation ambiguity's origins and its prevalence in large normative studies (e.g., Basnight-Brown et al. 2020) as well as the negative consequences of translation ambiguity on language processing. We also present a norming method used to capture the semantic similarity of meanings of the translations of a word, translation semantic variability (TSV; Bracken et al. 2017). TSV is a continuous measure of semantic similarity, on which the lowest TSV rating indicates that the meanings of the translations are "completely different", and the highest rating indicates that the meanings of the translations are "exactly the same". Bracken et al. found that words with higher TSV ratings are responded to faster and more accurately than words with lower TSV ratings. These findings suggest that cross-linguistic ambiguity is associated with slower and less accurate lexical processing when translations differ more in meaning. To conclude, we discuss the

importance of using more sensitive measures of lexical ambiguity such as subjectively-normed NOS and TSV.

## References

- Basnight-Brown, D. M., Kazanas, S. A., & Altarriba, J. (2020). Translation ambiguity in Mandarin-English bilinguals: Translation production differences in concrete, abstract, and emotion words. *Linguistic Approaches to Bilingualism*, 10(4), 559-586.
- Bracken, J., Degani, T., Eddington, C., & Tokowicz, N. (2017). Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words. *Bilingualism: Language and Cognition*, 20(4), 783-794.
- Degani, T., Prior, A., Eddington, C. M., Da Luz Fontes, A. B. A., & Tokowicz, N. (2016). Determinants of translation ambiguity. *Linguistic Approaches to Bilingualism*, 6(3), 290-307.
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, 113(2), 256-281.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.
- Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, 39(4), 1029-1038.
- Rice, C. A., Tokowicz, N., Fraundorf, S. H., & Liburd, T. L. (2019). The complex interactions of context availability, polysemy, word frequency, and orthographic variables during lexical processing. *Memory & Cognition*, 47, 1297-1313.
- Tokowicz, N. (2014). Translation ambiguity affects language processing, learning, and representation. In M. Ryan T, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, ... D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges between Disciplines*. Somerville, MA: Cascadia Proceedings Project

## **WS12 Mistaken beliefs**

**Caroline Gentens, William B. McGregor & Stef Spronck**

# Mirative rejection: When mistaken beliefs trigger surprise

Caterina Mauri & Antonia Russo  
(University of Bologna & University of Bergamo)

Keywords: <mirativity, rejection, grammaticalization, spoken Italian, dialogicity>

The aim of this paper is to discuss a family of constructions expressing *mirative rejection* (Mauri and Russo 2025), i.e. a complex function characterized by the expression of (i) surprise at another speaker's beliefs (Martinez 2023), and (ii) the falsity of those very beliefs (McGregor 2024). We will consider the case of Italian, as exemplified by examples (1) and (2):

- (1) **Altro che**      *noia,*      *la*      *vela*      *entusiasma*      *i*      *giovani*  
**Altro che**      boredom,      DEF      sailing      excite:PRS:3SG      DEF      young.people  
'Forget boredom, sailing excites young people.' (itTenTen20)
- (2) "[...] *Hai*      *esagerato*      *con*      *il*      *vino*      *del*      *Tata*  
Have:PRS:2SG      exaggerate:PRTC.PST      with      DEF      wine      of.DEF      Tata  
*vero?*" "**Macché**      *vino*      *e*      *vino!*      *Magari*      *fosse*      *colpa*  
right?" "**Macché**      wine      and      wine!      I wish      be:PST.PFV:3SG      fault  
*del*      *vino!*"  
of.DEF      wine!"  
"[...] You drank too much of the Tata's wine, didn't you?" "**What** wine and wine! I wish it were  
the wine's fault!" (itTenTen20)

Through the two constructions *altro che* (lit. 'other than', (1)) and *macché* (univerbation of lit. 'but what', (2)) the speaker rejects a given SoA, for which there was some expectation of validity by the hearer or in the context, at the same time expressing surprise at the expectation itself, thus underlining the wrong belief behind it.

First, we aim to provide a typology of the constructions attested in contemporary Italian for mirative rejection, through the analysis of the spoken Italian corpus KIParla (Mauri et al. 2019) (1.991.903 words). We will show that mirative rejection is conveyed by a wide, though not random, set of constructions including diversity expressions (e.g. *altro che*) and rhetorical echo questions (e.g. *macché* lit. 'but what', *ma dove* lit. 'but where', *quando mai* lit. 'when never').

Second, we will focus on the constructions *altro che* and *ma che*, which ultimately univerbized into response particles (respectively *altroché!* and *macché!*): by means of a representative sample of texts since the 13th century taken from the CODIT corpus (Micheli 2021) (29.495.514 words), we will compare the paths that led to the emergence of the mirative rejection function, highlighting the factors at work.

In both cases, dialogicity and correction play a role. The diachronic analysis shows that *altro che* began to show the mirative meaning in exclamative, corrective contexts from the 17<sup>th</sup> century onward, anaphorically referring to some prior mistaken belief. The mirative rejection uses of *ma che*, by contrast, start to be attested later on, around the 19<sup>th</sup> century, mainly in dialogical contexts: in this case, the evolution seems to have originated in rhetorical questions (e.g., *ma che dici?* 'but what do

you say?') and metalinguistic ones (e.g., *ma che dico?* 'but what do I say?'), where *ma che* was originally used to ask for clarification or to signal repair. We will conclude by presenting a unified account of mirative rejection constructions in Italian, explaining the variety of strategies used to convey this function in discourse as a result of the highly interactional and intersubjective nature of its meaning, which is closely tied to the speakers' expectations, beliefs, and stance.

## References

- Martinez, Randi. 2023. Mirativity in English response particles: An analysis from the syntax/semanticsinterface. *Yale Working Papers in Grammatical Diversity* 5(3). 1–25.
- Mauri, Caterina, Ballarè, Silvia, Goria, Eugenio, Cerruti, Massimo & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*, 1–7. Available at <https://kiparla.it/>.
- Mauri, Caterina & Antonia Russo 2025. The mirative values of Italian *altro che*. In Susana Rodriguez Rosique (ed.), *Expressing Surprise at the Crossroads: Mirativity, Exclamativity and Cooptation in Romance Languages*, edited by Susana Rodriguez Rosique. Berlin/Boston: De Gruyter Mouton, 143–178.
- McGregor, William B. 2024. On the expression of mistaken beliefs in Australian languages. *Linguistic Typology*. 101–145.
- Micheli, Silvia M. 2021. CODIT: Corpus diacronico dell'italiano. Ústav Českého národního korpusu FF UK, Prague. Available at <http://www.korpus.cz/>.

## Exploring Reality-Refuting Particles: The Multifunctionality of Ende *Ka* and Areal Parallels in Komnzo and Idi

Kate Lynn Lindsey (Boston University)  
Catherine Scanlon (University of California, Santa Barbara)

A group of languages in southern Papua New Guinea each have a counterfactual particle with diverse functionality. The multifunctional Ende particle *ka* appears in diverse contexts such as counterfactuals, hypotheticals, potentials, rhetorical questions, and corrective negation (Tighe and Gabbay 2022). With *ka*, speakers manage shifts in knowledge and perspective, aligning well with this workshop's focus on particles that mediate perspectives and correct assumptions.

In Idi, cognates like *ka* and *gä* similarly manage the common ground (Schokkin, p.c.), while Komnzo *kma* marks counterfactuals, potentials, and negations (Döhler 2018). Ende *ka* and Idi *gä* are observed more often in the speech of young women, potentially indicating a socially indexed meaning as well (Gast, p.c.). Parallels suggest possible areal and typological trends in how regional languages manage knowledge, inference, and belief, underscoring the need for further studies. In this paper, we investigate the particular use of *ka* in Ende by analyzing examples from the Ende Language Corpus (Lindsey 2015).

These varied uses of *ka* share a common thread: in Ende, *ka* signals a stance toward belief, expectation, or reality. In counterfactual and hypothetical constructions, *ka* marks scenarios as contrary to known reality or imagined possibilities, often framing situations of mistaken belief or unrealized expectations. The mistaken belief can involve new information (e.g., in (1), *ka* signals that the turtle is mistaken), or it can indicate a separate reality from the one held in the common ground (e.g., in (2), *ka* marks the event of chasing a prankster pretending to be a ghost). In this way, *ka* behaves similarly to the belief verb *yiwei* in Mandarin (Glass 2021), which signals that the common ground should not be updated with questionable or false information.

- (1) *Kottllam a [...]*    *gongnomenyän*    *ada*    *ka*    *ddia da*    *zäme*  
 turtle=NOM    thought.REM.3sgS    like\_this    CNTF deer=NOM    ALR  
*ngattong agan,*    *be bogo de*    *amne me gotarnän [...]*  
 first    aux.REC.3sgS    but    he=EMPH    middle=LOC sleep.REM.IMPF.3sgS  
 ‘Turtle thought [wrongly] that Deer already won, but he was in the middle sleeping.’ (Sowati 2016)

- (2) *Däbe*    *lla*    *kuddäll*    *anyke de*    *ada*    *ka*    *koenmäll*  
 that    man    dead    spirit=ACC    like\_this    CNTF chase  
*erallo,*    “Ya!”  
 aux.PRS.3pl>3sg go\_away  
 ‘They chased the dead man’s spirit [actually the prankster] away, shouting, “Ya!”’ (Kurupel 2017)

*Ka*’s use in rhetorical questions hints at the speaker’s expectation that the addressee should already know or infer the answer (3). In corrective negation, *ka* counters incorrect assumptions or statements (4).

- (3) *Ada*    *sisor*    *bin di*    *ada*    *eran*    *ka?*    *Ttongo*    *bin a*  
 like\_this    other    name=ACC    say    aux.PRS.3sg>3sg    Q    another    name=NOM  
*ddone*    *dan*    *ko.*  
 NEG    cop.PRS.sgS    also  
 ‘Do I have any other names [rhetorical]? I don’t have any other names, no.’ (Nama 2018)

- (4) *Ngämo yae era gudne gudaē gozegän, be ka*  
 1sg.POSS mother=FOC long\_ago long\_ago born.REM.3sgS but NEG  
*ngämo yae era still ddoḃae māngall=ang dan.*  
 1sg.POSS mother=FOC loanword very strong=ATR cop.PRS.sgS  
 ‘My mother was born a long time ago, but no, my mother is still very strong.’ (Sali 2018)

In potential expressions, *ka* conveys hope or attempts, overlaying possible future scenarios with an element of uncertainty. The potential meaning stands out from the other meanings in that it is the only one that represents a non-negative or possible interpretation (Spronck, p.c.). However, many of the potential examples co-occur with an event that does not take place. For example, in (5), Zakae conveys that the follow-up interview cannot or will not occur now (as expected) but could take place the following day.

- (5) *Ka, malla sisri. Abo ka kănazbag.*  
 NEG NEG now but POT tomorrow  
 ‘No, not now. But hopefully tomorrow.’ (Zakae 2018)

These varied uses of the single particle *ka* support an analysis where these types of constructions are classified as “polarity reversal constructions” (la Roi 2024), where counterfactuality is exemplified as just one example of a polarity reversal (Karttunen 1971).

**Language context:** Ende and Idi are Pahoturi River languages. Komnzo is a Yam language spoken to the west of the Pahoturi River languages.

## References

- Döhler, Christian. 2018. *A grammar of Komnzo* (Vol. 22). Language Science Press.
- Gast, Volker. Personal communication about his fieldwork on Sibidiri Idi.
- Glass, Lelia. 2022. The negatively biased Mandarin belief verb YIWEI. *Studia Linguistica* 77 (1). 1–46.
- Karttunen, Lauri. 1971. Subjunctive Conditionals and Polarity Reversals. *Paper in Linguistics* 4 (2). 279–298.
- Kurupel (Suwede), Warama. 2017. Auma we ibiatt ttoen (retelling). In Lindsey, Kate L. (Ed.) 2015. [http://catalog.paradisec.org.au/collections/LSNG08/items/SE\\_PN026](http://catalog.paradisec.org.au/collections/LSNG08/items/SE_PN026).
- la Roi, Ezra. 2024. Polarity reversal constructions and counterfactuals in Ancient Greek: Between implicature and conventionalization. *Journal of Historical Linguistics* 14 (2). 335–375.
- Lindsey, Kate L. 2015. *The Language Corpus of Ende and other Pahoturi River languages*. PARADISEC. <http://catalog.paradisec.org.au/collections/LSNG08/>.
- Nama, Al. 2018. Sociolinguistic Questionnaire - Al Nama. In Lindsey, Kate L. (Ed.) 2015. [http://catalog.paradisec.org.au/collections/LSNG08/items/SE\\_PI081](http://catalog.paradisec.org.au/collections/LSNG08/items/SE_PI081).
- Sali, Erabal. 2018. Sociolinguistic Questionnaire - Erabal Sali. In Lindsey, Kate L. (Ed.) 2015. [http://catalog.paradisec.org.au/collections/LSNG08/items/SE\\_PI064](http://catalog.paradisec.org.au/collections/LSNG08/items/SE_PI064).
- Schokkin, Dineke. Personal communication about her fieldwork on Dimsisi Idi.
- Sowati (Kurupel), Maryanne. 2016. Ddia da wa Kottlam a (Deer and Turtle). In Lindsey, Kate L. (Ed.) 2015. [http://catalog.paradisec.org.au/collections/LSNG08/items/RE\\_EN005](http://catalog.paradisec.org.au/collections/LSNG08/items/RE_EN005).
- Tighe, Amelia and Shaked Gabbay. 2022. *Particles ka and ke in Ende*. Unpublished manuscript.
- Zakae, Bibiae. 2018. Sociolinguistic Questionnaire - Bibiae Zakae. In Lindsey, Kate L. (Ed.) 2015. [http://catalog.paradisec.org.au/collections/LSNG08/items/SE\\_PI050](http://catalog.paradisec.org.au/collections/LSNG08/items/SE_PI050)



## Distinguishing types of ‘mistaken belief’ in reported thought and presupposed complement constructions

Kristin Davidse  
(University of Leuven)

Keywords: Semiotic Grammar, reported thought, presupposed complement constructions

The notion of ‘mistaken belief’ is traditionally restricted to constructions of reported thought, e.g. (1). I extend it to constructions with a presupposed complement, as in (2)-(4), arguing for reconceptualization of the factive presupposition in the process.

- (1) Doctors wrongly believed I was terminally ill. (WB)
- (2) Saddam did not accept that he lost the 1991 Gulf war. (WB)
- (3) i just accepted it that i’m always wrong. (Google)
- (4) Mr Blair did not accept that Mr Brown was to blame. (WB)

Reported thought constructions represent the cognizant’s *creation* of the idea in the complement clause (Davidse 1991, McGregor 1997), which the speaker may assess as ‘mistaken’, e.g. by *wrongly* in (1). A negated matrix, e.g. *Doctors never believed that I was ill*, entails *non-existence* of that idea in the *doctors’* minds.

The matrices in (2)-(4) contain the predicate *accept* traditionally listed as factive, whose negation does *not* entail non-existence of the proposition in the complement clause, as in (2) and (4). For this reason, the complement propositions of (*not*) *accept* are said to be *presupposed*. However, only (2) instantiates the received definition of factivity according to which the complement proposition is presupposed *true* by the *speaker* (Kiparsky and Kiparsky 1970). In (3)-(4), the complement proposition is not presupposed true by the speaker-now. In (3), it is other people who held that the speaker was always wrong, which the speaker-then as represented cognizant wrongly used to believe (*just accepted*). In (4) it is some third party who is committed to the truth of the proposition.

I argue that the matrix predicates in (2)-(4) trigger the presupposition that *a* proposition exists *prior* to the described mental operation -- not that this proposition is always true for the speaker. This is because, unlike in reported thought, the proposition *does not originate* in the mental activity described by the matrix verb (Halliday 1985). This opens it up for authorship by the speaker (2), the represented cognizant (3), a contextually given third party (4), or authorship may be left underspecified (Davidse 2003). Structurally, the complement clause is a nominalized argument of the matrix predicate -- reflected in its ability to be extraposed, as in (3) -- which conveys some type of ‘mental manipulation’ of the pre-existent, entitized proposition (Langacker 1991, Gentens 2020). With speaker-presuppositions the contextually triggered meaning of *not accept* is ‘not recognize the truth’, e.g. (2), while with third-party-presuppositions *accept* means ‘not reject’, e.g. (3), both implied to be mistaken by the speaker-now. We thus see how complementation constructions with presupposition-triggering cognition predicates can also convey speaker stance regarding mistaken belief.

This paper aims to develop a typology of ‘mistaken belief’ in the extended sense in terms of clusters of semantic-pragmatic features: (i) creation vs (subtypes of) manipulation of the proposition, (ii) authorship of the proposition, (iii) speaker-now different from or identical with represented cognizant, (iv) ‘mistake’ coded, inferable or left open. I will illustrate this typology, focussing on the various verb senses that can be conveyed by (*not*) *believe* / (*not*) *accept* – either activated

constructionally (as in reported thought versus presupposed complement constructions with *believe*) or contextually (as in presupposed complement constructions with *accept*). Using all logically possible feature combinations as heuristic, illustrative data will be collected from WordbanksOnline (abbreviated as WB) and the Internet.

## References

- Davidse, Kristin (1991), *Categories of Experiential Grammar*, PhD Thesis, Linguistics Department: University of Leuven.
- Davidse, Kristin (2003), A corpus check of the factive presupposition, in A. Remael and K. Pelsmaekers (eds), (2003), *Configurations of Culture: Essays in Honour of Michael Windross*, Antwerpen: Garant, 115-126.
- Gentens, Caroline (2020), *The Factive-Reported Distinction in English*, Berlin/Boston: Walter de Gruyter.
- Kiparsky, Paul and Carol Kiparsky (1970), Fact, in M. Bierwisch and K. Heidolph (eds), *Progress in Linguistics*, . Den Haag: Mouton de Gruyter, 143-173.
- Langacker, Ronald (1991), *Foundations of Cognitive Grammar. Part II. Descriptive application*. Stanford: Stanford University Press.
- Halliday, M.A.K. (1985), *An introduction to Functional Grammar*. London: Arnold.
- McGregor, William. (1997), *Semiotic Grammar*. Oxford: Clarendon.

# Mistaken belief constructions and egophoricity in a Sinitic language: The case of Baoding Mandarin

Na Song

(INALCO and CRLAO-CNRS)

**Keywords:** mistaken beliefs, egophoricity, epistemic modality, Baoding, Sinitic languages

This paper explores the interaction between mistaken beliefs and egophoricity in Baoding Mandarin, a Sinitic variety spoken in northern China. It examines how mistaken beliefs are grammatically encoded and how their expression is constrained by egophoric alignment.

As in many languages in Asia (McGregor 2022), Baoding employs a matrix verb  $xr^{22}$ - $t\zeta o$  ‘think (by mistake)’ in the complement construction, similar to Standard Mandarin, but with an egophoric constraint. In Baoding,  $xr^{22}$ - $t\zeta o$  ‘think (by mistake)’ can only occur with first-person singular subjects in declarative clauses and second-person singular subjects in interrogatives. In contrast, a second verb,  $t\tilde{a}^{45}$ , also meaning ‘think (by mistake)’, has no such restrictions and is compatible with all persons. For instance,

- (1)  $u\gamma^{213}$   $t\tilde{a}^{45}$   $\zeta\gamma^{51}$   $t^ha^{213}$   $ma^{45}$   $t\epsilon ie^{45}$   $t^ha^{45}=l\epsilon=\tilde{a}$ ,  
 1SG think (by mistake) COP 3SG mother pick.up 3SG=COME&DO=MIR  
 $n\zeta^{51}$   $p\tilde{a}^{51}$   $t^hi\tilde{a}^{45}$   $\zeta\gamma^{51}$   $t^ha^{213}$   $l\zeta^{213-21}l\zeta$   
 finally COP 3SG grandmother  
 ‘I thought it was her mother who came to pick her up, but it turned out to be her grandmother.’

The example (1) expresses a mistaken belief, with  $t\tilde{a}^{45}$ , which can also appear in the second person or third person context.

- (2)  $u\gamma^{213}$   $xr^{22}$ - $t\zeta o$   $\zeta\gamma^{51}$ - $k\gamma$   $ta^{51}$   $ko^{213}=\tilde{a}$   
 1SG think (by mistake)-STA COP-CLF big dog=MIR  
 $n\zeta^{51}$   $p\tilde{a}^{51}$   $t^hi\tilde{a}^{45}$   $\zeta\gamma^{51}$ - $k\gamma$   $lu^{51}$   
 finally COP-CLF deer

‘I thought it was a big dog, but it turned out to be a deer.’

In the example (2)  $xr^{22}$ - $t\zeta o$  is used only with the first person singular in declarative sentences.

The egophoric constraint observed with  $xr^{22}$ - $t\zeta o$  aligns with systems in languages with dedicated egophoric markers. However, unlike Wutun or other contact-influenced varieties (e.g., Janhunen et al. 2008, Sandman 2018), there is no clear evidence of external influence in Baoding. The system appears to have evolved internally (Song 2019).

To address the diachronic dimension (cf. McGregor 2024), this paper proposes that Baoding's mistaken belief construction may have developed through the convergence of two grammatical pathways: (i) reinterpretation of epistemic predicates with restricted person alignment, and (ii)

semantic bleaching of evidential mirativity markers like =ã, often used in correcting false assumptions.

The findings are based on personal fieldwork and elicitation with native speakers. This study contributes to the typology of mistaken belief constructions by documenting a previously undescribed interaction between egophoricity and epistemic marking in a Sinitic language. It also raises broader questions about how person-based epistemic systems may give rise to grammaticalized mistaken belief constructions.

## References

- Janhunen, Juha, R. L. Miller, S. Itaru, & Victor H. Mair (2008), *Wutun - A Chinese-Based Language of Amdo Qinghai*. Introduction.
- McGregor, William B. (2022), Expressions of mistaken belief: a functional and typological approach. Presented at XXXVèmes Journées de Linguistique d'Asie Orientale, Paris, 9th July 2022.
- McGregor, William B. (2024), On the expression of mistaken beliefs in Australian languages. *Linguistic Typology*, 28(1), 101–145.
- Sandman, Erika (2018), *A Grammar of Wutun*. Brill.
- Song, Na (2019), Egophoric marking in a Sinitic language: The case of Baoding. *Journal of pragmatics* 148 :88–110.

# Mistaken belief *wh*-constructions in Korean

Okgi Kim & Jong-Bok Kim  
(Kyung Hee University & Kyung Hee University)

Keywords: mistaken belief *wh*-construction, non-argumental *wh*-word, rhetorical, HPSG, Korean

We investigate mistaken belief *wh*-constructions (MBWCs) in Korean in which a *wh*-word is used to mark an addressee's belief as mistaken (Cheung 2008, Cheung 2009, Park & Kang 2020, and Kim & Kim 2022). Consider (1).

- (1) Mimi-ka {mwe-ka/eti/ettehkey} selun sal i-ni?  
Mimi-NOM what-NOM/where/how 30 years.old COP-QUE  
'You mistakenly believe that Mimi is 30 years old.'

The non-argumental *wh*-phrase has no information-seeking force as in interrogative *wh*-questions, but asserts the falsity of the addressee's belief conveyed by the question, just like other mistaken belief expressions (McGregor 2024).

MBWCs are licensed only in a context where the speaker and the addressee have conflicting beliefs about the event in question (Cheung 2009 and Kim & Kim 2022):

- (2) A: Mimi-nun an chakha-y.  
Mimi-TOP not honest-DECL  
'Mimi is not honest.'  
B: #Mimi-ka {mwe-ka/eti/ettehkey} chakha-ni?  
Mimi-NOM what-NOM/where/how honest-QUE  
'You mistakenly believe that Mimi is honest. (In fact, she is not.)'

The MBWC in (2B) is infelicitous in the given context where both the interlocutors have the same belief that Mimi is not honest. This validates the use of MBWCs as marking the addressee's belief as mistaken.

Both MBWCs and rhetorical *wh*-questions are interpreted as an assertion of the opposite polarity from what is apparently asked. However, the two behave differently in other respects. For example, unlike rhetorical *wh*-questions, MBWCs cannot be uttered discourse-initially or out of blue (Kim & Kim 2022):

- (3) [Mimi approaches John in the heavy rain and says:]  
a. nwu-ka o-keyss-ni?  
who-NOM come-KEYSS-QUE  
'No one will come. (because of the heavy rain)'  
b. #pi-ka mwe-ka o-ko iss-ni?  
rain-NOM what-NOM come-CONN COP-QUE  
'You wrongly believe that it is raining.'

This contextual constraint is attributable to MBWCs' function to assert the falsity of the addressee's belief expressed in the previous discourse.

To explain the unique properties of MBWCs, we suggest that the *wh*-phrase in MBWCs syntactically modifies an independent ([IC +]) polar question and semantically yields a rhetorical question by asserting the reverse polarity value of the polar question. This can be represented in the feature structure format of HPSG (Head-driven Phrase Structure Grammar):

$$(4) \left[ \begin{array}{l} \text{SYN} \left[ \begin{array}{l} \text{MOD } S \left[ \begin{array}{l} \text{IC } + \\ \text{SEM } \lambda\{ \} p \end{array} \right] \end{array} \right] \\ \text{SEM } \text{ASSERT } \neg[p] \\ \text{PRAG } \text{believe}(\text{speaker}, \neg) \end{array} \right]$$

Given this lexicalist assumption, the sentence in (5A) asserts that Mimi is not honest. The speaker's belief (*believe(speaker,  $\neg[\text{honest}(m)]$ )*) evoked pragmatically can be cancelled like the following:

- (5) A: Mimi-ka      mwe-ka      chakha-ni?  
          Mimi-NOM   what-NOM   honest-QUE  
          'No way is Mimi honest.'  
      B: mollasse?   chakhanil-ul      manhi      ha-ko      iss-e  
          dont.know   honest.thing-ACC   many      do-CONN   COP-DECL  
          'You don't know? She's been doing lots of honest things.'

The lexicalist view explains the ungrammaticality of (6).

- (6) \*nwu-ka      mwe-ka      yeypu-ni?  
          who-NOM   what-NOM   pretty-QUE  
          '(lit.) You wrongly think who is pretty?'

The modified sentence is a *wh*-question expressing the open proposition ( $\lambda\{x\}[\text{pretty}(x)]$ ). Open propositions cannot be the object of someone's belief, since they are not the type of thing that can be true or false. Thus, the MBWC cannot take the open proposition to assert the falsity of the addressee's belief.

## References

- Cheung, Lawrence Yam-Leung (2009), Negative *wh*-construction and its semantic properties. *Journal of East Asian Linguistics* 18, 297–321.  
 Cheung, Yam-Leung (2008), *The negative wh-construction*. University of California, Los Angeles.  
 Kim, Okgi & Jong-Bok Kim (2022), Negative *wh*-constructions in Korean: A discourse-based approach. *Studies in Generative Grammar* 32(1), 49–74.  
 McGregor, William B (2024), On the expression of mistaken beliefs in Australian languages. *Linguistic Typology* 28(1), 101–145.  
 Park, Myung-Kwan & Arum Kang (2020), The rhetorical adjunct *wh mwue(-l)* 'what' in Korean. *The Journal of Linguistic Science* 95, 111–135.

# ***Wähnen* is a false belief predicate**

Simon Wimmer

Heinrich-Heine-University Duesseldorf

Keywords: False belief predicates, Contrafactuals, Neg-raising, Attitude ascriptions, German

I argue that the German predicate *wähnen* roughly means ‘believe falsely and for no good reason’. My description of some of *wähnen*’s key properties contributes to recent work on the extent to which natural languages have false belief predicates (e.g. McGregor 2024; Roberts and Özyildiz 2023).

Holton (2017: 245-7, 263-4) has argued that *wähnen* is not a ‘contrafactive’, i.e. not a false belief predicate that encodes its falsity inference as a presupposition, is morphologically and conceptually atomic, and licenses the kinds of complement knowledge predicates license. I agree that, at least as used by my informants for contemporary German, *wähnen* is not a contrafactive. (Sander 2020, 7; Sander, forthcoming discuss past varieties of German.) For one, as 1a shows, *wähnen* does not license embedded questions; and even declarative complements, as in 1b, are only licensed for some informants. But, whilst for Holton, *wähnen*’s falsity inference is cancellable and thus neither an entailment nor a presupposition, I claim that it is an entailment.

1. a. Ayesha {*\*wähnt/weiß*}, ob Benedikt schwimmt.  
‘Ayesha *wähnt*/knows whether Rebecca swims’  
b. Benedikt {*wähnt/weiß*}, dass Ayesha Recht hat.  
‘Benedikt *wähnt*/knows that Ayesha is right’

I focus on *wähnen* with a reflexive or accusative followed by an adjective or prepositional phrase, as in 2, and *wähnen* with Konjunktiv I, as in 3a.

2. a. Chloe *wähnt* sich sicher.  
‘Chloe *wähnt* herself safe.’  
b. David *wähnt* die Delegation auf Reisen.  
‘David *wähnt* the delegation travelling.’
3. a. Ewald *wähnt*, Chloe sei schon eingetroffen.  
‘Ewald *wähnt* Chloe has already arrived.’

I argue that, in these constructions:

- i) *Wähnen* entails belief. E.g., 3a entails that Ewald believes that Chloe has already arrived.
  - ii) *Wähnen* has a falsity entailment, rather than implicature, presupposition, anti-presupposition, or post-supposition because *wähnen*’s falsity inference is neither cancellable nor projective. As for projection, 4a does not trigger the falsity inference that Ayesha is not responsible. As for cancellation, 4b shows that *wähnen* is incompatible with reporting that the proposition the matrix subject believes, here that Chloe is safe, is true.
4. a. Floris *wähnt* nicht Ayesha verantwortlich, sondern Benedikt.  
‘Floris does not *wähnt* Ayesha responsible, but Benedikt.’  
b. # Chloe *wähnt* sich sicher, aber sie ist sicher.  
‘Chloe *wähnt* herself safe, but she is safe.’

Notably, *Wähnen*'s falsity inference can appear cancellable, particularly if *wähnen* occurs in past tense, as in 5a. But this effect parallels the apparent cancellability of the truth inference of factive predicates like *wissen*, shown in 5b. The appearance can thus be explained away along similar lines (cf. Colonna Dahlman and van de Weijer 2022).

5. a. David währte die Delegation auf Reisen. Aber es wusste eh jeder, dass die Delegation gerade in Kanada unterwegs war.  
'David *währt*-past the delegation safe. But everyone knew anyway that the delegation was currently travelling around Canada.'
- b. Jeder wusste, dass Stress Geschwüre verursachten bis zwei australische Ärzte in den frühen 80er bewiesen, dass Geschwüre tatsächlich von bakteriellen Infektionen verursacht worden.  
'Everyone knew that stress caused ulcers, before two Australian doctors in the early 80s proved that ulcers are actually caused by bacterial infections.' (ex. 1 in Hazlett 2010)
- iii) *Wähnen* has a 'no-good-reason' entailment that the matrix subject does not believe for good reason, e.g., because they believe based on self-deception. This corresponds to *wähnen*'s apparent Indo-European root *\*uēn(ə)* 'strive', later 'wish, love, be satisfied' ('*wähnen*' 2024).
- iv) *Wähnen* does not license neg-raising. E.g., 4a does not trigger an inference that Floris *währt* that Ayesha is not responsible. To explain this, I assume that belief predicates are neg-raising because they presuppose that one either believes p or believes not-p (e.g. Gajewski 2007; Mayr 2019; Theiler, Roelofsen, and Aloni 2019). Given this, *wähnen* is not neg-raising because its falsity and no-good-reason inferences make the presupposition that one either *währt* p or *währt* not-p unreasonable. For, typically, if one does not falsely and for no good reason believe p, one also does not falsely and for no good reason believe not-p, but instead either falsely believes p for a good reason, truly believes p for a good reason, truly believes p for no good reason, or believes neither p nor not-p.

## Acknowledgments

I thank Thorsten Sander, Julia Zakkou, and Alexander Dinges for bringing the importance of describing *wähnen*'s behaviour to my attention.

## References

- Colonna Dahlman, Roberta, and Joost van de Weijer. 2022. 'Cognitive Factive Verbs across Languages'. *Language Sciences* 90. <https://doi.org/10.1016/j.langsci.2021.101458>.
- Gajewski, Jon Robert. 2007. 'Neg-Raising and Polarity'. *Linguistics and Philosophy* 30 (3): 289–328. <https://doi.org/10.1007/s10988-007-9020-z>.
- Hazlett, Allan. 2010. 'The Myth of Factive Verbs'. *Philosophy and Phenomenological Research* 80 (3): 497–522.
- Holton, Richard. 2017. 'I—Facts, Factives, and Contrafactuals'. *Aristotelian Society Supplementary Volume* 91 (1): 245–66. <https://doi.org/10.1093/arisup/akx003>.
- Mayr, Clemens. 2019. 'Triviality and Interrogative Embedding: Context Sensitivity, Factivity, and Neg-Raising'. *Natural Language Semantics* 27 (3): 227–78. <https://doi.org/10.1007/s11050-019-09153-8>.
- McGregor, William B. 2024. 'On the Expression of Mistaken Beliefs in Australian Languages'. *Linguistic Typology* 28 (1): 101–45. <https://doi.org/10.1515/lingty-2022-0023>.
- Roberts, Tom, and Deniz Özyildiz. 2023. 'Bad Attitudes'. Presented at Sinn und Bedeutung 28, Ruhr-University Bochum.
- Sander, Thorsten. 2020. 'Fregean Side-Thoughts'. *Australasian Journal of Philosophy* 0 (0): 1–17. <https://doi.org/10.1080/00048402.2020.1795216>.



- . Forthcoming. 'A Puzzle about Anti-Factives'. *Journal of the American Philosophical Association*.
- Theiler, Nadine, Floris Roelofsen, and Maria Aloni. 2019. 'Picky Predicates: Why Believe Doesn't like Interrogative Complements, and Other Puzzles'. *Natural Language Semantics* 27 (2): 95–134. <https://doi.org/10.1007/s11050-019-09152-9>.
- 'wähnen'. 2024. DWDS. August 2024. <https://www.dwds.de/wb/w%C3%A4hnen>.

## Mistaken belief and the problem of indirect speech

The most prominent strategy for commenting on the words and thoughts of others in the languages of the world is reported speech, often classified into two types: direct and indirect speech, as in (1-2) (Coulmas 1986).

- (1) Alex said: “This is an example of direct speech”
- (2) Alex suggested that this was an example of indirect speech

With a construction like (1) a speaker conveys a reported message without qualification, a construction like (2) may co-express or imply an attitude towards the reported message (e.g. “I do not believe the content of the reported message”). Indirect speech appears to be both much rarer in the languages of the world and more difficult to define than direct speech (cf. Evans 2013).

While most Australian Aboriginal languages are claimed to lack a direct-indirect speech opposition, many allow rather complex combinations of perspective expressions, particularly in the realm of mistaken belief (McGregor 2023). For example, (3) and (4), co-express a reported message and a statement that the reported thought or utterance was untrue.

- (3) Wanyjirra (Australia, Pama-Nyungan)  
*gulangarra=n*                      *gang-an-i*                      *ngarin*                      *ward*  
wrongly.thought=2MIN.SBJ CARRY-CONT-PAST meat.ABS return  
“I thought that you had brought back meat (but you didn’t).” (Senge 2015: 492)

- (4) Ngarinyin (Australia, Worrorran)  
*birr-niyangarri-karra*                      *burr-ma-ø*                      *mangarri*  
3PL-good-EPIST                      3PL-do-PRS food  
‘They think it is good food [but it is not]’ (Spronck 2015: 177)

Consequently, many Australian languages combine a wealth of perspectival sentence types with a complete absence or very marginal presence of a direct-indirect speech opposition, creating a paradoxical treatment of these languages in the typological literature on attitude sentences: On the one hand, the (near) absence of a clear indirect speech construction suggests that speech reporting in Australian languages is (predominantly) done through a presumably unqualified representation of the utterance of the reported speaker. On the other hand, Australian Aboriginal languages can often signal qualifications regarding, e.g., the verity of a reported message in more subtle and economic ways than, e.g., European languages.

This paper starts with a detailed examination of the strategies available for mixing perspectives in two Australian Aboriginal languages, Ngarinyin and Wanyjirra, partially based on previously unpublished fieldwork data. The analysis contextualises the mistaken-belief expressions with other (sentential) modal construction types and shows the degree of variability in the domain of reported speech. Based on these observations, we suggest that mistaken-belief expressions present a useful test-case for studying perspectival

meanings across various morphosyntactic types and apply our language-specific findings to a larger variety sample (Miestamo, Bakker, and Arppe 2016) of Australian Aboriginal languages (n = 42).

We conclude with two claims with relevance for both syntactic typology and the theme of the workshop, mistaken belief: first, we suggest that the presumed (near-)universality of direct speech as opposed to indirect speech is predicated on an incorrect classification of perspective expressions and, second, that a closer look at the Australian data demonstrates that mistaken-belief expressions, rather than an exceptional construction type, constitute an integral part of the cross-linguistic repertoire of perspectival constructions.

### References

- Coulmas, Florian, ed. 1986. *Direct and Indirect Speech*. Berlin: Mouton de Gruyter.
- Evans, Nicholas. 2013. "Some Problems in the Typology of Quotation: A Canonical Approach." In *Canonical Morphology and Syntax*, edited by Dunstan Brown, Marina Chumakina, and Greville G. Corbett, 66–98. Oxford etc.: Oxford University Press.
- McGregor, William B. 2023. "On the Expression of Mistaken Beliefs in Australian Languages." *Linguistic Typology* 28 (1): 101–45. <https://doi.org/10.1515/lingty-2022-0023>.
- Miestamo, Matti, Dik Bakker, and Antti Arppe. 2016. "Sampling for Variety." *Linguistic Typology* 20 (2). <https://doi.org/10.1515/lingty-2016-0006>.
- Senge, Chikako. 2015. "A Grammar of Wanyjirra, a Language of Northern Australia." PhD thesis, The Australian National University.
- Spronck, Stef. 2015. "Refracting Views: How to Construct Complex Perspective in Reported Speech and Thought in Ungarinyin." *STUF – Language Typology and Universals* 68 (2): 165–85.

**The diachrony of mistaken belief markers in Ngumpin-Yapa languages (Australia)**  
**Dr Tom Ennever — Surrey Morphology Group, University of Surrey**

Recently, attention has been brought to the fact that many Australian and Amazonian languages have some dedicated linguistic means for expressing notions of the kind ‘X thought y was true, but y isn’t’, referred to as **mistaken belief (MB)** (Spronck and Vuillermet 2019; McGregor 2024). For some languages, such dedicated markers have been shown to have emerged from such conceptually-related markers as similatives of the kind ‘like x, as if x’, as well as from factive ‘know’ constructions (McGregor 2024: 110). However, no studies to-date have explored the converse scenario, that is: *what can/do MB markers evolve into?* This study addresses precisely such a question, presenting evidence from languages within a closely related subgroup of an erstwhile MB marker (re-)grammaticalizing into a marker of standard negation as well as other negative markers restricted to specific modal contexts.

This comparative study focusses on the Ngumpin-Yapa (NGY) subgroup of Pama-Nyungan (Australia), drawing on published materials as well as my own fieldwork with Ngardi speakers (2016–2019). Building on McConvell & Laughren’s (2004) analysis of a shared NGY particle *\*kula*, I present a diachrony of *\*kula* as an historic MB marker that accounts for its retained MB meanings exemplified by eastern Walmajarri *kula* (1), as well as its more widespread distribution as a marker of various negative (modal) meanings, cf. Wanyjirra *gula* (2). I show that for the majority of languages, MB meanings themselves have only been retained through fusion of *\*kula* with other particles/enclitics carrying certain modal and/or evidential meanings (see Table 1) e.g. a dubitative marker *nga(rra)* ‘might’ e.g. (3) or an evidential marker *nganta* ‘supposedly’ (4).

The study considers a number of important issues that arise from the proposed diachronic analysis with respect to how mistaken belief is pragmatically interpreted (viz. whether it is defeasible), as well as the flexibility with respect to whom the mistaken belief can be attributed to: contrast (1) and (4). In sum, the study contributes new data and new insights into the diachrony and synchronic variation in MB expressions and their place within the related domains of negation and modality-evidentiality.

- (1) Walmajarri (Richards and Hudson 2012)  
***Kula*** *wanjingarni, paparr-marn-i.*  
**MB** alive without thought-speak-PST  
 ‘I thought he [was still] alive (but he wasn’t), I said (that), without thinking.’
- (2) Wanyjirra (Senge 2015: 490)  
***Gula***=*liyarra bina garriny-ana.*  
**NEG=1UA.SBJ** know stay-PRS  
 ‘We don’t know her.’
- (3) Wanyjirra (Senge 2015: 493)  
***Gulangarra*** *ngu=yi baya-rra.*  
**MB** REAL=1MIN.O bite-IMP/IRR  
 ‘I thought he (the dog) would bite me (but he didn’t).’
- (4) Warlpiri (Laughren et. al, 2007: 232)  
*Pantu-rnu=jana kulanganta yapapatu-juku, kala mingkirripatu.*  
 spear-PST=3PL.O **MB** real\_people-STILL but termite\_mounds  
 ‘He speared them as though (≅wrongly thinking that) they were real people, but (they were) termite mounds.’

Table 1: Mistaken belief grams in Ngumpin-Yapa languages

	MB marker	Gloss	Source
Ngardi	<i>kulanga, kulangarra, kulanganta</i>	‘mistaken belief’	(Ennever 2021: 605–607)
Walmajarri	<i>kula, kulangarra, kularni</i>	‘it seemed, contrary to fact’	(Richards and Hudson 2012)
Jaru	<i>gulanga (+...=rra), gulangarra,</i>	‘as though, it looks’	(Tsunoda 1981: 206, 209)
Wanyjirra	<i>gulangarra</i>	‘wrongly thought’	(Hudson & Richards 1984)
Bilinarra	—	—	(Meakins & Nordlinger 2014)
Gurindji	<i>kata + ...=nga</i>	‘thought’	(Meakins and McConvell 2021, 550–551)
Mudburra	—	—	(Osgarby 2018)
Warlmanpa	<i>kulanganta</i>	‘deontic correction’	(Browne 2024: 332–335)
Warlpiri	<i>kulanganta</i>	‘counterfactual’	(Nash 1986: 239)

## References

- Browne, Mitchell. 2024. *A Grammar of Warlmanpa*. Canberra: Pacific Linguistics.
- Ennever, Thomas. 2021. *A Grammar of Ngardi: As Spoken by F. Tjama, M. Yinjuru Bumblebee, D. Mungkirna Rockman, P. Yalurngali Rockman, Y. Nampijin, D. Yujuyu Nampijin, M. Mandigalli, K. Padoon, P. P. Napangardi, P. Lee, N. Japaljarri, M. Moora, M. Mudgedell and P. Smith. A Grammar of Ngardi*. Berlin, Boston: De Gruyter Mouton. <https://www.degruyter.com/document/isbn/9783110752434/html>.
- Laughren, Mary, Kenneth Locke Hale, and Warlpiri Lexicography Group. 2007. Warlpiri-English Encyclopaedic Dictionary. University of Queensland.
- McConvell, P., & Laughren, M. (2004). The Ngumpin-Yapa subgroup. In C. Bowern & H. Koch (eds.), *Australian languages: Classification and the comparative method*, 169–196. Amsterdam: John Benjamins Publishing.
- McGregor, W. B. (2024). On the expression of mistaken beliefs in Australian languages. *Linguistic Typology*, 28(1), 101–145. <https://doi.org/10.1515/lingty-2022-0023>.
- Meakins, Felicity, and Patrick McConvell. 2021. *A Grammar of Gurindji: As Spoken by Violet Wadrill, Ronnie Wavehill, Dandy Danbayarri, Biddy Wavehill, Topsy Dodd Ngarnjal, Long Johnny Kijngayarri, Banjo Ryan, Pincher Nyurmiari and Blanche Bulngari. A Grammar of Gurindji*. De Gruyter Mouton. <https://doi.org/10.1515/9783110746884>.
- Meakins, F., & Nordlinger, R. (2014). *A grammar of Bilinarra: An Australian Aboriginal language of the Northern Territory*. Berlin: De Gruyter Mouton.
- Nash, David. 1986. *Topics in Warlpiri Grammar*. New York; London: Garland.
- Osgarby, D. (2018). *Verbal morphology and syntax of Mudburra: An Australian Aboriginal Language of the Northern Territory*. MPhil dissertation. University of Queensland.
- Richards, Eirlys, and Joyce Hudson. 2012. *Interactive Walmajarri—English Dictionary: With English—Walmajarri Finderlist*. 2nd ed. AuSIL Interactive Dictionary Series A-5. Darwin: Australian Society for Indigenous Languages (AuSIL).
- Spronck, Stef & Marine Vuillermet. 2019. Mistaken-belief constructions across two continents. Presented at ALT2019 – 13th Conference of the Association for Linguistic Typology, Pavia.
- Senge, Chikako. 2015. *A Grammar of Wanyjirra, a Language of Northern Australia*. PhD dissertation, Australian National University.
- Tsunoda, Tasaku. 1981. *The Djaru Language of Kimberley, Western Australia*. B 78. Canberra: Pacific Linguistics.

## **WS13 Morphological boundaries in Creole languages**

**Ana R. Luís & Susanne Maria Michaelis**

## Diachrony of morphology in Carriols

Aymeric Daval-Markussen, Peter Bakker, Joost Robbe (Aarhus University, Denmark) & Kristoffer Friis Bøegh (Utrecht University, Netherlands)

Keywords: Carriols; diachrony; morphology

This research is part of the project "Digital demography, creole creation, light on letters" funded by the Independent Research Foundation of Denmark, Grant 10.46540/4256-00026B

Virgin Islands Dutch Creole – called Carriols by its speakers (Stein 1982) and formerly spoken on the then Danish Virgin Islands – is the creole language with the most extensive documentation on its historical development, with records spanning over 250 years (Rossem and van der Voort 1996). From the 150 slave letters written from the 1730s, through dozens of printed religious texts from the 1760s to the 1830s and vernacular utterances, texts and dialogues preserved from between the 1760s to 1987, this material is ideal for studying the development of Carriols morphology. In older texts, we find more bound morphemes than in newer texts. In this illustrative sentence from a letter in Carriols from 1739, we can observe a genitive *-s*, infinitival *-en* and a case-marked article *den*, all of them inherited from 18th-century Dutch:

- (1) *Nú is ons hope, De Konings Majestait ons sal die order geven, dat ons durf voortgaan te leeren den Heere Jesus.*

‘Now it is our hope that His Majesty the King will give us the order that we may continue to teach the Lord Jesus.’

(from a letter by enslaved printed in Zinzendorf 1742: 483-484)

Yet, it is also creole-like with uniform pronouns and preverbal mood markers. The modern Dutch equivalent is presented in (2), with adjectival agreement on the first plural possessive *ons* and number agreement on the auxiliary *mogen* (the equivalent of Carriols *durven*), all of which are absent in the creole sentence:

- (2) *Nu is het onz-e hoop, dat Majesteit de Koning ons de order zal gev-en, dat we mog-en door-gaa-n over de Heer Jezus te leren*

In the modern creole, based on data from the 1920s (de Josselin de Jong 1926), there is no derivational or inflectional morphology at all, and no reduplication. The modern Carriols sentence in (3) illustrates this:

- (3) *Nou ons hoop dat di kiniñ Majesteit sa befeel ons, dat ons mu lastān lōpān fo leer fu here Jisas.*  
(constructed sentence)

In this paper, we focus on a number of Dutch inflectional and derivational processes, and the development of innovations in those areas in Carriols, such as: nominal plural marking, person marking on verbs, tense marking on verbs, causative and transitivizing marking on verbs, passive morphology, separable compound verbs (verb-particle combinations), case marking on pronouns, agreement morphology on adjectives and reduplication of verbs and nouns (de Josselin de Jong 1926, Sabino 2012, and van Sluijs, Muysken and Los 2017). We base our results on a selection of representative texts, both religious (hymns, catechisms) and vernacular (dialogues, letters, proverbs), covering the whole period of documentation.

The available language data display a development where Dutch-derived morphological markers disappear from the record, and not much morphological innovation can be observed in later texts either. Yet, Carriols was a complete language.

This reveals a process of change unlike those traditionally found in creoles, from a language with some lexifier morphology to a language without lexifier morphology and with very little evidence of morphological innovation. We discuss whether this is evidence of gradual yet rapid loss of all lexifier morphology, or lack of authenticity of the texts, or a symptom of diglossia, where the basilectal variety was not documented.

## References

- de Josselin de Jong, Jan Petrus Benjamin (1926), *Het huidige Negerhollandsch (Teksten en woordenlijst)*, Amsterdam: Koninklijke Academie van Wetenschappen te Amsterdam.
- Rossem, Cefas van and Hein van der Voort (1996), *Die Creol Taal: 250 years of Negerhollands texts*, Amsterdam: Amsterdam University Press.
- Sabino, Robin (2012) *Language contact in the Danish West Indies: Giving Jack his jacket*, Brill's Studies in Language, Cognition and Culture 1, Leiden & Boston: Brill.
- Stein, Peter (1982), Quelques dates nouvelles de l'histoire du mot créole, *Études Créoles* 5, 162–165.
- van Sluijs, Robbert, Pieter Muysken and Bettelou Los (2017), Verb particle combinations and word order change in Dutch-lexifier creole languages, In Bettelou Los and Pieter De Haan (eds.), *Word order change in acquisition and language contact: Essays in honour of Ans van Kemenade*, 265–290, Amsterdam: John Benjamins.
- Zinzendorf, Nikolaus Ludwig von (1742), *Büdingische Sammlung einiger in die Kirchen-Historie einschlagender sonderlich neuerer Schrifften, Erster Band*, Büdingen.



## Verbal stems in the morphology of Guadeloupean Creole

Delphine Tribout, Florence Villoing, Fabiola Henri & Maxime Deglas  
(University of Lille and STL; Paris Nanterre University and Modyco; University of Buffalo; & Paris 8 University)

Keywords: verbal morphology, stem allomorphies, Guadeloupean Creole, lexical morphology

Creoles are often considered as having no, or at least very poor, inflectional morphology (Valdman 1978, Seuren and Wekker 1986, McWhorter 1998, among others). In the case of Guadeloupean Creole, this assumption may appear to be true because, unlike Indian Ocean Creoles where verbs typically alternate between a short and a long form (Henri 2010), Guadeloupean verbs are usually considered to have one single form. There is only a handful of exceptional cases where verbs can have two or three forms (e.g. Hazaël-Massieux 2002), but the contexts in which these exceptional cases are used are usually dismissed as cases of free variation. Lexical morphology is never taken into account.

Yet, in French, which is its lexifier language, verbal morphology is rather important and shows many stem allomorphies (e.g. Bonami and Boyé 2002, 2003).

The present paper will analyze what happened to these allomorphies when French verbs have been inherited in Creole. Building on Villoing et al. (2024), we will use real data collected in recent databases (Deglas 2023) and corpora (Missud 2018) in order to study both inflectional and lexical verbal morphology. We will confirm the authors' finding that two stems need to be postulated for Guadeloupean verbal morphology (Stem A and Stem B in Table 1). We will also study how French verbal allomorphies have been reanalyzed in Guadeloupean. For example, the French verb ADMIRER 'to admire' is usually analyzed as having an allomorphic stem *admirat-* /*admiras-* that is used to form the deverbal noun ADMIRATION /*admirasjõ*/ with the suffix *-ion* (/jõ/). However, this inherited deverbal noun is reanalyzed in Guadeloupean Creole as the stem *admir-* and the suffix *-asyon* in the noun ADMIRASYON /*admirasjõ*/ . We will also show how verbal morphology has developed its own system in Guadeloupean. Finally, we will see that this specific morphological organization is also visible within the part of the lexicon that is not inherited from French and is made of real Guadeloupean innovations (e.g. DÉKALFOUKÉ, LYANNÉ, POUPOULÉ, ROPYANNÉ).

Verb	Stem A	Stem B	Infl. Form	A derivative	B derivative
ADMIRÉ 'admire'	admir	admiré	<i>admiré</i>	ADMIR-ASYON 'admiration'	
CHANTÉ 'sing'	chant	chanté	<i>chanté</i>	CHANT-È 'singer'	CHANTÉ-CHARI 'ploughing song'
DÉKALFOUKÉ 'act violently'	dékalfouk	dékalfouké	<i>dékalfouké</i>	DÉKALFOUK-ASYON 'act of violence'	
LYANNÉ 'unit'	lyann	lyanné	<i>lyanné</i>	LYANN-AJ 'association'	DELYANNÉ 'disunité'
POUPOULÉ 'tease'	poupoul	poupoulé	<i>poupoulé</i>	POUPOUL-MAN 'teasing'	
RAPYANNÉ 'live frugally'	rapyann	rapyanné	<i>rapyanné</i>	RAPYANN-È 'miser'	
VIRÉ 'turn'	vir	viré	<i>viré</i>	VIR-AJ 'bend'	VIRÉ-DO 'abandon' lit. 'turn the back'

Table 1 – Verbal stems used for inflexional and lexical morphology

## References

- Bonami, Olivier and Boyé, Gilles (2002), Suppletion and dependency in inflectional morphology, in F. Van Eynde, L. Hellan and D. Beerman (eds), (2002), *Proceedings of the HPSG'01 Conference*, Stanford: CSLI publications, 51-70.
- Bonami, Olivier and Boyé, Gilles (2003), Supplétion et classes flexionnelles dans la conjugaison du français, *Langages* 152, 102-126.
- Deglas, Maxime (2023), *Lexique 2023* (personal communication).
- Hazaël-Massieux, Marie-Christine (2002), Les créoles à base française: une introduction, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)* 21, 63-86.
- Henri, Fabiola (2010), *A constraint-based approach to verbal constructions in Mauritian*, Doctoral dissertation, University of Mauritius and Université Paris Diderot.
- McWhorter, John (1998), Identifying the Creole prototype: Vindicating a typological class, *Language* 74, 788-818.
- Missud, Alice (2018), Un corpus de créoles à base française issu du web pour l'extraction de paires base-dérivé, *Séminaire des Grammaires Créoles*, Structures Formelles du Langage, UMR 7023 (<http://www.UMR7023.cnrs.fr/grammaires-creoles-alice-missud-nanterre>).
- Seuren, Pieter and Wekker, Herman (1986), Semantic transparency as a factor in Creole genesis, in P. Muysken and N. Smith (eds), (1986), *Substrata versus universals in Creole genesis*, Amsterdam: Benjamins, 57-70.
- Valdman, Albert (1978), *Le créole : structure, statut et origine*, Paris: Klincksieck.
- Villoing, Florence, Tribout, Delphine, Henri, Fabiola and Deglas, Maxime (2024), La complexité morphologique des verbes du créole guadeloupéen, *Études créoles* 41.

# The long and short of it: contact and morpho-syntactic variation in Chagossian Creole verb forms

Jonathan Kasstan, Michelle Sheehan & Anand Syea  
(University of Westminster, Newcastle University, University of Westminster)

Keywords: Chagossian Creole, language contact, language change, morphosyntax, Indian Ocean Creoles

In this talk we marshal data from a pilot study to track morphosyntactic change in Chagossian Creole (CC), an endangered French-lexified Indian Ocean creole (IOC) spoken previously by the indigenous population of the Chagos Archipelago. Between 1965-1973, the British government forcibly displaced the entire population of the archipelago to Mauritius and the Seychelles, bringing speakers into contact with Mauritian and Seychellois Creole, along with other superordinate language varieties. Many Chagossians have since migrated to or have grown up in parts of the UK when they became entitled to citizenship in 2002 (Kasstan et al. 2024). We consider: (RQ1) to what extent CC is structurally aligned with Mauritian or Seychellois Creole, given some fifty years of forced displacement and language contact; and (RQ2) what inferences can be drawn from pilot data in terms of how contact-induced change has taken place.

Data come from semi-structured sociolinguistic interviews and a sentence-translation task administered to a small sample of 1st and 2nd generation Chagossians (n=8). We consider the production of long/short-verb form alterations, a well-documented feature in IOC varieties (cf. Corne 1977, Papen 1978, Syea 1992). While verbs in CC do not inflect for person, number, tense or aspect, most (but not all) verbs come in two forms distinguished by the presence or absence of final vowel, as in (1-2) below.

- |     |    |                      |     |      |       |     |
|-----|----|----------------------|-----|------|-------|-----|
| (1) | a. | mo                   | pe  | māze |       |     |
|     |    | I                    | ASP | eat  |       |     |
|     |    | 'I'm eating'         |     |      |       |     |
|     | b. | mo                   | pu  | vini |       |     |
|     |    | I                    | MOD | come |       |     |
|     |    | 'I will come'        |     |      |       |     |
| (2) | a. | mo                   | pe  | māz  | dipě  |     |
|     |    | I                    | ASP | eat  | bread |     |
|     |    | 'I'm eating bread'   |     |      |       |     |
|     | b. | mo                   | pu  | vin  | par   | bis |
|     |    | I                    | MOD | come | by    | bus |
|     |    | 'I will come by bus' |     |      |       |     |

In Mauritian/Seychellois Creoles, the choice of long or short verb forms depends on syntactic structure: transitive verbs followed by a DP complement take (near-categorically) the short form, whereas intransitive and verbs followed by a WH-trace surface with the long form (Corne 1977, Syea 1992, 2014). For Papen (1978), the two languages differed, however, with respect to PP/CP complements. In Seychellois, PPs of location triggered optionality, but we find an almost categorical preference for the short form in Mauritian (Papen 1978: 408-409). The reverse pattern was observed with CP complements introduced by *pur* ‘for’ which surface almost categorically with long-form verbs in Seychellois, but show optionality in Mauritian (Papen 1978: 408-409). In this work in progress, and adopting Papen (1978)’s data as a baseline, we use these differences to track the potential impact of language contact on this morphological domain among speakers with nuanced language-contact profiles.

### **Acknowledgements:**

This research is supported by the British Academy/Leverhulme Small Grants Scheme (SRG2324/240017).

### **References:**

- Corne, Chris (1977), *Seychelles Creole Grammar. Elements for Indian Ocean Proto-Creole Reconstruction*, Tübingen: Narr.
- Kasstan, Jonathan R., Patrick Allen, Pascalina Nellan & Michelle Sheehan (2024), Chagossian Creole, *Language Documentation and Description* 24(1), doi: <https://doi.org/10.25894/ldd.2492>.
- Syea, Anand (1992), The short and long form of verbs in Mauritian Creole: Functionalism versus Formalism, *Theoretical Linguistics* 18(1), 62-97.
- Syea, Anand (2014), *The Syntax of Mauritian Creole*, London: Bloomsbury.
- Papen, Robert Antoine (1978), *The French-based Creoles of the Indian Ocean: an Analysis and Comparison*, PhD thesis, University of California, San Diego.

# Innovative features in creole morphology

Marlyse Baptista, Soo-Hwan Lee & Ousmane Cisse  
(University of Pennsylvania, Gyeonsang National University & Boston University)

Keywords: morphological innovation, congruence, pattern & matter mapping, passivization

Word limit: 494

This paper focuses on reanalysis and innovation in Creole morphology and uses a systematic comparison between Creoles and their contributing lexifiers *and* substrates to argue for morphological inheritance on the one hand and morpho-phonological and semantic innovation on the other.

This presentation builds on Baptista (2020), which showcased the role of congruence in Creole formation and development using a competition-and-selection framework (Mufwene, 2001; 2008). Using empirical data from Kihm (1990) and Corne (1999), Baptista (2020)'s original proposal and model of Pattern and Matter Mapping (PMM) stipulate that morphosyntactic and semantic features are more likely to be selected into the grammatical makeup of a given Creole when they preexist and are shared by some of the source languages present in its linguistic ecology. Baptista originally based this proposal and the PMM model on the congruent forms that had been reported in the scholarly literature.

A new phase of this research (see Baptista and Cisse, in press) tests the congruence reported in the past literature by systematically comparing a set of features (passive marking involving the morphemes *ta* and *-du* for this particular presentation) in Portuguese, Mandinka and Wolof, three of the source language of *Kriolu* (Cabo Verde) by collaborating with native speakers of these languages. The current analysis shows that while the functional features of *ta* and *-du* could be inherited and potentially traced back to a particular language (*ta* < Portuguese *ta* (and possibly derived from Portuguese *estar* > *sta* > *ta* 'to be'), and *-du* < Portuguese *-do*), they have been reanalyzed to form new functional features, allowing a novel functional exponent for the morphemes *ta* and *-du* in passive constructions. The current analysis of *ta* and *-du* points to divergence from source languages and innovation with respect to the function of these morphemes but also for the passive constructions where they can occur. Let's focus on the morpheme *ta* and consider the examples in (1) featuring Kriolu, Mandinka and Portuguese:

- (1) a. Karta **ta** skrebe-du. / **Ta** skrebe-du karta. KRIOLU  
letter **TA** write-DU / **TA** write-DU letter.  
'Letters are written.'
- b. Wot-ôo dádáá-**ta** (le). MANDINKA  
car-DET repair-**TA** FOC  
'The car has been repaired.'
- c. Fomos chamados à noite (pelos policiais). PORTUGUESE  
were called at night (by the officers)  
'We were called at night by the officers.'

When exploring potential commonalities between the preverbal marker *ta* in Kriolu and the postverbal marker *ta* in Mandinka, we observe that both markers allow a theme argument to precede the verb, resembling passivization. However, they do not permit an overt *by*-phrase associated with an agent which cannot be overtly expressed, in contrast to Portuguese where it can (see “by the officers” in (1c)). On the other hand, the reading induced by *ta* in Kriolu is generic (1a) whereas it does not have to be in Mandinka (1b).

This presentation covers both the overlaps in passive markers *ta* and *-du* between Kriolu, Mandinka, and Portuguese, and the ways in which Kriolu markedly departs from its source languages.

### References

- Baptista, Marlyse (2020), Competition, selection, and the role of congruence in Creole genesis and development. *Language* 96 (1)160–199.
- Baptista, Marlyse and Ousmane Cisse (in press), Cabo Verdean Creole and Mandinka DP Domain: A Comparative Analysis. To appear in *The Open Handbook of (In)definiteness: A Hitchhiker’s Guide to interpreting bare arguments*, Veneeta Dayal (ed.). MIT Open Handbooks in Linguistics.
- Corne, Chris (1999), *From French to Creole: The development of new vernaculars in the French colonial world*. (Westminster Creolistics 5.) London: Battlebridge.
- Kihm, Alain. 1990. Conflation as a directive process in creolization. Bochum Essener Kolloquium, 111–37. Bochum: Brockmeyer.

# Free forms, bound forms, and morphological reanalysis: A comparative approach of Upper Guinea Creoles

Nicolas Quint  
(LLACAN – UMR8135 – CNRS/EPHE/INALCO)

**Keywords:** Bound & free forms - Creole languages – Morphology – Reanalysis – - Upper Guinea

Upper Guinea Creoles (UGC), a family of Afro-Portuguese languages with a shared proto-UGC ancestor, comprises three branches: ABC (Papiamentu), Continental (Casamance and Guinea-Bissau Creoles), and Insular (Capeverdean). All three branches share a common Afro-Portuguese heritage and a common proto-UGC ancestor (Quint 2000, Jacobs 2012, Quint & Moreira Tavares 2019). However, despite their shared origin, UGC have diverged significantly over five centuries due to varying influences and evolutionary paths. These divergences are particularly striking in morphology, where the same original marker can be either free or bound<sup>1</sup>, displaying varying degrees of reanalysis, depending on the grammar of each branch or sub-branch. This is illustrated by the morphological behavior of the proto-UGC past marker *\*ba(η)* (< Portuguese *-va* ‘1<sup>st</sup> group past imperfective’ X Joola *baŋ* ‘finish’) in the sentences below:

Papiamentu (ABC):	<i>e</i>	<i>tabata</i>	<i>kant-é</i>
	3SG.SBJ	IPFV.PST.IPFV	sing-3SG.OBJ
Santiago Capeverdean (Insular):	<i>e</i>	<i>ta</i>	<i>kantába-el</i>
	3SG.SBJ	IPFV	sing.PST-3SG.OBJ
Casamance Creole (continental):	<i>i</i>	<i>ta</i>	<i>kantá-l</i> <b><i>baŋ</i></b>
	3SG.SBJ	IPFV	sing-3SG.OBJ <b>PST</b>
			‘s/he used to sing it’

These forms show varying degrees of boundedness: in Papiamentu, the past marker is bound (inserted in a fixed sequence of three morphemes) and preverbal; in Capeverdean, it is bound (inseparable from the verb, unstressed) and postverbal; while in Casamance Creole, it is free (stressed, clitic pronouns may be inserted between it and the verb) and postverbal. Additionally, /ba/ has undergone semantic reanalysis across all three branches: from imperfective to anterior (or past). This reanalysis is further exemplified in Santiago Capeverdean, where *ba* can combine with the perfective: *e*  $\emptyset$  *kantába-el* (‘s/he had sung it’).

Santiago Capeverdean:	<i>e</i>	$\emptyset$	<i>kantába-el</i>
	3SG.SBJ	PFV	sing.PST-3SG.OBJ
			‘s/he had sung it’

Building on this analysis, I will examine other morphological markers across UGC, more specifically two verbal markers and two nominal markers: subjunctive (Santantonense Capeverdean *bé* /‘bɛ/ ‘go’ > *béss* /‘bɛs/ ‘go-SBJV.PST’); causative (Guinea-Bissau Creole *firbí* ‘boil (intr.)’ > *firbintí* ‘boil (tr.)’); diminutive (Casamance Creole *miñjer* ‘woman’ > *miñjer siñu* ‘little woman’); and triplicative (Santiago Capeverdean *fora* ‘help oneself again (food)’ > *trisfora* ‘help again for the **third** time’). For each marker, I will examine its degree of boundedness and reanalysis in relation to its lexifier and the substrate languages. This comparative approach is liable to enable us to better understand the evolution of morphology in contact languages

<sup>1</sup> The free/bound behavior of each marker is determined according to its prosodic and morphosyntactic autonomy in relation to the lexical word (noun, verb...) it interacts with (for more detail, see Creissels 2006 : 28-33 ; Dixon 2010 : 12-22 ; Haspelmath & Sims 2013 : 19-22 ; 196-203).

## References

- Creissels, Denis. 2006. *Syntaxe générale, une introduction typologique. Vol. 1: catégories et constructions*. Paris: Lavoisier.
- Dixon, Robert M. W. 2010. *Basic Linguistic Theory, Vol. 2: Grammatical Topics*. Oxford/New York: Oxford University Press.
- Doneux, Jean-Léonce & Jean-Louis Rougé. 1993. Gramática das línguas do país, gramática do crioulo. *Papia* 2(2). 50–59.
- Haspelmath, Martin & Sims, Adam. D. 2013. *Understanding Morphology*. London/New York: Routledge.
- Jacobs, Bart. 2012. *Origins of a creole: The history of Papiamentu and its African ties*. Berlin: De Gruyter.
- Quint, Nicolas & Ana Karina Moreira Tavares. 2019. The common African lexical core of Upper Guinea Creoles and its historical significance. *Journal of Ibero-Romance Creoles* 9. 115–161.
- Quint, Nicolas. 2000. *Le cap-verdien: Origines et devenir d'une langue métisse*. Paris: L'Harmattan.
- Rougé, Jean-Louis. 1994. À propos de la formation des créoles du Cap-Vert et de Guinée. *Papia* 3(2). 137–149.



# Sources of morphology in creole languages

Peter Bakker  
(Aarhus University)

Keywords: creoles, morphology, grammaticalization, language contact, change

There is a widespread idea about creoles that they have little or no morphology, beyond compounding and reduplication. The discussion in creole studies is about the presence of derivational and inflectional morphology (Kihm 2003; Farquharson 2007; Good 2012).

There are four possible sources for morphology in creoles: **inheritance** from ancestral languages, spontaneous development through merger of morphs via **grammaticalization**, **borrowing** of substrate/adstrate bound morphemes, and **convergence** through copying patterns present in other languages (Bakker in press). Imaginary examples of each would be, for plural: *house-s* (inherited), *them house* (3d person pronoun as grammaticalized plural marker), *house-lar* (Turkish suffix borrowed) and *mu-house* (English “much” used as a Bantu-like prefix).

In my presentation, I will show that most of the derivational and inflectional morphology markers are spontaneous developments through grammaticalization (Michaelis & Haspelmath 2020). The inheritance of lexifier derivational morphemes is severely limited, as are borrowed morphemes and copied processes (Bakker, in press).

The almost complete loss of lexifier inflection across the board in creoles is most easily explained if a process of reduction in an earlier phase in the history of the creole, between the lexifier and the resulting complex language (McWhorter 2002; Becker & Veenstra 2003).

The supporting data material is from creole languages with a wide range of lexifiers, and from different continents, including APiCS (Michaelis et al. 2013). Due to spelling conventions, and prejudices among linguists, the quantity of morphological patterns in creoles is underestimated, but still on average more modest than non-creoles (Bakker in press). One should remember, however, that there is nothing wrong or lacking in languages with little or no morphology (Kihm 2003: 345 N. 40) like Chinese are perfectly normal languages.

Support: Det Frie Forskningsråd, Grant 10.46540/4256-00026B.

## References

- Bakker, Peter (in press), Creoles and mixed languages, in Peter Arkadiev & Franz Rainer (Hrsg.), *Handbook of Historical Morphology*. Oxford: Oxford University Press.
- Becker, Angelika & Tonjes Veenstra (2003), Creole prototypes as basic varieties and inflectional morphology, in Christine Dimroth & Marianne Starren (eds.), *Information Structure and the Dynamics of Language Acquisition*, 235–264. Amsterdam: John Benjamins Publishing Company.
- Farquharson, Joseph T. (2007), Creole morphology revisited. In Umberto Ansaldi, Stephen Matthews & Lisa Lim (Hrsg.), *Typological Studies in Language*, vol. 73, 21–37. Amsterdam: John Benjamins Publishing Company.
- Good, Jeff (2012), Typologizing grammatical complexities: or Why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages* 27(1). 1–47.
- Kihm, Alain (2003), Inflectional categories in Creole languages. In Ingo Plag (Hrsg.), *Phonology and Morphology of Creole Languages*, 333–364. Berlin: De Gruyter.
- McWhorter, John (2002), The rest of the story: Restoring pidginization to creole genesis theory. *Journal of Pidgin and Creole Languages* 17(1). 1–48.
- Michaelis, Susanne Maria & Martin Haspelmath (2020), Grammaticalization in creole languages: Accelerated functionalization and semantic imitation. In Walter Bisang & Andrej Malchukov

(eds.), *Grammaticalization Scenarios from Africa, the Americas, and the Pacific*, 1109–1128.  
De Gruyter.

Michaelis, Susanne, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.) (2013), *The atlas of Pidgin & Creole language structures* (Oxford linguistics). Oxford: Oxford University Press.

# Expansion of Betawi *ber-* *-nya* nominalization under prolonged contact with Indonesian

Pristina Koon  
(University of Pennsylvania)

Keywords: language contact, language change, morphology, Pidgins and Creoles, nominalization

Betawi is an endangered Malay-lexified Creole spoken in the greater Jakarta area that emerged in the 17th century (Knörr 2014: 11). Languages that were also part of the contact ecology include Hokkien, Arabic, Sundanese, Javanese, and Balinese (Kurniawan 2018: 6). This presentation focuses on verbal and nominal morphology in Betawi. In previous work on morphologically marked intransitive verbs in the active voice, Ikranagara (1980), Muhadjir (1981), Wouk (1999), and Chaer (2023) analyze *be/ber-* (depending on phonological environment) as the primary prefix for marking active, intransitive verbs (see summary of their findings in Table 1). In the nominal domain, Ikranagara (1980) and Muhadjir (1981) present *-nya* (derived from Indonesian) as a nominalizing suffix that can be applied to adjective, bare verbs, and passivized verbs (marked with *di-*) in Betawi.

Drawing from a brand-new Betawi-language Wikipedia (2024), a Python-scripted search for *-nya* suffixation coinciding with *be(r)* prefixation yields the following preliminary results: Betawi can append *-nya* to verbs marked with the active, intransitive prefix *be(r)*, illustrating that this circumfixation (until now unattested in the literature on Betawi) seems productive in contemporary Betawi. (Wikipedia pages that supplied the following examples additionally had their edit history verified to ensure that at least one L1 Betawi contributor had worked on the Wikipedia page.) The following are examples taken from the Wikipedia data:

- (1) Be-rojol-nya  
ACT.INTR-born-NMLZ  
'birth, being born'
- (2) Ber-enti-nya  
ACT.INTR-stop-NMLZ  
'stopping, cessation'

Results from both corpus data and speaker judgments point to a contact-induced change in Betawi morphology. Contact with Indonesian, specifically, is most likely the source of this new morphological form since *be(r)--nya* constructions are felicitous in Indonesian. Proximity between Betawi and Indonesian has only increased since the 1990s because of intermarriage, interethnic neighborhoods, and heightened Indonesian dominance in the media and education (Imelda & Halimatusa'diah 2021: 18; Musgrave 2014: 95).

The emergence of this construction into the Betawi lexicon highlights the innovations that morphemes in contact can undergo, as illustrated in (1)-(2). The constructions do not reveal new morphemes have entered the language but rather old morphemes that have been subject to novel combinations – allowing *be(r)-* to be productive in additional morphosyntax environments than previously recorded. Moreover, as there are multiple forms of nominalization documented in Betawi (e.g. *m—nya*, *-nya*, *ke—an*, *pe—an*, *-an*), the newly attested *be(r)—nya* form suggests further examination into why an Indonesian-influenced parallel is selected by speakers over the

original Betawi equivalents. For instance, since active, intransitive verbs in Betawi often lack an overt transitivity prefix, what about the constructions in (1) and (2) lead the speakers to choose *berojolnya* and *berentinya* as opposed to *rojolnya* or *entinya*? In other words, are speakers borrowing the entire nominalization structure from Indonesian (e.g., word-level code-switching), or are they applying an Indonesian morphosyntactic structure to Betawi morphemes? Future research will answer these questions.

Table 1: Descriptive accounts of Betawi morphology

Source	Intransitive Prefix	-nya Nominalization Candidates
Ikranagara 1980	ber-	adjective, <i>di</i> -passivized, bare verbs
Muhadjir 1981	be-	adjective, <i>di</i> -passivized, bare verbs
Wouk 1999	be-	N/A
Chaer 2023	<i>me-</i> in limited contexts	Only provides instances of <i>ke-</i> , <i>pe-</i> , <i>-an</i> , and <i>-in</i> nominalization (and those resulting from combinations of these morphemes).

## References

- Chaer, Abdul. 2023. *Kamus Bahasa Betawi*. (Ed.) I Hasbullah. Jakarta: Dinas Kebudayaan Provinsi DKI Jakarta.
- Knörr, Jacqueline. 2014. Jakarta, Batavia, Betawi. In *Creole Identity in Postcolonial Indonesia*, vol. 9, 41–73. New York: Berghahn Books, Incorporated.
- Knörr, Jacqueline (2018): Creolizations and Pidginization as Concepts of Language, Culture and Identity. In Wilson Trajano Filho & Jacqueline Knörr (Hrsg.), *Creolization and Pidginization in Contexts of Postcolonial Diversity: Language, Culture and Identity* (Brill's Studies in Language, Cognition and Culture), vol. 17, 418. Leiden: Brill.
- Kurniawan, Ferdinan Okki. 2018. *Phonological variation in Jakarta Indonesian: An emerging variety of Indonesian*. Cornell University.
- Ikranagara, Kay. 1980. *Melayu Betawi grammar*. Jakarta: Badan Penyelenggara Seri NUSA Universitas Atma Jaya.
- Imelda & Halimatusa'diah (2021): Minority language revitalization: Betawinese and Lampungese case study. *Linguistik Terjemahan Sastra (LINGTERSA)* 2(1). 16–21.
- Muhadjir. 1981. *Morphology of Jakarta dialect, affixation and reduplication*. (Ed.) Amran Halim, Soepomo Poedjosoedarmo, I. Suharno & John W.M. Verhaar. (Trans.) Kay Ikranagara. Vol. 11. Jakarta: Badan Penyelenggara Seri NUSA Universitas Atma Jaya.  
<http://sealang.net/archives/nusa/pdf/nusa-v11.pdf>.
- Musgrave, Simon (2014): Language shift and language maintenance in Indonesia. *Language, Education and Nation-Building: Assimilation and Shift in Southeast Asia*, 87–105. Basingstoke: Palgrave Macmillan.
- Wouk, Fay. 1999. Dialect contact and koineization in Jakarta Indonesia. *Language Sciences* 21. 61–86.

## Gender marking in Mauritian creole

Shrita Hassamal & Anne Abeillé  
(MIE, Mauritius & LLF Université Paris Cité)

Keywords: Creole, Gender, Morphology, French, English

This project is funded by PHC Le Réduit and LabEx Empirical Foundations of Linguistics.

While the study creole morphology has been renewed (Luis 2015), creole languages are generally regarded as lacking grammatical gender. Some authors have recently argued it may be present in some French-based creoles: DeGraff (2001), for Haitian adjectives (*fanm Ayisyenn/\*Ayisyen* ‘Haitian.<sub>F/\*M</sub> woman’); Térosier (2023) for Martinican human nouns (*pwofeser/ez* ‘professor.<sub>M/F</sub>’) and (some) adjectives (*ere/erez* ‘happy.<sub>M/F</sub>’). We argue that Mauritian, another French-based creole, is developing some gender-marking morphology for (human) nouns and adjectives.

In French (Bonami & Boyé 2019), grammatical gender tends to be interpreted as social gender, for human nouns (Corbett 2012), and can be morphologically marked (*fermier/ière* ‘farmer.<sub>M/F</sub>’) or not (*homme/femme* ‘man/woman’), with a recent tendency for common gender nouns: *le/la ministre* ‘the.<sub>M/F</sub> minister’ (male/female minister), *le/la médecin* ‘the.<sub>M/F</sub> physician’ (male/female doctor).

Drawing on Carpooran (2019)’s dictionary, and Lalit’s online dictionary ([www.lalitmauritius.org](http://www.lalitmauritius.org)) as well as native informants, we show that in Mauritian gender alternation is systematic for ethnic nouns, and that feminine marking is developing for role nouns and some (human) adjectives:

- a. For kinship nouns, the related pairs are inherited from French: *prins/prinses* ‘prince.<sub>M/F</sub>’, *kouzin/kouzinn* (‘cousin.<sub>M/F</sub>’), sometimes with an agglutinated determiner (*tonton/matant* ‘uncle/<sub>1SG</sub> aunt’), or hindi (*nani/nana* ‘maternal grandfather/mother’).
- b. For ethnic nouns, which are derived from adjectives, Mauritian has kept the (spoken) French suffix alternation: *Morisien/ienn* (Mauritian.<sub>M/F</sub>), *Franse/ez* (French.<sub>M/F</sub>), *Sinwa/az* (Chinese.<sub>M/F</sub>), but with some innovations: *Arzantinien/ienn*, instead of French *Argentin/ine* ‘Argentinian.<sub>M/F</sub>’, or *kreol/inn* (‘creole.<sub>M/F</sub>’) instead of common gender French *creole*. Interestingly, the feminine form is not possible with non-human nouns, even when derived from French feminine nouns (*lamisik franse/\*ez* ‘music French <sub>M/\*F</sub>’ ‘French music’).
- c. For some role nouns, Mauritian has kept some (spoken) French suffixes: *santer/tez* (‘singer.<sub>M/F</sub>’), *direkter/tris* (‘director.<sub>M/F</sub>’), turning some into infixes (*metdekol/metresdekol* ‘schoolmaster.<sub>M/F</sub>’). Some gendered pairs (*pwinter/ez* ‘boy/girlfriend’) have a very different meaning from corresponding French pairs (*pointeur/euse* ‘pointer.<sub>M/F</sub>’). In some cases, there may be competition between suffixes: *doktores* or *doktris* (‘doctor.<sub>F</sub>’ ‘female doctor’, from *doktor* ‘male doctor’), which both existed in classical French (*doctoresse*, *doctrice* ‘female doctor’), and also between suffixation (*polisier* ‘police.<sub>F</sub>’, from *polisie* ‘police.<sub>M</sub>’ ‘policeman’) and compounding (*fam-polisie* ‘woman-police’) for ‘policewoman’. Mauritian has also borrowed from English (*biznesmenn / bizneswoman* ‘businessman/woman’) and *speaker/speakrinn* (‘speaker.<sub>M/F</sub>’), with the last feminine form from French *speakerine* (‘female TV presenter’).
- d. Some adjectives also have two forms (*fou/fol* ‘mad.<sub>M/F</sub>’, *briyan/t* ‘brilliant.<sub>M/F</sub>’, *malin/inn* ‘clever.<sub>M/F</sub>’), as in French, but the feminine form is only used with (feminine) humans (*enn tifi intelizant* ‘an intelligent.<sub>F</sub> girl’) and not with non-humans (*enn remark intelizan/\*intelizant* ‘an intelligent.<sub>M/\*F</sub> remark’).

We classify these gender markers as morphologically bound based on positional consistency and obligatoriness (*enn fam malinn/\*malin* ‘a clever.<sub>F/\*M</sub> woman’).

## References

- Bonami, Olivier & Gilles Boyé (2019), Paradigm and uniformity in the French gender system. In J. Audring & F. Masini (eds), *Morphological perspectives*, Edinburgh: Edinburgh University Press, 169–192.
- Carpooran, Arnaud (2019), *Diksioner Morisien* (Trwaziem edision). Le Printemps.
- Corbett, Greville (2012), *Gender*. Cambridge: Cambridge University Press.
- DeGraff, Michel. (2001), Morphology in creole genesis: Linguistics and ideology. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, Cambridge, MA: MIT Press, 53–121.
- Luís, Ana R. (2015), Rethinking creole morphology. *Word Structure* 8(2).
- Térosier, Stéphane (2023), A few remarks on gender in Martinican Creole. In E. Lacheret-Dujour, Sylvie Hancil & G. Bonfante (eds), *La grammaire est une fête / Grammar is a moveable feast: Mélanges offerts à / A webschrift for Anne Zribi-Hertz*, 327–340.

# The free-to-bound cline in Berbice Dutch Creole and its substrate origins

Silvia Kouwenberg

University of the West Indies, Mona Campus

**Keywords:** morphological boundaries; creole genesis; substratist explanations; reanalysis; word order

Berbice Dutch Creole (BD) is unique among Caribbean creoles in the number of functional morphemes which originate in a West-African substrate language. The BD substrate source has been identified as one or several Eastern Ijo varieties (EI) (Smith, Robertson & Williamson 1987). The functional morphemes include some that have moved along the free > bound cline in BD, but also several which have the same status w.r.t. boundedness in BD and EI. The forms are illustrated below, using Kalabari to illustrate the substrate form.

Table 1. *BD bound morphemes of Eastern Ijo origin. Kalabari forms contain marks for implosive stops ɓ, ɗ and -ATR vowels ɪ, ɛ, ɔ, ʊ.*

Berbice Dutch (BD)	Kalabari (KA)	Comments
1. -a, -arɛ IMPERFECTIVE, suffixes to verbs	-árɪ IMPERFECTIVE, suffixes to verbs	Bound status: Similar Accompanied by retention of PFV. NB Mood and Factative suffixes were not retained.
2. -apu PLURAL, suffixes to nouns, demonstratives; also: ASSOCIATIVE PLURAL, suffixes to proper nouns, pronouns and kinship-denoting nouns	apɯ 'persons' [+HUMAN, +PLURAL] ENCLITIC NOMINALIZER, postposed to verbs and clauses	Bound status: Change from (relatively) free EI enclitic nominal > bound BD suffix. NB the [+HUMAN, -PLURAL] counterpart was not retained.
3. ɟɛ NOMINALIZER, enclitic, combines with [+N] categories (nouns, adjectives, pronouns) only	yé 'thing,' [-HUMAN, -PLURAL] ENCLITIC NOMINALIZER, postposed to verbs and clauses	Bound status: Similar, but with reduced semantic content and different subcategorization properties in BD. NB the [-HUMAN, +PLURAL] counterpart was not retained.
4. ka, kanɛ NEGATOR, clause final free form; reduced ka has higher frequency	ɔká < ɔkɔ-á [thus-NEG] 'isn't it so?' (question tag)  ɔkáā 'it isn't so!' (exclamative)	Bound status: not straightforward. The EI standard negator is bound clause-final -áā in KA, -ya in Nembe and Kolokuma (also in KA drum language, hence presumed to be the archaic KA form). This suffix is incorporated into the tag negator which forms the basis for the BD negator. BD kanɛ is a clause-final free form which combines EI-derived ka with Dutch- derived *nɛ 'no'. Negatives in Swaving (1827) and Dance (1881) – the earliest published attestations of BD – contain the short form ka.

5. -tē PERFECTIVE, suffixes to verbs	-tēē PERFECTIVE, suffixes to verbs	Similar bound status. Retention possibly supported by similarity to Dutch PAST IMPERFECTIVE inflection <i>-de/-te</i> [-də/-tə]. Accompanied by retention of IMPFV. NB Mood and Factative suffixes were not retained.
6. -ma CAUSATIVE, suffixes to verbs	-ma CAUSATIVE, suffixes to verbs	Similar bound status. BD new formations containing -ma are evidence of early productivity of the suffix, but it was ultimately lost as productive derivation.

Aside from the data shown here, it is notable that there are several cases where EI complex forms containing bound materials have been reanalysed as simplex forms in BD. This is illustrated for instance by unanalysable BD *pote* '(be/become) old', and *mangiapu* 'maroon, runaway slave', which incorporate PFV -tē and PL -apu, respectively. That the BD forms are opaque is supported by the fact that they can be combined with these same suffixes to produce *pote-tē* [old- PFV] and *mangiapu-apu* [maroon-PL]. Conversely, there are also instances of backformation, resulting in BD forms that lack morphological material that would be required in EI, as illustrated in BD *ala* 'boss', which contrasts with KA *álá ɓɔ* [chief +HUM,-PL] (Kouwenberg 2009).

In this paper, I will attempt to account for the different responses illustrated above to EI material in the formation of BD. In particular, I will explore the consequences of the loss of EI head-final word order (Jenewari 1977) for the recognition of enclitic material in an emergent head-initial language. EI head-final word order is illustrated here with Kalaɓarɪ examples, drawn from fieldwork recordings. In (4), *ɗukɔ* 'allow' is preceded by its object clause. In (5), *ɛɛɓɔ* 'woman' is preceded by a relative clause.

- 1) Ó [kɛ á tɔn] ɗukɔ-áa.  
2PL [KE 3SG.F measure] allow-NEG.  
*You have not allowed (us) to take her measurement.* [2504\_F.C30]
- 2) O sime ɓo wɛrarɪ ɓɛ ɛɛɓɔ ma, á jen ma awɔmɛ,  
3SG.M marry come ASPECT POSS woman DEF.F, 3SG.F other two children,  
kɛ sɔa, ɓo wɛrarɪ.  
KE put, come ASPECT.  
*The woman he married, she added another two children.* [2504\_F.C30]

In this paper, I argue that the word order change made the disintegration of the verbal suffix system and the nominalization system in which the inherited forms participated inevitable, resulting in partial retention of paradigms, and in varying degrees of reanalysis of what was retained. The word order change was assisted by the existence in Kalaɓarɪ of grammatical patterns which use postmodification, thus deviating from the canonical head-final order and providing an apparently left-headed model in an otherwise right-headed language (Kouwenberg 1992, 1996). The reanalysis of EI complex forms in BD shows that access to underlying representations was eroded during the formation of BD. I will consider the further implications of the shift from the EI-type system to the creole system for our understanding of the role of substrate speakers in these processes.

**Acknowledgement:** Fieldwork in Guyana with the last living speakers of Berbice Dutch Creole and in the Kalaɓarɪ kingdom in Nigeria was funded by grants awarded by the Netherlands Foundation for the Advancement of Tropical Research (WOTRO, grant reference W39-183).



## References

- Dance, Rev. Charles D. (1881), *Chapters from a Guianese log-book*, Georgetown: The Royal Gazette Establishment.
- Jenewari, Charles E. (1977), *Studies in Kalabari syntax*, Unpublished Ph.D Dissertation, University of Ibadan.
- Kouwenberg, Silvia (1992), From OV to VO. Linguistic negotiation in the development of Berbice Dutch Creole, *Lingua* 88, 263-299. [https://doi.org/10.1016/0024-3841\(92\)90044-J](https://doi.org/10.1016/0024-3841(92)90044-J).
- Kouwenberg, Silvia (1996). Grammaticalization and word order in the history of Berbice Dutch Creole, in Ph. Baker, and A. Syea (eds), *Changing meanings, changing functions: Papers relating to grammaticalization in contact languages*, Westminster Creolistics Series 2, London: University of Westminster Press, 207-218.
- Kouwenberg, Silvia (2009), The invisible hand in creole genesis: reanalysis in the formation of Berbice Dutch, in E. Aboh, and N. Smith (eds), *Complex Processes in New Languages*, Amsterdam/Philadelphia: John Benjamins, 115-158. <https://doi.org/10.1075/cli.35.09kou>.
- Smith, Norval, Ian Robertson and Kay Williamson (1987), The *ljo* element in Berbice Dutch, *Language in Society* 16, 49-90.
- Swaving, Justus G. (1827), *Swaving's Reizen en Lotgevallen*, vol.1, Dordrecht: Blussé en Van Braam.

# Grammatical markers in Creole languages: Separate words or affixes?

Susanne Maria Michaelis, Leipzig University & MPI-EVA, Leipzig

Keywords: Morphology, Creole languages, affix, word, spelling systems

Inflectional morphology has been a key ingredient in assessing complexity in Creole languages (cf. Arkadiev & Gardani 2020 for the multifaceted notion of complexity in morphology in general). It has been claimed that synchronically Creoles strikingly show a lack of inflectional morphology, thus giving rise to morphologically extremely simplified languages (see for instance McWhorter 2001, Good 2012, Daval-Markussen 2014, Siegel et al. 2014). In this presentation, I will challenge this view by drawing on data from the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013). I will consider two morpheme types, tense-aspect markers and case markers, as shown in examples (1) and (2), in a large variety of Creoles:

(1) Ternate Chabacano (Spanish-based, Philippines; Sippola 2013)

*Ta yudá éle su marído*  
IPFV help 3SG 3SG.POSS husband  
'She helps her husband.'

(2) Korlai (Portuguese-based, South Asia; Clements 2013)

*Pedru su kadz tido kadz Pedr su*  
Pedru GEN house both house Pedru GEN  
'Pedru's house' 'both of Pedru's houses'

The main misconception which has given rise to the idea that Creoles lack inflectional morphology is the explicit or implicit role of spelling in deciding whether a morpheme is an affix or a separate morpheme (e.g. Siegel et al. 2014). But the two morphemes *ta* and *su* written as separate words in (1) and (2) can be interpreted as affixes. I will consider two criteria for affixhood: (i) uninterruptibility (no item except for other affixes can intervene between the grammatical marker and the stem) and (ii) host selectivity (the grammatical marker in question always selects the same kind of host root, e.g. only verb roots, or only noun roots) (Haspelmath 2023). Accordingly, in example (1), nothing can intervene between the imperfective marker *ta* and the verb root *yudá* 'help'. The same holds for the genitive case marker *su* and the possessor *Pedru* in example (2). Furthermore, both grammatical markers, *ta* and *su*, only combine with the same kinds of hosts in terms of word classes: *ta* with verbal roots and *su* with noun roots. Under this definition, instances of tense-aspect affixes are abundant in the Creole languages documented and analyzed in *APiCS*, and affixed case markers also occur repeatedly in Creole languages, despite them being written separately with respect to their hosts. Thus, I argue that such Creoles have indeed lost most of their lexifiers' inflectional marking, but have at the same time built up a considerable degree of new inflectional morphology through unusual accelerated grammaticalization processes.

## References

- Arkadiev, Peter & Francesco Gardani (eds), (2020), *The complexities of morphology*. Oxford: Oxford University Press.
- Daval-Markussen, Aymeric (2014), First steps towards a typological profile of creoles. *Acta Linguistica Hafniensia* 45(2), 274–295.
- Good, Jeff (2012), Typologizing grammatical complexities or why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages* 27, 1–47.
- Clements, J. Clancy (2013), Korlai structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/40>.
- Haspelmath, Martin (2023), Types of clitics in the world's languages. *Linguistic Typology at the Crossroads* 3(2), 1–59.

- McWhorter, John H. (2001), The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3), 125–166.
- Michaelis, Susanne, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds), (2013), *The Atlas of Pidgin and Creole Language Structures*. Oxford: Oxford University Press.
- Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann (2014), Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1), 49–85.
- Sippola, Eeva (2013), Ternate Chabacano structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds), *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/44>

# Questioning morphological boundaries with the help of Automatic Speech Recognition

William N. Havard & Emmanuel Schang  
(Université d'Orléans)

Keywords: Haitian Creole, Automatic Speech Recognition, Morphological segmentation, Low-resource languages, Definiteness Marking

Advancements in machine learning techniques open new avenues for investigating the identification of morphological boundaries in spoken Creole languages (Ferrand et al 2023). While it is a well-established point that Automatic Speech Recognition models can deliver accurate transcriptions even when trained on small corpora, the units discovered by Automatic Speech Recognition models often differ from those expected by linguists (Adda et al 2018; Scharenborg et al 2018; Bartelds et al 2023). Language models favor units optimized for information compression (e.g., Byte Pair Encoding) over linguistic units, despite using terms like 'word' and 'subword'. In the case of automatically transcribed speech corpora, one can wonder if ASR-derived segmentations align with linguistic expectations (cf. ANR DeepTypo project).

This study investigates the potential of automatic transcription to elucidate morphological boundaries through the analysis of approximately 1,400 hours of Haitian Creole recordings from the **Radio Haiti Inter Corpus** (<https://repository.duke.edu/dc/radiohaiti>). The findings, derived from the ANR CREAM project on the automatic transcription of spoken Kreyòl, underscore the efficacy of using a native Haitian acoustic model (see Havard et al., 2024 for the initial version of the model). Notably, the transcriptions were produced without the integration of a language model, ensuring that the data are decoded exclusively based on acoustic features. This methodological choice allows for an unmediated exploration of the phonological and morphological characteristics of the language.

Among our findings, we show that automatic transcription, although generally consistent across the corpus, reveals non-uniform segmentation for the word-initial attachment of 'la', such as *la limyè/lalimyè/limyè* ('light'). Our study also shows that 'la' attachment in Haitian Creole reveals more diverse attachment patterns than Martinican Creole, where *l(a)-* and *lé* form 'semantically definite' DPs denoting specific items in the manner of proper names and forms without *la* imply indefiniteness or non-singularity (Zribi-Hertz & Jean-Louis 2013). However, in our corpus *la* cannot be reduced to a fixed role:

- (1) a. *la kilti popilè a*  
LA culture popular DEF  
'the popular culture'  
b. *lakilti popilè*  
'the popular culture'  
c. *kilti oksidantal (la)*  
'the Western culture'

As illustrated in (1a), *la* and the definite article can co-occur (or not, as in 1b), despite *la* traditionally functioning as a marker of specificity. And (1c) shows that the definite can be omitted despite the absence of *la*.

Note that both 'kilti', 'lakilti' and 'kiltirèl' were present in the ASR training set, but that 'kiltir', 'la kiltir' and 'la kilti' were generated by the model based on acoustic features only.

Additionally, we discuss the challenges of using ASR for segmentation of morphological boundaries in spoken language by analysing suffix attachment errors (e.g., *lang aj* for *langaj* 'language' or *siperyè ite* for *siperyorite* 'superiority'), unseen in the ASR fine-tuning data.

Acknowledgments: This research is funded by ANR CS38, CREAM Project.

## References

- Adda, Gilles, Martine Adda-Decker, Odette Ambourou, Laurent Besacier, David Blachon, Hélène Maynard, Pierre Godard, et al. (2018). BULB: Breaking the unwritten language barrier. In *Computational methods for endangered language documentation and description*. Paris.
- Bartelds, Martijn, Nay San, Bradley McDonnell, Dan Jurafsky & Martijn Wiering (2023). Making more of little data: Improving low-resource automatic speech recognition using data augmentation. In *The 61st Annual Meeting of the Association for Computational Linguistics*.
- Ferrand, Éric Le, Fabiola Henri, Benjamin Lecouteux & Emmanuel Schang (2023). Application of speech processes for the documentation of Kréyòl Gwadeloupéyen. In *Proceedings of the second workshop on NLP applications to field linguistics*.
- Havard, W. N., Govain, R., Teixeira, D. G., Lecouteux, B., and Schang, E. (2024). Technologies de la parole et données de terrain: le cas du créole haïtien. In *Actes de JEP-TALN-RECITAL 2024. 31ème Conférence sur le Traitement Automatique des Langues Naturelles, volume 1: articles longs et prises de position*, pages 686–694. ATALA & AFPC.
- Scharenborg, Odette, Laurent Besacier, Alan Black, Mark Hasegawa-Johnson, Florian Metze, Graham Neubig, Sebastian Stüker, et al. (2020). Speech technology for unwritten languages. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28. 964–975.
- Zribi-Hertz, Anne & Loïc Jean-Louis (2013). From noun to name: Definiteness marking in modern Martinikè. In Patricia Cabredo Hofherr & Anne Zribi-Hertz (eds.), *Crosslinguistic studies on noun phrase structure and reference*, 269–315. Leiden: Brill.