WS14 Omnipredicativity: its core and its fringes

Katharina Haude & Albert Álvarez

The decline of omnipredicativity in Taracahitan languages

Launey (1994, 2004) has proposed the concept of omnipredicativity for describing languages where members of all lexical word classes can function equally and without any marking as predicates, and in which the predicative use is primary and the referential use is morphosyntactically derived. Importantly, omnipredicativity also implies a same morphosyntactic status and behavior for all lexical items, which means for instance, that nouns and verbs can both be used predicatively and referentially in the same ways.

In omnipredicative languages, the mere presence of a bare noun is thus interpreted as a nominal predication, the mere presence of an adjective is interpreted as an adjectival predication, and the same applies to verbs, as illustrated in (1).

(1) MAN   means   HE IS A MAN
     BLUE   means   IT IS BLUE
     WORK   means   HE/SHE WORKS

Classical Nahuatl, a Nahuan language from the 16th century spoken in Central Mexico, is often presented as the prototype of omnipredicative languages, since Launey (2004) has identified the following features of omnipredicativity, based on the morphosyntactic system of this old Southern Uto-Aztecan language.

    (a) Lack of copula.
    (b) Same subject markers in nouns and verbs
    (c) Existence of a spatial verb meaning 'be (somewhere)'
    (d) Zero morpheme in 3rd person subject
    (e) Evidence for coindexation between nouns and verbs
    (f) Lack of case
    (g) Genitive head-marked
    (h) Nouns and verbs can be both used as noun modifiers
    (i) Tense copula in noun predicates
    (j) Vocative marking on nouns
    (k) Demonstratives cannot be used as predicates
    (l) Noun predicates are always indefinite

These features can be used for identifying strongly and weakly omnipredicative languages, and therefore supports a continuum approach to omnipredicativity. In this presentation, I will use the different omnipredicative features proposed for Classical Nahuatl, in order to determine the degrees of omnipredicativity observable in other Southern Uto-Aztecan languages, belonging to the Taracahitan branch and corresponding to the extinct †Tegüima, †Eudeve, and †Cahita, to the 17th-century Tarahumara and to the present-day Mayo, Yaqui, Guarijío, and Tarahumara, all spoken in northwestern Mexico.

Based on this comparison and on the reconstruction of Proto-Uto-Aztecan made by Langacker (1977) and Dakin (1991), I will hypothesize about the omnipredicative system of Proto-Southern Uto-Aztecan (PSUA), trying to determine its core features. It will thus be

argued that juxtaposed nominal predications and the animacy/inanimacy distinction are key aspects of this original omnipredicative system.

This diachronic study will also show the decline of onmipredicativity in Taracahitan languages, characterized by different processes of grammaticalization involving original NP-use markers in juxtaposed nominal predications, and causing the differentiation between denominal and deverbal functions, as well as the emergence of deverbal nominalizers. This evolution will be illustrated with the grammaticalization paths undergone by the main NP-use markers reconstructed for PSUA (see 2) :-*tɨ* for the absolute/non-possessed, -*mɨ* for noun plural (only with animates/humans), -*wa* for animate possession, -*ye* for inanimate inalienable possession, and -*ra* for unspecified inanimate inalienable possession. I will thus argue that when combined with verbs, these markers have been reanalyzed as verb valency markers and nominalizers, implying the decline of omnipredicativity.

(2) MAN-tɨ   'the/a man'
    MAN-mɨ   'the men'
    MY   MAN-wa   'my man'
    MY   HAND-ye   'my hand'
    HAND-ra   'the/a hand of someone'

References

Dakin, Karen. 1991. Nahuatl direct and mediated possession: A historical explanation for irregularities. *International Journal of American Linguistics* 57(3): 298–329.
Launey, Michel. 1994. *Une grammaire omniprédicative*. Paris: CNRS Editions.
Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals* 57(1). 49-69.
Langacker, Ronald W. 1977. *An Overview of Uto-Aztecan Grammar. Studies in Uto-Aztecan Grammar.* Vol. 1. Dallas TX: UTA & SIL Publication 56.

Omnipredicativity in Khoekhoe (Khoe-Kwadi)

Khoekhoe (a Khoe-Kwadi language spoken primarily in Namibia) has been previously analyzed as omnipredicative. Khoekhoe has four open word classes: verbs, nouns, adjectives, and adverbs. They are clearly distinguished morphologically by the derivation morphemes applicable to them: e.g. only verbs take valency-changing suffixes such as passive, reflexive, and reciprocal. These word classes also show different syntactic behavior: e.g. in an NP, adjectives precede nouns they modify, as in (1). However, nouns and verbs share a striking syntactic similarity: The unmarked constituent order in Khoekhoe is SOV with the verb in the clause final position, as in (1) and (2). This clause-final position can also be occupied by word from other word classes, e.g. nouns (3) and adjectives (4), which, in this case, act as predicates. This structure is accompanied by the presence of the element *a* (glossed as 'A') in the present tense, which is also used with a small class of stative verbs, such as ǂ*an* 'know', as in (2). Furthermore, nouns in the predicative position do not have the otherwise obligatory person-gender-number suffix, compare ǀ*gôa-b* 'child-3M.SG' in (1) and (2) and ǀ*gôa* 'child' in (3).

In multiple publications within the tradition of transformational-generative grammar, Haacke (1976, 1977, 1978, 1980) suggested that all Khoekhoe NPs are essentially derived from relative clauses. In a more recent study in the framework of Head-Driven Phrase Structure Grammar, Hahn (2014) essentially agrees with this earlier analysis and argues that all nouns function as predicates and all argument NPs are derived as projections of pronominal elements (person-gender-number suffixes, such as -*b* in (1) and (2)), modified by relative clauses, so that e.g. (1) should be best translated as 'The one who is a lazy child is thinking'.

The presentation has two goals. First, it will "translate" the two theory-specific accounts of Khoekhoe omnipredicativity into a morphosyntactic description accessible to wide linguistic audience. Second, it will outline the basics of Khoekhoe morphosyntax and highlight the role of the two phenomena in these accounts, viz. of the person-gender-number suffixes (Hahn's "pronominal elements"), which are present on all referential expressions and might be considered determiners, and of the morphosyntactic distinction between active and stative predicates, such as the use of copulas.

In contrast to the earlier studies exclusively based on elicitations, this study relies on a large corpus of conversational Khoekhoe. This allows us to assess semantic differences between the predicative and referential uses and to address questions about pragmatic effects of swapped lexical categories and whether the semantic differences between the predicative and referential use are systematic.

1. *ǀowesa ǀgôa-b   ge   ra   ǂâi*
    lazy   child-3M.SG DECL PROG think
    'The lazy child is thinking.'
2. *nē   ǀgôa-b   ge   a   ǂan*
    PROX child-3M.SG DECL A   know
    'This child knows.'
3. *sa-ts   ge   (a)   ǀgôa*
    2-2M.SG   DECL A   child
    'You are a child.'
4. *nē   ǀgôa-b   ge   (a)   !gôao!nâ*
    PROX child-3M.SG DECL A   rude
    'This child is rude.'

References

Haacke, Wilfrid H. G. 1976. *A Nama grammar: The noun-phrase*. University of Cape Town MA thesis.

Haacke, Wilfrid H. G. 1977. The so-called 'Personal Pronoun' in Nama. *Khoisan Linguistic Studies* 3(3). 43–62.

Haacke, Wilfrid H. G. 1978. *Subject deposition in Nama*. Colchester: University of Essex dissertation.

Haacke, Wilfrid H. G. 1980. Nama coreferential copulative sentences re-assessed. In Jannie Winston Snyman (ed.), *Bushman and Hottentot linguistic studies*, 80–106. Pretoria: University of South Africa.

Haacke, Wilfrid H. G. 1992. Dislocated noun phrases in Khoekhoe (Nama/Damara): Further evidence for the sentential hypothesis. *Afrikanistische Arbeitspapiere* 29. 149–162.

Hahn, Michael. 2014. Predication and NP structure in an omnipredicative language: The case of Khoekhoe. Proceedings of the International Conference on Head-Driven Phrase Structure Grammar. https://doi.org/10.21248/hpsg.2014.13.

Moi, yet another language lacking a noun-verb distinction?

Daniel Krauße

(Lattice-CNRS, École Normale Supérieure)

Keywords: omnipredicativity, lexical categories, serial verb construction, Papuan languages

Languages that seem to lack a clear noun-verb distinction have been presented as a challenge for typological studies (Evans & Osada 2005, Gil 2013, Himmelmann 1991, Launey 2004, François f/c), since morphosyntactic principles in linguistics typically rely on generalizations at the level of lexical categories (cf. the fundamental distinction between VP and NP). While Austronesian languages are often cited as typical examples of such a system, either referred to as omnipredicative (François f/c) or monocategorial (Gil 2013), Papuan languages are rarely if ever discussed in this context. This talk examines the case of Moi, a Papuan language spoken in Eastern Indonesia. Moi belongs to the West Bird's Head family, which is surrounded by Austronesian languages and other Papuan (i.e. non-Austronesian) families.

In Moi, the distinction between nouns, adjectives and verbs is a tricky one, as shown by the examples (1-3). Whether a root may be analyzed as a noun like *miye* 'child' (1), an adjective like *bok* 'good' (2), or a verb like *egen* 'see' (3), the same pronominal prefix (with vowel harmony) is used in all cases.

(1) *Te-**miye** dau.*
1SG-**child** NEG
'I have no **child**(ren).'

(2) *To-**bok** dau.*
1SG-**good** NEG
'I am not **good**.'

(3) *Te-**gen** dau ya kam kulu.*
1SG-**see** NEG REL thing INDEF
'I don't **see** anything.'

The distinction between categories becomes even more blurred in examples like (4) where the same pronominal prefix occurs on *egen* 'see' and on *miye* 'child' within the same clause. Syntactically, *egen* is the head of the VP, and *miye* is the head of the object NP. The position of the negator *dau* tells us that it is the NP that is negated, not the VP (cf. negation tests for Moi in Menick 1996). Non-verbal phrases like (5) can even be marked for aspect.

(4) ***Me**-gen **me**-miye dau kulu.*
3SG.F-see **3SG.F**-child NEG INDEF
'She didn't see her child.' (Menick 2000:11)

(5) *Me-miye dala mele pe-kedi **se** a Amis.*
3SG.F-child male one 3SG.M-name **PERF** REL PN.M
'Her only boy was called Amis.' (Menick 2000:9)

I argue that the syntax of Moi is a good candidate for an omnipredicative system. This assumption has implications for the analysis of other complex syntactic structures. For example, while a string of "verb-like" items marked with the same pronominal prefix as in (6) could be analyzed as a *serial verb construction* (Aikhenvald 2018:2), an identical structure as in (7), despite also consisting of a string of items marked with the same pronominal prefix, would normally not fall within the scope of serial verbs.

(6) *Y-ein*   *dadi*   **ya**-*faguk*   **ya**-*fasos*   **ya**-*bo~bowo.*
    3PL-PRO   all   **3PL**-gather   **3PL**-make   **3PL**-INTS~happy
    'They all work together happily.'

(7) *Jonatan*   **wi**-*bik*   **wa**-*kafu*   **we**-*ne*   *y-e.*
    PN.M   **3SG.M**-play   **3SG.M**-be.with   **3SG.M**-person   3PL-EMPH
    'Jonathan plays with his friends.'

Drawing on novel data from recent fieldwork, I will present evidence that the underlying principle of Moi syntax is omnipredicative, permitting zero conversion between major lexical categories (things, activities, properties, and manner), while this is not the case for minor categories (free pronouns, function words, interjections).

References

Aikhenvald, Alexandra Y. 2018. *Serial Verbs.* Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Evans, Nicholas & Toshiki Osada. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology 9*: 351-390.

François, Alexandre. f/c. *Non-verbal predication in Oceanic languages.* In P. Marco Bertinetto, L. Ciucci & D. Creissels: *Non-verbal Predication in the World's Languages: A Typological Survey.*

Gil, David. 2013. *Riau Indonesian: a language without nouns and verbs.* In J. Rijkhoff & E. van Lier (eds.): *Flexible Word Classes: Typological Studies of Underspecified Parts of Speech* (89-130). Oxford: Oxford University Press.

Himmelmann, Nikolaus P. 1991. *The Philippine Challenge to Universal Grammar* (Arbeitspapier 15, Neue Folge). Institut für Sprachwissenschaft.

Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF − Language Typology and Universals 57*(1). 49-69.

Menick, Raymond H. 1996. *Verb Sequences in Moi.* In G. Reesink: *Studies in Irian Languages: Part II* (pp. 41-60). Jakarta: Universitas Katolik Indonesia Atma Jaya.

Menick, Raymond H. 2000. *Moi.* In G. Reesink: *Studies in Irian Languages: Part II* (pp. 7-24). Jakarta: Universitas Katolik Indonesia Atma Jaya.

Omnipredicativity in focus

Jürgen Bohnemeyer

(University at Buffalo – SUNY)

Keywords: omnipredicativity; focus; typology; Yucatec Maya; exhaustivity presupposition

This paper examines the hypothesis that in omnipredicative languages, clefting represents an optimal strategy for syntactic focalization, in the following sense: Clefts express focus relying solely on expressive resources that most (if not all) languages have grammaticalized independently – predication, subordination, and extraction (or filler-gap dependencies). And whereas in monopredicative languages, clefts are significantly more complex than monoclausal focus constructions, omnipredicativity simplifies clefts by obviating the need for a copula. Moreover, omnipredicative languages appear to show an affinity for head-marking, and head-marking exploits the indexing of core arguments as a cost-free resumptive element in the expression of filler-gap dependencies. These two properties render omnipredicative clefts typically simpler than their counterparts in monopredicative syntax.

Data from my fieldwork on Yucatec Maya illustrates that it is possible at least in first approximation to analyze all focus constructions in omnipredicative languages as clefts (as argued by Launey 2004 for Classical Nahuatl) and that it may not be trivial to distinguish between clause-internal focus constructions and clefts in such languages. Indeed, it has long been debated whether all focus constructions in this omnipredicative language (Bohnemeyer 2002: 153-166; Vapnarsky 2013) are clefts or whether the simplest Yucatec focus constructions, instantiated in (1) and (2) below, are monoclausal, with Gutiérrez Bravo & Monforte (2009) and Verhoeven & Skopeteas (2015) arguing in favor and Bricker (1979), Bohnemeyer (2002: 116-128), Tonhauser (2003), and Vapnarsky (2013) pursuing clefting analyses.

One argument against the monoclausal analysis is the requirement for DPs in focus position to be augmented with the independent pronoun *leti'* illustrated in (1) (Tonhauser 2003; Vapnarsky 2013). Under a monoclausal analysis, the exclusion of DPs from focus seems unmotivated. Under a cleft analysis, this constraint is explained with reference to the exclusion of DPs from use as predicates, which has been noted independently as an apparent correlate of omnipredicativity (Launey 2004: 65-67; a similar constraint is noted for example by Jelinek & Demers (1994: 711) for Straits Salish).

(1) [Context: Talking about don Luciano: Didn't he pass away last year?]
Hàah,   hàah,   (*/**leti')** le=nohoch   máak
true(B3SG)   true(B3SG)   it(B3SG) DEF=big   person
h-kim   te=[àanyoh   h-máan=o'].
PRV-die(B3SG)   PREP:DEF=year   PRV-pass(B3SG)=D2
'Right, right, it was him, the old man who died last year.'

Further, unlike argument DPs, foci can be negated. I take polarity to be the key diagnostic of syntactic predication.

(2) [Context: Who was screaming? Pointing at Pablo:]
Ma'   **leti'**h-áawat-nah-ih.
NEG(B3SG)   it   PRV-scream-CMP-B3.CMP
'It wasn't he who was screaming/shouting.' [elicited]

Verhoeven & Skopeteas (2015) observe that plain focus constructions as in (1) do not trigger exhaustivity presuppositions, unlike European-style clefts. This, however, holds for non-sentential equative predications in Yucatec as well, whose semantics I argue to be modeled on that of ascriptive predications due to head-marking.

(3) [Context: A messenger is asking for somebody named Alberto. One person says 'I'm Alberto'. Another says:]

Bèey=xan	tèen,	**Alberto-en**	**xan**.
SIMIL=also	me(B3SG)	Alberto-B1SG	also

'Me too, I'm also Alberto (i.e., my name is also Alberto).' [elicited]

I plan to round out the presentation with a preliminary typological survey drawing primarily on the languages represented in the Omnipredicativity workshop. Capitalizing on the properties identified above, the survey aims to determine to what extent focus is expressed in omnipredicative languages through resources that can clearly be distinguished from clefts.

References

Bohnemeyer, J. (2002). *The grammar of time reference in Yukatek Maya*. Munich: LINCOM.

Bricker, V. R. (1979). Wh-questions, relativization, and clefting in Yucatec Maya. In L. Martin (Ed.), *Papers in Mayan linguistics, vol. 3*. Columbia, MO: Lucas Brothers. 107-136.

Gutiérrez Bravo, R. & J. Monforte. (2009). Focus, agent focus and relative clauses in Yucatec Maya. In H. Avelino, J. Coon, & E. Norcliffe (Eds.), *New Perspectives on Mayan Linguistics*. Cambridge, MA: MITWPL. 83-95.

Jelinek, E. & R. A. Demers. (1994). Predicates and Pronominal Arguments in Straits Salish Language. *Language* 70(4): 697-736.

Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals* 57(1). 49–69. https://www.degruyter.com/document/doi/10.1524/stuf.2004.57.1.49/html.

Skopeteas, S. & E. Verhoven. (2015). Licensing focus constructions in Yucatec Maya: An empirical study on the association with focus. *International Journal of American Linguistics* 81(1): 1-40.

Tonhauser, J. (2003). F-constructions in Yucatec Maya. In J. Anderssen, P. Menéndez Benito, & A.Werle (Eds.), *Proceedings of Semantics of Under-Represented Languages in the Americas II*. Amherst, MA: Graduate Linguistics Student Association. 203–223.

Vapnarsky, V. (2013). Is Yucatec an omnipredicative language? Predication, the copula, and focus constructions. *Sprachtypologie und Universalienforschung 66*(1): 40-86.

Key to morpheme glosses

1/3 – 1st/3rd person; B – Cross-reference Set-B; CMP – Completive; D2 – Indexical clause-final particle; DEF – Definite; NEG – Negation; PREP – "Generic" preposition; PRV – Perfective; SG – Singular; SIMIL – Similative.

A New Hypothesis of the Origins of Omnipredicativity in Nahuatl

Magnus Pharao Hansen
(University of Copenhagen)

Classical Nahuatl is the language originally used by Michel Launey to formulate the typological category of omnipredicativity (Launey 1986, 1994, 2004). However, the morphosyntactic profile of Nahuatl stands out as highly divergent from other languages in the Uto-Aztecan language family, which do not display omnipredicative traits. The divergent typology of Nahuatl within the Uto-Aztecan family is generally thought to be the result of intense contact with the languages of the Mesoamerican Sprachbund. Pre-Nahua lost the Uto-Aztecan verb-final syntax and the use of switch-reference markers and syntactic case marking on nouns, and introduced relational nouns, and an omnipredicative, polysynthetic morphosyntax.

Understanding the paths of language change through which omnipredicativity may arise, would be important for our understanding of how typologically uncommon patterns may develop from more common ones. Recently, Gildea (2023) has proposed that in several omnipredicative languages the trait developed as etymologically nominalized verbs became the nuclei of new main clause constructions, suggesting that reanalysis of nominalizations could also be a possible source for the omnipredicative patterns found in Nahuatl.

The present paper proposes that Nahuatl omnipredicativity is the combined result of three steps in the development of proto-Nahuatl beginning with language convergence:

1. Pre-Nahua developed holophrasis, as the SOV sequence of pronominal clitics, accusative pronouns and verbs (a.) fossilized and coalesced into a single polysynthetic verb complex (b.).

a. Pre-Nahua > b. proto-Nahuatl
*mehka=nɨ ma-tsi tseiya > *wehka ni-mits-tʃi:ya
far.away=1SGS 2sg-ACC watch > far 1SGS-2SGO-watch
"I see you far away" "I see you far away"

2. Because holophrasis did not permit external nominal arguments, subordinate clauses were used to introduce nominal arguments as adjuncts in focus constructions (c.). Pre-Nahua used the Uto-Aztecan switch reference marker *-tɨ (same subject) to mark nouns in focus position as subordinate predicates (d.).

c. Pre-Nahua > d. proto-Nahuatl
*ta:ka-t ikɨ tseiya > *ta:ka-tɨ Ø-ki-tʃi:ya
man-ABS 3SG watch > man-SS 3SGS-3SGO-watch
"the man watches it" "he (who) is a man, he watches it"

3. Switch reference markers were eventually reanalyzed as absolute suffixes marking all unpossessed nouns in argument position, extending predicative readings to *all* nouns (e).

e.   Proto-Nahuatl (and Current Nahuatl)
        ni-k-tʃi:ya   *in*   *Ø- tɬa:ka-tɬ*
        1sGS-3sGO-watch   DEF   3sgS-man-ABS
        "I see him, he (who) is a man/I see the man"

Under this hypothesis, Nahuatl omnipredicativity did not arise from nominalization. While the Uto-Aztecan switch-reference markers did originally have a nominalizing function, which the cognate morphemes in for example Ute and Wixárika retain, in Nahuatl their use appears to have been the opposite: that of creating denominal predicates.

This hypothesis solves two major questions in the development of Nahuatl, showing them to be interconnected: namely the origin and function of the absolute suffix *-tli/-tl* which is ubiquitous on free standing nouns in Nahuatl, and the omnipredicative properties that motivate an analysis seeing all nouns as subordinate predicates. The trigger for both of these developments appear to have been the adaptation of holophrasitic polysynthesis, which required a reorganization of the entire grammatical system.

The Nahuatl case suggests that there may be various paths of language development that lead to traits of omnipredicativity.

Gildea, Spike. (2023) *One path to the genesis of omnipredicativity*. Paper given in the seminar: "Equational Predications: Omnipredicativity, Non-verbal Predications or Zero Verb Constructions, and the Diachrony of Copulas", Universidad de Sonora.

Launey, Michel. (1986) *Catégories et opérations dans la grammaire nahuatl*. Paris: Université de Paris 7, thèse d'Etat.

Launey, Michel. (1994) *Une grammaire omniprédicative. Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.

Launey, Michel. (2004) *The features of omnipredicativity in Classical Nahuatl*. STUF-Language Typology and Universals, 57(1):49-69.

Existential predications in Tagalog and Totoli and what they tell us about omnipredicativity

Maria Bardají & Nikolaus P. Himmelmann

Universitat de Barcelona & Universität zu Köln

One grammatical context that is rarely discussed in work on omnipredicativity is existential constructions. In non-omnipredicative languages there typically are various constraints on what can be the complement of an existential operator. In English, for example, only mass nouns can be used without an article in this context (*There is food on the table* but not **there is nail on the table*). Finite verb forms are generally excluded (**there is goes*), only non-finite derivations are allowed (*there was (a lot of) shouting in front of the house*).

If omnipredicativity is primarily defined by syntactic uniformity (all content words may occur – without semantic or morphosyntactic changes – in all syntactic contexts) then it should be the case that in omnipredicative languages there are no differences in existential constructions that originate in word class differences. Specifically, there should be no constraints on the syntactic category of the complement of an existential operator. This seems to be the case in Tagalog, as seen in the following examples: in (1) the complement of the existential quantifier is a noun, in (2) it is a finite verb (in actor voice) with its thematic argument.

(1) **may libro** *sa* *mesa*
    EXIST book LOC table

    'there is a book on the table'

(2) **may nagaalaga'** *doon* *sa* *ibun*
    EXIST AV.RLS.RDP:care_for DIST.LOC LOC bird

    'there was already someone looking after the birds'

By adding a nominative argument – in (3) and (4) this is a pronoun (*kayo* and *ako*, respectively) – the existential construction becomes a possessive construction. Again, the existential complement can be any kind of content word.

(3) **may mga anak** *na* *ho* *ba kayo?*
    EXIST PL child already HON Q 2PL.NOM

    'Do you already have children?'

(4) **may ipapakita** *ako* *sa* *iyo*
    EXIST CV-RDP-CAUS-visible 1SG.NOM LOC 2SG.DAT

    'I have something to show you.'

In this contribution, we will review different options for analyzing the syntax of these constructions, with a special focus on the forms that appear to be finite verbs. We argue that these forms have participant-oriented meanings and in this regard resemble participles. More generally, we will show what it means to be syntactically uniform in the context of existential constructions. We do this by comparing the Tagalog data with data from Totoli, an Austronesian language which shares many features with Tagalog but which clearly

distinguishes nouns and verbs as major word classes. On first sight, there do not seem to be major differences between the existential constructions in Tagalog and Totoli. That is, examples 1-4 can be matched with Totoli examples which superficially look identical. Importantly, however, with verbs the semantics (and arguably also the syntax) is different from the one found in Tagalog, as illustrated by (5):

(5) *tapi* ***ingga daan notumbuli*** *kode* *<no>* *notumalib*
     tapi   ingga daan no-tumbuli   kode   <no>   no- -um-talib
     but   NEG   EXIST AV.RLS-stop.by   only   AV.RLS- -AUTO.MOT-pass.by

     'but he did not stop by, he just passed by'

As seen in (5), the construction consisting of existential *daan* plus a verb in actor voice denotes an event and does not introduce an individual ("there is an X" or "there is something/someone that/who is an X"). Instead, the existential operator emphasizes the truth value of the proposition, a reading that is not possible/does not arise in Tagalog.

We conclude with some observations on what the differences between Tagalog and Totoli imply for the concept of omnipredicativity.

Abbreviations

<> = false start; 1 = first person; 2 = second person; AUTO.MOT = autonomous motion; AV = actor voice; CV = conveyance voice; DAT = dative; DIST = distal; EXIST = existential ; LOC = locative; NEG = negation, negative; NOM = nominative; PL = plural; POT = potentive; PRX = proximal; RLS = realis; SG = singular; ST = stative.

Omnipredicativity and indexing in Oceanic languages and beyond

Matthew Micyk & Eva van Lier, University of Amsterdam

In this presentation we address the relation between omnipredicativity and no/optional verbal indexing. Launey (2004) asserts that omnipredicative languages are more likely to display no/optional subject indexing. We operationalise 'omnipredicativity' following Croft's (2001) parts of speech theory, as a language in which object/person-denoting words (semantic 'nouns') can be used in predicative function without additional coding (i.e. no copula), and may additionally express verbal features (potentially including indexing). We test Launey's claim using three datasets, zooming out from a family-specific survey (36 Oceanic languages), via a medium-scale multivariate typology (52 languages), to a large-scale world-wide survey based on binary Grambank data (1943 languages).

The first dataset consists of 36 Oceanic languages, which are often described in the literature as examples of omnipredicativity (Van Lier 2016, 2017a,b). For these, we coded the morphosyntactic expression of nominal predication and the presence/absence of (optional) S-indexing. We found that the majority of languages do not use a copula for nominal predication *and* have obligatory S-indexing on verbal predicates. Moreover, many allow the expression of S-indexing in nominal predicates. These findings, which are contrary to Launey's assertion, are likely related to Oceanic indexes typically being loosely integrated with their host, i.e. they are clitics or even independent forms, which facilitates their application across categories (cf. Hengeveld 2013). Consider for instance the Vamale example in (1):

(1) *Jacob   tha=**a**=juu   xavee*
    J.   ASS=**3SG**=really   fool
    'Jacob really is a fool.' (Rohleder 2023: 262)

Our second dataset contains 52 languages world-wide with optional S-indexing, i.e. S-indexing which is absent under some conditions (based on Walker & Van Lier, under review). For these languages, we again coded the morphosyntax of nominal predication as well as verbal reference. In this dataset, only 8 languages never do nominal predication with additional coding. On the other hand, 33 languages sometimes do nominal predication without extra coding. And, only 10 languages have some degree of overlap between conditions for no indexing and no copula in nominal predicates. Thus, it does not seem like a lack of indexing usually results in no copula. Furthermore, most languages require overt coding of verbs in referential function always or sometimes.

The third dataset consists of 1943 languages from Grambank (Skirgård et al. 2024) coded for three relevant features (two on S-indexing and one on copula use in nominal predicates). Via a Bayesian statistical analysis accounting for phylogenetic and areal effects with the methods outlined by Guzmán Naranjo & Becker (2021), we find no effect of S-indexing status on copula use in nominal predicates. However, we do find that copula use in nominal predicates is more effected by areal effects than phylogenetic effects (Micyk under review).

Summarily, based on these data , we suggest that the unexpected profile of Oceanic languages, which typically combine unmarked nominal predication with obligatory S-indexing, is due to the relatively independent status of indexing morphemes in these languages, a property that aligns well with lexical flexibility. Considering languages world-wide, our data do not show a clear relationship between optionality/absence of indexing and omnipredicativity.

References

Croft, W. 2001. *Radical Construction Grammar. Syntactic theory in typological perspective.* Oxford: Oxford University Press.

Guzmán Naranjo, M., & Becker, L. (2021). Statistical bias control in typology. *Linguistic Typology*, 26(3), 605–670. https://doi.org/10.1515/lingty-2021-0002

Hengeveld, K. 2013. Parts-of-speech system as a basic typological determinant. In J. Rijkhoff & E. van Lier (eds.), *Flexible word classes: Typological studies of underspecified parts of speech*. Oxford: Oxford University Press, pp. 31-55.

Launey, M. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF – Language Typology and Universals 57(1).* 49-69. https://doi.org/10.1524/stuf.2004.57.1.49

Micyk, M., under review. A Grambank typology of omnipredicativity: Copula use in nominal predication and subject indexing

Rohleder, J. 2021. A grammar of Vamale. A language of New Caledonia. PhD dissertation, Bern University.

Skirgård, H. et al. 2024. Grambank reveals global patterns in the structural diversity of the world's languages. *Science Advances* 9. DOI: 10.1126/sciadv.adg6175

Van Lier, E. 2016. Lexical flexibility in Oceanic languages. *Linguistic Typology* 20(2). 197–232. https://doi.org/10.1515/lingty-2016-0005

Van Lier, E. (ed.) 2017a. Special issue on Lexical flexibility in Oceanic languages. Studies in Language 42(2). https://doi.org/10.1075/sl.41.2

Van Lier, E. 2017b. The typology of property words in Oceanic languages. *Linguistics* 55(6). 1237–1280. https://doi.org/10.1515/ling-2017-0027

Walker, K. & E. van Lier, under review. A typology of conditional indexing.

The rise of nominal predication in Indo-Aryan -

Omnipredicativity, verb class flexibility and the role of language contact

Uta Reinöhl and Antje Casaretto (University of Freiburg)

Indo-Aryan verbal morphology transitioned historically from a complex set of finite categories to a much-simplified participial system (Masica 1991). Modern Indo-Aryan languages mostly rely on a contrast between imperfective (formerly present) and perfective (formerly, an undergoer-oriented deverbal adjective) participles, in part in combination with auxiliaries. This talk outlines this shift as one of emerging noun-verb flexibility and omnipredicativity (Launey 2004).

In early Old Indic (from ca. 1300 BC), the forms in question could already be used as main predicates, but not yet with all their arguments expressed, or even informationally foregrounded. With the deverbal adjective as main predicate, only undergoer arguments were overt. With the present participle, only agentive arguments were. The respective second argument surfaced only in modifying or secondary-predicative contexts (Jamison 1979). The rise of, specifically, the deverbal adjective into a fully-fledged main predicate is central in research into the origin of ergativity in Indo-Aryan (cf. Verbeke 2013 for an overview). Studies researching this development tend to focus on the labelling and appropriate characterization of the forerunner, i.e. pre-ergative, construction and its reanalysis (Klaiman 1978, Hock 1986, Dahl & Stroński, eds, 2016).

We here adopt a fresh perspective, showing that all ingredients for the modern, participial-based constructions were already available in early Old Indic. In contrast to most linguistic studies on early Old Indic which focus on the poetically structured Vedas, we analyse approximately one thousand clauses in the slightly younger and more naturalistic Vedic prose texts. Rather than focusing only on main predication, we advocate in our analysis for the larger morpho-syntactic system, as well as socio-linguistic factors, to be taken into account. The evidence of Indo-Aryan areal and social expansion points to a significant number of L2 learners in the transition from early Old Indic – Vedic Sanskrit – into the standardized, hegemonic Classical Sanskrit (Pollock 2009) and Early Middle Indo-Aryan (Peterson 1998). This suggests a scenario where the rise of ergativity can be understood as an expansion of predicative participial constructions from subordinate to main-clause contexts under pressure of an L2 learner bias for morphological simplicity (Lupyan & Dale 2008, Trudgill 2010). Given the importance of participial and similar forms in the rise of other nominally-dominant grammars, e.g. in Semitic and Austronesian languages (Sasse 1999, Himmelmann 2009), this contact scenario holds explanatory potential for understanding the diachronic developments of noun-verb flexibility and omnipredicativity beyond Indo-Aryan.

Dahl, Eystein & Krzysztof Stroński (eds.). 2016. *Indo-Aryan ergativity in typological and diachronic perspective*. Amsterdam & Philadelphia: John Benjamins.

Himmelmann, Nikolaus P. 2009. Notes on Tagalog Nominalism. *Theoretical Linguistics* 35. 115-123.

Hock, Hans Henrich. 1986. P-oriented constructions in Sanskrit. In B. Krishnamurti, C.P. Masica & A.K. Sinha (eds.), *South Asian languages: Structure, convergence and diglossia*, 15–26. Delhi: Motilal Banarsidass.

Klaiman, Miriam H. 1978. Arguments against a passive origin of the IA ergative. *Chicago Linguistic Society: Papers from the 14th regional meeting*, 204–216.

Launey, Michel. 2004. The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals* 57(1). 49–69.

Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PloS ONE* 5. 1-10

Jamison, Stephanie W. 1979. Remarks on the expression of agency with the passive in Vedic and Indo-European. *Zeitschrift für vergleichende Sprachforschung* 93 (2), 196-219.

Masica, Colin P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Peterson, John. 1998. *Grammatical relations in Pali and the emergence of ergativity in Indo-Aryan*. Munich: Lincom.

Sasse, Hans-Jürgen. 2009. Nominalism in Austronesian: a historical typological perspective. Comments on Daniel Kaufman's "Austronesian Nominalism and its consequences". *Theoretical Linguistics* 35(1). 167–181.

Trudgill, Peter. 2010. *Investigations in sociohistorical linguistics: Stories of colonisation and contact*. Cambridge: Cambridge University Press.

Verbeke, Saartje. 2013. *Alignment and ergativity in New Indo-Aryan languages*. Berlin/Boston: de Gruyter.

A discourse-based approach to categoriality in a Philippine language

Voltaire Q. Oyzon & Thomas E. Payne
(Leyte Normal University & University of Oregon)

Keywords: categoriality, parts of speech, multi-dimensional scaling, discourse, omnipredicativity

Waray is a major language spoken in the Eastern Bisayas region of the Philippines. It is a strongly omnipredicative language according to the criteria set out by Launey (1994, 2004). In this paper we describe a discourse study designed to determine the degree to which the categories "noun", "verb" and "adjective" are established in the lexicon of Waray.

Philippine languages have been claimed to be "precategorial" (Foley 2007), meaning that lexical roots are not inherently divided into word classes. Dictionaries, however, typically insist on strictly distinguishing nouns, verbs, adjectives, and other word classes. We see these positions as representing two extremes on a continuum: 1) Precategoriality (no categories exist in the lexicon), and 2) Absolute categoriality (all lexical entries are inherently categorized). Our intent was to test these positions against language in use.

The methodology was to select thirty common lexical roots, ten nouns, ten verbs, and ten adjectives, as characterized in three popular dictionaries of Waray (Tramp 1997, Makabenta & Makabenta 2004, and Abuyen 2005). We then examined the uses of these roots in a large body of Waray texts consisting of transcribed conversations, stories and published material (Oyzon n.d.). Three discourse functions were independently identified: reference (the prototypical function of nouns), predication (the prototypical function of verbs), and referent modification (the prototypical function of adjectives). We then correlated each instance of each lexical root with these discourse functions using multi-dimensional scaling. The results show that the lexicon of Waray exhibits neither precategoriality nor absolute categoriality. Rather, individual roots do have characteristic discourse-functional uses, but they are not required to function in one and only one way.

This study represents a challenge for theories that depend on a priori categorization of lexical items. It also identifies another potential characteristic of omnipredicative languages.

References

Abuyen, Tomas. A. (2005). *Dictionary English Waray-Waray/Tagalog*. Manila: National Book Store.

Foley, William. (2007). The place of Philippine languages in a typology of voice systems. In S. Musgrave & P. Austin (eds.), *Voice and grammatical relations in Austronesian languages*, 22–44. Stanford: CSLI Publications.

Launey, Michel. (1994). *Une grammaire omniprédicative: Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.

Launey, Michel. (2004). The features of omnipredicativity in Classical Nahuatl. *STUF - Language Typology and Universals* 57(1). 49–69.
https://www.degruyter.com/document/doi/10.1524/stuf.2004.57.1.49/html.

Makabenta, Eduardo. A. and Makabenta, Eduardo T. (2004). *Binisaya-English; English-Binisaya Dictionary* (2nd ed.). Quezon City: Adbox.

Oyzon, Voltaire, n.d. 3NS Corpora Project (corporaproject.org, accessed 17 October, 2024). National Network of Normal Schools.

Tramp, George. D. (1997). *Waray-English Dictionary*. Decatur, GA: Dunwoody Press.

WS15 Origins and functions of future formations in the world's languages

Gilles Authier & Steven Kaye

What's in a future? : The Abkhaz-Abaza future tense flip-flop

Anton Antonov
(INALCO-CRLAO-CNRS-EHESS, Paris)

Goal of the talk   In this talk I will try to explain the functional flip-flop of two future tenses in two closely related North West Caucasian languages, Abkhaz and Abaza.(cf. Table 1).

The facts   Both languages have two types of FUTURE tense: while FUTURE 1 is a relatively unmarked form with low epistemic import, used in utterances concerning future events in general, FUTURE 2 is more epistemic and deontic than temporal in nature. Examples 1 and 2 illustrate their use in Abkhaz [cf. (Chirikba 2003: pp. 44–45, 53) and (Hewitt 2010: 137ff, 164ff)].

	Abkhaz	**Abaza**
Future I	*-p'*	*-št'*
Future II	*-št'*	*-p'*

Table 1: Future tense suffixes in Abkhaz and Abaza

(1) *s-ca-p'*
    1SG-go-FUT1
    'I shall go.'

(2) *s-ca-š-t'*
    1SG-go-FUT2-FIN
    'I shall (probably) go.'

The problem   Although Lomtatidze (2006: pp. 158–9) treats Abaza's future tenses as being (semantically) equivalent to their cognates in Abkhaz (cf. examples 3 and 4), both Tabulova (1976: pp. 151–2) and Arkadiev (2020: pp. 27–8) describe their actual use as different from the one they have in Abkhaz.

(3) *d-ca-p'*
    1SG-go-FUT1
    '(S)he will go.'

(4) *d-ca-wa-š-t'*
    1SG-go-IPFV-FUT2-FIN
    '(S)he will (probably) go.'

Tabulova (1976: pp. 151–2) follows the Abkhaz tradition in treating *-p'* as the exponent of Future I, and *-št'* as the marker of Future II. Nevertheless, she labels Future I as an 'indefinite' future whose use implies that 'the speaker is not certain that the action will take place', while Future II is a 'definite' future which implies that 'the speaker has no doubt the action will take place'.

As for Arkadiev (2020: pp. 27–8) he labels the future in *-št'* as Future I and the one in *-p'* as Future II. He goes on to specify that "Future I (i.e. *-št'*) is by far the most frequent form of the two, being used for neutral reference to events following the speech time such as predictions or plans". Interestingly, he states that "The semantic difference between Future I and Future II (i.e. *-p'*) is not yet well-understood." A quick check in a parallel corpus (the Abkhaz and Abaza translations of the Gospel of Matthew) seem to confirm that although both future tense morphemes are transparently cognate, there seems to have been a flip-flop in the actual use of the forms between Abkhaz and Abaza: the unmarked member of the pair is *-št'* in Abaza vs. *-p'* in Abkhaz.

A tentative solution After an overview of the modern use of the two FUTURE tenses in these two languages, I will propose a tentative diachronic explanation for this apparent flip-flop in their actual use. It hinges crucially on the fact that the Abaza *-št'* future is attached to the non-finite imperfective converb *-wa* (cf. ex. 4), whereas in Abkhaz it attaches to the bare verb stem (as does the *-p'* future in both languages).

This hypothesis seems borne out by the fact that the most common way in Abkhaz of expressing certainty in reference to a future action or state of affairs is to use the present tense (Hewitt 2010: pp. 87, 137), which is also formed on the imperfective converb in Abkhaz, but not in Abaza, where it is never used with future reference. This is confirmed by the usage found in the four Abkhaz translations of the Gospel of Matthew which seem to prefer the present tense whereas Abaza uses rather systematically the future in (*-št'*) wherever the Russian or English version have a future tense form.

References

Arkadiev, Peter. 2020. Abaza. 2nd draft. https://www.academia.edu/43306221/Abaza_A_grammatical_sketch.

Chirikba, V. A. 2003. *Abkhaz* (Languages of the World/Materials 119). München: LINCOM.

Hewitt, George. 2010. *Abkhaz: a comprehensive self-tutor*. LINCOM EUROPA.

Lomtatidze, Ketevan. 2006. *Abazinskij jazyk (kratkoe obozrenie) [The Abaza language: A short overview]*. Tbilisi: Universali.

Tabulova, Nurja T. 1976. *Grammatika abazinskogo jazyka. Fonetika i morfologija [A Grammar of Abaza: Phonetics and Morphology]*. Čerkessk: Karačaevo-Čerkesskoe otdelenie Stavropol□skogo knižnogo izdatel□stva.

Language Contact as an Obstacle to the "Present > Future" Evolution

The grammaticalization pathway from 'progressive' to 'future' appears to have been discussed for the first time by BYBEE et al. (1994), who suggested that "the future use of a present is a late development." HASPELMATH (1998) has examined this evolution in more detail, proposing a scenario of diachronic development comprising two independent grammaticalization paths, one leading from progressives to habituals and the other from progressives to futures. In TATEVOSOV's approach (cf. BYBEE et al. 1994), the following schema is presented, including intermediate stages before reaching the future meaning: *progressive > habitual > ability > root possibility > epistemic possibility > predictive future > prospective/intentional future*. MAISAK (cf. AUTHIER & MAISAK, 2011), who studies this phenomenon in Lezgic languages, argues that "only in Lezgian has the Habitual/Future [= former present tense] become a "true" generalized Future, approaching the final stage of the "present > future" evolution".

In this presentation, we will show how contact with Azeri, a Turkic language, may have hindered this evolution in certain Lezgian dialects spoken in Azerbaijan. For instance, in the dialects of Sumağallı (spoken in the Ismayıllı region) and Qımıl (spoken in the Quba region), the former present tense (= *eventual*) has not acquired a future value, and their future domain has aligned with that of Azeri, characterized by a dichotomy between two opposing futures based on the feature *certain/uncertain*, as in (1) and (2).

These data will be compared with those from the Yargun dialect (spoken in the Qusar region of Azerbaijan), where the "present > future" evolution has unfolded without obstacles, and the former present functions as the main future tense, as in (3). In addition to the *eventual*, this dialect has three other forms of the future (3).

The Lezgi dialects of Azerbaijan have been very minimally studied to date. Our second objective is to explore the origins of these present and future forms, which have often undergone various morphophonological changes, making their equivalents unrecognizable across dialects.

(1) Sumağallı

Uncertain future:

haq'ıl	*q'il.i-z*	***x-qӧӧz***	*ada-n*
intelligence	head-DAT	REV-come.IPFV.FUT(I)	that.OBL-GEN

'He would regain his senses.'

Certain future:

ini-z	*erid*	*dıšman*	***qӧӧli***
here.OBL-DAT	seven	enemy	come.IPFV.FUT(II)

'Seven enemies will come here.'

(2) Qımıl

Uncertain future:

q'il-e-va-y	*haq'ıl*	***aqatkki-j***	*ada-n*
head-IN-be.in-PRT	intelligence	go_out.IPFV-FUT(I)	that.OBL-GEN

'He would lose his wits.'

Certain future:

kačal	*za*	***req'iila***
bald	1SG.ERG	kill.IPFV.FUT(II)

'I shall kill the Bald Man myself.'

(3) Yargun

Generalized Future:

kkačal	*za*	*juv-a*	***req'i-da***
bald	1SG.ERG	self1/2-ERG	kill.IPFV-EVT

'I shall kill the Bald Man myself.'

Hoped-for future:

ida	*ayal*	*ik'*	*xhii-la,*
this.OBL(GEN)	child	so	become.AOP-TEMP

zu	*ayal=ni*	*ik'*	***je-di***
1SG.GEN	child=ADD	so	become.IPFV-FUT

'If her child has had all of this, mine will have it too.'

Intentional future:

ini-z	*örüd*	*dušman*	***qve-r-val***	*ya*
here.OBL-DAT	seven	enemy	come.IPFV-AIMPP-MAN	COP

'Seven enemies are expected to come here.'

Indefinite future:

belki	*zun*	***fi-n***
maybe	1SG	go. IPFV/OP-HORT

'Maybe I will go tomorrow.'

References

ALEKSEEV, MIXAIL E. (1985). *Voprosy sravnitel'no-istoricheskoj grammatiki lezginskix jazykov. Morfologija. Sintaksis.* Moscow: Nauka.

AUTHIER, GILLES & MAISAK, TIMUR A. (2011). *Tense, aspect, modality and finiteness in Est Caucasian languages.* Bochum: Universitätsverlag Dr. N. Brockmeyer.

BYBEE, JOAN; PERKINS, REVERE & PAGLIUCA, WILLIAM (1994): *The evolution of grammar: Tense, aspect and modality in the languages of the world.* Chicago: University of Chicago Press.

HASPELMATH, MARTIN (1993). *A Grammar of Lezgian.* Berlin: Mouton de Gruyter.

HASPELMATH, MARTIN (1994). "The tense system of Lezgian". In: THIEROFF, Rolf & BALLWEG, JOACHIM, eds. *Tense systems in European languages.* Linguistische Arbeiten 308. Tübingen: Niemeyer.

HASPELMATH, MARTIN (1998). "The semantic development of old presents: new futures and subjunctives without grammaticalization". *Diachronica* 15/1, pp. 29–62.

Future in Enggano

Charlotte Hemmings
(University of Oxford)

This paper discusses the expression of future in Enggano, an Austronesian language spoken in Sumatra, Indonesia, by comparing a contemporary corpus collected as part of a language documentation (Sangian et al 2024, henceforth *Contemporary Enggano*) with a corpus of legacy materials collected in the 1930s (Kähler 1940, 1955-64, 1975, henceforth *Old Enggano*).

In Old Enggano, future and volitionality were expressed via the suffix -*a*, which is possibly cognate with the PMP subjunctive marker (Edwards 2015):

(1)   a.   *'ua   ki-pudu-**a**   kia*
             1SG   KI-kill-FUT   3SG
             'I will kill him.' (Kähler 1940 Grammar, 51.2)

       b.   *KE'anaha   ka-ø-kõkõnã=hã   e-koyo   ki-nõõ-**ã**   e-ko'E'E*
             then   3-BU-appear=EMPH   DIR-pig   KI-eat-FUT   DIR-devil
             'At that the pigs came out, who wanted to eat the devil' (Kähler 1955, 14.4)

Although Enggano generally permits independent verbs to occur in *ki-, bu-* or bare form, future/volitional verbs with -*a* only occur with the *ki-* prefix.

In Contemporary Enggano, the final vowel of every word is regularly lost (Yoder 2011, Smith 2020). Consequently, the -*a* suffix no longer surfaces. Instead, the difference between the future and the base form is marked only by the resurfacing of the Old Enggano root vowel (and any support consonants):

(2)   a.   *Ki   ki-puak*
             3SG   KI-go.back
             'He went back.' (Verbal Morphology 04, 130)

       b.   *No'man   ki   ki-**poka-h***
             tomorrow   3SG   KI-go.back-FUT
             'Tomorrow he will go back.' (Verbal Morphology 04, 135)

In Old Enggano, the root form for 'go back' was *puaka* and the future form *kipuakaha* with a regular allomorph of the suffix attached to roots ending in /a/. In Contemporary Enggano, however, morphophonological changes have made the connection between the root *puak* and the future form *kipokah* less transparent, as shown in (2).

Perhaps for this reason, Contemporary Enggano has innovated a regular means of expressing future using the auxiliary verb *buh*:

(3) a. *ki   ki-pua-h   no'man*
        3SG   KI-run-FUT   tomorrow
        'He will run tomorrow' (Verbal Morphology 03, 236)

    b. *ki   ki-**buh**   pu   no'man*
        3SG   KI-FUT   run   tomorrow
        'He will run tomorrow' (Verbal Morphology 03, 237)

We believe that this is grammaticalized from the Old Enggano nominalization *e'obuho* 'thing that must be done', which was formerly used to express deontic modality:

(4) *E'ana   **e-obuo-ho**   bu-pee   i'ioo   ka-'ano   u-kabake*
     DEM   DIR-do-PAT.NOM BU-give   OBL   HUM.PL-friend OBL-dead.person
     'That he must give to the friends of the (person) killed.'

This is a common pattern of grammaticalization (Kuteva et al, 2019). However, it is also reinforced by language contact with Indonesian, which has a periphrastic future construction with *akan*.

(5) Mereka   **akan**   tiba   pada   tanggal   enam   Maret
     3PL   FUT   arrive on   date   six   March
     'They will arrive on the sixth March.' (Sneddon et al, 2012: 136)

Consequently, this paper argues that the existence of two different patterns of future marking in Contemporary Enggano results from the specific context of sound change, grammaticalization and the reinforcing effects of language contact. This has interesting implications for our understanding of the diachronic development of future forms cross-linguistically.

References

Edwards, Owen (2015), The position of Enggano within Austronesian, *Oceanic Linguistics* 54(1), 54-109.

Kähler, Hans (1940), Grammatischer Abriß des Enggano, *Zeitschrift fur Eingeboren-sprachen* 30, 81-117, 182-210, 296—320.

Kähler, Hans (1955-1964), Texte von der Insel Enggano [Texts from the island of Enggano], *Afrika und Ubersee* 39 (1955), 89-94; 41 (1957), 153-156; 42 (1958), 179-189; 43 (1960), 299-312; 44 (1960), 41-71; 45 (1961), 205-225; 46 (1962), 127-145; 48 (1964), 46-58.

Kähler, Hans (1975), *Texte von der Insel Enggano (Berichte uber eine untergehende Kultur) [Texts from the Island of Enggano (Reports of a Declining Culture)]*, Berlin: Verlag von Dietrich Reimer.

Kuteva, Tania, Bernd Heine, Bong Ho, Haiping Long, Heiko Narrog & Seongha Rhee (2019), *World lexicon of grammaticalization*, 2nd ed, Cambridge: Cambridge University Press.

Sangian, Engga Zakaria, Dendi Wijaya, Charlotte Hemmings, I Wayan Arka, Mary Dalrymple. *2024), *Documentation of Contemporary Enggano*, Unpublished corpus, https://enggano.ling-phil.ox.ac.uk/

Smith, Alexander D. (2020), Nasalization in Enggano Historical Phonology, *Oceanic Linguistics* 59(1/2), 347-365.

Sneddon, James Neil, Alexander Adelaar, Dwi N Djenar & Michael C Ewing (2012), *Indonesian: A comprehensive grammar*, London: Routledge.

Yoder, Brendon (2011), Phonological and phonetic aspects of Enggano vowels,   North   Dakota: University of North Dakota MA Thesis.

Future time reference in Georgian: A diachronic study on preverbation

Diego Luinetti

(Università degli Studi Guglielmo Marconi)

Keywords: Grammaticalization, Preverbation, Future time reference, Diachrony, Georgian.

The Georgian tempo-aspectual system has been described as an intersection between Romance-style and Slavic-style systems (Tomelleri & Topadze 2015; Bondarenko 2017), as both perfectivity and telicity have independent grammatical status. However, little attention has been paid to the role of these two parameters in the expression of future time reference.

Modern Georgian (MG) features a dedicated Future tense, whose formal marking varies depending on the lexical aspect of the verb. Telic verbs form the Future by adding an originally spatial preverb to the corresponding Present tense form (e.g., MG *v-cer* 'I write' vs. *da-v-cer* 'I will write'), whereas atelic verbs form the Future by means of the circumfix *i/e-...-eb* (e.g., MG *v-tamaš-ob* 'I play' vs. *v-i-tamaš-eb* 'I will play') (Aronson 1982; Hewitt 1996; Makharoblidze 2012). Notably, the first morphological strategy for Future marking resembles the synthetic Future found in Slavic languages (e.g., Russian *ja na-pišú* 'I will write'), a typologically rare phenomenon (Dahl 1985).

The picture is different in Old Georgian (OG), which lacks a dedicated Future tense. Future time reference is instead expressed contextually through tenses whose primary temporal reference is the Present, or whose main function is modal (Conjunctive I, Conjunctive II, and Mixed Conjunctive) (Fänrich 1982). Notably, in OG, preverbs carry a purely lexical meaning, with no grammatical function (Veshapidze 1967; Harris 2003; Makharoblidze 2018).

This talk investigates the evolving relationship between preverbs and future time reference across the diachrony of Georgian in a twofold manner. First, it aims to describe the grammaticalization path of preverbs; second, it explores whether there is a correlation between perfectivity and preverbation in Old Georgian (OG).

To pursue the first objective, we compare a portion of the OG New Testament with its corresponding Modern Georgian (MG) translation, focusing on the presence or absence of a preverb in verbal forms expressing future time reference. Preliminary results indicate that in only 15% of cases does an unpreverbed OG form correspond to a preverbed MG form. Mismatches are predominantly found with low-telicity verbs, which do not entail a change of state in the undergoer. We argue that high-telicity verbs, which already show a strong tendency toward preverbation in future contexts, may have served as a model for low-telicity verbs in the process of preverb grammaticalization.

To address the second objective, we examine the compatibility of preverbed forms in OG with aspectually restrictive temporal adverbials, following the approach adopted by Bertinetto (1986) for Italian and Dahl (2010) for Sanskrit. Our hypothesis is that if preverbed forms are found to be exclusively compatible with a perfective reading, then the preverbed Future in MG may have developed from a Perfective Present, in a manner analogous to the origin of the Perfective Future in Russian.

References

Aronson, Howard I. (1982), *Georgian. A reading grammar*, Bloomington: Slavica Publisher.

Bondarenko, Tat'jana I. (2017), Two aspectual subsystems of Georgian [*Dve asspektual'nye podsistemy v gruzinskom jazyke*], *Voprosy jazykoznanija* 65(4), 74-100.

Bertinetto, Pier Marco (1986), *Tempo, aspetto e azione nel verbo italiano: il sistema dell'indicativo*, Firenze: Accademia della Crusca.

Dahl, Eystein (2010), *Time, Tense and Aspect in Early Vedic Grammar: Exploring Inflectional Semantics in the* Rigveda, Leiden / Boston: Brill.

Dahl, Östen (1985), *Tense and Aspect Systems*, Oxford / New York: Basil Blackwell.

Fänrich, Heinz (1982), *Grammatik der Altgeorgischen Sprache*, Tbilisi. [translation of Šaniʒe, Aḳaḳi, *Elemenṭaruli çigni ʒveli kartuli enis šesçavlad*]

Harris, Alice C. (2003), Preverbs and their origins in Georgian and Udi, in G. Booij, J. Van Marle (eds.), (2003), *Yearbook of Morphology 2003*, Dordrecht: Springer, 61-78.

Hewitt, George (1995), *Georgian. A structural reference grammar*, London / New York: Routledge.

Holisky, Dee Ann (1981), Aspect theory and Georgian aspect, in P. Tedeschi, A. Zaenen (eds.), (1981), *Tense and Aspect*, Leiden / Boston: Brill, 127-144.

Makharoblidze, Tamar (2018), On Georgian Preverbs, *Open Linguistics* 4, 163-183.

Makharoblidze, Tamar (2012), *The Georgian Verb*, Munich: LINCOM Studies in Caucasian Linguistics.

Tomelleri, Vittorio S. & Topadze Gäumann, Manana (2015), Aspectual pairs in Georgian: some questions, *STUF* 68(1), 49-85.

Veshapidze, Irakli (1967), *The preverb in Old Georgian* [*Zmnisçini ʒvel kartul enaši*], Tbilisi State University.

When there is no future: A corpus study of future time reference in Ingrian

Fedor Rozhanskiy & Elena Markus
(University of Tartu & University of Tartu)

Keywords: future tense, Finnic languages, Ingrian, corpus study, morphology

In Finnic languages, there is no specific verbal marker to express the future reference. Nevertheless, these languages possess a range of lexical and grammatical means to convey this semantic function (see, for example, Metslang 1996, Norvik 2013, 2015a, 2015b, and Ylikoski 2016).

This paper focuses on expressing future reference in Ingrian, one of the least documented Finnic languages. Our research is a corpus study based on Soikkola Ingrian texts, collected from the last generation of Ingrian native speakers in 2006–2014.

In Ingrian, the future reference can be expressed in the following ways:

1. By the verb *leej(j)ä* 'to be (in the future),' which is limited to existential semantics, e.g., *maa-munad kažvoid, kaľlaa **lęęnöö*** 'The potatoes have grown, there **will be** fish.'

2. By using a construction consisting of the verb *noišša* 'to become' and the supine of the main verb. This construction is only felicitous with imperfective actions, e.g., *lehto **noiššoo kažvamaa*** 'The underbrush **will grow**.'

3. Through DOM with a transitive verb in the present tense. If the object is encoded with a total case, the action most likely refers to the future, e.g., *oššan kera šen maťteerin* **buy.PRS.1SG** also that.**GEN** material.**GEN** 'I **will** also **buy** that material' (cf. *oššan šidä maderia* 'I **am buying** that.**PART** material.**PART**').

4. By using lexical means, primarily temporal adverbs, e.g., *miä **illaašt** šiiž šiulle šaunan tęęn* 'Then **in the evening** I will make you a sauna.'

5. With no overt marking, so the temporal reference is inferred solely from context, e.g., *šiä jälešt tǫǫ, miä mään ęęž* 'You go behind (them), and I **will go** in the front.'

One might expect that with such a rich choice of linguistic means for expressing future reference, their use would be frequent, and the question to investigate is which variant is chosen in a particular context. In fact, our study draws a different picture. In our corpus, 176 contexts expressing future reference were attested. The hierarchy of means based on their frequency was the following:

No overt marking: 108 cases (61%) > imperfective constructions with *noišša* 'to become': 42 cases (24%) > *leej(j)ä* 'to be in the future': 15 cases (9%) > temporal adverbs: 6 cases (3%) > construction with the total object: 5 (< 3%).

Apparently, in most cases, the meaning of the future tense is not explicitly marked and can only be inferred from context. If this meaning is marked, the most common construction involves an auxiliary verb *noišša* 'to become' and the supine form of the lexical verb, a structure found also in other Finnic languages (with various auxiliary verbs). All other means are significantly less frequent. Although the Finnic DOM is known as a way of expressing future reference (Laakso 2022: 248), it is used very rarely in Ingrian. Our study suggests that the absence of the future tense in a system of grammatical categories does not entail that the language involved has difficulty in expressing the corresponding meaning, but merely that it does not treat its expression as a grammatical necessity, and as a result employs the available means rather episodically than systematically.

References

Laakso, Johanna (2022), Finnic, in M. Bakró-Nagy, J. Laakso, and E. Skribnik (eds), *The Oxford guide to the Uralic languages*, Oxford: Oxford University Press, 240–253.

Metslang, Helle (1996), The Developments of the Futures in the Finno-Ugric Languages, in M. Erelt (ed), *Estonian Typological Studies I*, Tartu: University of Tartu, 123–144.

Norvik, Miina (2013), Future time reference in the Finnic languages: LEE(NE)-verbs, *Suomalais-Ugrilaisen Seuran Aikakauskirja* 94, 125–164.

Norvik, Miina (2015a), The past participle constructions LEE(NE)- + PTCP and SAA- + PTCP as future time reference devices: the example of Livonian against a Southern Finnic background, *Suomalais-Ugrilaisen Seuran Aikakauskirja* 95, 201–236.

Norvik, Miina (2015b), *Future Time Reference Devices in Livonian in Finnic Context*, Tartu. (Dissertationes Linguisticae Universitatis Tartuensis 25).

Ylikoski, Jussi (2016), Future time reference in Lule Saami, with some remarks on Finnish, *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 7(2), 209–244.

Future marking in Romani: Areal innovations across dialect groups

Kirill Kozhanov

(University of Potsdam)

Keywords: future, aspect, Romani, language contact, areal linguistics

Romani is an Indo-Aryan language spoken primarily in Europe since the Middle Ages (Matras 2002; Matras and Tenser 2020). Due to its use by multilingual communities across extensive territories, Romani is an ideal subject for studying linguistic variation and contact-induced innovations.

Romani has no traces of Old or Middle Indo-Aryan synthetic future forms (Beníšek 2020: 37). In Late Proto-Romani—a stage of the language spoken in the Byzantine Empire prior to its spread across Europe in the 15th century—the future was likely not morphologically distinct from the present forms (Matras 2002: 157). However, the verbal affix -*a*, of uncertain origin, existed during this stage, and its initial function was to mark "indicative mood" (Matras 2001, 2002), or "progressive aspect" (Scala 2023).

Over time, in some dialects, the "long" forms with the affix -*a* became specialized for indicating future events, while the same forms without the suffix were restricted to expressing the present tense, as shown in (1).

(1)   Servitika Romani (fieldwork data)

a.   *Šun-áv   ďil-í*

         listen-PRS.1SG   song-ACC.SG

         'I am listening to a song'

b.   *Šun-áv-a   ďil-í*

         listen-1SG-FUT   song-ACC.SG

         'I will listen to a song'

Furthermore, some Romani dialects copy future marking from contact languages. This includes the 'want'-future found in the Balkan dialects and an analytical future based on the auxiliaries such as 'take' or 'be' in the Romani dialects of Eastern Europe (Boretzky 1999). Ex. (2) illustrates the latter.

(2)   Servitika Romani (fieldwork data)

         Av-áva   te   šun-é   ďil-í

         AUX-1SG.FUT   COMP   listen-INF   song-ACC.SG

         'I will be listening to a song'

As shown in examples (1b) and 2, the same variety can employ multiple strategies for marking the future, with additional semantic distinctions. For instance, (1b) describes a perfective event in the future, whereas (2) is used to denote imperfective events. This particular Romani variety is spoken in the Ukraine and southern Russia and partially replicates the TAM-system found in East Slavic.

In this talk, I will address the following research questions:

1) What is the overall dsitribution of future marking across Romani dialects? Does it correlate with areal patterning or dialect classification?

2) When multiple marking strategies are available, what distinctions do they convey?

To answer the first question, I draw on the data from the Romani morphosyntax database (RMS), a questionnaire-based resource that contains information on 120 locations in Europe. By selecting a subset of 24 sentences that include future reference, I tagged future marking for all varieties. The results reveal that initial areal innovations in future marking are well preserved within dialect groups.

To address the second question, I present two case-studies based on transcribed narratives from two distinct varieties: Kishinevare Romani (20,000 tokens) and Russian Romani (52,000 tokens). In Kishenevare Romani, zero marking of the future is the default strategy, while *kam*(*a*)-marked future events tend to have additional modal meaning of possibility. In contrast, in Russian Romani, the difference between two markings reflect aspectual interpretation: *-a* forms typically denote perfective events, whereas the analytical future is reserved for imperfective future events.

References

Beníšek, Michael (2020), The historical origins of Romani, in Y. Matras, and A. Tenser (eds), (2020), *The Palgrave Handbook of Romani Language and Linguistics*, Cham: Palgrave Macmillan, 13–47.

Boretzky, Norbert (1999), Grammatical Interference in Romani: Loan Formations for Foreign Categories, *Acta Linguist. Hung*. 46 (3–4), 169–200.

Matras, Yaron (2001), Tense, Aspect and Modality Categories in Romani, *STUF - Language Typology and Universals* 54(2), 162-180.

Matras, Yaron (2002), *Romani: A linguistic introduction*, Cambridge: Cambridge University Press.

Matras, Yaron and Anton Tenser (eds), (2020), *The Palgrave handbook of Romani language and linguistics*, Cham: Palgrave Macmillan.

RMS — Romani morpho-syntax database. Currently available at https://romani.dch.phil-fak.uni-koeln.de/

Scala, Andrea (2023), Romani short and long present: some historical and typological aspects. Handout of the paper presented at the International conference on Romani linguistics, Milan.

Anchiq Karata future system and its Andic context

Konstantin Filatov

Linguistic Convergence Laboratory, HSE University

Keywords: future grams, infinitive, old present drift, Andic languages, Nakh-Daghestanian languages

This talk seeks to (i) provide a detailed descriptive and typological analysis of the future forms in the Anchiq dialect of Karata — an underdescribed Andic language of the North-West Daghestan (< Avar-Andic-Tsezic < Nakh-Daghestanian, see the table of forms below), and (ii) compare the Anchiq system with other Andic systems to explore potential diachronic processes that have influenced the variation in these systems.

In Anchiq, two main forms are used to denote future time reference: Synthetic Future and Periphrastic Future. Their uses overlap to a large extent: they both can be used in intentional and predictive contexts. The difference between their semantics in these contexts is best described by what Bybee et al. (1994) call "future possibility vs. future certainty" respectively. Synthetic Future is avoided in prospective and planned future contexts. Moreover, Synthetic Future tends to be used in perfective contexts, while Periphrastic Future seems to be aspectually neutral. Several peripheral forms are also in use. To emphasize the processual nature of a future event the Future Imperfective form can be used. Prospective future (which is, in fact, a combination of the 'APUD' spatial form of participle and copula) is exceedingly peripheral and is seemingly restricted to non-volitional verbs like 'to fall' or 'to rain'.

Regarding the diachrony of the forms discussed, it is known that futures in Andic are not stable (Alekseev 1988), and two main channels of grammaticalization (among several minor ones) can be identified: grammaticalization of the infinitive as a carrier form and "old present drift".

In Anchiq, it is the infinitive-based forms that dominate the system: Synthetic Future is apparently a former attributivized infinitive form (as Alekseev 1988 suggests). The inclination of Periphrastic Future (also Infinitive-based) towards prospective uses assumes that it is an "Old Prospective" later developed into a general future. The connection between infinitives and futures is attested also in many other Andic languages (and in the wider Nakh-Daghestanian context — in Lezgic languages cf. (Maisak & Merdanova 2003: 82)). However, this diachronic connection is rarely discussed and poorly understood, as Kuteva et al. (2019: 87) admit. It can be hypothesized, although, that it is the infinitive form that makes the main contribution to the future semantics, considering the fact that grammaticalized infinitives tend to have 'irrealis-prospective' and 'irrealis-potential' meanings (Haspelmath 1989).

Another important grammaticalization path that is attested among Andic — "old present drift" (Haspelmath 1998) — in Anchiq is not too advanced, as the Habitual form can refer to the future mainly in hypothetical protases, while in Tindi or Chamalal the cognate form *-ida* is grammaticalized into the main future gram.

	Verb form	**Meaning**
Synthetic Future	*gah-aː-sːe* do-INF-FUT	'will possibly do'
Periphrastic Future	*gah-aː gira* do-INF COP	'will certainly do' or 'will be doing'
Prospective Future	*b-ekːʷ-a-ɬi-q gira* N-fall-PTCP-OBL-APUD COP	'is going to fall'

Future Imperfective	*gah-ir b-ik'ʷ-a: gira* do-IPFV N-be-INF COP	'will be doing'
Habitual	*gah-ir-a* do-IPFV-HAB	'if does […]'

References

Alekseev, M. E. 1988. Sravnitel'no-istoričeskaja morfologija avaro-andijskix jazykov. [The historical-comparative morphology of Avar-Andic languages] M.: Nauka.

Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Haspelmath, Martin. 1989. From purposive to infinitive — a universal path of grammaticalization. Folia Linguistica Historica 10(1–2).

Haspelmath, Martin. 1998. The semantic development of old presents: new futures and subjunctives without grammaticalization. *Diachronica* 15(1). 29–62.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World Lexicon of Grammaticalization*. 2nd edn. Cambridge University Press.

Maisak, T. A. & S. R. Merdanova. 2003. Buduščee vremja v agul'skom jazyke v tipologičeskoj perspektive [The future in Agul in typological perspective]. Voprosy jazykoznanija [Topics in the Study of Language] (6). 76–107.

Asymmetric coding of future tense in the world's languages

Martin Haspelmath

(Max Planck Institute for Evolutionary Anthropology, Leipzig)

keywords: communicative efficiency, future tense, periphrastic encoding, grammaticalization

This contribution presents a small worldwide study of future-tense coding (in 50 languages worldwide) compared with present-tense and past-tense coding. We can see that future tenses are generally marked by longer forms, as expected on the basis of general regularities of efficient coding (future tense is generally less frequent; see Haspelmath 2021). By contrast past-tense forms are not substantially shorter than present-tense forms, even though they were also argued to be "marked" in contrast to the present (Greenberg 1966). Whether past or present is more frequent is hard to determine (because most larger texts are narratives or otherwise atypical of ordinary speech), but that the future is much less frequent than both is beyond doubt. The focus of my empirical study will be on the typological questions, and I will devote substantial attention to the question of how to identify future-tense, present-tense, and past-tense forms in a wide range of different languages. The 50 languages are from 50 different language families, from all continents.

In addition to these simple results and explanations, I will ask why it is that future-tense forms are often periphrastic (Bybee et al. 1994, Dahl & Velupillai 2005), much more so than past-tense forms (let alone present-tense forms). The most commonly heard explanation is that shorter forms are the result of a longer history and greater attrition in grammaticalization, but I will offer an alternative perspective here: I will suggest (or argue) that shortness of coding is not primarily the result of a long history, but of adaptation to the level of expectedness. In addition, because of a general restriction of affixes to short forms, longer forms are less often affixes, and more commonly auxiliaries. The proposed causality is thus rather different from the traditional proposals: Length of coding is the cause of periphrastic marking, not the result of incomplete attrition due to a short history. This is admittedly difficult to prove, but I will cite evidence from generalizations in lexification typology (e.g. Kemp et al. 2018) to support the proposed explanation. I hypothesize that in general, less commonly expressed meanings are less likely to be lexified as simplexes.

References

Bybee, Joan L. & Perkins, Revere & Pagliuca, William (1994), *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: The University of Chicago Press.

Dahl, Östen & Velupillai, Viveka (2005), The future tense. In Haspelmath, Martin & Dryer, Matthew S & Gil, David & Comrie, Bernard (eds.), *The world atlas of language structures*, 270-. Oxford: Oxford University Press. (https://wals.info/chapter/67)

Greenberg, Joseph H. (1966), *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.

Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability. *Journal of Linguistics* 57(3). 605–633. (doi:10.1017/S0022226720000535)

Kemp, Charles & Xu, Yang & Regier, Terry. 2018. Semantic typology and efficient communication. *Annual Review of Linguistics* 4(1). 109–128. (doi:10.1146/annurev-linguistics-011817-045406)

Recurrent properties of futures from hypoanalysis
in Indo-European languages

Martin Kümmel[1], Jan-Niklas Linnemeier[2], Daniel Kölligan[2] & Björn Wiemer[3]
([1]University of Jena, [2]University of Würzburg & [3]University of Mainz)

Keywords: hypoanalysis, future tense, present tense, subjunctives, Indo-European

Indo-European (IE) languages are rich in futures arisen via hypoanalysis. Following Croft (2000: 126f.), hypoanalysis consists in the reanalysis of "a contextual semantic/functional property as an inherent property of the syntactic unit"; a meaning implied by (grammatical) contexts is "attributed to the syntactic unit, and so the syntactic unit in question gains a new meaning or function", i.e. invited implicatures conventionalize into new defaults. Among Croft's examples are subjunctives and futures reinterpreted from earlier present indicatives. Similar cases were discussed in Bybee et al. (1994: 232-236), Haspelmath (1998), Reinöhl/Himmelmann (2017: 406f.), also in IE linguistics (McCone 1991, Rix 1998, Hackstein 2004, Hill 2004, de Melo 2007). However, only Fries et al. (2023) and Wiemer et al. (2024) acknowledge hypoanalysis as a distinct type of language change which cannot be subsumed, e.g., under regrammation (Andersen 2006) or 'secondary grammaticalization' (critically revised in Breban 2014), since no 'primary grammaticalization' can even be reconstructed.

The talk presents an inventorization of "hypoanalysis futures" (HA-futures) across all attested IE languages, guided mainly by Wiemer et al. (2024) and taking advantage of the rich subbranching, enormous time depth (~3500 years) and spread across Eurasia of this language family. On this basis, it will be shown

(i)   whether, and to what extent, HA-futures differ from futures arisen via grammaticalization (centering around lexical source etyma) not only in the involved steps of change, but also in their functional range (e.g., for modal meanings) and/or w.r.t. issues raised in (iv-v) below.

(ii)   how properties associated to hypoanalysis and to grammaticalization, respectively, may interact to form a future tense (e.g., agglutination of copula forms to *nomina agentis* in Vedic, to participles in Nuristani), or how hypoanalysis and grammaticalization lead to parallel futures, e.g., yielding allomorphy (compare Latin *ē-* vs *b(i)-futures*).

(iii)   how HA-futures distribute among IE languages according to (a) periods, (b) areas, (c) subbranches, (d) "input" categories (present, subjunctive, etc.). These dimensions will be correlated with each other. On a preliminary account, HA-futures are practically exclusive until the time of the Great Migrations, but they are also biased to areas (at the western and eastern peripheries) and subbranches in which they occur repeatedly period after period (Celtic, Iranian, Armenian). It seems difficult to explain these biases by language contact, while inherited features might play a more important role.

Jointly, clarification is provided as for

(iv)   the varying diachronic stability of HA-futures. In particular, the HA-futures in East Baltic and in North Slavic are extraordinarily stable, but probably for different reasons which have to do with their input categories.

(v)   whether HA-futures simply vanish, or develop into something else (e.g., subjunctive, imperative). Alternatively, single forms may undergo exaptation (e.g., Lat. *eris*.FUT.2SG > Span. *eres*.PRS.IND.2SG of 'be').

(vi)   how hypoanalysis present > future can be distinguished from *praesens pro futuro*. Relevant criteria also allow to identify tipping points toward HA-futures which have not been overcome, e.g., in Germanic languages.

The conclusions promise insights for a better understanding of mechanisms of change and of the (in)dependence of genealogical and areal groupings.

References

Andersen, Henning (2006), Grammation, regrammation, and degrammation, *Diachronica* 23, 231–258.

Breban, Tine (2014), What is secondary grammaticalization? Trying to see the wood for the trees in a confusion of interpretations, in F. von Mengden, and H. Simon (eds), (2014), *What is it then, this Grammaticalization? (= Special issue Folia Linguistica 48[2])*, 469–502.

Bybee, Joan L., Revere D. Perkins, and William Pagliuca (1994), *The Evolution of Grammar. Tense, aspect, and modality in the languages of the world*, Chicago: University of Chicago Press.

Croft, William (2000), *Explaining Language Change. An Evolutionary Approach*, London: Longman.

de Melo, Wolfgang D.C. (2007), The sigmatic future and the genetic affiliation of Venetic: Latin *faxō* 'I shall make' and Venetic *vha.g.s.to* 'he made', *Transactions of the Philological Society* 105(1), 1–21.

Fries, Simon, Jakob Halfmann, Eugen Hill & Denise Hübner (2023), From agent noun to future tense. The functional diachrony of the *l*-future in the Nuristani languages and its typological background, *Sprachtypologie und Universalienforschung* 76(1), 53–85.

Hackstein, Olav (2004), Zur Entwicklung alter Präsentien und Perfekta zu Konjunktiv- und Futur-Formen in einigen altindogermanischen Sprachen, T. Poschenrieder (ed), (2004), *Die Indogermanistik und ihre Anrainer*, Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 81–113.

Haspelmath, Martin (1998), The semantic development of old presents: new futures and subjunctives without grammaticalization, *Diachronica* 15(1), 29–62.

Hill, Eugen (2004), Die sigmatischen Modusbildungen der indogermanischen Sprachen I. Das baltische Futur und seine Verwandten, *International Journal of Diachronic Linguistics and Linguistic Reconstruction* 1, 69–172.

McCone, Kim (1991), *The Indo-European Origins of the Old Irish Nasal Presents, Subjunctives and Futures*, Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Reinöhl, Uta & Nikolaus P. Himmelmann (2017), Renewal: A figure of speech or a process sui generis?, *Language* 93(2), 381–413.

Rix, Helmut (1998), Bemerkungen zu den lateinischen Verbformen des Typs *faxo, faxim*, in J. H. Jasanoff, H. Craig Melchert, and L. Oliver (eds), (1998), *Mír Curad. Studies in honor of Calvert Watkins*, Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 619–634.

Wiemer, Björn, Eugen Hill, Daniel Kölligan, Jan-Niklas Linnemeier (2024), *Between the birth and death of future tenses (Related languages as a natural lab for research into grammatical change)*, München: Lincom Europa.

jan-niklas.linnemeier@uni-wuerzburg.de
wiemerb@uni-mainz.de
daniel.koelligan@uni-wuerzburg.de

Future, Future-in-the-Past, and Conditionals: Evidence from Ulcha (Tungusic)

Natalia Stoynova
(University of Hamburg)

The paper deals with the future tense marker *-ilə* in Ulcha (Tungusic; Khabarovsk Krai, Russia), (1).

(1)   Aaa,   əsi=lə   xor-**ila**-m=tani,   wən-di-n
          INTRJ   now=EMPH   save-**FUT**-1SG=COORD say-PRS-3SG
          'Aaa, now, I'll be saved, he says.' (Sunik 1985: text-3)

This marker is not typical for Tungusic languages and is only found in Ulcha and its close sister Uilta. Along with its uses as a future tense marker, as in (1), *-ilə* has another one in combination with the past tense marker *-xən*: *osi-xa-n(i)* [become-PST-3SG] 'he became' ~ *os-ila-xa-n(i)* [become-FUT-PST-3SG] 'he would become'. This use is restricted for the apodosis of counterfactual conditional constructions (2) and some other irreal contexts.

(2)   esli   do_six_por   bi-či   bimčən,
          if.R   still.R   be-PST   if.IRR
          <...>   sama   os-**ila**-xa-ni   min   am-bi
                  shaman   become-**FUT**-PST-3SG my   father-1SG
          'If he were still alive, my father would become a shaman.' (Field records)

At first glance, the Ulcha FUT-PST forms are similar to well-known future-in-the-past forms attested in languages of Europe, cf. *he will ~ would read* in English; *il lira ~ lirait* in French. The expansion of conditional/irreal forms in the future-in-the-past contexts was described as well (see, e.g., Daidone & Zahler 2016 on Spanish; Rottet 2010 on Louisiana French). However, the Ulcha PST-FUT combination unexpectedly lacks the core future-in-the-past interpretation, i.e. is not attested in such contexts, as 'He hoped he would see her soon' or 'He was happy. He would see her soon' and does not signal that the event follows a time point in the past.

In the talk, I will discuss whether *-ilə* can be treated as the proper tense marker, try to put the Ulcha data into a more general picture of possible FUT-PST combinations, and propose a probable semantic link between the FUT (1) and FUT-PST (2) uses of *-ilə*.

In the future semantic domain, the dedicated *-ilə* competes with the optative *-ŋə* and the present tense form. The present tense is the most neutral and frequent, while the use of *-ilə* is restricted. It belongs to prediction-based rather than intention-based future grams (in terms of Dahl 2000). Moreover, it seems to specifically signal that the future event is explicitly or implicitly conditioned by another event. For instance, in (1), the assumption that the hero can be saved follows from the just reported fact that he ran into someone's house on his way. The apodosis of real condition constructions (3) appears to be an especially frequent context for *-ilə*.

(3)   bu-di   očin=də,   ti   naː-du   bu-**dilə**-m=əmdə
          die-PRS   if=ADD   so   earth-ESS/DAT die-**FUT**-1SG=QUOT

‘If I die, at least I'll die on earth, he says.’ (Field records)

Thus, my claim is that the conditional rather than temporal component links the FUT and FUT-PST uses of -ilə.

The study is mostly corpus-based: the data come from my field collection, the archives of Kalinina et al. and Sem (ca. 100,000 tokens in total), and texts published in Petrova (1936); Avrorin (1981); Sunik (1985); Kazama (2002, 2006, 2008, and 2010). Elicited data are included as well.

References

Avrorin, Valentin A. (1981), Uljčskie teksty. Materialy dlja grammatičeskix i etnografičeskix issledovanij [Ulcha texts. Materials for studies in grammar and ethnography], *Morfologija imeni v sibirskix jazykax* [Morphology of nominals in languages of Siberia], Novosibirsk.

Dahl, Östen (2000), The grammar of future time reference in European languages, Dahl Ö. (ed.), *Tense and Aspect in the Languages of Europe*, Berlin / New York: Mouton de Gruyter.

Daidone, Danielle and Sara Zahler (2016), The future is in the past: A diachronic analysis of variable future-in-the-past expression in Spanish, A. Cuza, L. Czerwionka, & D. Olson (eds.), *Inquiries in Hispanic linguistics: From theory to empirical evidence*. Philadelphia, PA: John Benjamins, 317-334. http://dx.doi.org/10.1075/ihll.12.17dai

Kalinina et al. archive – *Ulcha texts collected by Elena Kalinina, Valentin Gusev, Svetlana Toldova, and Nina Sumbatova*. https://www.elararchive.org/uncategorized/SO_2ca69265-52b5-48df-a838-263f8bb6aaba/

Kazama, Shinjiro (2002), *Ulcha oral literature* 2, Publications on Tungus Languages and Cultures, 20, Osaka: ELPR.

Kazama, Shinjiro (2006), *Ulcha oral literature 3*, Publications on Tungus Languages and Cultures, 30, Tokyo: Graduate School of Letters, Hokkaido University.

Kazama, Shinjiro (2008), *Ulcha oral literature 4*, Publications on Tungus Languages and Cultures, 43, Tokyo: Tokyo University of Foreign Studies.

Kazama, Shinjiro (2010), *Ulcha oral literature 5*, Publications on Tungus Languages and Cultures, 49, Tokyo: Tokyo University of Foreign Studies.

Petrova, Taisia I. (1936), *Uljchskij dialekt nanajskogo jazyka* [The Ulcha dialect of Nanai], Moscow / Leningrad: Uchpedgiz.

Rottet, Kevin (2010), The future in the past in Louisiana French, *Congrès Mondial de Linguistique Française*, Paris, 2010, Institut de Linguistique Française Sociolinguistique et écologie des langues. DOI 10.1051/cmlf/2010164.

Sem archive – *Unpublished archive of Ulcha audio-recordings collected by Lidia Sem*, Institute for Linguistic Studies, Saint-Petersburg.

Sunik, Orest P. (1985), *Ulchskij jazyk: issledovanija i materialy* [Ulcha: Studies and materials], Leningrad: Nauka.

Aspect and irrealis in Mian and other Mountain Ok languages

Sebastian Fedden

LMU München, Université Sorbonne Nouvelle, Surrey Morphology Group

Typologically, the aspectual contrast between perfective and imperfective is often restricted to past tense in languages that make this distinction (e.g., perfect vs imperfect in Latin). In Mian, and other Mountain Ok languages (all Papuan languages of the Trans New Guinea family spoken in Papua New Guinea), this aspectual contrast can be found not only in the past but also with forms referring to the future.

Aspect is at the heart of the Mountain Ok verbal system. Substantial numbers of verb stems make a perfective-imperfective distinction (often called punctual-continuative in the traditional descriptions of these languages). In Mian, about two thirds of verbs follow this pattern, e.g., *fa/faka* 'make a fire' (with a suffix *-ka* in the imperfective) and *baa/o* 'say' (suppletive). Marking patterns are highly irregular with only a handful of verbs following any given pattern. For more than half of these the perfective or the imperfective stem is missing, e.g., *mâa'/—* 'stand up' or *—/dlan* 'last', with semantic factors often accounting for the gap. The rest of the verbal vocabulary is biaspectual, so the same stem can have either aspectual meaning, e.g., *fu* 'cook'.

In Mian, future time reference is one of the functions of the modal category 'irrealis', which is also used to express predictions, desires and past intensions which have been thwarted by adverse circumstances. As in the past, verb forms in the irrealis can be either perfective or imperfective. Witness the distinction in examples (1) and (2).

(1)   ase   fa-n-amab-i=be
       fire   make.fire.PFV-PFV-IRREALIS-1SG.SBJ=DECL
       'I'll make a fire.'

(2)   ase   faka-mab-i=be
       fire   make.fire.IPFV-IRREALIS-1SG.SBJ=DECL
       'I'll be making a fire.'

The perfective form in (1) describes a future situation as a whole, without making its internal temporal structure explicit. The corresponding imperfective form in (2) expresses that the situation—still in the future—is temporally structured, i.e., the focus of attention is on the duration of the situation, the fact that it consists of multiple individual steps involved in 'making a fire', or the possibility of something else happening while the situation of making a fire holds. The discussion focusses on Mian, but data from other Mountain Ok languages, such as Telefol (Healey 1965) and Tifal (Boush 1975), as well as further comparative data from the Lowland Ok language Muyu (Zahrer 2023) where only residues of this aspect system survive are also considered.

The paper contributes to our understanding of the relation between aspect and other morphosemantic categories, such as mood and tense, and presents new data from a family of Papuan languages in a comparative perspective.

References

Boush, Alfred. 1975. Tifal grammar essentials. Ms., Ukarumpa, Summer Institute of Linguistics.

Healey, Phyllis. 1965. Telefol verb phrases. In *Linguistic Circle of Canberra Publications*. (A-5.), 27–53. Canberra: Pacific Linguistics.

Zahrer, Alexander. 2023. A grammar of Muyu. Westfälische Wilhelms-Universität Münster PhD dissertation.

Futures, Modality, and Negation in Southern Bantu Languages

Thera Marie Crane
University of Helsinki
thera.crane@helsinki.fi

The 'potential' (possibility) marker *nga* and its cognates occur widely in Southern Bantu (SB) languages, both prefixally and as an auxiliary. In a cross-Bantu study, Nurse describes *nga* as "concessive, conditional, potential, irrealis, may, if/when" and points to its positional variability as evidence of "ongoing and recent grammaticalization" (Nurse 2008: 251–252). In SB, the potential occurs, with varying frequencies, in all modal flavours and with functions typically described as 'postmodal' (van der Auwera & Plungian 1998), such as concessive, although the precise semantic pathways of *nga* are not so straightforwardly traced (Nurse & Devos 2019: 227–228). Despite semantic links between the notion of 'potential' and futurity, *nga* is not documented as a future marker in SB. However, potential markers interact significantly with futurity in SB, especially under negation, and in more and less grammaticalized ways.

After a general introduction to future marking in SB and some of the temporal, spatial, and modal contrasts involved, this paper will give an overview of potential–future interactions and their relationship to the 'core' future systems – many of which also encode mutiple futures that are themselves distinguished partly based on modal semantics (Crane & Mabena 2019). Data come from elicitation and analysis of spoken texts, compared with grammatical descriptions where available.

A less-grammaticalized case of *nga* marking futurity is illustrated in (1), from Tsonga. Across SB, the potential is commonly prefixed to other modal expressions, with both expressions contributing to the contextually sensitive interpretation of the utterance. One common interpretation is future possibility.

(1) *Ndzi-**nga**-kota   ku-y-a   e-xitolo.*
     SP$_{1SG}$-**POT**-ABLE   INF-go-FV   LOC-store
     'I can go to the store [later].' (lit. 'I can be able to go to the store') (Tsonga)

The potential plays a more grammaticalized role in future marking under negation in many SB languages. To give just three examples:

- The negated potential auxiliary *ngeke* 'cannot' in Zulu, apparently formed with the negated potential and experiential marker *ke*, is also used as a negative future marker meaning 'will not' or 'will never', depending on the context. In Zulu, *ngeke* alternates with negated potential prefix *nge-*, but contrasts with it both in the former's future meanings and in its scope of negation: *nge-* has variable scope ('not possible' OR 'possibly not', e.g. ***u**ngedlali* 'you may not play' / 'you don't have to play'), while *ngeke* can only convey external negation ('not possible', e.g. ***u**ngeke wadlala* 'you may not play'), a fact likely interrelated with its negative futur(at)e meanings.

- Southern Ndebele uses as future / modal / negative / counterfactual markers both *ngekhe* (see Zulu *ngeke*) and *ngeze*, with *ze* likely related to *za* 'come from', also a source of two of Southern Ndebele's three futures.

- In Tsonga, potential *nga* may be involved in three negated future forms (du Plessis et al. 1995: 240; Ouwehand 1965: 52), at least two of which are also used to indicate impossibility.

The semantic overlap of (im)possibility and future (non-)instantiation is apparent, and many such connections appear to be robustly integrated into the grammatical systems of SB languages.

Futures, Modality, and Negation in Southern Bantu Languages

Thera Marie Crane

University of Helsinki

thera.crane@helsinki.fi

References

Baumbach, E.J.M. 1987. *Analytical Tsonga grammar*. Pretoria: University of South Africa.

Crane, Thera Marie & Msuswa Petrus Mabena 2019. Time, Space, Modality, and (Inter)subjectivity: Futures in isiNdebele and other Nguni Languages. *South African Journal of African Languages* 39(3): 291–304.

Du Plessis, J.A., N.E. Nxumalo & Marianna Visser. 1995. *Tsonga syntax*. (Stellenbosch communications in African languages, 2.) Stellenbosch: University of Stellenbosch.

Nurse, Derek. 2008. *Tense and aspect in Bantu*. Oxford: Oxford University Press.

Nurse, Derek & Maud Devos. Aspect, tense, and mood. In Mark Van de Velde, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 2nd ed., 204–236. London: Routledge.

van der Auwera, Johan & Vladimir Plungian. 1998. Modality's semantic map. *Linguistic Typology* 2: 79–124.

Burushaski future in an areal and cross-linguistic perspective

Vladimir Plungian
Institute of Linguistics (RAS), Moscow

Keywords: Burushaski, Hindukush linguistic area, verbal systems, future, aspect

Burushaski, a linguistic isolate of Hindukush, has preserved a peculiar verbal system with a curious mix of areal and specific features. Our data represent its westernmost Yasin variety and are based both on the existing descriptions (as Zarubin 1927, Berger 1974, Tiffou & Pesot 1989) and on consultations with a native speaker Ahmad Ali Shah in September 2024 and May 2025. Yasin Burushaski has a dedicated future obtained from the imperfective stem (morphologically derived from the basic, or perfective stem). Below (1) represents a small fragment of the verbal inflectional paradigm illustrating these correlations.

(1)   *girmín-um b-u* write-CNV be-PRS.3SG.F 'she has written' (periphrastic perfect with perfective converb and 'esse'-auxiliary)

girmín-č-um bu write-IPF-CNV be-PRS.3SG.F 'she writes / is writing' (periphrastic present with imperfective converb and 'esse'-auxiliary)

girmín-č-umu write-IPF-3SG.F 'she will write' (synthetic future morphologically built upon imperfective stem).

Synchronically, however, the meaning of the future form is rather aspect-neutral, various perfective and imperfective uses being equally possible: cf. a clearly perfective construal in (2) (adapted from Berger 1974: 72; the glosses and the English translation are ours).

(2) <…> gusé   har   áne   qasaí.étas-en-a   **ú-č-an**,   sén-i:
     this   ox   where   butcher-IND:H-DAT   give-FUT-1PL   say-AOR:3SG
     i-khár-e   **ṣí-ç-imi**,   mí-ɣa   paisá   **mi-čí-č-i**.
     3SG-REFL-ERG   slaughter-FUT-3SG   1PL-DAT money 1PL-give.Y-FUT-3SG
     {unser Ochse ist jetzt krank geworden} 'Wir **werden** ihn irgendwo einem Metzger **geben**, der **wird** ihn selber **schlachten** und uns Geld **geben'** / {our ox has now become ill} '**We will give** this ox to a butcher somewhere**,** he said, **he will slaughter** him himself and **give us** money'.

Diachronically, these future forms seem (as suggested already in Tiffou & Pesot 1989: 34) to have derived from an "old present" (Haspelmath 1998) ousted by a more recent periphrastic construction [imperfective converb + present auxiliary]; the latter has developed, in its turn, a large inventory of meanings including present habitual and peripheral future uses (they will be elaborated on in the presentation).

Thus, Burushaski represents a relatively rare system with a "marked present" and "unmarked future" (though diachronically imperfective-based); a similar – though not fully identical – pattern is attested in other Hindukush languages as well (most importantly, in neighboring Khowar).

REFERENCES
Berger, Hermann (1974), *Das Yasin-Burushaski (Werchikwar): Grammatik, Texte, Wörterbuch* (Neuindische Studien 3), Wiesbaden: Harrassowitz.
Haspelmath, Martin (1998), The semantic development of old presents: New futures and subjunctives without grammaticalization, *Diachronica* 15(1), 29–62.

Tiffou, Etienne & Jurgen Pesot (1989), *Contes du Yasin: introduction au bourochaski du Yasin avec grammaire et dictionnaire analytique* (SELAF 303), Paris: Peeters.

Zarubin, Ivan (1927), Veršikskoe narečie kandžutskogo jazyka [Wershik variety of Kanjut language]. *Zapiski kollegii vostokovedov*, t. II, vyp. 2, 275–364.

Multiple asymmetries in the tense system: A case study mainly in Chadic

Zygmunt Frajzyngier
University of Colorado at Boulder

Keywords: asymmetry of functions, future tense, past tense, granularity of tenses, pragmatic dependency

Terms: 'Tense' denotes the coding of time in the grammatical system. The study demonstrates that tense systems in Chadic languages (as listed in references) display multiple types of semantic asymmetries affecting future and past tenses. The study also explains one of these asymmetries.

Asymmetry 1: The number of past tenses (other than hodiernal) is smaller than the number of future tenses:

Number of past tenses	Number of future tenses	Language
Five (borrowed, hodiernal)	Two	Mupun,
Two	Three	Mina
One	Two	Hdi
None	One	Pero, Pévé
None	Two	Hausa, Gidar

Asymmetry 2: In hodiernal systems, the number of distinctions in past tenses is larger than the number of distinctions in future tenses (Mupun, Moba, (Gur)). This asymmetry is addressed in a different study.

Asymmetry 3: Different temporary characteristics are coded by the past and future tenses. Hdi has only a 'relative past', coding the event that occurred at some specific time in the past (ex.1-2). Future tenses in Hdi have no such distinction. In Lele, future tenses make a distinction between unspecified time versus specific time. There is no such distinction in the past tense. The asymmetrical systems are in contrast with symmetrical systems within the same family. Giziga has one past tense and one future tense.

Asymmetry 4: Tenses and the coding of pragmatic dependency. In some languages, distinct tense forms, while coding the same temporal value, also code the distinction between propositions that can be interpreted on their own (pragmatically independent) and propositions that **must** be interpreted in connection with another proposition, in connection with some event, or in connection with some specific time. In Hdi, the future tenses make a distinction between pragmatically dependent and pragmatically independent propositions. The asymmetric system of Hdi is in contrast with the symmetric system in Mina, where both past and future tenses distinguish between pragmatically independent and pragmatically dependent propositions.

Focus of the presentation is the explanation of asymmetry 1: All Chadic languages code completive or perfective aspects which allow to **infer** (but do not code!) the time of the event as past. Hence the information about the past time does not require a distinct marker. Not a single aspect in Chadic languages, however, allows to infer the time of the event as future. Hence, the future time of the event must be overtly marked. Future tense markers in Chadic are derived from the verbs 'go', 'come', from nouns denoting 'time' (Mupun), from the hypothetical modality markers (Hausa) and, unexpectedly, from the locative preposition (Mina).

Implication: The study demonstrates the existence of a **systemic** motivation in the emergence of a grammatical category. If a function cannot be inferred from other markers it is overtly marked (a motivation not envisaged in (Kuteva et al. (2019).

Examples:
Specific past tense:

(1) ***sí***   *tà*   *dv-áy-xә̀n*   *tá*   *łí-á-mú*

     R.PAST   IPFV   want-PO-3SG   OBJ   leave-GEN-1PL.INCL

     '[**at that time**] they wanted us to leave'

Compare past time reference, but not dependent on specific time of the event:

(2) *mbàɗ*   *ká*   *pákáwá ghúvì*   *kà*   *xvá-tá*   *xvá*

     then   COMP   hyena   SEQ   farm-REF   farm

     'Hyena had already farmed.'

References
Binnick, Robert I. 2011. *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press.
Botne, Robert. Remoteness distinction. In Binnick 2011, 536-562.
Comrie, Bernard. 1985. *Tense*. Cambridge: Cambridge University Press.
Frajzyngier, Zygmunt. 1989. *A Grammar of Pero*. Berlin: Reimer.
Frajzyngier, Zygmunt. 1993. *A Grammar of Mupun*. Berlin: Reimer.
Frajzyngier, Zygmunt. 2001. *A Grammar of Lele*. Stanford Monographs in African Linguistics. Stanford: CSLI.
Frajzyngier, Zygmunt. 2008. *A Grammar of Gidar*. Frankfurt: Peter Lang.
Frajzyngier, Zygmunt. 2012. *A Grammar of Wandala*. Berlin and New York: Mouton de Gruyter.
Frajzyngier, Zygmunt, Eric Johnston, with Adrian Edwards. 2005. *A Grammar of Mina*. Berlin/New York: Mouton de Gruyter.
Frajzyngier Zygmunt, with Erin Shay. 2002. *A Grammar of Hdi*. Berlin: Mouton de Gruyter.
Frajzyngier Zygmunt. Fieldnotes on Moba
Hewson, John. 2011. Tense. In Binnick 2011, 507-535.
Kuteva, Tania, Bernd Heine, Bo Hong, Haipin Long, Heiko Narrog, and Seongha Rhee. 2019. *World lexicon of grammaticalization*. Cambridhe: Cambridge University Press.
Newman, Paul. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. New Haven and London: Yale University Press.
Shay, Erin. 2021. *A Grammar of Giziga.* Leiden and Boston: Brill.
Shay, Erin, with Lazare Wambadang. 2019. *A Grammar of Pévé.* Leiden and Boston: Brill.

WS16 Pragmatic Language Development in Young Children

Tove Nilsson Gerholm, Maria Rosenberg & Linda Sandström

Normative pragmatics in early language development

Antonio Scarafone
LMU Munich

I introduce a novel view on child communication and early word learning, which goes under the heading of normative pragmatics. I illustrate the advantages of adopting normative pragmatics by considering existing empirical studies and suggesting new directions for research, but without presenting new data.

According to normative pragmatics, speech acts serve primarily to negotiate commitments (Geurts 2019, Jary 2021), and pragmatic reasoning consists in reasoning according to how speech acts, so conceived, update the common ground. From this point of view, prelinguistic communication can be studied in terms of how children become increasingly more prepared to share commitments (Scarafone & Michael 2022).

Pointing can be analysed as a gesture with which the producer aims at sharing a commitment to attend to something. Commitments are expectation management devices, and indeed infants sometimes point to express their expectations (e.g., pointing to the bathroom door), managing them according to the response they receive from the adult (e.g.: 'Yes, we'll go to the bathroom in a moment.'). This response also provides infants with a standard to refine their own accompanying vocalizations into words, which helps with expressing expectations more explicitly on subsequent occasions. This view, I argue, makes sense of findings about the relationships between infant pointing and vocalizations, the adult's response, and infant subsequent vocabulary (Donnellan et al. 2020).

Young children count as commitment sharers because, by around when they are 18 months of age, they start meeting two, key preconditions (Scarafone 2025): (i) knowing how to signal acceptance and rejection of candidate commitments (Kettner & Carpendale 2013); (ii) knowing how to behave in accordance with those commitments, including knowing how to intervene normatively in the face of violations (Schmidt et al. 2019). These preconditions are open-ended, which reflects the fact that cognitive skills needed for pragmatic reasoning are heterogeneous and acquired piece-meal. Though central, these phenomena are understudied. Therefore, normative pragmatics has the potential to disclose important, though rarely beaten, tracks for empirical research.

References
Donnellan, E.; Bannard, C.; McGillion, M. L.; Slocombe, K. E.; Matthews, D. (2020), Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production, *Developmental Science*, 23 (1).

Geurts, B. (2019), Communication as commitment sharing: Speech acts, implicatures, common ground, *Theoretical Linguistics*, 45 (1-2): 1-30.

Jary, M. (2021), *Actos de habla*, in A. Ahern; J. Amenós Pons; V. Escandell-Vidal (eds.), *Pragmática*, Madrid: Akal.

Kettner, V. A., Carpendale, J. I. M. (2013), Developing gestures for no and yes: head shaking and nodding in infancy, *Gesture*, 13 (2): 193–209.

Scarafone, A. & Michael, J. (2022), Getting Ready to Share Commitments, *Philosophical Topics*, 50 (1), 135-160.

Scarafone, A. (2025), Where Infants Dare: Teasing and Headshaking, *Topoi*.

Schmidt, M. F. H.; Rakoczy, H.; Tomasello, M. (2019), Eighteen-month-old infants correct non-conforming actions by others, *Infancy*, 24 (4): 613–635.

Assessing pragmatic and morphosyntactic abilities in Danish 3-to-5-year-olds' comprehension and use of object-first clauses

Ditte Boeg Thomsen
(University of Copenhagen)

In many languages, specific word orders are used for specific discourse-pragmatic purposes, and for children to acquire constructions with these word orders requires pragmatic skills of integrating information from the broader context in sentence interpretation. Crosslinguistically, word orders putting the object before the subject are often used for discourse-pragmatic purposes of signaling topicality and contrast, and in comprehension experiments, children have been found to misinterpret object-first clauses up to surprisingly high ages (Chan et al. 2009, Dittmar et al. 2008, Hakuta 1982, McDonald 1989, Slobin & Bever 1982). Since the object-first clauses tested are known to be context-sensitive, children's errors are often explained as arising from discourse-pragmatic immaturity and a lack of ability to integrate contextual information with intrasentential morphosyntactic cues (e.g., Bates & MacWhinney 1989, Chan et al. 2009). This interpretation is problematic, however, because the comprehension tests present object-first clauses *in isolation* and are therefore not appropriate for assessing awareness of the contextual requirements of the clauses. On the contrary, misunderstanding of object-first clauses in pragmatically infelicitous isolation contexts could reflect sophisticated sensitivity to the discourse-pragmatic functions of object-first clauses.

Boeg Thomsen & Poulsen (2015) therefore compared Danish 5-to-6-year-olds' comprehension of case-marked object-first clauses in supportive vs. unsupportive contexts and in isolation, and they found significantly better comprehension of object-first clauses in supportive contexts with appropriate topicality and contrast structures. This result indicates that studies examining acquisition of context-sensitive constructions in isolation are likely to underestimate children's discourse-pragmatic abilities.

As this single-age experiment cannot tell us whether sensitivity to the contextual requirements of object-first clauses is late-developing, emerging around age 5, the current study investigated comprehension and production of object-first clauses across age. For comprehension, an act-out experiment with 24 Danish kindergarteners (3-, 4- and 5-year-olds), using the materials from Boeg Thomsen & Poulsen (2015), demonstrated that at least from age 3, children utilize their understanding of topicality and contrast to interpret case-marked object-first clauses successfully in felicitous contexts. The production results converge. In a 99-hour video corpus of peer conversations (32 children, age range: 1;9-6;3 years), object-first clauses were frequent (2076 tokens), and in all age groups, they were used in the discourse contexts expected for mature language users. Together, the production and comprehension results thus provide converging evidence that Danish children's sensitivity to the discourse-pragmatic functions of object-first clauses is not a developmental achievement at age 5, but an integral part of their constructional knowledge from early in development.

The study further suggests an influence of children's pragmatic abilities on their development of mature morphosyntactic strategies. Whereas Danish adults depend on the morphosyntactic cue

case for comprehension of object-first clauses, also in isolation contexts (Boeg Thomsen & Kristensen 2014), the 3-to-5-year-olds in this study only trusted case as a cue for comprehending object-first clauses if the context licensed an object-first interpretation. This suggests that Danish children's experience with and awareness of the discourse-pragmatic functions of object-first clauses support and stabilize their initially fragile awareness of the morphosyntactic cue case that they will come to rely on in adulthood.

References

Bates, Elisabeth and Brian MacWhinney (1989), Functionalism and the competition model, in B. MacWhinney, and E. Bates (eds), (1989), *The Crosslinguistic Study of Sentence Processing*, Cambridge: Cambridge University Press, 3-73.

Boeg Thomsen, Ditte and Line Burholt Kristensen (2014), Context needed: Semantic role assignment in Danish children and adults, *Acta Linguistica Hafniensia* 46(2), 159-198.

Boeg Thomsen, Ditte and Mads Poulsen (2015), Cue conflicts in context: Interplay between morphosyntax and discourse context in Danish preschoolers' semantic role assignment, *Journal of Child Language* 42(6), 1237-1266.

Chan, Angel, Elena Lieven and Michael Tomasello (2009), Children's understanding of the agent-patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English, *Cognitive Linguistics* 20(2), 267-300.

Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven and Michael Tomasello (2008), German children's comprehension of word order and case marking in causative sentences. *Child Development* 79(4), 1152-1167.

Hakuta, Kenji (1982), Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Developmental Psychology* 18(1), 62-76.

McDonald, Janet (1989), The acquisition of cue-category mappings, in B. MacWhinney, and E. Bates (eds), (1989), *The Crosslinguistic Study of Sentence Processing*, Cambridge: Cambridge University Press, 375-396.

Slobin, Dan I. and Thomas G. Bever (1982), Children use canonical sentence schemas: A crosslinguistic study of word order and inflections, *Cognition* 12, 229-265.

The emergence of the illocutionary complementizer *că* in child Romanian

Ioana Stoicescu
University of Bucharest

Keywords**:** illocutionary complementizers, acquisition, child Romanian

An illocutionary complementizer is a connector that introduces a main clause rather than a subordinate clause and plays a discourse-related role (Corr 2016). Its functionality is not related to subordination rather to linking utterances or speech acts in discourse. In a new line of research on the acquisition of the syntax/pragmatics interface, evidence has become available of the early development of knowledge related to a domain located higher than the functional domain, namely the Speech Act/utterance domain, which hosts constituents that are relevant for the advancement of conversation and signal various illocutionary forces (Roeper 2007, Bosch 2023).

This study investigates the emergence of a complementizer associated with the Speech Act domain in child Romanian, namely *că* "that". The research question is whether, in the acquisition of Romanian by typically-developing children, the illocutionary complementizer *că* (Corr 2016) emerges early, before/after the subordinating complementizer. Given previous findings for other Romance languages (Bosch 2023), it is predicted that the illocutionary complementizer emerges before the subordinating complementizer, signalling the early activation of the utterance domain.

The methodology was the following: all occurrences of *că* in two longitudinal child Romanian corpora were identified along with the contexts in which they were produced and coded for complementizer type (subordinating/illocutionary). The utterances preceded by illocutionary *că* pattern with root clauses, and are prosodically independent from previous clauses, indicating a new conversational move and the addition of new information to the conversational background. Ambiguous examples were excluded from the count, as well as formulaic speech, repetitions, or unintelligible utterances. The data came from two longitudinal corpora of child Romanian belonging to a girl, Bianca (1;5.12 - 2;11.22) (Avram 2001), and a boy, Iosif (1;10.23 - 3;1.13) (Stoicescu 2013).

The results of the analysis showed that the illocutionary complementizer *că* emerges early in child Romanian, before subordinating *că*. Both children started using *că* as an illocutionary complementizer, and were more productive in its use than in the use of subordinating *că*. Bianca produced illocutionary *că* at 1.33 MLUw (2;0), while subordinating *că* emerged during the two-word stage at 2.8 MLUw (2;5). Iosif produced the first illocutionary *că* at 1.7 MLUw (2;1) and started using subordinating *că* at 2.5 MLUw (2;2). The majority of *că* instances that the children produced were illocutionary - Iosif: 63%; Bianca: 77%, while the instances of subordinating *că* were less numerous - Iosif: 37%; Bianca: 23%. In child Romanian, the illocutionary complementizer *că* was a marker of informational enrichment, explanation, but also correction and rejection of adult comments.

The early emergence of illocutionary *că* shows that the utterance-layer is activated from the onset of acquisition, contra bottom-up approaches to syntactic development such as the Growing Trees Hypothesis (Friedmann et al. 2021). The results are compatible with the approach in Biberauer (2018) and

Bosch (2023), according to which some illocutionary or speaker-hearer elements indicate the edges of syntactic domains and, due to this role, are prioritized in acquisition.

References

Avram, Larisa (2001), Early omission of articles in child Romanian and the emergence of DP. *Revue Roumaine de Linguistique* XLVI(1-4), 105-123.

Biberauer, Theresa (2018), Peripheral significance: a phasal perspective on the grammaticalisation of speaker perspective. Talk presented at DiGS 20 (York), June.

Bosch, Núria (2023), Not all complementizers are late: a first look at the acquisition of illocutionary complementizers in Catalan and Spanish, *Isogloss. Open Journal of Romance Linguistics* 9(1), 1-39.

Corr, Alice (2016), Ibero-Romance and the syntax of the utterance, PhD dissertation, University of Cambridge.

Friedmann, Naama, Belletti, Adriana, Rizzi, Luigi (2021), Growing trees: The acquisition of the left periphery, *Glossa: A Journal of General Linguistics* 6(1), 131.

Roeper, Tom (2007), *The Prism of Grammar: How Child Language Illuminates Humanism*, Cambridge, MA: MIT Press.

Stoicescu, Ioana (2013), *The Acquisition of Tense and Aspect in Romanian*, București: Editura Universității din București.

Learning how to converse: a longitudinal study of the acquisition of backchannels

Although language is learned in conversation (Clark, 2018), little is known about the acquisition of forms dedicated to managing it. We present data from a longitudinal case study of a monolingual English-speaking child (Ellie from the Sekali corpus (Beaupoil-Hourdel, 2015)) to identify when children begin to manage conversation using backchannels, such as *hm, oh, nice*. Among adults, backchannels signal understanding or attitudes by the addressee without claiming a turn; instead, they often constitute encouragements for the first speaker to continue (Yngve, 1970). Importantly, backchanneling happens at predictable points in conversation, namely at turn-constructional units (TCU) (Sacks et al., 1974). Hence, to learn when and how to backchannel is not a trivial task: a child needs to learn to predict the boundaries of TCUs, and to have an awareness of turn-taking mechanisms. It is therefore not surprising that the previous literature has assumed their acquisition to be rather late, viz. age 4-6 (Schatz & Gelman, 1973; Bodur et al., 2023). Our aim is to identify the onset and growing complexity in backchannels compared to other forms of interactional language. This will allow us to understand when children start to manage conversation.

Our paper presents the developmental path of a range of different units of language that can serve as backchannels. In Ellie's data, interactional language is present in adjacency pairs from 0;10 (years; months) in the form of repetitions and non-verbal backchannels. Verbal backchannels (and more varied conversational strategies) are first observed from 1;09. Figure 1 (on page 2) shows an increasing use and diversification of backchannels used. At age 3;05 we note a strong increase in the use of back channels that suggests stabilization and a secure command of the ability to manage conversation.

Our findings are relevant for understanding the role of conversation in child language development because they signal an early availability of interactional language, including backchannels. The early ability to manage conversation well before the onset of complex sentences suggests that children have a metalinguistic awareness of conversation even before they have a fully developed Theory of Mind (De Villiers, 2007). Many of the backchannels emerging before 4;0 (e.g. *eh, you think, I know*) signal agreement or surprise and require at least some mental representation of what individual interlocutors know to. This observation is in line with two recent proposals in language development: i) that turn-taking is learned very early and possibly even independently of language (Hilbrink et al., 2015); and that interactional language is an important window into the early stages of language development (Heim & Wiltschko, *in print*). In our paper, we present further examples of Ellie's interactional language and spell out the implications of their developmental stages for theories of language development. If what we find here is representative of language development, conversation management can no longer be an afterthought in our theories of language development but must be front and center.
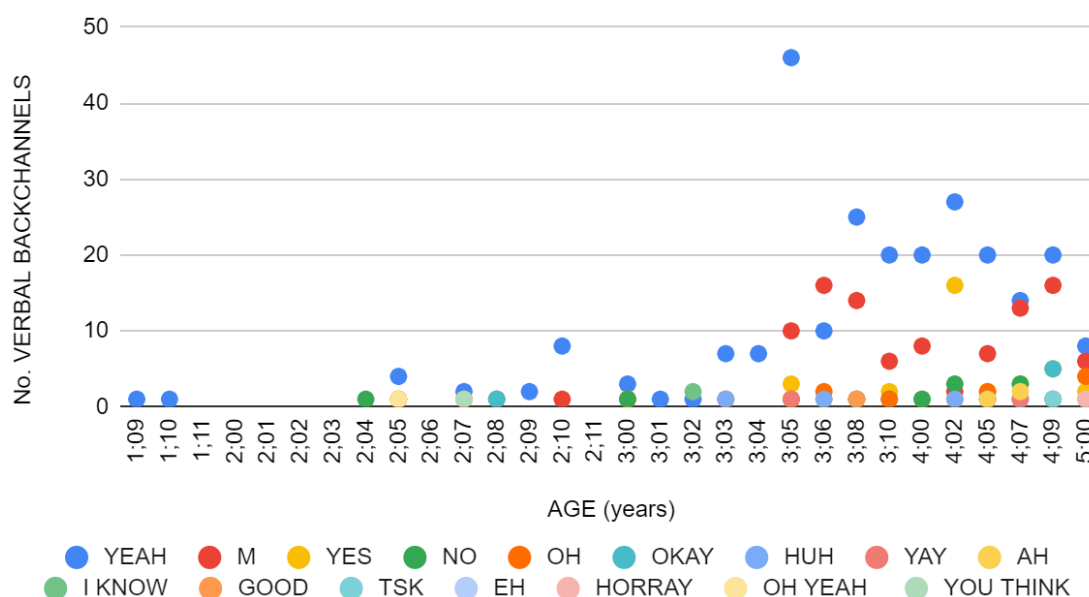
Figure 1: Different verbal back-channels by frequency per month recorded

References

Beaupoil-Hourdel, Pauline (2015). Multimodal acquisition and expression of negation. Analysis of a videotaped and longitudinal corpus of a French and an English mother-child dyad. (Ph.D. Dissertation). Sorbonne Nouvelle University, Paris.

Bodur, K., Nikolaus, M., Prévot, L., & Fourtassi, A. (2023). Using video calls to study children's conversational development: The case of backchannel signaling. *Frontiers In Computer Science*, 5. https://doi.org/10.3389/fcomp.2023.1088752

Clark, E. V. (2018). Conversation and language acquisition: A pragmatic approach. *Language Learning and Development*, *14*(3), 170-185.

De Villiers, J. (2007). The interface of language and theory of mind. *Lingua*, 117(11), 1858-1878.

Heim, J. & Wiltschko, M. (in print). Rethinking structural growth: Insights from the acquisition of interactional language. To appear in *Glossa*.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology*, 6, 1492.

Sacks, H., Schlegoff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking in conversation. *Language*, 50, 696-735.

Shatz, M., and Gelman, R. (1973). The development of communication skills: modifications in the speech of young children as a function of listener, in *Monographs of the Society for Research in Child Development*, 1–38.

Yngve, V. H. (1970). "On getting a word in edgewise," in *Chicago Linguistics Society, 6th Meeting*, Vol. 1970 (Chicago, IL), 567–578.

Pragmatic types of utterances and the acquisition of person category in Estonian

Reili Argus
(Tallinn University)

While the acquisition of person category has received attention from a semantic perspective (e.g., Perez-Pereira 1999: 656; Mazzaggio 2016: 56), pragmatic factors, such as the communicative types of utterances, remain underexplored. The acquisition of the Estonian person category has so far examined only in one study which focused on the relationship between the acquisition of pronouns and verbs (Argus 2023).

This study aims to describe the interplay between pragmatic, semantic, and grammatical factors in the acquisition of the person category in Estonian. Specifically, it addresses the following research questions: (1) Which grammatical forms are acquired first—verbal inflections or personal pronouns? (2) Are there differences between zero-reference constructions and those with pronouns regarding verb semantics? (3) Are constructions with verbs of certain semantic types associated with specific utterance types? (4) What characterizes the first and most typical pragmatic types of utterances for each person in terms of verb semantics and the presence or absence of personal pronouns (e.g., grammatical devices like person inflection, pronouns, or forms of address)?

The qualitative study employs longitudinal data from spontaneous speech recordings (63 h) of three monolingual Estonian-speaking children aged 1;3–3;1 transcribed in chat-format (MacWhinney 2000). Approximately 4,000 utterances are coded for person, pragmatic communicative type (e.g., requestives, directives, questives, constatives), verb semantics, presence or absence of personal pronouns, and negation.

The results reveal that the acquisition of the person category reflects a complex interplay of different factors. Despite minor differences between children in the acquisition of person category specific types of person constructions can be detected. Pragmatic types of utterances appear to play the most prominent role in the acquisition of the person category: first person constructions are predominantly requestives, while second-person constructions are mostly directives, and third-person constructions constatives in the speech of all three children. Each person-related utterance type is characterized by a unique combination of grammatical, semantic, and pragmatic features. For instance, second-person constructions without pronouns emerge with action verbs in directive utterances, whereas first-person constructions are predominantly requestives featuring the verb *want* and the first-person pronoun.

References

Argus, Reili (2023), *Mina tahan, aga sina vaata*. Eri tähendusega verbide ja isikuliste asesõnade omandamisest eesti keeles [*I want but you look*. Acquisition of personal pronouns and verbs in Estonian], *Eesti ja soome-ugri keeleteaduse ajakiri* = Journal of Estonian and Finno-Ugric Linguistics, 14 (2), 111–128. DOI: 10.12697/jeful.2023.14.2.05.

Mazzagio, Greta (2016), The Theory of Mind's role in pronoun acquisition: The phenomenon of pronoun reversal in typically developing children, *Studies in the Linguistic Sciences* 2016, 55–69.

MacWhinney, Brian (2000), The CHILDES Project: Tools for Analyzing Talk. Mahwah, NJ: Lawrence Erlbaum Associates.

Perez-Pereira, Miguel (1999), Deixis, personal reference, and the use of pronouns by blind children, *Journal of Child Language* 26(03), 655–680. 10.1017/S0305000999003955.

The relation between early interactional skills and later language competence

Stina Andersson, Helena Engström, Elin Fryleskog, Rebecca Finndell, Elisabet Eir Cortes, Signe Tonér
& Tove Nilsson Gerholm
(Department of Linguistics, Stockholm University)

Keywords: Joint Attention, Parent-Child Interaction, First Language Acquisition, Core Language, Pragmatics

Joint Attention (JA) is an early pragmatic ability, reflecting an understanding of – or interest in – social interaction. In several studies, JA has been related to later vocabulary (Mundy et al. 2007, Markus et al 2001, Morales et al. 2000 and Tenenbaum et al. 2015) and grammar (Eriksson, 2019). However, it is unclear if the predictive power of early JA holds in longitudinal data spanning more than the child's first 2–3 years. Further, the relation between pragmatic abilities and core language skills is hard to disentangle: is it the interest in social interactions that favor core language, or are pragmatic skills in themselves some kind of water mark of linguistic talent, regardless of specific interest in socialization or the parent's interactional attempts? This study investigates whether it is possible to make far-reaching predictions from early interactive behaviours to language skills up to 10 years ahead.

Research questions and hypotheses. The present study aims to address these questions by investigating longitudinal associations between parental and child interactive behaviors at 1;0 and the child's core language skills at 3;0, 4;0 and 11;0 and his/her pragmatic skills at 7;0 and 11;0. The measured behaviors and assessments along with the hypothesized relations are as follows:

A child that spends much time in and initiates JA at 1;0 is hypothesized to develop better core language skills (Morphology at 3;0; SCDI at 4;0; CCC-2 at 11;0) as well as stronger pragmatic skills (Audience design at 7;0; CCC-2 at 11;0). Accordingly, a child that spends less time in or initiates less JA at 1;0 is hypothesized to develop lower core language skills and weaker pragmatic skills.

Additionally, a child whose parent talk more with him/her at 1;0, has more opportunities to practice. This is hypothesized to lead to better core language skills at 3;0; 4;0; and 11;0 and improved pragmatic skills at 7;0 and 11;0. Correspondingly, a child with a less talkative parent has fewer opportunities to practice, which is hypothesized to result in lower core language and pragmatic skills at the same ages.

Material and methodology. The data consists of 51 children followed from age 1 to 11, while interacting with a parent (The MINT project, https://www.su.se/english/research/research-projects/mint-the-role-of-interaction-and-parental-input-in-the-language-acquisition-process). Child assessment: frequency of JA at 1;0, productive language at 3;0 (morphology, TTR, etc; Tonér and Gerholm, 2021), SCDI at 4;0 (SCDI-III; Eriksson, 2017), Audience Design at 7;0 (Dahlgren and Sandberg, 2008), and CCC-2 at 11;0 (Bishop, 2003). Parental measures consist of time spent vocalizing to the child at 1;0. Possible relations were investigated by correlation and regression analysis.

Expected outcomes. If the assumed relations hold, the result is a clear and neat view of the relationship between pragmatics and core language skills. However, are the results inconclusive and/or indicative of more entangled relationships between the measures and skills, the result is fuel for a further discussion aiming to enhance the theoretical development within the field.

References:

Bishop, D. (2003). Children's Communication Checklist, 2nd ed. Pearson.

Dahlgren, S., & Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism*, *12*(4), 335–348.

Eriksson, F. (2019). J*oint attention and language acquisition: A longitudinal study of joint attention in parent-child interaction*. Master thesis, Dept of linguistics, Stockholm University.

Eriksson, M. (2017). The Swedish Communicative Development Inventory III: Parent reports on language in preschool children. *International Journal of Behavioral Development, 41*(5), 647-654

Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., and Yale, M. (2001). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development, 9*, 302–315.

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R., & Kern, H. (2000). Responding to joint attention across the 6- to 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, 21, 283–298.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Vaughan van Hecke, A., & Venezia Parlade, M. (2007). Individual Differences and the Development of Joint Attention in Infancy. *Child Development, 78*(3), 938-954.

Tenenbaum, E.J., Sobel, D.M., Sheinkopf, S.J., Shah, R.J., Malle, B.F., & Morgan, J.L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language, 42*(6), 1173-90.

Tonér, S., & Nilsson Gerholm, T. (2021). Links between Language and Executive Functions in Swedish Preschool Children: A pilot study. *Applied Psycholinguistics, 42*(1), 207-241.

Practicing register variation: children's imitation of adults in role play.
A pilot study

Veronika Mattes
(University of Graz)

Keywords: child-speech, adult-speech, register, imitation of adult speech, role play

It had been observed that children start to adapt to speech register during their 3rd year of life, mainly by modifying the prosodic structure (Dunn & Kendrick 1982, Hoff 2010), and that from age 6 on children choose the appropriate lexical and morpho-syntactic features for different registers (Aronsson & Thorell 2002, Blum-Kulka 2004, Sachs & Devin 2008).

The main subject of studies investigating the development of spontaneous child language by analyzing and comparing child-speech (CS) and child-directed speech (CDS), are utterances in direct child-parent interaction. Children's interactions with other communication partners, such as siblings, peers or other adults, are much less considered (cf. Havron et al. 2019, Levie et al. 2019), and adult-directed speech (ADS) overheard by children is usually not included in the analyses at all, following the assumption that it does not (essentially) contribute to the children's input (see Shneidman et al. 2013).

The aim of the presented paper is to identify morphological and morphosyntactic features (such as modality marking) of children's different registers in different communicative situations and with different interaction partners, as well as to suggest and discuss a methodology for investigating this topic. The results of a quantitative and qualitative analysis of spontaneous children's conversations in a longitudinal corpus of two monolingual German speaking siblings with an age difference of 3 years, in interaction with parents, peers, and among each other are presented. For the quantitative analysis, two data points were compared: age 3;0 (younger child) & age 6;3 (older child) and age 6;0 (younger child) & age 9;3 (older child).

We separately analyzed and compared the lexical and morphosyntactic features in
a) the children's utterances that are spontaneously directed at their parents and their siblings,
b) the children's utterances in role play situations, mimicking adults (e.g., teachers, parents, policemen) or children of different ages (e.g., teenagers, babies)
c) adults' child directed speech

The main research questions are:
- Which lexical, morphological and morphosyntactic features can be identified in different "mimicked" registers produced by children in role plays?
- How do the distributions of these features differ from spontaneous child speech and from adult's child-directed speech?

The data shows that certain features, for example imperative or subjunctive verb forms, are clearly associated with specific registers, and that children apply them adequately when mimicking the respective register. Furthermore, complex morphological and morphosyntactic structures (for example subjunctive modality or abstract nominalizations) that are only rarely found in child-adult or child-child conversations are well used by the children when imitating adults' ADS or CDS in role play.

The results highlight, that the usage of complex and low frequency structures is not only a matter of cognitive and linguistic development, but also strongly depends on the children's experience with different registers in different social interactions including those in which they are not directly involved. Furthermore, it suggests that overheard speech (see Oshima-Takane et al. 1996, Akhtar et al. 2001) might play a more prominent role for language acquisition than usually acknowledged.

References:

Akhtar, Nameera, Jennifer Jipson, and Maureen A. Callanan (2001), Learning words through overhearing, *Child Development*, 72, 416-430.

Aronsson, Karin, and Mia Thorell (2002), Voice and collusion in adult-child Talk: toward an architecture of intersubjectivity, in S. Blum-Kulka, and C. E. Snow (eds), (2002), *Talking to Adults: The Contribution of Multiparty Discourse to Language Acquisition*, Mahaw, NJ: Lawrence Erlbaum Associates, 277-293.

Blum-Kulka, Shoshana (2004), The role of peer interaction in later pragmatic development. The case of speech representation, in R. Berman (ed), *Language Development across Childhood and Adolescence*, Amsterdam: John Benjamins, 191-210.

Dunn, Judy, and Carol Kendrick (1982), The speech of two-and three-year-olds to infant siblings: "baby talk" and the context of communication, *Journal of Child Language* 9, 579-595.

Havron, Naomi, Franck Ramus, Barbara Heude, Anne Forhan, Alejandrina Cristia, and Hugo Peyre (2019), The effect of older siblings on language development as a function of age difference and sex, *Psychological Scienc*e 30, 1333-1343.

Hoff, Erika (2010), Context effects on young children's language use: The influence of conversational setting and partner, *First Language* 30(3-4), 461-472.

Levie, Ronit, Elitzur Dattner, Racheli Zwilling, Hadas Rosenstein, Shirley E. Stanzas, and Dorit Ravid (2019), Complexity and density of Hebrew verbs in preschool talk. The effect of socio-economic background, *The Mental Lexicon* 14(2), 237-273.

Oshima-Takane, Yuriko, Elizabeth Goodz, and Jeffrey L. Derevensky (1996), Birth order effects on early language development: Do secondborn children learn from overheard speech?, *Child Development* 67, 621-634.

Sachs, Jaqueline, and Judith Devin (1976), Young children's use of age-appropriate speech styles in social interaction and role-playing, *Journal of Child Language* 3(1), 81-98.

Shneidman, Laura, Michelle Arroyo, Susan C. Levine, and Susan Goldin-Meadow (2013), What counts as effective input for word learning?, *Journal of Child Language* 40(3), 672-686.

WS17 Reanalysis in cross-linguistic perspective: Theoretical and empirical implications

Sonia Cristofaro & Andrej Malchukov

Reanalysis of allative as dative in Ryukyuan

Aleksandra Jarosz

(Adam Mickiewicz University in Poznań)

Keywords: dative; allative; Goal; Recipient; Ryukyuan

In this presentation, I will consider the patterns and causes of allative > dative reanalysis in Ryukyuan, a subfamily of Japonic. The reanalysis concerns a multitude of allative case markers with diverse etymological backgrounds on the one hand, such as Okinawan and Miyako *-nkai*, Yaeyama *-kai*, *-nka*, *-(n)ga*, *-tti*, Kunigami *-tɕi*, *-kati*, Kikai-Amami *-e*, and the homogenous historical dative marker descending from Proto-Japonic *-ni on the other (cf. Jarosz 2025).

A conceptual extension from Goal to Recipient relies on a simple metaphor of Recipient, the sentient target of an event, as a spatial destination of that event, and it seems rather frequent cross-linguistically, being reported e.g. for Lezgian, Tamil, Tibetan, English or Romance (Heine et al. 2019: 266–267). Malchukov and Narrog's (2008) map of semantic roles predicts a facile extension of semantic domains of case markers from Goal to Recipient, from Recipient to Beneficiary, and from Beneficiary to Possessor. Once a marker acquires the encoding of Recipient, it establishes a precedent of allative encoding of roles typically associated with dative. The allative encoding may then pour out to other dative-marked roles, which in the case of Japonic languages include the passive agent in passive voice and the causee in causative voice clauses.

In some Ryukyuan languages, like Shuri-Okinawan and Shika-Ishigaki (East Yaeyama), the historical dative has been entirely replaced by an allative marker, and the reanalysis process can be considered complete (Jarosz 2025). In others, the change is ongoing, with three main patterns to be observed.

a) Allative > Recipient. Between verbs of physical transfer ('to send', 'to sell'), transfer of information ('to tell', 'to teach') and benefactive verbs, it is the verbs of physical transfer in which allative marking is most common and forms an implicational relationship with the other two. This clarifies the diachronic order within the Goal > Recipient (*ergo* allative > dative) reanalysis as starting from Recipient of a physical transfer, through addressee of a speech act, to Beneficiary.

b) Allative > passive agent. While some topolects use the allative freely to mark any kind of passive agent, in others (North Ryukyuan Okinoerabu, Tokunoshima) only the agent with features [-sentient], [-volitional] can host allative marking.

dʑidoːɕa-tɕi	ɕik-aj-un
car-ALL	drag-PSV-NPST

'He was hit by a car.' (Wadomari-Okinoerabu; Hirayama 1986, 880)

This might be the first diachronic step in replacing dative with allative marking in passive clauses. Notably, this pattern contradicts Malchukov and Narrog's (2008) prediction that the path from Goal to Agent leads through Recipient, at the same being semblant of "substitution of the goal for the source marker" as discussed by Ikegami (1987).

c) Allative > causee. Also here, some topolects, e.g. in Miyako, use the allative marking only under specific circumstances. In Irabu-Miyako (Shimoji 2018: 120–121), allative is considered acceptable on condition the causation event includes multiple participants in a dynamic

relationship, such as a transfer of the P argument from one causee to another. A direct conceptualization of causee as Goal may be the initial cause of causee permitting allative marking.

An implicational relationship is observed between the patterns a)–c) insofar as the possibility of allative marking on causee or passive agent also assumes the availabilty of allative Recipient marking. This underscores the hypothesis that what incites the reanalysis is a reconceptualization of Recipient as animate Goal. Passive agent and/or causee can then receive allative marking per analogiam.

Nevertheless, if allative marking on passive agent is first enabled on inanimate participants, strictly speaking it is difficult to postulate a direct extension of the allative marking from Goal through Recipient to Agent. Allative marking of Recipient is a facilitating factor, but not a cause, of passive agent also adopting allative marking. Likewise, allative marking of causee may be a product of an independent conceptualization process Goal > causee, rather than Goal > Recipient > causee. The question of to what extent the grammaticalization of allative in diverse participant-marking functions is dependent on the initial Goal > Recipient reconceptualization calls for an in-depth cross-linguistic testing.

References

Hirayama Teruo 平山輝男 (author/editor) (1986), *Amami hōgen kisogoi-no kenkyū* 奄美方言基礎語彙の研究 [A study of the basic vocabulary of Amami dialects], Tokyo 東京都: Kadokawa Shoten 角川書店.

Ikegami, Yoshihiko (1987), 'Source' vs. 'Goal': a case of linguistic dissymetry, in: R. Dirven, and G. Radder (eds.), (1987), *Concepts of case*, Tübingen: Gunter Narr Verlag Tübingen, 127–146.

Jarosz, Aleksandra (2025), Space in time: diachrony of Goal-participant marking in Ryukyuan languages, in: E. dal Corso, and E. Ragnanin (eds.), (2025), *Time in East Asian endangered languages – grammar, history, and society*, Venice: Edizioni Ca'Foscari, 131–162.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog, and Seongha Ree (2019), *World lexicon of grammaticalization. Second, extensively revised and updated edition*, Cambridge: Cambridge University Press.

Malchukov, Andrej, and Heiko Narrog (2008), Chapter 34. Case polysemy, in: A.

Malchukov, and A. Spencer (eds.), (2008), *The Oxford handbook of case*, Oxford: Oxford University Press, 518–534

Shimoji Michinori 下地理則 (2018), *Minami Ryūkyū Miyakogo Irabu hōgen* 南琉球宮古語伊良部方言 [Irabu dialect of Miyako, a Southern Ryukyuan language], Tokyo 東京都: Kuroshio Shuppan くろしお出版.

Generalizing reanalysis paths: challenges and prospects

Andrej Malchukov

(University of Mainz)

Keywords: reanalysis, grammaticalization, endocentric grammaticalization, exocentric grammaticalization

Unlike proposals in the literature trying to reduce grammaticalization to reanalysis, or reanalysis to grammaticalization, there are reasons to view the domains of the two phenomena as overlapping rather than coextensive (cf. Lehmann 2004). In my talk I discuss phenomena on the margins of grammaticalization research, which are more naturally described in terms of reanalysis. For example, changes in word order are conventionally described in terms of reanalysis rather grammaticalization, although they do involve some "grammaticalization parameters" such as obligatorification and (reduction of) syntagmatic variability (see Nübling and Kempf 2020: 134-139 on 'grammaticalization of word order' in Germanic). Some of such phenomena on the margins of grammaticalization research are briefly discussed in (Bisang et al. 2020: 75-77) in terms of "exocentric" grammaticalization paths. "Exocentric grammaticalization" differ from the conventional "endocentric grammaticalization" by lacking a single 'nucleus' (cf. "construction marker" in Himmelmann 2005) with a clear source-target relation on a grammaticalization path. Cases of "nucleus mismatches" pertain to cases when the target concept does not derive its function from its primary lexical source, like in cases where person pronouns develop from nouns with corresponding possessive person endings (see Janhunen 2020: 387 on dummy pronouns with personal possessive suffixes replacing personal pronouns in some Samoyedic), or development of "new demonstratives" from positional verbs in Hoocak (Helmbrecht 2020: 917-920). Another example of a path which is difficult to capture under the nucleus-based approach is (poly)grammaticalization of copula/auxiliaries constructions into variety TAM constructions dependent on the context. So, while context dependency is generally acknowledged in the studies of grammaticalization, it is important to acknowledge that this dependency is a matter of degree (cf. 'ordinary grammaticalization' vs 'constructional grammaticalization' as discussed for Daghestanian in Maisak 2020:353-354). On this view, endocentric grammaticalization, exocentric grammaticalization and reanalysis can be conceived as a cline (Bisang et al 2020: 76). The talk will review a number of such 'exocentric' developments which are more natural to describe in terms of reanalysis rather than grammaticalization.

Further, it is also interesting to consider the challenge of generalizing paths of reanalysis. While research of grammaticalization succeeded in producing an impressive catalogue of grammaticalization paths (Heine & Kuteva 2002 and its offshoots), no such catalogue has been proposed and even attempted for reanalysis. Rather here there have been partial successes for selected domains such as alignment change (see, e.g. Zuniga 2018 and Cristofaro 2024 for recent overviews). Another "success story" is that of insubordination (Evans 2007), although even here the range of possibilities is hardly

exhausted (see Malchukov 2013 on cross-linguistic variation in paths of complement insubordination). What these changes have in common and what sets them apart from the conventional cases of grammaticalization is that in the words of Harris & Campbell (1995: 65), multiple changes are happening at once. Thus, the talk takes up a challenge of constructing a typological catalogue of reanalysis, without doubling the catalogue of grammaticalization paths (in the work by Heine, Kuteva and colleagues).

486 words

References

Bisang, Walter and Andrej Malchukov (eds.) (2020). *Grammaticalization Scenarios: cross-linguistic variation and universal tendencies*, vol. 1-2. Berlin, Boston: De Gruyter Mouton.

Bisang, Walter, Andrej Malchukov and the Mainz Grammaticalization Project team (I. Rieder, L. Sun, M. Martiny, S. Luell). (2020). Universal and areal patterns in grammaticalization. In: Bisang &. Malchukov, 1-88.

Evans, Nicholas (2007). Insubordination and its uses, in Irina Nikolaeva (ed.), *Finiteness: Theoretical and Empirical Foundations*, Oxford: Oxford University Press, 366–431.

Cristofaro, Sonia (2024). Diachronic pathways to case marking alignment and what they mean for the explanation of synchronic cross-linguistic patterns. *Journal of Historical Linguistics* 14, 142–177.

Harris, Alice & Lyle Campbell (1995). *Historical Syntax in Cross-linguistic Perspective*. Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva (2002). *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press

Helmbrecht, Johannes (2020). Grammaticalization in Hoocak, In Bisang &. Malchukov, 903-943.

Himmelmann, Nikolaus (2005). Gram, construction, and class formation. In C. Knobloch & B. Schaeder (eds.), *Wortarten und Grammatikalisierung*. Berlin & New York: Mouton de Gruyter, 79-93.

Janhunen, Juha (2020). Grammaticalization in Uralic viewed from a general Eurasian perspective. In Bisang &. Malchukov, 361-399.

Lehmann, Christian (2004). Theory and method in grammaticalization. *Zeitschrift für germanistische Linguistik* 32. 152-187.

Maisak, Timur (2020). Grammaticalization in Lezgic (East Caucasian). In Bisang &. Malchukov, 309-360.

Malchukov, Andrej (2013). Verbalization and insubordination in Siberian languages, in M. Robbeets and H. Cuyckens (eds). *Shared Grammaticalization: With Special Focus on the Transeurasian Languages*. Amsterdam: John Benjamins. 177–208.

Zuniga, Fernando (2018). The Diachrony of Morphosyntactic Alignment. *Language and Linguistics Compass* 12:9.

The complex relationship between grammaticalization and reanalysis: some insights from a comparison between Romance languages

Anne Carlier
Sorbonne University, Paris (France)

Key-words: grammaticalization, reanalysis, persistence, partitive article, gerund constructions

Romance languages, while genetically related, exhibit significant diversity in the grammatical structures and markers they employ. Notably, even when the same source construction is present across different Romance languages, its diachronic evolution can take entirely divergent paths.

This paper aims to investigate the divergent development of such source constructions through processes of grammaticalization and/or reanalysis across these languages, offering valuable insights into the broader mechanisms underpinning reanalysis. The focus will be on two specific grammatical patterns from a comparative perspective: (i) the so-called partitive article, which incorporates the preposition *DE* ('from/of') within its internal structure and conveys non-singular indefiniteness (Carlier 2007); and (ii) the reanalysis of posture and motion verbs, when combined with a gerund, into verbal auxiliaries encoding features of tense, aspect, and argument structure (Bybee & al. 1994; Squartini 1998; Bertinetto 2000, Fanego 2020; Carlier & al. 2025).

(i) Fr. *Marie*   *boit*   ***de l'***   *eau*
       Mary   drink.PRS.3SG   of_DEF.ART   water
       'Mary drinks/is drinking water'

(ii) It. *Il potere delle istituzioni tradizionali **si*** ***andò***   ***indebolendo***
                                                    go.PST.3SG   weaken.GER
       'The power of traditional institutions has been weakening.' (gpedia.com)

The following research questions will be addressed:
- To what extent is frequency a necessary parameter for triggering reanalysis? What is the respective impact of an increase in *type frequency* vs *token frequency* (Bybee 2007)?
- To what extent can reanalysis be explained by alignment with existing constructions in the language? At what level of abstraction are these constructions situated? What is the possible role of analogy (Fischer 2007)?
- In the case of reanalysis, what is the importance of the phenomenon of 'persistence' (Hopper 1991, Cristofaro 2019)?

This study aims to contribute to the ongoing debate through a detailed, comparative, corpus-based analysis of the divergent evolution of two grammatical patterns, pertaining to distinct areas of grammar: the development of nominal determination and the emergence of verbal auxiliaries. The comparative approach seeks to generate hypotheses about why reanalysis takes place in some languages and fails to take place in others. It also invites reflection on why reanalysis, originating from the same source construction, can result in differing outcomes. The analysis of the partitive article will focus on Ibero- and Gallo-Romance languages, including French, Occitan, Catalan, and Spanish, while the examination of gerund constructions will compare Spanish, Italian, and French.

References

Bertinetto, P. M. (2000), The Progressive in Romance, as Compared with English, In Östen Dahl (ed.), *Tense and Aspect in the Languages of Europe*, 559–604, Berlin: De Gruyter Mouton.

Bybee, J. (2007), *Frequency of Use and the Organization of Language*, Oxford: Oxford University Press.

Bybee, J., R. Perkins, & W. Pagliuca (1994), *The Evolution of Grammar*, Chicago and London: University of Chicago Press.

Carlier, A. (2007), From preposition to article: the grammaticalization of the French partitive, *Studies in Language* 31:1, 1-49.

Carlier, A. & B. Lamiroy (2022), Partitive articles in Romance, in: F. Gardani and M. Loporcaro Eds. *Oxford Research Encyclopedia of Romance Linguistics*. Oxford: OUP.
<https://doi.org/10.1093/acrefore/9780199384655.013.638>.

Carlier A., L. Pigère & J. Vangaever (2025, forthc.), The progressive gerund construction in Spanish, Italian and French: comparative and historical perspectives, in: M. Garachana, R. Enghels & M. Sol Sansifiena (eds). *Periphrastic constructions in Romance: insights from synchronic and diachronic studies.* Berlin: De Gruyter.

Cristofaro, S. (2019), Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In K. Schmidtke- Bode, N. Levshina, S. M. Michaelis, & I. A. Seržant Eds., *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 25–46. Berlin: Language Science Press.

Fanego, T. (2020), On the history of the English progressive construction 'Jane came whistling down the street'. *Journal of English Linguistics* 48(4): 319-354.

Fischer, O. (2007), *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: OUP.

Harris, A. C. & L. Campbell (1995), *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: CUP.

Hopper, J. (1991), On some principles of grammaticization, in: E. C. Traugott & B. Heine (eds). *Approaches to Grammaticalization, I. Theoretical and Methodological Issues*, 17–36. Amsterdam: Benjamins.

Squartini, M. (1998), *Verbal Periphrases in Romance: Aspect, Actionality, and Grammaticalization*. Berlin: De Gruyter.

The adposition-to-applicative pathway: The diachrony of preverbs in Dutch

Ann-Sophie Vrielynck, Dirk Pijpops & An Van Linden

(University of Liège and University of Antwerp; University of Antwerp; & University of Liège)

Keywords: reanalysis, preverbs, Dutch, diachrony, corpus linguistics

This paper will investigate the role of reanalysis in the development of adpositions with spatial meaning towards applicative markers in the history of Dutch. Dutch compound verbs divide into separable (1) and inseparable (2) ones. Their non-verbal elements are often of spatial origin (e.g., *onder* ('under') and *over* ('over')), and may change the valency of the base verb. For example, while in (1), the spatial element is valency-neutral, it functions as an applicative marker in (2).

> (1) *De zon **gaat onder**.* ('The sun **goes down**.') (separable, non-applicative)

> (2) *Hij **ondergaat** zijn lijden.* ('He **undergoes** his suffering.') (inseparable, applicative)

Adpositions such as *onder* and *over* indeed instantiate a well-known spatial source of applicatives, as do adverbs. These often developed into so-called (applicativizing) preverbs in ancient Indo-European languages, but also in Germanic, Slavic and Baltic languages, where they appear as prefixes or particles (see Zúñiga et al. 2024). For the diachrony of Dutch preverbs, Blom (2004: 47-49) proposes a development in two steps that each involve reanalysis, a first from postposition to separable particle, to be found in OV contexts, and a second from separable particle to inseparable prefix, also in OV contexts. However, the plausibility of reanalysis is only assessed on the basis of self-constructed examples from Present-day Dutch. Furthermore, Van der Horst (2008: 39-40), for instance, goes against Blom (2004), arguing that separable and inseparable verbs typically did not evolve from each other; rather, each class developed independently from postpositional constructions. This lack of consensus highlights the need for further investigation into the diachrony of Dutch compound verbs.

To address this, we study the diachrony of two compound verbs, namely *ondergaan* ('go down; undergo') and *overlopen* ('overflow; go through'). Synchronically, these verbs occur in both separable and inseparable forms, with the separable forms being spatial and intransitive, and the inseparable forms being transitive and applicative. By tracing the diachrony of these two verbs, we aim to investigate the potential reanalysis of postpositions into particles or prefixes, and whether the inseparable forms developed out of the separable ones. Our analysis draws on data from two diachronic corpora spanning Middle Dutch to contemporary Dutch: the Corpus of Middle Dutch of poetry and prose texts (1300–1550; 9,707,462 tokens), and a subset of the DBNL literary corpus (1501–2000; 9,116,714 tokens; following Van Olmen 2019). The extracted attestations are annotated for separability, semantics of the preverb in the compound verb (spatial vs. non-spatial), semantics of the base verb in the compound verb (literal vs. figurative), and applicative value of the preverb.

Preliminary findings for *overlopen* reveal that both separable and inseparable forms coexisted in Middle Dutch. Unlike in contemporary Dutch, the separable forms in Middle Dutch were sometimes used transitively, primarily to express non-spatial and figurative meanings, which are now exclusively associated with inseparable forms. After the Middle Dutch period, the transitive use of separable forms disappeared and was fully taken over by inseparable forms.

References

Blom, Corrien (2004), On the diachrony of complex predicates in Dutch: Predicative and nonpredicative preverbs, *Journal of Germanic Linguistics* 16(1), 1–75.

Corpus Middelnederlands, https://corpusmiddelnederlands.ivdnt.org.

DBNL: *Digitale Bibliotheek voor de Nederlandse Letteren* ('Digital library for Dutch literature') <https://www.dbnl.org/>

van der Horst, Joop (2008), Vaste werkwoordelijke verbindingen in het Nederlands en het Afrikaans, *Atn&a*, 29–43.

Van Olmen, Daniël (2019), A diachronic corpus study of prenominal *zo'n* 'so a' in Dutch: pathways, analogy and (inter)subjectification, *Functions of Language* 26, 217–248.

Zúñiga, Fernando, Arkadiev, Peter & Hegedűs, Veronika (2024), Applicativizing preverbs in selected European languages, in F. Zuniga, and D. Creissels (eds), (2024), *Applicative Constructions in the World's Languages*, Berlin / Boston: De Gruyter Mouton, 419–472.

Reanalysis paths to alternativity connectives

Mira Ariel & Caterina Mauri
(University of Tel Aviv & University of Bologna)

The aim of this paper is to identify and discuss the recurrent paths of reanalysis leading to **alternativity connectives**, based on a diversity sample of 102 languages, for which information on the diachronic sources is available, complemented by a focus on a smaller set of diachronically well-documented languages, like Italian and Hebrew.

Alternativity connectives, such as English *or*, link two elements construed as (i) irrealis (cf. Mauri 2008) (ii) mutually replaceable, and (iii) neutral with respect to a speaker stance (a)symmetry (Ariel & Mauri, in press 2025). Such fully grammaticized connectives convey 'alternativity' as an encoded procedural meaning, presented in a nontransparent, compact form rather than being derived through inferential or compositional processes.

Our analysis shows that 'or's evolve from source constructions characterized by either symmetric or asymmetric stance. The source constructions begin with only a subset of the properties (i)-(iii) above (Table 1), acquiring missing features via a limited number of recurrent reanalysis patterns.

	(I) IRREALIS	(II) MUTUALLY REPLACEABLE	(III) STANCE (A)SYMMETRY
Symmetric *sources* ex. *If X… If Y*	coded	inferred	symmetric
Asymmetric *sources* ex. *X. If not, Y*	coded	coded or inferred	asymmetric
Connective (*target*)	**CODED**	**CODED**	**NEUTRAL**

Table 1. Properties of source and target constructions.

Example (1) presents a symmetric source, doubly marked by an irrealis marker (e.g., 'if') modifying each element. Irrealis is encoded for both elements, but mutual replaceability is not. The stance is symmetric (i.e. no preference for one of the alternatives). Once the first irrealis marker is reanalyzed as scoping over both elements, the second marker can be reinterpreted as an alternativity connective. A later change liberates the [X or Y] sub-construction from the initial irrealis marker, and symmetry is no longer necessarily imposed.

1) [IRR x]**,** [IRR y]   **>>**   [IRR [x OR y]]

    Hebrew (Biblical Hebrew, Exodus 21, 31) – source: IF
    O ben   **o**   *bat*   *yigaX…*
    If son   if >> or daughter (it) gores…
    'If it gores a boy or a girl…'

Example (2) presents an asymmetric source, here marked by a negative conditional construction. Typically, the first element is asserted, but the second potentially replaces it. Thus, irrealis is encoded, mutual

replaceability may be encoded, but stance is asymmetric. Once the marker is reanalyzed as scoping over both elements, it is reinterpreted as an alternativity connective. A later compactization process may obliterate the asymmetry between the original alternatives.

2) [x]**.** [NEG COND y].   **>>**   [x OR y]

Cavineña (Tacanan, Guillaume 2004: 114) – source: IF NOT THUS.

*Tuekedya =pa   ekana$_S$ tere-ya   kwejipa=eke   **jadyaamajuatsu***
then =REP   3PL   finish-IMPFV   strong.wind=PERL being.not.thus >> or
e-tiki=eke
NPF-fire=PERL
'… would die from the strong winds or from the fire.'

We will argue that the syntactic and semantic reanalysis of specific symmetric and asymmetric source constructions enables the evolution of compact and nontransparent (procedural) 'or' connectives, which introduce mutually replaceable irrealis alternatives, unmarked for symmetry stance.

References

Ariel, Mira, & Mauri, Caterina (in press, 2025), "Or" cycles, In M.-B. Hansen Mosegaard & R. Waltereit (Eds.), *Cyclic Change in Grammar and Discourse*, Oxford: Oxford University Press.
Mauri, Caterina (2008), The irreality of alternatives: Towards a typology of disjunction, *Studies in Language*, *32*(1), 22–55. https://doi.org/10.1075/sl.32.1.03mau

Word order change in Old English: functional pressures, discrete systems?

In the past, word order change has mostly been explained in terms of reanalysis – to the point where synchronic variation has been attributed to the coexistence of competing grammars (e.g. Pintzuk 1999). At the same time, the role of functional constraints such as weight and information structure have increasingly been taken into consideration, not only to explain synchronic variation, but also as possible contributors to change (e.g. De Bastiani 2021). Despite this trend, accounts of word order change predominantly remain couched in formal syntactic terms and ultimately depend on some discrete syntactic reanalysis.

Against this background, this paper explores whether accounts are possible that assume no abrupt changes to underlying syntax. To that end, it presents an analysis of subordinate clause word order in Old English. The focus is on the writings of Ælfric, whose grammar allowed for both the old Germanic OV order and the new, characteristically English VO order.

Subordinate clauses have been randomly sampled from Ælfric's *Catholic Homilies* (n=500), and from each clause one non-verbal and non-subject constituent has been randomly selected. Each selected constituent represents a single observation of the variable context. Each observation has been annotated for the following parameters: (i) pre- or post-verbal position (dependent variable); (ii) syntactic function (e.g. direct object); (iii) weight (light vs heavy); (iv) definiteness; (v) clause type of the subordinate clause (e.g. relative clause); (vi) presence of modals/auxiliaries in the clause; (vii) number of other non-clausal constituents in the clause; and (viii) presence of preverbal non-subject constituents in the clause.

Logistic regression analysis of Ælfric's usage shows that the choice between the two orders is subject to a variety of functional constraints, some of which likely contributed to the longer-term word order shift. These include the degree of semantic integration in adverbial clauses (lower integration correlates with main clause-like VO order) and dependence on modals and emerging auxiliaries (the less verb-like the auxiliary, the more likely main-clause like VO order in the 'dependent' infinitive clause). While this supports the idea that word order change may be driven by functional forces gradually reshaping the grammar (De Smet 2016), it is nevertheless striking that Ælfric's choices are categorical in a different respect: he rarely mixes OV and VO structures in a single clause, typically choosing either fully OV or fully VO syntax. This could be taken as evidence of distinct grammars coexisting in a single author and some radical reanalysis between them. Alternatively, it reflects the presence of a powerful analogical model shaping word order decisions in the subordinate clause: viz. the main clause.

De Bastiani, C. (2021). The reanalysis of VO in the history of English: Evidence for a language-internal account. In: B. Los, C. Cowie, P. Honeybone, G. Trousdale (Eds.), *English historical linguistics: Change in structure and meaning*, 115-136. Amsterdam: John Benjamins.

De Smet, H. (2016). Change drives change: Word order and the auxiliaries in Ælfric's Catholic Homilies. In: A. Bannink, W. Honselaar (Eds.), *From variation to iconicity: Festschrift for Olga Fischer*, (105-119). Amsterdam: Pegasus.

Pintzuk, S. (1999). *Phrase structures in competition: Variation and change in Old English word order*. New York: Garland.

The role of reanalysis in grammaticalization and degrammaticalization

Boye & Harder (2012) proposed that the lexical-grammatical contrast represents a conventionalization of syntagmatic discourse prominence potential such that lexical elements have the potential to represent the attentional foreground of a syntagm, whereas grammatical elements are by convention associated with attentional background status. Based on this, Boye (2023) and Westergaard & Boye (2023) proposed to define grammaticalization as the conventionalization of attentional background status.

We begin the present theoretical paper by arguing that this proposal not only entails a close relationship between grammaticalization and reanalysis, but also allows us to be precise about the nature of this relationship. We understand reanalysis in line with other usage-based approaches as a conventionalization of a non-conventional form-meaning association emerging in language use (e.g. Croft 2000: 117). This means that on the definition given above, grammaticalization is a special case of reanalysis: like reanalysis, it consists in conventionalization, and like reanalysis, it presupposes usage patterns that prefigure the element or structure to be conventionalized. What is special about grammaticalization is that the non-conventionalized (i.e. context-dependent) input meaning is attentional background in language use. Grammaticalization, as a special case of reanalysis, consists in the conventionalization of the association of this meaning with a given expression.

We then turn to degrammaticalization. Degrammaticalization is often understood as a process where a construction changes from having more grammatical features to fewer grammatical features (e.g., Lehmann 2015; Norde 2002). Instead, in line with the rejection of the idea of a lexical-grammatical continuum in Boye (2023), we propose to define degrammaticalization as the conventionalization of a potential for attentional foreground status, that is, a reanalysis resulting in lexical status.

Based on this definition, we consider a number of putative cases of degrammaticalization, and argue that most of these are really cases of lexical meaning change or of regrammaticalization. This holds even for most of the rather few examples identified by Haspelmath (2004) as "antigrammaticalization". We conclude that degrammaticalization is largely limited to the speaker-initiated innovation type that comprises cases like *-ism > ism*.

Subsequently, we present an account of the fact that degrammaticalization is so rare. The reason, we argue, has to do with the input meaning presupposed by degrammaticalization: the input meaning must be attentional foreground, but at the same time it must be grammatical. On the definition of grammatical elements as elements that are attentional background by convention, these two requirements are basically incompatible. As long as conventions are adhered to, grammatical elements cannot be used with attentional foreground status. Thus, degrammaticalization requires that conventions are overridden in a way in which a context-dependent meaning with attentional foreground status becomes associated with an expression conventionally associated with background status. This scenario is not unthinkable, but it is hard to imagine it in combination with a hearer-initiated change based on bridging contexts. It is rather more easy to imagine it in combination with speaker-initiated metalinguistic innovations of the dominating *-ism > ism* type.

References

Boye, Kasper. 2023. Grammaticalization as conventionalization of discursively secondary status: Deconstructing the lexical-grammatical continuum. *Transactions of the Philological Society* 121.2. 270-292.

Boye, Kasper & Peter Harder. 2012. A usage-based theory of grammatical status and grammaticalization. *Language* 88.1. 1-44.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow, England/New York: Longman.

Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In: Fischer, Olga, Muriel Norde & Harry Perridon (eds.), *Up and down the cline: The nature of grammaticalization*, Amsterdam: John Benjamins. 17-44.

Norde, Muriel. 2002. The final stages of grammaticalization: affixhood and beyond. In Ilse Wischer & Gabriela Diewald (eds) *New reflections on grammaticalization*. Amsterdam/Philadelphia: John Benjamins, 45-65.

Westergaard, Lennart & Kasper Boye. 2023. On semantic change in grammaticalization: Why it is never metaphoric. Paper presented at SLE 2023, Athens.

On the Semantic Extension of the Existential/Possessive Negator *mau³³tæ²¹* in Rucheng (Sinitic)

HE Lisha & LÜ Shanshan

(Centre for Linguistic Research on Eastern Asia [Centre de recherches linguistiques sur l'Asie orientale]; School for Advanced Studies in the Social Sciences [École des hautes études en sciences sociales]; & Shanghai International Studies University)

Keywords: Semantic extension, existential/possessive negator, negative causative verb, negative modal verb, Sinitic languages

The Rucheng language is the *lingua franca* of Rucheng County, which is under the administration of Hunan Province in China. It is a Sinitic language whose precise affiliation is a controversial issue (cf. *Language Atlas of China* 2012; Chen 2002; Coblin 2018). This language has two main negators: the general negative marker *n̩⁵⁵* and the perfective negative marker *mau³³*. This paper focuses on the negative verb *mau³³tæ²¹* 'to not have; there is not' in the Rucheng language, aiming to provide an overview of its multifunctionality and examine its semantic extension.

mau³³tæ²¹ is composed by the perfective negator *mau³³* and the verb *tæ²¹* 'obtain'. This compound negator can be used in three different contexts. First, like the existential negator in many Sinitic languages (Chappell & Peyraube 2016), *mau³³tæ²¹* can be used as a verbal negator expressing non-possession as well as non-existence, note that its positive form is *jou³³* 'there be; to have'. Second, the negator *mau³³tæ²¹* 'to not have, there is not' can serve as a negative modal verb denoting that internal or external circumstances do not allow an event to take place, with a meaning equivalent to 'cannot' in English. Third, it can be used as a negative causative verb denoting 'not let'.

We propose that the modal use of the negative verb *mau³³tæ²¹* is derived from its negative existential use. Our argument is as follows: First, the meaning of 'cannot' can be interpreted as 'there is no possibility that an event would take place'. Second, there is an important context which allows, syntactically, the modal function of *mau³³tæ²¹*, i.e. the context in which existential negator *mau³³tæ²¹* can take a verb as its complement. Only under this syntactic condition can *mau³³tæ²¹* develop into a negative modal verb. In fact, in Rucheng the morpheme *tæ²¹*, derived from the verb 'obtain', can also be used as a modal verb denoting 'can', even though this function is quite limited. However, we do not consider the modal verb *tæ²¹* as the source form of the modal use of *mau³³tæ²¹* 'cannot', because the negative form of *tæ²¹* is *n̩⁵⁵tæ²¹* 'cannot; unlikely', which turns out to be more frequently used than its positive form. Finally, the fact that internal or external circumstances do not allow an event to take place can give rise to a new meaning 'not let', from which the negative causative verb use of *mau³³tæ²¹* is derived.

Furthermore, we find that existential/possessive negators in some other Sinitic languages, particularly southern Sinitic languages, have functions similar to those in Rucheng and appear to have undergone the same process of semantic extension. For example, *hn̩²⁴⁻²²təʔ⁴* in the Haimen dialect of Wu (Wang 2011) and *mau²²ti⁴²* in the Pingjiang dialect of Gan. This semantic extension is rare and has not been well-studied from a typological perspective (Liu 2005, Chappell & Peyraube

2016). Our study can hopefully enrich the existing typology by providing a new semantic category of negation.

References:

Chappell, Hilary & Peyraube, Alain (2016), A typological study of negation in Sinitic languages: synchronic and diachronic views, in Pang-hsin Ting et al. (eds), New Horizons in the Study of Chinese: Dialectology, Grammar, and Philology – Studies in Honor of Professor Anne Yue, Hong Kong: T.T. Ng Chinese Language Research Center, Chinese University of Hong Kong, 483–534.

Chen, Lizhong (2002), Shilun Hunan Ruchenghua de guishu 試論湖南汝城話的歸屬 [A Discussion on the Classification of the Dialect of Rucheng, Hunan], Fangyan [Dialectes] 3, 227–231.

Coblin, W. South (2018), Neo-Hakka, Paleo-Hakka, and Early Southern Highlands Chinese, Yuyan yanjiu jikan 語言研究集刊 [Bulletin of Linguistic Studies] 21 (Special number in honor of Jerry Norman), Shanghai: Shanghai Cishu Chubanshe, 175–238.

Institute of Linguistics of the Chinese Academy of Social Sciences, Institute of Ethnology and Anthropology of the Chinese Academy of Social Sciences, and Language Information Sciences Research Centre of the City University of Hong Kong (2012), Zhōngguó yǔyán dìtú jí (dì 2 bǎn): Hànyǔ fāngyán juǎn 中國語言地圖集 (第二版): 漢語方言卷 [Language Atlas of China (2nd edition): Chinese dialect volume], Beijing: The Commercial Press.

Liu, Danqing (2005), Hanyu foudingci xingtai jufa leixing de fangyan bijiao 漢語否定詞形態句法類型的方言比較 [A dialectal comparison of syntactic types of negative sentences in Chinese], Chūgokugo Gaku 252, 1–22.

Wang, Hongzhong (2011), Hanmen Fangyan Yufa Zhuanti Yanjiu 海門方言語法專題研究 [Specialized Study on the Grammar of the Haimen Dialect], Anhui: Anhui Shifan Daxue Chubanshe.

The reanalysis of reflexives as transitivity and voice modulators in Romance

This paper discusses the reanalysis of reflexives as transitivity and voice modulators in the passage from Latin to Romance, focusing on the interplay of root and templatic aspects of verb meaning with the relational (e.g., the continuum of control, Lehmann 1988) and inherent (e.g., animacy) properties of verbal arguments in the reinterpretation of the reflexive morpheme (SE) as a passive and split intransitivity marker.

It is shown that the reanalysis of the reflexive as a marker of the passive voice is triggered by a change in the aspectual classes of verbs occurring in the sequence S(ubject) *se* V(erb), spreading from achievements (e.g., Latin *scindere* 'crack') (1a) and accomplishments (e.g., *aperire* 'open') (1b), including also degree achievements (e.g., *mutare* 'change') (1c), — with which SE most typically occurs in anticausative function (signalling the lack of an external causer and the spontaneous manifestation of an event), — to activity verbs (e.g. *vocare* 'call, name', *vendere* 'sell'), with which *se* signals an externally induced process and the construction has a passive interpretation (2a-b).

(1) a. *lutamenta*   *scindunt*   *se*

    plaster.N.PL   crack.PRS.IND.3PL RFL

   'Plaster cracks' (Cat. *Agr.* 128)

  b. *neque*   *se*   *Luna*   *quoquam*   *mutat*
    neither   RFL   moon   at.all   change.PRS.IND.3SG
    'And the moon does not change at all.' (Pl. 273)

  c. *valvae*   *se*   *ipsae*   *aperuerunt*
    doors.NOM.PL   RFL   themselves   open.PRF.IND.3PL
    'The doors suddenly opened of their own accord' (Cic. Div. 1, 34, 74)

2) a. *qui se*   *vocat*   *palude* ...
    that RFL   call.PRS.IND.3SG
    'That is called marsh …' (*Cod. Dipl. Bar*. 9)

  b. … *que*   *se*   *vendiderit*

    that   RFL   sell.PRF.SBJV.3SG

    'That will be sold'

    (Port. Mon. Hist. *Leges* 350; Bastardas Parera 1957: 120, note 1)

The lexico-aspectual properties of verbs also play a role in the reanalysis of the reflexive as a split intransitivity marker in Late Latin, whereby the accusative (*se*) and dative (*sibi*) forms of the reflexive

morpheme come to differentiate two subclasses of intransitives (unergatives and unaccusatives, respectively) (*sibi nasci* 'die' vs *se periurare* 'perjure'). (Dahlén 1964; Cennamo 1999). The distinction resurfaces in the Romance languages, albeit in different ways and to a different extent, still reflecting the core syntax and function(s) of the original Latin pattern, and is molded by the interplay of the verb's aspectual properties with the continuum of control (i.e., the thematic properties of the subject), leading to so-called 'pleonastic reflexives' (Reichenkron 1933; Hatcher 1942, among others), with both transitive and intransitive verbs (cf. Romanian *I se face de plimbare* 'He feels like walking', European Portuguese *já se aconteceu* 'It has already happened', Spanish *se me ha olvidado la cartera* 'I forgot my bag', *se murió* 'He died') (Cennamo 2016: 971 and references therein).

It will be shown that the data investigated provide interesting insights on the mechanisms and paths of change involved in well-known and widely discussed issues such as the reanalysis of reflexives and on the general semantic constraints at work in the syntactic domains considered in the present study.

References

Cennamo, M. 1999. Late Latin pleonastic reflexives and the Unaccusative hypothesis. *Transaction of the Philological Society* 97.1: 103–50.

Cennamo, M. 2016. Voice. In M. Maiden & A. Ledgeway (eds), *The Oxford Handbook of the Syntax of the Romance Languages*. Oxford: Oxford University Press, 967-980.

Cyrino, S. 2013. Argument promotion and se-constructions in Brazilian Portuguese. In E. van Gelderen, M. Cennamo & J. Barðdal (eds), *Argument Structure in Flux*. The Naples-Capri papers. Amsterdam/Philadelphia: Benjamins, 284–306.

Dahlén, E. 1964. *Etudes Syntaxiques sur les Pronoms Reflechis Pleonastiques en Latin.* Uppsala: Almqvist & Wiksell.

Hatcher, A.G. 1942. *Reflexive Verbs in Latin, Old French, Modern French*. Baltimore: The John Hopkins Press.

Lehmann, Ch. 1988. Predicate classes and participation. In *Studies in General Comparative Linguistics*. Köln: Institut für Sprachwissenschaft Universität zu Köln, 33–77. (Arbeiten des Kölner Universalien-Projekts 71).

Reichenkron, G. 1933. *Passivum, Medium und Reflexivum in den romanischen Sprachen*, Jena/Leipzig: Gronau (Berliner Beiträge zur romanischen Philologie III.1).

Prosodic re-phrasing

Workshop *Reanalysis* at 58th SLE

Nikolaus P. Himmelmann, Universität zu Köln

Changing the boundaries of grammatical units is at the very heart of (morphosyntactic) reanalysis, often illustrated by rebracketing: a suffix becomes part of the stem, the argument of a matrix predicate becomes a constituent of the embedded clause, and so on. In some (but not all!) instances, prosody has a major role to play in changing (prosodic and morphosyntactic) constituent structure. Here we will be concerned with two fairly frequent scenarios: the integration of extra-clausal constituents into the core clause, distinguishing processes at the left edge of the core clause (topic integration) from those at the right edge (afterthought integration). Schematically, following Givón (1976):

(1)   *The girls, they ran uphill -> The girls they-ran uphill*

(2)   *They met them again, the other boys -> They met-them again the other boys*

The first change has been claimed to play a role in word order change from VSO/VOS to SVO and the rise of subject marking. The second change has been claimed to play a role in the emergence of object indexing (pronominal markers on the verb co-referential with the object). Both changes often involve clitic doubling stages (core arguments being expressed by both a clitic on the verb and a noun phrase). And, more importantly for present purposes, they involve the deletion of a prosodic boundary (indicated by the comma).

The contribution explores the issue of what exactly happens when prosodic boundaries disappear. More specifically, it looks at the variability of these boundaries in different scenarios and languages, primarily Western Austronesian and Germanic. A typical issue is as follows. In Totoli, a western Austronesian languages, preverbal arguments allow for different types of prosodic packaging. One option is that there is a clear prosodic break after the preverbal argument as in (3) where *kalibombang* 'butterfly' ends on a continuation rise and is followed by a longish pause. This invites the analysis as a topical constituent. But another option, seen in Figure 4, is for the pre-verbal argument (*oto terek* 'truck') to end on a rise, but not to be followed by a long pause. Rather, the following verb is produced as a direct prosodic continuation of the contour started on *oto*. The interpretation of these cases is more ambiguous. There is evidence both for a topic analysis (the rise), but also for a subject analysis (lack of a clear prosodic break). Further options exist, as to be discussed in the presentation.

(3)   bali   **kalibombang**   no-gutu=mo=ko   aka'
        so   butterfly   AV.RLS-make=CPL=AND   cheat
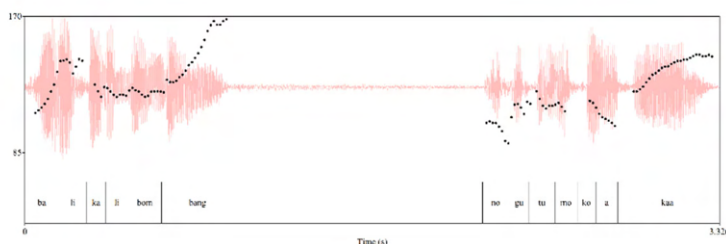   'so the butterflies cheated on him (the monkey)'   [monkey_butterfly.057]

Figure 1: F0 of example (1)

(4) **oto** **terek** saasake saapi teetengge
car truck RDP2:get.on cow RDP2:back
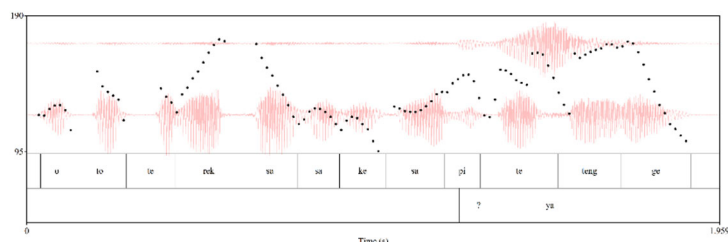'a truck is transporting a cow at the back' [spacegames_sequence1_KSR-SP.013]

Figure 2: F0 of example (2)

References

Ariel, Mira. 2000. The Development of Person Agreement Markers: from Pronouns to Higher Accessibility Markers. In Michael Barlow& Suzanne Kemmer (eds.), *Usage-based Models of Language*. Stanford: CSLI Publications, 197-260.

Bisang,Walter. & Andrej Malchukov (eds). 2020. Grammaticalization Scenarios: cross-linguistic variation and universal tendencies. Vol. 1. Berlin: De Gruyter Mouton.

Chafe, Wallace L. 1994. *Discourse, consciousness, and time*. Chicago: The University of Chicago.

Givón, Talmy. 1976. Topic, Pronoun, and Grammatical Agreement. In Charles N. Li (ed.), *Subject and Topic*. New York: Academic Press, 149-185.

Haig, Geoffrey. 2018. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56:781-818.

Haig, Geoffrey & Diana Forker. 2018. Agreement in grammar and discourse: A research overview. *Linguistics*, 56:715-734.

Himmelmann, Nikolaus P. 2022. Prosodic phrasing and the emergence of phrase structure. *Linguistics* 60: 715-743. https://doi.org/10.1515/ling-2020-0135

Lehmann, Christian. 1982. Universal and typological aspects of agreement. In: Hansjakob Seiler & Franz-Josef Stachowiak (eds), *Apprehension. Das sprachliche Erfassen von Gegenständen*, Teil II.: *Die Techniken und ihr Zusammenhang in Einzelsprachen*, Tübingen: Narr, 201-267.

Schnell, Stefan. 2018. Whence subject-verb agreement? Investigating the role of topicality, accessibility, and frequency in Vera'a texts. *Linguistics* 56:735-780.

Reanalysis of the French converb *en passant* as seen through diachronic and cross-linguistic data

Olga Nádvorníková & Laure Sarda

(Charles University of Prague & Lattice, CNRS – ENS-PSL and Université Sorbonne Nouvelle)

Keywords: converb, reanalysis, spatial semantics, cross-linguistic analysis, parallel corpus

This study examines the reanalysis of the French converb *en passant* from diachronic and cross-linguistic perspectives, using data from FRANTEXT and a large multilingual corpus annotated by *Universal Dependencies* (InterCorp). These corpora enable the analysis through equivalents in a typologically different language (Czech), serving as a semantic and functional mirror of the source construction (Dyvik 2004, and Nádvorníková 2024).

Converbs are defined as non-finite (dependent) verb-forms conveying adverbial subordination (Haspelmath 1995, Ross 2021, and Nedjalkov 1995). Their adverbial nature often leads to reanalysis as adverbs (Ylikoski 2003), e.g., cs. *takřka* 'almost' and fr. *en passant* 'quickly'. Adverbialized converbs may further develop into discourse markers through pragmaticalization (fr. *en attendant*, Vigier 2012). Grammaticalization pathways are also attested, with converbs developing into adpositions (cs. *vyjma* 'excepting') or conjunctions (fr. *en attendant que*, Kortmann 1997, and Halmøy 1982). Interestingly, Czech lacks examples of converbs grammaticalizing into conjunctions (Dvořák 1983), unlike French.

In spatial semantics (Talmy 2000, and Aurnague 2011), the verb *passer* conveys the movement of the Figure (F) – entity to be located relative to the Ground (G) – reference entity. We argue that two spatial configurations underpin the reanalysis of *en passant*:

(1) A distant and dynamic Figure/Ground relationship that exists momentarily as the Figure moves past the Ground ([…] *musiciens, qui chantoient en passant devant l'eschaffaut du roy*). This emphasizes brevity or negligence in reanalysis.

(2) A closer Figure/Ground relationship involving contact, overlap, or (momentary) inclusion (e.g., *aller de X a Z en passant par Y*). In this case, reanalysis focuses on the intermediate step, particularly in enumerative series.

The results show that the reanalysis pathways of *en passant* are mirrored in its Czech equivalents. Czech, as a satellite-framed language, encodes the meaning of *en passant* with a motion verb only when the motion meaning is preserved (fr. *en passant devant la maison* > cs. *když šel kolem domu* 'when passing in front of the house'). In cases of reanalysis, two pathways are revealed, highlighting specific semantic facets of the two scenarios:

(i) Adverbialization of the motion verb (e.g., fr. *dit-elle en passant > cs. řekla mimochodem* 'she said thoughtlessly'), especially with verbs of communication.

(ii) Reanalysis into a complex preposition (*en passant* par, Stosić 2012). In this case, Czech equivalents typically involve a preposition (*přes* 'across'), while initial and final phases may be encoded by converbs (*počínaje* 'starting' and *konče* 'ending').

In the analysis of examples drawn both from diachronic and multilingual corpora, we observe various factors potentially influencing reanalysis (syntactic position, subject control, properties of figure and Ground). Quantitative analysis examines factors such as the discourse frequency of the constructions. Using the UD annotation, we further analyse the properties of the head verb, the type of adjunct

construction (or its absence), etc. The results identify not only the different stages and types of reanalysis of *en passant,* and various bridging contexts of its evolution, but also the broader principles governing the reanalysis of converbs.

References

Aurnague, Michel (2011), How Motion Verbs Are Spatial: The Spatial Foundations of Intransitive Motion Verbs in French, *Lingvisticæ Investigationes* 34, 1–34. https://doi.org/10.1075/li.34.1.01aur.

Dvořák, Emil (1983), *Přechodníkové konstrukce v nové češtině* [Transgressives in New Czech], Praha: Universita Karlova.

Dyvik, Helge (2004), Translations as Semantic Mirrors: From Parallel Corpus to WordNet, *Language and Computers* 49, 311–326.

Halmøy, Odile (1982), *Le gérondif: Eléments pour une description syntaxique et sémantique*, Trondheim: Tapir.

Haspelmath, Martin (1995), The converb as a cross-linguistically valid category, in M. Haspelmath, and E. König (eds.), (1995), *Converbs in Cross-Linguistic Perspective*. *Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds*, Berlin / NY: Mouton de Gruyter, 1–57.

Kortmann, Bernd (1997), *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*, Berlin: Mouton de Gruyter.

Nádvorníková, Olga (2024), French, Polish and Czech Converbs: A Contrastive Corpus-Based Study, *Languages in Contrast* 24(2), 197–225. https://doi.org/10.1075/lic.00044.nad

Nedjalkov, Vladimir P. (1995), Some typological parameters of converbs, in M. Haspelmath, and E. König (eds.), (1995), *Converbs in Cross-Linguistic Perspective*. *Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds*, Berlin / NY: Mouton de Gruyter, 97–137.

Ross, Daniel (2025), The life cycle of converbs: A diachronic typology, in P. Cotticelli-Kurras, E. Dahl, and J. Živojinovič (eds.), (2025), *Diachronic, Typological, and Areal Aspects of Converbs*, Berlin / Boston: Mouton de Gruyter, 317–361.

Stosić, Dejan (2012), *En passant par* : une expression en voie de grammaticalisation ? *Corela. Cognition, représentation, langage*, no. HS-12 (December). https://doi.org/10.4000/corela.2844.

Talmy, Leonard (2000), *Toward a Cognitive Semantics: Concept Structuring Systems*, MIT Press.

Vigier, Denis (2012), *En attendant* : Un cas de pragmaticalisation. *Travaux de Linguistique : Revue Internationale de Linguistique Française* 64(1), 143–160.

Ylikoski, Jussi (2003), Defining Non-Finites: Action Nominals, Converbs, and Infinitives, *Journal of Linguistics* 16, 185–237.

Corpora

FRANTEXT. *https://www.frantext.fr/* ATILF, *Base textuelle Frantext* (En ligne), ATILF-CNRS & Université de Lorraine.

InterCorp v16ud (2024), https://wiki.korpus.cz/doku.php/en:cnk:intercorp. Institute of the Czech National Corpus, Prague: Charles University.

Causative Verbs as Object Markers: A pathway attested in Central Plains Mandarin

Shanshan Lü & Hilary Chappell

Shanghai International Studies University & Ecole des Hautes Etudes en Sciences Sociales, Paris

Sinitic object-marking constructions have long been studied from various perspectives and are generally considered to be highly transitive (Hopper & Thompson 1980). While a canonical Sinitic word order is VO, a typical Sinitic object-marking (OM) construction can be represented as [(S)-[OM-O]-V]. Three major sources for the OM in Sinitic include TAKE/HOLD verbs, GIVE/HELP verbs, and comitatives (Chappell 2013), while causative verbs (Huang & He 2016) and allatives (Chappell 2024) are considered as two minor sources.

Based on a corpus composed of over 12 hours of short-form videos of the Xuchang Central Plains Mandarin, spoken in Henan province, and taken from Douyin (the Chinese version of Tiktok), the present paper elaborates an argument in favour of the reanalysis from causative verb to object marker. In Xuchang, we observe that the two causative verbs *tɕiau³¹* 叫 and *ʐaŋ³¹* 让, both derived from speech act verbs, have developed into object markers:

$$[(S)\text{-}[\textbf{tɕiau}^{31}\textbf{/ʐaŋ}^{31}_{OM}\text{-}O]\text{-}VP].$$

We propose that the reanalysis can be modelled in terms of syntactic, semantic, and pragmatic factors. Syntactically, the source of both Sinitic causative and object-marking constructions is in serial verb constructions of the form [NP$_S$ V$_1$ NP V$_2$ X], which share the semantic feature of causativity (Bisang 1992). Nonetheless, we argue that predicate types, principally the domain of labile experiencer, and self-agentive verbs, are equally important in this process: they lead the way to an OM schema for this unusual diachronic pathway (see also Creissels 2024:ch.16.3; Croft 2022:212 on these special verb classes).

The example below shows one type of the three bridging contexts identified that can trigger the mechanism of reanalysis. The labile verb *ɕia³¹* 'scare, be scared' can give rise to two different syntactic analyses, depending on the context in question.

(1) tɕiau³¹   wɔ⁵⁵   ɕia³¹   liᵒ   jaᵒ

     let/OM   1SG   scare/be.scared   PRT PRT

     Causative verb: '(The big turtle) makes me so scared.'

     OM: '(The big turtle) scares me so much.'   (@两个小段 20220430)

First, the causative verb and the object marker can be clearly seen to occupy the same structural position. Second, the use of a labile, experiencer verb, 'scare/be scared' provides the right syntactic conditions for *tɕiau³¹* to develop into the more grammaticalized function of an object marker under transitivization of the predicate, required in the canonical OM construction. Hence, when *ɕia³¹* acts as an unaccusative intransitive verb denoting 'be scared', *tɕiau³¹* receives its interpretation as a causative verb. By contrast, when *ɕia³¹* is perceived as a transitive verb, *wɔ⁵⁵* '1SG' is understood as the object of *ɕia³¹*.

In the documented history of standard forms of the Chinese language, object-marking

constructions with *bǎ* and *jiāng* as OMs acquired causative meanings only after the extension of their object-marking function attained the stage of conventionalization (Wu 1997). However, the case of *tɕiau³¹* and *ʐaŋ³¹* in Xuchang shows the inverse direction of reanalysis.

REFERENCES

Bisang, Walter. 1992. *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer.* Tübingen: Gunter Narr Verlag.

Chappell, Hilary. 2013. Pan-Sinitic Object Marking: Morphology and Syntax. In Cao Guangshun, Hilary Chappell, Redouane Djamouri and Thekla Wiebusch (eds.), *Breaking Down the Barriers: Interdisciplinary Studies in Chinese Linguistics and Beyond.* Taipei: Academia Sinica. 785-816.

Chappell, Hilary. 2024. From oblique to core case in the Southern Min Languages: The role of topic in the emergence of optional object marking in Sinitic. *Journal of Historical Linguistics* 14(1). 66-107.

Creissels, Denis. 2024. *Transitivity, valence, and voice*. Oxford: Oxford University Press.

Croft, William. 2022. *Morphosyntax: Constructions of the world's languages*. Cambridge: Cambridge University Press.

Hopper, Paul J., and Thompson, Sandra A. 1980. 'Transitivity in Grammar and Discourse'. *Language* 56 (1). 251–99.

Huang, Xiaoxue, and Xuehui He. 2016. Cong *Qi Lu Deng* Kan Guanhua Zhong 'Jiao' Biao Zhuzhi de Xianxiang [On the phenomenon of causative *Jiao* as an object marker in *Qi Lu Deng*]. *Zhongguo Yuwen* 2016-6. 695-702.

Wu, Fuxiang. 1997. Tang Song Chuzhishi Ji Qi Laiyuan [On the disposal constructions in the Tang and Song dynasties and their origins]. *Cahiers de linguistique Asie orientale*, vol. 26(2). 201-220.

Explaining accusative alignment in case marking cross-linguistically: Reanalysis and its implications

Accusative alignment in case marking is traditionally explained in terms of principles of optimization in the encoding of A, S, and P arguments. A and S arguments are both topical, hence they are encoded in the same way because of their similarity. P arguments must be disambiguated from A arguments, hence they have a distinct form (Du Bois 1985, Comrie 1989, Dixon 1994, Mithun and Chafe 1999, among others).

This paper discusses the available diachronic evidence about the emergence of accusative alignment cross-linguistically, drawing on data from grammaticalization studies, historical linguistics, and the specialized literature on individual languages. Although generally unsystematic, this evidence shows that accusative alignment recurrently arises through processes of reanalysis that do not appear to be related to such optimization principles.

Some of these processes involve metonymization (Traugott and Dasher 2005): elements that initially do not encode grammatical relations (topic markers, focalizers, serial verbs) are reanalysed as markers for co-occurring arguments. This creates a distinct form for specific arguments (either P, or A and S), while the other arguments retain a form originally used for all arguments (see, e.g., Lord 1993, König 2008, or Melis 2021).

Other processes involve reanalysis of argument structure. A and S arguments evolve into one another, ending up being encoded in the same way. P arguments develop a form distinct from that of A and S arguments as they evolve from oblique or possessor NPs and retain the form of these NPs. These processes involve multiple source constructions: different types of intransitive constructions where an S argument and an oblique or possessor NP encode notional agents and patients, respectively (Harris and Campbell 1995, Gildea 1998, among others); transitive constructions where an oblique NP encoding a beneficiary is reanalysed as a P argument (Chappell, Peyraube, and Wu 2011, Chappell 2013); or transitive constructions with light verbs, which are reanalysed as intransitive (Harris 2002, Creissels 2008).

These various processes reflect factors unrelated to principles of optimization in the encoding of A, S, or P arguments. These factors include, for example, the fact that some source element co-occurs with particular arguments, leading to its being reanalysed as a marker for those arguments, or the fact that specific NPs encode notional agents or patients, leading to their being reanalysed as A or P arguments. In the resulting alignment pattern, the fact that A and S argument are encoded in the same way and differently than P arguments is likewise independent of such principles. Instead, this originates from inheritance (arguments retaining the form of the NPs from which they derive) and residue (some arguments retaining a form originally used for all arguments as new forms develop for others).

These facts suggest that, insofar as accusative alignment or other cross-linguistic patterns result from reanalysis, their explanation requires understanding multiple factors that may drive reanalysis in diverse source constructions across languages. Such factors are likely tied to the properties of individual constructions and their contexts of use, rather than optimization principles concerning the resulting patterns in themselves.

References

Chappell, H. (2013). Pan-Sinitic object markers: morphology and syntax. In *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*, pp. 785–816. Taipei: Academia Sinica.

Chappell, H., A. Peyraube, and Y. Wu (2011). A comitative source for object markers in Sinitic languages: kai55 in Waxiang and kang7 in Southern Min. *Journal of East Asian Linguistics 20*, 291–338.

Comrie, B. (1989). *Language universals and linguistic typology. 2nd edition*. Oxford: Basil Blackwell.

Creissels, D. (2008). Direct and indirect explanations of typological regularities: the case of alignment variations. *Folia Linguistica 42*, 1–38.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.

Du Bois, J. A. (1985). Competing motivations. In J. Haiman (Ed.), *Iconicity in syntax*, pp. 343–66. Amsterdam and Philadelphia: John Benjamins.

Gildea, S. (1998). *On reconstructing grammar: Comparative Cariban morphosyntax*. Oxford: Oxford University Press.

Harris, A. C. (2002). *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

Harris, A. C. and L. Campbell (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

König, C. (2008). *Case in Africa*. Oxford: Oxford University Press.

Lord, C. (1993). *Historical change in serial verb constructions*. Amsterdam and Philadelphia: John Benjamins.

Melis, C. (2021). From topic marking to definite object marking: Focusing on the beginnings of Spanish DOM. In J. Kabatek, P. Obrist, and A. Wall (Eds.), *Differential Object Marking in Romance: The third wave*, pp. 39–64. Berlin and Boston: De Gruyter.

Mithun, M. and W. Chafe (1999). What are S, A, and O? *Studies in Language 23.3*, 569–96.

Traugott, E. C. and R. B. Dasher (2005). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Reanalysis is not enough in processes of grammaticalization—
The case of Chinese

Walter Bisang
(University of Mainz)

As is well known, many markers of grammatical categories in Chinese and mainland Southeast Asian languages are based on verbal elements expressing grammatical functions such as tense-aspect-modality-evidential (TAME), phasal or deictic motion (directional verbs; DIR) and semantic roles ('coverbs' or adpositions; P). The present paper aims to model the grammaticalization of these markers on the basis of data from Chinese with its rich diachronic sources. As representative examples will show, the modeling of these structures requires the combination of reanalysis, analogy, and constructionalization.

Taking classical Chinese ($5^{th} - 3^{rd}$ centuries BC) as a point of departure, the use of verbs in grammatical functions is attested to a certain extent for TAME (1). The other functions developed later. An important starting point is the resultative construction, consisting of two verbs with V_1 being the main verb and V_2 expressing a resultant state (2). According to Xu (2006), it developed around the 1st century BC, when Chinese had largely lost its morphology and the expression of an action and its reaching a stable state had to be expressed by two separate verbs (2). As this structure has a new form (two verbs instead of one) and a new meaning (reaching a terminal state), it fulfills the condition of being analysed as a new construction in terms of constructionalization (Traugott & Trousdale 2013).

Although resultative constructions oscillate between the lexicon and full productivity to this day, its V_2-position is the basis for the development of the more strongly grammaticalized markers of directional verbs (DIR), postverbal TAME markers and adpositions (P). In each case, new markers have evolved in association with specific positions relative to the main verb. This is illustrated here by 'give'-verbs in P-function (Peyraube 1988 and many others since). At a first stage, a few verbs of giving (*yǔ* 與 'give', *yǔ* 予 'give', 遺 *wèi* 'offer') in V_2 were used for introducing datives/benefactives (3). Since about 400 AD, dative/benefactive 'give' also occurs after the object in V_2/P (V_1 NP V_2/P NP) (4). This is in analogy to prepositions with no verbal background. Finally, the preverbal occurrence of [V_2/P NP] in (5) can be seen in analogy to *zài* 'be at', which occurred earlier in that position as a locative marker.

In later periods, other 'give'-verbs are found in these positions, including 给 *gěi* 'give' since the 17th century. These processes, which also operate with positions for TAME and DIR, can either be seen in terms of analogy to earlier 'give'-verbs or as instances of the reanalysis of new 'give'-verbs for dative/benefactive (cf. Fisher 2007 on analogy vs. reanalysis). No matter which mechanism applies, the relevant positions attract new lexical items of compatible semantics to mark functions of TAME, DIR and P, respectively.

Conclusion: Reanalysis is an important factor in processes of grammaticalization in Chinese. However, the development of function-related positions for TAME, DIR and P and their use cannot be modeled without additional mechanisms such as analogy, constructionalization and, in addition, contact, since similar structures exist in many mainland Southeast Asian languages.

Examples

(1)   Classical Chinese (Zuo, Xuan 2), the verb *yǐ* 'do away with, stop', preverbal use:
己殺孔父而弒殤公

yǐ	*shā*	*Kǒng fù*	*ér*	*shì*	*Shāng gōng.*
stop/PFV	kill	Kong father	and	kill	Shang duke

'He killed Kong's father and then he killed duke Shang.'

(2)   Chinese from *Zhànguó zònghéng jiāshū*, 195 BC, (Xu 2006: 163):
秦戰勝魏，走孟卯。

Qín	*zhàn-shèng*	*Wèi,*	*zǒu*	*Mèng Mǎo.*
Qin	V₁:battle-V₂:win	Wei	expel	Meng Mao

'[The state of] Qin (battled and) defeated [the state of] Wei and expelled Meng Mao.'

(3)   Chinese from Shiji about 100 BC (Peyraube 1988: 146):
安國君及夫人因厚餽遺子楚。

Ānguó- jūn	*jí*	*fūrén*	*yīn*	*hòu*	*kuì*	*wèi*	*Zǐchǔ*
Anguo	and	wife	reason	generously	V₁:give.present	V₂:offer	Zichu

'This was why prince Anguo an his wife generously gave a present to Zichu.'

(4)   Chinese from *Shishuoxinyu* 400 AD (Peyraube 1988: 186):
送一車枝與和公。

sòng	*yī*	*chē*	*zhī*	*yǔ*	*Hé gōng.*
send	one	waggon	branches	V/P:to	He duke

'[He] sent one wagon-load of branches to duke He.'

(5)   Chinese from Dunhuang Bianwen, about 1.000 AD (Peyraube 1988: 289):
與維摩不教些些。

yǔ	*Wéimó*	*bú*	*jiào*	*xiē*	*xiē*
P/give	Vimalakirti	NEG	teach	thing	thing

'Do not teach these things to Vimalakirti (= Weimo).'

References

Fischer, Olga (2007), *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: Oxford University Press.

Peyraube, Alain (1988), *Syntaxe diachronique du chinois. Évolution des constructions datives du XIVᵉ siècle av. J.-C. au XVIIIᵉ siècle*. Paris: Colège de France: Institut des hautes études chinoises.

Traugott, Elisabeth C. & Trousdale, Graeme (2013), *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Xu, Dan (2006), *Typological Change in Chinese Syntax*. Oxford: Oxford University Press.

WS18 Taking time seriously: The temporal dynamics of language

Guido M. Linders, Catalina Torres, Stefan Schnell & Frank Seifart

Syntax in Time

Balthasar Bickel
(University of Zurich)

Comparative research increasingly demonstrates that the difference between humans and animals rests not so much in capacities for hierarchical structure (from knowledge taxonomy to hierarchical event cognition) but in the specific capacity of humans to regiment such structure into a temporally linearized string ready for communication with language (Zuberbühler & Bickel 2022; Brocard et al. 2024; Wilson et al. 2024). As a result, questions of time move center-stage for our understanding of syntax in language. Here I explore the consequences of this for the mapping between semantic role hierarchies and syntactic expression.

Neurophysiological evidence suggests that comprehension is affected by the high rank of agents because this drives incremental interpretation over time (Sauppe et al. 2023; Huber et al. 2024): when the comprehension system encounters an initial unmarked noun phrase it transiently infers an agent reading, long before the rest of the sentence is parsed, an effect that is independent of the eventual meaning of the sentence once it's over, its grammatical structure (word order, case), and the contextual frequencies of its words (surprisal).

This temporal dynamic is critically different during sentence planning (Sauppe et al. 2021; Giglio et al. 2024). Here, speakers plan noun phrases based on their current salience (Christianson & Ferreira 2005; Branigan et al. 2008; MacDonald 2013) and for this, the thematic role hierarchy is only one out of many competing factors. What matters decisively more is the ability to start an utterance at a speed sufficient for the fast pace of human face-to-face conversation (Levinson 2016). As a result of this, the neurophysiology of sentence planning does not evidence an agent preference but instead reveals neuronal activity suggesting delays in commitment to a fully-fledged plan (Sauppe et al. 2021; Egurtzegi et al. 2022). This strategy is best executed by placing an unmarked NP as early as possible, regardless of its role as topic, agent or patient, thus keeping syntactic continuations as much open as possible.

These temporal perspectives explain a global biases in linguistic evolution for unmarked NPs to denote agents, not patients (Bickel et al. 2015), and for sentences to preferably start with unmarked NPs, not verbs (Dryer 2013). The production perspective makes a further prediction on rates of change. Since initial verbs and initial ergative-marked NPs constrain syntactic continuations more than unmarked NPs, we expect less variation in how speaker execute sentence plans and, because of this, an overall slow-down of syntactic change. Change is expected to speed up again only through contact with languages dominated by sentences starting with unmarked NPs. These predictions are consistent with earlier observations that verb-initial and ergative languages seem to form diachronically remarkably stable but areally confined hotbeds (Nichols 1992; Nichols 2003).

Overall, these observations suggest that taking a temporally explicit perspective on syntax helps explain trends and preferences in linguistic evolution that remain mysterious and arbitrary choices from the point of view of argument mapping in the abstract terms used by grammatical theories. Differences in case alignment and word order are equally good mappings; their differences only appears in the light of how comprehension and production operate in real time.

References

Bickel, Balthasar, Alena Witzlack-Makarevich, Kamal K Choudhary, Matthias Schlesewsky & Ina Bornkessel-Schlesewsky. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10. e0132819. http://dx.doi.org/10.5167/uzh-112203.

Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189. doi:10.1016/j.lingua.2007.02.003.

Brocard, Sarah, Vanessa A. D. Wilson, Chloé Berton, Klaus Zuberbühler & Balthasar Bickel. 2024. A universal preference for animate agents in hominids. *iScience*. 109996. doi:10.1016/j.isci.2024.109996.

Christianson, Kiel & Fernanda Ferreira. 2005. Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98(2). 105–135. doi:10.1016/j.cognition.2004.10.006.

Dryer, Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures*. Leipzig: Max Planck Institute for Evolutionary Anthropology [http://wals.info/]. http://wals.info/chapter/81.

Egurtzegi, Aitor, Damián E. Blasi, Ina Bornkessel-Schlesewsky, Itziar Laka, Martin Meyer, Balthasar Bickel & Sebastian Sauppe. 2022. Cross-linguistic differences in case marking shape neural power dynamics and gaze behavior during sentence planning. *Brain and Language* 230. 105127. doi:10.1016/j.bandl.2022.105127.

Giglio, Laura, Markus Ostarek, Daniel Sharoh & Peter Hagoort. 2024. Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. *Proceedings of the National Academy of Sciences* 121(11). e2310766121. doi:10.1073/pnas.2310766121.

Huber, Eva, Sebastian Sauppe, Arrate Isasi-Isasmendi, Ina Bornkessel-Schlesewsky, Paola Merlo & Balthasar Bickel. 2024. Surprisal From Language Models Can Predict ERPs in Processing Predicate-Argument Structures Only if Enriched by an Agent Preference Principle. *Neurobiology of Language* 5(1). 167–200. doi:10.1162/nol_a_00121.

Levinson, Stephen C. 2016. Turn-taking in human communication – origins and implications for language processing. *Trends in Cognitive Sciences* 20. 6–14.

MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 1–16. doi:10.3389/fpsyg.2013.00226.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: The University of Chicago Press.

Nichols, Johanna. 2003. Diversity and stability in language. In Richard D. Janda & Brian D. Joseph (eds.), *Handbook of Historical Linguistics*, 283–310. London: Blackwell.

Sauppe, Sebastian, Kamal K. Choudhary, Nathalie Giroud, Damián E. Blasi, Elisabeth Norcliffe, Shikha Bhattamishra, Mahima Gulati, Aitor Egurtzegi, Ina Bornkessel-Schlesewsky, Martin Meyer & Balthasar Bickel. 2021. Neural signatures of syntactic variation in speech planning. *PLoS Biology* 19. e3001038. doi:10.1371/journal.pbio.3001038.

Sauppe, Sebastian, Åshild Næss, Giovanni Roversi, Martin Meyer, Ina Bornkessel-Schlesewsky & Balthasar Bickel. 2023. An Agent-First Preference in a Patient-First Language During Sentence Comprehension. *Cognitive Science* 47(9). e13340. doi:10.1111/cogs.13340.

Wilson, Vanessa A. D., Sebastian Sauppe, Sarah Brocard, Erik Ringen, Moritz M. Daum, Stephanie Wermelinger, Nianlong Gu, Caroline Andrews, Arrate Isasi-Isasmendi, Balthasar Bickel & Klaus Zuberbühler. 2024. Humans and great apes visually track event roles in similar ways. (Ed.) Leyla Isik. *PLOS Biology* 22(11). e3002857. doi:10.1371/journal.pbio.3002857.

Zuberbühler, Klaus & Balthasar Bickel. 2022. Transition to language: From agent perception to event representation. *WIREs Cognitive Science*. e1594. doi:10.1002/wcs.1594.

Separate in time but united by structure: The case of cross-serial and nested dependencies

Catalina Torres, Sebastian Sauppe & Balthasar Bickel
University of Zurich; NCCR Evolving Language

Formal language theory (FLT) predicts that crossed and nested dependencies are processed differently because they separate in time what belongs together in hierarchical structure. This study asks whether the increased memory demand posited for crossed dependencies (Jäger & Rogers, 2012; De Vries et al., 2012) also implies greater difficulty, as measured in response accuracy. Additionally, if this processing differs, we ask whether intonation helps, as it has been argued to do in other complex structures (Schafer, 1997). The intonational structure of discontinuous phrases is known to vary considerably across languages (Féry & Fanselow, 2020; Fanselow & Féry 2006; Swerts & van Wijk, 2005; Sanfelici et al., 2020). We report data from the production (Exp. 1) and comprehension (Exp. 2) of adjoined, nested, and crossed dependencies in Swiss German (SG). SG provides a unique test case because it allows variation in different dependency types, while keeping compositional semantics and information structure constant (Schieber, 1985).

We use sentence-final clusters of an embedding verb and a full verb, each with their own object (in pseudo-English translation of 'that Maria saw Manu marinate the lamb', **crossed**: that Maria Manu$_1$ the lamb$_2$ saw$_1$ marinate$_2$; **nested**: that Maria Manu$_1$ the lamb$_2$ marinate$_2$ saw$_1$; **adjoined**: that Maria Manu$_1$ saw$_1$ the lamb$_2$ marinate$_2$). Speech stimuli in Exp I were selected to allow for measurement of a continuous fundamental frequency (F0) contour and 20 participants were recorded performing a controlled reading task. Auditory stimulus sentences in Exp II were acoustically manipulated by resynthesizing duration and F0 to reduce the acoustic strength of prominences. Forty participants listened to 180 unique critical sentences (50% prosodically manipulated) and responded to comprehension questions.

The combination of word-order variation allows us to assess the relationship between syntax and prosody in speech production and comprehension. In Exp I we investigate the intonational contours of words in the verb clusters using generalised additive mixed-effects models. Results show that objects consistently carry pitch accents while embedding verbs are deaccented and full verbs can be accented depending on their position. The majority of word order variations are realised as single Intonation Phrases (IP) with an obligatory IP final boundary tone. In Exp 2, we use binomial generalized linear mixed models to assess the probability of accurate responses to comprehension questions finding that crossed dependencies realised with a natural prosody are more likely to elicit correct responses.

This study finds evidence for two predictive rules structuring the intonational structure of linearised verb clusters. First, syntactic category is predictive of pitch accent assignment. Second, prosodic phrasing determines the placement of an IP-final boundary tone. Additionally, contra predictions of FLT, we find that crossed dependencies were not more difficult to comprehend. Instead, we find that a natural prosody in crossed dependencies aligns with more accurate responses. However, this effect doesn't apply to all dependency types calling into question the relationship between prosody and syntax more broadly. This study contributes to our understanding of real-time syntactic processing by investigating variable linearisation. Our experimental paradigm allows us to disentangle syntactic from prosodic parsing. Additionally, it allows us to discuss response accuracy in a meaningful way since all participants are confronted with the three linearised verb clusters, and we don't compare across different speaker groups, as previously done (Bach et al., 1986). We suggest that human brains can equally well accommodate structures above and below the context-sensitive boundary of the Chomsky hierarchy, although it remains an open question whether they do so using the same neural resources and strategies.

References:

[1] Jäger, G., & Rogers, J. (2012). Formal language theory: refining the Chomsky hierarchy. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1598), 1956-1970.

[2] De Vries, M. H., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: Evidence from sequence learning. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1598), 2065-2076.

[3] Schafer, A. J. (1997). Prosodic parsing: The role of prosody in sentence comprehension. University of Massachusetts Amherst.

[4] Féry, C., & Fanselow, G. (2020). Prosody of discontinuous nominal phrases in Indian languages. Journal of South Asian Linguistics, 10, 60-87.

[5] Fanselow, G., & Féry, C. (2006). Prosodic and morphosyntactic aspects of discontinuous noun phrases: A comparative perspective. Ms., University of Potsdam.

[6] Swerts, M., & van Wijk, C. (2005). Prosodic, lexico-syntactic and regional influences on word order in Dutch verbal endgroups. Journal of Phonetics, 33(2), 243-262.

[7] Sanfelici, E., Féry, C., & Schulz, P. (2020). What verb-final and V2 have in common: evidence from the prosody of German restrictive relative clauses in adults and children. Zeitschrift für Sprachwissenschaft, 39(2), 201-230.

[8] Shieber, S. M. (1985). Evidence against the context-freeness of natural language. In The Formal complexity of natural language (pp. 320-334). Dordrecht: Springer Netherlands.

[9] Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. Language and Cognitive Processes, 1(4), 249-262.

Information transmission and speech timing: A cross-linguistic time-series analysis

Coupé Christophe, Oh Yoonmi, Seifart Frank, Dediu Dan, Stave Matt & Pellegrino François
(The University of Hong Kong; Ajou University; CNRS lab Structure et Dynamique des Langues and Humboldt-Universität zu Berlin; ICREA & University of Barcelona; CNRS lab Dynamique du Langage; & CNRS lab Dynamique du Langage)

Information transmission is the landmark of all communication systems. In linguistic communication, speakers emit sequences of linguistic units – from phonemes to sentences – according to complex rules, many of which are language-specific. Information appears to be distributed strategically over these units, with regulatory processes preventing spikes of information. This is expressed by the *Uniform Information Density Hypothesis* (e.g., Meister et al. 2021), and information trade-offs have been highlighted within languages across speakers or texts, as well as across languages. Pimentel et al. (2021) found for instance an inverse relation between the quantity of information borne by phones and their acoustic durations in around 600 languages, while Coupé et al. (2019) highlighted a trade-off between syllabic rate and syllabic information density across 17 languages.

Most studies focusing on such trade-offs adopt statistical approaches – usually regression models – which break down the speech signal into consecutive units then considered independently of each other. Here, we adopt another approach and study speech rate and information content as *time series*. We apply this to the DoReCo corpus (v2.0) (Seifart et al. 2024), which contains time-aligned spoken data from an areally and genealogically balanced, world-wide sample of over 50 languages (see reference section for a list of languages).

We used the English translation of the annotation units in the original languages to estimate their amount of information. We relied on the predictions of a GPT2 language model (Radford et al. 2019) since it predicts self-paced reading well (Wilcox et al. 2020) and better than larger models (Oh and Schuler 2022). We divided the data in five-second-long sections – as a tradeoff between having enough context and a suitable temporal resolution – and for each computed the (average) syllabic rate (SR), the (average) syllabic information density (SD), and the (total) amount of information (SURPRISAL). With these sections as elementary time units, we focused on time series with at least 40 sections and conducted a time-frequency analysis with wavelet transforms. This approach allows to explore the relationships between relevant variables at different frequencies / periodicities, and we considered two complementary approaches: i) for each time series, a computation of the average phase and lag between SD and SR at different frequencies, and ii) generalized additive models predicting SR with a range of predictors (sex, age, type of speech, cumulated duration, SD, SURPRISAL), across the different time series and across different frequency bands thanks to partial reconstructions of the signal in these bands.

Both approaches support our main result, which is that across a range of typologically diverse languages, the trade-off between SR and SD holds true at lower and higher frequencies of the speech signal. This suggests that regulation of information takes place at different time scales of the organization of speech

(with the limitation of our 5-second threshold required by our computation of SURPRISAL). Further, the amount of information (SURPRISAL) positively correlates with SR at different frequencies. Future work will rely on our framework to focus on cross-linguistic differences.

References

Oh, Byung-Doh, and William Schuler (2023), Why Does Surprisal From Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times?, *Transactions of the Association for Computational Linguistics* 11, 336–350.

Coupé, Christophe, Yoonmi Oh, Dan Dediu, and François Pellegrino (2019), Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche, *Science Advances* 5(9), eaaw2594.

Meister, Clara, Tiago Pimentel, Patrick Haller, Lena Jäger, Ryan Cotterell, and Roger Levy (2021), Revisiting the Uniform Information Density Hypothesis, in Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds), (2021), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana: Association for Computational Linguistics, 963-980.

Pimentel, Tiago, Clara Meister, Elizabeth Salesky, Simone Teufel, Damián Blasi, and Ryan Cotterell (2021), A surprisal–duration trade-off across and within the world`s languages, in Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds), 2021, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing,* Online and Punta Cana: Association for Computational Linguistics, 949-962.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever (2019), Language Models are Unsupervised Multitask Learners. *OpenAI*. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Seifart, Frank, Ludger Paschen, and Matthew Stave (eds.), 2024, *Language Documentation Reference Corpus (DoReCo) 2.0*. Lyon: Laboratoire Dynamique Du Langage (CNRS & Université Lyon 2).

Wilcox, Ethan G, Jon Gauthier, Jennifer Hu, Peng Qian, and Roger P. Levy (2020), On the Predictive Power of Neural Language Models for Human Real-Time Comprehension Behavior, *Proceedings of the Annual Meeting of the Cognitive Science Society*, 42.

Corpus references

Language	Family	Area	Reference
Anal	Sino-Tibetan	Eurasia	Ozerov, Pavel. 2024. Anal DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.0dbazp8m.
Arapaho	Algic	North America	Cowell, Andrew. 2024. Arapaho DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.36f5r1b6.
Asimjeeg Datooga	Nilotic	Africa	Griscom, Richard. 2024. Asimjeeg Datooga DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.f77c7m72.

Baïnounk Gubëeher	Atlantic-Congo	Africa	Cobbinah, Alexander Yao. 2024. Baïnounk Gubëeher DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.a332abw8.
Beja	Afro-Asiatic	Africa	Vanhove, Martine. 2024. Beja DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.edd011t1.
Bora	Boran	South America	Seifart, Frank. 2024a. Bora DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.6eaf5laq.
Cabécar	Chibchan	North America	Quesada, Juan Diego, Stavros Skopeteas, Carolina Pasamonik, Carolin Brokmann & Florian Fischer. 2024. Cabécar DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ebc4ra22.
Cashinahua	Pano-Tacanan	South America	Reiter, Sabine. 2024. Cashinahua DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.a8f9q2f1.
Daakie	Austro-nesian	Papu-nesia	Krifka, Manfred. 2024. Daakie DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.efeav5l9.
Dalabon	Gunwi-nyguan	Australia	Ponsonnet, Maïa. 2024. Dalabon DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.fae299ug.
Dolgan	Turkic	Eurasia	Däbritz, Chris Lasse, Nina Kudryakova, Eugénie Stapert & Alexandre Arkhipov. 2024. Dolgan DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.f09eikq3.
English (Southern England)	Indo-European	Eurasia	Schiborr, Nils Norman. 2024. English (Southern England) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.9c271u5g.
Evenki	Tungusic	Eurasia	Kazakevich, Olga & Elena Klyachko. 2024. Evenki DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.5e0d27cu.
Fanbyak	Austro-nesian	Papu-nesia	Franjieh, Michael. 2024. Fanbyak DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.02084446.
French (Swiss)	Indo-European	Eurasia	Avanzi, Mathieu, Marie-José Béguelin, Gilles Corminboeuf, Federica Diémoz & Laure Anne Johnsen. 2024. French (Swiss) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.3520l685.
Goemai	Afro-Asiatic	Africa	Hellwig, Birgit. 2024. Goemai DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.b93664ml.
Gorwaa	Afro-Asiatic	Africa	Harvey, Andrew. 2024. Gorwaa DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.a4b4ijj2.
Gurindji	Pama-Nyungan	Australia	Meakins, Felicity. 2024. Gurindji DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ab2d89mx.
Hoocąk	Siouan	North America	Hartmann, Iren. 2024. Hoocąk DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.b57f5065.
Jahai	Austro-Asiatic	Eurasia	Burenhult, Niclas. 2024. Jahai DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.6a71xp0p.
Jejuan	Koreanic	Eurasia	Kim, Soung-U. 2024. Jejuan DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.06ebrk38.
Kakabe	Mande	Africa	Vydrina, Alexandra. 2024. Kakabe DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.d5aeu9t6.
Kamas	Uralic	Eurasia	Gusev, Valentin, Tiina Klooster, Beáta Wagner-Nagy & Alexandre Arkhipov. 2024. Kamas DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.cdd8177b.

Komnzo	Yam	Papu-nesia	Döhler, Christian. 2024. Komnzo DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.c5e6dudv.
Light Warlpiri	Mixed Language	Australia	O'Shannessy, Carmel. 2024a. Light Warlpiri DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.7452803q.
Lower Sorbian	Indo-European	Eurasia	Bartels, Hauke & Marcin Szczepański. 2024. Lower Sorbian DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.6c6e4e9k.
Mojeño Trinitario	Arawakan	South America	Rose, Françoise. 2024. Mojeño Trinitario DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.cbc3b4xr.
Movima	Isolate	South America	Haude, Katharina. 2024. Movima DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.da42xf67.
Nafsan (South Efate)	Austro-nesian	Papu-nesia	Thieberger, Nick. 2024. Nafsan (South Efate) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ba4f760l.
Nisvai	Austro-	Papu-nesia	Aznar, Jocelyn. 2024. Nisvai DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.2801565f.
Northern Alta	Austro-nesian	Papu-nesia	Garcia-Laguia, Alexandro. 2024. Northern Alta DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.efea0b36.
Northern Kurdish (Kurmanji)	Indo-European	Eurasia	Haig, Geoff, Maria Vollmer & Hanna Thiele. 2024. Northern Kurdish (Kurmanji) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ca10ez5t.
N‖ng	Tuu	Africa	Güldemann, Tom, Martina Ernszt, Sven Siegmund & Alena Witzlack-Makarevich. 2024. N‖ng DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.f6c37fi0.
Pnar	Austro-Asiatic	Eurasia	Ring, Hiram. 2024. Pnar DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.5ba1062k.
Resígaro	Arawakan	South America	Seifart, Frank. 2024b. Resígaro DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ffb96lo8.
Ruuli	Atlantic-Congo	Africa	Witzlack-Makarevich, Alena, Saudah Namyalo, Anatol Kiriggwajjo & Zarina Molochieva. 2024. Ruuli DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.fde4pp1u.
Sadu	Sino-Tibetan	Eurasia	Xu, Xianming & Bibo Bai. 2024. Sadu DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.3db4u59d.
Sanzhi Dargwa	Nakh-Daghestanian	Eurasia	Forker, Diana & Nils Norman Schiborr. 2024. Sanzhi Dargwa DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.81934177.
Savosavo	Isolate	Papu-nesia	Wegener, Claudia. 2024. Savosavo DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.b74d1b33.
Sümi	Sino-Tibetan	Eurasia	Teo, Amos. 2024. Sümi DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.5ad4t01p.
Svan	Kartvelian	Eurasia	Gippert, Jost. 2024. Svan DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.9ba054c3.
Tabaq (Karko)	Nubian	Africa	Hellwig, Birgit, Gertrud Schneider-Blum & Khaleel Bakheet Khaleel Ismail. 2024. Tabaq (Karko) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.eea8144j.

Tabasaran	Nakh-Daghe-stanian	Eurasia	Bogomolova, Natalia, Dmitry Ganenkov & Nils Norman Schiborr. 2024. Tabasaran DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ad7f97xr.
Teop	Austro-nesian	Papu-nesia	Mosel, Ulrike. 2024. TeVop DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.9322sdf2.
Texistepec Popoluca	Mixe-Zoque	North America	Wichmann, Søren. 2024. Texistepec Popoluca DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.c50ck58f.
Totoli	Austro-nesian	Papu-nesia	Bardají i Farré, Maria, Christoph Bracks, Claudia Leto, Datra Hasan, Sonja Riesberg, Winaro S. Alamudi & Nikolaus P. Himmelmann. 2024. Totoli DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.c8b6ei29.
Urum	Turkic	Eurasia	Skopeteas, Stavros, Violeta Moisidi, Nutsa Tsetereli, Johanna Lorenz & Stefanie Schröter. 2024. Urum DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.ac166n10.
Vera'a	Austro-nesian	Papu-nesia	Schnell, Stefan. 2024. Vera'a DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.3e2cu8c4.
Warlpiri	Pama-Nyungan	Australia	O'Shannessy, Carmel. 2024b. Warlpiri DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.042dv614.
Yali (Apahapsili)	Nuclear Trans New Guinea	Papu-nesia	Riesberg, Sonja. 2024. Yali (Apahapsili) DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.9d91nkq2.
Yongning Na	Sino–Tibetan	Eurasia	Michaud, Alexis. 2024. Yongning Na DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.abe65p95.
Yucatec Maya	Mayan	North America	Skopeteas, Stavros. 2024. Yucatec Maya DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.9cbb3619.
Yurakaré	Isolate	South America	Gipper, Sonja & Jeremías Ballivián Torrico. 2024. Yurakaré DoReCo dataset. In Seifart et al. https://doi.org/10.34847/nkl.7ca412wg.

The effects of cross-linguistic word order differences on semantic role interpretation in human sentence processing: an ontogenetic perspective

Eva Huber, Balthasar Bickel & Sabine Stoll
(University of Zurich; University of Cologne, University of Zurich & University of Zurich)

Keywords: Semantic Roles, Language Acquisition, Language Processing, Computational Modelling, Word Order

Sentence comprehension is an intrinsically incremental process, and comprehenders are sensitive to the linear order of words (Bates and MacWhinney 1982) and to the probabilistic relationship between them (Elman, Hare, and McRae 2004; Levy 2008). Languages strongly vary in how they linearise the linguistic signal (Greenberg 1963) and to what degree they allow word order flexibility (Payne 1992; Meir et al. 2017; Levshina 2019). Given these differences, speakers of different languages need to vary their parsing strategies when incrementally building up a representation of 'who does what to whom'. An unresolved question is how children learn to adapt their parsing strategies to the language(s) they are exposed to. One possibility is that semantic roles are interpreted at similar points in a sentence across languages, allowing children to rely on universal parsing strategies in determining semantic roles. Alternatively, differences would suggest early adaption of parsing strategies during development. To evaluate these possibilities, we study three languages differing in basic word order, i.e., their pragmatically neutral and unmarked (Siewierska 1988; Dryer 2007), word order: Tagalog (verb-initial), English (verb-medial) and Turkish (verb-final). The data comes from (longitudinal), naturalistic child language corpora (Tagalog: children aged 2;0-4;0 years: Garcia and Kidd (2022), English, 1;8-3;2 years: Lieven, Salomo, and Tomasello (2009) and Theakston et al. (2001), Turkish, 1;0-3;0 years: Küntay, Koçbaş, and Taşçı (n.d.)). We extract utterances with transitive verbs (5382 for Tagalog, 3065 for English, 4847 for Turkish) directed to children and annotate them with protoroles (Dowty 1991), i.e., agent or patient. We implement a recurrent neural network model. Like humans, such models incrementally predict whether an argument is an agent or a patient. This model allows us to estimate the probability of each role at every time point in the sentence. We fit Bayesian hierarchical beta models to estimate the probability of the correct semantic role at different sentence positions with *language*, *sentence position* and *role* as main predictors of interest.

We find that the agent is predicted above chance across positions and languages, shown by the the posterior probability differences between languages and across positions. The patient is not well predicted without additional information; for Tagalog and Turkish the verb is required, while in English both the verb and the agent are required for accurate patient prediction. This suggests that children likely interpret the agent role correctly, regardless of its position in the sentence and the specifics of the input language. The correct interpretation of the patient by contrast relies on the presence of an other argument and/or the verb, and it varies across languages. For children, the agent role may thus function as a universally available anchor point, facilitating accurate sentence processing and identifying the cues that indicate patient roles. Our results suggest that children may rely on a universal parsing strategy in identifying the agent but gradually adapt their parsing strategies to accommodate language-specific demands when interpreting the patient.

References

Bates, Elizabeth and Brian MacWhinney (1982), Functionalist approaches to grammar, *Language acquisition: The state of the art*, 173–218.

Dowty, David (1991), Thematic Proto-Roles and Argument Selection, *Language* 67(3), 547–619, DOI: DOI: 10.2307/415037.

Dryer, Matthew S (2007), Word order, *Language typology and syntactic description* 1(61-131), 1–1.

Elman, Jeffrey L, Mary Hare, and Ken McRae (2004), Cues, constraints, and competition in sentence processing, in (eds), (2004), *Beyond Nature-Nurture*: Psychology Press, 161–188.

Garcia, Rowena and Evan Kidd (2022), Acquiring verb-argument structure in Tagalog: a multivariate corpus analysis of caregiver and child speech, *Linguistics* 60(6), 1855–1906, DOI: DOI: 10.1515/ling-2021-0107.

Greenberg, Joseph H. (1963), Some universals of grammar with particular reference to the order of meaningful elements, in Greenberg, Joseph H. (eds), (1963), *Universals of human language*, Cambridge, Mass: MIT Press, 73–113.

Küntay, Aylin Copty, Dilara Koçbaş, and Süleyman Sabri Taşçı (n.d.). *Koç university longitudinal language development database on language acquisition of 8 children from 8 to 36 months of age*. unpublished.

Levshina, Natalia (2019), Token-based typology and word order entropy: A study based on Universal Dependencies, *Linguistic Typology* 23(3), 533–572, DOI: DOI: 10.1515/lingty-2019-0025.

Levy, Roger (2008), Expectation-based syntactic comprehension, *Cognition* 106(3), 1126–1177.

Lieven, Elena, Dorothé Salomo, and Michael Tomasello (2009), Two-year-old children's production of multiword utterances: A usage-based analysis, *Cognitive Linguistics* 20(3), 481–508.

Meir, Irit, Mark Aronoff, Carl Börstell, So-One Hwang, Deniz Ilkbasaran, Itamar Kastner, Ryan Lepic, Adi Lifshitz Ben-Basat, Carol Padden, and Wendy Sandler (2017), The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages, *Cognition* 158, 189–207, DOI: DOI: 10.1016/j.cognition.2016.10.011.

Payne, Doris L. (1992), *Pragmatics of Word Order Flexibility*, Amsterdam: John Benjamins Publishing Company.

Siewierska, Anna (1988), Word order rules (Croom Helm linguistics series) London, *UK: Croom Helm*.

Theakston, Anna L., Elena V. M. Lieven, Julian M. Pine, and Caroline F. Rowland (2001), The role of performance limitations in the acquisition of verb-argument structure: an alternative account, *Journal of Child Language* 28(1), 127–152, DOI: DOI: 10.1017/S0305000900004608.

Rhythmic types vs. continuum: An empirical test on 53 languages

Frank Seifart & Volker Dellwo

(CNRS-SeDyL Villejuif and Humboldt Universität Berlin & Universität Zürich)

Keywords: Rhythm-class hypothesis, Typology, Corpus phonetics, Vowel and consonant durations

Languages differ regarding regularities in timing at the level of syllables or stressed intervals, as suggested by the rhythm-class hypothesis, proposing that languages can be classified into syllable-timed (e.g. French, Telugu, Yoruba) vs. stress-timed languages (e.g. English, Russian, Arabic) (Pike 1945; Abercrombie 1967). This hypothesis assumes a tendency towards isochrony of syllable units vs. stress units, and—in stress-timed languages—reduction of unstressed syllables, even though later studies have failed to find evidence for isochrony (e.g., Bertinetto 1989). The prevalent view nowadays is that across languages the distinction between stress- vs. syllable-timed is continuous, rather than categorical (Dauer 1983, White & Malisz 2020). But empirical evidence for this, based on larger language samples and consistency in methods and data compared (e.g. read vs. spontaneous), is still missing.

Here, we address this gap by analyzing vowel and consonant durations in DoReCo (Seifart et al. 2024), a corpus of spontaneous spoken language from an areally and genealogically balanced world-wide sample of 53 languages (see "corpus references", below). It contains close to 10,000 words on average per language that had been transcribed and annotated by language experts. Within DoReCo, these had been time-aligned at the phone level (Figure 1) using automatic procedures with manual corrections (Paschen et al. 2020). We will apply a set of commonly used metrics (Ramus et al. 1999, White & Malisz 2020) to compare different aspects of rhythmicity across these languages, controlling for inter-speaker variation and speech rate (Dellwo 2010).

Preliminary results, without rate normalisation, for two metrics (deltaC, the standard deviation of consonantal intervals, and percentV, the percentage over which speech is vocalic, Figure 2) suggest that, typologically, languages do not fall into clearly distinct rhythmic types, but rather form a continuum, reaching from higher consonantal and vocalic variability (top-left of the graph) to lower variability (bottom-right). Highlighted in the graph are results for English ("stress-timed") and French ("syllable-timed") from DoReCo and from a previous study (Ramus et al., 1999). These two results are consistent in that English has higher deltaC and lower percentV. The differences in magnitude along these dimensions are likely due to differences in speaking styles (read vs. spontaneous), dialects of French and English data used, and in alignment methods. Interestingly, most languages stay well below the deltaC value obtained for DoReCo English, suggesting that stress-timing of English is typologically exceptional.

Pending results for further measures of rhythmicity, and the application of control factors for inter-speaker variation and speech rate, we expect to conclude that rhythmic variability between languages is indeed continuous rather than categorical. We will also discuss if and if so how exactly English is exceptional in a wider context of the timing properties of the languages of the world.
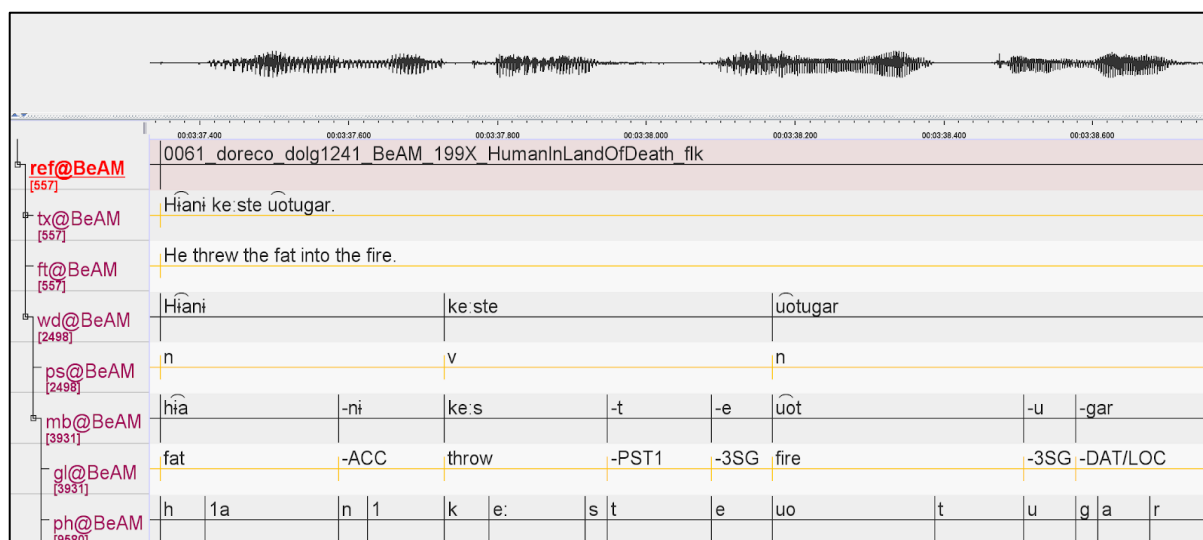
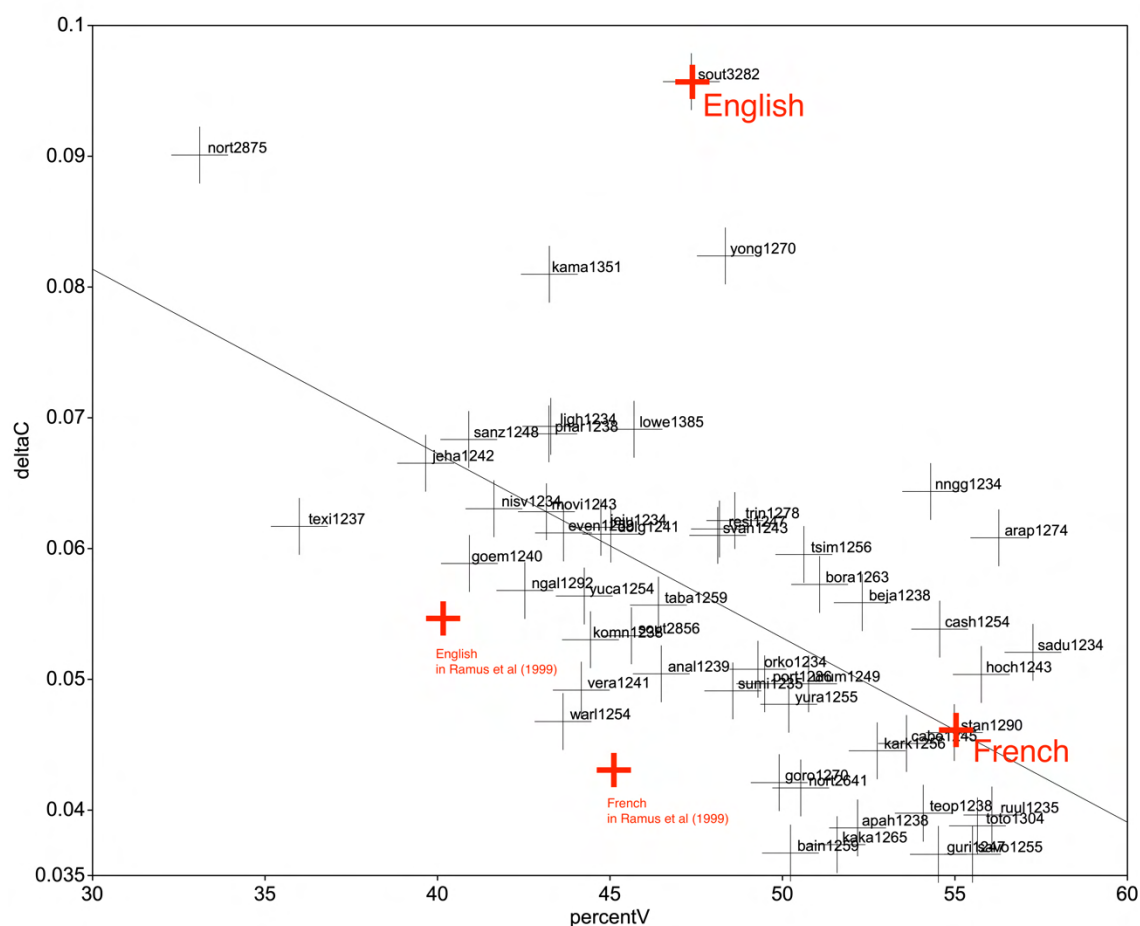Figure 1: DoReCo data for Siberian Dolgan (Däbritz et al. 2024)

Figure 2: DeltaC and percentV for DoReCo languages, identified by Hammarström et al.'s (2024) Glottolog codes (see "corpus references"), and from Ramus et al. (1999).

References (general)

Abercrombie, David. 1967. *Elements of General Phonetics*. Chicago: Aldine.

Bertinetto, Pier Marco. 1989. Reflections on the dichotomy "stress- vs. syllable-timing". *Revue de Phonétique Appliquée* 91/93. 99–130.

Dauer, R. M. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1). 51–62.

Dellwo, Volker. 2010. *Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence.* Universität Bonn PhD dissertation.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. *Glottolog 5.1*. Leipzig. https://doi.org/10.5281/zenodo.14006617.

Paschen, Ludger, François Delafontaine, Christoph Draxler, Susanne Fuchs, Matthew Stave & Frank Seifart. 2020. Building a Time-Aligned Cross-Linguistic Reference Corpus from Language Documentation Data (DoReCo). In *Proceedings of The 12th Language Resources and Evaluation Conference*, 2657–2666. Marseille, France: European Language Resources Association. https://www.aclweb.org/anthology/2020.lrec-1.324.

Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Ramus, Franck, Marina Nespor & Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3). 265–292.

Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2024. *Language Documentation Reference Corpus (DoReCo) 2.0.* Lyon: Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2).

White, Laurence & Zofia Malisz. 2020. Speech Rhythm and Timing. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford handbook of language prosody*, 167–182. Oxford: Oxford University Press.

Corpus references

Language	Glottolog	Family	Area	Reference (all in Seifart et al. (eds.) 2024)
Anal	anal1239	Sino-Tibetan	Eurasia	Ozerov, Pavel. 2024. Anal DoReCo dataset. https://doi.org/10.34847/nkl.0dbazp8m.
Yali Apahapsili	apah1238	Nucl. Trans New Guinea	Papunesia	Riesberg, Sonja. 2024. Yali (Apahapsili) DoReCo dataset. https://doi.org/10.34847/nkl.9d91nkq2.
Arapaho	arap1274	Algic	North America	Cowell, Andrew. 2024. Arapaho DoReCo dataset. https://doi.org/10.34847/nkl.36f5r1b6.
Baïnounk Gubëeher	bain1259	Atlantic-Congo	Africa	Cobbinah, Alexander Yao. 2024. Baïnounk Gubëeher DoReCo dataset. https://doi.org/10.34847/nkl.a332abw8.
Beja	beja1238	Afro-Asiatic	Africa	Vanhove, Martine. 2024. Beja DoReCo dataset. https://doi.org/10.34847/nkl.edd011t1.
Bora	bora1263	Boran	South America	Seifart, Frank. 2024a. Bora DoReCo dataset. https://doi.org/10.34847/nkl.6eaf5laq.
Cabécar	cabe1245	Chibchan	North America	Quesada, Juan Diego, Stavros Skopeteas, Carolina Pasamonik, Carolin Brokmann & Florian Fischer. 2024. Cabécar DoReCo dataset. https://doi.org/10.34847/nkl.ebc4ra22.
Cashinahua	cash1254	Pano-Tacanan	South America	Reiter, Sabine. 2024. Cashinahua DoReCo dataset. https://doi.org/10.34847/nkl.a8f9q2f1.
Dolgan	dolg1241	Turkic	Eurasia	Däbritz, Chris Lasse, Nina Kudryakova, Eugénie Stapert & Alexandre Arkhipov. 2024. Dolgan DoReCo dataset. https://doi.org/10.34847/nkl.f09eikq3.
Evenki	even1259	Tungusic	Eurasia	Kazakevich, Olga & Elena Klyachko. 2024. Evenki DoReCo dataset. https://doi.org/10.34847/nkl.5e0d27cu.

LANGUAGE	GLOTTOLOG	FAMILY	AREA	REFERENCE (ALL IN SEIFART ET AL. (EDS.) 2024)
Goemai	goem1240	Afro-Asiatic	Africa	Hellwig, Birgit. 2024. Goemai DoReCo dataset. https://doi.org/10.34847/nkl.b93664ml.
Gorwaa	goro1270	Afro-Asiatic	Africa	Harvey, Andrew. 2024. Gorwaa DoReCo dataset. https://doi.org/10.34847/nkl.a4b4ijj2.
Gurindji	guri1247	Pama-Nyungan	Australia	Meakins, Felicity. 2024. Gurindji DoReCo dataset. https://doi.org/10.34847/nkl.ab2d89mx.
Hoocąk	hoch1243	Siouan	North America	Hartmann, Iren. 2024. Hoocąk DoReCo dataset. https://doi.org/10.34847/nkl.b57f5065.
Jahai	jeha1242	Austro-Asiatic	Eurasia	Burenhult, Niclas. 2024. Jahai DoReCo dataset. https://doi.org/10.34847/nkl.6a71xp0p.
Jejuan	jeju1234	Koreanic	Eurasia	Kim, Soung-U. 2024. Jejuan DoReCo dataset. https://doi.org/10.34847/nkl.06ebrk38.
Kakabe	kaka1265	Mande	Africa	Vydrina, Alexandra. 2024. Kakabe DoReCo dataset. https://doi.org/10.34847/nkl.d5aeu9t6.
Kamas	kama1351	Uralic	Eurasia	Gusev, Valentin, Tiina Klooster, Beáta Wagner-Nagy & Alexandre Arkhipov. 2024. Kamas DoReCo dataset. https://doi.org/10.34847/nkl.cdd8177b.
Tabaq (Karko)	kark1256	Nubian	Africa	Hellwig, Birgit, Gertrud Schneider-Blum & Khaleel Bakheet Khaleel Ismail. 2024. Tabaq (Karko) DoReCo dataset. https://doi.org/10.34847/nkl.eea8144j.
Komnzo	komn1238	Yam	Papunesia	Döhler, Christian. 2024. Komnzo DoReCo dataset. https://doi.org/10.34847/nkl.c5e6dudv.
Light Warlpiri	ligh1234	Mixed Language	Australia	O'Shannessy, Carmel. 2024a. Light Warlpiri DoReCo dataset. https://doi.org/10.34847/nkl.7452803q.
Lower Sorbian	lowe1385	Indo-European	Eurasia	Bartels, Hauke & Marcin Szczepański. 2024. Lower Sorbian DoReCo dataset. https://doi.org/10.34847/nkl.6c6e4e9k.
Movima	movi1243	Isolate	South America	Haude, Katharina. 2024. Movima DoReCo dataset. https://doi.org/10.34847/nkl.da42xf67.
Dalabon	ngal1292	Gunwi-nyguan	Australia	Ponsonnet, Maïa. 2024. Dalabon DoReCo dataset. https://doi.org/10.34847/nkl.fae299ug.
Nisvai	nisv1234	Austro-	Papunesia	Aznar, Jocelyn. 2024. Nisvai DoReCo dataset. https://doi.org/10.34847/nkl.2801565f.
N‖ng	nngg1234	Tuu	Africa	Güldemann, Tom, Martina Ernszt, Sven Siegmund & Alena Witzlack-Makarevich. 2024. N‖ng DoReCo dataset. https://doi.org/10.34847/nkl.f6c37fi0.
Northern Kurdish (Kurmanji)	nort2641	Indo-European	Eurasia	Haig, Geoff, Maria Vollmer & Hanna Thiele. 2024. Northern Kurdish (Kurmanji) DoReCo dataset. https://doi.org/10.34847/nkl.ca10ez5t.
Northern Alta	nort2875	Austro-nesian	Papunesia	Garcia-Laguia, Alexandro. 2024. Northern Alta DoReCo dataset. https://doi.org/10.34847/nkl.efea0b36.
Fanbyak	orko1234	Austro-nesian	Papunesia	Franjieh, Michael. 2024. Fanbyak DoReCo dataset. https://doi.org/10.34847/nkl.02084446.
Pnar	pnar1238	Austro-Asiatic	Eurasia	Ring, Hiram. 2024. Pnar DoReCo dataset. https://doi.org/10.34847/nkl.5ba1062k.
Daakie	port1286	Austro-nesian	Papunesia	Krifka, Manfred. 2024. Daakie DoReCo dataset. https://doi.org/10.34847/nkl.efeav5l9.
Resígaro	resi1247	Arawakan	South America	Seifart, Frank. 2024b. Resígaro DoReCo dataset. https://doi.org/10.34847/nkl.ffb96lo8.

LANGUAGE	GLOTTOLOG	FAMILY	AREA	REFERENCE (ALL IN SEIFART ET AL. (EDS.) 2024)
Ruuli	ruul1235	Atlantic-Congo	Africa	Witzlack-Makarevich, Alena, Saudah Namyalo, Anatol Kiriggwajjo & Zarina Molochieva. 2024. Ruuli DoReCo dataset. https://doi.org/10.34847/nkl.fde4pp1u.
Sadu	sadu1234	Sino-Tibetan	Eurasia	Xu, Xianming & Bibo Bai. 2024. Sadu DoReCo dataset. https://doi.org/10.34847/nkl.3db4u59d.
Sanzhi Dargwa	sanz1248	Nakh-Daghestanian	Eurasia	Forker, Diana & Nils Norman Schiborr. 2024. Sanzhi Dargwa DoReCo dataset. https://doi.org/10.34847/nkl.81934177.
Savosavo	savo1255	Isolate	Papunesia	Wegener, Claudia. 2024. Savosavo DoReCo dataset. https://doi.org/10.34847/nkl.b74d1b33.
Nafsan (South Efate)	sout2856	Austro-nesian	Papunesia	Thieberger, Nick. 2024. Nafsan (South Efate) DoReCo dataset. https://doi.org/10.34847/nkl.ba4f760l.
English (Southern England)	sout3282	Indo-European	Eurasia	Schiborr, Nils Norman. 2024. English (Southern England) DoReCo dataset. https://doi.org/10.34847/nkl.9c271u5g.
French (Swiss)	stan1290	Indo-European	Eurasia	Avanzi, Mathieu, Marie-José Béguelin, Gilles Corminboeuf, Federica Diémoz & Laure Anne Johnsen. 2024. French (Swiss) DoReCo dataset. https://doi.org/10.34847/nkl.3520l685.
Sümi	sumi1235	Sino-Tibetan	Eurasia	Teo, Amos. 2024. Sümi DoReCo dataset. https://doi.org/10.34847/nkl.5ad4t01p.
Svan	svan1243	Kartvelian	Eurasia	Gippert, Jost. 2024. Svan DoReCo dataset. https://doi.org/10.34847/nkl.9ba054c3.
Tabasaran	taba1259	Nakh-Daghe-stanian	Eurasia	Bogomolova, Natalia, Dmitry Ganenkov & Nils Norman Schiborr. 2024. Tabasaran DoReCo dataset. https://doi.org/10.34847/nkl.ad7f97xr.
Teop	teop1238	Austro-nesian	Papunesia	Mosel, Ulrike. 2024. Teop DoReCo dataset. https://doi.org/10.34847/nkl.9322sdf2.
Texistepec Popoluca	texi1237	Mixe-Zoque	North America	Wichmann, Søren. 2024. Texistepec Popoluca DoReCo dataset. https://doi.org/10.34847/nkl.c50ck58f.
Totoli	toto1304	Austro-nesian	Papunesia	Bardají i Farré, Maria, Christoph Bracks, Claudia Leto, Datra Hasan, Sonja Riesberg, Winaro S. Alamudi & Nikolaus P. Himmelmann. 2024. Totoli DoReCo dataset. https://doi.org/10.34847/nkl.c8b6ei29.
Mojeño Trinitario	trin1278	Arawakan	South America	Rose, Françoise. 2024. Mojeño Trinitario DoReCo dataset. https://doi.org/10.34847/nkl.cbc3b4xr.
Asimjeeg Datooga	tsim1256	Nilotic	Africa	Griscom, Richard. 2024. Asimjeeg Datooga DoReCo dataset. https://doi.org/10.34847/nkl.f77c7m72.
Urum	urum1249	Turkic	Eurasia	Skopeteas, Stavros, Violeta Moisidi, Nutsa Tsetereli, Johanna Lorenz & Stefanie Schröter. 2024. Urum DoReCo dataset. https://doi.org/10.34847/nkl.ac166n10.
Vera'a	vera1241	Austro-nesian	Papunesia	Schnell, Stefan. 2024. Vera'a DoReCo dataset. https://doi.org/10.34847/nkl.3e2cu8c4.
Warlpiri	warl1254	Pama-Nyungan	Australia	O'Shannessy, Carmel. 2024b. Warlpiri DoReCo dataset. https://doi.org/10.34847/nkl.042dv614.
Yongning Na	yong1270	Sino-Tibetan	Eurasia	Michaud, Alexis. 2024. Yongning Na DoReCo dataset. https://doi.org/10.34847/nkl.abe65p95.
Yucatec Maya	yuca1254	Mayan	North America	Skopeteas, Stavros. 2024. Yucatec Maya DoReCo dataset. https://doi.org/10.34847/nkl.9cbb3619.
Yurakaré	yura1255	Isolate	South America	Gipper, Sonja & Jeremías Ballivián Torrico. 2024. Yurakaré DoReCo dataset. https://doi.org/10.34847/nkl.7ca412wg.

Temporal Clusters of Words in Conversations Depend on Genre and Frequency

Giuachin Kreiliger, Sebastian Krapp & Sabine Stoll
(University of Zürich)

Conversations are a blend of a few very frequent words, and a vast array of rare words. How do children accomplish learning these rare words?

Although frequency is the best predictor for learning, there is experimental evidence that temporal distribution is an important factor (Schwab and Lew-Williams 2016; Slone et al. 2023). In fact, linguistic units in adult conversations and child-directed speech are often temporally clustered (Altmann, Pierrehumbert, and Motter 2009; Katz 1996; Wojcik and Goulding 2025; Küntay and Slobin 1996; Lester et al. 2022; Waterfall 2006).

Here, we ask whether clustering in child–adult conversations partially compensates for the low frequency of occurrence, explaining why rare words can be learned by children despite their low frequency. In order to answer this question, we need to disentangle clustering from by-chance clustering because of frequency.

Available measures of temporal clustering, such as the dispersion measure in Gries (2008) and Wojcik and Goulding (2025), are correlated with frequency. The burstiness measure in Goh and Barabási (2008) and Slone et al. (2023) works well only for very long observation periods and large amounts of data.

To disentangle frequency from clustering, we introduce a new statistical model. We define appearance as a time-independent, and repetition as a time-dependent process.

The probability of a token to appear in a token-slot is the sum of both processes. There is a constant probability of appearance. Once the token appears, there is a certain probability of repetition.

With increasing distance to the last appearance or repetition, the probability of repetition decreases and eventually drops to the appearance probability. Here we use an exponential function to implement the decay of repetition probability (see Figure 1). The time-dependent repetition probability reflects the degree of temporal clustering independent of frequency.

We applied our model to two genres: adult–child and adult–adult conversations. We want to test whether clustering of rare words differs in the two genres. Our data come from the Manchester Child Language Corpus (Theakston et al. 2001) and the British National Corpus (BNC Consortium 2007).

For comparison, we used tokens that were observed in both genres and occurred at least 20 times. This left us with 801 tokens, 80% of them nouns, verbs, adverbs or adjectives.

We find that in both genres, raw frequency of a token and immediate repetition probability are negatively correlated. Rare tokens are hence in general more often immediately repeated (see Table 1 for regression analysis results).
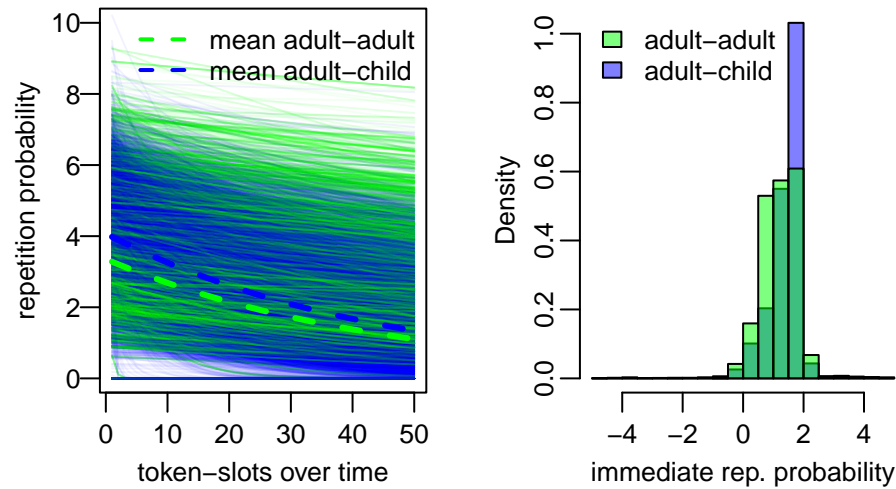
Figure 1: Repetition probabilities of 801 tokens after initial appearance. Left: Repetition probability decrease with increasing distance from initial appearance. Right: logarithm of immediate repetition probabilities at t=1.

We expected that child–adult conversations have a higher immediate repetition probability than adult–adult conversations, but we found that the immediate repetition probability in both genres is similar. We discuss this result in light of the child's contribution to the conversation.

Rare tokens are not immediately repeated more often in the child–adult genre. The general trend for higher immediate repetition probability in rare words compared to more frequent words might still explain how children can also learn rare words, since they are repeated more often, once they appear.

	Estimate	Q2.5	Q97.5
baseline adult–child	-0.51	-0.69	-0.34
adult–adult	0.02	-0.15	0.19
adult–child:log frequency	-0.17	-0.18	-0.16
adult–adult:log frequency	-0.02	-0.02	-0.01
word used by child from the start	0.17	0.14	0.19
word used by child at some point	0.11	0.09	0.13

Table 1: Regression coefficient mean and credible intervals of a model predicting the logarithm of immediate repetition probabilities. The ":" indicates an interaction of predictors. It can be seen that child-adult immediate repetition probability depends on whether the child uses the word as well (rows 5-6). Immediate repetition probability corresponds to the logarithm of the y-axis intercepts in Figure 1. The model was fit with the brms package in R (Bürkner 2021; R Core Team 2024)

References

Altmann, Eduardo G, Janet B Pierrehumbert, and Adilson E Motter (2009), Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words, *PLoS One* 4(11), e7678.

BNC Consortium (2007). *British National Corpus, version 3 BNC XML edition*. Oxford Text Archive. URL: http://hdl.handle.net/20.500.12024/2554.

Bürkner, Paul-Christian (2021), Bayesian Item Response Modeling in R with brms and Stan, *Journal of Statistical Software* 100(5), 1–54, DOI: DOI: 10.18637/jss.v100.i05.

Goh, K.-I. and A.-L. Barabási (2008), Burstiness and memory in complex systems, *Europhysics Letters* 81(4), 48002, DOI: DOI: 10.1209/0295-5075/81/48002.

Gries, Stefan Th. (2008), Dispersions and adjusted frequencies in corpora, *International Journal of Corpus Linguistics* 13(4), 403–437, DOI: DOI: 10.1075/ijcl.13.4.02gri.

Katz, Slava M. (1996), Distribution of content words and phrases in text and language modelling, *Natural Language Engineering* 2(1), 15–59, DOI: DOI: 10.1017/S1351324996001246.

Küntay, Aylin and Dan Isaac Slobin (1996), Listening to a Turkish mother: Some puzzles for acquisition. in (eds), (1996), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp.* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, 265–286.

Lester, Nicholas A., Steven Moran, Aylin C. Küntay, Shanley E. M. Allen, Barbara Pfeiler, and Sabine Stoll (2022), Detecting structured repetition in child-surrounding speech: Evidence from maximally diverse languages, *Cognition* 221, 104986, DOI: DOI: 10.1016/j.cognition.2021.104986.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Schwab, Jessica F. and Casey Lew-Williams (2016), Repetition Across Successive Sentences Facilitates Young Children's Word Learning, *Developmental Psychology* 52(6), 879–886, DOI: DOI: `10.1037/dev0000125`.

Slone, Lauren K., Drew H. Abney, Linda B. Smith, and Chen Yu (2023), The temporal structure of parent talk to toddlers about objects, *Cognition* 230, 105266, DOI: DOI: `10.1016/j.cognition.2022.105266`.

Theakston, Anna L., Elena V. M. Lieven, Julian M. Pine, and Caroline F. Rowland (2001), The role of performance limitations in the acquisition of verb-argument structure: An alternative account, *Journal of Child Language* 28(1), 127–152, DOI: DOI: `10.1017/S0305000900004608`.

Waterfall, Heidi (2006). "A little change is a good thing: feature theory, language acquisition and variation sets". PhD thesis. Chicago: University of Chicago.

Wojcik, Erica H. and Sarah J. Goulding (2025), Distribution of words across the first years of life: A longitudinal analysis of everyday language input to three English-learning infants, *Infancy* 30(1), e12622, DOI: DOI: `10.1111/infa.12622`.

Time guides referential choice in anaphora across diverse spoken languages

Guido M. Linders & Stefan Schnell
(University of Zurich)

Speakers are stipulated to keep their listeners in mind when producing referring expressions (Chafe, 1987, Clark and Murphy, 1982), deploying more verbose and informative forms for less accessible referents, and vice versa. Recency, i.e. the distance between a current referent mention and its antecedent mention, is an important co-determinant of referent accessibility (Ariel, 1990, Arnold, 2010), and has been shown to represent the single most important factor in determining full NP use across diverse languages (Kibrik et al., 2016, Linders and Schnell, 2025, Schiborr, 2023), as accessibility is assumed to *decay*. But what does it mean for accessibility to decay? How is antecedent distance to be measured? Traditional measures have been number of intervening clauses (Givón, 1983), referents — related to competition (Ariel, 1990) — or paragraphs ('unity' in Ariel, 1990), which have all been found to be of similar importance (see Same and van Deemter, 2020, for an overview). Most previous work is, however, based on written data, often from better studied languages like English, and it is thus unclear whether these results extend to the more common form of spoken discourse production, and to diverse languages (Evans and Levinson, 2009, Henrich et al., 2010). Crucially, one form of recency has thus never been considered systematically, namely *time*, which might be especially important in spoken language (Clark, 2002).

We here investigate spoken narratives from 8 typologically diverse languages, drawing on an overlap of production data from the Documentation Reference Corpus (DoReCo; Seifart et al., 2022),
and the Multilingual Corpus of Annotated Spoken Texts (Multi-CAST; Haig and Schnell, 2022), which together contain detailed segment-level time-aligned annotations as well as morphosyntactic annotations, including clause boundaries and syntactic functions (GRAID; Haig and Schnell, 2014) and referent tracking (RefIND; Schiborr et al., 2018). Comparing the number of syllables, referents, words, and clauses, and time (seconds), we show that including each measure separately into a baseline model with syntactic and antecedent function as predictors, generally improves the prediction accuracy over that baseline model and over a naive model that always predicts the most frequent referential form. Interestingly, time predicts referential choice equally well as other measures in terms of linguistic units across languages.

While we find all measures to be highly correlated with one another, our findings suggest that recency does not reduce to a single overriding factor and that time — as an entirely non-linguistic notion — does seem as relevant as any linguistic unit measure for speakers in their estimates of listeners to successfully retrieve an intended referent. This suggests that time itself may be an important aspect of speech planning in reference production.

References

Ariel, M. (1990). Accessing noun-phrase antecedents. Routledge. https://doi.org/10.4324/9781315857473

Arnold, J. E. (2010). How speakers refer: The role of accessibility. Language and Linguistics Compass, 4 (4), 187–203. https://doi.org/10.1111/j.1749-818X.2010.00193.x

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), Coherence and grounding in discourse (pp. 21–51). John Benjamins. https://doi.org/10.1075/tsl.11.03cha

Clark, H. H. (2002). Speaking in time. Speech Communication, 36 (1-2), 5–13. https://doi.org/10.1016/S0167-6393(01)00022-X

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. Advances in Psychology, 9, 287–299. https://doi.org/10.1016/S0166-4115(09)60059-5

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32 (5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Giv´on (Ed.), Topic continuity in discourse: A quantitative cross-language study (pp. 1–42, Vol. 3). John Benjamins. https://doi.org/10.1075/tsl.3.01giv

Haig, G., & Schnell, S. (2014). Annotations using GRAID [Version 7]. University of Bamberg. https://multicast.aspra.uni-bamberg.de/data/pubs/graid/Haig+Schnell2014_GRAID-manual_v7.0.pdf

Haig, G., & Schnell, S. (2022). Multi-CAST: Multilingual corpus of annotated spoken texts [Version 2211]. University of Bamberg. https://multicast.aspra.uni-bamberg.de/

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33 (2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Kibrik, A. A., Khudyakova, M. V., Dobrov, G. B., Linnik, A., & Zalmanov, D. A. (2016). Referential choice: Predictability and its limits. Frontiers in Psychology, 7, 1429. https://doi.org/10.3389/fpsyg.2016.01429

Linders, G. M., & Schnell, S. (2025). Functional factors predict referential choice similarly across languages: A cross-linguistic computational analysis. *PsyArXiv*. https://doi.org/10.31234/osf.io/23p4m_v1

Same, F., & van Deemter, K. (2020). Computational interpretations of recency for the choice of referring expressions in discourse. In C. Braud, C. Hardmeier, J. J. Li, A. Louis, & M. Strube (Eds.), Proceedings of the First Workshop on Computational Approaches to Discourse (pp. 113–123). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.codi-1.12

Schiborr, N. N. (2023). Lexical anaphora: A corpus-based typological study of referential choice. University of Bamberg Press.

Schiborr, N. N., Schnell, S., & Thiele, H. (2018). RefIND – Referent indexing in natural-language discourse [Version 1.1]. University of Bamberg. https://multicast.aspra.uni-bamberg.de/data/pubs/refind/Schiborr+etal2018_RefIND-guidelines_v1.1.pdf

Seifart, F., Paschen, L., & Stave, M. (2022). Language Documentation Reference Corpus (DoReCo) 2.0. Laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2). https://doi.org/10.34847/nkl.7cbfq779

When to ask and when to answer – studying the development of turn-taking in a Balinese sign language

Turn-taking is one of the fundamental skills children need to learn to become competent conversationalists. Across different spoken and signed languages, adults take turns effortlessly, monitoring their interlocutors' turn while planning and carefully timing their own response to achieve minimal gaps and overlaps between turns (Stivers et al. 2009, De Vos et al. 2015). Children's structure and timing of turns gradually develop alongside their linguistic and pragmatic skills (Casillas et al. 2016). Before age 6, response latencies are common; while adult interlocutors may manage slow responses, these delays may cause non-coordinated turns and misunderstandings when there are more than two conversation partners or in peer-to-peer conversations (Ervin-Tripp, 1979).

I investigate the development of turn-taking in Kata Kolok, the indigenous sign language of a single village in Bali, Indonesia, where multi-party conversations are the norm. Two key aspects of this signing community make it unusual. First, sustained high incidences of deafness have led to six generations of deaf and hearing signers, which provides children access to a **diversity** of signing interlocutors (Marsaja 2008). Second, communal child-rearing practices mean that most conversations available to children involve **multiple** signing interlocutors.

I analyse longitudinal corpus data from four deaf children from the Kata Kolok Child Signing Corpus, a collection of naturalistic recordings of children who acquire Kata Kolok from their deaf caregivers. Using conversation analysis, I compare structure and timing of 200 question-answer sequences from conversations among adults and 1,200 question-answer sequences with a child interlocutor (i.e., 300 questions per children distributed across 10 longitudinal recordings between 1-3 years old). For structure, I focus on question/response type, turn complexity and interactional context and for timing, I focus on gaps and overlap.

The annotations are ongoing and findings are preliminary. Initial data suggest four major observations:

1) Complexity of questions and responses gradually increases, proceeding from short to longer, information-seeking questions and responses.
2) Young children are often prompted with short polar questions that are embedded in familiar routines, such as EAT FINISH ('have you eaten'). These are repeated (over time and within the same interactive sequence) and deeply culturally entrenched, allowing children to hone the structure and timing of turns.
3) Possibly due to challenges of coordinating and timing visual attention, young children only gradually partake in multiparty settings, and if so, the conversations often involve peers and/or few adult interlocutors. Indeed, within medium-sized to large multiparty settings, children often are passive participants and/or initiate dyadic conversations.
4) The timing of question-answer sequences is influenced by conversational setting and age: conversations with younger children as well as among peers are generally slower (i.e. longer gaps between turns) with less overlap than when adult interlocutors are involved.

This study will present critical insights into how signing children acquire turn-taking skills. As one of the first ever to examine language acquisition in a small, rural sign language from the Global South, it significantly broadens our understanding language acquisition with insights from a community-centered culture.

References:

Casillas, M., Bobb, S., & Clark, E. (2016). Turn-taking, timing, and planning in early language acquisition. Journal of Child Language, 43(6), 1310–1337. https://doi:10.1017/S0305000915000689.

De Vos, C., Torreira, F., & Levinson, S. C. (2015). Turn-timing in signed conversations: Coordinating stroke-to-stroke turn boundaries. Frontiers in Psychology, 6. https://doi.org/10.3389/fpsyg.2015.00268

Ervin-Tripp, S. (1979). Children's verbal turn-taking. In E. Ochs & B. Schieffelin (eds), Developmental pragmatics, 391–414. New York: Academic Press.

Kata Kolok Child Signing Corpus (KKCSC) (2007-2024). The Language Archive. https://hdl.handle.net/1839/4d5f83b3-75b8-4dc3-bc16-3a398a37cf77.

Marsaja, I Gede. *Desa Kolok: A Deaf Village and Its Sign Language in Bali, Indonesia*. Nijmegen: Ishara Press, 2008.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. Proceedings of the National Academy of Sciences, 106(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

Intonation units form low frequency rhythms

Maya Inbar, Eitan Grossman & Ayelet N. Landau

The Hebrew University of Jerusalem; The Hebrew University of Jerusalem;
& The Hebrew University of Jerusalem and University College London

The foundations of the language system and its biological encoding remain fundamental questions in linguistics and neuroscience. Developing theories that account for the immense diversity of the world's languages presents a significant challenge (Evans & Levinson, 2009). Recent research focusing on the unfolding of speech in time has offered reconciliation to apparent diversities across languages (e.g., Coupé et al., 2019). In addition, the study of temporal structure in speech reveals meaningful interactions with temporally structured brain activity (e.g., Kaufeld et al., 2020; Chalas et al., 2024). This has inspired new models of speech processing and improved our understanding of the neural mechanisms underlying language function (Giraud & Poeppel, 2012). Despite the inherently dynamic nature of communication, few studies have thoroughly characterized different temporal structures in speech and even fewer have done so from a broad cross-linguistic perspective. In this study we systematically characterize the temporal structure of *intonation units* (IUs) across a genealogically and geographically diverse sample of the world's languages, and investigate the relationship between IU rate and syllable rate.

IUs are a universal building-block of human speech. IUs are defined by a set of auditory cues, including changes in syllable delivery rate, resets in pitch level, resets in volume and pauses. They are found cross-linguistically (Himmelmann et al., 2018) and are tied to important language functions such as the pacing of information in discourse (Chafe, 1994) and swift turn-taking (Ford & Thompson, 1996; Bögels & Torreira, 2015, 2021). We first introduce and validate an algorithm to annotate IUs cross-linguistically, thus solving a bottleneck in the field. Subsequently, we build on the substantial efforts of language documenters and the DoReCo corpus creators (Paschen et al., 2020; Seifart et al., 2022) and explicitly model the rate of IUs in spontaneous speech in 48 diverse languages from every continent and from 27 distinct language families. **Our key finding is that people spontaneously produce IUs at a similar low-frequency rate of 0.6-1 Hz, with minimal variation across demographics and life stages.**

We next relate these findings on IU rate to another measure of speech rate, syllable rate, which has already benefited from elegant cross-linguistic studies (Coupé et al., 2019; Ozaki et al., 2024). Specifically, it has been shown that speakers systematically balance their syllable delivery rate with language-specific informativeness properties of syllables, leading to substantial variation between languages in average syllable rate (Pellegrino et al., 2011; Coupé et al., 2019). We find that **syllable rate is only weakly related to the delivery rate of a unit hierarchically above the syllable – IUs, and no evidence that cross-linguistic variation in syllable rate accounts for cross-linguistic variation in IU rate.** These findings suggest that IUs fulfill a role in the pacing of language that is different from the role fulfilled by syllables. Syllables are low-level building blocks that are highly affected by the idiosyncrasies of articulation, while IUs are conceived of as "planning units" (Du Bois, 1987; Chafe, 1994; Croft, 1995), whose temporal structure possibly relies on cognitive mechanisms of attention and memory.

References

Bögels, S., & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics*, *52*, 46–57. https://doi.org/10.1016/j.wocn.2015.04.004

Bögels, S., & Torreira, F. (2021). Turn-end Estimation in Conversational Turn-taking: The Roles of Context and Prosody. *Discourse Processes*, *58*(10), 903–924. https://doi.org/10.1080/0163853X.2021.1986664

Chafe, W. (1994). *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press.

Chalas, N., Meyer, L., Lo, C.-W., Park, H., Kluger, D. S., Abbasi, O., Kayser, C., Nitsch, R., & Gross, J. (2024). Dissociating prosodic from syntactic delta activity during natural speech comprehension. *Current Biology*, *0*(0), 1–13. https://doi.org/10.1016/j.cub.2024.06.072

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, *5*(9). https://doi.org/10.1126/sciadv.aaw2594

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, *33*(5), 839–882. https://doi.org/10.1515/ling.1995.33.5.839

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, *63*(4), 805–855.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 134–184). Cambridge University Press. https://doi.org/10.1017/CBO9780511620874.003

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Himmelmann, N. P., Sandler, M., Strunk, J., & Unterladstetter, V. (2018). On the universality of intonational phrases: A cross-linguistic interrater study. *Phonology*, *35*(2), 207–245. https://doi.org/10.1017/S0952675718000039

Kaufeld, G., Bosker, H. R., ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, *40*(49), 9467–9475. https://doi.org/10.1523/JNEUROSCI.0302-20.2020

Ozaki, Y., Tierney, A., Pfordresher, P. Q., McBride, J. M., Benetos, E., Proutskova, P., Chiba, G., Liu, F., Jacoby, N., Purdy, S. C., Opondo, P., Fitch, W. T., Hegde, S., Rocamora, M., Thorne, R., Nweke, F., Sadaphal, D. P., Sadaphal, P. M., Hadavi, S., … Savage, P. E. (2024). Globally, songs and instrumental melodies are slower and higher and use more stable pitches than speech: A Registered Report. *Science Advances*, *10*(20). https://doi.org/10.1126/sciadv.adm9797

Paschen, L., Delafontaine, F., Draxler, C., Fuchs, S., Stave, M., & Seifart, F. (2020). Building a time-aligned cross-linguistic reference corpus from language documentation data (DoReCo). *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings*, *May*, 2657–2666.

Pellegrino, F., Coupé, C., & Marisco, E. (2011). A Cross-Language Perspecitve on Speech Information Rate. *Language*, *87*(3), 539–558. https://doi.org/10.1353/lan.2011.0057

Seifart, F., Paschen, L., & Stave, M. (2022). *Language Documentation Reference Corpus (DoReCo) 1.2* (F. Seifart, L. Paschen, & M. Stave (eds.)). Leibniz-Zentrum Allgemeine Sprachwissenschaft & laboratoire Dynamique Du Langage (UMR5596, CNRS & Université Lyon 2). https://doi.org/10.34847/nkl.7cbfq779

Temporal dynamics of the event network

Research into information in interaction has traditionally centred on the verbalisation of entities. Highly active referents in discourse are typically realised with shorter forms or not at all, while unpredictable referents are often realised with more segmental material, such as with full lexical phrases (e.g. Ariel 1990; Haig, Schnell, and Schiborr 2021; Mahowald et al. 2013). Little attention has been paid, however, to the variable realisation of events over time. While it is assumed that the verbalisation of events follows the same rules as for entities, I present evidence in this talk which suggests the opposite. Using corpus data annotated for event realisations, I explore forms associated with repeated realisation of the same event units in discourse and show that we cannot account for verbalisation of events in the same way as we do entities.

Events and entities are typically realised by referents and predicates. While referents prototypically construe concrete entities in the physical environment, predicates tend to be used for the verbalisation of events, which are inherently abstract, spatiotemporal, and relational (Langacker 1987). This results in perception of entities as "persistent" in discourse, compared to the more "transient" events (Chafe 1994, 66–68). Despite the difference in the temporal conceptualisation of events and entities, little attention has been paid to how this may affect their verbalisation over the course of an interaction.

Results from corpus data from three Trans-Himalayan languages from Northeast India show that the verbalisation of events does not follow the same pattern as that of entities. In other words, the forms used to verbalise events do not appear to be clearly linked to information status. For example, the first realisation of an event may be simpler and a later recapitulation of the event may be more complex (1–2). Additionally, instances of recapitulation strategies use the same exact forms for both realisations, despite the event being highly activated (3). This suggests that we cannot argue for a direct link between givenness and event verbalisation, contrary to expectations.

In addition, I will present initial findings from dynamic analyses of the event networks instantiated in the corpus texts. These analyses will investigate the verbalisation of events over time by using information about their temporal and interactional contexts. This novel approach simulates the process of language comprehension as it pertains to the event network (Zwaan and Radvansky 1998), giving us a sense of how activation within an event network may affect the verbalisation of events for the first time.

Accounting for the temporal dynamics of verbalisation requires more than just entities and referents. We must incorporate events and their unique properties to understand how we choose to realise information over time. Taking time seriously requires taking both interactional and cognitive factors seriously too.

1

Examples

(1) Galo: Lost Writing 59

joomb ekku bee dɨɨ? **dokaa ku***!*

joom=bə eK-ku-bee dɨɨ do-kaa-ku
how=DAT write-CMPL-EX.PRF WOND eat-PRF-CMPL

'How would they have been written after all indeed? **They ate it (the writing) up**!'

(2) Galo: Lost Writing 73

ŋunu… əgəm, **dopak tokʷəəmə** *[…]*

ngunu əgə-m do-pak-to-ku=əəm=əə
1PL ANA-ACC eat-rid-PFV-CMPL=ACC=TOP

'We… **because of having eaten it (the writing)**…'

(3) Duhumbi: Butterfly 11-14

mej **dʑakleda dʑubale***. mej* **dʑakleda dʑubakʰona***…*

mej dʑak-le-da dʑu-ba=le mej dʑak-le-da dʑu-ba-kho=na
bamboo wait-do-NF stay-NMLZ=COP.LE bamboo wait-do-NF stay-NMLZ-LOC=CFOC

'…(that old man) **stayed, taking a rest after carrying the bamboo**. **When staying taking a rest after carrying the bamboo**…'

References

Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.

Chafe, Wallace. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Haig, Geoffrey, Stefan Schnell, and Nils N. Schiborr. 2021. "Universals of Reference in Discourse and Grammar: Evidence from the Multi-CAST Collection of Spoken Corpora." *Language Documentation & Conservation*, Special Publication 25, 25: 141–77.

Langacker, Ronald W. 1987. "Nouns and Verbs." *Language* 63 (1): 53–94. https://doi.org/10/fjvs5n.

Mahowald, Kyle, Evelina Fedorenko, Steven T. Piantadosi, and Edward Gibson. 2013. "Info/Information Theory: Speakers Choose Shorter Words in Predictive Contexts." *Cognition* 126 (2): 313–18. https://doi.org/10.1016/j.cognition.2012.09.010.

Zwaan, Rolf A., and Gabriel A. Radvansky. 1998. "Situation Models in Language Comprehension and Memory." *Psychological Bulletin* 123 (2): 162–85.

2

Language-invariant patterns of using coordination markers for joint action navigation

Natalia Morozova, Sabine Stoll, & Adrian Bangerter

(Institute for the Interdisciplinary Study of Language Evolution, University of Zurich; Institute for the Interdisciplinary Study of Language Evolution, University of Zurich; & Institute of Work and Organizational Psychology (IPTO), University of Neuchâtel)

Keywords: joint action coordination, cognition in interaction, dual-tasking constraints, convergent language evolution

Humans conceptualise goal-directed activities (e.g. *building LEGO models together*) as hierarchies of larger tasks (*model one, model two*) and their smaller sub-tasks (*block one, block two*, etc; Zacks & Tversky, 2001). Navigating this hierarchy in joint activities implies that participants must agree on whether they elaborate on a specific sub-task (**horizontal transition**) or switch to the next sub-task (**vertical transition**; Bangerter & Clark, 2003). Previous research indicates that participants employ coordination markers such as *yeah, uh-huh, okay* to distinguish horizontal and vertical transitions acoustically (Morozova et al., 2024) and lexically, e.g. in English, with horizontal markers *yeah* and *uh-huh* and vertical markers *okay*, *all right* (Bangerter & Clark, 2003). Given that navigating horizontal and vertical transitions poses a generic coordination problem, these findings raise the questions of (1) whether speakers of other languages use similar types of words as specialised coordination devices and (2) which properties make some markers more suitable for horizontal vs. vertical transition contexts.

We hypothesised that horizontal and vertical transitions exert different functional pressures on the lexicality of coordination markers. As horizontal markers are produced in ongoing sub-tasks, the timing of their production co-occurs with simultaneous contributions to the non-linguistic task (e.g. model building), whereas vertical markers are deployed between sub-tasks and do not entail simultaneous contributions to the joint task. We thus predicted that participants might mitigate higher dual-tasking constraints by producing less lexicalised markers (e.g. *uh-huh, mm*) or verbatim repetitions in horizontal transitions.

We analysed the forms and functions of coordination markers in three typologically diverse languages (Swiss French, Vietnamese, and Shipibo-Konibo) with a cooperative game experiment (Clark & Krych, 2004). We tested 232 adults ($N_{French} = 78$, $N_{Vietnamese} = 80$, $N_{Shipibo-Konibo} = 74$) who constructed 10 LEGO models in dyads of directors and builders. The task involved navigating horizontal transitions through block sub-tasks, e.g. identifying and placing *block two*, and vertical transitions between blocks, e.g. switching from *block two* to *block three*. We manually extracted coordination markers ($N_{French} = 10,740$, $N_{Vietnamese} = 9,046$, $N_{Shipibo-Konibo} = 11,353$) and analysed their distributions between two transitions with Bayesian mixed-effects models.

We found that in each language, certain forms of coordination markers were consistently used as horizontal markers, while others as vertical, comprising an implicit system of transition-specific conventions. With regard to the lexicality of coordination markers, we found strong evidence suggesting that participants strategically deployed non-lexical markers (*uh-huh, mm*) and repetitions in horizontal transitions, while lexical markers (e.g. *yeah, okay, perfect*) in vertical transitions.

Our findings suggest that (1) coordination markers are potentially universal devices for navigating joint activities and (2) the forms of coordination markers could be shaped by the constraints and affordances of their interaction contexts. Specifically, in horizontal transitions, participants might prioritise maintaining attention on the task at hand by using non-disruptive markers that are easier to plan and produce (e.g. Knudsen et al., 2020; Bartolozzi et al., 2021; Dingemanse et al., 2022), while in vertical transitions, they could afford producing markers with more lexical meaning and higher incipient speakership (Jefferson, 1984; Drummond & Hopper, 1993) and acoustically less constrained than in horizontal transitions (Morozova et al., 2024). Our study provides new evidence of how different timing constraints in interactions might shape language use, possibly through the selective forces of convergent language evolution (Croft, 2000; Christiansen & Chater, 2008; Enfield, 2008; Dingemanse et al., 2013).

References

Bangerter, A. and H. H. Clark (2003), Navigating joint projects with dialogue, *Cognitive Science* 27(2), 195–225.

Bartolozzi, F., S. R. Jongman, and A. S. Meyer (2021), Concurrent speech planning does not eliminate repetition priming from spoken words: Evidence from linguistic dual-tasking, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 47(3), 466–480.

Christiansen, M. H. and N. Chater (2008), Language as shaped by the brain, *Behavioral and brain sciences* 31(5), 489–509.

Clark, H. H. and M. A. Krych (2004), Speaking while monitoring addressees for understanding, *Journal of Memory and Language* 50(1), 62–81.

Croft, W. (2000), *Explaining language change: An evolutionary approach*: Pearson Education.

Dingemanse, M., A. Liesenfeld, and M. Woensdregt (2022). "Convergent cultural evolution of continuers (mmhm)". In: *Joint Conference on Language Evolution (JCoLE)*, 61–67.

Dingemanse, M., F. Torreira, and N. J. Enfield (2013), Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items, *PloS ONE* 8(11), e78273.

Drummond, K. and R. Hopper (1993), Back Channels Revisited: Acknowledgment Tokens and Speakership Incipiency, *Research on Language and Social Interaction* 26(2), 157–177.

Enfield, N. J. (2008), Language as shaped by social interaction, *Behavioral and brain sciences* 31(5), 519–520.

Jefferson, G. (1984), Notes on a systematic deployment of the acknowledgement tokens "Yeah" and "Mm hm", *Papers in Linguistics* 17, 197–206.

Knudsen, B., A. Creemers, and A. S. Meyer (2020), Forgotten little words: How backchannels and particles may facilitate speech planning in conversation?, *Frontiers in Psychology* 11, 593671.

Zacks, J. M and B. Tversky (2001), Event structure in perception and conception, *Psychological Bulletin* 127(1), 3–21.

WS19 The diachrony of language geography: linking small-scale and large-scale perspective

Matthias Urban

# Two Melanesian pathways to linguistic diversity: West Papua vs. Vanuatu

Alexandre François & Daniel Krauße

(Lattice—CNRS & École Normale Supérieure)

Compared with the Polynesian triangle, an immense region with only 16 languages (Pawley 1967:260), Melanesia is linguistically much richer: it is home to >1300 languages (Landweer & Unseth 2012:2) – one fifth of all the world's – and includes the three countries with the highest linguistic diversity (UNESCO 2009:307).

But is Melanesia's wealth of languages of the same nature everywhere? Within this vast region [Map], our presentation will zoom in on two areas of our expertise: 1) the Bird's Head peninsula [BH] in West Papua (Indonesia); 2) the Torres & Banks islands [TB] in Vanuatu. While both zones are linguistically diverse, we will show that this diversity takes quite different forms there.

Map — Location of the two target areas within Melanesia
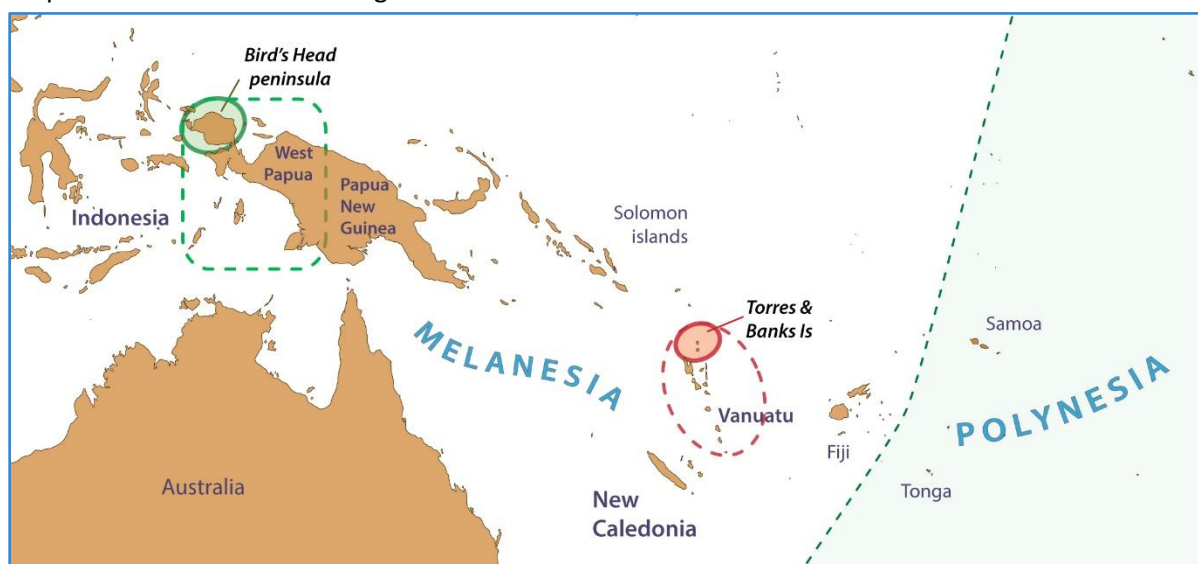
The BH area is home to 24 languages (Holten & Klamer 2018:570). For an indigenous population of just over 0.1M (Miedema & Reesink 2004), this is not extremely dense: the cliché whereby walking in New Guinea brings about new languages at every corner does not hold true here, as a language territorsy there can cover up to 5000 km². What is most striking in the BH is thus not its density, but its diversity, especially if defined in phylogenetic terms: its 24 languages belong to nine families (Hammarström et al. 2024). Recent counts list 80 different phyla for New Guinea (Palmer 2018:6), making it the most phylogenetically diverse island in the world; with its concentration of nine families, the BH region is a radical version of this diversity.

The Vanuatu situation is different, as all its languages belong to the same Oceanic subgroup of Austronesian. The country thus clearly lacks the phylogenetic diversity found in West Papua. However, with 138 languages for 0.3M people, it shows the world's highest linguistic density, defined in terms of languages per capita. With 17 languages for 9,400 people (François 2012), the TB area in the north

is a radical version of Vanuatu's linguistic density. The latter case reminds us that diversity can also develop over time, through a process of internal diversification.

Our talk will compare the two pathways by which Melanesian societies have brought about the linguistic mosaic we see today: 1) inherited linguistic diversity in West Papua; and 2) emergent linguistic diversity in Vanuatu.

The many families found in the BH can be explained by the historical depth of human presence in the region, going back >24 millennia (Pasveer et al. 2002:95). By contrast, the diversity observed in Vanuatu is innovated. It results from powerful trends of divergence, linked with a particular language ecology (François 2011, 2012): one where social ideologies promote small-size egalitarian communities, each endowed with its own emblematic language. In both cases, the final outcome is a linguistic mosaic — but with distinct characteristics, shaped by two different histories.

References

François, Alexandre. 2011. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics 1*(2), pp. 175–246.

—— 2012. The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. In L. Landweer & P. Unseth (eds.), pp. 85–110.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. *Glottolog 5.1.* Leipzig: Max Planck Institute for Evolutionary Anthropology.

Holten, Gary & Marian Klamer. 2018. The Papuan languages of East Nusantara and the Bird's Head. In B. Palmer (ed.), pp. 569–640.

Landweer, Lynn & Peter Unseth (eds.). 2012. *Language use in Melanesia*. Special issue of *International Journal of the Sociology of Language*, 214.

Miedema, Jelle & Ger Reesink. 2004. *One head, many faces: New perspectives on the Bird's Head Peninsula of New Guinea.* Leiden: KITLV Press.

Palmer, Bill (ed.). 2018. *The languages and linguistics of the New Guinea area: A comprehensive guide* (The World of Linguistics 4). Berlin: DeGruyter Mouton.

Pasveer, Juliette, Simon Clarke & Gifford Miller. 2002. Late Pleistocene human occupation of inland rainforest, Bird's Head, Papua. *Archaeology in Oceania 37*(2), pp. 92-95.

Pawley, Andrew. 1967. The relationships of Polynesian Outlier languages. *Journal of the Polynesian Society 76*(3), pp. 259–296.

UNESCO. 2009. *UNESCO World Report: Investing in Cultural Diversity and Intercultural Dialogue*. Paris: UNESCO.

Tibeto-Burman dispersal and linguistic diversity in Northeast India

Bingcong Deng
(Max Planck Institute of Geoanthropology)

Northeast India is a geographical region with great language diversity, both in the sense of its high number of spoken languages and phylogenetic diversity (the number of different language groups descending from a same ancestor). This region contains languages from at least four families: Tibeto-Burman, Austroasiatic, Tai-Kadai and Indo-European (Post & Burling, 2017). Although Tibeto-Burman languages in this region have been explored previously (e.g. Burling, 2007; Haokip, 2011; Van Driem, 2012), this paper takes a fresh look on this topic, incorporating multidiscplinary data to investigate the Tibeto-Burman dispersal and its role in language contact in Northeast India. Employing findings from lingusitics, archaeology and paleoclimate, this study aims to investigate the human-environment-language interaction with a focus on Tibeto-Burman speakers in Northeast India.

In particular, this paper focuses four Tibeto-Burman: Kho-Bwa, Tani, Sal and Kuki-Karbi. This research also examines the language contact of neighbouring languages with Tibeto-Burman as an effect of its dispersal, in addition to a comparison of the languages contact phenomenon exhibited by Sinitic languages and their neighbours.

This leads to the following research questions. First, what were the dispersal routes of the Tibeto-Burman groups in discussion of this research? What do they suggest about the Tibeto-Burman dispersal in general? Second, what were the pushing factors behind the Tibeto-Burman dispersal? Third, what is the evidence of early language contact between Tibeto-Burman languages and their neighbours in Northeast India? Lastly, what inferences can be drawn about the environment-human-language interactions in prehistoric Northeast India?

In order to answer these questions, this research employs a multidisciplinary method, drawing evidence from linguistics, archaeology and paleoclimatic studies. A database including lingusitic and archaeological findings was first collected, including the reconstructed vocabulary in the Tibeto-Burman groups under investigation in this study and the material culture in archaeological sites of this region. As this will yield a large dataset, this study chooses to focus on domains of agriculture, floral and fauna, which are indicative of the relationship of human and ecology. Then this database was put into the climatic data during the time period relevant to this study (ca. 5000-2000 BP) to draw inferences of the connection of human migration and its environmental background.

The preliminary results suggest that the speakers of the Tibeto-Burman branches investigated in this research most likely had taken the route from Northeast Sichuan, crossing the river valleys in Yunnan, and reaching their present-day locations. Possibly correlated with the spread of rice and millet farming package introduced from Sichuan into Yunnan around 2500 BC (Dal Martello et al., 2018), the migration might have been facilitated by the cooler and dryer environment (Xiao et al., 2018; Xue et al., 2022) and the demand for risk-resistant farming economy. The language contact between Tibeto-Burman languages and their neighbours did not limit to lexical borrowings but extended to moderate structural influence (e.g. an early phenomenon of metatypy of person-marking). In comparison to the Altaic influence on Sinitic revealed at a later period (e.g. Hashimoto, 1986; LaPolla, 2001), the contact phenomenon in the south implies a situation involving more intense contact with a less obvious dominant language in the region.

References

Burling, R. (2007). Language, Ethnicity and Migration in North-Eastern India. *South Asia: Journal of South Asian Studies*, *30*(3), 391–404. https://doi.org/10.1080/00856400701714039

Dal Martello, R., Min, R., Stevens, C., Higham, C., Higham, T., Qin, L., & Fuller, D. Q. (2018). Early agriculture at the crossroads of China and Southeast Asia: Archaeobotanical evidence and radiocarbon dates from Baiyangcun, Yunnan. *Journal of Archaeological Science: Reports*, *20*, 711–721. https://doi.org/10.1016/j.jasrep.2018.06.005

Haokip, P. (2011). *Socio-linguistic situation in north-east india*. Concept Publishing Company.

Hashimoto, M. J. (1986). The Altaicization of Northern Chinese. *Contributions to Sino-Tibetan Studies*, *5*, 76. https://doi.org/10.1163/9789004655409_004

LaPolla, R. J. (2001). The Role of Migration and Language Contact in the Development of the SIno-Tibetan Language Family. In A. I. Aïkhenval'd & R. M. W. Dixon (Eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford University Press.

Post, M., & Burling, R. (2017). The Tibeto-Burman languages of Northeast India. In *The Sino-Tibetan Languages* (p. 242). Routledge London.

Van Driem, G. (2012). Glimpses of the Ethnolinguistic Prehistory of Northeastern India. In T. Huber & S. H. Blackburn (Eds.), *Origins and migrations in the extended eastern Himalayas*. Brill.

Xiao, X., Haberle, S. G., Li, Y., Liu, E., Shen, J., Zhang, E., Yin, J., & Wang, S. (2018). Evidence of Holocene climatic change and human impact in northwestern Yunnan Province: High-resolution pollen and charcoal records from Chenghai Lake, southwestern China. *The Holocene*, *28*(1), 127–139. https://doi.org/10.1177/0959683617715692

Xue, J., Wang, W., Shao, J., Dai, X., Sun, Z., Gardner, J. D., Chen, L., Guo, X., Di, N., Pei, X., Wu, X., Zhang, G., Cui, C., Cao, P., Liu, F., Dai, Q., Feng, X., Yang, R., Ping, W., … Fu, Q. (2022). Ancient Mitogenomes Reveal the Origins and Genetic Structure of the

Neolithic Shimao Population in Northern China. *Frontiers in Genetics*, *13*, 909267. https://doi.org/10.3389/fgene.2022.909267

The linguistic diversity of Northern Asia:

the impact of environment, demography, and sociocultural factors

Brigitte Pakendorf[1], Olesya Khanina[2], Valentin Gusev[3]

[1]CNRS & Université Lumière Lyon 2, Laboratoire Dynamique Du Langage, 69007 Lyon, France

[2]Department of Finnish, Finno-Ugrian and Scandinavian Studies, University of Helsinki, Finland

[3]Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia

On the linguistic map of the world, Northern Asia stands out by its low linguistic diversity, both in numbers of language families and of languages. Furthermore, large areas are occupied by individual languages with unexpectedly little dialectal variety. Typological diversity, too, is fairly restricted, with the Uralic, Turkic, Mongolic, and Tungusic languages sharing cross-linguistically common features such as vowel harmony, verb-final word order, agglutinative morphology, large case systems, and subordination by non-finite verbs (Anderson 2006). Only a few small language families along the periphery, commonly grouped under "Paleosiberian", stand out typologically.

In this talk we will address the question of how these patterns of diversity came about by bringing together insights from a variety of disciplines. We show that demographic and sociocultural factors, both ultimately conditioned by the environment, shaped the patterns.

The climate in Northern Asia is extremely harsh, with long cold winters and short summers. Permafrost soils cover most of its eastern ranges, where the carrying capacity for riverine fish, for example, is eightfold lower than in Western Siberia (Skakun et al. 2016). The differences in permafrost range and food resources between Western and Eastern Siberia match differences in numbers of languages, suggesting a potential causal relationship. The low primary productivity also precludes intensive food production, resulting in extremely low population densities (cf. Yegorov 2016: Fig.4).

In such an environment, survival depends on extensive social networks, resulting in shared cultural codes along thousands of kilometres, e.g. ritual practices, folklore, importance of silence and non-verbal teaching, etc. (e.g. Dobzhanskaya 2008, Laptander 2020). These codes also include linguistic ideologies that value competence in the languages of one's neighbours over maintenance of one's patrimonial language (e.g. Khanina 2021). This sociocultural context is particularly beneficial to linguistic convergence, which in turn decreases typological diversity.

Furthermore, as demonstrated by genetic studies, the region has been home to repeated waves of migrations: although modern human occupation is known from the Upper Paleolithic (Fu et al. 2014, Raghavan et al. 2014, Sikora et al. 2019), a first wave of replacement is detected after the Last Glacial Maximum (Kılınç et al. 2021). Further migrations from the south to the north and northeast are known from the Early and Late Neolithic, with later spreads towards the west (Kılınç et al. 2021, Kuzmin et al. 2022: 65-67, Zeng et al. 2023, Gill et al. 2024). Crucially, new waves of migration replaced previous inhabitants; fundamental reasons for the repeated population turnover were likely small population size and thus vulnerability to environmental hazards.

As a result, all language families currently known in the region date at most to 6800-4500 BP (Gill et al 2024, Khanina, Forthc., Zeng et al. 2023). Their arrival erased previous linguistic diversity and they had

only limited time to diversify (cf. Nichols 2024: 743-749). In addition, the communities who used them often converged with neighbours in their cultural practices, further reducing linguistic diversity.

References

Anderson, G.D.S. 2006. Towards a typology of the Siberian linguistic area. In Matras, McMahon & Vincent (eds.), *Linguistic areas. Convergence in historical and typological perspective*, 266–300.

Dobzhanskaja, Oksana. 2008. *Shaman music of Samoyedic people in Krasnoyarsk region* [Shamanskaja muzyka samodijskih narodov Krasnojarskogo kraja]. Norilsk: APEKS

Fu et al. 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445-449.

Gill et al. 2024. Reconstructing the Genetic Relationship between Ancient and Present-Day Siberian Populations. *Genome Biol Evol*. 16(4):evae063

Khanina, Olesya. 2021. Language and ideologies at the Lower Yenisei (Siberia): reconstructing past multilingualism. *International Journal of Bilingualism* 25(4). Special issue 'Typology of small-scale multilingualism' guest edited by Nina Dobrushina, Olesya Khanina & Brigitte Pakendorf, 1059–1080.

Khanina, Olesya. Forthc. Language spreads in the Arctic, Grenoble, Lenore & Khanina, Olesya (eds.). *Language change in the Arctic*. Oxford: OUP.

Kılınç et al. 2021. Human population dynamics and Yersinia pestis in ancient northeast Asia. *Science Advances* 7: eabc4587

Kuzmin et al. 2022. Humans and Nature in Siberia: From the Palaeolithic to the Middle Ages. In Bocharnikov & Steblyanskaya (eds), *Humans in the Siberian Landscapes*, 59-87.

Laptander, Roza. 2020. *When we got reindeer, we moved to live to the tundra: the spoken and silenced history of the Yamal Nenets*. PhD thesis. University of Lapland. (Acta Electronica Universitatis Lapponiensis 278).

Nichols, J. 2024. Northern Asia as a linguistic area. In Vajda (ed), *The Languages and Linguistics of Northern Asia*, 741-766.

Raghavan et al. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87-91.

Sikora et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570(7760):182-188.

Skakun, V.A., S. Yu. Brazhnik, A.P. Pedchenko, I. Yu. Makarenkova, E.I. Barabanshchikov, T.V. Vasilyeva, A.N. Gadinov, G.M. Ustiuzhinskii, S.I. Dudkin & L.G. Bondarenko. 2016. Analysis of use of fish stock in fresh water bodies of Russia in 2013 [Анализ использования рыбных запасов внутренних пресных водоёмов России отечественным рыболовством в 2013 г.]. In: *Trudy VNIRO*. Vol. 160. 212-229

Yegorov, Y. 2016. Population Structures in Russia: Optimality and Dependence on Parameters of Global Evolution. *REGION*. 3/1, 89-106.

Zeng et al. 2023. Postglacial genomes from foragers across Northern Eurasia reveal prehistoric mobility associated with the spread of the Uralic and Yeniseian languages. *biorxiv*. October 04, 2023.

Linguistic distances between dialects are indicative of the history of resettling

George Moroz, Michael Daniel, Nina Dobrushina, Konstantin Filatov, Anastasija Ivanova, Maksim Melenchenko & Anastasija Alekseeva

(HSE University, University of Tübingen, Laboratoire Dynamique du Langage (CNRS, Lyon), HSE University, University of Munich, HSE University, HSE University)

Dialectometry, Rutul, Nakh-Daghestanian, language diversification, geographic distances

Dialect diversification is the first step towards, and ultimately leads to, language diversification. We try to reconstruct dynamics of language divergence in the dialects of Rutul, a minority language without a clear attractor dialect (cf. Pfeiler & Skopeteas 2022) by methods of dialectometry. To our knowledge, the smallest language size to which dialectometry has been applied is by an order larger than Rutul (cf. 15,000 speakers of Rutul and 300,000 speakers of Sui in Stanford 2012).

Rutul, Nakh-Daghestanian, is spoken in the highlands of the southern Caucasus. The Rutul villages of Daghestan lie along Samur and in the valleys adjacent to it. The walking distance between most adjacent Rutul villages ranges between 30 and 60 minutes. Linguistic distances between village lects range from very small to very significant. While standard Rutul is based on the variety spoken in the village of Rutul, the standard did not exert a strong influence on other varieties, because the village lects remain predominantly spoken, with limited exposure of their speakers to Rutul literacy.

The field study of Rutul dialects was conducted in July 2022. It traces a large number of features (425) in twelve highland villages. Unlike most previous dialectometry studies that focus on lexical comparison (Wieling et al. 2011, Pfeiler & Skopeteas 2022, Huisman et al 2019, Stanford 2012), but similarly to Morozova et al (2023), our features cover various linguistic domains, including vocabulary, phonology, morphology and syntax (Alekseeva et al 2023). Particularly important for the purposes of this study is that our data covers almost all existing settlements in a continuous geographical area. Similarly to Huisman et al. (2019), we aim at reconstructing scenarios leading to the observed linguistic divergence. Using statistical methods, we assess linguistic distances between the village lects and look at their correlation with geographic distances.

The statistical results are as follows. The beta-regression on Fig. 2 shows the expectation of the increase of linguistic distance as a function of geographic distance. On Fig. 3, geographically unexpected linguistic distances for each pair of neighboring villages are shown on a graph.

These results comply with conventional assessments of the status of the varieties of Rutul and what is known of their history. Khnov is an old village (Lavrov 1962: 114); Khnov Rutul is sometimes considered a separate language (Koryakov et al 2023). Even though its geographical distance from its closest neighbour, Kiche, is much higher than in any other pair, its linguistic distance is still higher than expected. Ikhrek is another old village speaking a highly divergent variety of Rutul. On the other hand, Kiche and Kufa are believed to be recent out-settlements from Rutul, and Dzhilikhur to be an out-settlement from Myukhrek (Lavrov 1962 and oral interviews with the villagers). In case of Yucatec Maya, for smaller-than-expected differences in the areas close to more densely populated settlements, Pfeiler et al. (2022) proposes a convergence interpretation, suggesting that these settlements acted as centers of convergence. For Rutul, we propose an alternative explanation: a divergence interpretation based on the relative timing of resettlement.
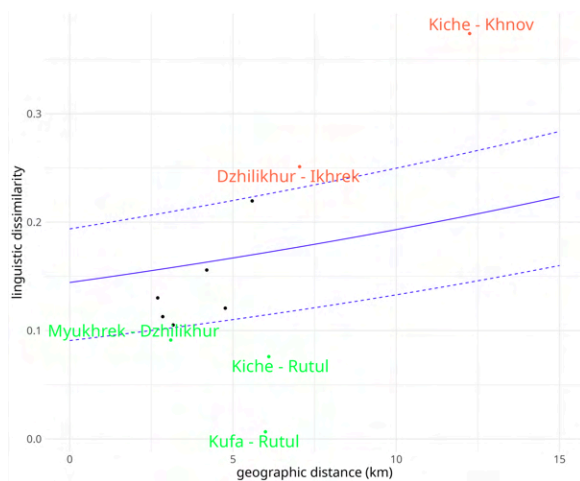
Fig. 1. Relation between linguistic (Y) and geographic (X) distances observed within village pairs (beta-regression fit, betareg package [Grün et al 2012]); only pairs of closest neighbours are shown. For closest neighbour pairs, red outliers are closest neighbors in which the linguistic difference is higher than expected; green outliers are closest neighbours in which linguistic differences are lower than expected.
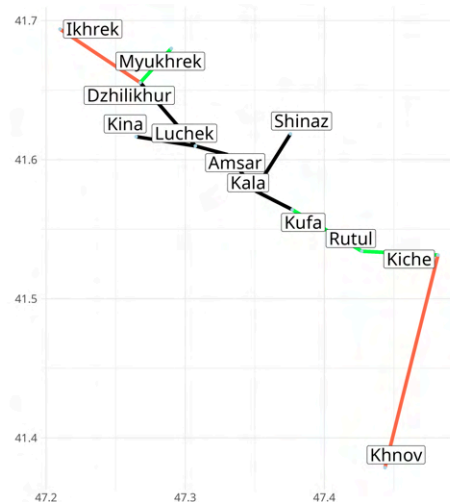
Fig. 2. Closest neighbours graph. Red edges show linguistic distances that are unexpectedly high wrt geographic distances; green edges show linguistic distances that are unexpectedly low wrt geographic ones.

References:

Alekseeva, Anastasija, N. Beklemishev, M. Daniel, N. Dobrushina, K. Filatov, A. Ivanova, T. Maisak, M. Melenchenko, G. Moroz, I. Netkachev, I. Sadakov (2023). Atlas of Rutul dialects. Moscow: Linguistic Convergence Laboratory, HSE University. https://lingconlab.github.io/rutul_dialectology/index.html

Pfeiler, Barbara B. and Skopeteas, Stavros (2022). Sources of convergence in indigenous languages: Lexical variation in Yucatec Maya. PLOS ONE 17(5).

Huisman, John L.A., Asifa Majid and Roeland van Hout (2019). The geographical configuration of a language area influences linguistic diversity. PLOS ONE 14(6)

Grün, Bettina, Ioannis Kosmidis, Achim Zeileis (2012). Extended Beta Regression in R: Shaken, Stirred, Mixed, and Partitioned. Journal of Statistical Software, 48(11), 1-25. doi:10.18637/jss.v048.i11

Koryakov, Yuri, Tatiana Davidjuk, Anastasija Evstigneeva and Arzhaana Sjurjun (2023). *Spisok jazykov Rossii*. https://jazykirf.iling-ran.ru/list_2023.shtml

Lavrov, Leonid I. (1962). Rutulcy v prošlom i nastojaŠŠem. Kavkazskij etnografičeskij sbornik 3. Moscow – Leningrad, 110-157.

Morozova, Maria, Anastasija Escher and Alexander Rusakov (2023). Linguistic complexity of South Slavic dialects: a new perspective on old data. Linguistics Vanguard, 9(s1), pp.133-154.

Stanford, James N. (2012). One size fits all? Dialectometry in a small clan-based indigenous society. Language Variation and Change. 24(2): 247-278.

Wieling, Martijn, John Nerbonne and R. Harald Baayen (2011). Quantitative social dialectology: explaining linguistic variation geographically and socially. PLOS ONE 6. e23613. doi:10.1371/journal.pone.0023613.

Linguistic ecology and language change in Albanian varieties of Kosovo

Maria Morozova & Alexander Rusakov

(Institute for Linguistic Studies, Russian Academy of Sciences, Saint Petersburg & Institute of Language and Literature, Academy of Sciences of Albania, Tirana)

Our research addresses the question of how the observable language properties and their evolution and development are affected by the corresponding linguistic ecologies. The present paper considers Albanian in Kosovo in its many dialectal varieties. Kosovo occupies a near-central position in the Balkan region and has played an important role in Balkan history. Its territory is surrounded by high mountains (up to 2,500 m) and forms a clear-cut geographical unit. The interior of Kosovo is relatively flat and raised, with a lower range of hills which divides it into two parts. Despite the ring of mountains, Kosovo has always been a crossing-place for trade routes and an area of population movements from and to the neighboring territories (Malcolm 2002).

The data for our study are drawn from the Dialectological atlas of the Albanian language, or DAAL (Gjinari et al. 2007), with 25 of 131 localities situated in Kosovo. Gheg Albanian varieties spoken in Kosovo are of diverse origin and have existed in a variety of settings, ranging from high contact to strong geographic and/or social isolation. The main variable considered is linguistic, or language complexity, defined after Nichols (2009: 111) as "complexity of the strictly linguistic domains of phonology, morphosyntax, lexicon, etc. and their components". For Albanian varieties, we measured the complexity on the basis of 28 phonological and morphological features from DAAL.

In our approach, complexity is viewed as a dynamic property showing geographical variation and subject to diachronic change which can be either contact-induced or associated with language-internal processes. Our goal is to identify which linguistic processes contributed to the variation of complexity across Albanian varieties in Kosovo and which external factors could affect this development. Such factors include but are not limited to geographic proximity or separation of the local Albanian-speaking communities, loose/dense social networks (Trudgill 2011), exogamous/endogamous marriage ties, migrations, and contacts with other language speakers.

Our analysis shows that Albanian varieties in Kosovo demonstrate a recent simplification and at the same maintain some of the earlier common Northern Gheg complexifying innovations. Their complexity ranges from 17 to 21 (of 28 possible). The most complex varieties (with complexity of 20–21) are spoken in northern Kosovo, where the local linguistic ecology is shaped by the mountainous landscape and considerable isolation of speech communities from one another, favoring complexification and preservation of "inherited" complexity. Considering that Albanian communities of Kosovo did not necessarily involve adult bilingualism (in Trudgill's terms, one of the main prerequisites of simplification related to language contact), simplification of varieties spoken in the interior of Kosovo may result from dialect leveling and mixing in loose-knit migrational communities formed after massive Albanian population movements from the mountainous north of Albania to Kosovo in 16–18 centuries (Elsie 2015). At a small-scale level, these considerations are illustrated by comparing groups of Albanian population, which are hypothesized to have originated from one ancestral community and diverged after a series of migrations, e.g. Shala of Bajgora in Kosovo and Dukagjin Shala in Albania.

References

Elsie, Robert (2015), *The Tribes of Albania: History, Society and Culture*, London / New York: I.B. Tauris.

Gjinari, Jorgji, Bahri Beci, Gjovalin Shkurtaj & Xheladin Gosturani (2007), *Atlasi dialektologjik i gjuhës shqipe*, Vëll. 1, Napoli: Università degli studi di Napoli L'Orientale.

Malcolm, Noel (2002), *Kosovo: A Short History*, London: Pan Books.

Nichols, Johanna (2009), Linguistic complexity: a comprehensive definition and surve, in: G. Sampson, D. Gil, and P. Trudgill (eds), *Language Complexity as an Evolving Variable* (= Studies in the Evolution of Language 13), Oxford: Oxford University Press, 110–125.

Trudgill, Peter (2011), *Sociolinguistic typology: Social determinants of linguistic complexity*, Oxford: Oxford University Press.

Language geography of the Northern Samoyedic area in Siberia

This paper provides a description of a vast area in Western Eurasian Arctic and Subarctic in terms of its language geography. We focus on the continuous area from the White sea in the west to the Tajmyr peninsula in the east, where related languages, Northern Samoyedic (< Uralic), have been spoken by nomadic reindeer herders, hunters, and fishers at least in the last 500 centuries (as reflected in written records). We ask what the linguistic diversity of this area is from the point of view of a worldwide typology, and what environmental and sociolinguistic factors have shaped it. To answer the question, we integrate earlier research results, stemming from different types of evidence (linguistic, sociolinguistic, geographic, ethnographic, historic).

We take into account:
- geographic spread of each Northern Samoyedic language/dialect in the area and known changes thereof in the last 500 years (see Figure 1);
- extent of divergence between all the languages and dialects involved (based on Authors et al. In Prep., see Figure 2):
    - for a measure of divergence in lexicon, the share of cognate lexical items in Uralex is taken (Uralex has 311 items, de Heer et al. (2021); Northern Samoyedic part by Koryakov et al. (2025));
    - for a measure of divergence in phonology and morphology, relative amounts of shared phonological/morphological innovations since Proto-Samoyedic are taken, (sets of 50+ innovations were taken for each domain, Authors et al. Submitted);
- traditional social relations between inhabitants of the area and language ideologies governing language use, mentioning the existing contrast between a multilingual subarea at the Yenisei river and monolingual subareas to the east and west of it (based on Khanina 2021 and Authors et al. Submitted).

Finally, we use two contrasting cases to focus on the complexity of interplay between geographic factors, language ideologies, and linguistic change. On the one hand, as shown by Koryakov (2025), Tundra Nenets (TN) extends across 2000+ km, but despite the existence of clear territorial groups with separate histories in the last 500 years, the pan-TN identity preserves and the TN dialects show relatively little divergence. On the other hand, as shown by Khanina et al. (2018) and Khanina & Shluinsky (2022), the two Enets languages, occupying a territory several times smaller than the TN, exemplify that distinct social identities and divergent linguistic features can be consciously upkept for centuries. Taken together, the two cases demonstrate how very similar environmental and economic settings could produce intriguingly different linguistic results, presumably stemming from language ideologies. The latter shows the significance of speakers' agency that needs to be accounted for in large-scale perspectives on language geographies.

We end with a review of recurrent patterns in how the changes of Northern Samoyedic language territories correspond to the extent of linguistic divergence and to the sociolinguistic details. We see these patterns as our contribution to the typology of local language ecologies worldwide.
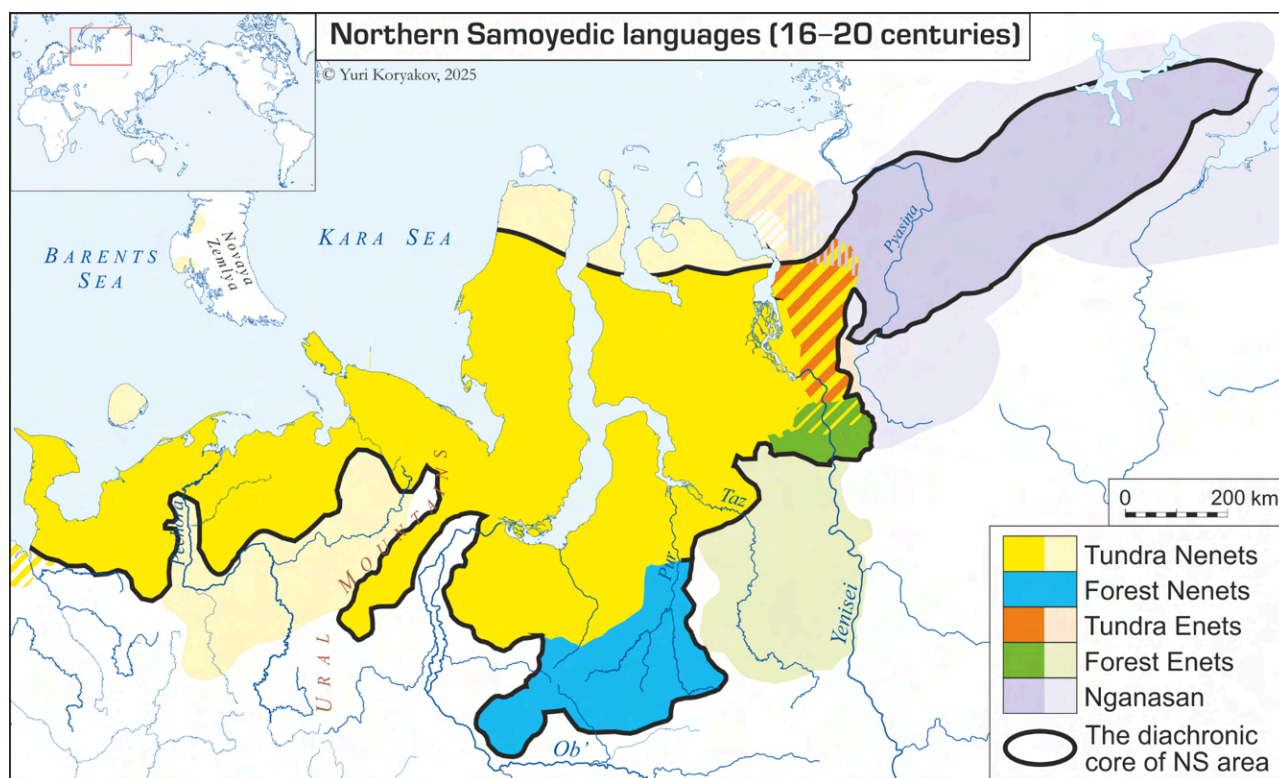
Figure 1. Map showing the core and the peripheries of the NS area (by brighter colors the core NS territory is shown, where these languages were in use without interruption for at least the last 500 years)
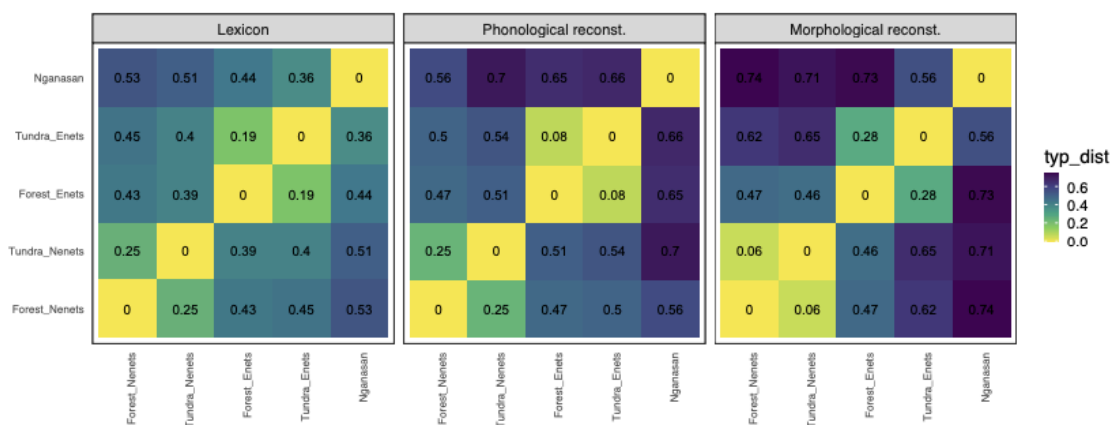

Figure 2. Linguistic distances between Northern Samoyedic languages by three linguistic domains (lexical cognates, phonological innovations, morphological innovations), based on Authors et al. (In Prep.)

References

Khanina, Olesya. 2021. Language and ideologies at the Lower Yenisei (Siberia): reconstructing past multilingualism. *International Journal of Bilingualism* 25(4). Special issue 'Typology of small-scale multilingualism' guest edited by Nina Dobrushina, Olesya Khanina & Brigitte Pakendorf, 1059–1080.

Khanina, Olesya, Yuri B. Koryakov, Andrey Shluinsky. 2018. Enets in space and time: a case study in linguistic geography. *Finnisch-Ugrische Mitteilungen* 42: 1-28.

Khanina, Olesya & Andrey Shluinsky. 2022. Forest Enets and Tundra Enets: how different are they and why? In: Kaisla Kaheinen, Larisa Leisiö, Riku Erkkilä & Toivo E. H. Qiu (eds.), *Hämeenmaalta Jamalille: kirja Tapani Salmiselle 07.04.2022*, 181-200. Helsinki: Helda Open Books, University of Helsinki.

Koryakov, Yuri B. 2025. Linguistic geography of Tundra Nenets: the history of the western groups. *International Journal of Eurasian Linguistics* 7.1.

Koryakov, Yuri, Valentin Gusev & Olesya Khanina. 2025. Northern Samoyedic wordlists in the Uralex database, In Szeverényi, Sándor (Ed.). *Proceedings of the 10th workshop on Samoyedic languages and cultures / 6th Mikola Conference, September 2024* (*Studia Uralo-Altaica* 57). Szeged: University of Szeged.

de Heer, Mervi, Mikko Heikkilä, Kaj Syrjänen, Jyri Lehtinen, Outi Vesakoski, Toni, Suutari, Michael Dunn, Urho Määttä, & Unni-Päivä Leino. 2021. *Uralic basic vocabulary with cognate and loanword information* (Version v2.0) [Data set]. Zenodo. https://zenodo.org/records/6392555

Matches and mismatches in the distribution of linguistic diversity in South America

Rik van Gijn (Leiden University)
Zoe Poirier Maruenda (Leiden University)
Nicholas Q. Emlen (University of Groningen)
Sietze J. Norder (Utrecht University)

The South American continent presents a microcosm of complex linguistic diversity patterns, which mostly do not align. For instance, while language richness (the number of languages) is generally low in South America as compared to other macroareas, genealogical richness (the number of distinct language families) is generally high (Van Gijn forthc.). Furthermore, structural diversity (the range of typological profiles represented among the languages) does not seem to correlate with genealogical diversity: the genealogically highly diverse western part of the continent has been suggested to show less structural diversity than the genealogically less diverse eastern part of the continent (Muysken et al. 2014). This is consistent with results presented by Van Gijn & Norder (2024), which showed that, on a global level, genealogical diversity is a poor predictor for structural diversity. This suggests that different types of diversity may be sensitive to different drivers of diversification and convergence, or that similar drivers have differential effects. For instance, in South America, structural convergence has been suggested to result from conscious identity preservation in the context of widespread multilingualism (Epps 2020).

In this contribution, we identify matches and mismatches between the distributions of genealogical and structural diversity in South America, and assess to what extent the explanatory model by Epps (2020) can explain the results, or whether other explanatory models are called for. We use two different measurements of linguistic diversity. The first is richness, which operationalizes diversity as the number of different units (languages, language families, structures, depending on the type of linguistic diversity is to be measured) in a given geographical area. This is a commonly used measure of diversity. The second measure, developed in Norder and Van Gijn (forthc.), we refer to as linguistic endemism. This measure quantifies the degree of uniqueness of a given geographical unit in terms of its languages, language families or linguistic structures. These measures allow us to quantitatively assess the predictive value of genealogical richness and genealogical endemism for explaining structural richness and structural endemism on continental and subcontinental scales. In turn, this may allow us to identify more subtle differences in the distribution of richness and endemism, which stay below the radar when regarded from a global perspective. A more regional focus also allows for a connection to more concrete sociohistorical processes and mechanisms from the regional anthropological literature, including different kinds of sociopolitical organization and kinship systems, subsistence practices, and participation in trade or marriage networks.

For structural diversity, we focus on the distribution of phonemes, based on Moran and McCloy (2019). Genealogical diversity is based on the consensus classifications in Hammarström et al. (2024).

References

Epps, Patience. 2020. "Amazonian Linguistic Diversity and Its Sociocultural Correlates." In *Language Dispersal, Diversification, and Contact*, edited by Mily Crevels and Pieter Muysken, 275–90. Oxford University Press. https://doi.org/10.1093/oso/9780198723813.003.0016.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2024. *Glottolog 5.1*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.14006617

Moran, Steven, and Daniel McCloy, eds. 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. http://phoible.org.

Muysken, Pieter, Harald Hammarström, Joshua Birchall, Swintha Danielsen, Love Eriksen, Ana Vilacy Galucio, Rik van Gijn, et al. 2014. "The Languages of South America: Deep Families, Areal Relationships, and Language Contact." In *The Native Languages of South America: Origins, Development, Typology*, edited by Loretta O'Connor and Pieter Muysken, 299–322. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107360105.017.

Norder, Sietze J, and Rik Van Gijn. Forthc. "Linguistic Endemism: The Continuum between Language Isolates and Family Expansions."

Van Gijn, Rik. forthc. "Linguistic Diversity and the South American Perspective." In *The Many Voices of Indigenous Latin America: Reconstructing, Describing, and Preserving Cultural and Linguistic Diversity*, edited by Françozo Mariana, Rik Van Gijn, and Martine Bruil. Leiden: Brill.

Van Gijn, Rik & Sietze J. Norder. 2024. "Linguistic endemism: Genealogical versus structural diversity." Paper presented at the 57th annual meeting of the Societas Linguistica Europaea, Helsinki, 21 August 2024.

Quantitative Analysis of Russian Dialects

Søren Wichmann, Matías Guzmán Naranjo & Darja Jonjić
(Kiel University & University of Freiburg & Kiel University)

This study investigates the dialectal landscape of the Russian language through a quantitative analysis of *The Dialectological Atlas of the Russian Language* (abbreviated DARJa based on its Russian title). Published in Moscow from 1986 to 2005, DARJa represents decades of linguistic data collection, and includes 313 maps – each corresponding to a specific linguistic feature – covering 4196 locations.

In 2015-16, researchers at Kazan Federal University extracted linguistic features and their values directly from the physical maps and created Excel files giving feature values across locations (Isaev et al. 2016). We have processed these materials further by georeferencing the map of covered locations and manually extracting latitude and longitude coordinates of every location. This integration of linguistic data with geospatial coordinates has enabled the application of advanced dialectometric methods, including interpolation of feature values using the Hamming distance and hierarchical clustering, to derive clusters that reflect dialectal divisions across the Russian linguistic landscape.

To investigate the historical and sociolinguistic foundations of Russian dialect geography, we first examined the relationship between dialectal divisions and historical administrative units. Our first results reveal an alignment between dialect clusters and the boundaries of *ujezds* – the administrative units established as early as the 15th century. This finding highlights the enduring influence of historical sociopolitical structures on linguistic variation. Beyond administrative correlates, we investigated the role of other factors in shaping the dialectal landscape, including population size. As in the study of Dutch data by Nerbonne & Heeringa (2006), our Russian data fails to show support for Trudgill's (1974) gravity model.

Aiming to identify large-scale patterns that arise from local dynamics, we introduce a new clustering method to determine the optimal number of dialect clusters. While approaches like Marchenko's MDS-based classification into six zones (2023) and Pšeničnova's structural-typological analysis resulting in five zones (1996) provide valuable frameworks, they lack a systematic way to objectively identify the optimal number of clusters. To address this gap, we introduce a new method, applying a process in which we select a number of k clusters and evaluate the fit of the classification by calculating the Adjusted Rand Index (ARI) for each linguistic feature separately. By averaging the ARI across all features, we get a measure of how well each choice of k clusters fits the data. This iterative ARI-based method enables us to identify the optimal number of clusters, which in the case of DARJa is five.

This study advances language geography by presenting a new, objective clustering approach for analyzing dialect continua. By examining historical, sociolinguistic, and geographic factors, we show how these elements shape linguistic variation in Russia. Our new method for determining optimal dialect clusters offers a replicable framework for studying language geographies in other regions.
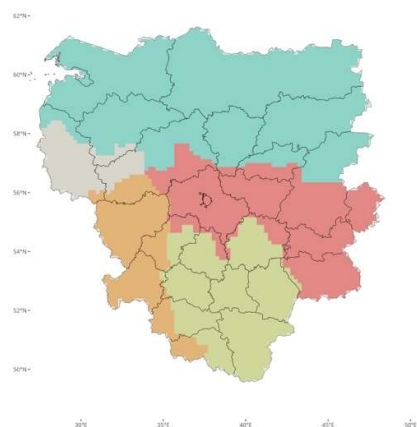
Figure 1: Map using interpolation and clustering on the DARJa data showing the optimal number of five clusters.

References

Isaev, Igor', Valerii Solov'ev, Farid Salimov, Aleksandr Piljugin & Venera Bairaševa (2016). Creating a database of Russian dialects and prospects for dialectometric studies, *Herald of the Russian Academy of Sciences* 86(6): 448–453.

Marchenko, Igor (2023). Building a classification of Russian dialects using multidimensional scaling. BA thesis, Higher School of Economics, Moscow. DOI: 10.13140/RG.2.2. 31726.38727.

Nerbonne, John & Wilbert Heeringa (2007). Geographic distributions of linguistic variation reflect dynamics of differentiation, in Sam Featherston, and Wolfgang Sternefeld (eds), *Roots: Linguistics in Search of its Evidential Base*, 267–297. Berlin/New York: Mouton de Gruyter.

Pšeničnova, Nadežda (1996). *Tipologija russkix govorov* [Typology of Russian dialects]. Moscow: Nauka.

Trudgill, Peter (1974). Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography, *Language in Society* 3(2): 215–246.

WS20 The syntax and semantics of perception

Clémentine Raffy, Michelle Sheehan, Giulia Mazzola & Liam Garside

From taste perception to flavors of epistemicity: Italian *mi sa*

Andrea Miglietta & Eva-Maria Remberger
(University of Vienna)

Keywords: perception, evidentiality, epistemicity, grammaticalization, parenthetical

In colloquial Italian the fixed expression *mi sa (che)* encodes the epistemic and evidential positioning of the speaker regarding the truth conditions of the propositional content of the sentence (1). This expression is composed of two invariable elements: the first person singular dative clitic *mi*, and the third person singular present tense form of the verb *sapere*, a verb which inherits its ambiguous meaning between 'taste/smell' and 'know' from the Latin verb *săpĕre.* According to Serianni (2012) the epistemic verbal expression *mi sa* (+CP) stems diachronically from the use of the construction *mi sa* (+PP) where *sapere* expresses a perceptual experience (2) (i.e. a "non-visual sensory" source for direct evidentiality, cf. Aikhenvald 2004; also Viberg 1983 for a typology of perception verbs, where 'taste' is at the lower end of a hierarchy). This diachronic change shows some of the features characteristic of grammaticalization patterns as described by Heine (1993): fixation (from a completely productive syntactic structure to an invariable expression); semantic bleaching (from a concrete meaning to a more abstract one); change in complementation (from a PP to a CP); impossibility to be under the scope of negation (*La zuppa non mi sa di sale* vs **Non mi sa che è andata a Verona*).

(1) Mi   sa   che è andata   a Verona.   (KiParla)
     me.DAT know.3SG   that is gone.F.SG to Verona
     'I think she went to Verona.'

(2) La zuppa   mi   sa   /gli   sapeva   di sale.
     The soup   me.DAT know.3SG   him.DAT know.IPF.3SG   of salt
      'The soup tastes salty to me/tasted salty to him.'

In this study we compare the distribution, in spoken Italian, of epistemic/evidential *mi sa* with the distribution of *credo* (Giorgi and Pianesi 2005), one of the most frequent epistemic complement-taking Italian verbs. This comparison is carried out with the help of an extended data set from the corpus *KIParla* (cf. also Riccioni et al. 2022) as well as judgments by native speakers. Our analysis is embedded in a generative grammar model and aims at answering the following two main research questions:

a. How does the original perceptual meaning of *mi sa* influence its epistemic value following the typology for the epistemic domain proposed e.g. by Pietrandrea (2004)?
b. How does the fixed internal morphosyntax of *mi sa*, with the complement clause in the function of subject and the dative experiencer inherited from the theta grid of perceptual *sapere*, influence its distribution?

From a semantic point of view *mi sa*, in comparison with *credo*, is expected to show a meaning closer to inferential evidentiality ("first hand experience") rather than epistemicity and a lower degree of speaker's certainty. From a syntactic perspective, *mi sa* is expected to show some of the functional

features proper to *credo* and to be concerned with phenomena connected to the left periphery (Cinque 1999), such as complementizer deletion. Because of its internal morphosyntax and its bleached meaning it is also expected to be found in parenthetical positions, a key feature concerned with the development of pragmatic values.

References:

Aikhenvald, Alexandra (2004), *Evidentiality*. Oxford: OUP.

Cinque, Guglielmo (1999), *Adverbs and Functional Heads*, New York: OUP.

Giorgi, Alessandra & Fabio Pianesi (2005), *Credo* (I Believe): Epistemicity and the syntactic representation of the speaker, *Working papers of Linguistic* 15, University of Venice.

Heine, Bernd (1993), *Auxiliaries: cognitive forces and grammaticalization*, Oxford: OUP.

Pietrandrea, Paola (2004), L'articolazione semantica del dominio epistemico dell'italiano, *Lingue e Linguaggio* 3.

Riccioni, I. & Zuckowski, A. & Burro, R. & Bongelli, R. (2022), The Italian epistemic marker *mi sa* [to me it knows] compared to *so* [I know], *non so* [I don't know], *non so se* [I don't know whether], *credo* [I believe], *penso* [I think], *PLoS ONE* 17, e0274694.

Serianni, Luca (2012), *Mi sa*, *Bollettino di italianistica* 9(2): 18–23.

Viberg, Åke (1983). The verbs of perception: a typological study. *Linguistics* 21/1, 123–162.

Pseudo-relative clauses in Australian languages: A typological survey

Dana Louagie, Ellison Luk

(University of Liège & University of Namur; independent)

This paper studies pseudo-relative clauses (PsRCs) in a balanced sample of 50 Australian languages. PsRCs encode verb complements (including perception), yet are formally very close to or indistinguishable from restrictive relative clauses. Compare examples (1a-b) from Mawng, where (1a) shows a relative clause (RC) functioning to restrict the head's reference, and (1b) a PsRC conveying event perception. Both clauses are introduced by an article showing noun class agreement with a noun: the head for RCs ('the man'), and the subject of the complement predicate for PsRCs ('the ground'). PsRCs are attested in diverse languages, including Romance languages (Van der Auwera 1985, Dik & Hengeveld 1991), Lori (Iranian; Noonan 2007), Korean (Kim 2008) and Mawng (Iwaidjan; Singer 2007), but it is unclear how common they really are cross-linguistically. This study identifies such structures in the languages of Australia, thus extending Singer's (2007) study of Mawng, and investigates their formal and functional characteristics.

(1) Mawng (Iwaidjan, Singer 2007: 274–276; her brackets)

a. *ngi-wurru-n* [*ja arrarrkpi* [***ja*** *ati-ma-ny* *parak ta kurnpi*]].
1SG/3M-know-NPST [M man [M 3M/3ED-take-PST.PFV away ED green.plum]]
'I know the man who took the green plums.'

b. *malany la aw-alyu-ng* [*ta kunak **ta** an-jirrngu-nang*] *pata ngakngak.*
so_then 3PL-feel-PST.CONT [LL ground LL 3LL-shake-PST.CONT] PL bird.type
'Then the Ngakngak birds felt the ground shaking.'

We identify potential PsRCs in 18 languages across Australia, with no clear geographical or phylogenetic patterning. We label a complex sentence construction as a potential PsRC if it could be used to encode both a restrictive relative clause and perception complement clause with no formal difference, as in the Mawng examples (1). Formally, PsRCs are quite diverse: finite and non-finite clauses are about equally represented, and marking may involve verb inflections, free-form conjunctions, clitics and agreement marking (typically case). An example of a PsRC marked by a verb inflection is given in (2b) from Anindilyakwa; compare with the RC in (2a), which uses the same marker *-ma*. PsRCs may show minor formal distinctions from restrictive RCs, e.g. in object marking on the main verb (e.g. Mawng; Singer 2007) or in the optionality of markers (e.g. Danggati; Lissarrague 2000, 2007: examples).

(2) Anindilyakwa (Gunwinyguan; van Egmond 2012: 305, 176)

a. *narrv-m-angka-rnv-ma* *mvnhvnga* *m-ibina* *narrv-ma-lyakukwa-**ma***
3A-VEG-collect-PST2-*ma* VEG.burrawang VEG-that 3A-VEG-soak.PST2-**DEP**
'They collected the burrawang nuts that they had soaked (in water).'

b. *nenv -rrvngka* *wurr-ambilyuma* *wurrajija* *nuw-angkarree-yi-na-**ma***
3M/COLL-see.PST2 COLL-two COLL.bird COLL-run-RECP-NPST2-**DEP**
'He saw the two birds flying away.'

Functionally, we investigate which types of event perception (Dik & Hengeveld 1991) PsRCs may encode in our sample. For Mawng, Singer (2007) highlights one cross-linguistic difference: French only uses PsRCs for immediate perception of events, while Mawng also uses them for mental perception (e.g. 'I heard of the fact that…'). In our sample, of all the types of perception, sight and hearing are the most commonly found, while mental and tactile types ('knowing', 'feeling') are quite marginal (cf. Evans & Wilkins 2000). Additionally and in as far as the available data allow us, we consider Singer's (2007: 282-285) hypothesis that PsRCs serve a particular discourse function, viz. introducing new referents and predicating on them at the same time (cf. Lambrecht 1994, 2002). Finally, we relate our findings to the broader issue of multifunctional subordinate structures in Australian languages (Nordlinger 2006), and thus contribute to our understanding of typical relative clauses in those languages as well (Hendery 2023; Louagie & Luk in prep.).

Abbreviations

A 'augmented', COLL 'collective noun class', CONT 'continuous', DEP 'dependent marker', ED 'edible gender', LL 'land gender', M 'masculine gender', N 'neuter' NPST 'non-past', PFV 'perfective', PL 'plural', PST 'past', RECP 'reciprocal', SG 'singular', VEG 'vegetable noun class'. The slash (/) indicates an agent (left) acting on a patient (right).

References

Auwera, Johan van der (1985), The predicate relatives of French perception verbs, in A.M. Bolkestein, C. de Groot, and J.L. Mackenzie (eds), (1985), *Predicates and terms in Functional Grammar*, Dordrecht: Foris, 219–234.

Dik, Simon and Kees Hengeveld (1991), The hierarchical structure of the clause and the typology of perception-verb complements, *Linguistics* 29; 231–259.

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.

Hendery, Rachel (2023), Relative clauses, in C. Bowern (ed.), *The Oxford Guide to Australian Languages,* 457–467. Oxford: Oxford University Press.

Lambrecht, Knud (1994), *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents,* Cambridge: Cambridge University Press.

Lambrecht, Knud (2002), Topic, focus and secondary predication: The French presentational relative construction, in C. Beyssade, R. Bok-Bennema, F. Drijkoningen and P. Monachesi (eds), *Romance Languages and Linguistic Theory*, Amsterdam / Philadelphia: John Benjamins, 171–212.

Lissarrague, Amanda (2000), A salvage grammar of Dunghutti, University of New England BA thesis.

Lissarrague, Amanda (2007), *Danggati grammar and dictionary with Dhanggati storie,* Nambucca Heads: Muurrbay Aboriginal Language and Culture Co-operative.

Noonan, Michael (2007), Complementation, in T. Shopen (ed), *Language Typology and Syntactic Description*, Cambridge: Cambridge University Press, 52–150.

Nordlinger, Rachel (2006), Spearing the Emu Drinking: Subordination and the Adjoined Relative Clause in Wambaya, *Australian Journal of Linguistics* 26(1), 5–29.

Singer, Ruth (2007), Events Masquerading as Entities: Pseudorelative Perception Verb Complements in Mawng (Australian) and Romance languages, in A.C. Schalley & D. Khlentzos (eds), *Mental States vol. 2: Language and cognitive structure,* Amsterdam/Philadelphia: John Benjamins, 267–288.

Tersko perception complementation: a Romance-like system in Slavic?

Francesco Costantini & Jelena Živojinović
(University of Udine; University of Graz)

Keywords: clausal perception verb complements, language contact, Tersko, Friulian

Clausal perception verb complements (see Declerck 1983; Van der Meer 1994 a.o., e.g. English *I heard* [*the kids screaming*]) denote sensorily perceived events; they include at least two syntactic constructions: the 'event perception construction' (EPC) and the 'individual perception construction' (IPC) (Casalicchio 2016a). EPC emphasizes the event itself, while IPC focuses on the individual satisfying the external θ-role in the 'perceived' event. Syntactically, languages can vary as for how IPC and EPC are expressed; in Italian for instance IPC is expressed via an infinitive clause (e.g. *Ho visto Gianni salutare Maria* 'I saw Gianni greet Maria') or via a pseudo-relative (e.g. *Ho visto Gianni che salutava Maria* 'I saw Gianni that greeted Maria'); EPC is only expressed via an infinitive clause (e.g. *Sento bussare* 'I hear (someone) knocking'; **sento che bussa*, meaning 'I hear (someone) knocking').

Tersko, a Slovenian variety spoken in the region of Friuli, Italy, has at least three available structures to express perceived events: the infinitive clause (1), the pseudorelative (2), as defined by Casalicchio (2021), and a gerundive structure (3).

(1) *Si   videu Jana <u>jesti</u>   jabok*.
    AUX.2SG seenJan   eat.INF apple
    'You saw Jan eating an apple'

(2) *Si   obritou Jana   <u>ki   u   spau</u>*.
    AUX.2SG found   Jan   that he.CL slept
    'You found Jan$_i$ sleeping$_i$'

(3) *Si   videu sonce <u>ustajoć</u>.*
    AUX.2SG seen sun   rise.GER
    'You saw the sun rise'

In order to determine their syntactic nature, we tested all three constructions by varying the argument structure (transitive, unaccusative, unergatives and null subject) and the type of perception verb (lexical vs. semi-functional, e.g. *to see* vs. *to find*). The preliminary results show a complementary distribution between gerundial structures for IPCs, which are attested in all conditions, except for null subjects), and infinitives for EPCs, while the pseudorelative spans across both syntactic constructions, null subject structures included.

We suggest that language contact with Romance varieties, primarily the historical contact with Friulian, has played a role in shaping the syntax of nowadays clausal perception in Tersko: first, (2) replicates a pattern found in Friulian, where the subject relative pronoun and the subject clitic pronouns cooccur; second, although nowadays Friulian does not exhibit gerundial structures in IPCs, 16th-century Friulian did (see ex. (4); cf. Casalicchio 2016b).

(4) *[...]O che sinte currint lisiarte o sbōrs*
       or that hear run.ger lizard or green_lizard
       'Or he hears the lizard or green lizard run'
                    (Trav. dell'Orl. Fur. I 33; cf. Joppi 1878; retrieved from Casalicchio 2016b)

Additionally, Tersko's syntax mirrors Romance in using gerunds for IPCs (3), infinitives for EPCs (1), as in 16th c. Friulian (see Tab.1 below), and pseudo-relatives in both contexts (Modern Friulian). Building on these data, we ponder over the nature of contact as a trigger for syntactic change.

	IPCs	EPCs
16th c. Friulian	**GERUND**	**INFINITIVE**
Tersko	**GERUND** (no null sbjs) **Pseudo-Relative** (null subjs included)	**INFINITIVE** (no null sbjs) **Pseudo-Relative** (null sbjs included)

Tab. 1: Clausal perception verb complements in 16th c. Friulian and Tersko

References

Casalicchio, Jan (2016a), The use of gerunds and infinitives in perceptive constructions, in E. Bidese, F. Cognola, M. Moroni (eds), *Theoretical Approaches to Linguistic Variation*, Amsterdam: John Benjamins, 53-87.

Casalicchio, Jan (2016b), Ricostruire la diacronia della sintassi ladino-dolomitica con Joppi. Il caso dei costrutti percettivi in: F. Vicario (ed), *Ad limina Alpium. VI Colloquium Retoromanistich, Cormons, dai 2 ai 4 di Otubar dal 2014*, Udine: Società Filologica Friulana, 97-126.

Casalicchio, Jan (2021), Expressing perception in parallel ways: Sentential Small Clauses in German and Romance, in S. Wolfe and C. Meklenborg (eds), *Continuity and Variation in Germanic and Romance*, Oxford: OUP, 70-96. DOI:10.1093/oso/9780198841166.003.0004.

Declerck, Renaat (1983), On the Passive of Infinitival Perception Verb Complements, *Journal of English Linguistics, 16(1),* 27-46.

Joppi, Vincenzo (1878), Testi inediti friulani dei secoli XIV al XIX, *Archivio Glottologico Italiano 4*, 185-342.

Van der Meer, Geart (1994), Verbs of perception and their complementation, *English Studies 75(5),* 468-480.

The syntax of perception along the Binding Hierarchy: an exploratory comparative corpus study of Romance languages

Giulia Mazzola, Clémentine Raffy, Liam Garside & Michelle Sheehan

(Newcastle University)

Keywords: syntax, semantics, complementation, perception, Romance

In generative and cognitive studies of Romance causatives, parallels have often been drawn with perception verbs taking clausal complements (Kayne 1975; Burzio 1986; Raposo 1989; Guasti 1993; Soares Da Silva 2012). So far, however, there has never been a systematic study of the syntax and semantics of the complementation of Romance perception verbs, including: finite clauses (1), Exceptional Case-Marking (ECM, (2)), clause union of the Faire-Infinitive (FI) or Faire-Par (FP) type, (2) and (3), pseudo-relatives and gerunds (4), prepositional (5) and inflected infinitives.

Although this rich variation has been observed (Casalicchio 2013; Sheehan 2016), the relationship between structure and interpretation, and its cross-linguistic distribution, remains unclear. In this study, we conduct a parallel corpus study of Italian and European French, Portuguese and Spanish using the TenTen corpora (Kilgarriff et al. 2014), extracting 2,000 random occurrences for each verb in each language (equivalents of *see, hear, listen* and *watch* – where available). Our aim is to consider the semantics of the attested patterns in relation to Givón's (2001) Binding Hierarchy and Wurmbrand and Lohninger's (2023) elaboration thereof, contributing to theoretical debates on argument binding and syntactic variation.

We analysed the data by using Multiple Correspondence Analysis to find structure in multivariate data through the observation of groupings (Desaguliер 2020:438), including the variables: language, complement type, semantic interpretation of the perception verb (e.g. pure, indirect, metaphorical perception), transitivity of subordinate verb, animacy of the matrix and embedded subject.

Preliminary results show that, in the first two dimensions, complement types cluster together with the semantic interpretation or the transitivity of the verb, regardless of the language. As expected, finite complements encode non-literal and indirect readings across all languages. Unaccusatives favour FIs due to the preference of these verbs for the postposition of the subject. The semantic-syntactic clustering confirms the predictions of the Binding Hierarchy. Whether transitivity systematically overrides semantics remains an open question. We also discuss some unexpected language-specific patterns, like the occurrence finite complements with the agentive verb *escutar* 'to listen' in Portuguese, and the remarkable absence of FI with a dative NP (like *al pescatore* in example (3)), in Italian and Portuguese.

This exploratory study offers a systematic view of syntactic variation across perception verbs in Romance and helps us generate finer research hypotheses to further investigate the syntax and semantics of this verb class.

Examples:

Finite clause - [Italian]

(1) *Adriano,*   <u>*ved-o*</u>   ***che***   ***ti***   ***sei***
     A.   see-PRS.1SG   COMP   REFL.2SG   be.PRS.2SG

     secc-at-o   ***e***   ***ti***   ***chied-o***   ***scus-a.***
     get.annoyed-PTCP.PST-M.SG   and   DAT.2SG   ask-PRS.1SG   apology.(F)-SG

     'Adriano, I see that you got annoyed, and I apologise for that' (ItTenTen, 500forum.it)

FI and ECM (intransitive) - [Spanish]

(2) *Así*   <u>*ve-rás*</u>   ***crec-er***   ***tu-s***   ***venta-s***   /
     this.way   see-FUT.2SG   grow-INF   your-PL   sale(F)-PL   /

     tu-s   ***venta-s***   ***crec-er*** *en region-es*   *más*   *lejan-a-s*
     your-PL   sale(F)-PL   grow-inf   in region(F)-PL   more   far-F-PL

     'This way you will see your sales grow in regions further away' (adapted from EsTenTen, hubspot.es)

FP and FI (transitive) - [Italian]

(3) *mi*   *è*   *capit-at-o*   *di*   <u>*ved-ere*</u>
     DAT.1SG   be.PRS.3SG   happen-PTCP.PST-M.SG   of   see-INF

     tagli-are   ***le***   ***cim-e***   ***da-l***   ***/ a-l***   ***pescator-e***
     cut-INF   DET.F.PL   rope(F)-PL   by-DET.M.SG   /   to-DET.M.SG   fisherman.(M)-SG

     'I happened to see the ropes cut by the fishermen / the fisherman cutting the ropes' (adapted from ItTenTen, amicidellavela.it)

Pseudo-relative and gerund - [French]

(4) *Mais*   *à*   *chaque* *relais*   *elle*   <u>*entend-ai-t*</u>   ***l'Anglais***
     but   at   each   station(M)   she   hear-IPFV-3SG   DET.M.SG=English.M.SG

     qui   ***se***   ***promen-ai-t***   ***/***   ***se***   ***promen-ant***
     REL.NOM   REFL.3SG   wander-IPFV-3SG   /   REFL.3SG   wander-GER

     autour *des*   *deux*   *voiture-s.*
     around   of.DET.PL   two   car-PL

     'But at each station, she could hear the Englishman that walked/walking around the two cars.' (adapted from FrTenTen, ebalzac.com)

Prepositional infinitive - [European Portuguese]

(5)   *Vitória*   *[…]*   <u>*ouve=o*</u>   ***a***   ***cont-ar***   *à*   *mãe*
Vitória   hear.PRS=3.M.SG.ACC at   tell-INF   to.DET.F.SG   mother

o   *que Carlos*   *lhe=*   *disse*   *sobre*   *Margarida*
DET.M.SG   REL Carlos   3.SG.DAT=   tell.PST.3SG   about Margarida

'Vitória […] hears him telling her mother everything that Carlos told him about Margarida.'
(adapted from PtTenTen, www.tv7dias.pt)

References

Burzio, Luigi (1986), *Italian syntax. A Government-Binding approach*. Dordrecht: Reidel.

Casalicchio, Jan (2013), *Pseudorelative, gerundi e infiniti nelle varietà romanze: affinità (solo) superficiali e corrispondenze strutturali*. Padova: Università di Padova PhD Thesis. https://www.research.unipd.it/handle/11577/3423609. (19 June, 2024).

Desagulier, Guillaume (2020), Multivariate Exploratory Approaches, in M. Paquot and S.Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*, Cham: Springer International Publishing, , 435–469. https://doi.org/10.1007/978-3-030-46216-1_19.

Givón, Talmy (2001), *Syntax: an introduction*, vol. 2. Amsterdam: John Benjamins.

Guasti, Maria Teresa (1993), *Causative and Perception Verbs: A Comparative Study*, Turin: Rosenberg & Sellier.

Kayne, Richard S (1975), *French Syntax: The Transformational Cycle*, Cambridge, Massachusetts: MIT Press.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel (2014), The Sketch Engine: ten years on, *Lexicography* 1, 7–36.

Raposo, Eduardo (1989), Prepositional Infinitival Constructions in European Portuguese, in O. A. Jaeggli and K. J. Safir (eds.), *The Null Subject Parameter,* Dordrecht: Springer Netherlands, , 277–305. https://doi.org/10.1007/978-94-009-2540-3_10.

Sheehan, Michelle (2016), Complex predicates, in A. Ledgeway and M. Maiden (eds.), The Oxford Guide to the Romance Languages, Oxford: Oxford University Press, 981–994. https://doi.org/10.1093/acprof:oso/9780199677108.003.0061.

Soares Da Silva, Augusto (2012), Stages of grammaticalization of causative verbs and constructions in Portuguese, Spanish, French and Italian, *Folia Linguistica* 46(2), 513–552. https://doi.org/10.1515/flin.2012.018.

Wurmbrand, Susanne and Magdalena Lohninger (2023), An implicational universal in complementation: Theoretical insights and empirical progress, in J.M. Hartmann and A. Wöllstein (eds.), *Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues,* Tübingen: Narr Verlag, , 183–229.

Seeing that in preschool French: Investigating clausal complementation as a function of evidentiality

Katerina Palasis

(Université Côte d'Azur, CNRS, BCL, France)

Keywords: perception verbs, complementation, evidentiality, acquisition, French

Research background: Evidential markers are described mostly as morphological and lexical in the world's languages and to a lesser extent as syntactic (on child and adult French, see Bassano, 1996; Benveniste, 1958; Cournane and Tailleur, 2020; Dendale and Izquierdo, 2014; Dendale and Van Bogaert, 2012; Granville Hatcher, 1944; a.o.). Nevertheless, recent studies have highlighted that when children move from simple to complex syntax with perception verbs, the shift entails that the initial narrow semantics of the verb (i.e., direct perception only) expands to more complex representations (i.e., indirect perception and inference, Davis and Landau, 2021). Furthermore, recent research has also suggested correspondences between different types of complementation and different types of information source. Angelopoulos, Bagioka and Terzi (2023), for instance, have shown that completive and pseudo-relative clauses differ in Standard Modern Greek (SMG), and that children above 9 years of age are sensitive to the divide. In their study, the completive structure (1a) is described as unmarked with regard to evidentiality, whereas the pseudo-relative type (1b) can only be uttered if the speaker actually witnessed the event. This line of research was recently extended to adult French with a hypothesized distribution between completive clauses, which express indirect information source (1a), versus relatives and infinitives (1b-1c), which convey direct witnessing (Baunaz, Palasis and Arslan, 2025).

(1)   Hypothesized distribution of embedded clauses as a function of evidentiality:
   a. Completive clauses:

Ida	[oti	kapios	efige]	(SMG)
J'ai vu	[que	quelqu'un	est parti]	(French)
I saw	that	someone	has left	

   b. Pseudo-relatives:

Ida	kapion	[pu	efige]	(SMG)
J'ai vu	quelqu'un	[qui	est parti]	(French)
I saw	someone	who	has left	

   c. Infinitives:

J'ai vu	quelqu'un	[partir]	(French)
I saw	someone	leave	

Methodology: This contribution extends the syntactic hypothesis on the completive/relative-infinitive divide by testing it in child French. The current study examined a corpus of 36,063 semi-naturalistic utterances in 16 native preschoolers (2;5-5;11). The children's biclausal sentences with the verb *voir* 'to see' in the matrix clause were extracted and filtered (final *n* = 38, 12 children), and the type of embedded clause (completive, relative or infinitive; 28.9%-53.3%-15.8%, respectively) was examined as a function of information source (direct, indirect or inference). Additional features, such as tense and person, were examined in order to further document the divide that exists in French between

proper relative clauses and pseudo-relatives (e.g., Pozniak, Hemforth, Haendler, Santi and Grillo, 2019).

Results: The data show that: (i) relative clauses convey direct evidentiality only; (ii) indirect evidentiality is expressed in completives and infinitives; and (iii) inference is expressed in completives only. Tense concordance between the main and the embedded verbs in relative constructions indicates a strong tendency for tense match, whether past or present, possibly identifying the relative clauses as pseudo-relatives rather than proper relatives.

Conclusions: The preliminary results seem to support the hypothesis of the existence of a syntactic strategy for evidentiality in child French, notably because when a child utters more than one type of evidentiality (i.e., direct witnessing and hearsay or inference), s/he distributes the different types over different syntactic structures. Interestingly however, the child and adult distributions are not strictly identical in French. These findings thus pave the way for further investigations on the developmental aspects of this syntactic strategy in French.

References

Angelopoulos, Nikos, Bagioka, Dafni-Vaia and Arhonto Terzi (2023), Exploring syntactically encoded evidentiality, *Language Acquisition,* 30(1), 50-75, https://doi.org/10.1080/10489223.2022.2087525

Bassano, Dominique (1996), Functional and formal constraints on the emergence of epistemic modality: A longitudinal study on French, *First Language,* 16(46), 77-113, https://doi.org/10.1177/014272379601604605

Baunaz, Lena, Palasis, Katerina and Seçkin Arslan (2025), *Investigating the syntax of evidentiality in French: Three information-source judgement experiments*, 58th Annual Meeting of the Societas Linguistica Europaea (SLE 2025), Bordeaux, August 26-29 2025.

Benveniste, Émile (1958), De la subjectivité dans le langage, in *Problèmes de linguistique générale 1* (1966), Paris: Gallimard, 258-266.

Cournane, Ailís and Sandrine Tailleur (2020), La production épistémique chez l'enfant francophone: complexité syntaxique et ordre d'acquisition, *Arborescences,* 10, 47-72, https://doi.org/10.7202/1081888ar

Davis, E. Emory and Barbara Landau (2021), Seeing vs. seeing that: Children's understanding of direct perception and inference reports, in A. Beltrama, F. Schwarz and A. Papafragou (eds.), *Proceedings of the first Experiments in Linguistic Meaning (ELM) Conference* (Vol. 1), Pennsylvania: Open Journal Systems, 125-135.

Dendale, Patrick and Damaso Izquierdo (2014), Les marqueurs évidentiels ou médiatifs en français: une approche bibliographique, *Cahiers de lexicologie,* 105, 79-97.

Dendale, Patrick and Julie Van Bogaert (2012), Réflexions sur les critères de définition et les problèmes d'identification des marqueurs évidentiels en français, *Langue française,* 173(1), 13-29, https://doi.org/10.3917/lf.173.0013

Granville Hatcher, Anna (1944), Je le vois sourire; Je le vois qui sourit; Je le vois souriant, *Modern Language Quarterly,* 5(3), 275-301, 387-405, https://doi.org/10.1215/00267929-5-3-275

Pozniak, Céline, Hemforth, Barbara, Haendler, Yair, Santi, Andrea and Nino Grillo (2019), Seeing events vs. entities: The processing advantage of Pseudo Relatives over Relative Clauses, *Journal of Memory and Language,* 107, 128-151, https://doi.org/10.1016/j.jml.2019.04.001

Investigating the syntax of evidentiality in French: Three information-source judgement experiments

Lena Baunaz, Katerina Palasis & Seçkin Arslan
(Université Côte d'Azur, CNRS, BCL, France)

In this paper, we evaluate a novel hypothesis regarding evidential marking in French. We argue that evidentiality in French, beyond its lexical and morphological manifestations (see Benveniste 1966, Dendale and Van Bogaert 2012, Squartini 2008, a.o), exhibits notable interactions with syntactic structures. To substantiate this claim, we report findings from three experimental studies conducted on French subordinated clauses. The results of these studies provide compelling evidence for a correlation between evidentiality and syntax in French. The implications of these findings suggest a syntactic dimension to evidential marking that has previously been mentioned but underexplored in the context of French (e.g., Granville Hatcher 1944).

Recent research has broadened our understanding of evidentiality markings, demonstrating that it can be conveyed not only morphologically or lexically but also syntactically, through diverse clausal structures (e.g., Angelopoulos et al. 2023 for Standard Modern Greek and Asudeh and Toivonen 2012 for English). We extend the investigation to French by examining three types of subordinate clauses under perception verbs: tensed complement clauses (1a), (pseudo-)relatives (1b), and ECM infinitives (1c).

(1) a. Tensed complement:

J'	ai	vu	que	quelqu'un	est	parti.
I	have	seen	that	someone	is	left

'I saw that someone has left.'

 b. (Pseudo-)relative:

J'	ai	vu	quelqu'un/Marie	qui	est	parti/e.
I	have	seen	someone/M	who	is	left

'I saw someone/Marie who has left.'

 c. ECM infinitive:

J'	ai	vu	quelqu'un/Marie	partir.
I	have	seen	someone/M	leave

'I saw someone leave.'

Experiment 1 was a witnessing rating in source identification task. We tested the structures in (1), and the complement clause was additionally tested with the main clause *on m'a dit* 'I was told' (control condition). Participants (*n* = 51) read sentences and had to decide whether the speaker witnessed the event or not on a scale from 1 (not witnessed, without doubt) to 5 (witnessed, without doubt). Experiment 2 was a forced-choice in source identification task. Participants (*n* = 81) read the same types of sentences as in Experiment 1, and had to choose one of three information sources: direct witnessing, inference or hearsay. Experiment 3 was a forced-choice discourse completion task in which the participants (*n* = 89) read one of two contexts (visual direct or indirect), and had to select which of two sentences they would produce to match the context. Experiment 3a submitted a choice between

a complement and a pseudo-relative clause, whereas 3b submitted a complement versus an ECM infinitive clause.

Our results indicate that ECM infinitives under perception verbs overwhelmingly mark direct evidentiality, and pseudo-relatives and relatives mainly convey direct evidentiality. As for complement clauses with *voir*, they are clearly associated with indirect evidentiality.

Our three information-source judgment experiments conducted with French speakers therefore show that the choice of syntactic structure and the evidential interpretation are strongly associated. We suggest that the diverging behaviours between infinitive ECM and (pseudo-)relative clauses on the one hand and complement clauses on the other hand could pertain to differences in their syntactic structure. Infinitive ECM and (pseudo-)relative clauses involve Small Clauses (Casalicchio 2016, Cinque 1992, a.o.) (an event is selected by a perception verb), whereas complement clauses involve CP-selection (by a cognitive verb).

References

Angelopoulos, Nikos, Bagioka, Dafni-Vaia and Arhonto Terzi (2023), Exploring syntactically encoded evidentiality, *Language Acquisition,* 30(1), 50-75, https://doi.org/10.1080/10489223.2022.2087525

Asudeh, Ash and Ida Toivonen (2012), Copy raising and perception, *Natural Language & Linguistic Theory,* 30, 321-380, https://www.doi.org/10.1007/s11049-012-9168-2

Benveniste, Émile (1958), De la subjectivité dans le langage, in *Problèmes de linguistique générale 1* (1966), Paris: Gallimard, 258-266.

Casalicchio, Jan (2016), Pseudo-relatives and their left-periphery: A unified account, in E. Carrilho, A. Fiéis, M. Lobo, and S. Pereira (eds.), (2016), *Romance Languages and Linguistic Theory. Selected papers from 'Going Romance' 28, Lisbon* (Vol. 10), Amsterdam: John Benjamins, 23-42.

Casalicchio, Jan and Peter Herbeck (2024), Pseudorelative clauses, infinitives, and gerunds with Spanish perception verbs: A comparative view, *Syntax: a journal of theoretical, experimental and interdisciplinary research,* https://doi.org/10.1111/synt.12271

Cinque, Guglielmo (1992), The Pseudo-Relative and Acc-*ing* Constructions after Verbs of Perception, *University of Venice Working Papers in Linguistics,* 2, 1-31.

Dendale, Patrick and Damaso Izquierdo (2014), Les marqueurs évidentiels ou médiatifs en français: une approche bibliographique, *Cahiers de lexicologie,* 105, 79-97.

Dendale, Patrick and Julie Van Bogaert (2012), Réflexions sur les critères de définition et les problèmes d'identification des marqueurs évidentiels en français, *Langue française,* 173(1), 13-29, https://doi.org/10.3917/lf.173.0013

Granville Hatcher, Anna (1944), Je le vois sourire; Je le vois qui sourit; Je le vois souriant, *Modern Language Quarterly,* 5(3), 275-301, 387-405, https://doi.org/10.1215/00267929-5-3-275

Squartini, Mario (2008), Lexical vs. grammatical evidentiality in French and Italian, *Linguistics,* 46(5), 917-947, https://doi.org.10.1515/LING.2008.030

From perception to speech acts: A study on Brazilian Portuguese perception verb *ver* 'see'

Marcus Lunguinho & Patrícia Rodrigues
(Brasilia University & Federal University of Paraná)

This paper examines two uses of the Brazilian Portuguese verb *ver* 'see', in (1), that has not been extensively studied. In these sentences, *ver* functions as: (i) a full verb selecting a finite *se* 'if' complement with the meaning 'to verify', or (ii) a pragmatic marker conveying either a direct order (1a) or the speaker's perspective (surprise or denial) (1b).

(1) a. Vê   se você me escuta!
        see-IMP   if you   me listen
        'Verify if you can hear me.'   (lexical verb)
        'Listen to me!'   (pragmatic marker)
     b. Vê   (lá)   se Maria vai chegar atrasada!
        see-IMP (there) if Maria will arrive   late
        'Verify if Maria will be late.'   (lexical verb)
        'I don't think that Maria will be late!'   (pragmatic marker)

The perception associated with verification, like other readings of *ver* (e.g., direct or indirect), depends on the complement form. Specifically, it is the finite *se* 'if' complement – an indirect interrogative clause – that defines *ver* as 'to verify', since the indirect interrogative conveys content that can be either true or false, requiring perception to assess its truth value. These constructions induce replies using the verb *ver* (2a) and display a complete morphological paradigm (2b).

(2) a. Vejo, sim!
        'Ok, I will check it.'
     b. Ela vai ver / já viu / sempre via se a tradução estava correta.
        'She will check / already checked / always checked if the translation was correct.'

Therefore, this meaning derives from a biclausal structure, in which *ver* is a full verb that selects two arguments – a DP and a ForceP headed by the complementizer *se* 'if':

(3)   [_{vP} DP [_{v'} *v* [_{VP} *ver* [_{ForceP} *se* [_{TP}

Regarding the expressive meanings, sentences in (1) induce replies with the second verb (4a) and lack a complete morphological paradigm (4b).

(4)   a. Não, me escuta você!
         'No, you listen to me!'
      b. *Ela vai   ver / já   viu   / sempre via   se me   escuta!
          she will   see / already saw   / always see   if me   listen

These meanings would be derived from a mono-clausal structure, where *vê*, a fixed form originating from the imperative, has undergone a process of pragmaticalization (cf. Dostie 2004) becoming a pragmatic marker. We propose to explain these constructions based on the proposals of Speas and Tenny (2003) and Hill (2014), according to which discursive-pragmatic phenomena are part of the syntactic structure and are encoded in the domain of a syntactic projection: the Speech Act (SA) category. We argue that this pragmatic marker is externally merged into the Speech Act phrase (SAP). Furthermore, these pragmaticalized expressions c-select a ForceP complement headed by *se*:

(5) [SAP   *vê* (*lá*) [ForceP   *se* [TP

The data examined in this work provide empirical evidence for the Speech Act category and can contribute to the discussion about the syntacticization of discourse (Haegeman and Hill 2013).

References

Dostie, Gaétane (2004), *Pragmaticalisation et Marqueurs Discursifs:* Analyse sémantique et traitement lexicographique, Brussels: De Boeck/Duculot.

Haegeman, Liliane; Hill, Virginia (2013), The syntacticization of discourse, in: R. Folli, R. Truswell and C. Sevdali (eds.), (2013), *Syntax and its Limits*, Oxford: Oxford University Press, p. 370-390.

Hill, Virginia (2014), *Vocatives: How syntax meets with pragmatics*, Leiden: Brill Publishers.

Speas, Peggy; Tenny, Carol (2003), Configurational properties of point of view roles, in: A. M. Di Sciullo (ed), (2003), *Asymmetry in Grammar*, Amsterdam: John Benjamins, p. 315-344.

Gerundive complements and direct vs. imaginative readings of *ver* 'see' in Brazilian Portuguese

Patrícia Rodrigues
(Federal University of Paraná)

Keywords: perception verbs; gerundive complements; direct reading; imaginative reading; Brazilian Portuguese

This paper analyzes Brazilian Portuguese sentences like (1), where the perception verb *ver* 'see' is ambiguous between a direct and an imaginative reading in the presence of a gerundive construction.

(1)　Pedro vê　　　Maria　　voltando　　para　casa.
　　　Pedro see-PRS　Maria　　return-PROG　to　　home
　　　'Pedro sees Maria returning home.'

While the direct interpretation, which implies simultaneity between the perception and the perceived event, has been widely discussed in the literature, the imaginative interpretation, which places the perception within in the experiencer's imagination (cf. Boivin 1998), remains underexplored. I will show that these two readings result from distinct structural configurations of the gerundive construction (this ambiguity arises exclusively in the presence of a gerund clause).

A key distinction between the two readings is that, under the direct reading, the gerundive construction allows three analyses: a complex DP, a DP object with a secondary predicate, or a small clause complement, where the DP is the gerund's subject (cf. Rodrigues, 2006). In contrast, under the imaginative reading, it can only be a small clause complement. For example, when the matrix verb is passivized (2), the imaginative reading becomes unavailable, while the direct reading remains. This occurs because, under the direct reading, the gerund is a secondary predicate (cf. Casalicchio and Herbeck, 2024), whereas under the imaginative reading, its small clause structure blocks a long passive.

(2)　Maria foi　　　vista　　voltando　　pra　casa.　(only direct reading)
　　　Maria be-PST　　see-PTCP　return-PROG　to　　home
　　　'Maria was seen returning home.'

Moreover, in imaginative perception, the following properties are observed that are absent in the direct reading: tense mismatches can occur (3), the gerund can be negated (4), individual-level predicates are accepted (5), and the complement clause can feature a topic (6).

(3)　Maria vê　　　hoje　os　meninos　brincando amanhã. (*direct; ᵒᵏimaginative)
　　　Maria see-PRS　today　the　boys　　play-PROG　tomorrow
　　　'Maria sees today the boys playing tomorrow.'

(4) Maria vê   o   professor não respondendo a-os   alunos. (*direct; ᵒᵏimaginative)
    Maria see-PRS the   teacher   not respond-PROG to-the students
    'Maria sees the teacher not responding to the students.'

(5) Maria vê   Pedro possuindo   uma   casa. (*direct; ᵒᵏimaginative)
    Maria see-PRS Pedro own-PROG   a   house
    'Maria sees Pedro owning a house.'

(6) Eu vejo   este livro$_i$ a   Maria lendo   ele$_i$. (*direct; ᵒᵏimaginative)
    I   see-PRS this book the Maria read-PROG it

Assuming that tense mismatch, negation and the possibility of an individual-level predicate indicate the presence of TP, I argue that, in the imaginative reading, the gerundive complement projects TP with an additional functional layer above it, allowing a topic position, while, in the direct reading, it lacks TP and projects AspP. These findings support studies on Romance languages suggesting that the meaning of the perception verb depends on the size of its complement.

References

Boivin, Marie Claude (1998), Complementation and interpretation: the concrete and imaginative readings of 'visual' perception verbs, *MIT Working Papers in Linguistics*, vol. 25, p. 103-123.

Casalicchio, Jan; Herbeck, Peter (2024), Pseudorelative clauses, infinitives, and gerunds with Spanish perception verbs: A comparative view, *Syntax,* https://doi.org/10.1111/synt.12271

Rodrigues, Patricia (2006), Les compléments infinitifs et gérondifs des verbes de perception en portugais brésilien, PhD-dissertation – Université du Québec à Montréal.

The syntax and semantics of Indonesian taste/touch verb

Perception verbs can undergo intrafield or transfield extensions (Viberg 1983; San Roque et al. 2018). The former involves extensions within the domain of perception, while the latter extends into the domain beyond perception, such as cognition. The paper focuses on the Indonesian perception verb from the "lower sense" denoting taste and touch, with semantic extensions into the domains of thinking and feeling. The core meaning of the root *rasa* is 'taste' (Stevens & Schmidgall-Tellings 2008; KBBI 2016). *Rasa* also encompasses the perception of touch in a broad sense (Winter 2019: 14–15). Furthermore, the Indonesian perception verb *rasa* signifying 'to taste' or 'to touch' extends semantically into the domain of cognition encompassing 'to feel' (1) and 'to think' (2) in terms of 'to believe' and 'to opine' (Jansegers & Gries 2017).

(1) | *Keluarga* | *Zaenal* | *me-**rasa*** | *curiga* | *dengan* | *kematian=nya* |
| --- | --- | --- | --- | --- | --- |
| family | NAME | AV-**rasa** | suspicious | with | death=3S |
| *yang* | *dinilai* | *tidak* | *wajar.* | | |
| REL | considered | NEG | natural | | |

'Zaenal's family **felt** suspicious about his unnatural death.'

(2) | *Karena* | *jujur* | *saja* | *saya* | ***rasa*** | *yang* |
| --- | --- | --- | --- | --- | --- |
| because | honest | only | 1S | ***rasa*** | REL |
| *nama=nya* | *Badan* | *Anggaran* | *ada* | *otak-otak=nya.* | |
| name=3S | agency | budget | EXIST | RED-brain=3S | |

'Because to be honest, I **think** that there are stooges (lit. brains) in the Budget Committee.'

In contrast, other languages exhibit distinct patterns, with perception verbs linked to "higher senses" often carry cognitive meanings, as seen in English's 'to see' (Sweetser 1990) and Australian auditory verbs (Evans & Wilkins 2000). The closely related meanings of 'to feel' and 'to think' conveyed by the Indonesian gustatory and haptic perception verb *rasa*, are rooted in the Indonesian cultural model that regards *rasa* as the cognitive faculty used to describe the "intuitive aspects of reality" (Stange 1984: 114). Using the Indonesian corpus from Leipzig Corpora Collection (ind_mixed_2013) containing 74,329,815 sentences and 1,206,281,985 tokens (Goldhahn et al. 2012), we examine the distribution of semantic domains for ten different verb forms based on *rasa*, categorizing them into COGNITION (including the sub-domains 'to feel' and 'to think') and PERCEPTION (including the sub-domains 'to taste' and 'to touch'). The paper aims to explore the semantic properties of these ten different verb forms containing *rasa* using behavioral profile analysis (Divjak & Gries 2006) and cluster analysis (Levshina 2015: 301–321) for gaining insights into their categories and characteristics. We aim to answer the following research questions: (a) what is the internal structure of a cluster of the verb forms, (b) how can we distinguish the sub-clusters and the individual verb forms from each other, and (c) how can we pinpoint a prototypical situation conveyed by each verb form. Our preliminary results show that different voices, with distinct constructions, convey varied meanings. For instance, the middle voice tends to express the domain of perception. The active voice is typically followed by an adjective conveying the domain of emotion, while the passive voice tends to convey the domain of cognition in the sense of opinion, like the zero (voice) prefix, usually followed by a subordinate sentence.

References

Divjak, Dagmar & Stefan T. Gries. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2. 23-60.

Evans, Nicholas & David Wilkins. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*. 546–592.

Goldhahn, Dirk, Thomas Eckart & Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the 8th Language Resources and Evaluation Conference (LREC) 2012*, 759–765.

Jansegers, Marlies & Stefan T. Gries. 2017. Towards a dynamic behavioral profile: A diachronic study of polysemous sentir in Spanish. *Corpus Linguistics and Linguistic Theory* 16(1). 145–187.

KBBI. 2016. Kamus Besar Bahasa Indonesia (Comprehensive Indonesian Dictionary). https://kbbi.kemdikbud.go.id. (15 January, 2023.)

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins Publishing Company.

San Roque, Lila, Kobin H. Kendrick, Elisabeth Norcliffe & Asifa Majid. 2018. Universal meaning extensions of perception verbs are grounded in interaction. *Cognitive linguistics* 29(3). 371–406.

Stange, Paul. 1984. The logic of rasa in Java. *Indonesia* (38). 113–134.

Stevens, Alan M. & A. E. Schmidgall-Tellings. 2008. Kamus Lengkap Indonesia-Inggris [A Comprehensive Indonesian-English Dictionary]. *Bandung: PT Mizan Pustaka*.

Sweetser, Eve. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure* (54). Cambridge University Press.

Viberg, Åke. 1983. The Verbs of perception: a typological study in Explanations for Language Universals. *Linguistics* 21(263). 123–162.

Winter, Bodo. 2019. Sensory linguistics. *Language, perception and metaphor. Amsterdam & Philadelphia (Benjamins)*.

Variable agreement constructions in Spanish: between perception modalities and conceptual foregrounding

Renata Enghels

(Ghent University)

Keywords: perception verbs, agreement, infinitive construction, profiling, Spanish

The Spanish pronominalized infinitival construction is characterized by variable singular and plural verb agreement between the perception verb, the perceived participant and the subordinate infinitive:

(1) a. *se   ve(n)   caer   los   árboles*
         SE   see.3SG(PL)   fall.INF the.PL   tree.PL
         ('one sees the trees falling'/ 'the trees are seen falling')
     b. *se   oye(n)   arrancar   las   máquinas*
         SE   hear.3SG(PL) start.INF   the.PL   machine.PL
         ('one hears the machines starting up' / 'the machines are heard starting up')

Grammatical accounts have associated this variability with passive (variable singular vs. plural concord) versus impersonal (consistently singular) interpretations (e.g. Mendikoetxea 1999). However, in previous research, variable agreement has been related to different conceptualizations of the perception stimulus suggesting an agreement ad sensum widespread in Spanish (Enghels 2019). This constructional variability was linked to how speakers foreground their perception of the participant or the event as a whole, influenced by the perception act's nature (object vs. event perception) and the modality (visual vs. auditory).

To investigate linguistic agreement in a controlled yet naturalistic setting, we conducted an acceptability judgment task, particularly suited to test whether speakers accept constructions featuring non-standard agreement. This task involved 100 native Spanish speakers, who completed the survey via LimeSurvey. Participants evaluated 40 sentences that served as fillers in the broader study, systematically manipulated versions of real corpus examples. The focus was exclusively on constructions featuring a postverbal second noun phrase (SN2), allowing for the isolation of agreement as a phenomenon. We used two perception verbs – *ver* ('to see') and *oír* ('to hear') – and constructed 20 sentences per verb, all with a plural SN2: half of the items used a verb in the singular form and the other half in the plural. Each of these combinations was further crossed with the animacy of the referent, resulting in ten sentences with a human SN2 and ten with an inanimate one per verb. Participants rated the acceptability of each sentence on a 7-point Likert scale, allowing for a fine-grained measure of gradient acceptability. Three independent variables were systematically manipulated: perception modality (*ver* vs. *oír*), agreement (singular vs. plural verb), and animacy of SN2 (human vs. inanimate). Two additional factors were also taken into account: the type of infinitive (unaccusative vs. unergative) and the presence or absence of differential object marking (DOM) with a.

The outcomes from a multilevel model show that, in the analysis of plural agreement constructions, no significant differences emerge across different perception modalities. However, a notable distinction arises in the use of singular agreement. Specifically, verbs associated with auditory perception are followed by a plural object more frequently that with visual perception. This suggests that when singular agreement is used, it tends to align with the infinitive complement as a whole

rather than with the specific object in sentences with auditory perception verbs. This pattern aligns with previous findings from corpus studies, reinforcing the hypothesis of the influence of the nature of the perception verb on agreement mechanisms.

References

Enghels, R. 2019. "Linguistic Reflections of Object Perception Versus Event Perception. Agreement Ad Sensum in the Pronominal Infinitive Construction in Spanish." *Syntaxe et Sémantique* 20, 107-123.

Mendikoetxea, A. 1999. "Construcciones con 'se': medias, pasivas e impersonales." En: *Gramática Descriptiva de la Lengua Española*. Bosque, I. y V. Demonte (eds.). Espasa Calpe Mexicana, 1631-1722.

WS21 Towards a better understanding of analogy: challenges, methods, and perspectives

Lorenzo Moretti & Marianne Hundt

Semantics drives morphological change in Germanic strong verbs: A phylogenetic study

Alexandru Craevschi, Sarah Babinski & Chundra Cathcart

(Institute for the Interdisciplinary Study of Language Evolution, University of Zurich)

This study examines the morphological change within Germanic languages. We address the hypothesis that the semantics of tense, aspect and mood influence patterns of stem allomorphy within the verbal paradigm— specifically that on lineages of the Germanic phylogeny where the past participle expresses narrative past tense, verbs show a stronger trend towards alternation patterns where the past participle and preterite forms share a stem to the exclusion of present forms (the so-called "ABB" pattern typified by Dutch *brengen/bracht/gebracht*; Dammel, Nowak, and Schmuck 2010; De Smet and Van de Velde 2019), which has been proposed within individual languages but not tested on a broader evolutionary scale.

We gathered a dataset of strong verbs from Germanic languages using sources such as Wiktionary and UniMorph, coding each verb according to its vowel alternation pattern in infinitive, past and past participle forms. Using a hierarchical phylogenetic approach (Cathcart 2024), we model the phylogenetic evolution of cognate verbs in 14 languages of the Germanic clade (Chang et al. 2015), allowing them to transition between alternation patterns under two different "regimes" representing (1) lineages that preserve a functional separation of past tense and past perfect (e.g., English and Swedish) versus (2) lineages where past perfect constructions can express narrative past (e.g., German and Dutch).

We find that there is a greater long-term preference for the ABB pattern on lineages where past perfect constructions express narrative past, in line with predictions from the literature. We inspect the frequencies of inter-pattern transitions inferred by our model to clarify the mechanisms underlying this result, and find that this configuration appears to obtain primarily due to the fact that verbs are more likely to remain in the ABB state in the relevant regime, preventing verbs from undergoing complete weakening, but not necessarily because other verbs are attracted to this state.

References

Cathcart, Chundra A. (2024), Multiple evolutionary pressures shape identical consonant avoidance in the world's languages, *Proceedings of the National Academy of Sciences* 121(27), e2316677121, DOI: DOI: 10.1073/pnas.2316677121.

Chang, Will, David Hall, Chundra Cathcart, and Andrew Garrett (2015), Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis, *Language* 91(1), 194–244, DOI: DOI: 10.1353/lan.2015.0005.

Dammel, Antje, Jessica Nowak, and Mirjam Schmuck (2010), Strong-Verb Paradigm Leveling in Four Germanic Languages: A Category Frequency Approach, *Journal of Germanic Linguistics* 22(4), 337–359, DOI: DOI: 10.1017/S1470542710000097.

De Smet, Isabeau and Freek Van de Velde (2019), Reassessing the evolution of West Germanic preterite inflection, *Diachronica* 36(2), 139–180, DOI: DOI: `https://doi.org/10.1075/dia.18020.des`.

Analogy in multilingual language acquisition: The case of code-mixing in morphological patterns

Antje Quick, Stefan Hartmann & Nikolas Koch
(Leipzig University, HHU Düsseldorf, LMU München)

Language contact phenomena such as code-mixing and transfer are ubiquitous in multilingual language use, especially so in multilingual first language acquisition (see e.g. Vihman 2018, Koch & Günther 2021, Quick et al. 2021, among many others). It seems intuitively obvious that analogy is a key mechanism underlying many of those phenomena, and indeed there is ample evidence that analogical reasoning plays a major role in language acquisition (e.g. Gentner 2003, Behrens 2017, Ibbotson 2020: 45–75). But to what extent can transfer phenomena be attributed to analogy, and how can we assess the role of analogy in a corpus-based way? In this paper, we try to tackle some aspects of these questions by combining theoretical considerations on the role of analogy in language acquisition, processing, and use with an empirical study of one particular code-mixing phenomenon in longitudinal data from three German-English bilingual children.

In particular, we analyze cases of morphological code-mixing such as *jumpen* 'jump + German infinitive marker', *wrinkelig* 'wrinkle + German adjective marker', and *feete* 'feet + German plural marker' drawing on the naturalistic speech of three German-English bilingual children, "Fion", "Silvie", and "Lily", between the ages of 2 and 3 (n=109,815 utterances) growing up in households with a similar socio-economic status but with different input situations. Our main research questions are a) which types of morphological mixes can be found in the children's data, b) how they relate to the children's overall use of the two languages, c) to what extent the mixes can be explained by direct analogy to phonologically similar words in the source language, and d) which individual differences we find between the children, and whether and how they can be related to different input situations. We extracted all code-mixed utterances (n=10,040) and checked whether they contained a morphological mix, e.g. an English verb with a German inflectional marker or vice versa. The mixes were then annotated for whether the word is a cognate of the corresponding word in the other language, and the phonological distance between the word form and the cognate word form was assessed by calculating the Levenshtein distances between the SAMPA (Wells 1997) transcriptions of both words using a combination of automatic annotation (BAS services, Reichel & Kisler 2014; PLD2flex, Wedig et al. 2024) and manual correction. In addition, samples of the data were annotated for morphological forms and inflection classes to assess to what extent direct analogy between corresponding forms and/or the frequency of morphological schemas are better predictors of morphological mixes. We predict that forms with a close formal correspondence to a cognate or a word with similar meaning in the other language will be more prone to occur in morphologically mixed forms.

On a more theoretical level, we argue that multilingual acquisition offers a highly relevant testbed for gaining a better understanding of analogy and its potential role as gateway to linguistic productivity in early stages of language acquisition.

References

Behrens, Heike. 2017. The Role of Analogy in Language Processing and Acquisition. In Marianne Hundt, Sandra Mollin & Simone E. Pfenninger (eds.), The Changing English Language, 215–239. Cambridge: Cambridge University Press.

Gentner, Dedre. 2003. Why we're so smart. In: Gentner, Dedre & Goldin-Meadow, Susan (eds.), Language in mind – Advances in the study of language and thought. Cambridge, MA: MIT Press, 195–236.

Ibbotson, Paul. 2020. What it takes to talk: exploring developmental cognitive linguistics. Berlin, Boston: De Gruyter.

Koch, Nikolas & Katharina Günther. 2021. Transfer Phenomena in Bilingual Language Acquisition: The Case of Caused-Motion Constructions. Languages 6(1). 25.

Quick, Endesfelder A.; Backus, A.; & Lieven, E. (2021). Entrenchment effects in code-mixing: Individual Differences in German-English bilingual children. *Cognitive Linguistics* 32 (2): 319 – 348. doi.org/10.1515/cog-2020-0036

Reichel, Uwe D. & Thomas Kisler. 2014. Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In Rüdiger Hoffmann (ed.), Elektronische Sprachverarbeitung 2014, 42–49. Dresden, Germany: TUDpress.

Vihman, Virve-Anneli. 2018. Language Interaction in Emergent Grammars: Morphology and Word Order in Bilingual Children's Code-Switching. Languages 3(4). 40.

Wedig, Helena, Felix Theodor, Joshua Wieler & Eva Belke. 2024. PLD2flex: Establishing the Phonological Levenshtein Distance for Pairs or Groups of (Pseudo)Words. Journal of Open Research Software 12. 12. https://doi.org/10.5334/jors.510.

Wells, John C. 1997. SAMPA computer readable phonetic alphabet. In Dafydd Gibbon, Roger Moore & Richard Winski (eds.), Handbook of standards and resources for spoken language systems, 684–738. Berlin, New York: Mouton de Gruyter. https://doi.org/10.1515/9783110804508-008.

Analogy as a situated speaker's act

Ferdinand von Mengden
(Freie Universität Berlin)

My claim is that linguistic analogy needs to be seen as small-scale, situated act. Accordingly, rather than based on conceptual or categorial similarities, analogy is a subtle adaptive transfer of an utterance type to a novel communicative setting.

In order to motivate this I connect analogy with 'recontextualization', a concept which describes a 'flexible transfer' of tool functions (e.g., in primate behavior) or of material culture (in anthropology) onto new tasks. While constituting analogical transfers, recontextualizations are dependent on the specific context of a single act rather than on stored conceptual units.

For linguistics this would open the door for a focus on the specific communicative environment as crucially contributing to the linguistic choices speakers make in a specific instance. Analogical recontextualization can be observed in both lexical semantics and morphosyntactic structures.

For instance, the increasing experience with vehicles with combustion engines crucially contributes to the intransitive use of *drive* (1b). As a prerequisite, different aspects of the same event type (1a/1a') compete for prominence in the experience of speakers – e.g., farmers who frequently drive animals for various purposes (1a) vs. travellers for whom horses are only a means to an end (1a').

(1) a *drive*   'make horses move (pulling a vehicle)'
     a' *drive*   'make a vehicle move (pulled by horses)'
     b *drive*   'travel in a vehicle'

These preferences are however not categorial, but instantaneous, subjective and dependent on the varying communicative conditions of each conversation. Yet, (1b) can only be a accounted for as a variant of (1a'), not of (1a), whereas only (1a) is compatible with the most ancient meanings of *drive*. The linguistic sign therefore requires a context-dependent flexibility which isomorphic approaches to meaning cannot capture. The analysis via recontextualization is therefore also more fine-grained than e.g. bridging contexts, where contextual factors are considered transitional and context-free coded meanings assumed to be the default.

In morphosyntactic change, analogical recontextualization can be observed in cases of different (but similar) settings of the communicative situations (2).

(2) a *Let us go in.*   'let her and me / you and me go in.'
     b *Lets go in.*   'let you and me go in.'
     c *Lets fight it out.*   'You and me / you both fight it out'

The difference between the two readings of (2a) applies solely to whether the addressee is included into the request or not. Again they describe the same event, but in different situational configurations. The higher frequency of the communicative settings (not of the construction) in which a request includes rather than excludes the addressee leads to the phonetic reduction only in the former situation type (2b) – and ultimately to /lets/ becoming an independent imperative marker (2c).

While compatible with common notions of language change (metaphor, bridging contexts, reanalysis), 'recontextualization' has the advantage of focusing on specific, context-driven communicative acts. Moreover, it links cognitive speaker activity with patterns that are

observed outside the study of human language. It therefore contributes to explaining linguistic structures on the basis of an interactive, emergent approach to language.

References:

Evans, Nicholas & David Wilkins, (2000), In the mind's ear: The semantic extensions of perceptions verbs in Australian languages, *Language* 76, 546-592.

Hopper, Paul J., (1987), Emergent Grammar, in Jon Aske, Natasha Beery, Laura Michaelis, Hana Filip (eds.), *Berkeley Linguistic Society. Proceedings of the Thirteenth Annual Meeting, February 14-16, 1987: General Session and Parasession on Grammar and Cognition*, Berkeley: Berkeley Linguistic Society, 139-157.

Heine, Bernd, (2002), On the role of context in grammaticalization, in: Ilse Wischer, Gabriele Diewald, *New Reflections on Grammaticalization*, Typological Studies in Language 49, Amsterdam / Philadelphia, Pa.: Benjamins. 83-101.

Kuhle, Anneliese, (2019), *Tool Intelligence as an Explanation of Cross-Linguistic Variation and Family Resemblance: An Evolutionary and Typological Investigation*, Lanham, My.: Lexington.

von Mengden, Ferdinand & Anneliese Kuhle, (2020), Recontextualization and language change, *Folia Linguistica Historica* 41, 253-281.

How analogy drives language comprehension:
Interpreting novel noun-noun compounds

Giulia Rambelli & Marianna Bolognesi
(University of Bologna)

Understanding novel creative expressions is a central issue in linguistics. For instance, how do people interpret expressions such as *avocado chair*? Such combinations are used every day, and people typically construct plausible interpretations even without contextual clues (Lynott & Connell, 2010). Indeed, combining two lexemes to form a new word, i.e., **compounding** (Bauer, 2017), represents a productive linguistic process, yet how people derive meanings for novel compounds remains unclear, especially because their meaning cannot be constructed through compositional mechanisms driven by superficial structure (Jackendoff, 2016).

A promising answer is offered by usage-based constructionist approaches. The core tenet of these theories is that language consists of constructions—form-meaning pairings of varying schematicity and complexity, which emerge and are stored through repeated exposure (Goldberg, 2019; Hoffmann, 2022; Ungerer, & Hartmann, 2023). Within this framework, analogy is considered the cognitive process that serves to generate novel utterances based on prior linguistic experiences (Bybee, 2010; Rambelli, 2024). That is, instead of interpreting an utterance word-by-word, compositionally, speakers map existing structural patterns (constructions) onto new instances, leveraging previous linguistic experiences (Bybee & Moder, 1983; Baayen, R. H., 2003; Ambridge, 2020).

The present study investigates how analogy guides the interpretation of novel noun-noun compounds. For example, interpreting

> *birthday dessert*

involves accessing the concepts denoted by the words and selecting an implicit relation ('a dessert intended for birthday') to form a unified conceptual representation (Gagné & Spalding, 2006). How is the relation selected? Prior research suggests that people rely on past experiences with similar combinations and extend them through **analogy** (Gagné, 2001; Gagné & Shoben, 1997; Krott, 2009). For example, *mud man* could be paraphrased as a man 'who delivers mud' (by analogy with *milk man*) or 'who collects mud' (by analogy with *garbage man*; van Jaarsveld, et al., 1994).

However, no systematic analysis has explored how analogy specifically drives novel compound interpretation. To address this challenge, we present a dataset of English novel compounds created from lexicalized compounds (i.e., the analogical base) taken from Tratz (2011) and Muraki et al. (2024). Novel combinations are created by systematically replacing the head or the modifier of the lexicalized compounds with their corresponding synonyms and hypernyms. We collect human judgments to assess: (i) the cognitive effort involved in processing novel compounds and (ii) how consistently participants converge on a specific interpretation (Schmidtke et al., 2016). We expect that greater semantic distance between component words of lexicalized and novel compounds will increase processing difficulty and lead to more varied interpretations. We also explore how linguistic properties of compounds influence the interpretation, such as concreteness. Finally, we evaluate how recent

large language models exhibit analogical behaviour in interpreting novel compounds (Coil & Shwartz, 2023; Rambelli et al., 2024).

The results will serve as a basis for a broader discussion of analogy as a cognitive mechanism underlying **linguistic productivity**, specifically by examining how it facilitates the interpretation of unfamiliar noun-noun compounds in both humans and large language models, thereby shedding light on the parallels and divergences in their processing strategies.

Acknowledgments

References

Ambridge, Ben (2020), Against Stored Abstractions: A Radical Exemplar Model of Language Acquisition. *First Language*, 40 (5–6): 509–59.

Baayen, R. Harald (2003), Probabilistic approaches to morphology. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 229-287), Cambridge: MIT Press.

Bauer, Laurie (2017), *Compounds and Compounding*, Cambridge: Cambridge University Press.

Bybee, Joan L., & Moder, Carol Lynn (1983), Morphological classes as natural categories, *Language*: 251-270.

Bybee, Joan L. (2010), *Language, Usage and Cognition*, Cambridge University Press.

Coil, Jordan & Shwartz, Vered (2023), From chocolate bunny to chocolate crocodile: Do language models understand noun compounds?, in *Findings of the Association for Computational Linguistics: ACL 2023*: 2698–2710.

Goldberg, Adele E. (2019), *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*, Princeton, NJ: Princeton University Press.

Hoffmann, Thomas (2022), *Construction grammar: The structure of English*, Cambridge: Cambridge University Press.

Jackendoff, Rey (2016), English noun-noun compounds in Conceptual Semantics, in P. Ten Hacken (Ed.), *The Semantics of Compounding* (1st ed., pp. 15–37), Cambridge: Cambridge University Press.

Krott, Andrea (2009), The Role of Analogy for Compound Words. In James P. Blevins & Juliette Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 18–136), Oxford: Oxford University Press.

Gagné, Christina L. (2001), Relation and lexical priming during the interpretation of noun–noun combinations, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1): 236–254.

Gagné, Christina L., & Shoben, Edward J. (1997), Influence of thematic relations on the comprehension of modifier–noun combinations, *Journal of experimental psychology: Learning, memory, and cognition*, 23(1): 71-87.

Gagné, Christina L., & Spalding, Thomas L. (2006), Using Conceptual Combination Research to Better Understand Novel Compound Words, *SKASE Journal of Theoretical Linguistics*, 3(2): 9–16.

Lynott, Dermot & Connell, Louise (2010), Embodied Conceptual Combination, Frontiers in Psychology, 1:212.

Muraki, Emiko J., Abdalla, Summer, Brysbaert, Marc, & Pexman, Penny M. (2023), Concreteness Ratings for 62,000 English Multiword Expressions, *Behavior Research Methods*, 55(5): 2522–2531.

Rambelli, Giulia (2024), Constructions and Compositionality: Cognitive and Computational Explorations, Cambridge: Cambridge University Press.

Rambelli, Giulia, Chersoni, Emmanuele, Collacciani, Claudia, & Bolognesi Marianna (2024). Can Large Language Models Interpret Noun-Noun Compounds? A Linguistically-Motivated Study on Lexicalized and Novel Compounds, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*: 11823–11835.

Schmidtke, Daniel, Kuperman, Victor, Gagné, Christina L., & Spalding, Thomas L. (2016), Competition between conceptual relations affects compound recognition: the role of entropy, *Psychonomic Bulletin & Review*, 23(2): 556–570.

Tratz, Stephen (2011), *Semantically-enriched Parsing for Natural Language Understanding*. Ph.D. thesis, University of Southern California.

Ungerer, Thobias & Hartmann, Stefan (2023), *Constructionist Approaches: Past, Present, Future*, Cambridge: Cambridge University Press.

van Jaarsveld, Henk J., Coolen, Riet, & Schreuder, Robert (1994), The role of analogy in the interpretation of novel compounds, *Journal of Psycholinguistic Research*, 23(2): 111–137.

Phonotactic similarities and relational analogies
in the acquisition of the dative case in German

Heike Behrens
(University of Basel)

German nouns are inflected for number, gender and case. Only number is marked on the noun, gender and case on determiners or pronouns. Number and gender are acquired in the third year of life, based on perceivable surface analogies between the phonotactics of the noun and its plural ending, and transitional probabilities between determiners and the noun. Case marking, in particular the dative, is harder to learn and more error prone, and takes well into school-age (Szagun 2019).

Case is an abstract relational category with different overt morphological realizations that cannot be acquired by surface analogies alone. Overt encoding differs on determiners, pronouns and adjectives, and their gender, number, and (in)definiteness. In usage-based approaches, complex relational are supposed to be learned bottom-up in a network-like fashion, based on local analogies. Extended trajectories of item-specific patterns with local generalizations are expected (Ambridge 2020a,b), before more variability provides evidence for network relations within the system that allow for top-down applications of inductively derived schemata.

The trajectory of different forms of dative marking are analyzed in longitudinal and morphologically annotated corpora of six German children (age 1;11-7) and their input. The total corpus-size is 3.5 million words, and includes 128000 words (determiners, adjectives, (pro-)nouns) in dative contexts by the adults, and 33500 by the children). Dative is first encoded as lexical prepositional case, and develops in an item-specific fashion with just a few prepositions initially, before it is extended to other prepositions and more variable instantiations of the dative object. Analyses of the few high frequency verbs requiring the dative (like *give* or *take*) that require the dative as structural case show that the object is often realized by unmarked proper nouns or a limited set of personal pronouns. Hence, the child is confronted with a system of vast variation regarding possible instantiations of the dative, but learns by complexity reduction by proper nouns and pronouns, which are less variable case and gender marked determiners.

A common measure for emerging morphological paradigms and syntactic network relations is increasing variability as shown by the overlap of different forms or lexical items in the same construction. In order to analyze what has to develop we will look at the later data of child(-directed) speech first: What is the endstate regarding dative acquisition in terms of lexical or constructional specificity or flexibility before formal education sets in? Second, what is the nature of the analogical links between these constructions, and how do they develop?

References

Ambridge, Ben, 2020a, Abstractions made of exemplars or 'You're all right, and I've changed my mind': Response to commentators, *First Language,* 40(5-6), 640-659.

Ambridge, Ben, 2020b, Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5-6), 509–559.

Szagun, Gisela, 2019, *Sprachentwicklung beim Kind: Ein Lehrbuch*, 7 edn*,* Weinheim/Basel: Beltz.

Stress position in English nouns and verbs – towards an analogy-based account

The semi-regular nature of lexical stress position is a well-known problem in English phonology. Although it is generally agreed that phonological (e.g. syllable weight), morphosyntactic (part-of-speech) and morphological properties of words play a role, it is also clear that these influences are far from categorical, and there is a lot of seemingly unpredictable variation in the lexicon. A large part of this variation is often attributed to historical developments in the English word stock, esp. the influx of a large number of Romance roots in Middle English, which have led to competition between Germanic-type and Romance-type stress assignment (e.g. Dresher & Lahiri 2022). Such accounts, however, cannot explain the apparent persistence of the variation. What is more, analogy-based computational models that are agnostic to abstract structural properties or historical strata have been surprisingly successful in predicting English stress assignment (Arndt-Lappe et al. 2023). How and on what basis this success is achieved, however, is not fully transparent, to the effect that it is difficult to relate what computational models do to previous generalisations based on structural features. In the present paper we will use a fully transparent computational analogical model ('AML', 'Analogical Modeling of Language', (Skousen, Stanford & Glenn 2013) to predict stress position in a large database of English verbs and nouns extracted from standard pronouncing dictionaries (Jones 2006; Wells 2008), which excludes words that have obvious stress-affecting morphology (N ≈ 7,000). This will allow us to test to what extent stress is predictable from analogy only, i.e. on the basis of recurrence of (strings of) sounds, without any explicit coding of more abstract phonological or morphological structure. We will then systematically analyse the structure of similarity relations that the algorithm deems relevant, and relate emerging generalisations from this to previous claims about structural factors influencing stress assignment. For example, AML's accuracy of stress classification in (previously unseen) verbs is between 84% and 94% (N = 3,033). Inspection of the sets of analogues used by the algorithm reveals that classification is heavily influenced by 'hubs' of highly similar lexical items, suggesting that the English lexicon is generally heavily skewed in terms of disproportionately frequent, recurring strings of segments, and that this fact is helpful for stress prediction. Many of such reoccurring hubs are based on the word-final syllable, which, for example, correspond to etymological Romance roots or Germanic suffixes. A parallel simulation study for nouns is currently underway. On a theoretical level we argue that considering synchronic stress assignment as based on analogy (recurrence) offers a straightforward explanation of how historical developments that have shaped the lexicon are still reflected in synchronic phonology.

References

Arndt-Lappe, Sabine, Robin Schrecklinger & Fabian Tomaschek. 2023. Stratification effects without morphological strata, syllable counting effects without counts – modelling English stress assignment with Naive Discriminative Learning. *Morphology* 33(4). 433–457. https://doi.org/10.1007/s11525-022-09399-9.

Dresher, B. Elan & Aditi Lahiri. 2022. Chapter 2. The foot in the history of English: Challenges to metrical coherence. In Bettelou Los, Claire Cowie, Patrick Honeybone & Graeme Trousdale (eds.), *Current Issues in Linguistic Theory*, vol. 358, 42–59. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.358.02dre.

Jones, Daniel. 2006. *Cambridge English Pronouncing Dictionary*. Cambridge: CUP.

Skousen, Royal. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer.

Skousen, Royal, Thereon Stanford & Nathan Glenn. 2013. *Algorithm::AM*. https://github.com/garfieldnate/Algorithm-AM.

Wells, John. 2008. *Longman pronunciation dictionary*. 3rd revised edition. Pearson Longman. http://www.worldcat.org/oclc/944210907.

Morphological gangs in verbal morphology: allomorphy and analogy in Italian

Sabine Heinemann
(University of Graz)

Verbs with a high token frequency often tend to have a high degree of irregularity (caused, for example, by unusual shortenings, which can lead to amorphous forms (cf. it. *ho*, *ha*), irregular assimilations or weakenings, etc.). Possible modifying changes are mainly caused by interparadigmatic analogies aiming at greater differentiation within the single paradigm with regard to individual forms, which thus achieve a greater degree of autonomy. Such developments are mainly found in certain tenses and modes (present, indicative; cf. Nübling 2000, Bybee 1985). However, multiple forms based on individual ones may also be involved (*gang effect*), whereby the verbs in question make up a cluster; the family similarity within such a cluster strengthens the microclass thus created, but at the same time distinguishes it from other verbs.

The contribution seeks to shed light on the diachronic development of the paradigms of the Italian verbs *andare* 'to go', *avere* 'to have', *dare* 'to give', *fare* 'to do', *sapere* 'to know', *stare* 'to be/stay' and the genesis of the respective microclass in (Old) Tuscan. It seems, that the verbs *dare* and *stare* are of particular importance for the microclass. However, this evidentially does not apply supra-regionally, as evidence such as ven. *dago* 'I give', *stago* 'I am/stay' etc. (cfr. *digo* 'I say') or old rom. *daco* 'they give', *staco* 'they are/stay' (cfr. *faco* 'they do') show. Thus, based on the microclass present in Tuscan (and to a certain degree in the standard language), the contribution aims to analyse the interdependence of the verbs mentioned also for two particularly interesting dialects, i.e. *romanesco* and *veneziano*. *Romanesco*, the dialect of Rome, originally showed a greater proximity to the southern Italian dialects (such as Neapolitan) and has been subject to a profound Tuscanisation since the 16th century, not least for socio-political reasons, which is why the developments that can be observed historically for *romanesco* are interesting. *Veneziano*, as the idiom of the Republic of Venice, instead reached a high level of development (*Ausbau*) with the rule of Venice and, as a prestigious idiom, overlaid the (not only Venetian) dialects of the dominion. Despite the Tuscanisation that was also visible as a result of the standardisation of Italian, *veneziano* was largely able to preserve many dialect-specific characteristics.

The diachronic development can be traced very well for Tuscan, and with limitations for *romanesco* and *veneziano*, via the data from the *TLIO* corpus (cfr. also *AGLIO* for the elaboration oft he data). In addition, early grammars are used and especially presentations with regard to (irregular) verbs which also refer to other than purely diatopically marked forms (cfr. among others Pistolesi 1761) are taken into consideration. For the modern dialects, apart from grammars, the explanations in various dialectological studies (cf. generally Rohlfs 1968) and the data of the AIS are important sources for the analysis.

References

AGLIO = Atlante Grammaticale della Lingua Italiana delle Origini, online: http://aglio.ovi.cnr.it.

Bybee, Joan L. (1985), *Morphology. A study of the relation between meaning and form*, Amsterdam/Philadelphia: Benjamins.

Nübling, Damaris (2000), *Prinzipien der Irregularisierung. Eine kontrastive Untersuchung zu zehn Verben in zehn germanischen Sprachen*, Tübingen: Niemeyer.

Pistolesi, Giovanni Batista (1761/²1813), *Prospetto di verbi toscani tanto regolari che irregolari*, Roma/Pisa: Capurro.

Rohlfs, Gerhard (1966/1968), *Grammatica storica dell'italiano e dei suoi dialetti*, vol. 1: *Fonetica*, vol. 2: *Morfologia*, Torino: Einaudi.

TLIO = *Corpus TLIO* (*Tesoro della lingua italiana delle origini*), online: http://tlioweb.ovi.cnr.it.

Analogy in innovation: The intersection of psycho- and historical linguistics

Szilvia Daczó

(Bielefeld University)

Keywords: analogy, processing, priming, innovation, language change

My presentation addresses how analogy as a cognitive mechanism provides a link between psycho- and historical linguistics. I demonstrate this through a case study on the novel *the hair is hairing* construction, which is formed on the basis of the pre-existing *the baby is sleeping$_V$/stunning$_{Adj}$* structure. This construction is unique due to the obligatory retention of the source noun(phrase) as the Agent, while its overt copy is part of the predicate.

After I collected ≈160 examples using the Internet, I investigated the construction in affirmative and negated contexts (*the hair is (not) hairing*) and with concrete and abstract parent nouns (*the mood is (not) mooding*).

My research questions are the following:

 (RQ1) Do different empirical methods reveal distinct findings about analogy, especially with regards to factors affecting analogy?

 (RQ2) What is the connection between analogy and priming?

I report the findings of a Qualtrics survey (n=147) and an in-person self-paced reading experiment (experimental=15, control=15). The former elicited acceptability ratings and multiple-choice responses besides open commentaries on the structure, meaning, and use of the construction. The latter elicited reaction times during exposure.

(RQ1)

The survey shows analogy being constrained by factors such as conscious processing, prescriptivism, (metalinguistic or self)-awareness, and social evaluation. These impacted analogy such that comparisons were made with the near-proportional *the baby is sleeping* alongside the *it's giving X* construction, but further analogical products were blocked. Conversely, the experiment shows analogical extension being facilitated by unconscious processing and psycholinguistic phenomena like surprisal, adaptation, and crucially, priming.

These indicate that divergent methods show opposing but reconcilable results, painting a more complete picture about analogy. Taken together, the results show that the amount of previous exposure indirectly affects the recourse to analogy. Furthermore, I argue that analogues, analogization, and analogical change are supported by bottom-up automatic psycholinguistic phenomena like structural priming (De Smet, 2018), but top-down external factors are non-negligible either.

Moreover, the style of elicitation may bring about contrasting implications for change: the results of the survey imply that the propagation of the investigated construction will stop, while the experiment indicates that this innovation can spread. Therefore, off-line and on-line responses should be considered side-by-side.

(RQ2)

There are various identifiable connections between analogy and priming. They share characteristics such as structural-relational alignment or imitation, and both transfer/repeat a mapping from base/prime to target (analogy: Holyoak & Thagard, 1995; priming: Pickering & Ferreira, 2008). For instance, *the baby is sleeping > the hair is hairing* (actuation) and *the hair is hairing > the internet is interneting* (actualization). Additionally, both are conservative (reuse a structure or proportion) *and* supportive of creativity (generalisation to previously unattested exemplars) (analogy: Bergs, 2018; analogy/priming: De Smet, 2016). Crucially, by increasing availability, priming facilitates the access to analogy (viz. Leech et al., 2008). By implication, this supports Walkden's (2021) claim that analogy as a traditional mechanism of language change is in itself not explanatory.

In conclusion, while the mind may utilise analogy in innovation, the inclusion of elementary psycholinguistic processes like priming provide a more complete conceptualisation of this prominent mechanism in both synchrony and diachrony.

(word count: 500)

References

Bergs, A. (2018). Learn the Rules like a Pro, so you can Break them like an Artist (Picasso): Linguistic Aberrancy from a Constructional Perspective. *Zeitschrift für Anglistik und Amerikanistik*, 66(3), 277–293.

Daczó, Sz. (2024). *'My hair is hairing today': an exploratory case study and the assessment of how psycholinguistic mechanisms are involved in language change* [Unpublished MPhil dissertation]. University of Cambridge.

De Smet, H. (2016). How gradual change progresses: The interaction between convention and innovation. *Language Variation and Change*, 28, 83–102.

De Smet, H. (2018). Unwitting Inventors: Speakers Use -ly-Adverbs More Creatively when Primed. *Zeitschrift für Anglistik und Amerikanistik*, 66(3), 329–340.

Holyoak, K. J., & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: MIT Press.

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioural and Brain Sciences,* 31(4), 357–414.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3), 427–459.

Walkden, G. (2021). Against Mechanisms: Towards a Minimal Theory of Change. *Journal of Historical Syntax*, 5(33), 1–27.

How much can analogy explain? A survey study

The historical linguistics literature holds extremely diverse views on the role of analogy in language and language change, ranging from near-dismissal (e.g. Lass 1997; Lehmann 2004; Newmeyer 2014) to full and enthusiastic embrace (e.g. Itkonen 2005; Fischer 2007; Ambridge 2020). Much of the discussion, however, revolves around isolated (and potentially cherry-picked) case studies, or theoretical argumentation based more in principle or non-historical evidence than in systematic historical evidence. One way to properly assess the involvement of analogy in language change is through larger-scale studies that incorporate and compare larger numbers of changes. This paper reports on the results of such a survey study on grammaticalization in English.

Analogy can be informally defined as the mechanism by which the grammatical behaviour of an item or construction changes under the influence of similar items or constructions, with increased similarity as a result. Our survey study was conducted to obtain a large and unbiased sample of changes, in which we could systematically check for likely cases of analogically-motivated grammatical change, as well as counterexamples, consistently applying a single definition of analogy. The survey was based on a collection of changes collected from the existing literature on grammaticalization in English. Grammaticalization was chosen because of the particularly contentious role of analogy in this area of change (Meillet 1912 famously distinguished between two types of change: grammaticalization and analogy). The selection of changes was compiled by sampling the *Linguistics and Language Behavious Abstracts* database. On the most conservative count, 33 instances of grammaticalization were identified. For each of these cases, we tested whether observed changes in the formal behaviour of grammaticalizing expressions did or did not reflect formal behaviour in contemporary existing and functionally similar expressions.

The results of the survey show that analogy is most apt at explaining how grammaticalizing items adopt innovative syntactic behaviour, as opposed to the loss of older behaviours through decategorialization. Both positive evidence (the occurrence of innovative syntactic behaviour) and negative evidence (the absence of innovative syntactic behaviour) point in this direction. At the same time, some changes show other sources of innovative syntactic behaviour. Three types emerge from the survey. Sometimes innovative syntactic behaviour probably results from blending. In other cases, innovative syntactic behaviour seems to be linked to collocational expansion. A final group of changes are reminiscent of pragmaticalization and involve what appears to be extreme productivity. The first two types

could arguably still be conceived of as variants of analogical change. The last type is probably impossible to reconcile with analogy-based accounts, but fits well with the notion of co-optation (Kaltenböck et al. 2011; Heine 2013).

Ambridge, Ben. 2020. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 40(5-6): 509-559.

Fischer, Olga. 2007. *Morphosyntactic Change: Functional and Formal Perspectives.* Oxford: Oxford University Press.

Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6): 1205-1247.

Itkonen, Esa (2005). *Analogy as Structure and Process: Approaches in Linguistics, Cognitive Psychology and Philosophy of Science*. Amsterdam: Benjamins.

Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35(4): 848-893.

Lass, Roger. 1997. *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.

Lehmann, Christian. 2004. Theory and method in grammaticalization. *Zeitschrift für Germanistische Linguistik* 32: 152-87.

Meillet, Antoine. 1912. L'évolution des formes grammaticales. *Scientia* 12(26): 130-148.

Newmeyer, Frederick. 2014. Where do motivations compete?, in Brian MacWhinney, Andrej Malchukov, and Edith.A. Moravcsik (eds), *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press, 299-314.

WS22 Transitivity and labile verbs in typological and diachronic perspectives: Indo-European and beyond

Leonid Kulikov, Tim Ongenae & Daria Chistiakova

P-lability in Albanian: An attempt of a diachronic study

Alexander Rusakov & Maria Morozova

(Institute of Language and Literature, Academy of Sciences of Albania, Tirana & Institute for Linguistic Studies, Russian Academy of Sciences, Saint Petersburg)

Keywords: Albanian, P-lability, diachrony, lexical semantics, Albanian National Corpus, early Albanian monuments

Albanian, like other Balkan languages (except for Romani dialects), belongs to the group of languages where decausative derivation predominates over causative derivation (Nichols 2020). It has also been noted that Balkan languages are characterized by the phenomenon of lability, i.e. the same verb can be used transitively and intransitively without changing its form (Friedman & Joseph 2025). Recent studies attempt to rank the Balkan languages according to the prevalence of this phenomenon. In this ranking, Macedonian and Greek occupy the top positions, while Albanian is among the languages with the least representation of lability (Bužarovska & Mitkovska 2022, see also Makartsev et al. 2024).

Meanwhile, the number of labile verbs in Albanian is quite substantial. It is worth mentioning that for four out of the 18 pairs of contrasting transitive-intransitive verbs in the diagnostic list by Nichols et al. (2004), Albanian *can* use ambitransitive verbs (*die – kill, learn – teach, boil – boil, sleep – be asleep*). Dhrimo (1965) mentions in modern standard Albanian ca. 30 verbs that can be employed both transitively and intransitively. The only existing study of lability in Albanian provides a list of about 50 Albanian verbs demonstrating lability (Diveeva 2013). In our work, we primarily address P-lability in Albanian as exemplified in (1).

(1) a.   *ata   vdis-nin*
           they   die-IPF.3PL
           'they were dying'

     b.   *e=rrah-u   sa   e=vdiq*
           3SG.ACC=beat-PST.3SG   as.much   3SG.ACC=kill.PST.3SG
           'he beat him so that he killed him'

The most significant groups demonstrating P-lability in Albanian are phasal verbs and motion verbs, with a notable group of verbs exhibiting conversive lability also present.

There are no studies examining the diachrony of Albanian lability, and the present study aims to fill this gap. Two temporal "snapshots" are compared: the language of early Albanian monuments from the 16th-17th centuries and the contemporary Albanian language of the 20th-21st centuries. The first is obtained through a comprehensive analysis of texts from these monuments using the existing verb concordances, e.g. (Schumacher & Matzinger 2014). The second is based on the test version of the Albanian National Corpus, comprising over 110 million words (Morozova et al. 2024). For all verbs in our diagnostic list based on Dhrimo (1965) and Diveeva (2013), both transitive and intransitive usages are examined, along with the presence of parallel morphologically inactive forms. For older texts, a comprehensive sampling method is used, while for the contemporary language the proportion for the first 300 randomly ordered examples is counted.

The proposed approach allows for a certain degree of determination of the dynamics of lability development in Albanian and the direction of semantic derivation for different groups of labile verbs. Preliminary results show that at the observable stages of the development of Albanian, labile verbs

exhibit a degree of variability, e.g. verbs that used to be primarily transitive may increase the number of intransitive usages (*afroj* 'bring closer; get closer' is used as a transitive verb in Buzuk's, Budi's, and Variboba's writings with the corresponding nonactive form expressing the decausative meaning, while in modern Albanian the active form may be used both transitively and intransitively). The most stable group are phasal verbs like *filloj* 'begin' and *pushoj* 'stop', demonstrating both transitive and intransitive usages at all stages.

References

Bužarovska, Eleni & Liljana Mitkovska (2022), *On the lability of verbs in Balkan languages*, Paper presented at the 22nd Balkan and South Slavic Conference*,* The Ohio State University, 7–10 April 2022.

Dhrimo, Ali (1965), Një vështrim mbi foljet me më se një reksion në shqipen e sotme letrare, *Studime filologjike* 2, 139–151.

Diveeva, Varvara (2013), *Labil'nost' albanskikh glagolov v balkanskom kontekste*, Course paper, Saint Petersburg State University.

Friedman, Victor A. & Brian D. Joseph (2025), *The Balkan Languages*, Cambridge: Cambridge University Press.

Makartsev, Maxim, Max Wahlström & Anastasia Escher (2024), Lability in Balkan Slavic, *Studies in Language*, Online first. https://doi.org/10.1075/sl.23062.mak

Morozova, Maria, Alexander Rusakov & Timofey Arkhangelskiy (2024), *Albanian National Corpus*. Available online at: albanian.web-corpora.net.

Nichols, Johanna (2020), The Balkan sprachbund in typological-geographical space, *Acta Linguistica Petropolitana* 16(2), 306–330.

Nichols, Johanna, David A. Peterson & Jonathan Barnes (2004), Transitivizing and detransitivizing languages, *Linguistic Typology* 8(2), 149–211.

Schumacher, Stefan & Joachim Matzinger (2014), *Die Verben des Altalbanischen: Belegwörterbuch, Vorgeschichte und Etymologie*, Wiesbaden: Harrassowitz Verlag (Albanische Forschungen, Band 33).

Valency Change in Chinese: The Case of the (Labile) Verb 敗 Bài

Anni Wang

(Ghent University)

Keywords: historical Chinese, labile verbs, diachronic change, argument structure, Chinese morphosyntax

Labile verbs are a notable feature of Chinese, yet their study has only recently gained researchers' attention (e.g., Basciano 2010, Zhang 2017, 2019, and Du, Zuo & Li 2024). Even less is known about the mechanisms behind the development of Chinese labile verbs. This paper takes the labile verb *bài* (敗) 'to lose~cause to lose' and investigates the diachronic shift in its transitivity from Old Chinese to Modern Chinese. The study focuses on three research questions: (1) What are the diachronic patterns of transitivity in *bài*? (2) How can existing frameworks, particularly Ramchand's (2008) framework using a syntactic approach to argument structure (as applied to diachronic change by Meisterernst 2023), explain these patterns? and (3) How do the diachronic patterns align with broader changes in Chinese morphosyntax?

Using data from the Academia Sinica Ancient Chinese Corpora (2021) and CCL Corpus (Zhan et al. 2019), this study combines a corpus-based diachronic approach with syntactic analysis to investigate the verb *bài* in four historical stages of Chinese, including Old Chinese (11th c. BCE–2nd c. CE), Middle Chinese (2nd c. CE–10th c. CE), Early Mandarin Chinese (11th c. CE–19th c. CE), and Modern Chinese (20th c.–present). Morphophonological factors in Old Chinese are considered. A normalized and randomized set of samples from each stage was extracted and manually coded for multiple linguistic features, such as transitive and intransitive uses, serial verb constructions, resultative compounds, and collocations with aspect markers. This study, with data from an extended historical period, provides insights rarely explored in existing research on valency change in Chinese.

The data demonstrates a significant decline in the verb's transitive use, accompanied by an increase in its intransitive usage. Additionally, the use of serial verbs and verbal compounds involving *bài* rises as its valency decreases, especially since the Early Mandarin Chinese period, mirroring broader syntactic changes in the language, such as the emergence of analytic structures, and in particular periphrastic causative constructions replacing originally labile verbs such as *bài*.

The paper argues, based on the structure proposed by Ramchand (2008), that the decrease in transitivity of *bài* can be viewed as a loss of V-to-v movement, originally giving rise to lability (Fig. 1a), and the lexicalization of v by a causative light verb instead (e.g., 打 *dǎ* or 擊 *jī*, 'to hit'), to form bisyllabic compounds like 打敗 *dǎbài* or 擊敗 *jībài* ('to defeat') (Fig. 1b). Syntactic change and grammaticalization resulting from the loss of syntactic movement is a well-discussed mechanism (e.g., Roberts & Roussou 2003, McFadden 2015 for Indo-European languages, and Meisterernst 2023 for the Chinese language). However, the dynamics between valency shifts and morphosyntactic change are so far rarely discussed. This paper seeks to address this gap by examining the specific case of *bài*, offering insights into how these dynamics played out in the diachrony of Chinese.
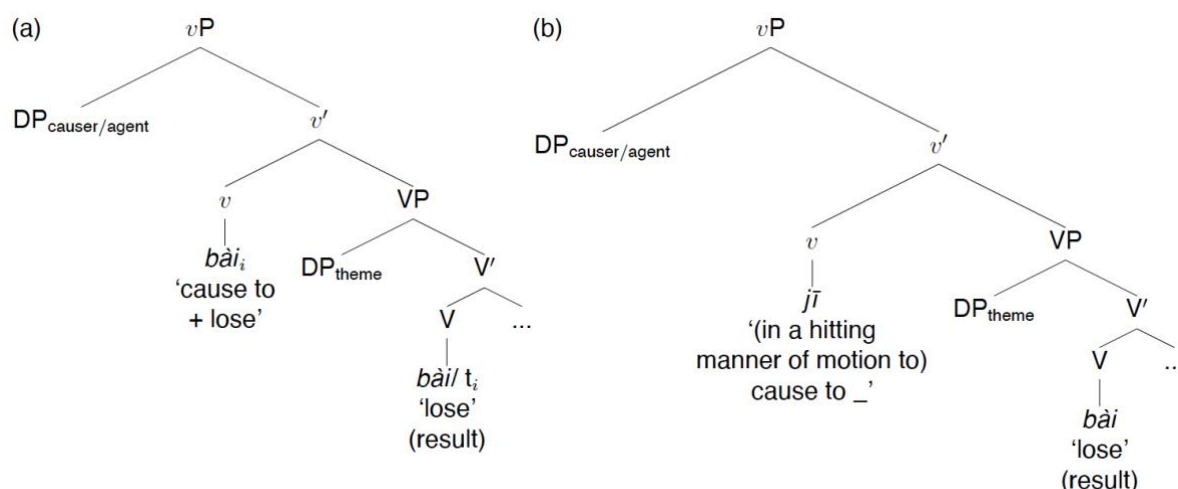
Fig. 1: Labile *Bài* as V-to-v-movement (a); Intransitive Bài and Causativization by Compounding (b)

References

Basciano, Bianca (2010), *Verbal compounding and causativity in Mandarin Chinese*. Doctoral dissertation, Ca' Foscari University of Venice.

Du, Jing; Zuo, Shan and Li, Fuyin Thomas (2024), Attraction or differentiation: diachronic changes in the causative alternation of Chinese change of state verbs. *Cognitive Linguistics*, *35*(4), 603-638.

Institute of Linguistics, Academia Sinica (2021), *Academia Sinica Ancient Chinese Corpora*. http://lingcorpus.iis.sinica.edu.tw/cgi-bin/kiwi/akiwi/kiwi.sh

McFadden, Thomas (2015), Preverbal ge-in Old and Middle English. *ZAS Papers in Linguistics*, *58*, 15-48.

Meisterernst, Barbara (2023), The loss of morphology and the emergence of analytic structures in Chinese. *Journal of Historical Syntax*, *7*(6-19), 1-55.

Ramchand, Gillian Catriona (2008), Verb meaning and the lexicon: A first phase syntax. *Cambridge University*.

Roberts, Ian and Roussou, Anna (2003), *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge University Press.

Zhan, Weidong; Guo, Rui; Chang, Baobao; Chen, Yirong and Chen, Long (2019), The building of the CCL corpus: Its design and implementation, *Corpus Linguistics, Vol.6, No.1, pp.71-86*

Zhang, Liulin (2017), *Lability of Verbs and the Change-of-State Construction in Chinese* [Doctoral dissertation, University of Hawai'i at Manoa].

Zhang, Liulin (2019), A study into the prototypicality of Chinese labile verbs. *Cognitive Semantics*, *5*(1), 1-31.

Residual lability in Baltic

Axel Holvoet

(Vilnius University)

The Baltic branch of Indo-European, which has close genealogical and areal links to Slavic (with which it is often assumed to share a common Balto-Slavic ancestor language), now comprises two languages, Lithuanian and Latvian. In the course of its separate history, Baltic has developed a strong tendency to provide transitivity oppositions with formal marking. This manifests itself, e.g., in the reappropriation of ablaut alternations originally connected with tense/aspect oppositions as transitivity markers, as in Lithuanian *kilti* 'rise' vs *kelti* 'raise' (Stang 1966: 331–333, 356). Additionally, Baltic has developed a rather robust and productive system of morphological causativisation (Lith. *vysti* 'wither' → *vyt-inti* 'cause to wither'), cf. Arkadiev & Pakerys (2015), Nau (2015). On the other hand, Baltic shares with Slavic the replacement of the IE middle with reflexive-marked verb forms, which yields a productive anticausative derivation (Latvian *pildīt* 'fill (tr.)' → *pildit-ies* 'fill (intr.)'). Baltic can thus be said to occupy an intermediate position between mainly intransitivising Slavic and mainly transitivising Fennic (Nau & Pakerys 2016). The result of these developments is that the domain of lability is now small and basically residual, i.e., labile verbs represent a state of affairs predating the rise of the above-mentioned marking devices. Moreover, in all cases involved one of the pair of labile verbs stands alongside a formally marked (anticausative or causative) verb that is functionally differentiated from the labile verb, e.g., alongside Lith. *virti* 'boil, cook (intr./tr.)' there is a causative *vir-inti* 'bring to boiling'. In the talk we will look at the lexical domain in which lability has proved resistant (taking into account not only a verb's position on the spontaneity scale as defined by Haspelmath 2016 but all parameters that could be relevant) and at the nature of the alternations between labile and formally marked verbs. In labile vs causative alternations, the factors involved are assumed to be related to, but not necessarily identical with, those involved the dictinctions between lexical and morphological (or morphological and syntactic) causitivisation (on which cf., e.g., Shibatani 2002), e.g., in pairs like (1)–(2) it is the nature of the causation that will be relevant:

(1) *Žmonės*   *deg-a*   *žibintus*.
     people.NOM.PL   burn-PRS.3   lantern.ACC.PL
     'People are burning lanterns.'

(2) *Žmonės*   *deg-in-a*   *šiukšles*.
     people.NOM.PL   burn-CAUS-PRS.3   garbage.ACC.PL
     'People are burning garbage.'

In labile vs anticausative alternations, the relevant factors should rather be sought in peripheral subdomains of middle/reflexive semantics, where, e.g., aspectual distinctions appear, as in (3)–(4):

(3) *Žvakė*   *deg-ė*   *ant*   *stalo*.
     candle.NOM.SG   burn-PST.3   on   table.GEN.SG
     'A candle was burning on the table.'

(4) *Žvakės*   *už-si-deg-ė*   *languose.*
     candle.NOM.PL   PFV-REFL-burn-PST.3   window.LOC.PL
     'Candles lit up in the windows.'

Whether the factors involved in the retreat of lability are different from those involved in its expansion, and to what extent lexical entrenchment plays a role here, remains to be established.

The data will be taken from the corpora of modern Lithuanian and Latvian and, for a diachronic perspective, from the data bases of 16th- and 17th-century texts for Lithuanian (https://seniejirastai.lki.lt/db.php) and Latvian (https://senie.korpuss.lv).

References

Arkadiev, Peter, and Jurgis Pakerys (2015), Lithuanian morphological causatives. A corpus-based study, in A. Holvoet, and N. Nau (eds), (2015), *Voice and Argument Structure in Baltic*, Amsterdam / Philadelphia: John Benjamins, 39–97.

Haspelmath, Martin (2016), Universals of causative and anticausative verb formation and the spontaneity scale, *Lingua Posnaniensis* 58(2), 33–63.

Nau, Nicole (2015), Morphological causatives in contemporary Latvian, in A. Holvoet, and N. Nau (eds), (2015), *Voice and Argument Structure in Baltic*, Amsterdam / Philadelphia: John Benjamins, 99–145.

Nau, Nicole, and Jurgis Pakerys (2016), Transitivity pairs in Baltic: Between Finnic and Slavic, *Lingua Posnaniensis* 63(2), 83–126.

Shibatani, Masayoshi (2002), Some basic issues in the grammar of causation. in M. Shibatani (ed), *The Grammar of Causation and Interpersonal Manipulation,* Amsterdam / Philadelphia: John Benjamins, 1–21.

Stang, Christian S. (1966), *Vergleichende Grammatik der baltischen Sprachen,* Oslo–Bergen–Tromsö: Universitetsforlaget.

Lability in Eastern Iranian languages

Arseniy Vydrin & Daria Chistiakova

(Institute of Linguistics, RAS & University of Liège; KU Leuven)

This paper presents an analysis of lability in some Eastern Iranian languages, with a focus on Ossetian, Yaghnobi, and the Shughni-Rushani languages of the Pamir group. The study offers insights into the diachronic and morphosyntactic dynamics of lability in Eastern Iranian languages, contributing to a broader understanding of typological patterns in lability across languages.

For example, in the Shughni-Rushani group, non-canonical lability emerges, fitting the definition of partial lability (Letuchiy 2013: 57). In past tenses, verbs agree with the subject with second-position agreement clitics, while verb stems in past and perfect remain invariable. These languages have 17 verb pairs that share identical infinitives and past/perfect stems, though present forms often differ (see Table 1 for Rushani). Additionally, while transitive verbs do not show gender agreement with the subject, intransitive verbs in the past tense agree with the subject in a feminine and a plural form (see ex.1a for transitive usage, 1b-1c for intransitive usages). For Shughni, this phenomenon has led to the proposal of the term "infinitive-labile verbs" (Parker 2023: 383).

(1) Shughni (Indo-European, Eastern Iranian, Shughni-Rushani group)

a) *wisêrn=um*   ***viruǎ-t***
    jug=1SG   break-PST
    'I broke the jug' (Makarov et al. 2022).

b) *yu*   *būtal*   ***viruǎ-t***
    D3.M.SG   bottle   break-PST
    'The bottle broke' (Makarov et al. 2022).

c) *yā*   *mošin* ***viraǎ-t***
    D3.F.SG car   break.F-PST
    'The car broke down' (elic. 2025).

Table 1. Paradigm for *viriǎtow* 'to break' (itr/tr) in Rushani (Indo-European, Eastern Iranian, Shughni-Rushani group), based on Karamshoev (1978: 152), Sokolova (1959).

	Rushani	
	intransitive	transitive
Present stem	*viraw-*	*virand-*
Past (masc)	*viruǎt*	
Past (fem/pl)	*viraǎt*	*viruǎt*
Perfect (masc)	*viruǎč*	
Perfect (fem)	*viriǎc*	*viruǎč*

Perfect (pl)	*virax̌č*	

In 7 of 17 labile verbs in Shughni-Rushani, a morphological causative coexists with the labile form in transitive contexts. In example (2), 2a shows a causative verb, 2b a labile verb used transitively, and 2c the same verb used intransitively.

(2) Bartangi (Indo-European, Eastern Iranian, Shughni-Rushani group)

a) *mun   yax   az   um   tač   firaw-d*
   my   sister   DOBJ   D3.F.SG.OBL   bowl   wash.CAUS-PST
   'My sister washed this bowl' (elic. 2025).

b) *xu   ǰoyga=yum   firu-d*
   REFL   dishes=1SG   wash-PST
   'I washed the dishes' (elic. 2025).

c) *ta   ǰoyga   na-firu-d*
   your   dishes   NEG-wash-PST
   'Your dishes didn't get clean' (elic. 2025).

In Shughni-Rushani languages, past stems and infinitive forms mostly derive from Proto-Iranian perfect participles, with *virix̌tow* specifically tracing back to Proto-Iranian **brušta-* (Rastorgueva, Edelman, 2000: 175). The matching past stems alongside divergent present-tense forms suggest the development of **secondary lability**, based on participial forms — a unifying-type lability characteristic of non-finite forms (Letuchiy 2013: 61).

However, in modern Shughni-Rushani languages, many labile verbs have become obsolete or reanalyzed:

- seven verbs retain both transitive and intransitive uses;
- six are reanalyzed as either only transitive or only intransitive;
- four with an additional causative form retain only the causative form.

The reduction of newly formed lability aligns well with a trend observed in Iranian languages, where the number of labile verbs is steadily decreasing (Kulikov et al. 2024). It can be argued that alongside this general decline, secondary partial lability – developed in the non-finite verb forms of Pamir languages – is also disappearing.

In Ossetian, transitivity is overtly marked in the counterfactual mood with the suffix *-t*, and in the past tenses with the suffix *-t* combined with different sets of personal endings. Some labile verbs have both conjugation types, e.g., *žilən* 'to spin': itr. *žəld*, tr. *žata*. Most verbs, however, use either the transitive or intransitive type in all contexts, for example, *tašən* 'to bend' uses the transitive form *tašəd* in both transitive and intransitive contexts.

In the talk, more data on lability in Ossetian and Yaghnobi will also be presented.

Abbreviations

CAUS – causative; D – demonstrative; DOBJ – direct object marker; F – feminine; M – masculine; NEG – negation; OBL – oblique case; PL – plural; PST – past; REFL – reflexive; SG – singular.

References

Karamshoev, Dodkhudo (1978), Kategoriya roda v pamirskikh yazykakh (shugnano-rushanskaya gruppa) [The category of gender in Pamir languages (Shughni-Rushani group)]. Dushanbe: Donish.

Davari, Shadi & Kulikov, Leonid & Navid Naderi (2024), Decline of lability in Iranian: A diachronic typological perspective. SLE conference, Helsinki, 2024.

Letuchiy, Alexander (2013), Tipologiya labil'nykh glagolov [Typology of Labile Verbs]. M.: Yazyki slavyanskoy kultury.

Makarov, Yury & Melenchenko, Maksim & Dmitry Novokshanov (2022), Digital Resources for the Shughni Language. Proceedings of The Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-Resourced Languages in Eurasia within the 13th LREC, 61–64.

Parker, Clinton (2023), A Grammar of the Shughni Language. PhD Thesis, McGill University.

Rastorgueva, V.S. & J.I. Edelman (2000), Etymologicheskiy slovar' iranskikh yazykov. T.1 [Etymological Dictionary of Iranian Languages. Vol.1]. Institut yazykoznaniya RAN, M.: "Vostochnaya literatura", 2000.

Sokolova, V.S. (1959), Rushanskie i khufskie teksty i slovar' [Rushani and Khufi Texts and Dictionary]. Institut yazykoznaniya. M.-L.: Izdatel'stvo AN SSSR.

Lability drift in Modern Aramaic languages

Elena Shvedova

(HSE University of Moscow, Linguistic Convergence Laboratory)

This paper studies labile verbs in Modern Aramaic languages (<Semitic <Afro-Asiatic) from a diachronic and typological perspective. Labile verbs can be used both transitively and intransitively without morphological change, as in example (1) from Christian Urmi (<North-Eastern Neo-Aramaic).

(1) a. *beta* *tləx-lə*
         house   break.PST-LS.3M
         'The house collapsed'.
     b. *ginavə*   *tləx-le*   *beta*
         robbers   break.PST-LS.3PL   house
         'The robbers destroyed the house'.

I compare the encoding of the transitivity opposition in several Aramaic languages of different periods and show that labile verbs are widespread in different branches of Modern Aramaic, in contrast to earlier Aramaic varieties where anticausative or causative marking was more common. I also classify Modern Aramaic labile verbs into three groups based on their historical development.

In recent decades, several typological works on the classification of labile verbs have appeared (e.g. Letuchiy 2013; Creissels 2014), but only a few of them are diachronically oriented, so there are still many unresolved questions about the evolution of labile verbs (Kulikov & Lavidas 2014). As far as Modern Aramaic languages are concerned, previous studies on labile verbs focus on individual varieties, e.g. Mengozzi (1999) and Göransson (2015) on North-Eastern Neo-Aramaic, Fassberg (2021) on Modern Western Aramaic and Kuzin (2024) on Ṭuroyo, leaving a gap in comparative and diachronic analysis that my study aims to fill.

Neo-Aramaic languages can be divided into two main genealogical groups: Eastern and Western Aramaic, which separated during the first millennium BC. Using Haspelmath's (1993) methodology, I examined the transitivity profiles of four Eastern Modern Aramaic languages (Christian Urmi, Bohtan, Maməday and Ṭuroyo), Modern Western Aramaic (MWA) and two Middle Aramaic languages (Christian Palestinian Aramaic and Classical Syriac). The translations of 31 causal-noncausal verb pairs into these languages were elicited from native speakers or obtained from corpora and dictionaries.

I found that some verbs, such as 'freeze', 'fill' and 'begin', were already labile in earlier varieties of Aramaic, suggesting that their lability is a **shared retention** in both Eastern and Western Modern Aramaic varieties. The second group includes the verbs 'open', 'break', 'close', 'split', 'spread', 'rock' and 'roll'. They had anticausative marking in Middle Aramaic languages, but are coded by labile verbs in most Modern Aramaic varieties, indicating a **parallel development** or **drift** in the two branches. The third group of labile verbs is unique to MWA and includes the verbs *a-ʁəl* 'boil', *naʃʃaf* 'dry' and *a-rkeʃ* 'wake up'. In other Middle and Modern Aramaic languages these meanings are expressed in causatively marked pairs, so the lability of these verbs in MWA is a **morphological innovation**.

In my talk, I will present a more detailed corpora- and dictionary-based analysis of the verbal systems of Christian Urmi and MWA, focusing on the semantics and morphosyntax of labile verbs. I will also propose some explanations for the lability drift in Modern Aramaic languages, such as the phonetic loss of the anticausative marker, the expansion of verbs with four root consonants that cannot be causativized, and possible areal factors.

References

Creissels, Denis. 2014. P-lability and radical P-alignment. *Linguistics* 52.4: 911-944.

Göransson, Kathrin. 2015. Causative-Inchoative Alternation in Northeastern Neo-Aramaic. In *Neo-Aramaic and Its Linguistic Context*, Lidia Napiorkowska (ed), 207-231. Piscataway: Gorgias Press.

Haspelmath, Martin. 1993. More on the Typology of Inchoative/Causative Verb Alternations. In *Causatives and Transitivity*, Bernard Comrie and Maria Polinsky (eds), 87-120.

Kulikov, Leonid & Nikolaos Lavidas. 2014. Introduction (Typology of labile verbs: Focus on diachrony). *Linguistics* 52.4: 871-878.

Kuzin, Nikita. 2024. Labile verbs in Ṭuroyo: Marginal phenomenon or structured innovation. In *Current Research in Semitic Studies. Proceedings of the Semitic Studies Section at the 34th DOT at Freie Universität Berlin*, Simona Olivieri and Shabo Talay (eds.), 115-132. Wiesbaden: Harrassowitz Verlag.

Letuchiy, Alexander. 2013. *Tipologija labil'nyx glagolov* [A typology of labile verbs]. Moscow: Jazyki slav. kul'tur.

Mengozzi, Alessandro. 1999. "The Mountains Will Be Covered with Snow": On the fringes of transitivity in the Neo-Aramaic of Alqosh (North Iraq). In *Afroasiatica Tergestina: Papers from the 9th Italian Meeting of Afro-Asiatic (Hamito-Semitic) Linguistics, Trieste, April 23–24, 1998*, M. Lamberti & L. Tonelli (eds), 83-99. Padova: Unipress.

Fassberg, Steven E. 2021. On the Afel Stem in Western Neo-Aramaic. In *Studies in the Grammar and Lexicon of Neo-Aramaic*, G. Khan & P. M. Noorlander (eds), 287-300. Cambridge: Open Book Publishers.

Abbreviations

LS   L-series of suffixes
M   Masculine
MWA   Modern Western Aramaic
PL   Plural
PST   Past tense

Valency alternations, lability types and semantic verb classes in Dargwa

Gasangusen Sulaibanov

(École Pratique des Hautes Études - Université PSL, Paris)

Keywords: valence change, lability, Dargwa, semantic verb classes, transitivity, East Caucasian

This study investigates how valency alternations are expressed in Dargwa languages, with a focus on Tsugni Dargwa, and their relation to semantic verb classes. It draws on an oral corpus of the Tsugni Dargwa language (Tsugni Dargwa < Dargwa languages < East Caucasian), compiled by the author through fieldwork in the Republic of Daghestan, Russia, as part of a PhD dissertation aimed at documenting and describing Tsugni Dargwa previously unstudied and undocumented. In addition to the language corpus data, the study also employs materials obtained through elicitation and introspection (since the author is a native speaker). The application of synchronic and diachronic descriptive analysis enables the identification of valency change strategies and their typological significance, offering new insights into lability types.

In Dargwa languages, only a few verbs like 'kill' and 'die' form suppletive transitivity pairs, and transitivity alternation can be expressed by different devices linked to actionality, imperfective aspect, mood, and morphological type. In Tsugni Dargwa for instance, semantic verb classes are strongly linked to valence change possibilities. Transitivity alternation can be expressed by means of

- P-lability for just a small class of verbs that have no necessary Agent in their semantic frame, like 'open', 'close' or 'fill' as in (1);

- A-lability (antipassive), in the imperfective tenses only, for a large class of transitive verbs which must keep a semantic Agent in their semantic frame, like 'eat' or 'drink' in (2);

- a small class of periphrastic causatives for lexically atelic intransitive verbs like 'cry' or 'boil' in (3);

- a large class of 'prosodic' causatives (Sulaibanov & Sumbatova 2022), in which valence increase is expressed by shifting the stress onto the ending instead of the stem, for instance 'break' in (4);

- a recent (and the only productive) class of equipollent verbal compounds using different verbalizers 'be' vs 'do' (5).

From a diachronic point of view

- periphrastic formations are more recent and transparent, interacting with complex aspectual formations;

- stress shift probably goes back to a lost suffix (from a univerbalized causative auxiliary) which underwent total segmental attrition;

- different conjugations in the imperative, prohibitive, 'eventual' and other irrealis moods rely in part, for the intransitive forms, on detransitivizing suffixes having cognates in other East Caucasian branches (cf. Authier 2012).

Unlike Andic languages (Rochant 2018), Tsugni Dargwa's limited P-labile verbs highlight its unique valency system.

Examples

1. a. k'uč'u (nig-li)   b-ic'-ib
    cup   milk-INS   N-fill.PFV-PRET(3)
    'The cup filled (with milk).'
    (noncausal context)

b. rursːi-li k'uč'u (nig-li)   b-ic'-ib
    girl-ERG cup   milk-INS N-fill.PFV-PRET(3)
    'The girl filled the cup (with milk).'
    (causal context)

2. a. rursːi-li nig   d-uč:-ib
    girl-ERG milk(NPL) NPL-drink.IPFV-PRET(3)
    'The girl used to drink milk.'
    (causal verb)

b. rursːi (nig-li)   r-uč:-ib
    girl   milk-INS   F-drink.IPFV-PRET(3)
    'The girl used to drink (milk).'
    (detransitivized via antipassive constr.)

3. a. hin   ruˤrq-ib
    water   boil.IPFV-PRET(3)
    'The water boiled.'
    (noncausal verb)

b. rursːi-li hin   ruˤrq-iˤ   iʁ-ib
    girl-ERG water boil.IPFV-INF chase.IPFV-PRET(3)
    'The girl boiled water.'
    (transitivized via periphrastic causative)

4. a. k'uču b-aˤč'-un
    cup   N-break.PFV-PRET(3)
    'The cup broke.'
    (noncausal verb)

b. rursːi-li k'uču b-aˤč'-un
    girl-ERG cup   N-break.PFV-PRET(3)(CAUS)
    'The girl broke the cup. '
    (transitivized via prosodic causative)

5. a. rursːi aq r-iχʷ-ub
    girl   high F-be.PFV-PRET(3)
    'The girl got up.'
    (noncausal compound verb with ITR light verb)

b. rursːi-li   naˤq   aq   b-arq'-ib
    girl-ERG   hand   high   N-do.PFV-PRET(3)
    'The girl raised her hand.'
    (causal compound verb with TR light verb)

References

Authier, Gilles (2012), The detransitive voice in Kryz, in G. Authier and K. Haude (eds), *Ergativity, Valency and Voice*, Berlin: De Gruyter Mouton, 133–164.

Rochant, Neige (2018), Lability in Andi, *Acta Linguistica Petropolitana* 15(2), 229–264.

Sulaibanov, Gasangusen & Sumbatova, Nina (2022), On one typological rarity: The causative in Tsugni Dargwa, *Voprosy Jazykoznanija* 3, 109–131.

External causation in anticausativization and lability in Latin: Towards a passive-anticausative continuum

Giovanbattista Galdi & Tim Ongenae

(Ghent University & Ghent University)

Keywords: anticausative, lability, Latin, passive, semantic roles

This paper investigates the influence of external causation or its absence on the development of anticausativization and lability in Latin. *Anticausativization* transforms a *causal* event (externally caused: "John opens the door") into a *noncausal* event (occurring spontaneously: "The door opens"). Verbs undergoing anticausativization lack an Agent-oriented meaning component (Haspelmath 1993). This allows semantic flexibility in the subject of the causal alternant, as it bears the generalized role of *Effector*, which can be subdivided into (Van Valin & Wilkins 1996, and Næss 2007: 107–110):

i.   *Agent* (+control, -affected): (1)
ii.   *Instrument* (-control, +affected): (2)
iii. *Force* (-control, -affected): (3).

(1) *Graeci (…)*   *hordeum*   *siccant.*
     Greek.NOM.M.PL   barley.ACC.N.SG   dry.IND.PRS.3PL.ACT
     "Greeks dry barley." (PLIN. nat. 18, 72)

(2) *Spongeae (…)*   *ulcera (…)*   *inpositae*   *siccant.*
     sponge.NOM.F.PL   ulcer.ACC.N.PL   applied.NOM.F.PL   dry.IND.PRS.3PL.ACT
     "Applied sponges dry ulcers." (PLIN. nat. 31, 126)

(3) *Haec (…)*   *sol*   *siccat.*
     DET.ACC.N.PL   sun.NOM.M.SG   dry.IND.PRS.3SG.ACT
     "The sun dries these things." (PLIN. nat. 21, 84)

This paper examines how the role of Force impacts the syntax of *detransitivization* in Latin. Detransitivization with Force, as in (4) and (5), is sometimes labelled as *passive* (causal) (Comrie 1985, and Siewierska 1986), and sometimes as *anticausative* (noncausal) (Kulikov 1998, and Zúñiga & Kittilä 2019: 43–48). However, the Latin data suggest that neither classification proves satisfactory. This paper argues that the development of anticausativization and the rise of lability in Latin (see Gianollo 2014, Cennamo, Eythórsson & Barðdal 2015, Cennamo 2022, and Ongenae 2024), is influenced by the presence of a Force. Data are extracted from the *Library of Latin Texts*, a corpus of Latin texts ranging from the third century BCE to the sixth century CE (approximately 10,000,000 words). Noncausal examples were annotated for the following variables:

- Construction (the three anticausative strategies):
    i.   mediopassive (in *-r*, syncretic with the passive marker): (4),
    ii.   labile (active intransitive): (5),
    iii. reflexive (with *se* REFL): (6),
- Deagentivization: implicit or explicit presence of an external Agent,
- Causalness degree: the proportion of causal uses of a verb, calculated as (causal) / (causal + noncausal) (following Haspelmath et al. 2014),
- Presence of a Force, as *vento* (4) or *frigore* (5),
- Century

(4) *Frumenta   vento   separantur.*
grain.NOM.N.PL wind.ABL.N.SG   separate.IND.PRS.3PL.MPASS
"Grain separates through the wind." (COLVM. 2, 20, 5)

(5) *Flumina (…)   frigore   durant.*
river.NOM.N.PL cold.ABL.N.SG   harden.IND.PRS.3PL.ACT
"Rivers harden through the cold." (AETNA 498)

(6) *vitium (…)   ruperit   se*
disease.NOM.N.SG   break.IND.FUT.3SG.ACT   REFL.ACC
"The disease will have broken." (CHIRON 384)

The data reveal that reflexivity, as in (6), precludes a Force, which confirms that it is mainly reserved for *autocausatives*. Furthermore, Figure 1 shows that, in Late Latin, the labile strategy is preferred for verbs with a low causalness degree and extends to verbs with a higher causalness degree when a Force is absent, while the mediopassive remains frequent when a Force is present. These findings support the reconsideration of the passive-anticausative distinction as a continuum based on form-frequency and the presence of external causation.
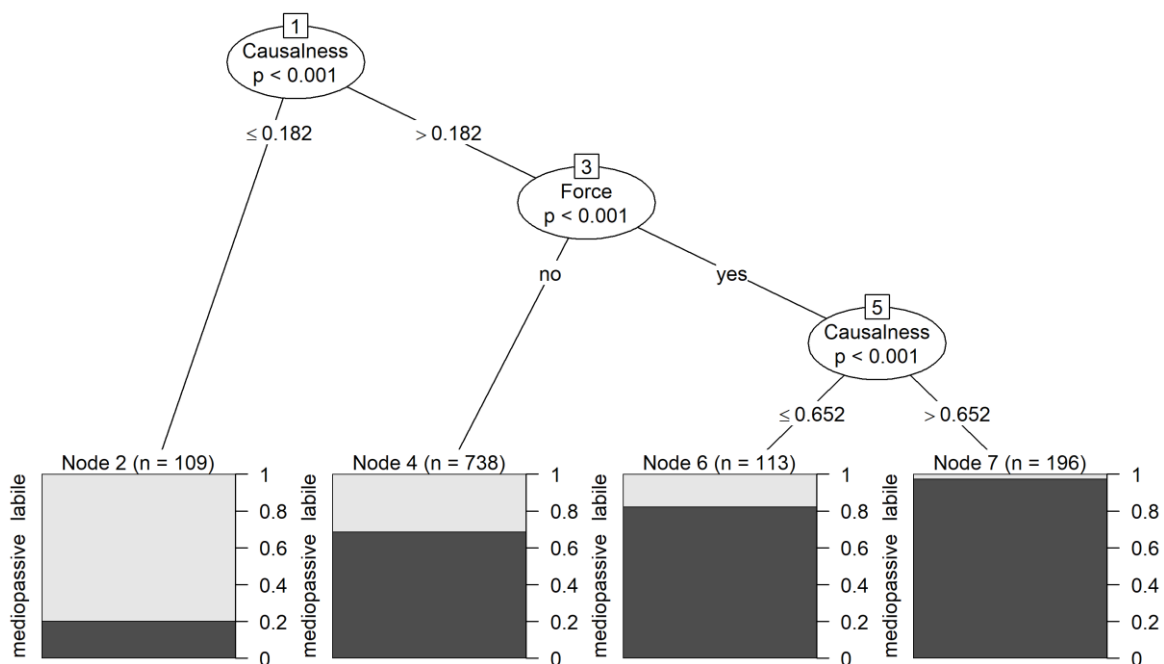

Figure 1: Conditional Inference Tree: Construction ~ Causalness + Force (Construction = mediopassive and labile with +deagentivization in Late Latin)

References

Cennamo, Michela (2022), Anticausativization in Latin and Early Italo-Romance: The semantics of predicates and the syntax of voice, in D. Romagno, F. Rovai, M. Bianconi & M. Capano (eds),

Variation, Contact, and Reconstruction in the Ancient Indo-European Languages, Leiden: Brill, 243–273.

Cennamo, Michela, Eythórsson, Thórhallur & Barðdal, Jóhanna (2015), Semantic and (morpho)syntactic constraints on anticausativization: Evidence from Latin and Old Norse-Icelandic, *Linguistics* 53(4), 677–729.

Comrie, Bernard (1985), Causative verb formation and other verb-deriving morphology, in T. Shopen (ed.), *Causative verb formation and other verb-deriving morphology* (*Language Typology and Syntactic Description* 3), Cambridge: Cambridge University Press, 309–348.

Gianollo, Chiara (2014), Labile verbs in Late Latin, *Linguistics* 52(4), 945–1002.

Haspelmath, Martin (1993), More on the typology of inchoative/causative verb alternations, in B. Comrie & M. Polinsky (eds.), *Studies in Language Companion Series*, vol. 23, Amsterdam: John Benjamins, 87–120.

Haspelmath, Martin, Calude, Andreea, Spagnol, Michael, Narrog, Heiko & Bamyaci, EliF (2014), Coding causal–noncausal verb alternations: A form–frequency correspondence explanation, *Journal of Linguistics* 50(3), 587–625.

Kulikov, Leonid (1998), Passive, anticausative and classification of verbs: The case of Vedic, in L. Kulikov & H. Vater (eds.), *Typology of Verbal Categories*, Tübingen: Niemeyer, 139–168.

Næss, Åshild (2007), *Prototypical transitivity* (*Typological Studies in Language* 72), Amsterdam: John Benjamins.

Ongenae, Tim A. F. (2024), *Permittito aperiat oculum*: Typological considerations on P-lability and its interaction with morphosyntactic alignment in Latin medical texts, *Folia Linguistica Historica* 45(1), 79–113.

Siewierska, Anna (1986), *The passive: A comparative linguistic analysis* (*Croom Helm Linguistics Series*), London / Sydney / Wolfeboro, N.H.: Croom Helm.

Van Valin, Robert D. & Wilkins, David P. (1996), The case for 'Effector': Case roles, agents, and agency revisited, in M. Shibatani (ed.), *Grammatical Constructions*, Oxford: Oxford University Press, 289–322.

Zúñiga, Fernando & Kittilä, Seppo (2019), *Grammatical voice* (*Cambridge Textbooks in Linguistics*), Cambridge: Cambridge University Press.

Typological Shift of Chinese: The Diachronic Evolution of Causatives and Anticausatives

Jing Du, Zifan Li & Yu Shen

(University of Chinese Academy of Sciences, Beihang University & Zhejiang University)

Keywords: lability, causative-anticausative alternation, Chinese, typological shift, diachrony

Lability refers to a verb's ability to alternate between different valency structures (Kulikov 2014, Heidinger 2015, and Zhang 2019). One common form of lability is causative-anticausative alternation, which Haspelmath (1993) categorized into five typological types: causative, anticausative, labile, equipollent, and suppletive. This study hypothesizes that Chinese has undergone a typological shift from a labile language, where causative and anticausative forms are identical, to a causative language, where causatives are formed through compounding, and anticausatives remain monomorphemic. Historically, ancient Chinese was classified as labile (e.g., Sun 1999, and Wei 2000), as seen in the following examples where causative and anticausative forms share the same structure (1a-b).

(1)a. 石工破山石。

    *shí   gōng   **pò**   shān   shí*

    stone worker **break** mountain rock

    'Stone workers broke the mountain rocks.'

  b. 石破山崩。

    *shí **pò**   shān   bēng*

    stone **break** mountain collapse

    'Stones broke and mountains collapsed.'

Modern Chinese, by contrast, exhibits distinct causative and anticausative forms. Causatives are typically compound verbs (action verb + result verb), while anticausatives are monomorphemic result verbs (2a-b).

(2)a. 他打破了杯子。

    *tā dǎ **pò** le bēi-zi*

    he hit **break** PFV cup

    'He broke the cup.'

  b. 杯子破了。

    *bēi-zi **pò**   le*

    cup **break** PFV

'The cup broke.'

Regarding the causative-anticausative alternation, previous synchronic studies have primarily focused on the derivational relationship between the two patterns (Levin and Rappaport Hovav 1995, Yang 2015, and Song 2023), and the syntactic or semantic factors constraining their alternation (Wright 2002, Rappaport Hovav 2014: 26, Alexiadou 2014, Levin 2015, and Heidinger and Huyghe 2024). In contrast, relatively few diachronic studies have systematically investigated the evolution of causatives and anticausatives, although some research has briefly addressed changes in verb transitivity (Wu 1999, Jiang 2000, Wei 2000, Xu 2005, and Hu 2005).

Situated against this research background, this study tests the hypothesis of typological shift using a corpus-based research methodology. To this end, five keywords, including *po* 'break', *kai* 'open', *bai* 'defeat', *chu* 'exit' and *jin* 'enter' were annotated in terms of their morphological structures (monomorphemic or compound) and their construction patterns (anticuasative, causative or passive).

Our major findings are: (i) Over time, compound forms of both causative and anticausative verbs have increased, while monomorphemic forms have declined. However, monomorphemic anticausative forms have remained relatively stable with only a slight decrease, suggesting that Chinese has shifted from a labile language to one dominated by causative constructions. (ii) Among the five verbs, *po* exhibits a rapid increase in causative compound forms, indicating a faster shift to a causative typology. *Jin*, while not as rapid, shows a steady increase in compound autonomous forms and a corresponding decrease in monomorphemic forms, suggesting a moderate shift. Conversely, *bai* and *chu* maintain a strong preference for monomorphemic forms, reflecting slower progression toward a causative language. *Kai* shows a more balanced development, with growth in both compound and monomorphemic forms across different uses.

In conclusion, the findings suggest that Chinese has undergone a shift from a labile to a causative-dominant language. However, the typological shift is uneven across different verbs, with some transitioning more rapidly than others.

References:

Alexiadou, Artemis (2014), The problem with internally caused change-of-state verbs, *Linguistics* 52(4), 879–909.

Haspelmath, Martin (1993), More on the typology of inchoative/causative verb alternations, in C. Bernard, and P. Maria (eds), (1993), *Causatives and Transitivity*, Berlin: John Benjamins, 87-120.

Heidinger, Steffen (2015), Causalness and the encoding of the causative–anticausative alternation in French and Spanish, *Journal of Linguistics* 51 (3), 562-594.

Heidinger, Steffen and Richard Huyghe (2024), Semantic roles and the causative-anticausative

alternation: Evidence from French change-of-state verbs, *Linguistics* 62(1), 159-202.

Hu, Chirui (2005), Dòngjiéshì de zǎoqī xíngshì jí pàndìng biāozhǔn [The early forms of resultative construction and the relevant criterion], *Studies of the Chinese Language* 3, 214-225.

Jiang, Shaoyu (2000), Hànyǔ dòngjiéshì chǎnshēng de shídài [The emerging period of Chinese resultative constructions], in S. Jiang(ed), *Collection of Papers on The History of Chinese Vocabulary and Grammar*, (2000), Beijing: The Commercial Press, 240-262.

Kulikov, Leonid (2014), The decline of labile syntax in Old Indo-Aryan: A diachronic typological perspective, *Linguistics* 52(4), 1139-1165.

Levin, Beth (2015), Semantics and pragmatics of argument alternations. *Annual Review of Linguistics* 1, 63-83.

Levin, Beth & Malka Rappaport Hovav (1995), *Unaccusativity: At the syntax-lexical semantics interface*, MA: MIT Press.

Rappaport Hovav, Malka (2014), Lexical content and context: The causative alternation in English revisited, *Lingua* 141, 8-29.

Song, Wenhui (2023), Xiàndài hànyǔ suǒwèi fēibīngé – fēi zuò gé dòngcí fēn lèi de xìngzhì [On the nature of so-called unaccusative-unergative verbs in Modern Chinese], *Contemporary Linguistics* 25(6), 810-829.

Sun, Chaofen (1999), The origin of the Chinese verbal suffixes. In A. Peyraube & C. Sun (eds), *Studies on Chinese Historical Syntax and Morphology: In Honor of Mei Tsu-Lin*, (1999), Paris: EHESS, CRLAO, 183-202.

Wei, Peiquan (2000), Shuō zhōnggǔ hànyǔ de shǐchéng jiégòu [A study on causative structures in Middle Chinese], *China Academy of Science Collected Papers on Diachronic Semantics* 71(4), 807-947.

Wright, Saundra K. (2002), Transitivity and change of state verbs, *BLS* 28, 339-350.

Wu, Fuxiang (1999), A tentative study on the origin of resultative constructions in Modern Chinese, in L. Jiang and J. Hou (eds), (1999), *Studies on Chinese Present and History— Collected Papers from First Chinese Linguistic International Seminar*, Beijing: Chinese Academy of Social Sciences Press, 317-345.

Xu, Dan (2005), Tán "pò": Hànyǔ mǒuxiē dòngcí de lèixíng zhuǎnbiàn [Typological changes of some verbs in Chinese: the case of *pò* (to break > broken)], *Studies of the Chinese Language* 4, 333-339.

Yang, Daran (2015), A study on the causative alternation from a non-derivational perspective, in X. Su & T. He (eds), (2015), *Chinese Lexical Semantics*, Dordrecht: Springer, 161-169.

Zhang, Liulin (2019), A study into the prototypicality of Chinese labile verbs, *Cognitive Semantics* 5 (1), 1-31.

Labile verbs in Irish. Descriptive, typological and historical perspectives

John Peterson
(Kiel University)

Keywords: *Irish, labile verbs, middle voice, transitivization, intransitivization*

In my talk I discuss labile verbs in Irish (Glottocode iris1253, ISO 639 gle). Irish is rich in labile verbs: E.g., of the nine "animate" categories mentioned in Nichols et al. (2004), five can express both the plain and induced meanings through labile verbs, as can seven of the nine "inanimate" categories. This is thus the primary valency orientation of Irish (cf. Table 1, labile verbs are shaded and ((1)-(2)).

As the vast majority of these verbs end in *-(a)igh*, stemming from a productive means of deriving verbs from nouns and adjectives in Old Irish, I trace the development of these verbs through Old Irish, following Griffith (2010; 2013), and putting this into the larger typological/historical-comparative perspective. E.g., whereas in Latin the transitive meaning of labile verbs is marked by the active and the intransitive by the middle voice (Xu et al., 2007: 137-139), in Irish both transitive and intransitive meanings are marked by the category which continued the Indo-European middle. I argue this was because this was no longer a middle voice in Old Irish, according to the definition given in Inglese (2022), but a lexically determined inflectional paradigm to which these derived verb stems belonged (cf. also Cowgill, 1983: 73; Stüber, 2017: 1209), with no oppositional functions such as "anticausative", "reflexive", etc. These and further related topics will be discussed.

Table 1: Plain and induced forms in Modern Irish (after Nichols et al., 2004)

Pair	Plain	Induced
1. laugh / make laugh, amuse, strike as funny	*gáir*	*cuir ag gáire*
2. die / kill	*éag* / *básaigh*	*maraigh* / *básaigh*
3. sit / seat, have sit, make sit	*suigh*	*suigh*
4. eat / feed, give food	*ith*	*biathaigh, beathaigh, cothaigh*
5. learn, know / teach	*foghlaim*	*foghlaim, múin, teagasc*
6. see / show	*feic*	*taispeáin*
7. be/become angry / anger, make angry	*feargaigh*	*feargaigh*
8. fear, be afraid / frighten, scare	*eaglaigh* / *uamhnaigh*	*eaglaigh* / *uamhnaigh*
9. hide, go into hiding / hide, conceal, put into hiding	*folaigh (é/í féin* 'him-/herself' as OBJ; SUBJ = OBJ)	*folaigh* (+ OBJ; SUBJ ≠ OBJ) / *cuir i bhfolach*
10. (come to) boil / (bring to) boil	*fiuch* / *bruith*	*fiuch* / *bruith*
11. burn, catch fire / burn, set fire	*dóigh*	*dóigh*
12. break / break	*bris*	*bris*
13. open / open	*oscail*	*oscail*
14. dry / make dry	*triomaigh*	*triomaigh*
15. be/become straight / straighten, make straight	*dírigh*	*dírigh*
16. hang / hang (up)	*bí crochta*	*croch*
17. turn over / turn over	*iompaigh* ('turn over') / *cas* ('turn')	*iompaigh* / *cas*
18. fall / drop, let fall	*tit*	*lig titim*

(1)   ***Bhailigh*** *an rang timpeall uirthi.*
        gather.PST   DEF   class   around   on.3SG.F
        'The class gathered around her.'

[Ní Mhuimhneacháin, 2020: 57]

(2)   *...* ***bhailigh***   *siad*   *a*   *gcuid*   *bia, ...*
        gather.PST   3PL   POSS.3PL   share.of   food
        '... they gathered their food, ...'

[Logan et al., 2000: 31]

References

Cowgill, Warren (1983), On the prehistory of Celtic passive and deponent inflection. *Ériu* 34: 73-111.

Griffith, Aaron (2010), Old Irish denominal and deadjectival deponents. Handout for the conference "The Indo-European Verb", UCLA, September 13-15, 2010.

Griffith, Aaron (2013), The decline of the Old Irish deponent, in Elisa Roma, and David Stifter (eds), *Proceedings of the XIV International Congress of Celtic Studies*. Lewiston, NY: Edwin Mellen Press. 53-78.

Inglese, Guglielmo (2022), Towards a typology of middle voice systems. *Linguistic Typology* 26(3), 489-531.

Logan, Austin, Sheila Tinsley, Rhonda Glasgow, Vivien Kelly & Maurice Todd. 2000. *An saol sa Chianaimsir.* Béal Feirste: An tÁisaonad Coláiste Naomh Muire.

Ní Mhuimhneacháin, Siobhán (2020), *An gadaí dubh. Kieran Black a mhaisigh.* Baile Átha Cliath: An Gúm.

Nichols, Johanna, David A. Peterson & Johnathan Barnes. 2004. Tranisitivizing and detransitivizing languages. *Linguistic Typology* 8. 149-211.

Stüber, Karin (2017), The morphology of Celtic. In: Jared Klein, Brian Joseph & Matthias Fritz (eds.), in cooperation with Mark Wenthe, *Handbook of comparative and historical Indo-European linguistics.* [Handbücher zur Sprach- und Komunikationswissenschaft, 41.2]. Berlin / Boston: De Gruyter Mouton. 1203-1218.

Intransitive Auxiliary as the encoding of non-canonical reflexive and reciprocal voice in Basque

Kristina Bilbao

(University of the Basque Country UPV/EHU)

Typological research on reflexive and reciprocal constructions has primarily focused on the morphological encoding of reflexivity/reciprocity, distinguishing between nominal and verbal reflexives/reciprocals (Faltz 1985, König and Kokutani 2006, Nedjalkov 2007, and Janic et al. 2023). Verbal reflexives/reciprocals are associated with reflexive/reciprocal voices, as they typically involve detransitivized constructions encoded by a verbal voice marker (Kulikov 2011). However, some inherently reflexive/reciprocal verbs express these meanings without specific morphological encoding (Zúñiga and Kittilä 2019). This pattern is cross-linguistically attested in grooming verbs for reflexivity (Haspelmath 2023) and social interaction verbs for reciprocity (Haspelmath 2007).

This study investigates non-canonical reflexive (1a) and reciprocal (1b) constructions in Basque. Although derived from transitive verbs like *ikusi* 'see', these constructions involve a single argument marked with the absolutive case, triggering absolutive agreement with the intransitive auxiliary *izan* 'be':

(1) a. Ane   ispiluan   ikusi   da.
     Ane(ABS)   mirror.in   see.PFV be.(3ABS)
     'Ane has seen herself in the mirror.'

    b. Ane   eta Miren   kalean   ikusi   dira.
      Ane(ABS)   and Miren(ABS)   street.in   see.PFV   be.3PL.ABS
      'Ane and Miren have seen each other in the street.'

Thus, the constructions in (1) are intransitive forms where no morphological voice marker is used. In this respect, they behave exactly like grooming verbs such as *dutxatu* 'shower' (2a) and verbs of social interaction like *batzartu* 'meet' (2b):

(2) a. Ane   dutxatu   da.
     Ane(ABS)   shower.PFV   be.(3ABS)
     'Ane has taken a shower.'

    b. Ane   eta Miren   batzartu   dira.
      Ane(ABS)   and Miren(ABS)   meet.PFV   be.3PL. ABS
      'Ane and Miren have met.'

Although both (1) and (2) lack morphological encoding of reflexivity/reciprocity, I argue they employ distinct strategies to express those meanings: the former relies on a morphosyntactic strategy, while the latter conveys them lexically. This study demonstrates that, despite the

absence of overt reflexive/reciprocal marking, verbs in constructions like (1) are not inherently reflexive/reciprocal, unlike verbs in (2). Furthermore, I propose that constructions like (1) are reflexive/reciprocal voices encoded through the intransitive auxiliary, rather than through a dedicated voice marker.

Three observations support these claims. First, the canonical counterparts of (1) are transitive constructions involving anaphors, while the verbs in (2) are canonical in their intransitive form and cannot occur with anaphors (Etxepare 2003). This indicates that the former, but not the latter, require explicit encoding for a reflexive/reciprocal interpretation. Second, verbs like *ikusi* (1) are not interpreted as reflexive/reciprocal in non-finite clauses, whereas grooming verbs and social interaction verbs are. Third, verbs behaving as *ikusi* (1) require the intransitive auxiliary to convey reflexivity/reciprocity, also in ditransitive constructions with an explicit object. These facts may suggest that the intransitive auxiliary is essential for achieving a reflexive/reciprocal interpretation in (1) and is not selected solely due to the intransitive nature of these constructions, unlike in (2).

In conclusion, I argue that Basque non-canonical reflexive/reciprocal constructions represent reflexive/reciprocal voices, challenging the conventional view that such voices necessarily involve a verbal morpheme as a voice marker. This work offers new insights into the typology of reflexive/reciprocal constructions, suggesting that constructions like (1) may arise from syntactic processes involving auxiliary selection rather than morphological voice markers.

References

Etxepare, Ricardo (2003), Reflexives and reciprocals, in J. I. Hualde, and J. Ortiz de Urbina (eds), (2003), *A Grammar of Basque,* Berlin: Mouton de Gruyter, 380–385. https://doi.org/10.1515/9783110895285.363

Faltz, Leonard M. (1985), *Reflexivization: A study in universal syntax*, New York: Garland.

Haspelmath, Martin (2007), Further remarks on reciprocal constructions, in V. Nedjalkov (ed), (2007), *Reciprocal Constructions*, Amsterdam: John Benjamins, 2087–2115. https://doi.org/10.1075/tsl.71.74has

Haspelmath, Martin (2023), Comparing reflexive constructions in the world's languages, in K. Janic, N. Puddu, and M. Haspelmath (eds), (2023), *Reflexive constructions in the world's languages*, [*Research on Comparative Grammar* 3], Berlin: Language Science Press, 19–62. 10.5281/zenodo.7861660

Janic, Katarzyna, Puddu, Nicoletta, and Haspelmath, Martin (eds), (2023), *Reflexive constructions in the world's languages*, [*Research on Comparative Grammar* 3], Berlin: Language Science Press. 10.5281/zenodo.7861660

König, Ekkehard, and Kokutani, Shigehiro (2006), Towards a typology of reciprocal constructions: focus on German and Japanese, *Linguistics* 44(2), 271–302. https://doi.org/10.1515/LING.2006.010

Kulikov, Leonid (2011), Voice typology, in J. J. Song (ed), *The Oxford Handbook of Linguistic Typology,* Oxford: Oxford University Press, 368–398.

Nedjalkov, Vladimir P. (ed), (2007), *Reciprocal Constructions*, Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.71

Zúñiga, Fernando, and Kittilä, Seppo (2019), *Grammatical voice*, Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316671399

Decline of lability in Old Indo-Aryan and new labile verbs in Middle Indo-Aryan: two competing tendencies

Leonid Kulikov & Mayya Shlyakhter

(Ghent University and Université catholique de Louvain (UCL) & Hebrew University of Jerusalem)

This paper focuses on the evolution of the system of labile verbs in Indo-Aryan, i.e. verbs that can appear both in intransitive and transitive-causative usages) with no change in the form of the verb, as in the case of the early Vedic (Rgveda) 3rd person plural of the active perfect *vāvṛdhuḥ*, cf. RV 2.34.13 *rudrā́ ṛtásya sádaneṣu **vāvṛdhuḥ*** 'Rudras **have grown** [intransitive] in the residences of the truth' ~ RV 8.6.35 *índram ukthā́ni **vāvṛdhuḥ*** 'the hymns **have increased** [transitive-causative] Indra') in Old Indo-Aryan (Sanskrit) (cf. Renou 1924). The paper offers a general survey of labile verbs and their status within the system of causative oppositions in Old Indo-Aryan in a diachronic perspective, focusing on the history of labile verbs and elucidating the position of the Indo-Aryan branch within Indo-European.

We arguably observe the decline of the labile type already in the second most ancient Vedic text, the Atharvaveda, where the number of labile forms considerably decreases (for details, see Kulikov 2014: 1158f.). Thus, most of the active perfects that are labile in the Ṛgveda either occur in intransitive usages only, or in transitive-causative usages only, or are not attested at all.

This strong tendency towards the decrease of the number of labile verbs must be due to growing productivity of the morphological causatives (with the suffix -*aya*-), which eventually oust the unmarked (or weakly marked) causative members in anticausative/causative oppositions.

Yet, in spite of this obvious tendency, in late Old Indo-Aryan (late Vedic and early post-Vedic Sanskrit) several new labile verbs arise, such as *sūyate* 'produces; is produced' or *sṛjyate* 'makes; is made'. I argue that this phenomenon is due to a number of dramatic changes in the verbal system observed between the (late) Old Indo-Aryan and Middle Indo-Aryan periods, such as, first of all, the collapse of the Sanskrit morphological system. This 'new lability' could further be supported by the influence of the early Middle Indo-Aryan dialects (used as spoken languages in this period), where many of the Old Indo-Aryan morphological opposition were lost entirely. Thus, the existence of the Middle Indic (e.g. Pāli) labile verbs of the type *abhibhuyyati* 'overcomes, overpowers; is overpowered' (which arise due to the loss of the middle voice; see e.g. Oberlies 2003: 199) could trigger both (i) the rise of late Vedic labile verbs of the type *sūyate* 'produces, generates' / 'is produced, is generated' (in late Sūtras) and (ii) the emergence of several new morphological types of transitivity oppositions, based on analogical present formations (such as the -*ya*-present -*bhūya*-, impossible in early Vedic) and resulting in pairs of the type *abhibhūya**te*** [passive] 'is (being) overcome' ~ *abhibhūya**ti*** [transitive] 'overcomes' (for instance, in Maitrāyaṇī-Up. and other late Upaniṣads; see van Buitenen 1962: 129f.).

A detailed analysis of the evolution of the system of labile verbs in Old and Middle Indo-Aryan furnishes important evidence for a diachronic typology of syntactically unstable verbs and, particularly, for elucidation of classes of verbs where the emergence of lability was particularly common.

References

Buitenen, J.A.B. van. 1962. *The Maitrāyaṇīya Upaniṣad: a critical essay, with text, translation and commentary*. 's-Gravenhage: Mouton.

Kulikov, L. The decline of labile syntax in Old Indo-Aryan: a diachronic typological perspective. *Linguistics* 52(4): 1139–1165.

Oberlies, T. 2003. *A grammar of Epic Sanskrit* (Indian Philology and South Asian Studies 5). Berlin – New York: de Gruyter.

Renou, L. 1924. Védique *vāvṛdhé, vāvṛdhúḥ*. *Bulletin de la Société de Linguistique de Paris* 24: 185–188.

Anticausatives and lability in Latin: a quantitative study

This paper investigates the role played by the interplay of the *aspectual template of verbs*, the *verb's inherent meaning* (the 'root'), and the *nature of the –P subject* (e.g., animacy and control) in determining the distribution of the different strategies available in Latin to mark anticausativization (the (medio-passive) *-r* form, the reflexive pattern and the active intransitive. The different forms are usually described in the literature as interchangeable (Feltenius 1977). We will argue, instead, that the structural and lexical aspects of the verb meaning (following Levin & Rappaport Hovav 2005, int. al) as well as the inherent and relational properties of verbal arguments affect the use of the different patterns, interacting with the encoding of voice, both synchronically and diachronically.

 Building on Ongenae (2024), we will carry out a quantitative investigation, testing the hypothesis put forward in Cennamo, Eythórsson & Barðdal 2015, Cennamo 2022, whereby the selection of the reflexive strategy in Latin was initially confined to inherently telic predicates (achievements and accomplishments) (e.g., *frangere* 'break', *dum calor se frangat* - till   heat   RFL breaks - 'Till the heat goes down', *fervefacere* 'heat up', **se** *patinae* **fervefaciunt** RFL - pans heat-up -'The pans heat up'), whilst the active intransitive, labile strategy mainly occurred with verbs of variable/reduced telicity (e.g., *fuscare* 'darken', *gratia fuscavit* – grace blurred - 'Grace blurred'), with activities (e.g., *volutare* 'roll', *saxa volutant* - stones roll - 'Stones roll') and, marginally, with accomplishments lexicalizing a reversible state (e.g., *aperire* 'open', *foris aperit* - door opens - 'The door opens'). Gradually, in the course of time, the reflexive spreads to non-inherently telic (e.g., *minuere* 'decrease', **minuente se** *morbo* - decreasing RFL illness) and atelic predicates (e.g., *servare* 'keep'- *mala* **se servant** - apples RFL keep - 'Apples keep'), and the active intransitive expands to inherently telic predicates (e.g., *rumpere* 'break'), until in late texts the three anticausative forms become truly interchangeable (**rumpunt** *dentes*/**rumpuntur** *dentes*/*dentes* **se rumpunt** - break teeth/break-R teeth/teeth RFL break - 'Its teeth break' (sc. *equus* 'horse'). The analysis will also consider the interplay of the available strategies for anti-causativization with different types of verbal root (e.g., denominal and deadjectival).

The Latin data, therefore, appear to offer an interesting contribution to the current debate on the role played by the verb's inherent meaning and its interaction and integration with the event structure template of predicates in determining argument realization, showing the relevance of these notions for the synchrony and diachrony of anticausativization in Latin.

References

Cennamo, M, Eythórsson, T. & Barðdal, J. 2015. *Semantic and (morpho)syntactic constraints on anticausativization: Evidence from Latin and Old Norse Icelandic*. Linguistics 53.4: 677-729.

Cennamo, M. 2022. *Anticausativization in Latin and Early Italo-Romance: The Semantics of Predicates and the Syntax of Voice.* in F. Rovai, D. Romagno et alii (eds.). *Variation, Contact, and Reconstruction in the Ancient Indo-European Languages.* Brill, Mouton-De Gruyter.

Feltenius, L. 1977. *Intransitivizations in Latin.* Upsala: Almquist & Wiksell.

Kulikov, L. 2003. *The labile syntactic type in a diachronic perspective: the case of Vedic.* SKY Journal of Linguistics 16: 93-112.

Kulikov, L. & N. Lavidas. (guest-eds) 2014. *Typology of labile verbs: Focus on diachrony.* Linguistics 52.4: Special issue.

Levin B. & Rappaport Hovav, M. 2005. *Argument Realization.* Cambridge: Cambridge University Press.

Ongenae, T.A.F. 2024. *Permittito aperiat oculum: typological considerations on P-lability and its interaction with morphosyntactic alignment in Latin medical texts.* Folia Linguistica Historica 58 (s45-s1): 79-113.

Marescotti, C. 2024. *Denominal verbs in *-ye/o- at the syntax-semantics interface. Insights from ancient Greek and Latin.* PhD Thesis, Università di Pisa.

Towards a typology of passive lability

Peter Arkadiev
(University of Potsdam)

Keywords: lability, morphosyntax, passive, typology, uncoded voice alternations

Uncoded passive alternations, also known as passive lability (1), are only rarely mentioned and discussed in the typological and theoretical literature on either voice (Zúñiga & Kittilä 2019: 188–189) or lability (Letuchiy 2013: 136–145; Creissels 2014).

Kakabe (Vydrina 2011: 190)

(1) a. *Wùléè   bati   Sɛ́ɛku   kín*
         dog.ART   PRF   Seeku   bite
         'The dog bit Seeku.'

     b. *Sɛ́ɛku   bati   kín*
         Seku   PRF   bite
         'Seeku has been bitten.'

At the same time, they are known to be prominent in some language families (e.g. Mande, Cobbinah 2008) and linguistic areas (e.g. Western Africa, Cobbinah & Lüpke 2012), as well as in Creoles (Kouwenberg 2023) (2).

Jamaican Creole (LaCharité & Wellington 1999: 260)

(2) a. *Kieti   rait   Jaiz   di   leta*
         Katie   write   Joice   DEF   letter
         'Katie wrote Joyce the letter.'

     b. *Di   leta   rait.*
         DEF   letter   write
         'The letter was written.'

Based on this, it has been proposed that passive lability is associated with a "general lack of verbal morphology" (Cobbinah & Lüpke 2012: 154). In this study based on a world-wide convenience sample of ca. 30 languages, I look at the cross-linguistic variation of passive lability and argue that it is compatible with rich verbal morphology as well (3). The data mainly come from reference grammars and special publications.

Central Alaskan Yupik (Miyaoka 2015: 1177, 1184)

(3) a. *angute-m   neqa   ner-a-a*
         man-ERG.SG   fish.ABS.SG   eat-TR-IND.3SG>3SG
         'The man is eating the fish.'

     b. *neqa   ner'-u-q   ak'a*
         fish.ABS.SG   eat-INTR-IND.3SG   IAM
         'The fish is/has been eaten.'

The parameters of passive lability that I consider are as follows:
(i) Whether the passive variant allows an oblique expression of the agent (most languages disallow it, but still it is attested e.g. in Bambara (4), and some languages, like Manggarai, even require it).

Bambara (Creissels 2014: 920)

(4) a. *wùlu má sògo dún*
dog.DEF PFV.NEG meat.DEF eat
'The dog has not eaten the meat.'

b. *sògo má dún **wùlu fὲ***
meat.DEF PFV.NEG eat dog.DEF by
'The meat has not been eaten by the dog.'

(ii) Whether the passive variant expresses a dynamic situation (e.g. Basque and Mande languages) (1), (4), a resultant state (Abaza, Nama and Berber languages) (5), or some kind of modality (English *The book sells well*).

Tarifiyt (Gutova 2013: 76)

(5) a. *ḥasan y-bna t-addart*
Hasan 3SG.M-build F-house
'Hasan built a house.'

b. *t-addart nn-es t-bna*
F-house POSS-3SG 3SG.F-build
'His house is built.'

(iii) Which transitive verbs show passive lability, with variation ranging from apparently full productivity in some Mande and Creole languages via semantic restrictions related to telicity or agentivity (Abaza, Nama) to idiosyncratic closed classes (Koyraboro Senni).

(iv) Which morphological changes go together with passive lability; while in many languages there are indeed none, one also finds marking of person and/or transitivity (e.g. Basque, Central Alaskan Yupik), tense and aspect (Mandinka), and combinations thereof (Abaza, Nama).

In summary, I show that passive lability is somewhat more widespread than is usually believed and that its cross-linguistic variation largely fits within the typology of "canonical" morphologically marked passives and complements it.

References

Cobbinah, Alexander (2008), Das morphologisch unmarkierte Passiv. Typologische Untersuchungen am Beispiel des Bambara, Magisterarbeit, Universität zu Köln.

Cobbinah, Alexander, and Friederike Lüpke (2012), Not cut to fit — zero coded passives in African languages, in M. Brenzinger, and A.-M. Fehn (eds.) (2012), *Proceedings of the 6th World Congress of African Linguistics*, Köln: Rüdiger Köppe, 153–165.

Creissels, Denis (2014), P-lability and radical P-alignment, *Linguistics* 52(4), 911–944.

Gutova, Evgeniya (2013), Voice, lability and causatives in Berber, Part 1, Voice and lability in Berber, *Etudes et documents berbères* 32, 63–79.

Kouwenberg, Silvia (2023), The Creole passive, in: P. Cabredo Hofherr, H. Glaude, and E. Soare (eds.) (2023), *Grammar is a Moveable Feast: A Webschrift for Anne Zribi-Hertz*, 239–249. https://zenodo.org/records/8054044

LaCharité, Darlene, and Jean Wellington (1999), Passive in Jamaican Creole: Phonologically empty but syntactically active, *Journal of Pidgin and Creole Languages* 14(2), 259–283.

Letuchiy, Alexander (2013), *Tipologija labil'nyx glagolov* [Typology of labile verbs], Moscow: LRC Publishing.

Miyaoka, Osahito (2015), Valency classes in Central Alaskan Yupik, and Eskimoan Language, in: A. Malchukov, and B. Comrie (eds.), *Valency Classes in the World's Languages, Vol. 2, Case Studies from Austronesia and the Pacific, the Americas, and Theoretical Outlook*, Berlin / Boston: De Gruyter Mouton, 1165–1204.

Vydrina, Alexandra (2011), Labi'lnost' v jazyke kakabe [Lability in Kakabe], *Acta Linguistica Petropolitana* 7(2), 174–217.

Zúñiga, Fernando, and Seppo Kittilä (2019), *Grammatical Voice*, Cambridge: Cambridge University Press.

Lability and Classical and Cilician Armenian

Petr Kocharov

(University of Würzburg)

Keywords: lability, voice, Classical Armenian, Cilician Armenian

Classical Armenian verbal system combines forms marked by oppositional active/mediopassive endings with labile forms. The paper offers an overview of the evolution of Classical Armenian (CArm) labile forms in the post-classical period in comparison to the Cilician Middle Armenian (MidArm). The present study is based on the descriptive grammars of Classical and Cilician Armenian (cf. Aṙakʿelyan 2010; Karst 1901) as well as on the corpus data (cf. *Leiden Armenian Lexical Textbase*: https://www.sd-editions.com; *TITUS*: https://titus.uni-frankfurt.de). The oppositional vs. labile marking of valency and valency alternations is considered within the theoretical framework outlined in Malchukov, Comrie 2015 and Haspelmath 2016 with references.

In the CArm present tense, the oppositional voice is marked by the endings in two out of four conjugations, cf. act. *ayr-ē* '(s)he burns', mp. *ayr-i* 'it is burned; it burns (by itself)', as opposed to lab. *ban-ay* '(s)he opens; it opens (by itself); it is opened', lab. *ln-ow* '(s)he fills; it fills (by itself); it is filled'. By contrast, in Cilician Armenian, the passive voice is coded by a derivational suffix *-ow-*, while the endings no longer marked the oppositional voice in either of the conjugations: MidArm. *as-ē* '(s)he says' : *as-ow-i* 'it is said'; *awc-n-ē* '(s)he anoints' : *awc-n-ow-i* 'it is anointed'; MidArm. *əndown-i* '(s)he receives' : *əndown-ow-i* 'it is received'; MidArm. *tʿoł-ow* '(s)he leaves' : *tʿoł-ow-i* 'it is left', *gt-n-ow* '(s)he finds' : *gt-ow-i* 'it is found'. As a result of this change, the former mediopassive conjugation became confined to non-passive intransitives, cf. act. *ayr-ē* '(s)he burns', mid. *ayr-i* 'it burns (by itself)', pass. *ayr-ow-i* 'it is burned'. The rise of the new derivational passive marker increased the lability of non-passive forms, which tended to generalize the former active voice forms, cf. CArm. *karmr-i* > MidArm. *karmr-ē* 'it becomes red', CArm. *tʿoł-ow* > MidArm. *tʿoł-ē* '(s)he leaves', with some systematic exceptions, cf. CArm. *tes-an-ē* > MidArm. *tes-n-ow* '(s)he sees'. In the CArm. aorist, the oppositional voice is expressed in all verbs, including verbs with the labile present forms, by two sets of endings added to the aorist stem. By contrast, in Cilician Armenian, the passive aorist is typically built from the derived passive stem, to which the aorist suffix with the former mediopassive endings are added, cf. MidArm. pres. act. *kap-ē* '(s)he ties' : pres. pass. *kap-ow-i* 'it is tied' : aor. pass. *kap-ow-ecʿ-aw* 'it was tied'.

The changes outlined above resulted in a split of intransitives into non-passive ones, expressed by labile forms on a par with transitives, and derivationally marked passives. This data offers new evidence on the role of lability in a long-term drift from the predominantly equipollent marking of transitivity alternations in CArm. to the predominantly intransitivizing strategy of the Middle and Modern Armenian (cf. Kocharov 2023a, 2023b).

References

Aṙakʿelyan, Varag (2010), *Grabari kʿerakanutʿiwn* [A grammar of Old Armenian], Yerevan: Viamir.
Haspelmath, Martin (2016), Universals of causative and anticausative verb formation and the spontaneity scale, *Lingua Posnaniensis* 58(2), 33–63.
Karst, Josef (1901), *Historische Grammatik des Kilikisch-Armenischen*, Strassburg: Trübner.

Kocharov, Petr (2023a), Anticausatives in Classical Armenian, *Journal of Historical Linguistics* 13(2), 255–294.

Kocharov, Petr (2023b), Causatives in Classical Armenian, *Folia Linguistica* 57(3), 593–627.

Malchukov, Andrej, Comrie, Bernard, eds. (2015). *Volume 1. Introducing the framework, and case studies from Africa and Eurasia*. Berlin, Boston: De Gruyter.

Decline of lability in Old Indo-Aryan and new labile verbs in Middle Indo-Aryan: two competing tendencies

Leonid Kulikov & Mayya Shlyakhter

(Ghent University and Université catholique de Louvain (UCL) & Hebrew University of Jerusalem)

This paper focuses on the evolution of the system of labile verbs in Indo-Aryan, i.e. verbs that can appear both in intransitive and transitive-causative usages) with no change in the form of the verb, as in the case of the early Vedic (Rgveda) 3rd person plural of the active perfect *vāvr̥dhuḥ*, cf. RV 2.34.13 *rudrā́ r̥tásya sádaneṣu vāvr̥dhuḥ* 'Rudras **have grown** [intransitive] in the residences of the truth' ~ RV 8.6.35 *índram ukthā́ni vāvr̥dhuḥ* 'the hymns **have increased** [transitive-causative] Indra') in Old Indo-Aryan (Sanskrit) (cf. Renou 1924). The paper offers a general survey of labile verbs and their status within the system of causative oppositions in Old Indo-Aryan in a diachronic perspective, focusing on the history of labile verbs and elucidating the position of the Indo-Aryan branch within Indo-European.

We arguably observe the decline of the labile type already in the second most ancient Vedic text, the Atharvaveda, where the number of labile forms considerably decreases (for details, see Kulikov 2014: 1158f.). Thus, most of the active perfects that are labile in the Ṛgveda either occur in intransitive usages only, or in transitive-causative usages only, or are not attested at all.

This strong tendency towards the decrease of the number of labile verbs must be due to growing productivity of the morphological causatives (with the suffix *-aya-*), which eventually oust the unmarked (or weakly marked) causative members in anticausative/causative oppositions.

Yet, in spite of this obvious tendency, in late Old Indo-Aryan (late Vedic and early post-Vedic Sanskrit) several new labile verbs arise, such as *sūyate* 'produces; is produced' or *sr̥jyate* 'makes; is made'. I argue that this phenomenon is due to a number of dramatic changes in the verbal system observed between the (late) Old Indo-Aryan and Middle Indo-Aryan periods, such as, first of all, the collapse of the Sanskrit morphological system. This 'new lability' could further be supported by the influence of the early Middle Indo-Aryan dialects (used as spoken languages in this period), where many of the Old Indo-Aryan morphological opposition were lost entirely. Thus, the existence of the Middle Indic (e.g. Pāli) labile verbs of the type *abhibhuyyati* 'overcomes, overpowers; is overpowered' (which arise due to the loss of the middle voice; see e.g. Oberlies 2003: 199) could trigger both (i) the rise of late Vedic labile verbs of the type *sūyate* 'produces, generates' / 'is produced, is generated' (in late Sūtras) and (ii) the emergence of several new morphological types of transitivity oppositions, based on analogical present formations (such as the *-ya*-present *-bhūya-*, impossible in early Vedic) and resulting in pairs of the type *abhibhūyate* [passive] 'is (being) overcome' ~ *abhibhūyati* [transitive] 'overcomes' (for instance, in Maitrāyaṇī-Up. and other late Upaniṣads; see van Buitenen 1962: 129f.).

A detailed analysis of the evolution of the system of labile verbs in Old and Middle Indo-Aryan furnishes important evidence for a diachronic typology of syntactically unstable verbs and, particularly, for elucidation of classes of verbs where the emergence of lability was particularly common.

References

Buitenen, J.A.B. van. 1962. *The Maitrāyaṇīya Upaniṣad: a critical essay, with text, translation and commentary*. 's-Gravenhage: Mouton.

Kulikov, L. The decline of labile syntax in Old Indo-Aryan: a diachronic typological perspective. *Linguistics* 52(4): 1139–1165.

Oberlies, T. 2003. *A grammar of Epic Sanskrit* (Indian Philology and South Asian Studies 5). Berlin – New York: de Gruyter.

Renou, L. 1924. Védique *vāvṛdhé, vāvṛdhúḥ*. *Bulletin de la Société de Linguistique de Paris* 24: 185–188.

Two types of P-lability in Icelandic

In this paper we report on two types of patient-preserving (P-) lability in Icelandic. We identify the characteristics of these labile verbs, linking the discussion to morphological factors and syntactic alignment. In P-lability the patient (object) argument of the transitive variant is preserved as the subject of the intransitive (anticausative) variant (Kulikov & Lavidas 2014).

The two types of P-lability in Icelandic are exemplified below; the transitive and intransitive verb forms are identical in both types. Type 1 (Morphologically Unmarked) has a nominative subject of the anticausative (1b). In Type 2 (Case-Preserving Anticausativization, CPA) the oblique case (accusative, dative or genitive) of the transitive variant (2a) is preserved with the subject of the anticausative variant (2b). On oblique subjects in Icelandic, cf. the summary in Thráinsson (2007).

(1)   Type 1: Morphologically Unmarked
        a. *Stelpan*   *hleður*   ***símann***.
           girl.the.NOM   charges   phone.the.ACC 'The girl is charging the phone.'
        b. ***Síminn***   *hleður*   *ekki*.
           phone.the.NOM   charges   not 'The phone isn't charging.'

(2)   Type 2: Case-Preserving Anticausativization (CPA)
        a. *Vindurinn*   *rak*   ***bátinn***   *á land*.
           wind.the.NOM   drove   boat.the.ACC to land 'The wind drove the boat ashore.'
        b. ***Bátinn***   *rak*   *á land*.
           boat.the.ACC   drove   to land 'The boat drifted (lit., drove) ashore.'

The morphologically unmarked type is the norm in Mainland Scandinavian and English, where inflectional morphology has been lost to a significant degree. However, in Old Icelandic only a handful of verbs of this type are found, e.g., *breikka* 'broaden' and *minnka* 'lessen'. In Present-Day Icelandic, Type 1 has gained some additional members, through language contact and by a change termed Nominative Substitution, whereby oblique case with subjects is replaced by nominative (Svavarsdóttir 1982).

Type 2 (CPA) was productive in Icelandic until the 19th century and a number of verbs follow this pattern in Modern Icelandic. Pooth et al. (2019) and Barðdal et al. (2020) maintain that CPA represents an old layer in the language that can be traced to Proto-Indo-European (PIE). Barðdal & Eythórsson (2009) argue that oblique-subject constructions in Icelandic can be explained by referring to a putative Fluid-S alignment system in PIE. Although this assumption may be valid for PIE, it is problematic for the historical stages of Germanic; it neither accounts for the increase in oblique subjects throughout the history of Icelandic, nor does it agree with the different ways in which such constructions can emerge. Today, the productivity of CPA is very limited, with only a few new sporadic examples known to us. Among the factors detrimental to CPA is Nominative Substitution, chipping away the number of anticausatives formed according to Type 2 in favor of Type 1.

To conclude, by investigating selected P-labile verbs in the history of Icelandic we show that the productivity of the strategies in (1) and (2) depends on morphological changes in

the grammar. We claim that morphological richness and changes in the case system play an important role in determining which type of lability, Type 1 or Type 2, is productive in each period.

References

Barðdal, Jóhanna & Thórhallur Eythórsson. 2009. The origin of the oblique-subject construction: an Indo-European comparison. In Vit Bubenik et al. (eds.), *Grammatical Changes in Indo-European languages*, ráðstefnurit frá 18th ICHL, Montréal 2007, 1.-6. August (pp. 179-193). John Benjamins Publishing Company.

Barðdal, Jóhanna, Leonid Kulikov, Roland Pooth & Peter Kerkhof. 2020. Oblique anticausatives: a morphosyntactic isogloss in Indo-European. *Poznan Studies in Contemporary Linguistics* 56(3):413–449.

Kulikov, Leonid & Nikolaos Lavidas. 2014. Introduction: Typology of labile verbs. *Linguistics* 52(4):871–877.

Pooth, Roland, Peter Alexander Kerkhof, Lenoid Kulikov & Johanna Barðdal. 2019. The origin of non-canonical case marking of subjects in Proto-Indo-European: accusative, ergative, or semantic alignment. *Indogermanische Forschungen* 124(1):245–263.

Svavarsdóttir, Ásta. 1982. "Þágufallssýki" [Dative Sickness]. *Íslenskt mál og almenn málfræði* 4:19–62.

Thráinsson, Höskuldur 2007. *The syntax of Icelandic*. Cambridge: Cambridge University Press.

WS23 Turn out verbs and constructions crosslinguistically: Properties and boundaries, synchrony and diachrony

Patrick Dendale & Ana Stulic

'Turn out' constructions in Afrikaans

Adri Breed & Lande Botha

North-West University, South Africa

Afrikaans and Dutch are closely related languages with estimated lexical resemblance of 90 to 95 percent (Bradfort & Claughton 1995; Carstens 1989, Raidt 1972). Afrikaans has three epistemic verbs equivalent to those found in Dutch, namely *skyn* 'schijnen (Du)/appear (En)', *lyk* 'lijken (Du)/seem (En) and *blyk* 'blijken' (as in example 1) (see Mortelmans 2022 for a discussion of the Dutch verbs).

1   *Toe **blyk dat** die man net so hulpeloos is as die vrou.*   *(afTenTen24)*
    'It turned out that the man is just as helpless as the woman.'

Similar to Dutch *blijken*, Afrikaans uses the verb *blyk* as the most frequent rendition of TURN OUT constructions. While *blyk* conveys a high degree of certainty (see Mortelmans 2022:296), *skyn* and *lyk* are associated with lower degrees of certainty regarding the proposition in the clause. Afrikaans also has other constructions, similar to Dutch, that convey the sense of new information 'coming to light' (Dendale et al., 2024), such as *aan die lig kom* ('come to light') and *word duidelik* ('becomes clear'). The latter expressions seem to indicate a more gradual emergence of new information, while *blyk* constructions suggest a quicker or more sudden emergence of the information.

2   *Later het dit **aan die lig gekom** dat Duitsland en Japan 'n ooreenkoms gesluit het*   *(afTenTen24)*
    'Later, it came to light that Germany and Japan had reached an agreement.'

3   *Dit **word duidelik** dat SA ook reeds voor Covid-19 in 'n humanitêre krisis verkeer het*
    'It becomes clear that South Africa was already in a humanitarian crisis before Covid-19.'

Further, a calque of the English 'turn out' construction is also sometimes used in informal and spoken Afrikaans.

4   *Terwyl ek gaan draf het vandag het ek gehoor iemand klap vir my hande dit **het uitgedraai dat** dit my bene is*   *(Google)*
    'While I was jogging today, I heard someone clap for me, and it turned out that it was my legs

In Afrikaans, similar to Dutch, the construction with *blyk* typically takes the complementiser *dat* 'that' to introduce the subordinate clause containing the proposition's information, while *lyk* takes the complementiser *of* 'if'. However, a cursory perusal of online Afrikaans shows uses of the verb *blyk* in constructions normally associated with *lyk* (sometimes also with *of*, and/or the modal *wil* 'would'). This internal contamination can cause ambiguity in the interpretation of the level of certainty of the proposition.

5   *Dit **wil blyk of** hierdie stap egter net die teenstrydighede vererger het!*   *(afTenTen24)*
    'It seems that/turns out that this step, however, has only exacerbated the contradictions!'

6   *Toe dit vir hom **gelyk** het **dat** hy sou sterf, het hy te besorg daaroor geword om homself te regverdig*
                                                                            (afTenTen24)
    'When it seemed/became clear to him that he was going to die, he became too worried about it to justify himself.'

To supplement the work done on TURN OUT constructions in other languages, this paper will provide a corpus-based inventory of Afrikaans TURN OUT constructions, including its potential co-occurrence with an impersonal subject, complementiser, modal auxiliaries and other epistemic markers. We will also classify the emergent information in the proposition in terms of its duration (gradual vs sudden) and levels of certainty.

References

Brandford, W. & Claughton, J.W. 1995. In Mesthrie (ed). *Language and social history: studies in South African sociolinguistics.* Cape Town: David Philip, pp. 209-221.

Carstens, W.A.M. 1989. *Norme vir Afrikaans.* Pretoria: Academica

Dendale, P. Izquierdo Alegria, D. & Stulic, A. 2024. '*Turn out'* verbs in European languages: Are they evidentials or something else? Talk at the 2024 SLE conference, Helsinki, 21-24 August 2024.

Mortelmans, T. 2022. *Blijken*: een evidentieel-miratieve outsider: Een corpusanalyse op basis van het CGN. *Nederlandse Taalkunde*, 27, 3, p. 293-327.

Raidt, E.H. 1972. *Afrikaans en sy Europese verlede.* Kaapstad: Nasou Beperk.

Sketch Engine. 2025. *afTenTen24 Corpus*. Accessed January 14, 2025.
https://www.sketchengine.eu/corpora-and-languages/afrikaans-text-corpora/

Truth-revelation discourse markers

Agnès Celle, Université Paris Cité, Chad Howe, University of Georgia

This paper deals with a paradigm of discourse markers derived from raising verbs: *turns out, turned out, ends up* and *ended up*. It aims to determine to what extent their use as discourse markers is driven by their semantic and syntactic profile. This use in detached position is characterized by impersonal subject and complementizer omission. It is taken to reflect the last stage of a grammaticalization process, the parenthetical use (*as it turns out, it turns out,*) being an intermediate stage (Howe and Heller 2010). Semantically, these discourse markers encode the revelation of truth, as illustrated in (1) and (2). Following AnderBois (2018), we argue that this revelation causes the speaker to perform a speech act to update the common ground. This illocutionary update may take various forms, depending on the speaker's expectation of outcome.

We make a distinction between *end up* discourse markers, which do not necessarily encode the revelation as counterexpectational, and *turn out* discourse markers, which carry a nonveridical presupposition. We relate this difference to the development of each discourse marker. Although *turn out* and *end up* are both resultative constructions that have developed discourse uses specialized in the revelation of truth, this development did not follow the same pathway. *Turn out* is known to have developed from a resultative lexical verb *(Our expedition up Vesuvius turned out very well*. OED, quoted from Serrano-Losada 2017) into a raising catenative verb (Huddleston & Pullum 2002), giving rise to two related constructions – a raised subject construction (*The expedition turned out to be a success*) and an extraposed subject construction (*It turned out that the expedition was a success*). It is the latter construction that is undergoing further development, as it may be used parenthetically (*as it turns out, it turns out,*) and as a discourse marker (*turns out, turned out*). As argued by Howe and Heller (2010) and Kaltenböck (2014), the impersonal construction is grammaticalizing into a pragmatic marker and, according to Serrano-Losada (2017) has acquired evidential and mirative meanings. The development of *ends up / ended up* as a mirative discourse marker is more recent and more difficult to trace as the raising verb generally takes a gerundial complement (*this demon ended up doing it*), even if the extraposed subject construction is also attested (*It ends up that this demon did it*).

In his analysis of the emergence of the mirative *end up*-parenthetical, Serrano-Losada (2020) argues that the extraposed subject construction results from a cooptation process (Kaltenböck et al. 2011; Heine 2013) driven by analogical modeling. We hypothesize that the difference in the process that led to the emergence of discourse markers has important consequences that are reflected in the use of these constructions. Specifically, we maintain that *end up* discourse markers may have simple mirative extensions, while *turn out* discourse markers may have complex mirative extensions (Barés Gómez & Fontaine 2021), allowing for abductive inferential reasoning. Only *turn out* seems to be compatible with deferred realization, which, as argued by Aikhenvald (2004: 202), combines evidentiality with mirativity by way of a post-factum inference:

1. Lily yelled at her boss yesterday. Ends up, she was fired.
2. Lily yelled at her boss yesterday. Turns out, she was fired.

While the outcome that is targeted by the discourse marker in (1) may only be a logical consequence of the preceding proposition, another interpretation is preferred for (2), namely that the counterexpectational outcome is the reason for the preceding proposition.

In other words, the grammaticalization process might be responsible for a complex mirative discourse marker in the case of *turn out*, involving a counterexpectational outcome resulting from an abductive inference, while the cooptation process might account for the simple mirative overtone of *end up* discourse markers. This difference might explain why the uses of *turn out* discourse markers are much more widespread than those of *end up* discourse markers.

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford, New York: Oxford University Press.
AnderBois, Scott. 2018. Illocutionary Revelations: Yucatec Maya Bakáan and the Typology of Miratives. *Journal of Semantics* 35(1). 171–206. https://doi.org/10.1093/jos/ffx019.
Barés Gómez, Cristina & Matthieu Fontaine. 2021. How Surprising! Mirativity, Evidentiality and Abductive Inference. In Teresa Lopez-Soto (ed.), *Dialog Systems* (Logic, Argumentation & Reasoning), vol. 22, 115–136. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-61438-6_7.
DeLancey, Scott. 2012. Still mirative after all these years. *Linguistic Typology* 16(3). 529–564.
Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6). https://doi.org/10.1515/ling-2013-0048.
Howe, Chad & Brooke Heller. 2010. "Turns out they weren't much of a stretch": Variable patterns of structural persistence. Presented at the 39th New Ways of Analyzing Variation (NWAV 39), 4-6 November, University of Texas, San Antonio.
Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. 1st edn. Cambridge University Press. https://doi.org/10.1017/9781316423530.
Kaltenböck, Gunther. 2014. Processibility. In Karin Aijmer & Christoph Rühlemann (eds.), *Corpus Pragmatics*, 117–142. 1st edn. Cambridge University Press. https://doi.org/10.1017/CBO9781139057493.007.
Kaltenböck, Gunther, Bernd Heine & Tania Kuteva. 2011. On thetical grammar. *Studies in Language* 35(4). 852–897. https://doi.org/10.1075/sl.35.4.03kal.
Serrano-Losada, Mario. 2017. Raising "turn out" in Late Modern English: The rise of a mirative predicate. *Review of Cognitive Linguistics* 15(2). 411–437. https://doi.org/10.1075/rcl.15.2.05ser.
Serrano-Losada, Mario. 2020. Analogy-driven change: the emergence and development of mirative *end up* constructions in American English. *English Language and Linguistics* 24(1). 97–121. https://doi.org/10.1017/S1360674318000266.

The Semantics of '*Turn out* Verbs' Cross-Linguistically:
Evidence from French, Spanish, and Serbian

Ana Stulic & Patrick Dendale

(Université Bordeaux Montaigne, AMERIBER-GRIAL & University of Antwerp, GaP)

Keywords: *turn out* verbs, achievements, non-evidential, mirative, truth

CONTEXT AND RESEARCH QUESTION

'*Turn out* verbs' came into the picture in the early 90s, when scholars started categorizing them as *markers of evidentiality* (e.g. Nuyts 1994 for Dutch *blijken*, Tobback & Lauwers 2012 for French *s'avérer;* Cornillie 2007 for Spanish *resultar;* Trbojević Milošević 2018 for Serbian *ispostavilo se*), as *markers of mirativity* (e.g. Bomans 2019 for Dutch; Serrano-Losada 2017 for English and Spanish), or as both (Mortelmans 2022 for Dutch).

These categorizations fail, however, to capture the core semantics of these verbs: *turn out* verbs do not encode a particular type of information source (see Dendale *et al*. 2024), and mirativity, we will argue, is not a key property of them. So, what then are the key elements of their semantics?

METHODOLOGY AND DATA

To answer that question, we inventoried, on the basis of corpora within SketchEngine, the main properties of the *turn out* verbs *s'avérer*, *resultar* and *ispostaviti se* as they appear in sentences like (1)-(3) – all translatable with *turn out* and 'the place is lovely' as information:

> (1) **Il s'avère que** l'endroit est charmant. (*frTenten23*)
> (2) Este lugar **resulta** encantador. (*esTenTen23*)
> (3) Mesto je, **ispostavi se**, divno. (contrived)

(1)-(3) illustrate three constructions available in the languages studied: (A) an impersonal construction with an extraposed subject; (B) a copular or semi-auxiliary construction with raised subject; and (C) a parenthetical construction.

FINDINGS

In our analysis, *turn out* verbs *s'avérer, resultar,* and *ispostaviti se* all denote an event that consists in the public emerging of true information. These are their key semantic features.

Aspectually, *turn out* verbs qualify as *achievements* (Vendler 1957). They display properties of *telicity* (Rothstein 2004:6)*, instantaneousness* (Piñón 1997), and *dynamicity* (Fleischhauer 2016:182). We will show how these are reflected in the data. Achievements also *presuppose* (rather than *lexicalize* –as accomplishments do) a preparatory event (Martin 2011), which, in the case of *turn out* verbs, is the acquisition and verification of information. This, typically, is a durative and an agentive activity. Properties which are vague or unspecified in the lexical meaning of these verbs, can be complemented by elements of the linguistic context. We will present the main ones.

Like other achievement verbs, *turn out* verbs are dynamic, and signal a change of property, namely an *epistemic* property, which is intertwined with the verbs' aspectual properties. The key epistemic property is that the information that emerged is true, a special kind of truth: a collective, 'impersonal'

truth. The reason for this is the absence, in the thematic structure of *turn out* verbs, of an Experiencer, whose presence could restrict the scope of the truth. The 'true' status of the emerged information is understood as being in contrast with the epistemic status of the information *before* its emergence, a status that can be signalled or suggested by the context. This contrast also explains mirative analyses of these verbs. The non-evidential status of *turn out* verbs, finally, comes from the fact that their lexical meaning is vague as for the nature of the preparatory event of 'acquisition of information'.

REFERENCES

Bomans, C. (2019). On the development of *blijken* and *turn out*: How did these verbs become linguistic markers of surprise? Poster presented at *A Germanic Sandwich* 2019. Amsterdam, 23-24 May 2019.

Cornillie, B. (2007). *Evidentiality and Epistemic Modality in Spanish (Semi-)Auxiliaries*. Berlin: Mouton De Gruyter.

Dendale, P. Izquierdo Alégria, D. & Stulic, A. (2024). '*Turn out*' verbs in European languages: Are they evidentials or something else? Talk at the 2024 SLE conference, Helsinki, 21-24 August 2024.

Fleischhauer, J. (2016). *Degree Gradation of Verbs*. Düsseldorf, Düsseldorf University Press.

Martin, F. (2011). Revisiting the difference between accomplishments and achievements. In: J. Mortelmans, T. Mortelmans, and W. de Mulder (eds), *From Now to Eternity*. Amsterdam/New York, Rodopi, 43-64. (Cahiers Chronos 22).

Mortelmans, T. (2022). *Blijken*: Een evidentieel-miratieve outsider: Een corpusanalyse op basis van het CGN. *Nederlandse Taalkunde*, 27, 3, 293-327.

Nuyts, J. (1994). *Layered Structures and Control in Language and Conceptualization*. Antwerp, University of Antwerp.

Piñón, Chr. (1997). Achievements in an Event Semantics, *Semantics and Linguistic Theory*. *Proceedings of SALT 7,* 277-293.

Rothstein, S. (2004). *Structuring Events: A Study in the Semantics of Lexical Aspect*. Blackwell Publishing.

Serrano-Losada, M. (2017). On English *turn out* and Spanish *resultar* mirative constructions. A case of ongoing grammaticalization? *Journal of Historical Linguistics*, 7, 160-189.

Sketch Engine, Create and search a text corpus, https://app.sketchengine.eu.

Tobback, E. & Lauwers P. (2012). Une analyse en miroir de deux semi-copules évidentielles: *s'avérer* et *se révéler. Revue romane*, 47, 49-75.

Trbojević Milošević, I. (2018). Corpus evidence for evidentials in English and Serbian political interviews. *Belgrade English Language and Literature Studies*, 10, 1, 131-156.

Vendler, Z. (1957). Verbs and Times. In: Z. Vendler (1967). *Linguistics in Philosophy*. Ithaca: Cornell University Press, 97-121.

A contrastive view of the French verb *s'avérer* and its Estonian counterparts

Anu Treikelder

(University of Tartu)

Keywords: 'turn out' verbs, contrastive analysis, semantics, French, Estonian

The French verb *s'avérer* ('to turn out to be', 'to prove to be') has been described as an evidential marker expressing indirect evidentiality. According to Tobback & Lauwers (2012: 64), this verb indicates that the information is obtained by inference from various types of evidence. Recently, it has been suggested that this verb belongs to a separate cross-linguistic category of 'turn out' verbs that may have a mirative use, indicating that the information is new or unexpected for the speaker (Dendale et al. 2024). Syntactically, the verb *s'avérer* can take different constructions: While it is most frequently used with predicative adjectives or nouns *(s'avérer inefficace* 'to prove to be ineffective' *s'avérer un monstre* 'to turn out to be a monster')*, it can also appear with that-complement clauses in its impersonal form (*il s'avère que* 'it turns out that') (Tobback & Lauwers 2012).

In lexicographic sources (GDEF, *Sõnaveeb*) three verbs can be found as the closest Estonian counterparts to *s'avérer*. Two of them are formed with the affix *–u,* which is used to derive change-of-state verbs (EG 2023: 301): *selguma* (derived from the adjective *selge* 'clear'; see example 1 below) and *osutuma* (derived from the verb *osutama* 'to indicate, to point out'; example 2). The third is a phrasal verb *välja tulema* (lit. 'to come out'; example 3) that has also a 'turn out' meaning besides its multiple other metaphorical uses. These verbs are not mentioned among Estonian epistemic and evidential markers (EKS 2017: 142-159), and, in general, their semantics and use seem not to have received much attention in linguistic research.

This paper aims to explore the counterparts of the verb *s'avérer* in the Estonian-French Parallel Corpus (CoPEF) in order to get a first insight into the class of 'turn out' verbs in Estonian. The material comes from the sub-corpora of Estonian and French fictional and non-fictional texts (130 examples in source texts and their translations in both directions), and the sub-corpus of debates in the European Parliament with various source languages (1785 examples). The corpus data reveal that the three verbs mentioned above are indeed the most frequent Estonian counterparts of *s'avérer*:

(1) *Dans la pratique, il **s'avère que** certains sont plus égaux que d'autres.*

Selgub [become_clear.PR.3SG]*, **et** [that] tegelikkuses on mõni võrdsem kui teised.*

'It turns out that in reality some are more equal than others'

(2) *Comme tu vois, ton sacrifice **s'est avéré** superflu.*

*Nagu näed, sinu ohver **osutus** [prove_to_be.PST.3SG] ülearuseks [superfluous.TR].*

'As you can see, your sacrifice proved to be superfluous.'

(3) *Deux ans après, **il s'est avéré que** la production devait être limitée.*

*Kaks aastat hiljem **tuli välja** [come_out.PST.3SG]*, **et** [that] toodangut tuleb piirata.*

'Two years later, it turned out that production had to be limited.'

However, other verbs occur and there are noticeable differences in the use of verbs between Estonian source texts and translations. This paper will examine the distribution of these verbs, depending partly on their morphosyntactic properties, but also on differences in their semantic features appearing in various contexts. This comparative study will hopefully also contribute to specifying the semantics and use of the French verb *s'avérer*, and especially to evaluating its evidential and/or mirative properties.

References

CoPEF = *Corpus parallèle estonien-français de l'Association franco-estonienne de lexicographie,* http://corpus.estfra.ee (13 January, 2025).

Dendale, Patrick, Ana Stulic and Dámaso Izquierdo-Alegría (2024). 'Turn out' verbs in European languages: Are they evidentials or something else? Talk at the *2024 SLE conference*, Helsinki, 21-24 August 2024.

EG 2023 = Metslang, Helle, Mati Erelt, Külli Habicht, Tiit Hennoste, Reet Kasik, PireTeras, Annika Viht, Eva Liina Asu, Liina Lindström, Pärtel Lippus, Renate Pajusalu, Helen Plado, Andriela Rääbis, Ann Veismann (2023), *Eesti grammatika* [Estonian grammar], Tartu: Tartu Ülikooli Kirjastus.

EKS 2017 = Erelt, Mati and Helle Metslang (eds.) (2017), *Eesti keele süntaks* [Estonian Syntax], *Eesti keele varamu III*, Tartu: Tartu Ülikooli Kirjastus.

Sõnaveeb https://sonaveeb.ee/ (13 January, 2025)

GDEF = *Grand dictionnaire estonien-français,* https://www.estfra.ee/ (13 January, 2025).

Tobback, Els and Peter Lauwers (2012), Une analyse en miroir de deux semi-copules évidentielles: *s'avérer* et *se révéler*, *Revue romane* 47, 49-75.

***Saltare fuori, uscire fuori, venire fuori*:**
different motions, different paths to the truth?

This research examines three Italian verb-particle constructions which can be translated into English as 'turn out': *saltare fuori, uscire fuori, venire fuori*. These "syntagmatic verbs" (Simone, 1997), combine a movement verb ('jump', 'exit', 'come') with the locative adverb *fuori* ('out(side)'), reinforcing a telic interpretation (Iacobini & Masini, 2005).

Building on Dendale, Izquierdo and Stulic's (2024) work on other languages, this study firstly aims to challenge the idea that these Italian constructions are evidential, as 'turn out' verbs are described in various languages (Cornillie, 2007; Serrano-Losada, 2017; Tobback & Lauwers, 2012). We analyse what sources of information – a core feature of evidentiality (Dendale & Tasmowski, 2001) – can be associated to corpus examples containing these constructions. Preliminary results suggest that, with these verbs, the information source can vary, being identified as inference (1), report (2), directed experience (3), or remaining ambiguous (4). If confirmed, this would support Dendale and colleagues' (2024) proposal that 'turn out' verbs express a "come-to-light event" rather than evidentiality.

(1) Il parlarne con uno stile in particolare fa parte anche di come siamo noi. Da qui può **uscire fuori che** i Baustelle sono dandy […].
'Talking about it in a particular style is also part of who we are. This might lead people to think that Baustelle are dandy […].'

(2) […] Dalle indagini Istat **è venuto fuori che** si fanno sempre meno figli.
'[…] Istat surveys have revealed that people are having fewer and fewer children.'

(3) Confrontando il mio prodotto finale col racconto ne **è venuto fuori** che sono identici.
'Comparing my final product with the story, it turned out that they are identical.'

(4) Da dove **salti fuori** che l'avvocato è 10-20 volte più caro è un mistero.
'Where the idea comes from that the lawyer is 10-20 times more expensive is a mystery.'

The second research question draws on Miecznikowski's (2018) approach to other Italian 'turn out' verbs and addresses the semantic differences among these constructions in terms of information source (1-4), mirative tone (5) and speaker's high (6) or low (4) commitment.

(5) Sta a vedere che adesso **salta fuori** che hanno ragione!!
'Watch it now, it turns out they're actually right!'

(6) […] qualcuno che forse (ipotesi!) ha fatto qualcosa (firmarsi con altri pseudonimi) che poi **salta fuori** che hai fatto (certezza!) proprio tu per primo.
'[…] someone who perhaps (hypothetically!) did something (used other pseudonyms) that later turns out (certainly!) you were the first to do yourself.'

Provisional findings indicate that the three expressions occur with all types of information sources, though *saltare fuori* appears more strongly associated with mirative tone and a higher degree of commitment compared to the others. If confirmed through further analysis, *saltare fuori* could be interpreted as a pragmatically marked expression, reflecting the semantics of *saltare*, which encodes manner by describing a forceful movement (Aurnague, 2011; Stosic et al., 2023). In conclusion, we aim to explore whether, and how, the semantic properties of these motion verbs resonate in their metaphorical meanings and uses.

1

References

Aurnague, M. (2011). How motion verbs are spatial : The spatial foundations of intransitive motion verbs in French. *Lingvisticae Investigationes*, *34*(1), 1-34. https://doi.org/10.1075/li.34.1.01aur

Clark, E. V. (1974). Normal States and Evaluative Viewpoints. *Language*, *50*(2), 316-332. https://doi.org/10.2307/412440

Cornillie, B. (2007). The continuum between lexical and grammatical evidentiality : A functional analysis of Spanish parecer. *Rivista di linguistica*, *19*(1), 109-128.

Dendale, P., Izquierdo, D., & Stulic, A. (2024). *'Turn out' verbs in European languages : Are they evidentials or something else?* 57th Annual Meeting of the Societas Linguistica Europaea.

Dendale, P., & Tasmowski, L. (2001). Introduction : Evidentiality and related notions. *Journal of Pragmatics*, *33*(3), 339-348. https://doi.org/10.1016/S0378-2166(00)00005-9

Iacobini, C., & Masini, F. (2005). Verb-particle constructions and prefixed verbs in Italian : Typology, diachrony and semantics. *Mediterranean Morphology Meetings*, *5*, 157-184. https://mmm.library.upatras.gr/index.php/mmm/article/viewFile/2388/2647

Miecznikowski, J. (2018). Evidential and Argumentative Functions of Dynamic Appearance Verbs in Italian : The Example of Rivelare and Emergere. In S. Oswald, T. Herman, & J. Jacquin (Éds.), *Argumentation and Language—Linguistic, Cognitive and Discursive Explorations* (Vol. 32, p. 73-105). Springer International Publishing. https://doi.org/10.1007/978-3-319-73972-4_4

Morita, T. (2022). What does deixis tell us about motion typology? Linguistic or cultural variations of speakers'" here" space vis-à-vis perceived physical events". *Neglected Aspects of Motion-Event Description: Deixis, asymmetries, constructions*, *72*, 25.

Serrano-Losada, M. (2017). On English *turn out* and Spanish *resultar* mirative constructions : A case of ongoing grammaticalization? *Journal of Historical Linguistics*, *7*(1-2), 160-189. https://doi.org/10.1075/jhl.7.1-2.07ser

Simone, R. (1997). Esistono verbi sintagmatici in italiano? *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche*, 155-170.

Stosic, D., Minoccheri, C., & Aurnague, M. (2023). *DinaVmouv (V2) : Description, INventaire, Analyse des Verbes de MOUVement. An annotated lexicon of motion verbs in French. Freely available at : Http://redac.univ-tlse2.fr/lexicons/dinaVmouv_fr.html.* [Software].

Tobback, E., & Lauwers, P. (2012). Une analyse en miroir de deux semi-copules évidentielles : S'avérer et se révéler. *Revue Romane. Langue et Littérature. International Journal of Romance Languages and Literatures*, *47*(1), 49-75. https://doi.org/10.1075/rro.47.1.03tob

Unexpected or expected? Rethinking the role of mirativity in the semantics of 'turn out' verbs through the case of Spanish *resultar*

Dámaso Izquierdo-Alegría
(ICS, University of Navarra)

Keywords: 'turn out' verbs, mirativity, (un)expectedness, Spanish,

Cross-linguistically, the semantics of 'turn out' verbs (e.g., English *turn out*, French *s'avérer*, Spanish *resultar*, Dutch *blijken*) has received limited attention, leaving their categorial status unclear. While some scholars classify them as markers of indirect evidentiality (Cornillie 2007, Lauwers & Tobback 2012, Serrano-Losada 2017, and Mortelmans 2022), recent findings demonstrate that these verbs are compatible not only with indirect modes of access to information, but also with direct ones (Dendale et al. 2024). This calls into question their evidential status.

Additionally, these verbs have been described in terms of mirativity—that is, as marking states of affairs as surprising, unexpected, or new (Serrano-Losada 2017, and Mortelmans 2022). Nonetheless, certain 'turn out' verbs are also compatible with states of affairs that are entirely expected or even previously known to the speaker. In some languages, this applies particularly to peripheral 'turn out' or 'turn-out'-like verbs, such as French *se trouver* in the impersonal construction but is rare with prototypical members of this group (*s'avérer*).

However, in Spanish, the most prominent 'turn out' verb, *resultar*, is not only compatible with unexpected states of affairs, but also accepts very frequently expected ones (examples 1-2), and states of affairs already known to the speaker (example 3), as will be shown with corpus data (EsTenTen18&23 and Timestamped). This challenges previous descriptions of 'turn out' verbs, including the recent proposal by Dendale et al. (2024), which argues that they introduce *new* true information that emerged at some point and was not part of the speaker's previous knowledge. Furthermore, these examples do not align with a specific group of occurrences described by Pérez Béjar (2020: 351-352) that just mark the beginning of a particular sequence and establish its topic.

> (1) El viernes predije en Twitter que Brian Vera le ganaría a Julio César Chávez Jr. **Resulta que** yo tenía razón. (EsTenTen18)
> *'On Friday, I predicted on Twitter that Brian Vera would beat Julio César Chávez Jr. **It turns out** I was right.'*

> (2) Ayer se fallaron los Premios Canarias. [...] Y **resulta que**, como era de esperar, han pasado olímpicamente de concedérselo a José Antonio Ramos [...]. (EsTenTen18)
> *'Yesterday, the Canary Islands Awards were decided. [...] And **it turns out that**, as expected, they completely ignored giving it to José Antonio Ramos [...].'*

> (3) El New York Times publicó un reportaje que muestra cómo la policía estadounidense utiliza datos personales para resolver crímenes. **Resulta que**, como es conocido, Google rastrea todos nuestros movimientos. (Timestamped2014-2021)

'The New York Times published a report showing how U.S. police use personal data to solve crimes. ***It turns out that****, <u>as is well-known</u>, Google tracks all our movements.'*

This presentation aims to disentangle the semantic complexity of Spanish *resultar* as a 'turn out' verb in the impersonal construction and evaluate its status as a mirative marker. Special attention will be given to the semantic link between mirative or mirative-like uses and non-mirative ones. The implications of these findings for the broader semantic description of 'turn out' verbs across languages will also be discussed.

References

Cornillie, Bert (2007), *Evidentiality and Epistemic Modality in Spanish (Semi-)Auxiliaries*, Berlin: De Gruyter.

Dendale, Patrick, Izquierdo-Alegría, Dámaso, & Stulic, Ana (2024), 'Turn out' Verbs in European Languages: Are they Evidentials or Something Else?, Talk at the *2024 SLE conference*, Helsinki, 21-24 August 2024.

Mortelmans, Tanja (2022), *Blijken*: een evidentieel-miratieve outsider: Een corpusanalyse op basis van het CGN, *Nederlandse Taalkunde* 27/3, 293-327.

Pérez-Béjar, Víctor (2020), *Resulta que...* De cláusula subordinante a elemento procedimental del discurso, *RILCE. Revista de Filología Hispánica* 36(3), 1193-1217.

Serrano-Losada, Mario (2017), On English *turn out* and Spanish *resultar* Mirative Constructions. A Case of Ongoing Grammaticalization?, *Journal of Historical Linguistics* 7, 160-189.

Tobback, Els, & Lauwers, Peter (2012), Une analyse en miroir de deux semi-copules évidentielles : *s'avérer* et *se révéler*, *Revue Romane* 47, 49-75.

Turkish *Turn Out* verbs and remarks from a general perspective

Dorian Pastor

(École Normale Supérieure; PSL University; AOROC)

Keywords: turnout; verb; Turkish; typology; resultative.

This talk discusses:

A. Four Turkish verbs, all translatable as 'to turn out' and fitting the definition of *Turn Out* verbs by Dendale et al. (2024): *meydan-a cıkmak* ('square-DAT exit') (1), *orta-ya çıkmak* ('middle-DAT exit') (2), *belli olmak* ('clear become') (3), and *çıkmak* 'to exit' (4):

(1) *Rus   ajan-ı   ol-duğ-u   **meydan-a   çık-tı**.*
    Russian   agent-POSS3   become-PTCP-POSS3   square-DAT   exit-PST.3
    '**It turned out** that s/he was a Russian spy.' (MED)

(2) *Almanya-da   genç-ler-in   ekonomi-den*
    Germany-LOC   young-PL-GEN   economy-ABL
    *anla-ma-dık-lar-ı   **orta-ya   çık-tı**.*
    understand-NEG-PTCP-PL-POSS3   middle-DAT   exit-PST.3
    '**It turned out** that young people in Germany do not understand the economy.' (MED)

(3) *Kadın-lar-ın   cinsellik-te-ki   altın   çağ-ı-nın*
    woman-PL-GEN   sexuality-LOC-ADJZ   gold   era-POSS3-GEN
    otuz   yaş   band-ın-da   ol-duğ-u
    thirty   age   around-POSS3-LOC   become-PTCP-POSS3
    ***belli   ol-du**.*
    clear   become-PST.3
    '**It turned out** that women's golden age in sexuality occurs in their thirties.' (MED)

The constructions in (1–3) are impersonal. They occur in the main clause of a complex sentence, where the subordinate clause is marked by the participial suffix *-dIk* (underlined in the gloss), which corresponds to Indo-European *that*-clauses. Reversely, *çıkmak*, is subject-focused:

(4) *Şükür   bi   kez   daha   haklı **çık-tı-m**.*
    thanks   one   time   more   right exit-PST-1SG
    'Luckily, **I turned out** to be right once again.' (WEB)

Preliminary analyses suggest that *ortaya / meydana çıkmak* on one hand and *belli olmak* on the other complement each other: *ortaya / meydana çıkmak* introduce wholly new information, while *belli olmak* confirms or completes pre-existing knowledge. Furthermore, *ortaya / meydana çıkmak* carry a sense of suddenness, unlike *belli olmak*.

1

B. *Turn Out* verbs from a general perspective. I suggest that:

i. *Turn Out* verbs are **resultative**: they denote the result of a process. The construction in (5a), i.e., impersonal construction, expresses the end point to what was a process of uncertainty or waiting (cf. Turkish, 1–3). The construction in (5b), i.e., personal construction, refers to a resultant state only (cf. Turkish, 4):

(5) a. ***it turned out that*** *it was a hot day*
    b. *this day **turned out** to be hot*

This can be illustrated through English when a *Turn Out* clause is expressed as secondary predication. In (6), it refers to the fear of a result to be <u>viewed as</u> true, while in (7) it refers to a result:

(6) *I just don't want it to turn out that Sam's father's, like, mixed up with al-Qaeda or something.* (Miller 2010: 63)

(7) *[A]t least I have a friend. But **you want me to turn out to be just like you**. <u>Just staying in your kitchen all day long.</u>* (Miller 1997: 94)

ii. Moreover, I suggest that there are no *Turn Out* "verbs" but rather:

• **resultative "pseudo-copulas"** (cf. "pseudo-copula" Suleymanov 2024; "semi-copule" Tobback & Lauwers 2012) referring to a personal assertion, with verbs, which may (e.g., *to turn out*) or may not (e.g., *to prove*) be used in impersonal constructions;

• **information-revealing "constructions"**, viz. impersonal constructions built with resultative "pseudo-copulas" (e.g. English *to turn out*, French *s'avérer*) as well as with various verbs, which are not resultative "pseudo-copulas" (e.g. Turkish *ortaya çıkmak* among others). These constructions are fundamentally resultative but the use of the impersonal *it* neutralizes the main subject, giving the impression that the revelation arises from external circumstances rather than a personal assertion.

Abbreviations
<u>Glosses:</u> 1 (first person) – 3 (third person) – ABL (ablative) – ADJZ (adjectiviseur) – DAT (dative) – GEN (genitive) – LOC (locative) – NEG (negation) – PL (plural) – POSS (possessive) – PST (past) – PTCP (participle) – SG (singular)
<u>Sources:</u> MED (newspapers) – WEB (websites, social medias, blogs)

References
Dendale, Patrick, Izquierdo-Alégria, Dámaso & Stulic, Ana (2024), "'Turn out' Verbs in European Languages: Are they Evidentials or Something else?", Talk given at the *57th Annual Meeting of the Societas Linguistica Europaea*, Helsinki: University of Helsinki.
Suleymanov, Murad (2024), "Syntax of Scents and Tastes: A Unique Case of Double-Nominative Argument Marking in Azeri", Talk given at the *57th Annual Meeting of the Societas Linguistica Europaea*, Helsinki: University of Helsinki.
Tobback, Els & LAUWERS, Peter (2012). "Une analyse en miroir de deux semi-copules évidentielles: *s'avérer* et *se révéler*", *Revue Romane. Langue et littérature. International Journal of Romance Languages and Literatures* 47(1), 49–75.

Literature

Miller, Kathleen Daisy (1999), *Give Me Your Answer*, Ontario: Porcupine's Quill.

Miller, Lynn (2010), *Crossing the Line*, Bloomington: AuthorHouse.

1018

Turn out verbs and constructions in Latin

Elena Zheltova
(Saint Petersburg State University)

Turn out verbs (such as English *turn out*, French *s'avérer*, Spanish/Catalan *resultar*) have been sufficiently investigated in quite a few modern languages but have never drawn the attention of scholars in Latin.

In previous research, the *turn out* verbs have been categorized as 'evidential markers' (Cornillie 2007, and Miecznikowski 2018), but this was challenged in some recent studies that have brought forward the idea of mirative rather than evidential semantics of *turn out* constructions (Serrano-Losado 2017, and Dendale et al. 2024).

In this study I attempt to: 1) identify such verbs in Latin and observe them in synchrony and diachrony, 2) make a generalized overview of their meanings and single out the contexts in which they show the *turn out* semantics, 3) to highlight the morphosyntactic and pragmatic constraints on their use as *turn out* verbs, and 4) clarify their semantic nuances in different contexts.

The corpus study was conducted by means of Packard Humanities Institute database (https://latin.packhum.org). Seven candidates were selected, of which only four appeared to fit the definition of *turn out* verbs, viz. *apparere* 'to appear, be clear', *invenire* 'to find, discover', *reperire* 'to discover, get to know', and *evenire* 'to come out, happen' (Glare 1968, s.v.). A total of 3112 instances of the four verbs were retrieved and examined.

Each verb proved to have a whole range of meanings, but the *turn out* semantics, with its overtones of unexpectedness and surprise, is revealed only in some temporal-aspectual forms, or in the passive voice, or in a special syntactic construction. Thus, *apparere* is used in this function only in the perfect (*apparuit*), in a parenthetical construction *ut apparuit* 'as it turned out' (e.g., Liv. 44.31.11), and in a combination with the accusative and infinitive construction, as in ex. (1):

> (1) ...*post paucos menses in coniuratione Catilinae esse eum* **apparuit**. (Asc. *In Toga Candida* 73.15)
> 'A few months later **it turned out that** he had participated in Catiline's conspiracy.'

The *turn out* meaning of *invenire* and *reperire* proved to be restricted to only passive perfect forms, ex. (2–3):

> (2) *Caesaris coniunx ... sola* **est** *caelesti digna* **reperta** *toro*. (Ov. *Pont.* 3.1.118)
> 'Caesar's wife.. **turned out to be** the only one worthy of the heavenly chambers.'

> (3) *Tua quae fuit Palaestra, ea filia* **inventast** *mea!* (Plaut. *Rud.* 1364)
> 'This Palestra who was yours **turned out to be** my daughter!'

As regards *evenire*, it occurs with a *turn out* meaning very rarely and only in the impersonal construction *evenit ut* 'it happened that, it turned out that' (e.g., Liv. 1.22.3.1).

The study has shown that: 1) *turn out* verbs do exist in Latin, 2) the *turn out* meaning of such verbs coexisted with other meanings in all periods of Latin, 3) some lexical units (viz. adverbs, particles) can block the *turn out* meaning while others will favor it, 4) Latin *turn out* verbs allow both an evidential and a mirative reading depending on the syntactic-pragmatic conditions, the original semantics, and *Aktionsart* of a particular verb. The last point needs further discussion, as does the question of whether

the imperfect of delayed realization (Aikhenvald 2012) can be considered a good competitor of the verbs under consideration.

References

Aikhenvald, Alexandra (2012), The essence of mirativity, *Linguistic Typology*, 16(3), 435–485.

Cornillie, Bert (2007), *Evidentiality and Epistemic Modality in Spanish (semi-)auxiliaries*, Berlin: Mouton De Gruyter.

Dendale, Patrick, Izquierdo Alégria, Dámaso, Stulic, Ana (2024), 'Turn out' verbs in European languages: Are they evidentials or something else? *Talk at the 2024 SLE conference*, Helsinki, 21–24 August 2024.

Glare, Peter Jeffrey William (1968), *Oxford Latin Dictionary,* Oxford: Clarendon Press.

Miecznikowski, Johanna (2018), Evidential and Argumentative Functions of Dynamic Appearance Verbs in Italian: The Example of *Rivelare* and *Emergere*, in S. Oswald*,* T. Herman, and J. Jacquin (eds), (2018), *Argumentation and Language-Linguistic, Cognitive and Discursive Explorations*, Cham: Springer, 73–104.

Serrano-Losado, Mario (2017), Raising turn out in Late Modern English. The rise of a mirative predicate, *Review of Cognitive Linguistics* 15(2), 411–437.

Augmenting mirativity?
Pseudo-coordinations with (TURN) OUT particle verbs in Italian

Francesca Masini & Flavio Pisciotta

(University of Bologna & University of Salerno)

Keywords: mirativity; pseudo-coordination; particle verbs; TURN OUT verbs; Italian

Like many other European languages, Italian has (synthetic and analytic) TURN OUT verbs coming from the domain of motion that seem to convey grammatical meanings, namely evidentiality and mirativity:

(1) *Dall'inchiesta britannica sui pacchi-bomba **emerge che** la soffiata è venuta da un pentito di Al Qaeda.* [Miecznikowski 2018:91]
'British parcel bomb investigation reveals tip came from al Qaeda turncoat'

(2) *[…] l'operatrice verifica e **salta fuori che** hanno spedito tutto all'indirizzo sbagliato.* [Pisciotta 2025:2]
'the operator checks and it turns out that they sent everything to the wrong address'

In particular, motion particle verbs with *fuori* 'out' (*venire/uscire/saltare fuori* 'come/exit/jump out'), like (2), can behave as dynamic verbs of appearance, expressing that some information has been acquired unexpectedly (Pisciotta 2025). The emergence of this evidential/mirative meaning seems to trace back to (at least) the XVI Century and then slightly increases afterwards (Marion 2024).

In this paper, we explore a so far unnoticed aspect of their behaviour, namely their occurrence in (emerging) pseudo-coordinations (Ross 2016), and how this relates to their original spatial and TURN OUT meanings. Pseudo-coordination is renowned for carrying mirativity in various languages especially with TAKE verbs (e.g., Wiklund 2008), including Italian:

(3) *Avevamo deciso di non dirlo a nessuno, poi lui **prende e fa** un numero di telefono* [Masini et al. 2019:133]
'We had decided not to tell anyone, then he [unexpectedly] makes a phone number'

However, other types of pseudo-coordinations have been traced (Mauri & Masini 2022), some of which (still unexplored) feature precisely motion particle verbs with *fuori*:

(4) *Abbiamo scritto le regole, progettato i campi, abbiamo fatto tutto. Poi, dopo 50 anni, qualcuno è **saltato fuori e ha detto**: 'Siete dei bugiardi* […] [itTenTen20]
'We wrote the rules, we designed the fields, we did everything. Then, after 50 years, someone jumped out and said, "You guys are liars […]'

When embedded in pseudo-coordinations these verbs seem to express a further step towards the expression of mirativity, as they seem to be devoid of any spatial or cognitive meaning and serve as markers of unexpectedness. However, it is not clear whether this meaning is connected to their TURN OUT use or developed in parallel, favoured by their insertion in the mirative pseudo-coordination construction.

The aim of this paper is (i) to confirm that we are dealing with constructions conveying mirativity; (ii) to unveil other grammatical/textual functions these constructions may carry; and (iii) to understand if and how the meaning of V_{motion} *fuori* in pseudo-coordinations is connected to their TURN OUT meaning, considering that other (synthetic) TURN OUT verbs coming from the motion domain (like *emergere* (1)) don't participate in pseudo-coordinations.

To address these issues, we gather occurrences of the <V_{motion} *fuori e* V> constructions from itTenTen20, as we need a big corpus given the emergent status of these constructions (manual checking is currently in progress). We will then annotate the occurrences according to relevant formal/functional parameters (e.g., presence of mirativity and other values, semantics of second V, type of subject, etc.) and we will then analyze them both quantitatively and qualitatively.

References

Marion, Laura (2024), *Verbi sintagmatici in diacronia: uno studio corpus-based delle particelle fuori e via*. Unpublished MA Thesis, University of Bologna.

Masini, Francesca, Simone Mattiola & Greta Vecchi (2019), La costruzione "prendere e V" nell'italiano e contemporaneo. In Bruno Moretti, Aline Kunz, Silvia Natale, and Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana. Berna, 6-8 settembre 2018*, Milano: Officinaventuno, 115–137.

Mauri, Caterina & Francesca Masini (2022), Diversity, discourse, diachrony: A converging evidence methodology for grammar emergence. In Miriam Voghera (ed.), *From Speaking to Grammar*, Bern: Peter Lang, 101-150.

Miecznikowski, Johanna (2018), Evidential and argumentative functions of dynamic appearance verbs in Italian: The example of 'rivelare' and 'emergere'. In Steve Oswald, Thierry Herman, and Jérôme Jacquin (eds.), *Argumentation and language - Linguistic, cognitive and discursive explorations*, Cham: Springer, 73–105.

Pisciotta, Flavio (2025) [in press], Ai confini tra evidenzialità e miratività: usi d'apparenza dei verbi sintagmatici di movimento con fuori. In Massimo Cerruti, Cristina Onesti, Riccardo Regis, and Davide Ricca (eds.), *Continuo e discreto nelle scienze del linguaggio. Atti del LVI Congresso internazionale della Società di Linguistica Italiana, 14-16 settembre 2023, Torino*, Milano: Officinaventuno.

Ross, Daniel (2016), Between coordination and subordination: Typological, structural and diachronic perspectives on pseudocoordination. In Fernanda Pratas, Sandra Pereira, and Clara Pinto (eds.), *Coordination and subordination: Form and meaning*, 209-243. Newcastle upon Tyne: Cambridge Scholars Publishing.

Wiklund, Anna-Lena (2008), Creating surprise in complex predication. *Nordlyd: Tromsø University Working Papers on Language & Linguistics* 35, 163-187.

English 'turn out': Development and current usage

Diana Lewis
(University of Aix Marseille)

This corpus-based study examines the phrasal verb *turn out* in British English, with special focus on its 'become known [to speaker]' usage (1).

(1)   *we had to get the train ... well **it turned out** we got on the wrong train* [BNC2014]

What diachronic pathway has produced the present-day English (PDE) *turn out* constructions, and what are the usage patterns to-day? Corpus data come mainly from the OBC (1720s-1913) (Huber et al. 2016) and the BNC2014 (2010s) (Love et al. 2017). PDE *turn out* has evolved from a motion verb with directional particle, a development that includes features of grammaticalization. Semantically, *turn out* develops more abstract meanings, resulting, by the late eighteenth century, in a polysemy that includes occurrence ('happen') (2), result or change of state ('end up') (3), and new knowledge ('become known / be revealed') (4).

(2)   *he has been highly respected .. till this unfortunate affair **turned out*** [OBC, 1780]
(3)   *the medical assistance that she met with indeed **turned out** so successful ..* [OBC, 1791]
(4)   *I was let in by a woman that afterwards **turned out** to be Mrs Green* [OBC, 1782]

The diachronic paths and conceptual contiguities can be modeled in a semantic map of the polysemy development. Syntactically, there is expansion in complementation, from [*turn out* {Ø|AdjP|AdvP|NP}] to infinitival [*turn out to* {*be*+compl.|VP}] and clausal [*it turn- out* (*that*) clause]. Parallels are drawn with the trajectories of verbs like *seem* (Gisborne and Holmes 2007). In the PDE data, all these constructions occur. The most frequent (235/393,60%) is the clausal complement construction, corresponding in Creissels' typology of impersonals to the special lexico-semantic impersonal construction (2008: 25). The second most frequent is the infinitival (so-called 'raising') construction (88/393,22%). The 'become known' usage presupposes a non-knowing or false assumption state prior to a knowing state, as in (1), and this presupposition is sometimes made explicit in contrastive and counterexpectational discourse constructions (5), which are the most striking collocation of this *turn out*.

(5)   ***they thought** I was really good at product design and **it turns out** I'm not at all* [BNC2014]

The new knowledge often appears surprising to speaker and/or other participant, but I argue that the surprisal effect is better treated as discourse pragmatic than as semanticized mirativity, and may be linked to the use of *it turn- out* as a presentative sequence adding discourse prominence to the following proposition. *Turn out* has much in common with happenstance verbs expressing agentless events, like *seem*, *happen* or, in earlier periods of English, *gelimpan*, *bifallen*. It continues to evolve, however, and shows signs of incipient adverbialization along the lines of *methinks* (Palander-Collin 1997), *maybe* (López Couso and Méndez Naya 2016), etc. It occurs in the present tense at right

periphery and as a post-subject parenthetical (both are sentence-adverbial positions), with a potential split emerging between tensed matrix clause *it turn- out that* and invariant peripheral *(it) turns out*.

References

Creissels, Denis (2008), Impersonal and related constructions: A typological approach. Lectures presented at the University of Tartu, 2-3 June. MS. Lyon: University of Lyon 2. <http://deniscreissels.fr>

Gisborne, Nikolas and Jasper Holmes (2007), A history of English evidential verbs of appearance, *English Language and Linguistics* 11(1), 1-29.

Huber, Magnus, Magnus Nissel and Karin Puga (2016), *Old Bailey Corpus 2.0*. hdl:11858/00-246C-0000-0023-8CFB-2.

López Couso, María José and Belén Méndez Naya (2016), From clause to adverb: On the history of *maybe*. In Gunther Kaltenböck, Evelien Keizer and Arne Lohmann (eds), *Outside the Clause: Form and Function of Extra-clausal Constituents*, Amsterdam: John Benjamins, 157-176.

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), 319–344.

Palander-Collin, Minna (1997), A medieval case of grammaticalization, methinks. In Matti Rissanen, Merya Kytö and Kirsi Heikkonen (eds), *Grammaticalization at Work: Studies of Long-term Developments in English*, Berlin: De Gruyter Mouton, 371-404.

Turn out **verbs from an English-Swedish contrastive perspective**

Karin Aijmer, University of Gothenburg

Turn out verbs such as English *turn out* and *prove to* are closely associated with perception and appearance. They have therefore been compared with *seem*-verbs. However, there are many unclarities with regard to the types of meaning they have and how these meanings should be accounted for particularly when the verbs are used in impersonal constructions with a *that*-clause or with a raised subject followed by an infinitive (Dendale et al. 2024). According to these authors, turn out verbs express the emergence of new-knowledge. The verbs are also characterized by their lexical and syntactic features in different languages. The properties of the English verb *turn out* have, for instance, been discussed by Serrano-Losada (2017) and Miecznikowski (2018) has described the Italian verbs *emergere* and *rivelare* in contexts where they acquire the function of expressing emerging new knowledge.

The present study compares the English verbs *turn out* and *prove to* with their correspondences in Swedish using a parallel corpus methodology. The material for the cross-linguistic analysis comes from the fiction and non-fiction parts of the English-Swedish Parallel Corpus (about three million words) (Altenberg and Aijmern2000). The methodology involves the following steps:

- what are the lexical correspondences of *turn out* and *prove to* in Swedish?

- in what syntactic constructions are the Swedish verbs found and how should the types of verbs be described (e.g. as having a raised subject followed by an infinitive or as an impersonal verb?)

- what meanings and semantic nuances do the verbs express?

The findings show, for example, that the most frequent Swedish lexical correspondence of *turn out* is a reflexive verb *visa sig* 'show oneself'. *Visa sig* is found both as an impersonal verb followed by a *that*-clause *det visar sig att* 'it shows itself that' and with a raised subject followed by an infinitive *visar sig vara* 'shows itself to be'.

The correspondence *komma fram* lit. 'come forwards' covers the meanings 'come out', 'emerge', and 'reveal'. A mirative overtone is suggested by the rendering of *it turned*

out as *som det nu föll sig* lit. 'as it now fell itself', 'as it happened'. The types of correspondence also involve aspect e.g. *bli* 'become' and *komma att bli* 'come to be'.

By examining the correspondences of the Swedish *visa sig* in English we can add new lexical realizations of turn out verbs. The most frequent English correspondences of Swedish *visa sig* 'show itself' are *prove (to be)* and *turn out that (to be)*. Other lexical correspondences are *emerge, appear, be found to, it is found that, show to be, show oneself to be, reveal oneself to, seem to be, look, become clear, come to be*. Some of the correspondences have the meaning emergence of new knowledge while other correspondences refer to the result of the process of knowledge acquisition.

References

Altenberg, B. and K. Aijmer. 2000. The English-Swedish Parallel Corpus. A resource for contrastive research and translation studies. In Mair, C. and M. Hundt (eds), *Corpus linguistics and linguistic theory. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20).* Freiburg im Breisgau 1999. Amsterdam/Atlanta: Rodopi. 15-33.

Dendale, P., Izqierdo Alégria,D. and A. Stulic. 2024. 'Turn out' verbs in European languages: Are they evidentials or something else? Talk at the 2024 SLE conference, Helsinki, 21-24 August 2024.

Miecznikowski, J. 2018.Evidential and argumentative functions of dynamic appearance verbs in Italian: The Example of *rivelare* and *emergere*. In: Oswald, S*., Herman, T. and J. Jacquin (eds). *Argumentation and language-Linguistic, cognitive and discursive explorations*. Springer, p.73-104.

Serrano-Losada, M. 2017. Raising *turn out* in Late Modern English. The rise of a mirative predicate. *Review of Cognitive Linguistics* 15(2): 411-437.

'Turn out' and appearance verbs in Danish and Greenlandic: Grammatical or lexical?

The development of 'turn out' verbs and appearance verbs like 'seem' is often discussed in relation to grammaticalization (see e.g. Lauwers & Tobback 2013; Miecznikowski 2018; Serrano Losada 2020 on 'turn out' verbs). This may seem natural in so far as there is agreement that, for instance, appearance verbs with a non-propositional argument (1) are lexical, whereas variants used as raising verbs in construction with a propositional expression (2) are often taken to be grammatical.

(1) *They look like bastards*
(2) *They seem to be bastards*.

Yet, 'turn out' verbs and appearance verbs in construction with a propositional expression present a challenge to all theories of grammaticalization, and to theories of the lexical-grammatical distinction in the first place. On the one hand, at least some of them meet criteria of lexical status, such as modifiability (e.g. Bybee, Perkins & Pagliuca 1994: 7; Keizer 2007; Boye 2024). For instance, English *turn out* and *seem* can both be modified by *long ago*. Thus, in (3), *long ago* must be read as modifying the superordinate verb *turn out* rather than the subordinate predicate (*be bastards*).

(3) *It turned out long ago that they were bastards*.

On the other hand, they are found in constructions that have been associated with grammaticalization. They occur as parentheticals (4) as well as in subject-to-subject raising constructions (2), (5) (e.g. Serrano Losada 2020).

(4) *They were, it turned out, bastards.*
(5) *They turned out to be bastards*.

Parentheticals are known to give rise to grammatical elements (e.g., Schoonjans 2012; Boye & Harder 2021), and there is a long tradition for discussing raising verbs in relation to grammaticalization (e.g. Ross 1969; Traugott 1997).

The Danish 'turn out' verb par excellence, *vise sig*, lit. 'show itself', forms a paradigm together with appearance verbs like *synes* 'seem'. A defining feature of the paradigm is that its members occur in active voice with a nominative with infinitive construction. This is illustrated in (6) with *vise sig* 'turn out' (gloss abbreviations adhere to The Leipzig Glossing Rules).

(6) *De   vis-te   sig   at   være   nogle   skiderikker.*
     3PL   turn-PST REFL   to   be   INDF.PL bastards
     'They turned out to be bastards.'

In the present paper, we first show that the properties mentioned above are found also in Danish *vise sig* 'turn out' and appearance verbs, and building on the account in Boye (2010), we propose a theoretical basis for disentangling grammatical from lexical status in relation to these verbs, while simultaneously highlighting the special circumstances that may lead to confusion. Subsequently, we stress our points in a discussion of Greenlandic. In Greenlandic, the closest counterparts of 'turn out' and appearance verbs are suffixes such as *-gunar* 'look like, it seems, no doubt'. These affixes include both lexical and grammatical members, and

crucially, they display a clear distributional reflection of the distinction between 1) lexical predicate with non-propositional argument, 2) lexical predicate with propositional argument, and 3) grammatical element modifying a propositional expression.

References

Boye, Kasper. 2010. Raising verbs and auxiliaries in a functional theory of grammatical status. In: Kasper Boye & Elisabeth Engberg-Pedersen (eds.), *Language usage and language structure*, Berlin: Mouton de Gruyter. 73-104.

Boye, K. 2024. Evidentiality, discourse prominence and grammaticalization. *Studies in Language* 48.3. 575-607.

Boye, K. & P. Harder. 2021. Complement-taking predicates, parentheticals and grammaticalization. *Language Sciences* 88, 101416.

Bybee, Joan, Revere D. Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, Aspect, and Modality in the languages of the world.* Chicago: Chicago University Press.

Keizer, Evelien. 2007. The lexical-grammatical dichotomy in Functional Discourse Grammar. *ALFA: Revista de Linguística* 51.2.

Lauwers, Peter & Els Tobback. 2013. Copularisation processes in French: Constructional intertwining, lexical attraction, and other dangerous things. *Folia Linguistica Historica* 34.1. 115-147.

Miecznikowski, Johanna. 2018. Evidential and argumentative functions of dynamic appearance verbs in Italian: The Example of Rivelare and Emergere. In: Oswald, S., T. Herman, & J. Jacquin (eds), *Argumentation and Language-Linguistic, Cognitive and Discursive Explorations*. Springer. 73-104.

Ross, John R. 1969 Auxiliaries as main verbs. In: W. Todd (ed.). *Studies in Philosophical Linguistics*, 77-102. Evanston, IL: Great Expectations Press.

Schoonjans, Steven, 2012. The particulization of German complement-taking mental predicates. *Journal of Pragmatics* 44. 776-797.

Serrano Losada, Mario. 2020. Analogization at Work: Evidential and Mirative as-Parentheticals with Raising Verbs in Late Modern English. In: P. Guerrero-Medina, M. P. Gutiérrez & M. V. Redondo (eds.), *Advances in English and American Studies: Current developments, future trends*. Córdoba: UCOPress / Editorial Universidad de Córdoba. 163-173.

Traugott, Elisabeth C. 1997 Subjectification and the development of epistemic meaning: The case of *promise* and *threaten*. In: Toril Swan & Olaf J. Westvik (eds.), *Modality in Germanic Languages: Historical and Comparative Perspectives*, Berlin/NewYork: Mouton de Gruyter. 185–210.

When and how did Italian *turn out* phrasal verbs turn out?

Luisa Brucale, Irene De Felice, Maria Napoli & Francesca Strik-Lievers
(University of Palermo, University of Piemonte Orientale, University of Piemonte Orientale &
University of Genoa)

Keywords: phrasal verbs, lexicalization, diachrony, Italian, *turn out* constructions

In Italian, many *turn out* verbs (Dendale et al. 2024) are phrasal verbs formed by a predicate of motion followed by *fuori* ('out'). We analyze this class of *turn out* verbs in Italian from a diachronic perspective. Phrasal verbs are known to be a Romance innovation (Iacobini/Masini 2006; Masini 2006). In Latin, motion verbs could occur with the adverb *foras* ('out'), but they only expressed physical movement (Lorusso et al. 2022). To trace the diachronic development of these *turn out* phrasal verbs, we analyze all occurrences of *venire* 'to come', *uscire* 'to go out', *saltare* 'to jump', and *spuntare* 'to pop up' combined with *fuori* in major Italian corpora spanning from the 13th to 21st centuries (MIDIA, OVI, ItTenTen20); contemporary spoken Italian is also considered (KIParla). Each occurrence is manually annotated for syntactic (i.e. number and type of arguments) and semantic features to identify the key stages in the evolution of these verbs as full-fledged *turn out* verbs.

For instance, until the 18th century, *venire fuori* exclusively expressed motion in space. The subject was typically animate; the origin of the motion was often specified; *fuori* occurred both pre- and post-verbally:

(1)   *Però **fuor** della terra **era venuto*** (MIDIA, Boiardo, *Orlando innamorato*, 1483)
        'He had come out of the land'

By the mid-19th century, *venire fuori* seems to be fully lexicalized (Bernini 2012) as a telic phrasal verb, with *fuori* stabilized in post-verbal position, abstract subjects and/or origins, and metaphorical meanings:

(2)   *"Ho piacere che **sia venuto fuori** questo discorso"* (MIDIA, Fogazzaro, *Daniele Cortis*, 1885)
        "I'm glad this topic came up"'

In the 20th century, among non-spatial uses more complex syntactic structures emerged, where *venire fuori* is followed by infinitive clauses introduced by *a* 'to' (*venire fuori **a dire***, 'to come up saying'), gerunds (*venire fuori **dicendo***, 'to come up saying'), prepositional phrases introduced by *con* 'with' (*venire fuori **con un argomento***, 'to come out with an argument'), and predicative adjectives (*venire fuori **diverso***, 'to turn out different'). Finally, a complement clause introduced by *che* 'that' (3) and a parenthetical use (4) also appear, which fully align *venire fuori* with the definition of *turn out* verbs as expressing the emergence of new and true knowledge (Dendale et al. 2024), and often mirativity:

(3)   *Una palestra dove cioè **fosse venuto fuori che** qualcuno aveva fatto a botte* (KIParla)
        'A gym where, I mean, it ever came out that someone had gotten into a fight'

(4)   *Google ha anche rimosso - **è venuto fuori** - i link a parecchi articoli* (itTenTen20)
        'Google also removed - it turned out - the links to several articles'

Our diachronic analysis highlights potential correlations between the stages of lexicalization of these phrasal verbs in Italian and the gradual emergence of features often associated with *turn out* verbs across languages, such as mirativity and factuality (Serrano-Losado 2017; Miecznikowski 2018; Pérez-Béjar 2020). The data also provide insights into the main factors that shape this correlation, considering the semantics of the verb and *fuori*, telicity, syntactic constructions, and argument structure.

References

Bernini, Giuliano (2012), Lexicalization and grammaticalization in the area of multi-word verbs: A case study from Italo-Romance, in V. Bambini et al. (a cura di), *Linguaggio e cervello – Semantica / Language and the Brain – Semantics. Atti del XLII Congresso Internazionale di Studi della SLI* (Pisa, 25-27 settembre 2008), Roma: Bulzoni, 131–154.

Dendale, Patrick, Stulic, Ana & Izquierdo Alegría, Dámaso. (2024), *'Turn out' verbs in European languages: Are they evidentials or something else?* Talk at the 2024 SLE conference, Helsinki, 21–24 August 2024.

Iacobini, Claudio & Masini, Francesca (2006), The emergence of verb-particle constructions in Italian: Locative and actional meanings, *Morphology* 16, 155–188.

Lorusso, Paolo, Oniga, Renato & Re, Alessandro (2022), Le origini latine dei verbi sintagmatici romanzi: Osservazioni sull'uso dell'avverbio *foras*, *L'Analisi Linguistica e Letteraria* 30(3), 5–34.

Masini, Francesca (2006), Diacronia dei verbi sintagmatici in italiano, *Archivio Glottologico Italiano* 91(1), 67–105.

Miecznikowski, Johanna (2018), Evidential and argumentative functions of dynamic appearance verbs in Italian: The example of *rivelare* and *emergere*, in S. Oswald, T. Herman and J. Jacquin (eds), *Argumentation and language-linguistic, cognitive and discursive explorations*, Cham: Springer Verlag, 73–104.

Pérez-Béjar, Victor (2020), *Resulta que*... De cláusula subordinante a elemento procedimental del discurso, *Rilce. Revista De Filología Hispánica* 36(3), 1193–1217.

Serrano-Losado, Mario (2017), On English *turn out* and Spanish *resultar* mirative constructions. A case of ongoing grammaticalization? *Journal of Historical Linguistics* 7, 160–189.

Corpora

ItTenTen20 = *Italian Corpus from the Web* (https://www.sketchengine.eu/ittenten-italian-corpus/)
KIParla = *Corpus KIParla. L'italiano parlato e chi parla italiano* (https://kiparla.it)
MIDIA = *Morfologia dell'Italiano in DIAcronia* (https://www.corpusmidia.unito.it)
OVI = *Corpus OVI dell'Italiano antico, Istituto Opera del Vocabolario Italiano* (http://gattoweb.ovi.cnr.it/)

Does the English phrasal verb *turn out* express novelty *and* mirativity?

Marc Fryd, Université de Poitiers

The English *turn out* phrasal verb has been deemed (Serrano-Losada 2017a, 2017b) to show "evidential" properties and to express "mirativity", understood by Aikhenvald (2012: 437) to encompass the following sub-types: Sudden discovery; Revelation or realization; Surprise; Unprepared mind; Counterexpectation; and New information, where each type may apply to Speaker, Addressee, or a main protagonist.

The first claim, regarding evidential properties, is disputed by Dendale *et al.* (2024), who argue that, across languages, *turn out* verbs do not by themselves mark the source of information and so cannot be called "evidential".

I plan to discuss the second claim and the nature of any mirative overtones elicited by *turn out*.

According to Serrano-Losada (2017: 166), example (1) illustrates a "conventionalised" mirative implicature on the grounds of the suggested oddity of the added negative comments:

> (1) It turns out that elephants have an advanced sense of self, # not that this is a newsworthy, unexpected or surprising discovery.

On the basis of examples (2-3-4), I suggest that novelty alone may qualify as a core feature of *turn out*, but not surprise or unexpectedness:

> (2) {a.Unsurprisingly / b.Surprisingly}, it turns out that elephants have an advanced sense of self.
> (3) {a.Against our expectations / b.As we all suspected}, it turns out that elephants have an advanced sense of self.
> (4) {a. After careful consideration, / b.??As already established}, it turns out that elephants have an advanced sense of self.

While (2-3) show that surprise and unexpectedness are only optional with *turn out*, example (4b.) highlights the verb's incompatibility with prior (proven) knowledge and thus underlines novelty as a core feature of the verb in the context of occurrence. Example (5) further demonstrates how *turn out* is disallowed in non-novel contexts:

> (5) It was a glorious day and we all went to the beach. ??As it turned out, we had a lovely time.

The tautological contents of the clauses in (5) (*glorious day ≡ lovely time*) indeed violate the novelty constraint imposed by *turn out*.

By contrast, *turn out* is licensed in example (6), which exhibits both novelty and a polarity change: *foul day (foul time)* ➜ *foul day (lovely time)*:

> (6) It was a foul day but we all went to the beach nonetheless. As it turned out, we had a lovely time.

There is a wide body of research in cognitive psychology on "novelty". A recent trend (Foster & Keane (2015), Meyer *et al.* (2015), Reichardt *et al.* (2020), Skavronskaya *et al.* (2020)), distinguishes between "low" and "high" novelty, the latter alone leading to unexpectedness and

surprise. This is congruent with Reizenstein *et al*. (2019), where the experience of unexpected events is depicted in terms of *schema-discrepancy*, i.e. discordance between an experience at hand and the sum of an individual's mental schemata.

Such a pattern is evidenced in instances of *turn out* such as (6), which may adequately be labelled as mirative, but not in other instances such as (2a.) and (3b.), where novelty entails no attendant schema discrepancy. On the basis of these considerations, *turn out* can only be defined as compatible with mirativity but not as a core mirative.

References:

Aikhenvald, A. Y. (2012). "The Essence of Mirativity." In: *Linguistic Typology* 16(3): 435-485.

DeLancey, S. (1997). Mirativity: the grammatical marking of unexpected information. *Linguistic Typology* 1: 33-52.

DeLancey, S. (2001). The mirative and evidentiality. *Journal of Pragmatics* 33: 369-382.

DeLancey, S. (2012). Still mirative after all these years. *Linguistic Typology* 16: 529-564.

Dendale, P. (2019a). 'Non-evidentials': the case of verbs like *s'avérer*, *turn out*, *blijken*, *resultar… Talk at the 16th International Pragmatics Conference, Hong Kong, 9-14 June, 2019.*

Dendale, P. (2019b). Pourquoi *s'avérer* (et ses equivalents dans d'autres langues) ne sont pas des évidentiels. *Talk at the CIPLR conference, Copenhague, 1-6 July 2019.*

Dendale, P. Izquierdo Alegria, D. & Stulic, A. (2024). '*Turn out'* verbs in European languages: Are they evidentials or something else? *Talk at the 2024 SLE conference, Helsinki, 21-24 August 2024*.

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology* 81: 74-116.

Friedman, V. A. (1986) Evidentiality in the Balkans: Bulgarian, Macedonian, and Albanian. In: Wallace Chafe and Johanna Nichols (eds.), *Evidentiality: the linguistic coding of epistemology.* Norwood, NJ, 168-187.

Fryd, M. (2023). La notion de "pure" mirativité. *Verbum* 45(2): 295-325.

Guentchéva, Z. (2017). An enunciative account of admirativity in Bulgarian. *Review of Cognitive Linguistics* 15(2): 540-575.

Meyer, W. U., Reisenzein, R., & Schützwohl, A. (1997). Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion* 21(3): 251-274.

Reichardt, R., Polner, B., & Simor, P. (2020). Novelty manipulations, memory performance, and predictive coding: The role of unexpectedness. *Frontiers in Human Neuroscience*, *14*, 152.

Reizenstein, R., Horstmann, G., Schützwohl, A. (2019). The Cognitive-Evolutionary Model of Surprise: A Review of the Evidence. *Topics in Cognitive Science* 11: 50-74.

Serrano-Losada, M. (2017a). On English *turn out* and Spanish *resultar* mirative constructions. A case of ongoing grammaticalization? *Journal of Historical Linguistics* 7: 160-189.

Serrano-Losada, M. (2017b). Raising *turn out* in Late Modern English. *Review of Cognitive Linguistics* 15(2): 411-437.

Skavronskaya, L., Moyle, B., & Scott, N. (2020). The experience of novelty and the novelty of experience. *Frontiers in Psychology* 11.

Slobin D., Aksu A. (1982). Tense, Aspect, and Modality in the use of the Turkish Evidential. In: P.J. Hopper (ed), *Tense-Aspect: Between Semantics and Pragmatics*. Amsterdam: John Benjamins, 185-200.

Zeisler, B. (2000). Narrative conventions in Tibetan Languages: The Issue of Mirativity. *Linguistics of the Tibeto-Burman Area* 23(2): 39-77.

Zeisler, B. (2004). *Relative tense and aspectual values in Tibetan languages*. Berlin: Mouton de Gruyter.

Zeisler B. (2017). "Don't believe in a paradigm that you havent't manipulated yourself!" Evidentiality, speaker attitude, and mirativity in Ladakhi. *Review of Cognitive Linguistics* 15(2): 515-539.

Zhuang, L. (2023) *The Surprise Factor: A Semantic Theory of Mirativity*. Doctoral dissertation, Cornell University.

Mirative constructions on the rise: The development of *end up* and *wind up* in the recent diachrony of English

Mario Serrano-Losada
(Universidad Complutense de Madrid)

Mirative strategies, used to express information which is new, surprising or unexpected, are pervasive in English today. These encompass a range of linguistic features, including interjections such as *gosh*, disjuncts like *in the end*, and grammaticalized expressions such as the [*what a* NP] construction. Among these strategies, evidential (e.g. *seem*, *appear*) and mirative (e.g. *chance*, *prove*) raising verbs occupy a prominent position. While many of these evidential an mirative raising verbs (EMRVs) can be traced back to Middle English, others have emerged more recently, particularly from the Late Modern English period onward. Such an example is mirative *turn out* (e.g. *It turned out to be a false alarm*), which is first recorded in the mid-eighteenth century (Serrano-Losada 2017).

This corpus-based study examines the parallel development of *end up* and *wind up*, two twentieth century innovations that can be classified alongside EMRVs. The examples in (1)-(2) illustrate one of their uses:

(1)   In 1992, Cuomo **ended up** delivering the nominating speech for Bill Clinton at the Democratic convention in New York. (COCA:NEWS:2015)

(2)   Incredibly, the 1968 campaign **wound up** being […] one of the best in MLB history. (COCA:MAG:2019)

In raised subject constructions like (1) and (2), *end up* and *wind up* are used to introduce new, oftentimes surprising, information. Typically, these verbs are interpreted to imply an outcome that was unexpected, unintended, or reached through a series of events that may not have been entirely predictable. As such, their meanings are closely related to the concept of mirativity (Aikhenvald 2012, DeLancey 1997; cf. Dendale et al 2024).

The historical record suggests that such mirative-like readings developed from earlier uses in which these verbs expressed a sense of completion or finality, as in examples (3)-(4):

(3)   I shall be better directed in what manner to **wind up** the Catastrophe of the pretty Novel. (CLMET:1740)
(4)   the affair **ended up** with a grand ball at the Clarendon. (COHA:FIC:1872)

In (3) and (4), neither verb refers to the novelty or unexpected nature of the events. Instead, *wind up* in (3) refers to the act of bringing the novel to a conclusion, while *end up* in (4) indicates that the affair reached its conclusion in a particular way.

Mirative readings for *wind up* first emerge in the mid-nineteenth century and became entrenched by the early twentieth century. Similarly, mirative *end up* originated in the early twentieth century,

soaring in frequency towards the mid-twentieth century (Serrano-Losada 2020). This study adopts a constructional perspective to trace the historical development of both verbs and the constructions in which they take part, focusing on the emergence of mirative-related meanings. Moreover, the paper discusses the ongoing integration of these verbs into the class of English EMRVs. The rise of mirative *end up* and *wind up* constructions is attributed to pragmatic enrichment and paradigmatic analogy. Corpus evidence indicates that both verbs have gradually been analogically drawn into the EMRV category, influenced by semantically and formally related expressions like *turn out*. Data for this paper were drawn from various sources, including the OED, EEBO, CLMET, and COHA, for the historical data, and COCA and NOW, for more recent evidence.

Sources

CLMET = De Smet, Hendrik et al. The Corpus of Late Modern English Texts, version 3.1. https://perswww.kuleuven.be/~u0044428/.

COCA = Davies, Mark. 2008-. Corpus of Contemporary American English. https://www.english-corpora.org/coca/.

COHA = Davies, Mark. 2010-. Corpus of Historical American English. https://www.english-corpora.org/coha/.

EEBO = Davies, Mark. 2017. Early English Books Online Corpus. Available online at https://www.english-corpora.org/eebo/.

NOW = Davies, Mark. 2016-. Corpus of News on the Web: 3+ billion words from 20 countries, updated every day. https://www.english-corpora.org/now/.

OED = Oxford English Dictionary Online. Oxford University Press. http://www.oed.com/.

References

Aikhenvald, Alexandra Y. (2012), The essence of mirativity, *Linguistic Typology* 16(3), 435–485.

DeLancey, Scott (1997), Mirativity: The grammatical marking of unexpected information, *Linguistic Typology* 1(1), 33–52.

Dendale, Patrick, Dámaso Izquierdo Alegría & Ana Stulic (2024), 'Turn out' verbs in European languages: Are they evidentials or something else? Presentation at the 2024 SLE Conference, Helsinki, 21-24 August.

Serrano-Losada, Mario (2017), Raising *turn out* in Late Modern English: The rise of a mirative predicate, *Review of Cognitive Linguistics* 15(2), 411–437.

Serrano-Losada, Mario (2020), Analogy-driven change: the emergence and development of mirative *end up* constructions in American English, *English Language & Linguistics* 24(1), 97–121.

What is it that turns out to be true in French?

Philippe Kreutz

Université Libre de Bruxelles

Keywords: French "avéré", fact, ethos, social reality, truthmaking

This ongoing corpus-based (frTenTen23) research investigates two semi-schematic constructions:

(I) [NP+être+avéré]
(II) [NP+avéré]

From a preliminary analysis of 16.000 instances of construction (I), it turns out that *être+avéré* combines with four types of NPs:

(i)   *Le(s) fait(s)*, representing ±12% of the total number of occurrences
(ii)   Content-oriented NPs such as "l'information", "les affirmations", "le pronostic"…
(iii)   Axiological NPs like "le meurtre", "la fraude", "les abus", "cette faute", "l'innocence" …
(iv)   Non-evaluative NPs like "la présence", "les effets'', "la disparition", "les évolutions"…

The initial research questions are about the ontological category of the referent of each type of (definite) NP:

(Q1) Is there a single category of referent for the four types of (definite) NPs?
(Q2) What could be such an unified ontological category: fact, state-of-affairs, event, proposition…?

Beyond (Q1)-(Q2), we target a semantic issue (Q3):

(Q3) Does the ontological category of the NP have any impact on the meaning of "avéré"?

Our working assumptions regarding (Q1)-(Q2) are:

(A0)   Definite NPs in the above-mentioned structure are referential expressions
(A1)   Definite NPs under (iii) refer to events under an evaluative perspective
(A2)   Definite NPs under (iv) refer to events/processes only under some aspectual perspective
(A3)   NPs *le(s) fait(s)* are (at best) anaphoric expressions for referring expressions of (ii)-(iii)-(iv) types or sentences
(A4)   Facts are not truth-bearer
(A5)   Facts are not basic truthmakers nor even basic referents (Mulligan, 2006)

At first glance, content-oriented NPs (ii) express the most plausible entity liable to turn to be true, namely propositional content. Notice that "le mensonge" falls under both (ii) and (iii) types, but (1) can only mean that a lie occurred rather than its content being true. Social status of the event prevails on anything else.

(1) "Le mensonge est avéré"
     The lie is proven

Construction (II), [NP+avéré], also involves (i)-(ii)-(iv) types of NPs but also a lot of (iii')-type of NP ascribing social roles to humans.

(iii') Status-oriented NPs like "un meurtrier", "un voleur", "ce pilote", "mon ancêtre", "le chercheur"…

Ascription of status can rely on two different processes directly related to two distinct meanings of "avéré". For judicial authorities (and people recognizing their authority), "un voleur avéré" would be an *official/legally proved thief* but it would rather be a *competent* (ideally not yet convicted) *thief* for a recruiting mafia gang. The *competent* construal comes from the *thief-ethos* assigned by the gang to its new recruited member. Ethos is a type of character, a disposition to behave in a certain way (Maingueneau 2013). Depending on whether social reality relies on authoritative source or on shared common sense (within a domain), the meaning of "avéré" balances between (a) and (b) meanings:

(a) official/legal
(b) competent/legitimate

But social reality (anchored on "collective intentionality" and physical reality (Searle, 1995)) remains a form of reality. Indeed, our data show that (a)-(b) meanings are often subsumed by a more neutral (c) meaning:

(c) real/genuine

From its etymological base (latin "verus"), "avéré" remains a semantic notion but it pertains more to publicly recognized truthmaking (by physical or social reality) than to truth itself.

References

Dendale, P. (2019), Pourquoi s'avérer (et ses équivalents dans d'autres langues) ne sont pas des évidentiels. Talk at the CIPLR conference, Copenhagen, 1-6 July 2019.

Maingueneau, D. (2013), L'*éthos* : un articulateur », *COnTEXTES* [Online], 13 | 2013, URL : http://journals.openedition.org/contextes/5772 ; DOI : https://doi.org/10.4000/contextes.5772

Mulligan, K. (2006), Facts, Formal Objects and Ontology, *Modes of Existence. Papers in Ontology and Philosophical Logic*, Andrea Bottani & Richard Davies (eds.), Frankfurt: Ontos Verlag, 31–46.

Rodriguez-Pereyra, G. (2006), Truthmakers, *Philosophy Compass* 1/2, 186–200.

Rychter, P. (2014), Truthmaker Theory without Truthmakers, *Ratio*, Volume 27 (3), 276-290.

Searle, J. R. (1995), *The Construction of Social Reality*. Allen Lane.

'Turn out'-verbs in Finnish

Seppo Kittilä
(University of Helsinki)

Finnish uses four constructions (including five verbs *käydä ilmi*, *paljastua*, *selvitä*, *ilmetä* and *osoittautua*, all of which can be translated with *turn out*) to express 'turn out' -semantics:

1. (*asiakirjo-i-sta/asiakirjo-j-en*   *perusteella*)   *käv-i*   *ilmi, että*
    (document-PL-ELAT/document-PL-GEN   based)   turn-PST   out   that
    Kalle on   *petturi*
    PN   COP.PRS.3SG   traitor.NOM
    '(Based on the documents) it turned out that Kalle is a traitor'
2. *Kalle-n*   *petturuus*   *selvis-i*   (*asiakirjo-i-sta/*
    PN-GEN   treachery   become clear-3SG.PST   (document-PL-ELAT/
    asiakirjo-j-en   *perusteella*)
    Document-PL-GEN   based)
    'Kalle's treachery became clear (based on the documents)'
3. *Kalle paljastu-i*   *petturi-ksi*
    PN   reveal-3SG.PST   traitor-TRANSL
    'It turned out that Kalle is a traitor'
4. *Asiakirja-t*   *paljast-i-vat*   *Kalle-n*   *petturuude-n/*   *että*
    Document-PL reveal.TR-PST-3PL   Kalle-GEN   treachery-ACC/   that
    Kalle on   *petturi*
    PN   COP.PRS.3SG   traitor.NOM
    'The documents revealed Kalle's treachery/that Kalle is a traitor'

Generally, the constructions involve an intransitive verb, but transitive verbs are also possible, as shown in (4). The state-of-affairs referred to can be expressed by a complement clause, or it can be expressed as the subject (as in (2)), adjunct (3) or an object (as in (4)). The 'turn out'-constructions can also include an explicit reference to the information source, which can be expressed by an adjunct (as in (2)) or as a subject (in (4)).

The constructions above differ according to which verbs can be used in them. Construction in (1) is the most neutral one in this respect and can be used with all 'turn out'-verbs. In (2), *osoittautua* is not possible. (3) and (4) are much more restricted, and in (3), only *paljastua* and *osoittautua* are possible, while only *paljastua*, *osoittautua* and *selvitä* have a transitive counterpart. The constructions illustrated above also display semantic differences. The turn out-semantics is most obvious in (1), while in the transitive construction the 'turn out' -semantics (e.g., mirativity or high information value) is least clear (example (4) does not necessarily express any high information value, while this is the case in (1)).

Functionally, 'turn out'-constructions of Finnish cannot be viewed as evidentials, because they do not specify the speaker's information source, but they do resemble modals in that they evaluate information; the speaker finds it somehow relevant and newsworthy (see Cornillie 2007 and Lampert 2020). Pragmatically, their use implies that the expressed information has some a high information value. The expressed information is not necessarily mirative per se (cf. Serrano-Losada 2017), since the constructions can be used also in cases, where the speaker does not have any clear expectations as regards the occurrence of the given SoA. However, the constructions cannot be used with something very trivial, as in 'it turned

out that cars have wheels'. Finally, the constructions have a high truth value, and their use implies that the speaker believes what they refer to.

In this paper, the five 'turn out'-verbs and the four constructions of Finnish will be discussed both from a formal and a functional perspective. It will be shown that the discussed constructions and verbs display clear differences as regards their use, e.g. according to how mirative the denoted information is.

References

Cornillie, B. (2007), *Evidentiality and Epistemic Modality in Spanish (semi-)auxiliaries*, Berlin: Mouton De Gruyter.

Lampert, G. (2020), Seem: Evidential, Epistemic, or What Else? A Study in Cognitive Semantics. *International Journal of Cognitive Linguistics* 2(1): 1-24

Serrano-Losada, M. (2017), On English turn out and Spanish resultar mirative constructions. A case of ongoing grammaticalization?, *Journal of Historical Linguistics*(7), 160-189.

'Find' as an auxiliary in discovery constructions: data from East Caucasian

Sofya Glavatskikh, Timur Maisak & Michael Daniel

(National Research University Higher School of Economics, National Research University Higher School of Economics & Eberhard Karls Universität Tübingen)

Keywords: Tsez, presumptive inferentiality, conditional constructions, grammaticalization

In many East Caucasian (Nakh-Daghestanian) languages, verbs meaning 'find, be found' can function as auxiliaries in evidential/modal constructions describing discovery ('as someone found out', 'as it turned out') and presumptive inference ('presumably'). Daniel & Maisak (2018), based on the data from Bagvalal (Andic), Archi (Lezgic) and Mehweb (Dargwa), argue for the areal localization of the phenomenon in central/northern Daghestan. More recent studies of the auxiliary 'find' include Lyutikova (2023), based on her fieldwork on Khwarshi, and Glavatskikh (2024), who presents an extensive corpus study of 'find' in Tsez (both Tsezic). We aim at generalizing observations on auxiliary uses of 'find' based on corpus data from the genealogically diversified sample of four languages, Archi, Mehweb, Tsez and Tukita, all belonging to different branches of the family.

Example (1) illustrates auxiliary use of 'find' in the *situation discovery* construction. It describes a situation that is being observed by the subject (the main participant of the narrative episode). The situation discovered can be an ongoing process or a resultant state, which is reflected in the aspectual value of the lexical verb form which can be a converb or a resultative participle, respectively. These uses may have a value of unexpectedness, and are thus semantically comparable to 'turn out' constructions. The meaning of the discovery construction is different from direct evidentiality or miratives in that the acquisition of information is not restricted to the speaker: in non-personal narrative texts like legends or fairy-tales, it is common to describe a discovery made by the third person protagonist, as in (1).

(1)   Tsez (< Tsezic)

     aždaħ-ä   *kid*   *y-exur-äsi*   *y-esu-n*

     dragon-ERG   daughter   II-kill-RES.PRT   II-find-PST.UNW

     '(He) found that a dragon had killed (his) daughter.' [Abdulaev et al. 2022]

Another auxiliary use is found in *conditional protases*. The semantic impact of 'find' is not always clear (~ 'if it turns out that P' > 'if P'), but using 'find' as an auxiliary makes it possible to express aspectual characteristics of the situation by choosing the appropriate form of the lexical verb, which is impossible by using the conditional converb alone. Thus, in (2) the combination of the resultative participle with the auxiliary 'find' allows to express the perfect/resultative conditional meaning ('if' + 'have killed').

(2) Tsez (< Tsezic)

neła-s		*uži*	*el-a*	*exur-äsi*	*esu-näy*
DEM1.IISG.OBL-GEN1		son	we(I)OBL-ERG	kill-RES.PRT	find-CND.CVB

'If we have killed her son, [...]' [Abdulaev et al. 2022]

A distinctive use of the auxiliary 'find' is found in constructions with *presumptive inferential* semantics: 'based on natural assumptions, it is probable that P'. As a rule, the auxiliary occurs in future/general habitual; in some languages, this form is also used as a parenthetical 'probably'.

(3) Tsez (< Tsezic)

yeda	*ƛirba*	*mekoy-xo*	*äsu-ƛin*
DEM2.ISG	guest	be.hungry-IPFV.CVB	I.find.FUT.INDEF-QUOT

'The guest must be hungry (they thought).' [Abdulaev et al. 2022]

Based exclusively on naturalistic data, we attempt to draft the evolutionary path of the construction, suggesting that *situation discovery* use - which we suggest to be a type of 'turn out' construction - predates the grammaticalization of the two subsequent functions.

Glosses other than LGR:

I, II — genders; CND — conditional; DEM1, DEM2 — demonstrative series; INDEF — indefinite; IPFV — imperfective; OBL — oblique stem; QUOT — quotative; RES — resultative; UNW — unwitnessed.

References

Abdulaev et al. 2022 — A. K. Abdulaev, I. K. Abdullaev, André Müller, Evgeniya Zhivotova, & Bernard Comrie. The Tsez Annotated Corpus Project (v1.0) [Data set]. Zenodo. https://tsezacp.clld.org. 2022.

Daniel, M. A. & Maisak, T. A. 2018. Černaja koška grammatikalizacii: konstrukcii s glagolom 'najti' v dagestanskix jazykax. In D. A. Ryžova, N. R. Dobrušina, A. A. Bonč-Osmolovskaja, A. S. Vyrenkova, M. V. Kjuseva, B. V. Orexov, T. I. Reznikova (eds.), *EVRika! Sbornik statej o poiskax i naxodkax k jubileju E. V. Raxilinoj.* Moscow: Labirint. P. 120—149.

Glavatskikh, S. S. 2024. Upotreblenie konstrukcij s glagolom *-esu-* 'najti' v cezskom. In *Dvadcat' pervaja konferencija po tipologii i grammatike dlja molodyx issledovatelej. Tezisy dokladov*. St. Petersburg: ILI RAN. P. 43—46.

Lyutikova, E. A. 2023. Ot polipredikacii k èvidencial'nym uslovnym konstrukcijam: xvaršinskie naxodki. In N. A. Korotaev & N. R. Sumbatova (eds.), *Sostav nauki: Sbornik statej k jubileju Very Isaakovny Podlesskoj.* Moscow: Buki-Vedi. P. 108—129.

Turn-out verbs in Dutch and German: Similarities and differences

Tanja Mortelmans

(University of Antwerp)

The Dutch 'turn out' verb *blijken* has been studied by Vliegen (2010, 2011), Bomans (2019), Sanders & Spooren (1996) and Mortelmans (2022); Vliegen (2015) addresses the German equivalents of Dutch *blijken*: *sich heraustellen, sich erweisen, sich ergeben* and *hervorgehen*. Whereas *blijken* has been claimed to also have mirative meaning (Bomans 2019, Mortelmans 2022), no such claims have been made for the German verbs. What connects the German verbs with *blijken*, however, is the fact that they do not only function as each other's translation equivalents (as shown in Vliegen 2015), but also as translation equivalents of *turn out*, which is especially the case for 'sich herausstellen' and 'sich erweisen'.

(1) ENG It **turned out** he'd done it to lots of women. (Nicci French, Sunday Morning Coming Down, p. 95)
    GER Es **stellte sich heraus**, dass er das Gleiche schon vielen Frauen angetan hatte.
    DUT Het **bleek** dat hij het bij heel veel vrouwen had gedaan

Another unifying factor is that both the Dutch and the German verbs participate in a wide variety of constructions. German *sich herausstellen*, for instance, can combine with a NP or AdjP introduced by *als* (3), with a *dass*-clause (4) or occur in adverbial or parenthetical constructions (5); all examples are taken from Vliegen 2015.

(3) **Als** Fehlalarm hat sich […] ein Chlorgas-Alarm im […] Hallenbad **herausgestellt**. (FR99/SEP)
(4) Bei einer […] Untersuchung […] **habe sich herausgestellt**, **dass** der meiste Autoverkehr von den Anwohnern selbst stamme. (FR99/SEP)
(5) Die Eltern des Jungen, **stellt sich heraus**, sind […] massakriert worden. (FR99/FEB)

Interestingly, each German verb has a preference for a particular construction type: whereas *sich erweisen* combines with *als* + NP/AdjP in more than 90% of all its usages, *sich herausstellen* prefers the combination with a *dass*-clause (Vliegen 2015). An even higher constructional variability can be attested for *blijken* (Mortelmans 2022): it occurs as a main verb, as a copula, in an impersonal construction with *dat* (*het blijkt dat*), as a semi-auxiliary with a *te*-infinitive and in adverbial and parenthetical constructions. At the same time, *blijken* favours three of these construction types: the main verb use, the combination with a *dat*-clause or with a *te*-infinitive (Mortelmans 2022), whereby especially the latter invites a mirative reading.

In my contribution, I will address the following questions on the basis of a corpus-based analysis. First, can mirative uses also be attested for (some of the) German verbs? Second, Mortelmans (2022) has claimed for *blijken* that mirative uses are prevalent when *blijken* is not focalized and does not combine with an explicit evidential source. To what extent does this also apply to possible mirative uses of the German equivalents? And third, are there similarites regarding the actionality of the Dutch and German verbs? Whereas *blijken* has been argued to be an intransitive change of state verb (see Honselaar 1987), the German counterparts are mostly reflexive verbs with change of state semantics as well (Steinbach 2002). Hence, change of state semantics seems to be another factor uniting *blijken* with its German counterparts.

References

Bomans, Cas (2019), On the development of *blijken* and *turn out*: how did these verbs become linguistic markers of surprise? Poster presented at *A Germanic Sandwich* 2019. Amsterdam, 23-24 May 2019. https://www.meertens.knaw.nl/germanicsandwich2019/wp-content/uploads/2019/05/ Bomans_GS2019.pdf.

Honselaar, Wim (1987), Zijn vs. Hebben in het Samengesteld Perfectum. [Zijn 'be' vs. hebben 'have' in the composite perfect.], *De Nieuwe Taalgids* 80(1), 55–68.

Mortelmans, Tanja (2022), *Blijken*: een evidentieel-miratieve outsider: Een corpusanalyse op basis van het CGN, *Nederlandse Taalkunde* 27(3), 293-327.

Sanders, José and Wilbert Spooren (1996), Subjectivity and certainty in epistemic modality: A study of Dutch epistemic modifiers*, Cognitive Linguistics* 7(3), 241-264.

Steinbach, Markus (2002), *Middle voice. A comparative study in the syntax-semantics interface in German*, Amsterdam: John Benjamins.

Vliegen, Maurice (2010), Lexikalische Evidentialität: das niederländische *blijken*. Ein evidenter Fall der Grammatikalisierung, in A. ten Cate *et al.* (eds.), (2010), *Grammatik Praxis Geschichte*, Tübingen: Narr, 209-216.

Vliegen, Maurice (2011), Evidentiality. Dutch seem and appear verbs *blijken, lijken, schijnen*, *Linguistics in the Netherlands* 28, 125-137.

Vliegen, Maurice (2015), Das niederländische 'blijken' samt seinen lexikalischen Pendants im Deutschen. Ein kontrastives Problem, in J. Tóth (ed.) (2015), *Die Sprache und ihre Wissenschaft zwischen Tradition und Innovation*. Frankfurt/Main: Peter Lang, 97-107.

Spanish 'turn out' verb *resultar* in its predicative and atributive constructions

The Spanish *<resultar* + nominal subordinate clause>-construction (e.g. *Resulta que el documento era falso*), as it expresses a novel circumstance as something unexpected, represents a paradigmatic case of 'turn out' expressions and has been recognized as such by several authors (Serrano-Losada 2017; Dendale 2019). Bearing in mind the objectives of the SLE 2025 Workshop: *'Turn out' verbs and constructions crosslinguistically,* we propose to carry out a syntactic characterization of the 'turn out' uses of *resultar*, which -we will argue- would not be limited to its combinations with a subordinate clause. Our ultimate goal, thus, is to contribute to the determination of the characteristics of 'turn out' verbs and to the discussion whether these verbs should be treated as a grammatically relevant class or not. In order to achieve these general objectives, and on the basis of Morimoto/Pavón 2007 and Morimoto 2008, in our presentation we will review the different uses of the verb *resultar* both as a pseudocopulative verb and as a predicative verb from syntactic, semantic and aspectual perspectives, and we will point out that the pseudocopulative *resultar* also functions as a 'turn out' verb in certain contexts.

As a predicative verb, *resultar* can take as its subject a nominal subordinate clause headed by the conjunction *que*, (1) or a noun phrase expressing the consequence or result of something, (2). It is also possible for the subject to be an infinitive clause, in which case the subordinate verb is usually a copulative verb: *ser*, (3) or *estar*, (4).

(1) se produce un doble juego de contrastes simultáneos, de lo cual resulta que radiaciones iguales producen sensaciones claramente distintas (Antonio Paredes-Candia, *El banquete, su historia y tradición en Bolivia. CORPES*)
(2) De aquella reunión resultó un exitoso proyecto
(3) Así que esto resultó ser una reunión de trabajo (R. Chávez, C. Santajuliana, *El final de las nubes. CORPES*)
(4) Pero si resulta estar equivocada, los franceses me llamarán suizo (M. Seara Valero, *Hasta el infinito y más allá: Las últimas noticias sobre el cosmos. CORPES*)

Pseudocopulative *resultar* has an aspectual value, in which the attribute represents the result of an event, (5), and a non-aspectual value, in which, as noted by Morimoto and Pavón (2007) and Morimoto (2008), the attribute is presented as the result of a process of evaluation by the subject, (6):

(5) Dos pasajeros resultaron heridos en el accidente
(6) Ese color resulta demasiado llamativo

As we have indicated, the 'turn out' use of *resultar* is found not only in its predicative value, with a nominal subordinate clause as subject, (7), but also in its use as a pseudocopulative, (8). The latter is a case of non-aspectual *resultar*:

(7) Resulta que ahora soy yo quien tiene la culpa
(8) El documento resultó falso

Our talk will conclude with a close examination of the relationship between the pseudocopulative 'turn out' *resultar* and *<resultar* + nominal subordinate clause> construction, which is already well recognized as a 'turn out' expression.

Works cited in the abstract:

Dendale, P. (2019). 'Non-evidentials' : the case of verbs like s'avérer, turn out, blijken, resultar... Talk at *the 16th International Pragmatics Conference*, Hong Kong, 9-14 June, 2019. https://researchmgt.monash.edu/ws/portalfiles/portal/373885398/IPRA2019_Abstracts_Book.pdf

Morimoto, Y. (2008). *Resultar*: del significado aspectual al significado evaluativo. *ACTA FACULTATIS PHILOSOPHICAE UNIVERSITATIS PRESOVIENSIS*, p. 160-169.

Morimoto, Y. / Pavón, M. V. (2007). *Los verbos pseudocopulativos del español,* Madrid: Arco/Libros.

REAL ACADEMIA ESPAÑOLA: Banco de datos (CORPES XXI) [en línea]. *Corpus del Español del Siglo XXI (CORPES).* <http://www.rae.es>

Serrano-Losado, M. (2017), On English turn out and Spanish resultar mirative constructions. A case of ongoing grammaticalization? *Journal of Historical Linguistics,* 7, p. 160-189.

WS24 Unifying the comparative analysis of tonal systems

Kirill Maslinsky, Valentin Vydrin & Dmitry Gerasimov

Typologizing the Limburgish tone

Carlos Gussenhoven
Radboud University, Netherlands

Will the lexical tone of Limburgish (limb1263; ISO 639; Fig. 1 [1]) reveal the same dialect zones as were established on the basis of segmental isoglosses? Seven typological features were defined on the basis of accounts of nine dialects in the literature. The data not only confirm the segmentally defined zones, but also suggest a developmental history of the tone contrast from its tonogenesis in Cologne some 750 years ago.

In all nine dialects, the tone contrast appears to be embedded in the intonation system, a unique feature of Limburgish because, unlike the tones of Mandarin or Yoruba, their tone catogories cannot be given invariant phonetic descriptions, as these depend on the melodic and pitch-accentual conditions of the sentence. A more usual feature they share is that the lexical tone contrast is binary and privative, such that approximately half the vocabulary has a lexical tone ('accent 2'), while the other half lacks it ('accent 1'). Third, in all nine dialects the contrast is located in the final or penultimate stressed syllable of (simplex) words. In (1), some monosyllabic minimal pairs are given (on the Figure page).

The distinguishing typological features are (i) the number of nuclear intonation melodies, (ii) the value of the lexical tone, (iii) the nature of the Tone Bearing Unit (mora or syllable), (iv) the sequencing of the intonation and lexical tones within a syllable, (v) the contrast on monomoraic sonorant syllables, (vi) the contrast on syllables without intonational tone, and (vii) the number of phonological interactions between lexical and intonational tones. The dialects divide into two groups of four, Hasselt, Neerpelt, Borgloon and Tongeren, which largely fall in the West Limburgish area, and Helden, Roermond, Sittard and Venlo, located in the East Limburgish area. One of the two Central Limburgish dialects, Maastricht, presents a mixed picture, reflecting its intermediate location, while the other Central Limburgish dialect, Tongeren, is typologically West Limburgish and lies close to the border between those two zones. No pair of dialects has identical scores for the seven variables.

These data suggest a number of historical changes. The tonogenesis has been claimed to have arisen in monosyllabic nouns in sentence-final declarative focused positions [2]. Sharply falling pitch for accent 1 came to contrast with lengthened and weakly dipping pitch, which in the current Cologne dialect is closed off by a brief fall [3]. That origin suggests that subsequent developments must have occurred in interrogative and in non-focused contexts. As for the first of these, [4] reported a brief pitch fall after the sentence-final interrogative rise for accent 1 in a Central Franconian dialect, termed an 'epitone'. It can be interpreted as a transfer of the declarative accent-1 cue to the end of an interrogative rise and must be the source of the rising-falling question intonation in East Limburgish dialects (Fig. 2). Second, in the unfocused contexts, words with accent 1 should have no distiguishing features, while this with accent 2 should display the lexical tone by itself, with no intonational tones in the same syllable, except in intination phrase-final position, where there will be a final boundary tone. This is true for the southern dialects, while the more northern ones lost the contrast there.

The main split in the historical development of Limburgish may be due to different developments in the pitch shape in sentence-final focused declarative accent 2. Arguably, in West Limburgish, the peak before the dip was lowered and the second shifted rightward, while in East Limburgish the dip was deepened, leading to two full-fledged peaks. Both groups truncated these forms, in West Limburgish leading a late peak falling to mid pitch and in East Limburgish leading to a fall-rise, with considerable variation between dialects in the pitch range of the final rise (Fig. 3).
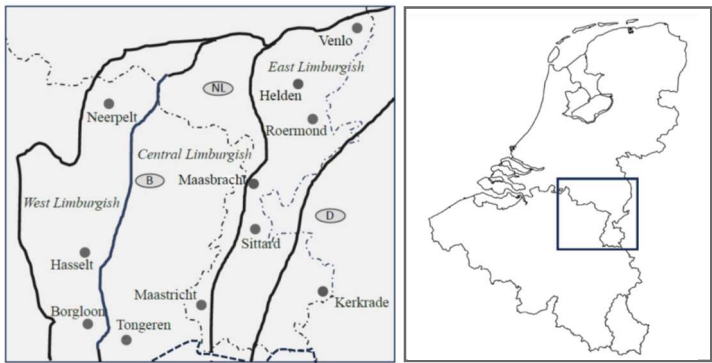
Figure 1. The West, Central and East Limburgish dialect areas (──), with national borders (─ . ─ .), the Germanic-Romance language boundary (▬▬▬) and a locator map for the Benelux.

(1) Hasselt ˈeːk 'vinegar' ˈeːkᴸ 'oak'

Maastricht ˈbɑl 'dance party' ˈbɑlᴴ 'ball'
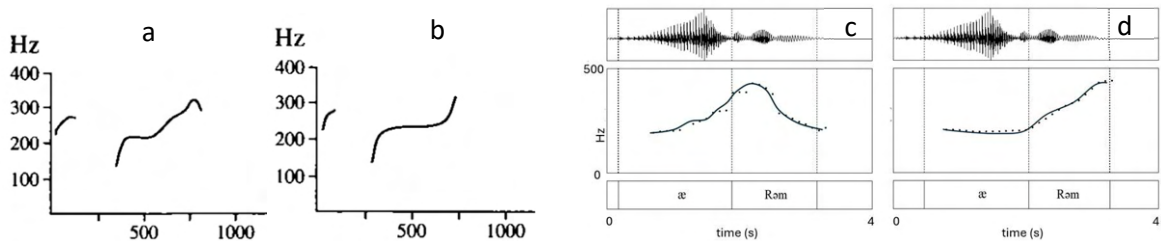
Sittard ˈkæːl 'throat' ˈkæːlᴴ 'bloke'

Figure 2. The 'epitone' after the interrogative rise on *Mand* [man^1] 'Basket?' (panel a) in the Central Franconian dialect of Mayen and its absence on *Mann* [man^2] 'Man?' (panel b) [3: 200], with their equivalents on e*rm?* [æʁm^1] 'Arms?' (panel c) and *erm* [æʁm^2] 'Arm?' (panel d) in Sittard Limburgish.

Figure 3. Reconstructed stages II and III between the original monosyllabic sentence-final focused declarative shape of accent 2 (today's attested form in the Cologne dialect, Stage I) and the present-day shapes in West and East Limburgish (Stage IV).

[1] Goossens, J. (1977). *Inleiding tot de Nederlandse Dialectologie*. Groningen: Wolters-Noordhoff.
[2] Gussenhoven, C. (2000). On the origin and development of the Central Franconian tone contrast. In A. Lahiri (Ed.), *Analogy, Levelling, Markedness.* Berlin: Mouton de Gruyter. 215-260.
[3] Werth, A. (2012). Perceptual evidence for contrast enhancement in tone-intonation interaction. In O. Niebuhr (Ed*.), Understanding prosody*, 187–200. Berlin: Mouton de Gruyter.
[4] Schmidt, J-E. (1986). *Die mittelfränkischen Tonakzente: rheinische Akzentuierung)*. Stuttgart: Steiner.

Dissimilatory tone can be tonemic: Evidence from Mundurukú

Dmitry Gerasimov

(INALCO, Paris)

In Mundurukú (Tupian; Brazil), many morphemes are subject to tonal changes that largely appear dissimilatory in nature, with a H tone surfacing after L, and vice versa. To account for this variation, Picanço (2005) proposes an OT analysis with an elaborate system of constraints. While the tonal system of Mundurukú indeed naturally lends itself to an OT analysis, it can also easily be accounted for under a more traditional rules-based approach. In what follows, I reframe Picanço's findings in a set of ordered rules and show that the Mundurukú data are of interest for the typology of tonal rules, of the kind envisioned in (Hyman & Schuh 1974).

In line with Picanço's original analysis, there are two phonological tones (tonemes): L and H. The tone-bearing unit is the syllable. Underlyingly, any syllable is either L, H or Ø (toneless). In addition, some toneless morphemes have a floating H toneme at their right edge. The main tonal rules (in order of application) are as follows:

1. Tone Dissimilation (OCP): Any syllable following a L tonal span receives a H toneme.
2. Floating H Association: A floating H tone docks to an adjacent syllable, provided that this syllable is (i) word-final; (ii) toneless: (iii) not preceded by a H tonal span; if these conditions are not met, the floating H tone is deleted.
3. Default L Insertion: Any remaining toneless syllables receive a default L tone.

Consider the contrast in (1a,b):

(1)   a.   *o̧-t-a^H-da^H*   →   [ò.**tà.dá**]

            1SG-3OBJ-CL-cook

            'I cooked it.'

    b.   *ò̧-t-a^H-da^H*   →   [ò̧.**tá.dà**]

            3-3OBJ-CL-cook

            'S/he cooked it.'   (Picanço 2005: 334)

Derivations of the two forms above illustrate the successive application of tonal rules:

(2)   a.   *o̧-t-a-da*   *ò̧-t-a-da*   underlying

                          |

             Ⓗ Ⓗ   L Ⓗ Ⓗ

    b.   *o̧.ta.da*   *ò̧.tá.da*   syllabification/Tone Dissimilation

                          | |

             Ⓗ Ⓗ   L H Ⓗ Ⓗ

    c.   *o̧.ta.dá*   *ò̧.tá.da*   Floating H Association

                | | |

              H   L H

|

c.　　　*ọ̀.tà.dà*　　　　*ọ̀.tá.dà*　　　　Default L Insertion
　　　　 |　 |　 |　　　　　 |　 |　 |
　　　　 L　L　H　　　　　 L　H　L

Most interestingly, as can be seen from the second derivation in (2), a H tone arising from Tone Dissimilation blocks realization of a floating H toneme on the following syllable (per condition (iii) of rule (2)). This shows that the dissimilatory H tone in Mundurukú is tonemic (phonologically active), despite being both predictable and absent from the underlying representation. This sets up a broader question about the phonological status of dissimilatory tone cross-linguistically.

Acknowledgements

References

Hyman, Larry M. & Russel G. Schuh. 1974. Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry* V/1: 81–115.

Picanço, Gessiane Lobato. 2005. *Mundurukú: phonetics, phonology, synchrony, diachrony*. PhD Thesis. University of British Columbia.

Elena Perekhvalskaya
INALCO, ERC Project "Theory of tone"
LHASA TIBETAN: PROSODIC WORD AND TONAL SPAN.

The paper is devoted to the analysis of the tonal system of Lhasa Tibetan using the terminology and approaches developed within the framework of the Thot project.

The following concepts and terms are used: Toneme, Tone span, Prosodic word (for definition see https://thot.huma-num.fr/terms-and-notions/).

PROSODIC WORD. The number of syllables in a prosodic word is theoretically unlimited .Syllables differ by prominence. The first syllable is always prominent. It is characterized by the following features: 1) in the onset position all consonant phonemes are possible; 2) it bears the toneme, this toneme can spread to the next syllable.

In multisyllabic prosodic words all syllables, except for the first one, are non-prominent. Non-prominent syllables have the following characteristics: 1) some consonants are not licensed in the onset position, in particular, aspirated consonant are disallowed (1); 2) the lexical tone of a syllable is neutralized (2); 3) In the word-final position, occlusive consonants are substituted by the glottal stop (3), sonorants, with the exception of /m/, tend to disappear, resulting in a compensatory vowel length.

(1) ཆུ་ *chu* [cʰú] 'water' + ཚོད་ *tshod* [tsʰɵ̂ʔ] 'measure' > ཆུ་ཚོད་ *chu tshod* [cʰú.tsɵ̂ʔ] 'time, hour';

(2) ཧ་ *ha* [há] 'breath' + ཡང་ *yang* [jàŋ] 'light' > ཧ་ཡང་ *ha yang* [há.jáŋ] 'aluminum'.

(3) ཁ་ལག་ *kha-lag* [kʰá.lâʔ] 'food' (mouth + hand) [H kʰá.laʔ]

A prosodic word necessarily begins with a prominant syllable which carries a toneme; the tonal span may spread on the second syllable. The remaining syllables are extra-tonal and are realized with a default low tone.

TONEME. There are two tonemes, High (Falling) and Low (Raising). Each toneme has two main surface representations : one for Light syllables, the other for Heavy syllables and bi-syllabic sequences. Each toneme has two main allotonemes: a level one and a contour one.

TONAL SPAN. The size of a tonal span may be one or two syllables.

Change of tonal span boundaries. All auxiliary morphemes (case markers, derivational nominal and adjectival suffixes, postpositions, verbal suffixes, copulas) are underlyingly toneless.

With nouns, adjectives or adverbs, auxiliary morphemes integrate into the tonal span if they appear as the second syllable of the prosodic word (4). But the same morphemes remain extra-tonal if they appear as the third, forth, etc. syllable of a prosodic word (5).

(4) བོད་ལ་ *bod-la* [pʰɵ̂ʔ.lá] [L pʰɵ̂ʔ.la] 'in Tibet';

(5) མཚམ་མཚམ་ལ་ *mtsams-mtsams-la* [tsám.tsám.là] [H tsám.tsam].la 'sometimes';

The Lhasa Tibetan tonal system can be characterized as a mixed level-contour tone system.

References

DeLancey, S. (2003). Lhasa Tibetan. In G. Thurgood & R. J. Lapolla (Eds.), *The Sino-Tibetan Languages* (pp. 270–286). Taylor & Francis.

Duanmu, S. (1992). An autosegmental analysis of tone in four Tibetan languages. Linguistics of the Tibeto-Burman Area, 15(1).

Gong, Xun. Word-level prosodic structure and stress in Old Tibetan. Special panel "MediaevalTibeto-Burman Languages". Annual Meeting of the Southeast Asian Linguistics Society, SOAS. 2014.

Jäschke, Heinrich August. Tibetan grammar. 2018 (1888) London: Dalton house.

Lim, Keh Shen. The Tonal and Intonational Phonology of Lhasa Tibetan. Dissertation. University of Ottawa, Ottawa. 2018.

Michailovsky, Boyd & Martine Mazaudon. Lost syllables and tone contour in Dzongkha (Bhutan). Prosodic analysis and Asian linguistics: to honour R. K. Sprigg, Pacific Linguistics, pp.115-136, 1988, Pacific Linguistics Series. 104.

Roerich, Jurij (Ю. Н. Рерих). Буддизм и культурное единство Азии. Сборник статей. М.: Международный Центр Рерихов. 2002.

Sprigg, R. K. The tonal system of Tibetan (Lhasa dialect) and the nominal phrase // Bulletin of the School of Oriental and African Studies, University of London , 1955, Vol. 17, No. 1 (1955), pp. 133-153

Tournadre, Nicolas & Sangda Dorje. Manual of Standard Tibetan. Language and Civilization. Ithaca, NY: Snow Lion Publ. 2003.

Zhang, Yubin. Central Tibetan (Lhasa). Illustration of the IPA. Journal of the International Phonetic Association (2024), 1–23.

Measuring the Functional Load of Tonal Melodies

Hugh Paterson III
Institute for Advanced Study

Keywords: Tone, Functional Load, Orthography, Phonology, Cognitive Load

The computation of functional load has been debated. We present a method which addresses the computation of the functional load of tonal oppositions by viewing tone as non-concatenative (non-linear) oppositions–tonal melodies (e.g., Snider 1990a, 1992). Current approaches (e.g., Oh 2013; Hall et al. 2022) classify combinatory attenuations as series of singleton attenuations, failing to accurately reflect oppositions. An accurate assessment of functional load must do two things:

1. Account for oppositions between sets of pitch heights across the domain to which the set attaches—account for non-linear distinctions.
2. Account for variations between underlying tonal melodies and their surface forms when changes represent a meaningful distinction—include morphological options in computational algorithms.

Promoted by Prague school linguists, the functional load hypothesis suggests that phoneme stability and consequential sound system stability (or inversely, sound system shifts) are the result of the relative importance (frequency) of phonological oppositions (Catford, 1988). Quantifying functional load, especially for tone, has been more challenging with suggestions from linear segmental (structuralist) approaches proposed by Greenberg (1959), Hackett (1966, 1967), Wang (1967b), Surendran (2003), and Surendran and Niyogi (2006). In addition to proposed impacts on language evolution, functional load remains a relevant concept among scholars involved with orthography development. They often justify orthography decisions for representing tone on the basis of perceived functional load, generally analyzing contrasts from a linear approach (Koffi 2014; Priestly 1992; Roberts 2009). This stands in contrast to the well-argued phonological importance of tonal melodies and their role in creating meaningful oppositions (Snider 1990a, 1990b, 1992).

Defining the contrastive units involving pitch in communication has been elusive. The context of conducting tonal analysis is often construed as the necessity to distinguish between underlying-forms and surface-forms (Snider 2014) or which set of phonological features support a given pitch within a larger set of phonological units and pitch height contrasts (among others: Anderson 1978; Fromkin 1972; Wang 1967a, Snider 1988, McPherson 2016). Much like text input processes in computing, the analytical methods for calculating functional load have assumed linearity in segments, computing pitch either in conjunction with specific segmental carriers or independent as its own segment (in addition to the carrier). In contrast, tone (Snider 1999, 2014; Yip 1993, 2002) as well as other segments in morphological constructs, e.g., segmental non-concatenative morphology (Davis & Tsujimura 2018), has been convincingly argued to operate in non-linear ways as both part of the base lexical unit and as part of a templatic morphological paradigm applied during post-lexical processes, where morpho-phonological processes still allow pitch heights to change. This further highlights the need for clarity when calculating functional load–is the calculation being conducted upon lexemes prior-to or after a morphological process has transpired? A non-linear view suggests that in both cases the contrast developed is not with specific other pitch segments (or their carriers) but is better analysed as a set of segments acting together

indicating meaning, contrasting with other sets of indicators within a given domain (e.g., verbs, nouns, etc.). A more accurate approach to computing functional load must include the comparison of all phonological units.

References

Anderson, Stephen R. 1978. "Tone Features." In *Tone*, edited by Victoria A. Fromkin, 133–75. Academic Press. https://doi.org/10.1016/B978-0-12-267350-4.50010-8.

Catford, John C. 1988. "Functional Load and Diachronic Phonology." In *The Prague School and Its Legacy*, edited by Yishai Tobin, 3–20. Linguistic and Literary Studies in Eastern Europe 27. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/llsee.27.04cat.

Corriente, Federico C. 1971. "On the Functional Yield of Some Synthetic Devices in Arabic and Semitic Morphology." *The Jewish Quarterly Review* 62 (1): 20–50. https://doi.org/10.2307/1453861.

Davis, Stuart, and Natsuko Tsujimura. 2018. "Arabic Nonconcatenative Morphology in Construction Morphology." In *The Construction of Words*, edited by Geert Booij, 315–39. Studies in Morphology 4. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-74394-3_12.

Fromkin, Victoria A. 1972. "Tone Features and Tone Rules." *Studies in African Linguistics* 3 (1): 47–76. https://doi.org/10.32473/sal.v3i1.107636.

Greenberg, Joseph. 1959. "A Method of Measuring Functional Yield as Applied to Tone in African Languages." In *Georgetown University Round Table on Languages and Linguistics*, 7–16. Georgetown University Monograph Series on Languages and Linguistics 12. Washington, D.C.: Georgetown University Press.

Hall, Kathleen Currie, Blake Allen, Edith Coates, Michael Fry, Serena Huang, Khia Johnson, Roger Lo, Scott Mackie, Michael McAuliffe, Stanley Nam. 2022. *Phonological CorpusTools*. Software.

Hockett, C F. 1966. "The Quantification of Functional Load - a Linguistic Problem." Santa Monica, California: The Rand Corporation. https://eric.ed.gov/?id=ED011649.

Hockett, Charles F. 1967. "The Quantification of Functional Load." *WORD* 23 (1–3): 300–320. https://doi.org/10.1080/00437956.1967.11435484.

King, Robert D. 1967. "A Measure for Functional Load." *Studia Linguistica* 21 (1): 1–14. https://doi.org/10.1111/j.1467-9582.1967.tb00545.x.

Koffi, Ettien N. 2014. "Towards an Optimal Representation of Tones in the Orthographies of African Languages." *Linguistic Portfolios* 3 (1): 163–90. http://repository.stcloudstate.edu/stcloud_ling/vol3/iss1/12.

McPherson, Laura. 2017. "Tone Features Revisited Evidence from Seenku." In *Diversity in African Languages*, edited by Doris L Payne, Sara Pacchiarotti, and Mokaya Bosire, 5–22. Contemporary African Linguistics 1. Berlin: Language Science Press. https://doi.org/10.17169/langsci.b121.282.

Oh, Yoon Mi, François Pellegrino, Christophe Coupé, and Egidio Marsico. 2013. "Cross-Language Comparison of Functional Load for Vowels, Consonants, and Tones." In Interspeech 2013, edited by Frédéric Bimbot. https://www.isca-archive.org/interspeech_2013/oh13c_interspeech.html

Priestly, Tom M.S. 1992. "Problems in the Creation of an Orthography: Functional Load, Interference, and Political Preferences." *The Slavic and East European Journal* 36 (3): 302–316. https://doi.org/10.2307/308584.

Roberts, David. 2009. "Visual Crowding and the Tone Orthography of African Languages." *Written Language & Literacy* 12:140–55. https://doi.org/10.1075/wll.12.1.07rob.

Snider, Keith L. 1988. "Towards the Representation of Tone: A Three-Dimensional Approach." In *Features, Segmental Structure and Harmony Processes*, edited by Harry Van Der Hulst and Norval Smith, 1:237–68. Linguistic Models 12. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110875812-009.

Snider, Keith L. 1990a. "Tone in Proto-Guang Nouns∗." *African Languages and Cultures* 3: 87–105. https://doi.org/10.1080/09544169008717712.

Snider, Keith L. 1990b. "Tonal Upstep in Krachi: Evidence for a Register Tier." *Language* 66 (3): 453–74. https://doi.org/10.2307/414608.

Snider, Keith L. 1992. "'Grammatical Tone' and Orthography." *Notes on Literacy* 18 (4): 25–30.

Snider, Keith L. 1999. *The Geometry and Features of Tone*. Dallas, Texas: Summer Institute of Linguistics.

Snider, Keith L. 2014. "On Establishing Underlying Tonal Contrast." *Language Documentation & Conservation* 8:707–31. http://hdl.handle.net/10125/24622.

Surendran, Dinoj. 2003. "The Functional Load of Phonological Contrasts." M.S. in Computer Science, Chicago, Illinois: University of Chicago.

Surendran, Dinoj, and Partha Niyogi. 2006. "Quantifying the Functional Load of Phonemic Oppositions, Distinctive Features, and Suprasegmentals." In *Current Issues in Linguistic Theory*, edited by Ole Nedergaard Thomsen, 279:43–58. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.279.05sur.

Wang, William S-Y. 1967a. "Phonological Features of Tone." *International Journal of American Linguistics* 33 (2): 93–105.

Wang, William S-Y. 1967b. "The Measurement of Functional Load." *Phonetica* 16:36–54.

Yip, Moira. 1993. "Tonal Register in East Asian Languages." In *The Phonology of Tone : The Representation of Tonal Register*, edited by Harry van der Hulst, Keith L. Snider, and Harry van der Hulst, 245–68. Linguistic Models. Berlin and New York: Mouton de Gruyter.

Yip, Moira. 2002. *Tone*. Cambridge; New York: Cambridge University Press.

Tonal complexity inversely correlates with onset/coda complexity in Sino-Tibetan

The presence or the complexity of tones is often associated to the complexity of consonants that can precede or follow a vowel, diachronically or synchronically. Diachronically, it is reported that tonogenesis most commonly arises as a compensatory process in regards to the loss of the complexity of onset or coda consonants, such as the loss of the voicing contrast (Hombert 1978; Kingston 2011). Synchronically, it has been observed that tonal lects often have less complex syllable structures, including complex onsets and codas: Maddieson's (2005) comparison between a ternary division of syllabic complexity (simple, moderate, and complex) and a ternary division of tonal complexity (atonal, binary, and ternary and beyond) in 543 lects shows that lects with complex syllable structures are less likely to have a large number of tones or be tonal at all. The diachronic and synchronic evidence lead us to hypothesize the compensatory nature between the number of tones and the number of distinctive onset and coda forms.

To test this hypothesis, I use the phonotactic information of 102 Sino-Tibetan lects available in Phonotacticon 1.0, a cross-linguistic database containing the possible onset, nucleus, and coda forms of 516 Eurasian lects (Joo and Hsu 2024). I have chosen the Sino-Tibetan family as it is one of the largest families in the database, consisting of both tonal and atonal member lects, as well as being areally restricted to Mainland Southeast Asia and nearby areas. I have excluded outlier lects with more than 100 onset or coda forms. The linear regressions between the number of possible onset/coda forms and the number of tones shows that both the number of onsets and the number of codas are inversely correlated to the number of tones ($p < 0.001$, Figure 1). This confirms the theory that tones may compensate the lesser degree of distinctive cues delivered by the lesser number of possible onsets and codas.

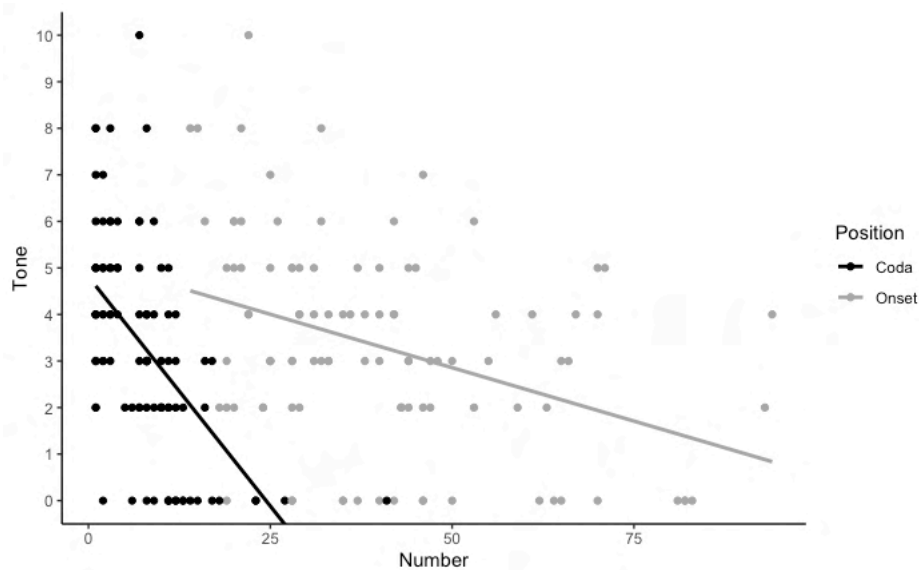**Keywords**: Tone; Sino-Tibetan; Phonotactics

1

Figure 1: Correlation between the number of possible onsets/codas (singletons and sequences) and the number of tones in Sino-Tibetan

References

Hombert, Jean-Marie (1978). "Consonant types, vowel quality, and tone". In: *Tone: A linguistic survey*. Ed. by Victoria Fromkin. New York: Academic Press, pp. 77–111. DOI: 10.1016/B978-0-12-267350-4.50008-X.

Joo, Ian and Yu-Yin Hsu (2024). "Phonotacticon: a cross-linguistic phonotactic database". In: *Linguistic Typology* (ahead of print). DOI: 10.1515/lingty-2023-0094.

Kingston, John (2011). "Tonogenesis". In: *The Blackwell Companion to Phonology*. Ed. by Marc van Oostendorp et al. Vol. 4. Malden, MA: Wiley-Blackwell, pp. 2304–2333. DOI: 10.1002/9781444335262.wbctp0097.

Maddieson, Ian (2005). "Issues of phonological complexity: Statistical analysis of the relationship between syllable structures, segment inventories and tone contrasts". In: *UC Berkeley PhonLab Annual Report* 1.1.

2

Negation and tone: A typological survey

Juan LI-NAAIJER

Bibliothèque nationale de France; Centre de recherches linguistiques sur l'Asie orientale

Major typological surveys have reported the use of tone to convey negation (Payne 1985; Dahl 1979, 2010; Dryer 2013; Miestamo 2017; Auwera & Krasnoukhova 2020). There is general consensus that this phenomenon is uncommon, with attested cases limited to a handful of African languages (cf. 7 out of 1,325 languages sampled in Dryer 2013), typically in conjunction with segmental coding such as affixation. For instance, in Degema [deg, Niger-Congo] (see example 1), negation involves both tonal alternation and morphological change in person proclitics (Kari 1997, 2001).

(1) Degema
a. *mʊ́=tá*   b. *ʊ́=ta*
    2SG=go   2SG.NEG=go
    'You will go.'   'You will not go.'   (Kari 2001: 185)

Typologists further agree that it is even rarer – though possible – for tone alone to encode negative polarity. Notably, when tone is the sole exponent of negation, it usually appears on a unit other than the semantically negated morpheme. A case in point is the Aboh dialect of Igbo [ukw, Niger-Congo] (see example 2), where the tone of the third-person subject pronoun distinguishes an affirmative sentence from its negative counterpart (Hyman 2011).

(2) Igbo (Aboh)
a.   */ò   jè   kò̩/   →   ò   jè   kò̩*   'S/he is going.'
     3SG   go   PROG

b.   */ó   jè   kò̩/   →   ó   jé   kò̩*   'S/he is not going.'   (H of /ó/ spreads onto /jè/)
     NEG.3SG go   PROG   (Hyman 2011: 203)

Drawing on data from a range of language families, this study compiles an extensive survey of the use of tone to mark negation, accessing the global distribution and diversity of this phenomenon. Crucially, it highlights the possibility for tone to directly and exclusively target the morpheme that is semantically negated. To the best of my knowledge, such behavior is attested only in Tai Lue [khb, Kra-Dai] (Li 2022) (see example 3), Tai Phake [phk, Kra-Dai] (Morey 2005, 2008), and Yoloxóchitl Mixtec [xty, Oto-Manguean] (Palancar, Amith & García 2016).

(3) Tai Lue
     ku⁵⁵   ha²⁵   pɔ⁴⁴   su⁵⁵,   ku⁵⁵   ha⁵⁵   mɛ⁴⁴   su⁵⁵
     1SG   NEG.look.for father   2SG   1SG   look.for   mother   2SG
     'I do not look for your father, I look for your mother.'   (Li 2022: 475)

In the survey, the marking of negation in each sampled language is coded for its segmental and/or tonal features, as well as its syntagmatic position. Furthermore, by examining whether the tonal

marking of negation is redundant (i.e., whether negation is marked by tone alone or in combination with segmental alternation), this study explores the interplay between phonology and morphology and considers the evolutionary stages of the attested systems. The findings identify implicational universals and outline the conditions that promote the emergence and further development of such systems, contributing to a deeper understanding of tonal negation and its typological implications.

References

Auwera, Johan van der & Olga Krasnoukhova. 2020. The typology of negation. In Viviane Déprez & M. Teresa Espinal (eds.), *The Oxford handbook of negation*, 91–116. Oxford: Oxford University Press.

Dahl, Östen. 1979. Typology of sentence negation. *Linguistics* 17. 79–106.

Dahl, Östen. 2010. Typology of negation. In Laurence R. Horn (ed.), *The expression of negation*, 9–38. Berlin, New York: De Gruyter Mouton.

Dryer, Matthew S. 2013. Order of negative morpheme and verb (v2020.4). 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/143. (10 April, 2025).

Hyman, Larry M. 2011. Tone: Is it different? In John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds.), *The handbook of phonological theory* (Blackwell Handbooks in Linguistics), 197–239. 2nd edn. Chichester: Wiley-Blackwell.

Kari, Ethelbert. 2001. On the grammar of clitics in Degema. *Journal of Asian and African Studies* 61. 179–192.

Kari, Ethelbert E. 1997. *Degema* (Languages of the World/Materials 180). München: Lincom Europa.

Li, Juan. 2022. *Grammaire descriptive du tai lü [A descriptive grammar of Tai Lue]*. Paris: Sorbonne Université Ph.D. dissertation. https://theses.hal.science/tel-03828164#. (10 April, 2025).

Miestamo, Matti. 2017. Negation. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *The Cambridge handbook of linguistic typology* (Cambridge Handbooks in Language and Linguistics), 405–439. Cambridge: Cambridge University Press.

Morey, Stephen. 2005. *The Tai languages of Assam: A grammar and texts*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, the Australian National University.

Morey, Stephen. 2008. The Tai languages of Assam. In Van Nostrand Diller, Jerold Edmondson & Yongxian Luo (eds.), *The Tai-Kadai languages*, 207–253. London: Routledge.

Palancar, Enrique L., Jonathan D. Amith & Rey Castillo García. 2016. 12. Verbal inflection in Yoloxóchitl Mixtec. In Enrique L. Palancar & Jean Léo Léonard (eds.), *Tone and inflection: New facts and new perspectives*, 295–336. De Gruyter Mouton.

Payne, John. R. 1985. Negation. In Timothy Shopen (ed.), *Language typology and syntactic description. Volume I: Clause structure*, 197–242. Cambridge: Cambridge University Press.

The role of tone mapping in Hausa

Lora Litvinova

(INALCO, LLACAN)

This paper discusses the role of tone mapping in Hausa, a Chadic language spoken in West Africa. The specifics of the Hausa tonal system have been extensively described in linguistic studies (e.g. Leben 1971, Schuh 1989, Newman 2000, Jaggar 2001). Hausa distinguishes between H(igh) and L(ow) level tonemes. One of the rules that play an important role in Hausa grammar is the rule of tone mapping (also referred to as the rule of tone assignment). This rule implies that each toneme in a tonal melody is assigned to syllables of a word from right to left. In example (1), the three toneme H-L-H melody of the plural suffix -aːCeː is assigned to the form *báкàːкéː* by replacing the lexical tonal melody of its singular form. When the number of syllables within a word exceeds the number of tonemes in a melody, the leftmost toneme extends towards the left edge of the word, as in (2).

(1)   [báкàːкéː]   (2)   [ɽíːgúnàː]

        /báкàːкéː/   /ɽíːgunàː/

        |báкiː-aːCeː\HLH| |ɽìːgáː-unaː\HL|

        black.SG-PL   gown.SG-PL

        'black (pl.)'   'gowns'

The tone mapping rule allows us to define the boundaries of the *toneme* assignment domain, viz. referred to as *tonal span*, and the boundaries of the *tonal melody* assignment domain, viz. referred to as *prosodic word*. Applying the right-to-left mapping rule, the plural locative noun *màsallaːtái* 'mosques' in (3) can be analyzed as one prosodic word, since the tonal melody L-H of the plural suffix -ai is mapped on the whole word. However, the plural agentive noun *máɽùbuːtáː* 'writers' in (4) constitutes two prosodic words: *má* and *ɽùbuːtáː*, since the L toneme of the L-H melody of the plural suffix -aː does not extend onto the leftmost syllable *má*.

(3)   [màsàllàːtái]   (4)   [máɽùbùːtáː]

        /màsallaːtái/   /máɽùbuːtáː/

        |ma-sàllátàː-ai\LH| |má-rúbùːtáː-aː\LH|

        LOC-perform_prayer-PL   AGT-write-PL

        'mosques'   'writers'

In this talk, I will demonstrate how tonemes and tonal melodies correspond with segmental units in various morphological environments. Additionally, I will discuss tonal processes involved in tone mapping, such as tonal melody shift, tonal melody replacement, toneme deletion, and toneme extension.

References

Jaggar, Philip J. (2001), Hausa. Amsterdam & Philadephia: Benjamins.

Leben, William R. (1971), The morphophonemics of tone in Hausa. In Papers in African Linguistics, ed. C.-W. Kim & Herbert Stahlke, pp. 201-218. Edmonton: Linguistic Research.

Newman, Paul (2000), The Hausa Language, An Encyclopedic Reference Grammar. New Haven: Yale University Press.

Schuh, Russell G. (1989), The reality of Hausa "low tone raising"; a response to Newman & Jaggar. Studies in African Linguistics, Vol 20, 253–262.

Toneme in Guinean Kpelle

Maria Konoshenko

(University of Helsinki)

In recent years, tonologists have readdressed the notion of tonal inventory, more specifically revisiting the notion of *toneme* as the minimal inventory unit (Vydrin & Maslinsky in prep). In this talk, I will discuss Vydrin & Maslinsky's (in prep.) understanding of toneme as applied to Guinean Kpelle, a Southwestern Mande language spoken by ca. 460 000 people in the Republic of Guinea (Konoshenko 2014, 2019).

Guinean Kpelle features a binary L vs. H tone contrast; the TBU is the syllable. Tonal elements make up fixed melodies associated with a foot, usually corresponding to a word: /H/, /L^H/, /HL/, /LHL/, /L/ and /LH/ as shown in (1).

(1) Kpelle

/H/	láá 'name'	kɛ́ɛ́ 'peanut'	kélé 'raffia palm'
/L^H/		kɛ̀ɛ̌ 'field space'	kèlě 'small bee'
/HL/		kɛ́ɛ̀ 'djabara percussion'	kélê 'shed'
/L/	làà 'there'		kèlèŋ 'bird species'
/LHL/		kɛ̀â 'now'	kèlêŋ 'African fox'

These melodies can be specified at the underlying level, both lexically and grammatically; they also constrain the possible outcomes of surface rules, i.e. H spread, floating ^{H} anchoring and contour simplification. Items having /L^H/ melody end with a phonological floating ^{H}. On the surface, this melody is realised as flat low [L^o] before pause, as opposed to /L/ surfacing as low falling tone before pause. The /HL/ melody is realised as falling [HL] on CVV, e.g. *kêɛ* 'percussion type', and as [H.HL] on CVCV, e.g. *yílê* 'dog'. There are replacive grammatical melodies, e.g. fully replacive {L} marking the verb in some TAM constructions, and fully replacive {L^H} marking relational nouns in compounds; on a par with additive tonal prefixes, e.g. H tone pronominal prefix marking 1sg and L tone prefix for 3sg.

Putting Guinean Kpelle data into the specific context of the workshop, I formulate the talk's central question as follows: Should the notion of toneme as defined by Maslinsky and Vydrin be associated with elementary H and L units or with larger melodies in Guinean Kpelle?

On the one hand, both H and L are phonologically active and independent in Guinean Kpelle. H has the following properties: (a) H can float; (b) H can spread rightwards; (c) H functions as a grammatical prefixal tone marking 1SG polyfunctional pronoun. L has the following properties: (a) L can delink and float as a result of obligatory contour simplification rule; (b) L can function as a replacive tone marking nominal and verbal stems in certain morphosyntactic contexts; (c) L functions as a grammatical prefixal tone marking 3SG polyfunctional pronoun. These arguments support the interpretation of H and L as tonemes in Guinean Kpelle. However, there are two considerations pointing towards tone melodies as single units: first, there is a limited amount of melodies suggesting that they make up an inventory at a higher level (foot, rather than TBU); and second, /L^H/ melody functions as a replacive tone on relative nouns suggesting that it does function as a single unit.

To sum up, I argue that both atomic tones (tonemes) associated with a syllable and larger toneme sequences realized on feet (melodies) are needed to account for the tonal inventory and tone rules in Guinean Kpelle.

Funding

References

Maslinsky, Kirill and Valentin Vydrin (In prep), Revisiting tonal inventories: Criteria for toneme identification.

Konoshenko, Maria (2019), Dictionnaire kpele de la Guinée (guerzé) – français. *Mandenkan* 62. 3–164.

Konoshenko, Maria (2014), The syntax of tone in Guinean Kpelle. In: Proceedings of the fortieth annual meeting of the Berkeley linguistics society. February 7–9, 2014. 233–252.

Low tone anticipatory spreading in Babanki: A typological anomaly

Pius W. Akumbu & Valentin Vydrin (LLACAN – CNRS – INALCO & LLACAN – INALCO)

In an early study of the typology of tonal processes (Hyman & Schuh 1974), it was claimed that "the spreading of tones takes place always to the right and apparently never (or at least rarely) to the left". Thirty years later, Hyman recognized that, even if "cases of perseverative HTS vastly outnumber cases of anticipatory HTS", the tonal anticipation occurs in many languages (Hyman 2007). He analyzes a number of Bantu languages and concludes that the anticipatory spreading always concerns the high tone; it occurs mainly in the languages with a privative /H, Ø/ tonal system; and that H tone is usually attracted to a metrically strong position.

Babanki (Central Ring Grassfields Bantu, Cameroon) is a /L, H/ language. Besides two rightward spreading processes (perseverative HTS and LTS), it also has a leftward LTS process.

The anticipatory LTS occurs in one specific context and under certain conditions: when in the initial position in the associative construction, H-toned monosyllabic nouns belonging to classes 1 and 9 and ending in a vowel or underlying -ŋ (on the surface, -ŋ is always elided in this context) lose their lexical tone which is replaced by the L tone spreading leftwards from the associative marker ə̀:

mpfì	*ə̀*	*kə̀kɔ̀s*
mpfí	ə̀	kə̀-kòs
1.mother	1.AM	7-slave
'mother of slave'		

This rule is blocked if a noun ends in a glottal stop or *-m*:

mbjéʔ	*ə̀*	*kə̀kɔ̀s*
mbéʔ	ə̀	kə̀-kòs
9.shoulder	9.AM	7-slave
'shoulder of slave'		

The anticipatory LTS runs counter to all the constraints mentioned by Hyman: the Babanki tonal system is not privative; the process concerns L tone (rather than high); and it is not related to a metrically strong position.

Keywords: Spreading, Low tone, anticipatory, Babanki, Grassfields Bantu

References

Hyman, Larry M. (2007), Universals of tone rules: 30 years later. In T. Riad, and C. Gussenhoven (eds), *Tones and Tunes* (Typological Studies in Word and Sentence Prosody), vol. 1, 1–34. Berlin/New York: Mouton de Gruyter.

Hyman, Larry M. & Russel G. Schuh (1974), Universals of Tone Rules: Evidence from West Africa. *Linguistic Inquiry* 5(1). 81–115.

When Welmers first wrote about "Sup'ide" (1950; 1973) he called particular attention to its complex tonal system. Carlson (1994) expands on that earlier work, but not with sufficient depth to enable current tonologists to situate Supyire (spp; Senufo; south-central Mali) within a typology of grammatical tone. Thus Rolle (2018) in the typology section of his work on grammatical tone mentions the "grammatical tone sandhi … found in Supyire" (2018:111), but does not say more than that the "complex [tone] changes in Supyire … straddle the grammatical tone / tone sandhi distinction" (2018:37). Under the narrow definition he uses in that work, Supyire has little that qualifies as grammatical tone. Even under the broader definition of grammatical tone in Lionnet et al. (2022) ("a tonological pattern (e.g. involving tone addition, replacement, shifting, assimilation, dissimilation, etc.), which is restricted to a specific morpheme or construction or a natural class of morphemes or constructions, and not attributable to the general tonal phonology"), most of the tonal processes of Supyire do not clearly qualify as "grammatical" (cf. also Konoshenko 2017). As Rolle (2018) puts it, Supyire seems to be on the borderline of grammatical tone. On the one hand, there are no tonal processes for which phonological conditioning factors are not at least partly responsible, yet on the other hand it is impossible to adequately describe grammatical constructions without specifying which tonal processes take place in them.

In its "general tonal phonology", Supyire has 3 tone levels, but with additional complexities such as: (i) M tone and L tone morphemes are each split into two morphological classes which differ from each other in how they participate in tonal processes; (ii) roughly a fifth of noun roots are followed by a floating L tone; (iii) a floating L also follows all definite noun class suffixes and all demonstrative pronouns. The floating Ls mostly cause the same sort of tone changes in the following word that a L noun does. The way the system is presented in Carlson (1994), arranged by tonal "changes" (e.g. "low-spread", "high-spread", "conversion of mid to high" and so on), does not systematically address the question of the distribution of tone processes across constructions and therefore to what extent a construction may be signaled by such processes whose grammatical function is therefore "emergent" (cf. Bybee 2001, Nesset 2008).

This paper aims to remedy that lack by describing the tone changes which occur in the most common constructions (possessor-possessum, verb-object, object-postposition, TAM-verb, subject-TAM, noun-modifier) both on a type level and at the token level in a corpus of Supyire texts comprising about 60,000 clauses.

REFERENCES

Bybee, Joan. 2001. *Phonology and language use.* Cambridge: Cambridge University Press.
Carlson, Robert. 1994. *A grammar of Supyire.*
Konoshenko, Maria B. 2017. Tone in grammar: what we already know and what we still don't. *Voprosy Jazykoznaniya* 4.101-114.
Lionnel, Florian & McPherson, Laura & Rolle, Nicholas. 2022. Theoretical approaches to grammatical tone. *Phonology* 39:3 385-398.

Nesset, Tore. 2008. *Abstract phonology in a concrete model: cognitive linguistics and the morphology-phonology interface.* Berlin: Mouton de Gruyter.

Rolle, Nicholas. 2018. *Grammatical tone: typology and theory.* Berkeley Papers in Formal Linguistics 1:1.

Welmers, William. 1950. Notes on two languages in the Senufo group: II Sup'ide. *Language* 26:494-531.

Welmers, William. 1973. *African language structures.* Berkeley: University of California Press.

Tone marking conventions: A cross-linguistic perspective

Sonja Dahlgren, Sami Honkasalo & Don Killian
(University of Helsinki)

Keywords: accent marks, tone letters, tone marking, tones, typology

Three widely used conventions exist for marking tones in tonal languages, causing confusion in typological comparison. Depending on the language in question and the assumed readership, different styles of tone marking are preferred. Tone is commonly marked with diacritic accent marks in most African languages, for example; in languages of Southeast Asia, however, tones tend to be marked by means of either numbers or tone letters (Chao 1930).

Having multiple conventions for marking tones causes confusion in typological comparison. From the three most common systems, accent marks are the most transparent for non-specialist readers, with tones directly marked on the tone bearing units (TBU). They cannot, however, handle all tonal contours across languages. Furthermore, in older grammars, accent marking has confusingly been used for the marking of both stress and tones. Tone letters show more contour flexibility than accent diacritics, but are placed after the TBU, which is a less transparent convention. Numbers are the simplest for denoting different contours, but have cross-linguistic variation in their denotations. Southeast Asia typically uses the system devised by Y. R. Chao, in which 5 represents the highest tone, and 1 the lowest. Robert Longacre (1952), in contrast, devised a system in Mesoamerica in which 1 represents the highest pitch, and 5 the lowest. Nowadays, even the same language can sometimes show differing transcription methods; Skilton (2023) uses the Chao representation for Ticuna tone, whereas earlier research on Ticuna used the Longacre method (e.g. Anderson 1962). Furthermore, the numerical values tend not to be well-known outside of areal specialists, so using numbers to represent tones is not transparent. In this paper, we present suggestions on simplifying the system.

A single system to mark tones would be ideal, but unlikely given the variation in tone systems. Most tone systems in Africa show maximally three pitch levels, making tone letters unnecessarily complicated. Southeast Asian languages, however, often show systems with many contours; some languages additionally incorporate voice registers into their tonal phonology (e.g. creaky voice). Accent marks are problematic for languages with more than three levels, but tone letters are usable for most tones. Largely following Maddieson (1990), we recommend keeping accents for simpler tone systems, and using tone letters for more complex ones. This would reduce the system to two conventions.

Tone letters would, however, benefit from some refinement. Currently, tone letters are often placed syllable- or word-finally, reinforcing the misleading notion of tones being peripheral aspects of phonology. Instead, tone letters could be placed directly next to the TBU. This would also avoid issues where word-final rhotics can be both TBUs as well as syllable-codas; cf. Wawa mbú˦ɾ˦ 'nuisance,' búɾ˦ 'door' (Martin 2012:56). Furthermore, voice register features in tone are typically marked separately from tone, despite the role they play in tonal systems (e.g. Burmese). Voice register feature marking could be placed underneath the tone letter, rather than the nucleus (e.g. x˩̰ for a creaky low tone and x˧˥̤ for a breathy mid-rising tone).

References

Anderson, Doris (1962), Conversational Ticuna, Yarinacocha, PE: Summer Institute of Linguistics. https://www.sil.org/resources/archives/30245.

Chao, Yuen-Ren (1930), ə sistim əv "toun-letəz" [A system of "tone-letters"], *Le Maître Phonétique* 30, 24–27.

Longacre, Robert E. (1952), Five phonemic pitch levels in Trique, *Acta linguistica* 7(1-2), 62-82.

Maddieson, Ian (1990), The transcription of tone in the IPA, *Journal of the International Phonetic Association* 20(2), 28-32.

Martin, Marieke (2012), A grammar of Wawa: an endangered language of Cameroon, University of Kent dissertation.

Skilton, Amalia (2023), Tone, stress, and their interactions in Cushillococha Ticuna, *Phonological Data and Analysis* 5(5), 1-44.

WS25 United by language contact: Linguistics and Translation-Interpreting Studies in search of a shared framework

Gert De Sutter, Maja Miličević Petrović, Silvia Bernardini, Adriano Ferraresi, Moritz Schaeffer & Paweł Korpal

A new method in the study of language contact: Variation-Based Distance & Similarity Modeling

Benedikt Szmrecsanyi, Jason Grafmiller & Laura Rosseel
(KU Leuven, University of Birmingham & Vrije Universiteit Brussel)

Keywords: comparative sociolinguistics, variation, probabilistic grammar, dialectometry, variationist linguistics

Inspired by work in comparative sociolinguistics (e.g. Tagliamonte 2001) and quantitative dialectometry (e.g. Nerbonne, Heeringa & Kleiweg 1999), this talk sketches a corpus-based, variationist method (Variation-Based Distance & Similarity Modeling – VADIS for short) to rigorously quantify the similarity between varieties and dialects as a function of the correspondence of the ways in which language users choose between "alternate ways of saying 'the same' thing" (Labov 1972: 188). In other words**,** the basic idea is to measure inter-speaker variability by assessing intra-speaker variability.

VADIS relies on three lines of evidence to determine relatedness between varieties:

1. Are the same constraints significant across varieties?

2. Do the constraints have the same strength across varieties?

3. Is the constraint hierarchy similar?

In comparative sociolinguistics, similarity according to these lines of evidence is often interpreted as historical and genetic relatedness. VADIS draws inspiration from this literature and adapts the comparative sociolinguistics method so that it can be scaled up to the study of more than a couple of lects, and to more than one variable phenomenon at a time. Technically speaking, the VADIS workflow consists of multivariate modeling (both regression and CRF) of the variation phenomena under study. The output of this multivariate modeling is subsequently used as input to distance/similarity calculations.

To showcase the potential of the method, we present a case study that investigates three syntactic alternations in the grammar of English: the dative alternation (1), the genitive alternation (2) , and the particle placement alternation (3).

(1) The genitive alternation
    a. *the country's economic crisis*   (the *s*-genitive variant)
    b. *the economic growth of the country*   (the *of*-genitive variant)

(2) The dative alternation
    a. *I'd given Heidi my T-Shirt*   (the ditransitive dative variant)
    b. *I'd given the key to Helen*   (the prepositional dative variant)

(3) The particle placement alternation
    a. *just cut the tops off*   (the 'split' verb-object-particle variant)
    b. *cut off the flowers*   (the 'continuous' verb-particle-object variant)

These in principle well-known alternations are studied in nine international varieties of English, including some well-known contact varieties and indigenized L2 varieties of English (British, Canadian, Irish, New Zealand, Hong Kong, Indian, Jamaican, Philippines, and Singapore English).

Key findings include the following: (a) probabilistic grammars are remarkably similar and stable across the varieties under study; and (b) there is a fairly robust split between L1 (a.k.a. Inner Circle) varieties, such as British English, and L2 (a.k.a. Outer Circle) varieties, such as Indian English. We will argue that VADIS has the potential to contribute to theorizing about how not only surface structure

but also more subtle variation patterns are subject to contact effects. VADIS also has application potential in regard to the comparison of interpretation and translation output to other lects.

References

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.

Nerbonne, John, Wilbert Heeringa & Peter Kleiweg. 1999. Edit Distance and Dialect Proximity. In *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*, v–xv. Stanford: CSLI Press.

Tagliamonte, Sali. 2001. Comparative sociolinguistics. In *Handbook of Language Variation and Change*, 729–763. Malden and Oxford: Blackwell.

Intratextual priming in translated versus original historical texts – the case of the dative alternation in Middle English

Gunnar Jacob and Mirja Maier

University of Mannheim

Keywords: Structural priming, language production, translation, writing, Middle English

There is substantial evidence that structural choices in spoken and written language production, such as the choice between whether to produce a prepositional-object (PO) or a double-object (DO) structure during the production of an English ditransitive sentence, are influenced by the structure of previous sentences the speaker has encountered or produced shortly before (e.g. Bock, 1986, and much subsequent work). Such *structural priming* effects have also been shown in texts, where structural choices are influenced by the structure of previous sentences in the same text (i.e. *intratextual priming*).

In the production of translated texts, however, structural choices are also substantially affected by the structure of the source sentence. Such *intertextual priming* effects in translation may potentially reduce or even entirely prevent intratextual priming in translated texts. De Sutter, Colleman, and Ghyselen (2021) investigated this possibility in a corpus study comparing intratextual priming effects for the overt realization of *that* in complement clauses in original versus translated English texts. The results showed significant intratextual priming for original, but not for translated texts, suggesting that the influence of the structure of the source sentence may largely overwrite intratextual priming effects.

If it is indeed the case that intertextual priming in translation reduces or entirely overwrites intratextual priming, the difference between translated and original texts should occur irrespective of time period of origin, text type, translation style, and structural alternation investigated. In the present study, we test this assumption by comparing intratextual self-priming effects for the dative alternation in 17 translated versus 12 non-translated Middle-English texts from M2 and M3 sections of the *Penn-Helsinki Parsed Corpus of Middle-English* (*PPCME2*).

The analysis relies on Gries' (2005) approach for investigating structural (self-)priming effects in corpus data. A corpus search for ditransitive sentences with a PO or DO structure yielded a total of 1489 ditransitive target sentences. All targets were annotated for the structure of the target (PO or DO), the structure of the preceding prime, i.e. the last previous ditransitive sentence in the same text (PO or DO), prime-target distance, text type (translated or non-translated text), and whether prime and target shared the same verb or not.

For both translated and non-translated texts, the results showed a higher proportion of PO target sentences following PO primes than following DO primes. In addition, both text types also showed *lexical boost* effects of a similar magnitude, i.e. stronger intratextual priming when prime and target shared the same verb. A generalized-linear-mixed effects model predicting target structure revealed significant intratextual priming effects and lexical boost effects across both text types, but no interactions between priming and text type. This suggests that intratextual priming was not significantly weaker in translated than in non-translated texts. We thus conclude that intratextual priming can in principle also occur in translated texts, and suggest that the interplay between intratextual and intertextual priming may be moderated by text type and translation style.

References

De Sutter, G., Colleman, T., & Ghyselen, A. S. (2021). Intra-and inter-textual syntactic priming in original and translated English. *Cognitive sociolinguistics revisited*, *48*, 410.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research*, *34*, 365-399.

Language variation, language change and translation through a usage-based lens: Language contact between Dutch/Afrikaans and English in South Africa

How does translation, as a form of language contact, relate to broader ongoing processes of (contact-influenced) variation and change? Existing studies attempting to answer this question are limited empirically by the availability of suitable diachronic corpora, and theoretically by the absence of a coherent linguistic framework for studying translation alongside other forms of contact-influenced language use (see Kotze 2020). In this paper we address these two limitations. We make use of a diachronic, register-differentiated, bidirectional combined parallel and comparable translation corpus of Afrikaans original texts and their English translations (and vice versa) across five registers, produced in South Africa in the period 1910 to 2016. We study the retention versus omission of the complementiser *that* in South African English and its equivalent in Afrikaans, *dat*. This feature is an example of clear cross-linguistic constructional variation and change, in both languages: The omission of the complementiser *dat* is a contact-influenced innovation in Afrikaans, and in both languages the feature varies across registers and time (Feinauer 1989; Kruger & Van Rooy 2016; Van Rooy 2021).

Our study combines frequentist and variationist approaches. We first analyse the frequency of *that/dat*-omission over time in original Afrikaans and English texts across different registers, as well as in Afrikaans translations (from English) and English translations (from Afrikaans). Subsequently we use random forests and conditional inference tree analysis to investigate the language-internal and language-external factors that condition the choice between overt or omitted *that/dat* in the two languages and production modes (originals and translations) over time (see also Kruger & De Sutter 2018; Kruger 2019). Linking our empirical findings to usage-based frameworks of constructional change (see Van Rooy 2021) we make a contribution to the further theorisation of translation as a form of language contact, within broader linguistic contact settings.

References
Feinauer, A. E. 1989. Plasing in Afrikaanse afhanklike sinne [Positionality in Afrikaans dependent clauses]. *Suid-Afrikaanse Tydskrif vir Taalkunde* 7: 30-37.
Kotze, Haidee. 2020. Translation, language contact and cognition. In Fabio Alves & Arnt Lykke Jakobsen, eds. *The Routledge Handbook of Translation and Cognition*. Routledge. pp. 113-132.
Kruger, Haidee. 2019. *That* again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures* 20(1): 1-33.
Kruger, Haidee & Gert de Sutter. 2018. Alternations in contact and non-contact varieties: Reconceptualising *that*-omission in translated and non-translated English using the MuPDAR approach. *Translation, Cognition and Behavior* 1(2): 251-290.
Kruger, Haidee & Bertus van Rooy. 2016. Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences* 56: 118-131.
Van Rooy, Bertus. 2021. A diasystematic construction grammar analysis of language change in the Afrikaans and English finite verb complement clause construction. In Hans Boas &

Steffen Höder, eds. *Constructions in Contact 2: Language Change, Multilingual Practices, and Additional Language Acquisition.* John Benjamins. pp. 109-138.

Bridging constrained communication and cognitive semantics:
Elicitation approaches to constraints in Chinese-French translation and L2 production

Lin Shen

(University of Cambridge)

Keywords: constrained communication, motion events, Chinese and French, cognitive constraints

Similarities between translated language and second language (L2) production have gradually emerged (Chesterman, 2004), and rather than isolating each discipline for separate analysis, it may be more productive to develop a unified theory of "constrained communication" (De Sutter & Lefer, 2020, p. 6). This is a concept proposed by Lanstyák and Heltai (2012) and further developed into a framework by following studies (Kruger, 2018).

Existing studies on constrained communication primarily use corpus analysis, focusing on formal features such as lexical diversity (Kajzer-Wietrzny & Ivaska, 2020). This study seeks to diversify the literature by introducing elicitation methods in the cognitive semantics studies of motion expressions (Hendriks & Hickmann, 2015). More specifically, this study addresses two research questions (RQs):

- RQ1: Are cognitive constraints shared in motion event encoding between L2 French translation and description tasks?
- RQ2: Do translation and description tasks differ in cognitive constraints and extent of L1 transfer?

This study integrates Talmy's (2000) typology and Kotze's (2022) constrained communication framework. Participants include 15 Chinese monolinguals, 15 French monolinguals, and 15 Chinese-French bilinguals. Monolinguals perform L1 video description tasks, while bilinguals complete both L2 description and translation tasks. Each task involves 24 motion events across five paths—DOWN, UP, OUT, INTO, and ACROSS.

The constrained communication framework models language production with five constraints. The tasks are controlled to ensure that the only difference lies in text production while keeping other cognitive constraints consistent.

- **Language motivation:** Bilingual (L1→L2)
- **Modality and register:** Written
- **Text production:**
  - L2 Description: Mediated (video stimuli)

- L2 Translation: Mediated (source texts)
- **Proficiency:** Learner
- **Task expertise:** Non-expert

The responses are coded in four aspects of motion event encoding: verb locus, other locus, semantic density, and syntactic packaging. Chi-square tests are conducted to compare the differences in four aspects between L1 and L2 descriptions, as well as between source texts (ST) and target texts (TT) in translation tasks.

For RQ1, findings indicate that cognitive constraints are shared between task types. For example, both tasks display partial shifts from the equipollently-framed patterns (manner + path verbs) in Chinese to the verb-framed patterns (path verbs) in French, as illustrated in (1) and (2).

(1) 篮球　　　从楼梯上　滚[Manner]　　落下来[Path] (L1 Chinese)
　　A basketball　from upstairs tumbled [Manner] fell [Path]
　　Un ballon de basket descend [Path] des escaliers. 　　　(L1 French)
　　Un ballon de basket descend [Path] de l'escalier. 　　　(L2 French)
(2) 一个篮球　　滚[Manner]　　　下[Path]了　楼梯　　　(ST Chinese)
　　A basketball　tumbled [Manner] fell [Path]　　the stairs
　　Un ballon est tombé [Path] de l'escalier en roulant. 　　(TT French)

For RQ2, translation tasks demonstrate greater shifts towards L2 attunement than description tasks. For instance, the translation tasks display higher frequencies of path verbs (48.33%) than the L2 description tasks (44.72%).

These findings support previous research suggesting that translation tasks promote deeper engagement with L2 structures (Lewandowski & Özçalışkan, 2024; Neumann et al., 2024). The structured nature of translation, guided by the richness of ST, may explain this difference.

References

Chesterman, A. (2004). Hypotheses about Translation Universals. In G. Hansen, K. Malmkjaer, & D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies: Selected Contributions from the EST Congress, Copenhagen 2001* (pp. 1–13). John Benjamins. https://doi.org/10.1075/btl.50.02che

De Sutter, G., & Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1), 1–23. https://doi.org/10.1080/0907676X.2019.1611891

Hendriks, H., & Hickmann, M. (2015). Finding One's Path Into Another Language: On the Expression of Boundary Crossing by English Learners of French. *MODERN LANGUAGE JOURNAL*, *99*, 14–31. https://doi.org/10.1111/j.1540-4781.2015.12176.x

Kajzer-Wietrzny, M., & Ivaska, I. (2020). A Multivariate Approach to Lexical Diversity in Constrained Language. *Across Languages and Cultures*, *21*(2), 169–194. https://doi.org/10.1556/084.2020.00011

Kotze, H. (2022). Translation as constrained communication. In S. Granger & M.-A. Lefer (Eds.), *Extending the Scope of Corpus-Based Translation Studies* (pp. 67–97). Bloomsbury Academic. https://doi.org/10.5040/9781350143289

Kruger, H. (2018). Expanding the third code: Corpus-based studies of constrained communication and language mediation. In S. Granger, M.-A. Lefer, & L. Penha-Marion (Eds.), *Book of abstracts. Using corpora in contrastive and translation studies conference (5th edition)* (pp. 9–12). Centre for English Corpus Linguistics/Université catholique de Louvain.

Lanstyák, I., & Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures*, *13*(1), 99–121. https://doi.org/10.1556/Acr.13.2012.1.6

Lewandowski, W., & Özçalışkan, Ş. (2024). Translating Motion Events Across Physical and Metaphorical Spaces in Structurally Similar Versus Structurally Different Languages. *Metaphor and Symbol*, *39*(1), 10–39. https://doi.org/10.1080/10926488.2023.2268685

Neumann, S., Kerz, E., & Heilmann, A. (2024). Comparing contact effects in translation and second language writing. In A. J. Van Rooy & H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings* (pp. 223–255). John Benjamins Publishing Company.

Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press.

Simulating the Translating Mind

Translation scholars seem to agree that several concurrent translation processes unfold synchronously in the translator's mind, each characterized by sets of functions and constraints that contribute to the final translation product in some ways. However, the complexity of interactions between these layers and constraints make it difficult for researchers to analyze and disentangle their contribution in the observable translation behavior.

In order to address this complexity and to advance the field of enquiry, I suggest a novel approach to cognitive translation and interpretation studies (CTIS) in which a multi-layered artificial agent simulates those translation processes and generates translation behavior that can be assessed and validated against human-generated empirical data. The agent's internal parameters can be tuned to various translator profiles, texts, language combinations, etc. to generate a range of different translation behaviors. A simulation approach is common in various fields of research where numerous variables interact in complex unknown ways too complex to be modeled through closed, analytical methods, but a simulation approach has not been proposed or explored in CTIS to date.

I suggest an ABC-approach for modelling the translating mind as a hierarchicy of three embedded processing layers: A: a phenomenal layer which reflects the translators affective/emotional states, B: a sensorimotor layer which produces behavioral data (keystrokes and gaze data), and C: a cognitive layer which simulates reasoning and reflective thought. These three layers are tightly interconnected but follow layer-specific constraints and timelines. Following up on previous work, I propose to model the hierarchical architecture within an embedded Bayesian framework, more specifically using active inference (AIF) and predictive processing (PP). AIF and PP posit that the mind constructs a generative model of the task at hand to guide the agent's behavior by continually generating and integrating predictions and sensory input.

The translation agent will be trained and evaluated on a large dataset of behavioral keystroke and gaze data, the CRITT Translation Process Research Database (TPR-DB). The TPR-DB contains several hundred hours of translation behavioral data (keystrokes and gaze data) that covers a large variety of translation configurations and translation settings. Previous research suggests that this behavioral translation data can be fragmented on various levels of granularity into different process and product units which illustrate/represent traces of the three embedded ABC layers. The aim of the translation agent is, then, to generated (simulate) sequences of these units that reflect specific translation configurations and thus different translation styles.

In this talk I focus on the shape and characteristics of the process and product units that the agent is supposed to generate/traverse. I outline how the process unints can be assembled - on various levels of granularity - into a Behavioral Translation Style Space (BTSS). In its current form, the BTSS is a multi-dimensional space that encodes properties of the most basic Activity Units (AUs). AUs are characterized by eight basic keystroke- and gaze-path related features which span the base-BTSS. Sequences of AUs combine into higher-level Task Segments and HOF states which reflect instances of the C and A layer respectively.

What can a diachronic annotated corpus of re-translated texts offer to the study of language contact?
Pleonastic object pronouns and cognate objects in the history of Greek

Nikolaos Lavidas & Sofia Chionidi
(National and Kapodistrian University of Athens)

Translations as a significant aspect of written language contact have been explored in various studies on contact—such as those concerning ancient languages and religious texts. Specific grammatical and stylistic features have also been viewed as connected to the impact of translations in the standardization and the establishment of the prestige of the target language. Additionally, there is agreement that translations have played a significant role in the incorporation of loanwords—such as in the case of Latin loanwords introduced into Early English (among others, Steiner 2008, Delisle & Woodsworth 2012, Lavidas 2021).

In the present study, we discuss data from a diachronic annotated corpus that includes details on the argument structure of (selected) verbs with a focus on genealogical features and contact-induced changes. The annotated corpus contains texts from various periods of the history of Greek and, mainly, re-translations of biblical and Classical Greek texts produced in different stages of the diachrony of Greek.

The following hypothesis found in the traditional literature serves as the starting point of the study: (i) The presence of pleonastic object pronouns in the Septuagint (the Koiné Greek translation of the Hebrew Old Testament) stems from the contact between Greek and Hebrew (see Ex. 1a vs 1b – Janse 2002, George 2010); (ii) Cognate objects are also commonly attested in the Septuagint (see Ex. 2); (iii) The absence of referential null objects in later (post-Koiné) Greek is also seen as having been influenced by the contact with Hebrew.

We contend against a direct correlation between the relevant change in transitivity in later Greek and the features of transfer from Hebrew in the Septuagint (which is the conventional perspective presented in the literature). Hence, we suggest a more intricate relation between written contact and its community- and system-level consequences through language change. Our discussion of the proposed relation is grounded in principles of language change and processes of language transfer seen in L2 acquisition and L1 attrition. We demonstrate that a mixed primary input for L1 in the community implies that speakers may transfer resumptive pronouns from another language (*Interpretability Hypothesis*; Tsimpli 2003 and elsewhere) while also attempting to reduce the semantic features of the derivation according to the principles of internal change as related to language acquisition (*Feature Economy Principle*; van Gelderen 2008 and elsewhere).

(1) a.   Koiné Greek - Septuagint

kaì	*anéstē*	*Kain*	*epì*	*Abel*	*tòn*
and	rose-up.3SG	Cain.NOM	against	Abel.ACC	ART.ACC

adelphòn	*autoû*	*kaì*	*apékteinen*	***autón.***	
brother.ACC	3SG.GEN	and	killed.3SG	3SG.ACC	

'And Cain rose up against Abel, his brother, and killed him.' (Ge.4:8)

b. vs Ancient Greek - Xenophon

Null objects are available

ho	*dè*	*empimplàs*	*hapántōn*	*tèn*
3SG.NOM	PTC	satisfied.PRT	every-one.GEN	ART.ACC

gnṓmēn	***apépempe***	**Ø**
expectation.ACC	dismissed.3SG	Ø

'And after having satisfied the expectations of every one of them, he dismissed them.' (Xen. An. 1, 7, 8; Luraghi 2003: 169)

(2) Koiné Greek - Septuagint

ploutḗsei	***ploûton***	*mégan.*
enrich.FUT.3SG	richness.ACC	big.ACC

'[…] (he) will become much richer (than all the others).' (Dan.11:2)

References

Delisle, Jean, and Judith Woodsworth (2012), *Translators through History*, Revised Edition, Amsterdam: John Benjamins.

van Gelderen, Elly (2008), Where did Late Merge go? Grammaticalization as feature economy, *Studia Linguistica* 62(3), 287-300.

George, Coulter H. (2010), Jewish and Christian Greek, in E. J. Bakker (ed.), *A Companion to the Ancient Greek Language*, Oxford: Blackwell, 267-280.

Janse, Mark (2002), Aspects of bilingualism in the history of the Greek language, in J. N. Adams, Mark Janse, and Simon Swain (eds), *Bilingualism in Ancient Society: Language Contact and the Written Word*, Oxford: Oxford University Press, 332-390.

Lavidas, Nikolaos (2021), *The Diachrony of Written Language Contact: A Contrastive Approach*, Brill's Studies in Historical Linguistics, Leiden / Boston: Brill.

Luraghi, Silvia (2003), Definite referential null objects in Ancient Greek, *Indogermanische Forschungen* 108, 167-194.

Steiner, Erich (2008), Empirical studies of translations as a mode of language contact. "Explicitness" of lexicogrammatical encoding as a relevant dimension, in P. Siemund and N. Kintana (eds), *Language Contact and Contact Languages*, Amsterdam / Philadelphia: John Benjamins, 317-346.

Tsimpli, Ianthi-Maria (2003), Clitics and determiners in L2 Greek, in J. Liceras, H. Goodluck, and H. Zobl (eds), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*, Somerville, MA: Cascadilla Proceedings Project, 331-339.

Bridging the gap between translation and translanguaging: towards a general framework for the study of transcultural communication

Adopting a transdisciplinary view of transcultural communication, this paper seeks to explore parallelisms between *languaging* processes (Love 2017) characterising translation and transcultural interactions. Translation and mediation are increasingly understood as fluid and dynamic processes rather than static acts of language transfer. For example, Cronin and Simon (2014) describe translation as a central communicative activity in the era of globalization. Similarly, the role of translation as negotiation of meaning across cultures and languages is emphasised in the updated version of the CEFR Companion Volume (2018). Although a disclaimer is added therein regarding the difference between mediation activities and the work of professional translators, in fact recent work suggests that significant overlaps exist, particularly in the cognitive and social mechanisms underlying these practices.

Following Lanstyák and Heltai's conceptualisation of translators as a sub-group of bilinguals (2012), in this contribution we analyse a dataset of spoken interactions among plurilingual speakers in transcultural settings – more specifically digital nomads (Woldoff & Lichtfield 2021) – with a view to identifying similarities between translation and transcultural communication. From a cognitive perspective, the revised *gravitational pool hypothesis* developed within translation studies (Halverson 2017) highlights the way in which bilinguals pair and systematise linguistic forms and meanings in an interconnected manner across language varieties. Similarly, in bilingualism studies, the concept of *translanguaging* was proposed to describe semiotic repertoires of plurilingual speakers as holistic, adaptive, idiosyncratic and socially-bound (García & Wei 2014). Despite pertaining to different traditions, these theories appear to share numerous similarities. For instance, the metaphor of the "third code", first introduced by Frawley (1984) with reference to translated language, and taken up by Kotze (2022), also describes well the pidgins and linguae francae typical of transcultural communication (Seidlhofer 2007; Holmes & Dervin 2016). Furthermore, in proposing a view of translation as 'constrained communication', Kotze (2022) adapts the variationist model proposed by Szmercsanyi and Kortmann (2009) to understand the (cross-)linguistic and sociocultural constraints affecting translation, thus demonstrating the relevance of sociolinguistic approaches to translation studies.

Digital nomad communities, composed of remote-workers with diverse linguistic and cultural backgrounds, have been chosen for this case study because their communicative behaviours well reflect contemporary changes in language use due to increasing mobility and cultural

fluidity. Relying on samples from a dataset of spoken interactions collected at social events of digital nomad communities in Europe, our contribution highlights the pertinence of features traditionally attributed to translation practice for the description of transcultural communication, and specifically: 1. the scope of mediation; 2. bilingual activation; 3. activation of language learning processes, including hypothesis checking; 4. metalinguistic reflection, including crosscultural comparison and ideological use of language; 5. use of external supports, i.e. technology-augmented communication (adapted by Kotze 2022 and Lanstyák and Heltai 2012).

Building on the results of our analysis, the ultimate aim of our contribution is to work toward a shared theoretical framework for the study of language produced by plurilingual speakers in transcultural settings, overcoming conceptual and disciplinary boundaries that have so far hidden the inherent commonalities. This shift in conceptualisation has the twofold potential of changing conceptions about the role and skills of professional translators, and of highlighting mediation and transcultural competences as salient abilities to be addressed in language learning and translator education syllabuses.

References

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume with new descriptors.* Council of Europe Publishing.

Cronin, M., & Simon, S. (Eds.). (2014). The City as Translation Zone. *Special Issue of Translation Studies, 7(2)*.

Frawley, W. (1984). Prolegomenon to a theory of translation. *Translation. Literary, Linguistic and Philosophical Perspectives*. University of Delaware Press.

García, O., & Li Wei. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.

Halverson, S. L. (2017). Gravitational pull in translation. Testing a revised model. *Empirical Translation Studies. New Methodological and Theoretical Traditions.* De Gruyter Mouton.

Holmes, P., & Dervin, F. (2016). Introduction–English as a lingua franca and interculturality: Beyond orthodoxies. *The cultural and intercultural dimensions of English as a lingua franca*, 1-30. Multilingual Matters.

Kotze, H. (2022). Translation as constrained communication: Principles, concepts and methods. *Extending the Scope of Corpus-Based Translation Studies*. Bloomsbury Academic.

Lanstyák, I., Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures* 13(1), 99-121.

Love, N. (2017). On languaging and languages. *Language Sciences* 61,113-147.

Woldoff, R.A., Litchfield, R.C., (2021). D*igital Nomads: In Search of Meaningful Work in the New Economy.* Oxford Academic.

Seidlhofer, B. (2007). Common Property: English as a Lingua Franca in Europe. In: Cummins, J., Davison, C. (eds) *International Handbook of English Language Teaching.* Springer International Handbooks of Education, vol 15. Springer.

Szmercsányi, B., & Kortmann, B. (2009). *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. Routledge.

Comparing L2 writing and translation: Insights from a corpus-based study of learner texts

Stella Neumann, Marie-Aude Lefer, Zoë Miljanović & Celina Brost

(RWTH Aachen University, UCLouvain, RWTH Aachen University & RWTH Aachen University)

Keywords: language mediation, L2 writing, translation, language contact, corpus research

Language mediation such as professional translation and interpreting, user-generated translation, mediation in foreign language teaching etc., is wide-spread in many speech communities. This also includes translation from the L1, the translator's dominant language, into the L2, their second most used language. Does this bilingual text production mode diverge in linguistic choice from other modes, particularly from L2 writing from scratch? Or are these modes alike due to the simultaneous activation of two languages as a shared constraint during production (cf. Kotze & Van Rooy, 2024)? While writing in the L2 involves some activation of the L1, especially for L2 learners, it is not the immediate re-encoding of an anterior text in another language (Hansen-Schirra & Steiner, 2012), where the language user needs to fully activate both languages while switching between processing the source language text and producing the translation in the target language. By contrast, L2 writing from scratch does not necessarily draw on other texts. Previous studies suggest that there is a subtle, but significant difference between linguistic choices in the different production modes (e.g. Rabinovich et al., 2016; Neumann et al., 2024). However, often data sets are not entirely comparable (e.g. in register), and/or L2 writing is compared with translation into the L1 rather than the L2.

This paper reports on a corpus study of texts by 37 advanced learners of English (L1 German) who wrote a review about a cultural experience, such as a concert or film, from scratch in their L2 English and translated a German review into English. A corpus of comparable L1 English reviews is included as a benchmark (Taboada & Grieve, 2004). We analyse the frequency of infinitives, a feature significantly more frequent in English than in German (Mahler 2024). If the presence of an anterior text in the mediation condition affects linguistic choice, the use of infinitives should differ in the three text production modes (L2 writing, L2 translation and L1 writing) in English under the varying influence of the source or first language, which is expected to be stronger in L2 translation than in L2 writing.

A linear regression model regressing the normalised frequency of infinitives on production mode and the additional predictors of sentence length and topic only retrieved associations with the latter two predictors. A second linear model without the L1 texts but with two additional predictors capturing the results of LexTALE proficiency tests in the L1 and the L2 (Lemhöfer & Broersma, 2012) for our participants yielded a marginally significant negative association between the frequency of infinitives and L2 translation. Higher proficiency in the L2 was also negatively associated with the frequency of infinitives.

These results suggest that the ability to adapt to L2 usage is influenced by proficiency and a potential transfer effect facilitated by the anterior text. Thus, mediation appears to play a marginal role. Drawing on a highly comparable dataset, this finding adds to the critical perspective that questions the exceptionalism of translation (Halverson & Kotze, 2022) in contrast to other (bilingual) text production modes.

References

Halverson, S. L., and Kotze, H. (2022), Sociocognitive Constructs in Translation and Interpreting Studies (TIS): Do We Really Need Concepts Like Norms and Risk When We Have a Comprehensive Usage-Based Theory of Language?, in S. L. Halverson, and Á. M. García (Eds.), *Contesting Epistemologies in Cognitive Translation and Interpreting Studies*, Routledge, 51–79.

Hansen-Schirra, S., and Steiner, E. (2012), Towards a typology of translation properties, in S. Hansen-Schirra, S. Neumann, and E. Steiner, *Cross-Linguistic Corpora for the Study of Translations*, DE GRUYTER, 255–280. https://doi.org/10.1515/9783110260328.255

Kotze, H., and Van Rooy, B. (2024), Introduction: The constrained communication framework for studying contact-influenced varieties, in B. Van Rooy, and H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings*, Benjamins, 1-28.

Lemhöfer, K. and Broersma, M. (2012), Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English, *Behavior Research Methods* 44(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Mahler, H. (2024), *Contrastive Grammar in Use: Quantitative Perspectives on the Verb Phrase in English and German* [PhD dissertation], Albert-Ludwigs-Universität Freiburg.

Neumann, S., Kerz, E., and Heilmann, A. (2024), Comparing contact effects in translation and second language learning, in B. Van Rooy, and H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings,* Benjamins, 223–254.

Rabinovich, E., Nisioi, S., Ordan, N., and Wintner, S. (2016), On the Similarities Between Native, Non-native and Translated Texts, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1870–1881. https://doi.org/10.18653/v1/P16-1176

Taboada, M., and Grieve, J. (2004), Analyzing Appraisal Automatically, *Exploring Attitude and Affect in Text: Theories and Applications,* American Association for Artificial Intelligence, Stanford.

Translation choices and their variability:
The role of network structure and entrenchment

Thomas Prinzie
(UCLouvain, Belgium)

Keywords: Translation choice; Translation variability; Salience; Entrenchment; Noun sequence

Halverson's (2003, 2017) Gravitational Pull hypothesis integrates insights from cognitive linguistics and bilingualism (Langacker 2008, Diessel 2019, and De Groot 2011) to provide a usage-based theoretical model for translational behavior. Given the notion that translators' bilingual knowledge is organized in networks whose structure is determined by the number of nodes and the salience of intra- and cross-linguistic connections, Halverson posits that the more entrenched connections (which are more readily activated) may attract translators to particular linguistic choices. Subsequent work has considered aspects of salience besides entrenchment, including recency and cross-linguistic similarity (Halverson 2024, Lefer & De Sutter 2022, and Heilmann et al. 2022).

My presentation discusses how this model helps to explain translator choices in French translations of English [N+N] sequences, e.g. *research firm → société de recherche, entreprise dédiée à la recherche, agence de notation* (among other solutions). These items pose particular difficulties (Lefer & De Clerck 2021) due to cross-linguistic formal differences, potentially ambiguous semantic relationships, and the availability of near-synonyms in French. These difficulties can lead translators to different translation solutions. Following Malmkjær's suggestion that "investigation might prove particularly fruitful" where translator's "varied realisations" differ most (Malmkjaer 1998: 539), I focus on *translation variability* as a phenomenon of interest in its own right, taking advantage of the recent Multilingual Student Translation corpus (MUST; Granger & Lefer 2020), which contains multiple translations for each source text. From 10 source texts (specialized in sustainable finance) and 168 translations, I have identified 117 noun sequences with 4,444 French translations.

The main research question concerns how translation choices and their variability are influenced by network structure, i.e. by the number and relative salience of different connections. I combine multiple operationalizations to examine the role of entrenchment. Monolingual and bilingual reference corpora provide frequency data for general and specialized language. As simple token frequencies do not capture entrenchment well, my analyses also consider dispersion, association strength, type frequencies (morphological family size), and surprisal and entropy (Gries 2022). These corpus data are complemented with smaller-scale, preliminary data drawn from elicitation and judgment experiments to test for individual entrenchment of [N+N] combinations, using Psytoolkit (Stoet 2017) to collect responses and reaction times. To date, I have collected such data for 4 participants (students who also contributed to the MUST corpus). This combination of corpus and experimental methods reflects a growing trend (Gilquin & Gries 2009, Dąbrowska 2016, De Sutter & Lefer 2020, and Serbina & Neumann 2021) and

obtains both collective and individual evidence of usage (cf. Schmid's (2020) Entrenchment and Conventionalization model).

My talk presents qualitative and preliminary quantitative results. Descriptions of multiple translations illustrate how semasiological and onomasiological patterns combine to produce the observed variability. The quantitative analyses are expected to confirm that the number and distribution of translation solutions will be congruent with the number and relative salience of connections within and between the source and target languages.

References

Dąbrowska, Ewa. 2016. Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics* 27(4). 479–491. https://doi.org/10.1515/cog-2016-0059.

De Groot, Annette M. B. 2011. *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York: Psychology Press. https://doi.org/10.4324/9780203841228.

De Sutter, Gert & Marie-Aude Lefer. 2020. On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28(1). 1–23. https://doi.org/10.1080/0907676X.2019.1611891.

Diessel, Holger. 2019. *The Grammar Network*. Cambridge University Press.

Gilquin, Gaëtanelle & Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5(1). 1–26. https://doi.org/10.1515/CLLT.2009.001.

Granger, Sylviane & Marie-Aude Lefer. 2020. The Multilingual Student Translation corpus: a resource for translation teaching and research. *Language Resources and Evaluation* 54(4). 1183–1199. https://doi.org/10.1007/s10579-020-09485-6.

Gries, Stefan Th. 2022. On, or against?, (just) frequency. In Hans C. Boas (ed.), *Directions for Pedagogical Construction Grammar*, 47–72. De Gruyter Mouton. https://doi.org/10.1515/9783110746723-002.

Halverson, Sandra L. 2003. The cognitive basis of translation universals. *Target* 15(2). 197–241. https://doi.org/10.1075/target.15.2.02hal.

Halverson, Sandra L. 2017. Gravitational pull in translation: Testing a revised model. In Gert De Sutter, Marie-Aude Lefer & Isabelle Delaere (eds.), *Empirical translation studies: New methodological and theoretical traditions*, 9–45. De Gruyter Mouton.

Halverson, Sandra L. 2024. The gravitational pull hypothesis: Status and prospects. In Josep Marco & Isabel Tello (eds.), *Towards an empirical verification of the gravitational pull hypothesis: evidence from the COVALT corpus*, 21–52. Berlin: Peter Lang. https://www.peterlang.com/document/1340141.

Heilmann, Arndt, Tatiana Serbina, Jonas Freiwald & Stella Neumann. 2022. Grammatical metaphor in translation: A corpus-based investigation of nominal of-constructions. In Sylviane Granger & Marie-Aude Lefer (eds.), *Extending the scope of corpus-based translation studies*, 158–181. London: Bloomsbury Publishing.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Lefer, Marie-Aude & Marie De Clerck. 2021. L'apport des corpus intermodaux en lexicologie contrastive : Étude comparative de la traduction écrite et de l'interprétation simultanée des séquences de noms. In Sylvie Hanote & Raluca Nita (eds.), *Morphophonologie, lexicologie et langue de spécialité: hommage à Jean-Louis Duchet et à Michel Paillard*, 145–162. Rennes: Presses universitaires de Rennes.

Lefer, Marie-Aude & Gert De Sutter. 2022. Using the Gravitational Pull Hypothesis to explain patterns in interpreting and translation: The case of concatenated nouns in mediated European Parliament discourse. In Marta Kajzer-Wietrzny, Adriano Ferraresi, Ilmari Ivaska & Silvia Bernardini (eds.), *Mediated discourse at the European Parliament: Empirical investigations*, 133–159. Language Science Press. https://doi.org/10.5281/zenodo.6977046.

Malmkjær, Kirsten. 1998. Love thy neighbour: Will parallel corpora endear linguists to translators? *Meta* 43(4). 534–541. https://doi.org/10.7202/003545ar.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford University Press.

Serbina, Tatiana & Stella Neumann. 2021. Translation Product and Process Data: A Happy Marriage or Worlds Apart? In Sandra L. Halverson & Álvaro Marín García (eds.), *Contesting Epistemologies in Cognitive Translation and Interpreting Studies*, 131–152. Routledge.

Stoet, Gijsbert. 2017. PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology* 44(1). 24–31. https://doi.org/10.1177/0098628316677643.

WS26 Wordhood in West African languages and beyond

Neige Rochant & Andrey Shluinsky

Incorporation, formal incorporation and "wordhood": the case of Mande (with special reference to Gban)

Alexander Zheltov

(St. Petersburg University)

In some Mande languages we can observe that the complex Spro-OBJ-V is syntactically closer than other syntactic constituents (nominal subject, oblique object, adjuncts). In some languages it can lead to the object incorporation into a verb. There are at least three Mande languages where object incorporation was attested: Mandinka (Creissels 2008; Creissels & Sambou 2013), Tigemaxo (Blecke 1996; 2011), Soninke (Creissels & Drame 2018). Analyzing the data concerning object + verb incorporation, we can see that in all three cases, we observe the "pure" incorporation.

For instance, in Tigemaxo (Blecke 1996: 88–9; Blecke 2011: 1), we can observe both incorporated (1) / non-incorporated (2) pairing and phonological indicators of compounding:

Tigemaxo (Blecke 2011: 1)

(1) *n ga xulu-**mɛnɛ***
1SG IPFV pirogue-look for[**INTR**]
'I "pirogue-look for".'
(2) *n ga xulu **mana***
1SG IPFV pirogue look for[**TR**]
'I look for a pirogue.'

So, we can see the evident distinction between one word VO incorporation (1) and two separate words for V and O (2).

In some other Mande languages various formal means to join pronominal subject, object, and the verb are attested: tonal assimilation, contractive (portmanteau) forms (e.g., in Gban: PROSBJ+TAM, PROSBJ+PROOBJ, PROSUBJ+TAM +PROOBJ). Nominal subjects, oblique objects, and adjuncts are never involved in such processes. These processes do not form a single incorporated word as in (1) but make the degree of syntactic independence of the constituencies of VP evidently lower than that of nominal subjects, oblique objects, adjuncts. If we consider the latter as separate lexemes, we should use a different term for the former ones, and it should be neither the same as "separate lexemes" nor as "object incorporating verb". They show the phenomenon that can be called "formal incorporation".

In the case of a pronominal object, we can find the full chain of boundedness (PROSBJ-TAM-PROOBJ-V).

(3) Gban (personal data)

Mà b̀ɛ
1SG.PRS.3SG take.PRS
'I take him.'

In this short example, we can see many grammatical elements "glued" together: *ma = ĩ* (1SG) + low tone + *a* (3SG.NSBJ); *b̀ɛ* = 'take' + low tone (the first syllable of the verb in the Present tense always copies the tone of the closest syllable of adjacent NP): a portmanteau pronoun as in (3) for a pronominal object, the last syllable of the nominal object NP (4), or a pronominal subject for intransitive verbs (5).

(4) Gban (personal data)

*Mű ɛ̀ bl**ú** b**ɛ́***
person 3SG.PRS bread take.PRS
'The person takes the bread.'
(5) Gban (personal data)
Ĩ̀ tà 1SG.PRS go.PRS
'I go.'

These examples show tonal progressive assimilation of the verb (assimilating to the tone of the closest syllable), confirming the closer link between object/pronominal subject and the verb than it is attested between the verb and nominal subject NP, etc.).

Summing up, we demonstrate the closer connection between pronominal subject, TAM marker (which is often tonal), pronominal object (all of them forming PPM) and the verb, on the one hand (examples (3) and (5)), and nominal object NP and the verb (4) on the other hand.

So, we can see that the notion of the "wordness" for Mande languages should be more complicated than the binary opposition "one word"/"separate words" having Gban-like complex Spro-OBJ-V as something in between.

References

Blecke, Thomas (2011), L'incorporation en tigemaxo (bozo). A presentation at the *3rd International Conference on Mande Languages & Linguistics*, Paris, 14–17 September 2011. http://llacan.vjf.cnrs.fr/fi chiers/mande2011/bublio/blecke.pdf (accessed 2020-06-28.)

Blecke, Thomas (1996), *Lexikalische Kategorien und grammatische Strukturen im Tigemaxo (Bozo, Mande)*. Koln: Koppe Verlag.

Creissels, Denis (2008), L'incorporation en mandinka. In Amiot, Dany (ed.), *La composition dans une perspective typologique*, 75–88. Lille: Artois Presses Universite. (Cited based on preprint version at
https://www.researchgate.net/profile/Denis_Creissels/publication/254487504_L'incorporation_en_mandinka/links/562499be08aed8dd1948cde7/Lincorporation-en-mandinka.pdf, accessed 2020-10-30.)

Creissels, Denis & Sambou, Pierre (2013), *Le mandinka: phonologie, grammaire, textes.* Paris: Karthala.

Creissels, Denis & Drame, Djibril (2018), Transitivity and incorporation in Soninke. In Anyanwu, Rose-Juliet (ed.), *Transitivity in African languages*, 37–54. (Frankfurter Afrikanistische Blatter 26.) (Cited based on preprint version at https://www.academia.edu/44174953/Transitivity_and_incorporation_in_Soninke_West_Mande, accessed 2020-10-30.)

Hesitations as a wordhood criterion: Evidence from Akebu

Andrey Shluinsky

(Humboldt University of Berlin)

Keywords: wordhood, hesitation, pauses, Akebu, West African

Akebu (< Ka-Togo < (Ghana-Togo Mountain) < Kwa) is an example of a language where the concept of word is not clear even in a working intuitive form. Such known criteria as vowel harmony (Makeeva & Kuznetsova 2022), tonal rules (Makeeva 2024), reduction (Shluinsky 2020) cut out significantly different domains. Other criteria, such as uninterruptibility, cannot be applied to its data because of the rigid word order. At the same time, Akebu has complex constructions, such as in particular noun-attribute complexes (Shluinsky 2022), which make the issue even more complicated.

One of the most traditional criteria of wordhood is related to potential pauses around a word. While it is hard to test it directly e.g. in elicitation sessions, spontaneous discourse data allow to go through hesitations attested there. While hesitation phenomena are reported to happen marginally within the words, such occurences are very rare (as shown both in old studies such Maclay & Osgood 1959 and in more recent work, such as Podlesskaya 2015 or Gósy et al. 2023).

Based on a sample of 4,5 hrs of recorded and transcribed Akebu texts, I have studied instances of attested hesitations of several formal types, namely (i) silent pauses, (ii) non-phonemic lengthening, (iii) filled pauses (with a non-lexical vocalic sound like ǝː or ɔ̃ː etc.) and (iv) the segmental hesitation marker *àmàmà ~ àmà*. While predictably most hesitations occur between items that are clear distinct words in any reasonable sense, there is a statistically significant difference between morpheme boundaries which are regularly separated by hesitations and those where hesitations happen extremely marginally if at all.

In particular, the following results are worth mentioning. First, postpositive grammatical markers are almost never attested to be separated by hesitations from their host lexical items they are phonologically and/or grammatically connected to. This fact is in line with cross-linguistic tendencies (see Himmelmann 2014), but importantly no potential Akebu suffixes are autonomous according to the criterion in focus, except for object personal pronouns. Second, prepositive grammatical markers behave differently. While some of them are only very marginally attested to be separated by hesitations and only by silent pauses (such as nominal noun class prefixes), others do it regularly. Third, hesitations are regularly attested to follow some grammatical markers that are subject to vowel harmony with the subsequent host stem. Sometimes, but not always such markers still harmonized, like the 3rd person 1st noun class possessive marker ˋ*nv́* in the example (1). Fourth, some grammatical elements are attested to be separated from notional items on both sides. In particular, the possessive marker *lv́* can be both preceded by a hesitation following the possessor NP and followed by a hesitation preceding the possessee NP. Finally, compounds of all types, including noun-attribute complexes are extremely rarely separated by hesitations, the example (2) being very marginal.

(1) ˋ*ná*   [0.19s]   *à-fɛ̀ɛ̀-kālà-yā*ˋ   'her sewn parts'
     CL1.POSS   CL5-place-sew-CL5

(2) ˋ*né*   ˋ*bʈíkísì*[0.20s]-*mūŋmūŋ-yā*ˋ   *sā*   'these big bricks of hers'
     CL1.POSS   brick-big-CL1   DEM

References

Gósy, Mária (2023), Occurrences and durations of filled pauses in relation to words and silent pauses in spontaneous speech, *Languages* 8, 79.

Himmelmann, Nikolaus P (2014), Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference, *Language* 90(4), 927-960.

Makeeva, Nadezhda, and Natalia Kuznetsova (2022), Phonological and acoustic properties of ATR in the vowel system of Akebu (Kwa), *Phonology* 39(4), 641-678.

Maclay, Howard, and Charles E. Osgood (1959) Hesitation phenomena in spontaneous English speech, *Word* 15(1), 19-44.

Makeeva, Nadezhda (2024), Sistema kombinatornyx tonal'nyx pravil jazyka akebu (kva) [System of combinatory tonal rules in Akebu (Kwa)], in N. Gromova, and M. Urb (eds), *Kul'turno-jazykovye processy v Afrike*, Moscow: Ključ-S, 71-79.

Podlesskaya, Vera (2015), A corpus-based study of self-repairs in Russian spoken monologues, *Russian Linguistics* 39(1), 63-79.

Shluinsky, Andrey (2020), Reduction of noun class suffixes in Akebu, *Language in Africa* 1(2), 65-88.

Shluinsky, Andrey (2022), Incorporation as a nominal attribute strategy in Akebu, *Language Sciences* 93, 1-19.

Where length meets Type : Wordhood and wordclass in Fe'fe' [970]

Yosseu Anne

(HHU Düsseldorf)

Keywords : Fe'fe', Wordhood, root, Aspect,serial verb

The question of wordhood and word boundaries in the preverbal complex of Fe'fe' (fmp, Grassfields, 970 in Bikoï *et al.* 2012) is briefly discussed by Ngangoum (2015). She notes that tense and aspect markers are generally written as one orthographic word unit with the verb root in literacy books, as in the following example (Ngangoum, 2015, 47, slightly adjusted):

(1) *Siani fhú-mȧ -ŋgɛ́ ntɛ́ ɛ́*
    Siani P1-PROG-go market
    "Siani was going to the market."

However, Ngangoum (2015) argues for a word boundary between the preverbal markers and the verb root, because it is possible to insert full lexemes in between, such as *mbȧ* "again" in the following example (Ngangoum, 2015, 47):

(2) *Siani fhú vúsí   mȧ   mbȧ   ŋgɛ́ ntɛ́ ɛ́*
    Siani P1 hurry PROG again go   market
    "Siani was hurrying back to the market."

This argument hinges on the analysis of *mbȧ* "again" as an adverb, but *mbȧ* is also a verb with the meaning "return", as shown in Ngangoum
(2015, 48):

(3) *Siani mȧmbȧ   njȧm*
    Siani PROG return back
    "Siani is coming back."

This means that in 2, both *vúsí* "hurry" and *mbȧ* may best be analysed as serial verbs, and there may not be a strong word boundary between them and the tense and aspect markers preceding them. For this talk, I therefore discuss the plausibility of the following assignment of word boundaries, as indicated by blank spaces:

(4) *Siani fhú-vúsí mɑ̀ -mbɑ́   ŋgɛ́ ntɛ̀ ɛ̀*
       Siani P1-hurry PROG-return go   market
       "Siani was hurrying back to the market."

   I will present new evidence addressing the wordhood properties of TAM markers and verb roots in Fe'fe', relative to commonly proposed criteria (Dixon & Aikhenvald, 2003; Haspelmath, 2017; Zingler, 2020; Tallman, 2020). I will introduce my preliminary work on serial verb constructions in the language (in the extend relevant to the current topic here), which have not been previously explored in the literature.

   The empirical basis of my research is a small, oral corpus I have collected during fieldwork in 2024, along with elicitations with speakers. Interpretations are informed by my own intuitions as a native speaker. It is fair to say that a conclusive analysis of Fe'fe' Wordhood will be the aim of the presentend talk. Due to the Absence of a substantial amount of data, caution was exercised while interpreting the results.

References

Bikoï, C.B., Ndongo-Semengue, M.A.B., Tourneux, A., Tourneux, H., & Centre régional de recherche et de documentation sur les traditions orales et pour le développement des langues africaines (Yaoundé, Cameroon). 2012. *Atlas linguistique du cameroun (alcam):*. Atlas linguistique de l'Afrique centrale. Editions du CERDOTOLA.

Dixon, Robert MW, & Aikhenvald, Alexandra Y. 2003. Word: a typological framework. *Pages 1–41 of:* Dixon, Robert MW, & Aikhenvald, Alexandra Y (eds), *Word: A cross-linguistic typology*. Cambridge University Press.

Haspelmath, Martin. 2017. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia linguistica*, **51**(s1000), 31– 80.

Ngangoum, Emilienne. 2015. *When synchrony meets diachrony*. Ph.D. thesis.

Tallman, Adam JR. 2020. Beyond grammatical and phonological words. *Language and linguistics compass*, **14**(2), e12364.

Zingler, Tim. 2020. *Wordhood issues: Typology and grammaticalization*. Ph.D. thesis, The University of New Mexico.

Word spacing and wordhood in Irish and beyond

Cormac Anderson

(University of Surrey)

Keywords:

This paper examines notions of the word in the Irish grammatical tradition and the relevance of these to debates on wordhood in contemporary linguistic typology.

The practice of inserting spaces between words when using Latin script was innovated by Irish scribes in the seventh century (Saenger 1997). However, their use of spacing differed depending on whether they were writing Latin or Old Irish (Bronner et al. 2018). In Latin (1a), the word was taken to correspond to a distinct part of speech, as codified in the *Ars Minor* of Donatus. In Old Irish (1b), it was understood to comprise a larger stress group, including an autosemantic element bearing primary stress and optionally also unstressed affixes. This analysis of the Old Irish word is explicit in Irish medieval grammar (Bergin ed. 1916-1955, Adams 1970, Ahlqvist 1974), which breaks from the classical model and asserts the superiority of the vernacular (Calder ed. 1917), reflecting an intellectual confidence born of political autonomy.

With the collapse of the Irish social and cultural order in the wake of colonisation (from 1600CE), the locus of Irish grammatical activity shifted to mainland Europe. In this context, Irish came to be written according to the Latin model. Modern Irish thus uses different principles of word segmentation to Old (and Classical) Irish, although the various stages of the language are structurally similar.

These different notions of what constitutes a word have obvious consequences for language comparison. For example, analyses of phenomena such as consonant mutation, which occur both in West Africa and in Celtic languages such as Irish (Ternes 1990), are dependent on the broader analysis of wordhood employed (Iosad 2010).

(1)
(a)
cum populis   (Reeves ed. 1857: 23)
with people.ABL
'with the people'

(b)
du-m-phopul-sa   (Stokes and Strachan 1901 eds.: 263, i.e. Ml77a13)
to-1SG.POSS-people.DAT-1SG.FOC
'to my people'

References

Adams, G. (1970). Grammatical analysis and terminology in the Irish bardic schools. *Folia Linguistica*, *4*(1-2), 157-166. https://doi.org/10.1515/flin.1970.4.1-2.157

Ahlqvist, A. (1974). Notes on 'Case' and Word-Boundaries. *Ériu* 25: 181-189

Bergin, O. ed. (1916-1955). Irish grammatical tracts [I-V]. *Ériu* 8, 9, 10, 14, 17.

Bronner, D., Busch, N., Fleischer, J. & Poppe, E. (2018). (Non-)separation of words in early medieval Irish and German manuscripts and the concept "word". In C. Ulbrich, A. Werth & R. Wiese (Ed.), *Empirical Approaches to the Phonological Structure of Words* (pp. 45-70). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110542899-003

Calder, G. ed. (1917). *Auraicept na n-Éces: The scholars' primer, being the texts of the Ogham tract from the Book of Ballymote and the Yellow book of Lecan, and the text of the Trefhocul from the Book of Leinster*. Edinburgh: John Grant.

Iosad, P. (2010). Right at the left edge: initial consonant mutations in the languages of the world. In J. Wohlgemuth & M. Cysouw (Ed.), *Rethinking Universals: How Rarities affect Linguistic Theory* (pp. 105-138). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110220933.105

Reeves, W. ed. (1857). *The Life of Columba, written by Adamnan*. Dublin.

Stokes, W. and Strachan J. eds. (1901). *Thesaurus Palaeohibernicus : a collection of Old-Irish glosses, Scholia prose and verse: Vol. 1*. Cambridge: Cambridge University Press.

Ternes, E. (1990). Initial mutations in Celtic and West African languages: synchrony and diachrony. *Afrika und Übersee* 73: 3–17.

Vowel harmony and other wordhood criteria in Jóola Fóoñi (Atlantic)

Denis Creissels
University of Lyon

The main reference descriptions of Jóola Fóoñi (Sapir (1965), Sambou (1983) and Hopkins (1995)) agree that the ten vowels of Jóola Fóoñi divide into two subsets of five vowels each (variously characterized as tense/lax or +ATR/-ATR), and that underlyingly +ATR vowels may spread their +ATR feature to neighboring vowels that are underlyingly -ATR (vowel harmony). As regards the domain of vowel harmony, Sambou (1983: 79) simply states that "within the limits of the word, a lax vowel becomes tense in the presence of a tense vowel". Both Sapir (1965: 12) and Hopkins (1995: 19-20) acknowledge that the situation is more complex, without, however, entering into the details. In fact, there are several morphemes that are uncontroversial verbal suffixes according to all the other possible wordhood criteria, and nevertheless do not undergo the vowel harmony rule. The clearest case is that of the reduplicative suffix that marks positive assertion in two tenses: completive positive, as in *nɪ-ja-jaw* /1SG-go-RDPL/ 'I went' and habitual positive *nɪ-jaw-ɛ-jaw* /1SG-go-HAB-RDPL/ 'I usually go'.

The suffixal nature of this reduplicative suffix follows from the fact that it can only be separated from the root by a small number of morphemes, which are all bona fide suffixes: the past marker, the venitive marker, the inclusive marker (if the subject is first person plural), and one or two human object indexes (depending on the valency properties of the verb).

Crucially, the reduplicative suffix interacts phonologically with the preceding formative in a very specific way. For example, the reduction of CVVC syllables to CV (as for example with the stem *raaf* 'suckle', which becomes *ra* in *na-ra-raaf* /3SG:clA-suckle-RDPL/ 'he suckled', occurs in no other possible configuration.

By contrast, when the reduplicative suffix reduplicates an inherently -ATR stem, such as *wɔnk* 'call', it does not undergo vowel harmony. For example, 'he called us', underlyingly *na-wɔnk-oli-wɔnk* /3SG:clA-call-1PL.EXCL-RDPL/, is realized *ne-wonk-oli-wɔnk*, which means that the +ATR feature of *oli* spreads to the preceding formatives (stem and prefix), but not to the reduplicative suffix. Consequently, the reduplicative suffix, whose suffixal behavior is uncontroversial as regards segmental processes, does not participate in a suprasegmental phenomenon traditionally presented as providing an unproblematic wordhood criterion in Jóola languages.

The historical explanation is that the reduplicative suffix marking positive assertion in two of the tenses that constitute the inflection of Jóola Fóoñi verbs probably results from the relatively recent grammaticalization of a cognate object in a "verb + cognate object" construction whose function was to highlight the lexical meaning of the verb. Presumably, the vowel harmony rule had already ceased to be an active phonological rule at the time when this construction underwent univerbation, but this did not prevent the former cognate object converted into a formative of the verbal word to develop a typical suffixal behavior in other respects.

Hopkins, Bradley L. 1995. *Contribution à une étude de la syntaxe diola-fogny*. Dakar: Société Internationale de Linguistique.

Sambou, Pierre-Marie. 1983. *Lexique diola fogny - français*. Dakar : CLAD.

Sapir, J. David. 1965. *A grammar of Diola-Fogny*. Cambridge University Press.

Stranger things of word (dis)integrity in Southern Mande

Dmitry Idiatov

(LLACAN – CNRS/INALCO/EPHE)

Keywords: endoclitics, Mande, phonotactics, reanalysis, word integrity

In several Southern Mande languages, the integrity of words can be violated by inserting other bona fide words inside them. This can happen to morphologically complex words [Root+Suffix] and morphologically simplex words [Root]. The Southern Mande language that appears to be most prolific in this respect is Tura [glottocode: tour1242] (Bearth 1971:171-176; Idiatov 2005, 2008). Thus, in (1) the morphologically simplex numeral *pì̋lɛ́* 'two' is split by the insertion of the restrictor *lèfʉ̀* 'even'. In (2), the morphologically complex verb *dɔ́-lɔ̋* 'stop' (from *dɔ́* 'stand' plus the deobliquative suffix *-lɔ̋*) is split by the insertion of the modifying expression *dɛ́ɛ́ ké* 'again' (lit. 'a certain new [X]'). In (3), *dɛ́ɛ́ ké* 'again' breaks up the morphologically simplex verb *gbɔ̋lɔ̋* 'thunder'. This kind of violation of word integrity is typologically extremely unusual as such. However, in addition to that, as argued by Idiatov (2005, 2008), in all these cases the second part of the original word actually functions as the syntactic head governing the first part of the original word, and this irrespective of whether it is an actual morpheme itself (2) or just a meaningless part of the word (1, 3). In this construction, the first part is the complement of the second part which confers on the first part the combinatorial possibilities of a noun, such as the possibility to be modified by adnominal modifiers, while the respective word as a whole is not capable of that on its own. Kushnir (2016) and Fedotov (2014) report similar but more restricted cases of violation of word integrity by other words for Yaure [glottocode:yaou1238] and Gban [glottocode:gagu1242], respectively.

In this paper, I present the relevant data in more detail. I further elaborate on the concepts of *quasi-words* (or *pseudowords*) proposed by Idiatov (2005) to refer to the parts of such split word constructions. I also situate these stranger cases of violation of word integrity of Southern Mande in a typological perspective by contrasting them with comparable word integrity violation phenomena in other languages, such as endoclisis in Udi (Harris 2002) and expletive insertion in English, recently argued by Zingler (2024) to be an example of discontinuous compounding. Finally, I discuss the hypotheses proposed for the diachrony of these word integrity violation phenomena and further elaborate on them by highlighting the similar patterns of reanalysis and the particular relevance of the typical word phonotactic templates found in the relevant languages (variously referred to in the literature as *syllabemes*, *syllabomorphemes*, *canonical morphemes* and *(featural) feet*).

(1) Wǎá   pì̋ lèfʉ̀ lɛ́
     3PL.SBJ.NEG.COP two₁ even two₂
     'They are not even two/ they are not two at all [but just one].' (Idiatov 2005:32)

(2) Ɔ̀ó   dɔ́   dɛ́ɛ́ ké   lɔ́-ɔ́
     3SG.SBJ.NEG.PFV stand₁ new a.certain stand₂\PFV-PFV
     'He did not stop again.' (Idiatov 2008:162)

(3) Lá=á   gbɔ̋   dɛ́ɛ́ ké   lá-á
     rain=AUX thunder₁ new a.certain thunder₂\PFV-PFV
     'It thundered again.' (Idiatov 2008:164)

References:

Bearth, Thomas. 1971. *L'énoncé toura (Côte d'Ivoire)*. Norman: Summer Institute of Linguistics, University of Oklahoma.

Fedotov, Maxim. 2014. Фокус, заглагольные зависимые и эндоклитики в гбан. In Valentin Vydrine & Natalia Kuznetsova (eds.), *От Бикина до Бамбалюмы, из варяг в греки. Экспедиционные этюды в честь Елены Всеволодовны Перехвальской*, 195–214. Saint-Petersburg: Nestor-Istorija.

Harris, Alice. 2002. *Endoclitics and the origins of Udi morphosyntax*. Oxford: Oxford University Press.

Idiatov, Dmitry. 2005. The exceptional morphology of Tura numerals and restrictors: endoclitics, infixes and pseudowords. *Journal of African Languages and Linguistics* 26(1). 31–78. doi: 10.1515/jall.2005.26.1.31.

Idiatov, Dmitry. 2008. Antigrammaticalization, antimorphologization and the case of Tura. In Elena Seoane, María José López-Couso & (in collaboration with) Teresa Fanego (eds.), *Theoretical and empirical issues in grammaticalization*, 151–169. Amsterdam: John Benjamins. doi: 10.1075/tsl.77.09idi.

Kushnir, Elizaveta L. 2016. Особенности топикализации в языке яурэ (южные манде). In Dmitry V. Gerasimov (ed.), *Acta linguistica Petropolitana. Труды Института лингвистических исследований РАН*, vol. 1, 106–112. Saint-Petersburg: Nauka.

Zingler, Tim. 2024. Expletive insertion: a morphological approach. *English Language & Linguistics* 28(1). 23–42. doi: 10.1017/S1360674323000254.

Mbam wordhood and Niger-Congo morphosyntactic reconstruction

Elisabeth J. Kerr
(Ghent University)

Keywords: wordhood, morphosyntax, Niger-Congo, West/Central Africa, historical syntax

Introduction. It is often noted that Niger-Congo languages show a large degree of variation in their morphosyntactic profiles, ranging from highly analytic patterns (as found across West Africa, e.g. Kwa) to highly synthetic patterns (as seen in Eastern and Southern Bantu). It is therefore clear that there has been analytic↔synthetic morphosyntactic change within the phylum, although the exact processes remain debated. The hypothesis proposed by Hyman (2004, 2017, a.o.) is that the variation can be understood through a synthetic→analytic direction of change, with languages varying in the degree of breakdown of the verbal word (cf. Güldemann 2022 for an alternative proposal). Here, the key boundary is between the inflectional material and verb root (to which derivational material may be attached as suffixes) (Hyman 2008), as in (1).

(1) INFL-VERB-DER →　　INFL#VERB-DER

An evergreen methodological issue in being able to (i) model how prefixes become independent words and (ii) situate languages within this scenario of change is the validity of the wordhood demarcations made in the available sources. As Nurse (2008:169-170) notes for Benue-Congo languages of Cameroon/Nigeria, authors differ in whether STAMP morphemes and verbs are written together or separately, even for the same language, a situation significantly influenced by anglophone versus francophone research traditions and thus not necessarily reflecting actual differences in linguistic systems. For example, in (2) the Tunen ([tvu], Cameroon) tense marker and verb stem form one orthographic word, while the missionary orthography in (3) writes them separately.[1]

(2) a **nákan**　　ebàk' ōmbɛl.
　　 SM.1　 PST.leave　 7.lizard 3.house
　　 'He went to the lizard's house.'　　　　　　　　 (Tunen; Dugast 1975:61, adapted)

(3) Yowánɛsɛ　　 a　　 **nɔ́**　 **akána** u　　 nioní.
　　 1.Jean　　　 SM.1　 PST　 leave　 PREP　 5.market
　　 'Jean went to the market.'　　　　　　　 (Tunen; Sartre et al. 2008:11, adapted)

Approach. This paper therefore uses new wordhood tests to control for inter-source variation in word boundaries by collecting comparable data across closely-related languages, to be used for bottom-up reconstruction. **Method**. I present an in-situ fieldwork study on the 4 languages of the Western Mbam subgroup of Cameroonian Bantu/Bantoid languages, namely Tunen [tvu], Nyokon [nvo], Nomaandé [lem], and Atomb [ttf]. These languages spoken close to the Bantu homeland are significant as they exemplify intermediate stages of (1). I devise phonological and morphosyntactic wordhood tests for a boundary between the inflectional material and verb stem. **Data.** The tests cover properties including syntactic interruptability and pre-stem ATR harmony.

Results. The Mbam test results show intermediate stages of (1). For example, in Tunen regressive ATR harmony from the verb root to inflection is optional (4a) (see also Boyd 2015:31), while word-internal harmony is obligatory (4b), showing partial evidence for an INFL#VERB word boundary.

---

[1] 1, 2, 3… = Bantu noun class; 1SG = 1st person singular; PREP = preposition; PST = past tense ; SM = subject marker.

(4) a. /mɛ-ná#hólíə́/ →   [mi ná hólíə́] *or* [mɛ ná hólíə́] 'thank you'
         SM.1SG-PST#thank
     b. /#hɛ-noní/ →   [hinoní], *[hɛnoní] 'bird'
         19-bird   (Tunen; own data)

I argue that areal phonological changes such as the development of stem-initial prominence feed the creation of an INFL#VERB word boundary. This in turn explains typologically-unusual morphosyntactic patterns of the region, e.g. innovative OV patterns (cf. Mous 2005; Kerr 2024).

Implications. This study supports the synthetic→analytic hypothesis, providing a more precise model by studying the intermediate stages of change. I introduce wordhood tests that can be applied to other languages (within and outside Benue-Congo). As the tests consider areally-relevant properties such as stem-initial prominence, they allow us to better explain how the morphosyntactic variation observed between West/Central African languages and other Niger-Congo arose.

References

Boyd, Ginger (2015), *The phonological systems of the Mbam languages of Cameroon with a focus on vowels and vowel harmony,* PhD thesis, Leiden University.

Dugast, Idelette (1975), *Contes, proverbes et devinettes des Banen*, Paris: SELAF.

Güldemann, Tom (2022), Predicate structure and argument indexing in early Bantu, in Koen Bostoen, Gilles-Maurice de Schryver, Rozenn Guérois, and Sara Pacchiarotti (eds), *On reconstructing Proto-Bantu grammar* (Niger-Congo Comparative Studies), Berlin: Language Science Press, 387–422.

Hyman, Larry M. (2004), How to become a "Kwa" verb. *Journal of West African Languages*, 30:2, pp. 89-108.

Hyman, Larry M. (2008), Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics*, 46:2, pp. 309-350.

Hyman, Larry M. (2017), Multiple argument marking in Bantoid: From syntheticity to analyticity, in | Walter Bisang, and Andrej Malchukov (eds), *Unity and diversity in grammaticalization scenarios*, Berlin: Language Science Press, 67–95.

Kerr, Elisabeth J. (2024), *Tunen syntax and information structure*, PhD thesis, Leiden University.

Mous, Maarten (2005), The innovative character of object-verb word order in Nen (Bantu A44, Cameroon), in Koen Bostoen, and Jacky Maniacky (eds.), *Studies in African comparative linguistics with special focus on Bantu and Mande*, Tervuren: RMCA, 411-424.

Nurse, Derek (2008), *Tense and aspect in Bantu*, Oxford: Oxford University Press.

Sartre, Scott A., Welaze, Jacquis K., Batouan, Joseph-Alain B., Bete, Samuel, & A. Flore Sebieni, (2008), *Précis d'orthographe pour la langue tunen*, Yaoundé, Cameroon: Association Camerounaise pour la Traduction de la Bible et l'Alphabétisation (CABTAL).

Prosodic foot in Gban (South Mande)

Maksim Fedotov
(LLACAN/INALCO, Paris)

Keywords: Gban, Mande, prosodic foot, phonological word, wordhood, tone, nasalization

In Gban (*gbá̰*; Glottocode: gagu1242; a South Mande language spoken by about 60 000 people in the central part of Côte d'Ivoire), a special intermediate prosodic unit can be identified: the **foot**, which is larger than the syllable and smaller than the phonological word.

The notion of prosodic foot has been shown to be relevant to the phonological systems of a number of stressless languages, notably Mande languages, cf. (Kuznetsova 2007; Vydrin 2008; 2010). Culhane (2023) emphasizes "the need for more study of prosodic foot <…> separate from the study of prosodic prominence".

I will present a description of the (morpho-)phonological phenomena involving feet in Gban, which can also be used for their identification. The study is based on my field data from 2011–2024.

Syllables in Gban can have the structure V, CV, or C*l*V. The only consonant possible in the medial position in a syllable is /l/.

Feet can be monosyllabic: |V|, |CV|, |C*l*V|, or disyllabic: |CV$_1$.V$_2$|. I.e. the only attested type of disyllabic feet has two different vowels (all tested cases with two identical vowels turned out to be composed of two feet: |CV$_1$|V$_1$|, cf. |bà|à| 'village') and no medial consonants. Cf. |kḭ̀.á̰| 'anger'.

Feet are involved in several phenomena (serving as criteria for their identification):

1. Behaviour of non-inherent tonemes. **Toneme copying/replacement** occurs on a whole foot: in various tonal verb forms and in the intensifying *mə̰́ə̰̀*-construction, which involves toneme copying on the second reduplicant.

Cf. |gbè.ǽ| 'fall' → *…gbèà* <…IPFV\fall> vs. |yè|kɛ́| 'do' → *…yɛ̀kɛ́* <…IPFV\do>;

|kḭ̀.á̰| 'anger' → *kḭ̀á̰-mə̰́ə̰̀-kḭ̀à̰* 'real anger' vs. |blè|ɛ́| 'game (kind)' → *blèɛ́-mə̰́ə̰̀-blèè* 'real game'

2. Distribution of **inherent tonemes** and of vowel **nasalization**. (Properly speaking, the foot (and not the syllable) is the minimal domain of both toneme and nasalization in Gban.) For the structures |V|, |CV|, |C*l*V|, |CV$_1$.V$_2$| (meeting the first criterion), only six possible tonal patterns are attested in the lexicon; they correspond to the six tonemes identified in Gban. In addition, either only nasal or only oral vowels occur in the same foot.

Cf. *g^{w}ɛ̰́ɛ̰́* (eH) 'boy', *kéá* (H) 'be heavy', *lìè* (L) 'be good', *mḭ̀ɛ̰̀* (eL) 'snake', *bḛ̌ǎ̰* (rising: eL-eH) 'work', *vḭ̀ɔ̰́* (half-rising: eL-H) 'airplane'.

3. **Truncation** of syllable structures in the *mə̰́ə̰̀*-construction. The reduplication in this construction may be partial, and the minimal unit remaining to the left and to the right of the particle is a foot.

Cf. one disyllabic foot |kḭ́.á̰| 'pepper' → *kḭ́á̰-mə̰́ə̰̀-kḭ̀à̰* (/ **kḭ́-mə̰́ə̰̀-kḭ̀à̰*) 'real pepper' vs. two monosyllabic feet: |sö|kó| 'running' → *sökó-mə̰́ə̰̀-sòkɔ̀* ~ *sö-mə̰́ə̰̀-sòkɔ̀* 'real running'; |gü|ù| 'round' → *güù-mə̰́ə̰̀-gùù* ~ *gü-mə̰́ə̰̀-gùù* 'really round').

A larger unit, the **phonological word**, is still necessary in Gban to describe the optional deletion of intervocalic consonants in some C(*l*)V.**C(*l*)**V structures, occurring only inside phonological words. Cf. #sè|fɰ́ɰ́# ~ #sè|ɰ́ɰ́# 'cat', #k͡plà|k͡plɑ́# ~ #k͡plà|ɑ́# 'cunning', #kè| yɛ́# ~ #kè| ɛ́# <each_other with>.

In the talk, I will elaborate on the relationship between the foot and the phonological word on the one hand, and the morphosyntactic word and the morpheme on the other.

Glosses

1/2/3 — 1ˢᵗ/2ⁿᵈ/3ʳᵈ person; HEST — hesternal; IPFV — imperfective; PL — plural; SG — singular.

References

Culhane, Kirsten (2023), The prosodic foot beyond prosodic prominence: a preliminary survey, *Linguistic Typology* 27(2), 313–339, https://doi.org/10.1515/lingty-2022-0039.

Kuznetsova, Natalia (2007), Le statut fonctionnel du pied phonologique en gouro, *Mandenkan* (43), 13–45.

Vydrin, Valentin (2008), Metricheskaja stopa v jazykakh mande (Метрическая стопа в языках манде) [Metric foot in Mande languages], in Alexander Arkhipov, V. Zakharov, Andrej Kibrik, Alexander Kibrik, Irina Kobozeva, Olga Krivnova, Ekaterina Ljutikova, and Olga Fedorova (eds.), *Fonetika i nefonetika. K 70-letiju Sandro Vasiljevicha Kodzasova (Фонетика и нефонетика. К 70-летию Сандро Васильевича Кодзасова) [Phonetics and non-phonetics. To the 70th anniversary of Sandro V. Kodzasov]*, Moscow: Jazyki slav'anskikh kul'tur, 308–317.

Vydrin, Valentin (2010), Le pied métrique dans les langues mandé, in Franck Floric (ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*, Lyon: ENS Éditions, 53–62, https://halshs.archives-ouvertes.fr/halshs-00715537 (19 November, 2019).

Wordhood in Koalib (Niger-Congo, Kordofanian, Sudan):
some criteria to define it and their limits

Massinissa Garaoun

LLACAN – UMR 8135 (CNRS/EPHE/INALCO)

Nicolas Quint

LLACAN – UMR 8135 (CNRS/EPHE/INALCO)

Keywords: Koalib – Heibanian – Kordofanian – Language Description – Wordhood

Koalib is a Niger-Congo language (Kordofanian branch, Heiban family) spoken by ca. 100,000 persons living in or originating from the towns and localities of Abri, Delami, Jebel Nyukur, Umm Berembeita, Umm Heitan, and their surroundings (Quint 2006, 2009). Like all Kordofanian languages, Koalib has a rich synthetic morphology (Quint & Allassonnière-Tang 2022; Quint & Ali Karmal Kokko forthcoming), and therefore one needs to resort to an array of different criteria to define what a word is in this language[1]. Basically, three main criteria allow drawing the boundaries of most Koalib words: tonal domain (1), vowel harmony (2), and noun classes (3).

(1a) *kwény-èɛcɛ́* |HLH| 'I will see him'

(1b) *kwény-ɛ́ɛcɛ́* |HHH| 'he will see me'

These two inflected forms of the verb *èɛcɛ́* 'see' are segmentally identical but tonally different, and, as their tonal melody associates both with the verb and the preverbal morphemes, one must assume that (1a) and (1b) make up one word each.

(2a) *kèttám* 'book.s' vs. *kèttámɛ́* 'book.ᴏ'

(2b) *lìttém* 'amulet.s' vs. *lìttémí* 'amulet.ᴏ'

The high-toned object marker of 'book' and 'amulet' agrees in vowel-harmony (low set in (2a), high set in (2b)) with the noun and is therefore an affix, not an independent word.

(3a)　**kw**-*ór*　　　**kw**-*ínyí*　　　**kwè**-*cào*

---

[1] For more detail about wordhood, see Creissels 2006 : 13-35 ; Dixon 2010 : 1-36 ; Haspelmath & Sims 2013 : 14-32 ; 197-206.

CL_{kw}-man   CL_{kw}-POSS1SG   CL_{kw}-be.nice.PFV
'my husband (lit. 'my man') is nice'

(3b)   *I-èpéntì*   *I-ínyí*   *lè-cào*
       CL_I-teacher   CL_I-POSS1SG   CL_I-be.nice.PFV
       'my teacher is nice'

For nouns, verbs and various determiners (here possessives), the class marker is a marker of wordhood. Therefore, in (3a) and (3b), one can legitimately consider that there are three different Koalib words in each sentence.

However, things are not always that straightforward, as shown by (4), (5), (6):

(4) kwé-**Kwókkò**-ɛ̀ɛcɛ́ '**Kwokko** will see him/her/it'

In some cases, a noun (here the proper name *Kwókkò*) can be inserted between the verb word and the preverbal morphemes, which questions the 'monowordhood' of the verb forms presented in (1).

(5) *kwónʈà-búny* 'shrew'

This noun has vowels belonging both to the low (/o, a/) and the high (/u/) set.

(6a)   *k-àkró-**k-è***
       CL_k-hen-CL_k-INS
       'with the hen'

(6b)   *k-èrccú-**k-ì***
       CL_k-ox-CL_k-INS
       'with the ox'

The instrumental marker agrees both in harmony (/e/ for the low set, /i/ for the high set) – an affixal feature - and in class – a feature shared with independent words - with the noun.

In other words: there are rules to define wordhood in Koalib, but these rules suffer a significant number of exceptions. In this presentation, I will give more details about each of these rules and try to account for their limits, adducing diverse examples extracted from my fieldwork data in order to enrich the discussion. Being still an underdescribed language, Koalib definitely provides a good case-study to test the limits and consistency of the very notion of wordhood.

References

Haspelmath, Martin & Sims, Adam. D. 2013. Understanding Morphology. London/New York: Routledge.

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.

Quint, Nicolas & Ali Karmal Kokko, Siddig. 2025. Verb extensions in Koalib: a first general overview followed by some brief comparative considerations, in Gertrud Schneider-Blum, Helene Fatima Idris and Jan Junglas (eds), *Nuba Mountain Language Studies - Contemporary Perspectives*, Cologne: USB Monographs, 105-148, https://monographs.ub.uni-koeln.de/usbk/catalog/view/38/61/363 (last accessed 2025/04/27).

Quint, Nicolas & Allassonnière-Tang, Marc. 2022. Inferring case paradigms with computational classifiers. *Corpus Linguistics and Linguistic Theory* 19(2), 237–269, https://doi.org/10.1515/cllt-2021-0028 (last accessed 2024/06/19).

Quint, Nicolas, 2009, *The Phonology of Koalib, a Kordofanian Language of the Nuba Mountains (Sudan)*, Cologne: Rüdiger Köppe.

Quint, Nicolas, 2006, *Phonologie de la langue koalibe*. Paris: L'Harmattan.

Domains of phonological rules and wordhood in Burak (Adamawa, Niger-Congo, Nigeria)

Miracle Oppong Peprah & Jakob Lesage
(LLACAN; CNRS; INALCO & Humboldt University of Berlin)

This paper studies the domains of phonological rules in Burak (Adamawa, Niger-Congo, Nigeria) to investigate their role in identifying word boundaries (see e.g. Bickel et al. 2009, Aikhenvald et al. 2020, Tallman 2020). Drawing on the first author's fieldwork data, we analyze a range of processes that occur across morpheme boundaries, including:

- Hiatus resolution strategies, such as vowel coalescence (*lwá* 'fire'+ *-í* 'DET' → *lwɛ́*
- 'the fire' vs. *ɗwá í*ː [dog steal] 'a dog stole'), reduction to a glide (*nʊ́* 'AFF'+ *à* 'Q' → *nwà* vs. *Màrɩ̀jámʊ́ ábà* [Mary open] 'Mary opened'
- ATR vowel harmony (*-lɛ́* 'perfect' becomes *-lí, -lé, -lí*, depending on the vowel in the root verb)
- Word-final neutralization of voicing contrasts
- Optional word-initial (possibly postpausal) devoicing (*ɓâŋ* [b̥ậŋ] 'song')
- Low tone spread (e.g. *árɩ̀wé*ː → *árɩ̀wě*ː 'many seeds')
- Tonal absorption (e.g. *bǎ*ː *mɛ́*ː →*bà*ː *mɛ́*ː 'cook rice')

Our findings suggest that segmental processes work in tandem, operating on similar domains. These domains often coincide with morphosyntactic units such as nouns and verbs with their associated functional morphemes, suggesting a strong alignment between phonological and grammatical wordhood. However, tonal processes seem to correlate less directly with grammatical wordhood, with preliminary evidence indicating that they operate on a phrasal domain instead.

To contextualize these findings and to invite detailed descriptions of phonological processes that define wordhood in other languages in the region, we compare Burak's phonological processes with those of neighboring distantly related languages such as Kam (Lesage 2021) and Kugama (Litvinova 2023). Our paper aims to contribute to the typology of the phonology of wordhood in this area, which is characterized by widespread erosion of segmental morphology and varying levels of analyticity and morphophonological complexity.

References

Aikhenvald, Alexandra Y., R. M. W. Dixon & Nathan M. White. 2020. The essence of 'word.' In *Phonological Word and Grammatical Word*, 1–24. Oxford University Press. https://doi.org/10.1093/oso/9780198865681.003.0001.

Bickel, Balthasar, Kristine A. Hildebrandt & René Schiering. 2009. The distribution of phonological word domains: A probabilistic typology. In Artemis Alexiadou, T. Alan Hall, Janet Grijzenhout & Baris

Kabak (eds.), *Phonological Domains: Universals and deviations*, 47–78. Berlin, New York: Mouton de Gruyter. https://doi.org/10.1515/9783110217100.1.47.

Lesage, Jakob. 2020. *A grammar and lexicon of Kam (àŋwɔ̀m), a Niger-Congo language of central eastern Nigeria*. Paris: INALCO PhD Thesis.

Litvinova, Lora. 2023. *A grammatical analysis and documentation of Kugama, an Adamawa language of Nigeria*. Paris: INALCO PhD Thesis.

Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass 14*(2). https://doi.org/10.1111/lnc3.12364

Phonological and morphosyntactic word in Guro in the planar-fractal model

Natalia Kuznetsova
(Università Cattolica del Sacro Cuore, Milan)

Keywords: Guro, Niger-Congo, phonological words, morphosyntactic words, parametric typology

The problem of wordhood has recently re-gained the attention of linguists and has been a hotly debated theoretical issue (e.g. Haspelmath 2023; Tallman 2024). Previous work on wordhood in Guro, a South Mande language of West Africa (Niger-Congo), has primarily touched upon the phonological wordhood problems (N. Kuznetsova 2007, 2021). In particular, that work has found mutual mismatches between the morphosyntactic word and the two prosodic domains correlating to the lexical level: the "featural foot" (Green 2015; Vydrin 2020) and the domain of the grammatical tonal change. Such a mismatch creates so-called "bracketing paradoxes" (Sproat 1988) and presents challenges for the Match Theory (Selkirk 2011; Elfner 2018). The Match Theory is a recent development of the Prosodic Hierarchy hypothesis which assumes that the prosodic categories starting from the word level are directly derived from the syntactic constituent structure and, therefore, expects no mismatch between prosodic and morphosyntactic constituents. The aforementioned Guro data, however, contradicts this expectation.

In the present talk, I will go beyond the state-of-art and propose a broader picture of correspondences and the lack thereof between the "phonological words" and "morphosyntactic words" of different size established in Guro by the consecutive application of various general and language-specific criteria for the wordhood. I will apply the planar-fractal method proposed by Tallman (2020, 2024; Auderset et al. 2024) to the basic sentence structure of Guro (as outlined in N. Kuznetsova 2024: 67), to the extent of detalisation possible. I will use my field data on Guro collected on site and online in different time periods between 2006 and 2021 and data from the new narrative corpus of Guro (O. Kuznetsova 2022).

For nouns, the preliminary results of the application of about two dozen phonological and morphophonological criteria and of about ten syntactic criteria show a statistical convergence, around a disyllabic unit. This unit occupies one "slot" on the planar-fractal grid and functionally usually corresponds to a Guro root and to the featural foot, but not always. For verbs, such a unit of major convergence includes, apart from the root, also the suffix of the imperfective (i.e. two "slots" on the grid).

However, I will also show that by many criteria the size of both the "phonological word" and the "morphosyntactic word" may be significantly larger, sometimes covering 8-9 slots. This frequent lack of convergence between different criteria is similar to the situation described by Tallman et al. (2024) for many of the languages of the Americas and is the main source of indeterminacy also for the establishment of clear orthographic principles for word spelling and word boundaries in Guro.

References

Auderset, Sandra, Caroline de Becker, Gladys Camacho Rios, Eric W. Campbell, Carol, Javier, Duzerol, Minella, Epps, Patience, et al. (2024), Constituency and convergence in the Americas - Results and discussion, in *Constituency and convergence in the Americas* [Topics in Phonological Diversity 1], Adam J. R. Tallman, Sandra Auderset & Hiroto Uchihara (eds), 735-777. Berlin: Language Science Press, https://doi.org/10.5281/ZENODO.13208572

Elfner, Emily (2018), The syntax-prosody interface: current theoretical approaches and outstanding questions. *Linguistics Vanguard* 4(1), https://doi.org/10.1515/lingvan-2016-0081

Green, Christopher (2015.), The foot domain in Bambara, *Language* 91(1), e1-26, https://doi.org/10.1353/lan.2015.0009

Haspelmath, Martin (2023), Defining the word, *WORD* 69(3), 283-97, https://doi.org/10.1080/00437956.2023.2237272

Kuznetsova, Natalia. 2007. Le statut fonctionnel du pied phonologique en gouro, *Mandenkan* (43), 13-45.

Kuznetsova, Natalia (2021), Phonological wordhood issues in Guro (South Mande), *L'Analisi linguistica e letteraria* 29(3), 43-54.

Kuznetsova, Natalia (2024), *Elements of Guro Grammar: Phonology and Morphology. Non-Verbal Lexical Classes. Determination, Deixis, Information Structure* [ACCADEMIA]. Milano: EDUCatt.

Kuznetsova, Olga (2022), A narrative corpus of Guro. CNRS-LLACAN; LACITO. *The SpeechReporting Corpus: Discourse Reporting in Storytelling*, 2022, <http://discoursereporting.huma-num.fr/corpus.html>

Selkirk, Elisabeth (2011), The syntax-phonology interface, in *The handbook of phonological theory* [Blackwell Handbooks in Linguistics], John A. Goldsmith, Jason Riggle & Alan C. L. Yu (eds), 2nd ed., 435-485. Malden: Wiley-Blackwell.

Sproat, Richard (1988), Bracketing paradoxes, cliticization and other topics: The mapping between syntactic and phonological structure, in *Morphology and modularity: In honour of Henk Schultink*, Martin Everaert, Arnold Evers, Riny Huybregts & Mieke Trommelen (eds), 339-360. Dordrecht: Foris.

Tallman, Adam J. R. (2020), Beyond grammatical and phonological words, *Language and Linguistics Compass* 14(2), https://doi.org/10.1111/lnc3.12364

Tallman, Adam J. R. (2024.), Introduction: Phonological and morphosyntactic constituency in cross-linguistic perspective, in *Constituency and convergence in the Americas* [Topics in Phonological Diversity 1], Adam J. R. Tallman, Sandra Auderset & Hiroto Uchihara (eds), 1-84. Berlin: Language Science Press, https://doi.org/10.5281/ZENODO.13208540

Tallman, Adam J. R., Auderset, Sandra & Uchihara, Hiroto (eds) (2024), *Constituency and Convergence in the Americas* [Topics in Phonological Diversity 1]. Berlin: Language Science Press, https://doi.org/10.5281/zenodo.10559861

Vydrin, Valentin (2020), Featural foot in Bambara, *Journal of African Languages and Linguistics* 41(2), 265-300, https://doi.org/10.1515/jall-2020-2012

On cliticization and prosodic units in Ʉt-Maꞌin

Rebecca Paterson

(Princeton University & SIL)

I explore the structure of prosodic units in Ʉt-Maꞌin (Northwest Kainji; Nigeria), with particular concern for patterns that defy morphosyntactic-word boundaries. Crucial to the analysis of nominal modification in Ʉt-Maꞌin is the form of the head noun and contiguous (ad)nominal class marking (cf. Shirtz & Paterson 2025; see also "clitic-like" adnominal marking in Güldemann & Fiedler, to appear). The morphological shape of the noun depends on which, if any, modifier immediately follows the head. For example, the class 5/6 'basket' obligatorily occurs in distinct forms depending on the accompanying modifiers (Table 1).

Table 1: Noun forms and co-occurring modifiers in Ʉt-Maꞌin

Form	Type of adnominal noun marker	Singular forms:	Plural forms:	Co-occurring modifiers
I	Mid-tone prefix	ə̄r-kɔ́ːr	ə̄t-kɔ́ːr	QUANT; No modifier (object)
II	Copy-tone suffix	kɔ́ːr-ə́r	kɔ́ːr-ə́t	DEF, DEM, A, POSS
III	Low-tone suffix	kɔ́ːr-də̀	kɔ́ːr-tə̀	No modifier (subject)
IV	No affix	kɔ́ːr	kɔ́ːr	NOMINAL, RELCL, INDEF-SPEC

Complex expressions like *jàʔ tə̀ stéʔ tɔ́* 'fruits' (1), can be analyzed as four morphosyntactic words and simultaneously two prosodic units: [CVC.CVC] and [CVC.CV]—splitting the morphologically obligatory agreement prefix *s-* from the root *té* of the modifier noun (cf., citation forms: *(ə̄)t-jàʔ* 'C6-fruits' and *(ə̄)s-téʔ* 'C4-trees').

(1)   [jàʔ.tə̀s # té.tɔ́]   [ʃámə̀t # mɔ́ŋɔ̀ré]

    [N$_i$ AG$_i$-ASSOC   C$_j$-N$_j$ DEF$_i$]

    jàʔ   *t*-L   *s-téʔ*   *tɔ́*   *ʃámə̀*   *t-mɔ́ŋɔ̀ré*

    fruit   AG6-ASSOC   CL4-trees   AG6.DEF   resemble   CL6-mango

    'the fruits resemble mangos'

Depending on the syllabic structure of particular roots and noun class morphology in combination, this construction (without DEF marking) is attested with three prosodic configurations: N$_i$=AG$_i$-ASSOC=C$_j$ #N$_j$; N$_i$ #AG$_i$-ASSOC=C$_j$ #N$_j$; and N$_i$ #AG$_i$-ASSOC=C$_j$-N$_j$. (Subscripts i and j indicate the same "noun class/agreement" value.)

The data suggest that the notion of syllable and the notion of phrase are both relevant in the domain of nominal phrases, but the notion of word is less critical. I show this phenomena with data from four NW Kainji languages.

1

References

Güldemann, Tom and Ines Feidler. To appear. The synchronic and diachronic diversity of adnominal gender-number marking in Niger-Congo.

Paterson, Rebecca Dow Smith. 2019. Nominalization and Predication in Ut̲-Maꞌin. PhD dissertation, University of Oregon. https://scholarsbank.uoregon.edu/xmlui/handle/1794/25259.

Shirtz, Shahar & Rebecca Paterson. 2025. ConstructiCon-based description: A nominal modification case study. *Constructions* 17. 1–35. https://doi.org/10.24338/CONS-649.

Welmers, William E. 1963. Associative *a* and *ka* in Niger-Congo. *Language* 39(3). 432–447. https://doi.org/10.2307/411125.

Abbreviations

#	word boundary
A	adjective
AG	agreement marker (plus agreement class number label: 4, 5, 6)
ASSOC	associative marker (ala Welmers 1963)
CL	noun class (plus class number label: 4, 5, 6)
C/V	consonant/vowel
DEF	definite marker
DEM	demonstrative
INDEF-SPEC	indefinite specific
N	noun
POSS	possessive
QUANT	quantifier
REL.CL	relative clause

On the wordhood of N + ADJ sequences in West African languages: The case of Gur/Mabia languages

Songfolo Lacina Silue & Vadim Dyachkov
(LACITO CNRS & LLACAN CNRS, Paris)

Keywords: Gur/Mabia, incorporation, nominal compounding, NP cartography, Senufo

Nouns in Gur/ Mabia languages typically belong to different classes expressed by nominal suffixes and agreement markers on noun modifiers. While in some attributive constructions, the noun appears with its class marker and the adjective with its agreement suffix as in Lamba (1), in other cases, the class suffix appears only on the stand-alone noun or the modifying adjective as in Gurmanche (2).

(1)   a.   hɔ̃ 'dog'   +   cɐ́pînô 'a black one'   → hɔ̃ cɐ́pînô 'a black dog'
        b.   hásâ 'dogs'   +   cɐ́pînásâ 'black ones'   → hásâ cɐ́pînásâ 'black dogs'
        (Creissels 2016: 6, citing Aritiba 1987: 108-109)

(2)   a.   tí-bū 'tree'   +   ciám- 'big'   → tī-ciám-bū 'big tree'
        b.   tīi-dí 'trees'   +   ciám- 'big'   → tī-cián-dī 'big trees'
        (Creissels 2016: 7, citing Ouoba 1982: 131-133)

The latter case is usually analyzed as compounding/ incorporation with the adjective being inserted between the noun and its class suffix (Nicole 1999: 18, Carlson 1994: 166, Creissels 2018: 736, and Nicole 2018: 59). However, following the ideas proposed in Nikitina & Silué (2022), we present evidence against this analysis based on our own fieldwork data (namely, from Senufo and Natioro). For instance, in Kafire (< Senufo group), the adjective order in combinations with nouns is free, cf. (3), and is determined by information structure (to be discussed in detail).

(3)   a.   túbéré   **wɔ́**   cɛ̀   vɔ̰́=ʔɔ̰̄
           shoe   black   pretty   new=CL2
           'a pretty BLACK new shoe'
        b.   túbéré   **cɛ̀**   wɔ́   vɔ̰́=ʔɔ̰̄
           shoe   pretty   black   new=CL2
           'a PRETTY black new shoe'

We claim that constructions with adjectives: 1) are not instances of incorporation because they do not have non-incorporated counterparts; 2) are not compounds because they are absolutely productive and not lexicalized. In addition, compounds tend to be right-headed (Williams 1981), which is not the case in the languages under study. Other typological criteria opposing compounds to N + ADJ phrases are: 1) idiomaticity ; 2) inseparability and the possibility of adjective reordering. For instance, in Kafire, when the combination of the noun 'man' and the adjective 'old' forms a compound, it has an idiomatic meaning ('uncle'). No adjective can intervene between the two elements, cf. (4). In contrast, when the combination is a phrase, the adjective order is not rigid, cf. (5).

(4)   a.   sɛ̄lɛ̄   cɛ̀=w   b.   #sɛ̄   cɛ̀   lɛ̄=w
           uncle   good=DEF1.SG   man   good   old=DEF1.SG
           'the good uncle'   'the good uncle'
(5)   a.   sɛ̄   lɛ̄   cɛ̀=w   b.   sɛ̄   cɛ̀   lɛ̄=w
           man   old   good=DEF1.SG   man   good   old=DEF1.SG
           'the good old man'   'the GOOD old man'

The other aspect of our alternative proposal aims to explain the non-suffixal nature of class markers. We propose that they are clitics attaching to the rightmost category within a certain syntactic domain. We support this with traditional criteria for identifying clitics (Haspelmath 2023). Moreover, we claim that within the sequences N + ADJ + CL, the order of constituents straightforwardly reflects the NP hierarchy N-ADJ-NUM-D or N-ADJ-D-NUM, where the class markers spell out D, NUM or NUM-D functional elements. We sketch a theoretical analysis of these constructions.

References

Aritiba, Adji Sardji (1987), *Le lamba de defale (langue gurunsi du togo): phonologie et morphologie*, Grenoble: Université Stendhal. (Doctoral dissertation).

Carlson, Robert (1994), *A grammar of supyire*, Berlin / New York: Mouton de Gruyter.

Creissels, Denis. 2016. Noun+ adjective compounding and quasi-compounding in the languages of sub-Saharan Africa, *Syntax of the World's Languages 7, Mexico, August* 17–19.

Creissels, Denis (2018), Current issues in African morphosyntactic typology, in T. Güldemann (ed.), *The Languages and Linguistics of Africa.* 712–821. Berlin: De Gruyter Mouton.

Haspelmath, Martin (2023), Types of clitics in the world's languages, *Linguistic Typology at the Crossroads* 3(2), 1–59.

Nicole, Jacques (1999), *Les classes nominales dans les langues voltaïques: esquisse d'un cadre de description*. Lome: SIL Togo.

Nicole, Jacques (2018), Grammaire du nawdm, *SIL e-Books* 69.

Nikitina, Tatiana & Songfolo Lacina Silué (2022), Noun class agreement in kafire (senufo): a lexical-functional grammar account, *Journal of Linguistics* 59(1), 1–28.

Ouoba, Bendi Benoît (1982), *Description systèmatique du gulmancema: phonologie, lexicologie, syntaxe*, Paris: Université de la Sorbonne Nouvelle (Paris 3). (Doctoral dissertation).

Williams, Edwin (1981), On the notions 'lexically related' and 'head of a word', *Linguistic Inquiry* (12), 245–274.

Tonal criteria of wordhood in West African languages

Valentin Vydrin

(INALCO & LLACAN, CNRS, Paris)

In the studies of the prosody, the word accent (or stress) is traditionally considered as a dedicated means to secure the word integrity. In tonal languages of West Africa stress is commonly absent, and tone is rarely culminative, for which reason one may think that tones are more or less irrelevant for the word delimitation.

Certainly, this opinion is incorrect. Tones provide numerous clues for the solution of the wordhood problems, however, these solutions are often much less straightforward than those stemming from the stress. Tone-related factors which can contribute to the establishing word boundaries are language-specific and highly variable, but still, they can be grouped into a limited set of types.

– Word melodies. A language may have a limited set of toneme combinations (melodies) within a word; such melodies are commonly extensible. In some languages, tonal melodies are of primary importance for the definition of wordhood. E.g. Hausa (Chadic) (Jaggar 2001; Litvinova 2024), Soninke (Western Mande) (Creissels 2016; Vydrin 2024).

– In some languages, a toneme span coincides (by default) with a prosodic word. With some reservations, it is true for most of Manding languages (Mande) (Vydrin 2019) and probably for Mende (Mande) and Chumburung (Guang) (Snider 2018). Otherwise, within a word, a sequence of TBUs bearing identical tones tends to represent one tonal span (i.e., to represent one toneme).

– Positional constraints on tones. A language may have restrictions on toneme occurences with respect to word boundaries. E.g., in Eastern Soninke, in the word-final position, two (or more) adjacent syllables cannot bear identical tones if preceded by a different tone (within the limits of the same word). I.e., sequences *LHH and *HLL are prohibited.

– Grammatical tone domain. In many languages, replacive tonal morphemes are mapped on entire words and can be therefore regarded as markers of word limits; e.g.: Hausa (Jaggar 2001; Litvinova 2024), Soninke (Creissels 2016; Vydrin 2024).

– Tonal processes are often sensitive to word boundaries and can therefore serve wordhood criteria. E.g., in Mwan (Southern Mande) the rule of L toneme dissimilation is applied across the word boundary (Perekhvalskaya & Vydrin 2024).

It is also true that the word boundaries established on the basis of tonal factors may differ from those established by other criteria.

References

Creissels, Denis. 2016. Phonologie segmentale et tonale du soninké (parler du Kingi). *Mandenkan* 55. 3–174.

Jaggar, Philip J. 2001. *Hausa* (London Oriental and African Languages Library 7). Amsterdam / Philadelphia: John Benjamins.

Litvinova, Lora. 2024. Hausa. Analytical report on the tonal system. Villejuif. https://doi.org/10.5281/zenodo.14749131.

Perekhvalskaya, Elena & Valentin Vydrin. 2024. Mwan. Analytical report on the tonal system. Villejuif. https://doi.org/10.5281/zenodo.14725476.

Snider, Keith W. 2018. *Tone analysis for field linguists*. Dallas, Texas: SIL International.

Vydrin, Valentin. 2019. *Cours de grammaire bambara*. Paris: Presses de l'INALCO.

Vydrin, Valentin. 2024. Eastern Soninke. Analytical report on the tonal system. Villejuif. https://doi.org/10.5281/zenodo.14725393.

Special outreach session: Communicating linguistics research

Marc Olivier & Alexandru Nicolae

How skilled are you with the tricks of political communication?
Cues, challenges and didactic developments

Mara Frascarelli & Viviana Masia
(Roma Tre University)

One of the most challenging frontiers of recent trends in applied and experimental linguistics is the assessment of so-called "transversal language skills" already in the pre-university school system (cf. Frascarelli 2021; Saussure 2005; Maillat and Oswald 2009). The contexts in which this occurs more often is political discourse and advertising (Lombardi Vallauri 2019; Masia 2021).

Among other visual or semiotic traits, what often makes political and commercial messages deceptive is the presence of subliminal communicative strategies that are likely to skip the addressee's awareness. These are by and large represented by presuppositions, implicatures, vague expressions (Grice 1975; Stalnaker 2002; Sorensen 2003), among others. Below, examples from propagandistic language uses are given.

(1) Renault Kadjar. *Stop* watching, *start* living (PRESUPPOSITION – change of state verb)
(2) It is not possible to go on with tactics and blank ballots (IMPLICATURE – post of an Italian politician)
(3) Apple. Think *DIFFERENT* (VAGUENESS)

Later developments in implicit communication research have sought to bring reflection on manipulative communication outside the academic walls and, specifically, to make it available to the average citizen in order to equip him with fundamental metalinguistic tools to "defend" themselves from tendentious communication practices. If this is true for any individual – regardless of their level of education, employment and/or social position – it is even more true for young people (still of school age) who are near their coming of age and thus very close to start exercising their right to vote. The potentially harmful effects of a lack of awareness of the way in which certain linguistic uses can manipulate impose a reversal of the trend.

This talk will discuss the outcome of a training experience in which 90 students took part in a 40-hour course aimed at strengthening their ability to detect implicit language in different types of manipulative texts. The training sessions mostly zoomed on implicit strategies such as presuppositions, implicatures, as well as (semantic or syntactic) vagueness, and got the students assess their impact on the construal of (c)overt sentence meaning.

For both theoretical and more practical sessions, the selected working materials consisted in real excerpts from politicians' speeches, Twitter/X's posts as well as slogans of commercial ads, either in Italian or in other languages (among those studied by the participants).

A pre-training and a post-training test allowed monitoring the remarkable and ongoing progress of the students, also thanks to the administration of hands-on activities designed to gauge the advancement of their comprehension levels. On debriefing interviews, the students reported of an extremely positive impact of the training on their capacity to read a message "between-the-lines",

thereby becoming increasingly more capable of getting the most out of the information load of an utterance.

Within the objectives of the SLE *"Linguistics in Schools Manifesto"*, we do believe that this undertaking might constitute a valuable gambit in the direction of constructing language competences for "life-long learning" and introducing ad-hoc curricula to foster the growth of learners' independence in democratic societies as well as construct a more desirable social awareness.

References

Frascarelli, Mara (2021), Valutare le competenze linguistiche in italiano L1 tra scuola e università. Criteri e obiettivi, in P. De Socio, Patrizia (ed.) *I Quaderni della Ricerca, 60. Cultura e scuola. Prima e durante l'emergenza: fare i conti con la distanza* (LES), 127-143. Torino: Loescher.

Grice, Herbert P. (1975), Logic and Conversation, in P. Cole, and J. Morgan (eds.), *Syntax and Semantics 3: Speech Acts*, New York: Academic Press, 41-58.

Lombardi Vallauri, Edoardo (2009), *La struttura informativa. Forma e funzione negli enunciati linguistici*. Roma: Carocci.

Maillat, Didier and Oswald, Steve (2009), Defining Manipulative Discourse: The Pragmatics of Cognitive Illusions, *International Review of Pragmatics* 1: 348–370.

Masia, Viviana (2021), *The Manipulative Disguise of Truth*. Amsterdam/Philadelphia: John Benjamins.

Saussure de, Louis (2005), Manipulation and Cognitive Pragmatics: Preliminary Hypotheses, in Louis de Saussure, and Peter Schulz (eds)*, Manipulation and Ideologies in the Twentieth Century: Discourse, Language, Mind*, Amsterdam: John Benjamins, 113-146.

Rigotti, Eddo (2005), Towards a Typology of Manipulative Processes, in Louis de Saussure, and Peter Schulz (eds)*, Manipulation and Ideologies in the Twentieth Century: Discourse, Language, Mind*, Amsterdam: John Benjamins, 61-83.

Sbisà, Marina (2007), *Detto non detto. Le forme della comunicazione implicita*. Roma-Bari: Laterza.

Sorensen, Roy A. (2003), Vagueness Within the Language of Thought. *The Philosophical Quarterly* 41(165): 389–413. https://doi.org/10.2307/2220076

Why language teaching needs linguistics: evidence from the UK

We present a cross-institution UK project which makes the case for the place of linguistics in school-based language teaching to enhance both uptake and results. In UK schools, languages are thought of purely as skills, with teaching and assessment focused on reading, writing, listening and speaking. This makes languages study distinct from all other disciplines in schools, including even physical education which post-age 16 includes a 70% focus on critical/analytical engagement in assessment. We contend that this focus on skills limits the appeal of studying a language, with detrimental effects on uptake both at school and university. We report on the design and conduct of two co-created interventions, providing evidence that the inclusion of the critical/analytical study of language(s) in the form of linguistics is: (i) attractive to students and teachers and; (ii) feasible without curricula/framework reform or extensive teacher-training.

In the first intervention, groups of academic linguists and experienced teachers co-created materials for 16-18 year olds on topics such as: linguistic variation and regional identity, language and digital media, language-trait-focused discrimination and attitudes towards language change in French, German and Spanish. These materials were tested in UK schools, with feedback gathered from 17 teachers and 65 pupils via online questionnaires and semi-structured interviews. The results (reported in Authors 2024) show that linguistics is perceived by teachers and students as being different from what is currently offered but nonetheless compatible with the existing curriculum. Notably, pupils found the materials enabled critical debates and – owing to the particular approach to materials design adopted – teachers reported that they were feasible for them to teach, even without an academic background in linguistics.

A second co-creation project (in progress) replicates the above but with younger students (aged 12-13). This is a crucial age as languages are compulsory in England until age 14. Thereafter, students choose which subjects to continue until age 16; languages are often a casualty. This project will assess whether an intervention can have a positive effect in arresting declining rates of languages selection. To do so, we again enrich the existing curriculum by adding a critical/analytical dimension to existing materials for French, German and Spanish. For example, the study of cognates in French can be significantly enriched by a discussion of: (i) the history of English and French; (ii) principles of borrowing and (iii) stress differences between the languages.

The aims of these interventions, along with our other engagement activities, are to co-create a set of high-quality, open-access linguistics-based materials which fit into existing UK syllabi; to test their efficacy and appeal from both a teacher and a learner perspective; to record any changes in pupils' attitudes to language and language study following exposure to the materials; to gather testimonies; and ultimately use these findings to build a body of evidence to lobby for the inclusion of linguistics in the languages curriculum. Success has also been facilitated by a close collaboration with textbook producers, exam boards, the Department for Education, as well as civil service linguists.

Linguistics unplugged: A five-year journey in public engagement

Francesca Masini, Nicola Grandi, Nicole Marinaro & Alex Piovan

(University of Bologna, University of Bologna, Ulster University & Independent researcher)

Keywords: science popularization, linguistics outreach, blog, social networks, Italy

In this talk aimed at the Special panel session "Communicating linguistics research", we present and discuss results from our practical experience as founders and directors of an outreach enterprise that has been popularizing linguistics in Italy since 2020: *Linguisticamente* (www.linguisticamente.org/). *Linguisticamente* is the first website/blog in Italy specifically dedicated to the popularization of language sciences. Over the last 5 years, we published 156 articles written by 131 authors (data update: May 2025), including advanced students and early career researchers (see Figure 1). The target is broad, ranging from (university) students to teachers to the general audience.

Most articles deal with specific domains/topics of language sciences often anchored to current research projects (41%) and with the practical side of linguistics, including its importance to better understand current events and news (27%). A fair number of articles are devoted to famous linguists (8%) and myths to be dispelled (5%). Finally, 19% of the articles have a mixed nature, comprising more than one the above-mentioned characteristics.

Figure 1. Number of outreach articles over the last 5 years

Articles are posted on the blog and then shared on Facebook (almost 5300 followers, 77% of which are women) and Instagram (more than 3000 followers). Over the last year, the blog obtained 43.126 total clicks and 2,43 Mln impressions, mostly from our own country but also (marginally) from other European countries as well as the USA. The most clicked articles concern topics like languages and dialects, semantics, corpora, language (faculty) vs. (specific) language, generative grammar, language and thought, grammatical categories, or how to survive your first general linguistics exam.

In the talk, we will discuss these data in further details, also with respect to the national context, and we will share our insights about linguistics outreach and the challenges linguists need to address to make linguistics communication projects more effective. To this end we will illustrate what *was* and *was not* successful in our 5-year experience, as briefly summarized in Table 1, and we will put forward some possible explanations that might help current and future projects.

Successful	Unsuccessful
Quantity of articles	Not all articles are "popular" enough
Quantity of authors involved	Authors do not volunteer easily
Good reception from scholarly community and students	Difficulty in becoming a reference for science journalists
Good ranking in search results/engines	Difficulty in managing social interactions

Table 1. Successful and unsuccessful aspects of our enterprise