WS19 The Arabic (Semitic) Lexicon and its words

Neuroimaging evidence that Arabic roots are syntactically underspecified

Suhail Matar, Samantha Wray, Sherine Bou Dargham, Linnaea Stockall & Alec Marantz
(Basque Center on Cognition Brain and Language, Dartmouth College, New York University Abu Dhabi, Queen Mary University of London & New York University)

Keywords: morphology, magnetoencephalography, neurolinguistics, morphosyntax, psycholinguistics

Contemporary models of lexical access in the brain posit distinct stages for semantically-blind decomposition of a complex word into its potential constituent morphemes, and recomposition stages that check syntactic and semantic compatibility of the morphemes (Neophytou et al. 2018, Stockall et al. 2019). These models are primarily informed by data from Indo-European languages with affixes appended to cohesive, contiguous stems.

The current study recorded brain activity via magnetoencephalography (MEG) for N=18 native Arabic speakers as they read real words and nonwords and asked if each was real or not (i.e. a visual lexical decision task). In addition to reading grammatical, attested words, speakers read two types of nonwords (following Neophytou et al. 2018): Syntactic Violation, in which roots only attested in tandem with nominal patterns were interleaved with a verbal pattern (e.g. ʕ-q-r-b 'scorpion' + ta-a--a- 'passive/reflexive' = '*scorpioned'); and Semantic Violation, with roots attested in verbsbut interleaved with a passive/reflexive verbal pattern in which they are unattested (e.g. z-ɣ-r-d 'trill' + ta-a--a- 'passive/reflexive' = '*was trilled').

Spatiotemporal cluster-based regressions at 10,000 permutations (following Maris & Oostenveld 2007) were performed to determine clusters of differences in brain activity in time and space associated with the different types of stimuli. For the decomposition stage, we probed the relationship between root frequency and whole word frequency (formalized as stem-to-word transition probability; TP) and its modulation of elicited activity in the fusiform gyrus 150-250ms after stimulus presentation. A significant cluster of activity demonstrating a correlation between TP and activity in left hemisphere fusiform gyrus was found from 165-225ms (p = 0.03), which is consistent with previous research showing that the putative Visual Word Form Area in fusiform gyrus is a seat of early semantically-blind morphological processing (Gwilliams & Marantz 2018, Stockall et al. 2019).

For recomposition, we examined a period of time 300-500ms after stimulus presentation in orbitofrontal cortex. Two clusters were found: in the left hemisphere, Semantic Violation items elicited significantly more activity than Syntactic Violation items from 426-492ms (p = 0.052); in the right hemisphere, Semantic Violation items also elicited more activity than Syntactic Violation items, in a comparable time window (421-478ms, p = 0.017).

This study builds on the long history of psycholinguistic research that has demonstrated the primacy of the Arabic root in the mental lexicon (Prunet et al. 2000, Boudelaa et al. 2010, among others) by showing that Semitic words, which are characterized by their non-concatenative root-and-pattern morphology, are decomposed using the same brain regions and in the same time scale as words in languages with concatenative morphology. In contrast, we show the recomposition stages to be divergent from these languages; only semantic – and not syntactic -- compatibility between the root and pattern is checked during this process. We interpret these results as being due to the fact that Arabic roots are traditionally

analyzed as being underspecified; a root does not receive its syntactic category until after interleaving with a pattern.

References

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., & Marslen-Wilson, W. (2010). Arabic morphology in the neural language system. Journal of cognitive neuroscience, 22(5), 998-1010.

Gwilliams, L., & Marantz, A. (2018). Morphological representations are extrapolated from morpho-syntactic rules. Neuropsychologia, 114, 77-87.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods, 164(1), 177-190.

Neophytou, K., Manouilidou, C., Stockall, L., & Marantz, A. (2018). Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. Brain and language, 183, 11-20.

Prunet, J. F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. Linguistic inquiry, 31(4), 609-648.

Stockall, L., Manouilidou, C., Gwilliams, L., Neophytou, K., & Marantz, A. (2019). Prefix stripping re-re-revisited: MEG investigations of morphological decomposition and recomposition. Frontiers in Psychology, 10, 455621.

Semantic complexity between root and template

Peter Hallman

(Austrian Research Institute for Artificial Intelligence)

Keywords: Templatic morphology, aspect, participles, argument structure, Arabic

In this talk, I claim that Arabic roots may be associated with semantic derivational operators prior to being given a syntactic category by template association. I argue for this on the basis of words that display properties that cannot be attributed to either the template they exemplify—because other words in the same template do not share the property—nor to the root—because other derivatives of the same root to not share the property. This suggests that it must be possible to semantically augment a root before its association with a template.

In English, the verb *weaken* meaning 'become weak' is transparently derived from the adjective *weak* by the inchoativizing and verbalizing suffix *-en*. In Arabic, the adjective *ḍaʕiːf* 'weak' has more prosodic structure (CVCVːC) than the verbal counterpart *ḍaʕuf* 'weaken' (CVCVC), suggesting that the latter is not dervied from the former (assuming derivation may add but not subtract prosodic structure), but rather that both are derived from the root *ḍʕf* by association to different templates. But while the adjective *ḍaʕiːf* describes a state of being weak, the verb *ḍaʕuf* describes a change of state from not weak to weak. It is tempting to attribute the inchoative part of the meaning of the verb to the verbalizing template. However, this template also characterizes verbs that are purely stative, such as *baqiy* (>*baqiː*) 'remain' or *xalaw* (>*xalaː*) 'be empty'. I conclude that the root *ḍʕf* combines with an inchoative operator BECOME prior to being verbalized by the CaCVC template, as shown in (1a).

Generalizing this conclusion, I propose that roots are given an aspectual profile prior to being verbalized by the CaCVC template, either by combining with BECOME, like *ḍʕf*, or with BE, like *xlw*, shown in (1b), but that the adjectivizing template CaCiːC combines with a root directly, shown in (1c).

(1)   a.   [$_V$ CaCVC   [$_{Asp}$ BECOME   [$_V$ ḍ ʕ f]]]   → ḍaʕuf 'become weak'

        b.   [$_V$ CaCVC   [$_{Asp}$ BE   [$_V$ x l y]]]   → xalaː 'be empty'

        c.   [$_A$ CaCiːC   [$_V$ ḍ ʕ f]]   → ḍaʕiːf 'weak'

The idea that the adjectival template CaCiːC combines with a root directly is supported by the fact that such adjectives typically describe a result state of the corresponding verb, e.g. *kasiːr* 'broken', *qatiːl* 'killed', *faqiːd* 'lost or deceased' etc., in contrast to the active participles of the corresponding verbs in the template CaːCiC, which are also adjectival but describe the event having that result, e.g. *kaːsir* 'breaking', *qaːtil* 'killing', *faːqid* 'being deprived (of) or bereaved'. This makes sense from the perspective of Embick's (2004) analysis of participles, which maintains that roots systematically denote result states and are augmented with inchoative or causative semantic formatives to form verbs, which in turn feed into participle formation. I conclude that what counts as a base for a template may in principle contain more semantic structure than what inheres in the root alone, meaning that some structure building processes are prior to template association.

References

Embick, David (2004). *On the structure of resultative participles in English*. Linguistic Inquiry 35(3):355-392.

The distribution of agreement markers in Arabic and Amharic:
The role of Vocabulary Insertion

Matthew Hewett & Ruth Kramer
(Georgetown University)

1. The distribution of agreement. In dialects of Arabic, both the main verb and the auxiliary display full subject agreement when they co-occur (*modulo* word order in Standard Arabic; Aoun et al. 1994):

(1)   beʃ   j-ku:n-u   j-ʃu:f-u
        FUT   3-AUX-PL   3-see-PL
        'they will be seeing' (Tunisian Arabic)

In contrast, in the Amharic complex gerund, suffixal agreement appears on both the gerund and the enclitic auxiliary in the third feminine singular ((2)) but only on the gerund in the second feminine singular ((3)).

(2)   säbrä-a   -all-ätʃtʃ   (>*säbrallätʃtʃ*)
        break.GER-3FSG   -AUX.NPST-3FSG
        'she has broken' (Leslau 1995:387)

(3)   säbrä-ʃ   -all
        break.GER-2FSG   -AUX.NPST
        'you.FSG have broken' (*ibid.*)

We propose that in both Arabic and Amharic there is full agreement on auxiliaries and main verbs in the syntax. Adopting the framework of Distributed Morphology, we show how the disappearance of subject agreement markers in Amharic follows from a constraint on Vocabulary Insertion that blocks multiple insertion of the same exponent, which we propose applies only within specific morphological domains.

2. Morphological domains. One important morphological domain is the morphosyntactic word (M-word): the highest segment of an X^0 not contained in an X^0, i.e., the topmost node of a complex head (Embick and Noyer 2001). There are various indications that the main verb and auxiliary comprise a single M-word in Amharic but two M-words in Arabic. In Amharic, freestanding words cannot intervene between the main verb and auxiliary. However, in Arabic, this is grammatical. Moreover, the complex gerund in Amharic behaves as a single domain for stress assignment (Sande and Hedding 2017) whereas the main verb and auxiliary have separate stress in Arabic. Finally, both Arabic and Amharic have robust written traditions; in Amharic, the compound gerund is written as a single unit whereas in Arabic the verb and auxiliary are orthographically separate.

3. Uniqueness Constraint. In Distributed Morphology, Vocabulary Insertion is the process by which abstract syntactic feature bundles are matched with morphophonological exponents. We propose that Vocabulary Insertion comprises two steps: (i) Select, in which an exponent is chosen, and (ii) Insert, in which the exponent is inserted. Crucially, we also propose a domain-specific Uniqueness Constraint on Insert: the same exponent cannot be Inserted more than once within an M-Word. This straightforwardly derives the entire Amharic complex gerund paradigm. For example, in (2), the exponent of agreement on the gerund (i.e. *-a*) is distinct from the exponent of agreement on *allä* (i.e. *-atʃtʃ*); hence, Uniqueness is not relevant. By contrast, in (3), Uniqueness blocks the insertion of auxiliary agreement (expected form: *-ʃ*) because it would be identical to main verb agreement and the two are in the same M-word. In Arabic, auxiliaries and main verbs do not form an M-word, so Uniqueness does not apply and full agreement persists from syntax to PF ((1)). Uniqueness is also predicted not to apply in Amharic if the gerund and auxiliary are unexpectedly separate M-words; this is borne out in relative clauses, where the gerund and the auxiliary are in distinct M-words and each shows subject agreement (Leslau 1995: 390).

4. Conclusion. We conclude by showing how this approach is more effective than previous proposals (Diertani and Eilam 2010; Kramer To appear) and how it extends to the Amharic complex imperfective.

References

Aoun, Joseph, Elabbas Benmamoun, and Dominique Sportiche. 1994. Agreement, word order, and conjunction in some varieties of Arabic. *Linguistic Inquiry* 195–220.

Diertani, C.E.A., and Aviad Eilam. 2010. How Amharic deals with multiple exponence. Unpublished Ms., LSA 2010.

Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–595.

Kramer, Ruth. To appear. Disappearing subject agreement in Amharic: A Local Dislocation analysis. In *Perspectives on Templatic Morphology*, ed. Ronny Meyer and Chris Reintges.

Leslau, Wolf. 1995. *Reference Grammar of Amharic*. Wiesbaden: Harrassowitz.

Sande, Hannah, and Andrew Hedding. 2017. Syllable weight in Amharic. In *Syllable Weight in African Languages*, ed. Paul Newman, 69–82. Amsterdam: John Benjamins.

Realizing syntactic domains: How synthetic is the Semitic passive?

Hagit Borer
(QMUL)

Keywords: periphrasis vs. synthesis, passive, locality, realization domains, phonological word

In a broad number of languages, including some that do not particular favour periphrasis, verbal passive is periphrastic, with an auxiliary that supports tense alongside non-finite form of the verb. Intuitively, it could be proposed that passive voice (P-Vc) defines a domain for phonological realization, thereby preventing a shared realization between the verbal form and T. An auxiliary is consequently necessary to support tense features.

The approach runs into conceptual problems, however, for languages such as Hebrew and Modern Standard Arabic (MSA), where passive, otherwise syntactically similar to that attested in English or Italian, nonetheless gives rise to a *single* realization (a single phonological word, PW) which includes tense features, and without periphrasis. The theoretical conundrum is self-evident. Should we abandon the claim that the realization domain for P-Vc is syntactically defined, or are we, conversely, to endorse the view that the syntax of passive in Hebrew and MSA is distinct from that attested in periphrastic passives? Without closer investigation no conclusion can be safely reached.

Based primarily on morpho-phonological selection and on locality effects, I will propose that contrary to appearance, P-Vc *does* define a realization domain in both Hebrew and MSA. However, while the domains are the same, the realizations in Hebrew and MSA are distinct. In Hebrew, the vocalic realization features of the passive verb are determined within the syntactic domain contained by P-Vc. In MSA, on the other hand, the realization excludes the complement of P-Vc. In both cases, we note, P-Vc acts as a barrier for a single realization, but is included in the realization (verbal spine simplified):

1)   Realization domains:
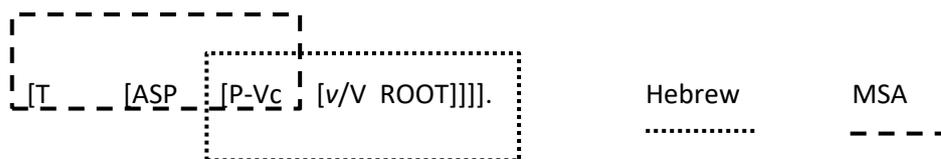
A number of important theoretical conclusions follow:

A.   Locality conditions on realization should be viewed in terms of domains within the spine, each of which may include more than a single functional head (see also Merchant 2015).

B.   While vocalic values within the Semitic verbal system are selected in syntactic contexts, the prosodic interleaving of consonants and vowels is post-syntactic and purely phonological. The emerging PW may thus mask syntactic complexity, including the presence of distinct, syntactically defined, realization domains.

C.   If P-Vc defines a realization domain, it must be syntactically distinct from non-active *Voice* (cf. Alexiadou and Doron 2012 i.a.), including unaccusatives, which does not define such a domain.

D   There is no correlation between PWs and syntactic structures. A single PW may certainly correspond to a complex structure. More importantly, a complex syntactic structure not only need not correspond to a single PW, but it may also correspond to an *impossible* PW.

References

Alexiadou, A. and E. Doron (2012), The syntactic construction of two non-active Voices: Passive and middle. *Journal of Linguistics* 48(01)

Merchant, J. (2015), How much context is enough? Two cases of span-conditioned stem allomorphy. *Linguistic Inquiry* Vol. 46, No. 2 (Spring 2015), pp. 273-303

LexiVault: A web tool for supporting research on the Arabic lexicon

Samantha Wray & Hind Saddiki
(Dartmouth College & New York University Abu Dhabi)

We introduce LexiVault (available at https://github.com/SAVANT-team/LexiVault) an open-source web tool with annotated lexicons and rich retrieval capabilities designed for the support of research into lexical statistics of less-studied languages, including Arabic. LexiVault aims to fill in the gap left between the current state-of-the-art tools and researcher needs. It aims to do this by focusing on multi-dialectal capabilities to enable research that focuses on the Arabic lexicon at the level of the colloquial dialect, or even cross-dialectal research. This goal aims to address the gap in Arabic lexicon tools since the introduction of the previously state-of-the-art tool AraLex (Boudelaa and Marslen-Wilson 2010) in the years since its release, as AraLex focused on Modern Standard Arabic. Although the number of available colloquial dialect corpora has exploded in the years since (see Ahmed et al. 2022 for a comprehensive review), these corpora still require additional processing at the hand of the researcher if they are focused on fine-grained elements of the lexicon that empower research at the morphological and phonological level.

LexiVault contains corpus-sourced Arabic lexicons annotated with word and morpheme frequency, part-of-speech tags, root-and-pattern information, transcription in the International Phonetic Alphabet (IPA), character and phoneme bigram and trigram frequency, minimal pair/phonological neighborhood information, and gloss in English (measures provided by Camel-Tools: Obeid et al. 2020 and Phonological Corpus tools: Hall et al. 2019). With LexiVault, it is possible to utilize the user-friendly search interface to generate data addressing the following example inquiries: Which roots appear in which patterns? Which phonemes appear most often in which root positions? How likely is it a given word contains a root vs. other words with that root (i.e. stem:whole word transition probability)? What are some minimal pairs along the dimension of a phonetic feature such as pharyngealization? How likely is it a given root appears in nouns more often than in verbs? Typically, researchers of Arabic interested in distributional properties of the lexicon would need to calculate these measures from scratch given a dictionary. The launch of LexiVault is intended to eliminate this challenge and allow generation of lexical statistics that support research in morphology, phonology, psycholinguistic stimuli design, and other domains.

LexiVault currently includes multiple major dialect groups of Arabic, including Modern Standard Arabic, Egyptian, and Gulf. Additional dialects, including Moroccan/Maghrebi and Levantine, are also in development. The tool is designed to be friendly to growth of additional dialects, and will accommodate additional lexicons from both the maintaining researchers and community users who are able to contribute. A tutorial and pipeline for the onboarding of new lexical materials is in development to facilitate researcher contributions and ensure depth and breadth of dialectal coverage.

References:

Ahmed, A., Ali, N., Alzubaidi, M., Zaghouani, W., Abd-alrazaq, A. A., & Househ, M. (2022). Freely available Arabic corpora: A scoping review. Computer Methods and Programs in Biomedicine Update, 2, 100049.

Boudelaa, S., & Marslen-Wilson, W. D. (2010). Aralex: A lexical database for modern standard Arabic. Behavior Research Methods, 42(2), 481-487.

Hall, K. C., Mackie, J. S., & Lo, R. Y. H. (2019). Phonological CorpusTools: Software for doing phonological analysis on transcribed corpora. International Journal of Corpus Linguistics, 24(4), 522-535.

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020, May). CAMeL tools: An open source python toolkit for Arabic natural language processing. In Proceedings of the Twelfth language resources and evaluation conference (pp. 7022-7032).

Comparative formation in Maltese

Michela Vella & Margherita Pallottino
(University of Malta & OFAI)

Keywords: comparatives; analytic forms; synthetic forms; triliteral roots; derivational rules

Lexical classes in Arabic varieties are characterized by a templatic structure, and comparative formation in the adjectival system represents a prototypical case of this property. The comparative formation primarily follows the template vCCvC as illustrated in 1):

1)   sˤaɣīr   ʔasˤɣar
        small   *smaller*

In Maltese, a language originally derived from Maghrebi Arabic and extensively influenced by Sicilian, Italian and English, two comparative formations co-exist: synthetic formations, exemplified in 2), which take the Arabic templatic pattern vCCvC/vCCv/vCvCC, and analytic formations, illustrated by 3), consisting of a periphrastic construction whereby the comparative adverb *aktar/iktar/iżjed* 'more' precedes the adjective in its basic form.

2)   sabiħ   isbaħ
        beautiful   *more beautiful*
3)   intelliġenti   aktar/iktar/iżjed intelliġenti
        intelligent   *more intelligent*

Maltese adjectives of Arabic origin have both periphrastic and synthetic comparative formations whereas those of Sicilian, Italian and English origin have the periphrastic comparative form only.

The present study focuses on a limited class of adjectives derived from Sicilian, Italian and English which do not comply with this generalization since they also take the synthetic comparative. Adjectives belonging to this class are found in Standard Maltese (Aquilina, 1987-1991), see 4); dialects spoken in various parts of Malta and Gozo (Aquilina, 1987-1991; Spagnol, 2023), see 5); and Maltraljan – the Maltese variety spoken in Australia (Bovingdon, 2001), see 6) and 7):

4)   fin   ifjen
        fine   *finer*
5)   baxx   abaxx/ibaxx, ibxax/ibxex, ibjax
        low   *lower*
6)   bravu   ibrav
        clever   *cleverer*
7)   sniki   isnek
        sneaky   *sneakier*

The overarching question we delve into is: Which property/ies allow Maltese adjectives of Sicilian, Italian and English origin to follow the synthetic comparative pattern generally restricted to those of Arabic origin?

Based on corpus evidence (Korpus Malti 3.0), we propose that the choice of the comparative form in Maltese is determined by the phonological structure of the adjective in question: adjectives of non-Arabic origin take the synthetic pattern whenever they can be reduced to abstract triliteral roots. The root is then fed to the templatic derivation which results in the formation of a synthetic comparative. We claim that the availability of this abstract base allows the derivation of a synthetic comparative form, and that the base form of the adjective does not play a role in the derivation. Compelling evidence of different kinds supports this analysis, for example the fact that a single adjective, such as *baxx* 'low' in 5), can have different synthetic forms depending on what a speaker identifies as the abstract root level, whether containing a glide (i.e., BJX) or a geminate consonant (i.e., BXX).

Multiple templatic realizations from a single base are only possible if we assume that comparative derivations in Maltese proceed from the root to the template, in line with what Davis (2016) claims for Egyptian comparative forms, and against Ratcliff's (1998) assumption that Arabic derivational rules are in all cases operations on words. As for periphrastic constructions, they do not entail any morphological derivation and, therefore, they are unrestricted and represent the only comparative form for adjectives which are incompatible with a reduction to a triconsonantal root.

References

Aquilina, J. (1987-1991). *Maltese-English dictionary*. Midsea Books.

Bovingdon, R. (2001). *The Maltese language of Australia, Maltraljan: A lexical compilation with linguistic notations & a social, political and historical background.* Lincom Europa.

Davis, S. (2016). The Arabic comparative and the nature of templatic mapping in Arabic. *Word Formation across languages*, 73-90.

Gatt, A., Micallef, K., Tanti, M., van der Plas, L. and Borg C. (2016). *Korpus Malti 3.0* (https://mlrs.research.um.edu.mt/CQPweb/)

Ratcliffe, R. (1998). *The "broken" plural problem in Arabic and comparative semitic: Allomorphy and analogy in non-concatenative morphology*. John Benjamins.

Spagnol, M. (2023). Ngħidlek id-dritt nitkellem bl-imgħawweġ. *Minn tagħna: Ġabra ta' artikli fuq il-Malti, 1*, 30-31.

A rooted classification of Arabic perception events

Huda Salem Taha
(UAE University)

Studies of lexical knowledge in natural languages, including perception (sense, and cognition), often make use of taxonomies or typologies based on the verb *category*, segregating the information it carries from that of its satellites (Viberg 1983, 2015, Gisborne 2010, van Gelderen 2018, along Talmy's 2000 verb-framed/satellite types, or enriched variants). Lexical items are taken to be *words*, already labeled with a *category*. We argue instead that a *root-based* approach to the lexicon, e.g. the minimalist distributed generative engine, as modeled by Chomsky (1995), Marantz & Halle (1993), Hale & Keyser (2002), Harley (2014), Marantz (2015), or Borer (2005, 2014), presents empirical and theoretical advantages. The lexical item is then first borne as a root, with a gradual growing complexity built in the syntactic tree. Affixes are root-adjoined in RootP, before categories like v, n, or a project vP, nP, or aP at a higher level (Lowenstamm 2012). Meanings are construed and constructed in syntax.

One advantage of this Root/'Semitic' model is its ability to get (compositionally) the right meaning at the right stage, by basic or augmented roots, long before categorization (Levinson 2014). Consider the four 'polysemous' meanings of the English root *taste*:

(1) Peter tasted the food.   EXPERIENCE (non-volitional)
(2) Peter was testing the food (on purpose). ACTIVITY/ACTION (agentive)
(3) The food tastes good.   STATIVE
(4) I (can) taste correctly.   ABILITY

Some of these meanings are equally available for nouns or adjectives (a good *taste*, a *taste*r, a *tasteful* meal).

The situation in Arabic is totally parrallel, independently of the category, although (a) morphology (b) category syntax and (c) vocabulary realization can vary. For example, to express the perceptual vision SEE, the two basic vocabulary roots *raʔay* and *naḏar* alternate (in addition to many others). We assume the Distributed Morphology (DM) separation between *lexical items*, processed early in the derivation, and *vocabulary items*, inserted late in the derivation. Examples (5)-(8) illustrate the four senses which parallel those of English above (with (6) as equivalent to (2) being cognitive, rather than perceptual action):

(5) raʔaa   ṭ-ṭuyyur-a   EXPERIENCE
    see.PAST.3 the-birds-ACC
    He saw the birds.

(6) r-t-aʔaa   l-ḥall-a   COGNITIVE ACTIVITY
    see-REFL-PAST.3 the-solution- ACC
    He envisioned the solution.

(7) ta-raaʔaa   ġaaḍib-an   STATIVE
    REFL-see.PAST.3 angry- ACC
    He looked angry.

(8) l-qiṭaṭ-u   t-araa   fii ḍ-ḍalaam-i. ABILITY
    the-cats-NOM FEM-see in the-dark-GEN
    Cats see in the dark.

The sensory activity is only available in (9), with a distinct root:

(9) naḍara   ʔilaa ṭ-ṭuyyur-i   SENSORY ACTION

He looked at the birds.

Arabic varieties such as Emirati, Iraqi, Moroccan, or others, tend to use only one vocabulary item *šaaf*, with distinct syntactic complementation:

(10) šef-t   l-bent

saw-I   the-girl

I saw the girl

(11) šef-t   fe-l-bent

saw-I   at-the-girl

I looked at the girl.

When *šaaf* is found in Standard Arabic, it is only in MANNER use. Deverbal noun *šawf* or participle *šaayef* are lately categorized. Universal potential lexicalizations are then available, subject to relevant externalizations or variation parameters (Chomsky 1995, Kayne 2005, Rizzi & Cinque 2016, among others).

These syntactically based senses are appropriately accounted for within the *Arabic Constructional and Variational Lexicon*, a generative cognitive model of lexical competence designed for describing and processing Arabic lexemes and words, using a DM style design (Fassi Fehri & Salem-Taha 2021, Fassi Fehri et al 2021).

References

Chomsky, N. *The Minimalist Program*. Cambridge: The MIT Press.

Hale, K. & J. Keyser. 2002. *Prolegomena to a Theory of Argument Structure*. Cambridge MA: MIT Press. Halle, M. & A. Marantz. 1993. Distributed morphology and the pieces of Inflection. In The view from building 20, eds. K. Hale & J. Keyser, pp. 111–176. Cambridge, MA: MIT Press.

Harley, H. 2014. On the identity of roots. Theoretical Linguistics 40.3–4: 225–276.

Gisborne, N. 2010. *The Event Structure of Perception Verbs*. Oxford: OUP.

Marantz, Alec. 1997. No escape from syntax. UWPL 4: 201-225.

Marantz, Alec. 2001. Phases and Words. Ms. New York University.

Talmy, L. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

Viberg, A. 1983. The verbs of perception: A typological study. In B. Butterworth, B. Comrie, & Ö. Dahl eds. Explanations for language universals, 123-162. Berlin: de Gruyter.

Fassi Fehri, A. & Salem-Taha, H. 2021. أوضاع الإدراك في المعجم العربي البنائي [Perception Eventualities in the Arabic Constructional and Variational Lexicon]. *Lisaaniyaat Arabiyyah [The Arabic Linguistics Journal]*, 13, 7-49.

Fassi Fehri, A. & others. 2021. المعجم العربي البنائي التنوعي: أسسه ونماذجه وقضاياه [the *Arabic Constructional and Variational Lexicon*]. Amman: Dar Konoz Almaarifa.

Fassi Fehri, A. & Salem-Taha, H. 2023. المفردات والعبارات النفسية العربية: الأصناف المعجمية، والتراكيب، والدلالات. [Arabic psych words and expressions: lexical classes, syntax, and semantics]. *Lisaaniyaat Arabiyyah [The Arabic Linguistics Journal]*17. 9-60.

Lexicalization Patterns of Prepositions and Axial Nouns in Arabic

Maather Alrawi & Nadia Aamiri
(King Abdulaziz University & Mohammed V University)

Lexicalization patterns of prepositional and axial noun expressions are hardly studied across languages, and typically in Arabic, compared to Germanic, Romance, or Slavic, along the lines of other categories (Talmy 1985). Arabic prepositions are arguably of a dual life, born as a lexical *root,* and then derived as a *category p* (case marker). A Place root such as *ʕlw* functions as a preposition *ʕalaa* 'on', as a noun *ʕulaa* 'greatness', as a verb *ʕalaa* 'make higher', or a comparative adjective *ʔaʕlaa* 'higher'. A closely related Place root *fwq* can function as a axial noun *fawqa* 'over', a verb *faaqa* 'exceed', or an adjective *fawq-iy* 'upper'. Both *ʕalaa* 'on' and *fawqa* 'over' express a first meaning related to height/superiority that makes them interchangeable in contests like (1a) or (1b):

(1) a. *l-kitaab-u   ʕalaa ṭ-ṭaawilat-i*   b. *l-kitaab-u   fawq-a ṭ-ṭaawilat-i*
      the-book-NOM on   the-table-GEN   the-book-NOM over-ACC the-table-GEN
      'The book is on the table.'   'The book is on the table.'

But the preposition and the axial noun carry other senses, like *projective*, *contact*, or temporal meanings that distinguish them, as illustrated in (2) and (3).

(2)   *marrat-i ṭ-ṭaaʔirat-u   fawq-a/*ʕalaa   l-jisr-i*
       passed-G the-plane-NOM over-ACC/on   the-bridge
       'The plane passed over the bridge.'
(3)   daxal-tu   *ʕalaa/*fawq-a s-saaʕat-i   r-raabiʕat-i*
       entered-I on/over-ACC   the-hour-GEN the-four-GEN
       'I entered at four o'clock.'

In this paper, we argue for a split between l-syntax and f-syntax, motivating a p-over-√ structure, inspired by the Distributed Morphology (DM) approach, a basic tenet of which is the separation of the root from the category (Svenonius 2010; Wood & Marantz 2017), in a sort of Larsonian (1988) 'shell structure', or Fassi Fehri's (2021) 'distributed lexicon'. The projections pP and √P correspond to the traditional single projection Place (or Path) (den Dikken 2003, 2010; Gehrke 2007, 2008; Koopman 2000; Svenonius 2008, 2010 ; Cinque 2010, among others) which is challenged in Arabic. The functional p head, as a case assignor, embeds a lexical Root √ head, which is endowed with semantic specifications:

(4)
```
        pP
      /    \
     p     √P
          /   \
         √     XP
```
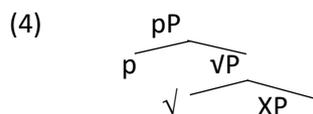
Our analysis neatly accounts for the tripartite locational/directional/motional classification attested in Arabic (Fassi Fehri & Alrawi 2023), in contrast to bipartite locational/directional found in other languages (see Gehrke 2008; Svenonius 2010; Cinque 2010, among others), and for the fine-grained distinction between synonymous prepositions such as *li-* and *Ɂilaa*, as well as other geometric specifications proposed in Herkovits (1986), Landau & Jackendoff (1993), among others. It adequately deals with the complex semantics of Arabic spatial prepositions and axial nouns that have been widely analyzed from a cognitive perspective (Jan 2018; Lentzner 1977; Esseesy 2010), but still lack more precise and elaborate characterization in terms of morpho-syntax as well as semantics. Our DM alternative, enriched with a fine-grained cartography, offers a more viable comparative account that can be theoretically and cross-linguistically motivated.

References

Cinque, G. 2010. Mapping Spatial PPs: An Introduction. In *Mapping Spatial PPs. The Cartography of Syntactic Structures* 6, eds. G. Cinque, L. Rizzi, 3–25. New York: Oxford University Press.

den Dikken, M. 2003. On the syntax of locative and directional adpositional phrases. Ms., CUNY.

den Dikken, M. 2010. On the functional structure of locative and directional PPs. In The Cartography of Syntactic Structure, vol. 6, ed. by Guglielmo Cinque and Luigi Rizzi, 74–126. New York: Oxford University Press.

Esseesy, M. 2010. *Grammaticalization of Arabic Prepositions and Subordinators.* Leiden: Brill.

Fassi Fehri, A. 1986. *The Arabic Lexicon* [*l-muʕjam l-ʕarabii*]. Casablanca: Tubqal Publishers.

Fassi Fehri, A. Alrawi, M. 2023. Arabic PPs in a Rooted Lexicon. Languages 8(2), 95: 1-23.

Gehrke, B. 2007. On directional readings of locative prepositions. In Proceedings of ConSOLE XIV, ed. by Sylvia Blaho, Luis Vicente and Erik Schoorlemmer, 99–120.

Gehrke, B. 2008. Ps in Motion: on the Semantics and Syntax of P Elements and Motion Events. Doctoral dissertation, Utrecht University.

Herkovits, A. 1986. *Language and spatial cognition*. Cambridge: Cambridge University Press.

Jan, H. 2018. *A cognitive linguistic approach to explaining the polysemy of 'alaa and fii in Modern Standard Arabic*. PhD thesis, Georgetown University, Washington, D.C.

Koopman, H. 2000. Prepositions, postpositions, circumpositions, and particles. In The Syntax of Specifiers and Heads, ed. by Hilda Koopman, 204–260. London: Routledge.

Lentzner, K. 1977. *Semantic and syntactic aspects of Arabic prepositions*. Doctoral dissertation, Georgetown University, Washington, D.C.

Svenonius, P. 2008. Projections of P. In Syntax and Semantics of Spatial P, ed. by Anna Asbury, Jakub Dotlacil, Berit Gehrke and Rick Nouwen, 63–84. Amsterdam: John Benjamins.

Svenonius, P. 2010. Spatial P in English. In *Mapping Spatial PPs, The Cartography of Syntactic Structures,* eds. G. Cinque & L. Rizzi, 127-160. Oxford: OUP.

Talmy, L. 1985. Lexicalization patterns: Semantic structure in lexical forms. In *Language Typology and Syntactic Description*, ed. T. Shopen, 36-149. Cambridge: Cambridge University Press.

Arabic masdar formation from a root-based approach

Bashayer Alotaibi
(Kuwait University)

Masdar is an event naming noun. It has been categorised as a derived nominal, which starts off its derivation as a verb and merges with a nominal affix at some point in the syntax. (Fassi Fehri 1993, Kremers 2003). However, masdars don't exhibit unified morphological or syntactic behaviours in Arabic – of which I will give two examples - and hence are problematic given previous theoretical approaches. In this study, I aim to show that a more systematic analysis can be given to the formation of masdar in Arabic through the Arabic constructional and variational lexicon framework (see Fassi Fehri et al. 2021). This framework is root-based as it takes into consideration that roots have semantic properties, and they are not necessarily simple syntactic units. They can be rather phrasal and create complex root constructions (in a root syntax) prior to categorization (in category syntax). A root-based analysis of masdar is motivated by interesting findings in masdar formation in Arabic, the first of which is the formation of the canonical triliteral form CaCC which appears to be sensitive to semantic event types. Fassi Fehri & Vinet (2008) show that the canonical form of the Masdar CaCC is available only for activities and accomplishments (1-2), but not achievements or states (3-4) as in the following examples:

1-	Jaraa	jary-an (activity)	2-	ʔakala	ʔakl-an (accomplishment)
	ran	run-ACC		eat	eat-ACC
	'he ran a running'			'he ate an eating'	
3-	balaġa	*balġ-an/ buluuġ-an (achievement)	4-	kariha	*karh-an/kurh-an (state)
	reach	reach-ACC		hate	hate-ACC
	'reach a reaching'			'hate a hating'	

In (3-4), the canonical form is not felicitous with achievements and states, and therefore, another 'eventuality noun' form is used to substitute for the unavailable canonical form. The form available for achievements (CuCuuC) is considered an internal (broken) plural form, which functions to express extension, progression or continuation, while the form for states is an eventuality noun (Fassi Fehri & Vinet, 2008: 65-66). Such facts are better explained in relation to event semantics, and properties of the event encoded in its root.

The second line of evidence comes from the difference between two types of masdars: 'result nominals' and 'process nominals' (Fassi Fehri 1993: 235-237). Result nominals can conjugate in the dual, plural, or diminutive forms, can be preceded by demonstratives, but cannot theta mark arguments (5a). Process nominals are to the opposite (5b). These differences are assumed not to be related directly to the root (see Borer 2014), as there are many cases of the same masdar being used as a result nominal in some constructions, and as a process nominal in others as shown in the following example with masdar *naqd* 'criticism' (Fassi Fehri 1993:236):

5-	a.	haaḏaa	nuqayd-un	ġariib-un
		this	criticism.diminutive-nom	strange-nom
		This is a (little) strange criticism.		

b.	* nuqayd-u	zayd-in	hind-an	ġariib-un
	criticism.diminutive-nom	Zayd-gen	Hind-ace	strange-nom
	Zayd's (little) criticism of Hind is strange.			

However, in a phrasal root construction the difference between process and result nominals can be accounted for within the root phrase structure, given the availability of different flavours of event root heads which may merge with the phonological root prior to categorization, as this study will show.

References

Borer, H. (2014). Derived nominals and the domain of content. *Lingua*, *141*, 71-96.

Fassi Fehri , Abdelkader. 1993. *Issues in the Structure of Arabic Clauses and Words*. Dordrecht: Kluwer Academic Publishers.

Fassi Fehri, A., & Vinet, M. T. (2008). Nominal and verbal classes in Arabic and Chinese. *Aspect et pluralité d'événements*, 55-83.

Fassi Fehri, A. et al. 2021. *al-muʕjam al-ʕarabii al-binaaʔii t-tanawwuʔii [The Arabic Constructional and Variational Lexicon]*. Amman: Daar Kunuuz Publishers.

Kremers, J. (2007). Masdar formation. In E. Ditters & H. Motzki (Eds), *Approaches to Arabic linguistics* (pp. 475-499). Brill.

Arabic psych nominals and their adicity

Abdelkader Fassi Fehri

(Linguistic Society of Morocco; Mohammed V University)

Event psych nominals (or 'nominalizations') such as *kurh* 'hate' *maqt* 'despise' basically mirror the structures of their SExp (Subject Experiencer) verbal counterparts, and in particular the fact that they can be dyadic or 'transitive'. Their subject is then marked as genitive, and their object can bear a morphological accusative, or a prepositional objective *li-* case, as in
(1); the latter paralleling the structure of the action nominal in (2); Fassi Fehri (1993):

(1) *kurh-u/maqt-u   r-rajul-I   l-ʕamal-a;   li-l-ʕamal-i*
     hate-nom/despise-nom the-man-gen the-work-acc; of-the-work-gen
     'The man's hate/despise of-the-work.'

(2) *hadm-u   l-ʕaduww-i   l-madiinat-a;   li-l-madiinat-i*
     destruction-nom the-enemy-gen the-city-acc;   of-the-city-gen
     'The enemy's destruction of the city'

But OExp (Object Experiencer) verbs *šaġala* 'preoccupy', *dahaša* 'astonish', *sarra* 'please, have no such dyadic option for psych nominals; they can only be monadic, which mirrors solely the structure of the intransitive SExp predicate (Iordachioaia 2019; Fassi Fehri 2023; for more complexity, Belletti & Rizzi 1988, Landau 2010, Pesetsky1995). Hence the nominal OExp nominal construction in (4) is ungrammatical, compared to the verbal one in (3):

(3) *y-ašġal-u l-xabar-u   l-mudiir-a*
     3-occupy the-news-nom   the- director-acc
     'The news preoccupies the director'.

(4) **šuġl-u   l-xabar-i   l-mudiir-a ;   li-l-mudiir-i*
     preoccupation-nom the-news-gen   the-director-acc; of-the-director-gen
     Intended to mean: 'The news's preoccupation (of) the director'.

Only the intransitive SExp construction is then possible in this case, in the form of the simple intransitive in (5a), or the inchoative-reflexive in (5b):

(5) a. *suruur-u   l-mudiir-i   (bi-l-ʔamr-i)*
        happiness-nom the-director-gen   with-the-matter-gen
        'The director's happineness (with the matter)'.
     b. *n-šiġaal-u   l-mudiir-i   bi-l-ʔamr-i*
        inch-preoccupation-nom   the-director-gen   with-the-matter-gen
        'The director's preoccupation with the matter'.

Why is this so? Suppose that OExp verb phrases are kinds of 'apophonic' derived transitives or causatives (Fassi Fehri 1987, 2012), then the following descriptive statement holds:

(6) Derived (apophonic) psych transitives do not have well-formed nominal counterparts.

The puzzle now is that other derived complex dyadic psych nominals are productively felicitous. For example, the verbal glottal causative in (7a), or the geminated causative in (7b), have well-formed nominal dyadic parallels in (8), without posing any problem:

(7) a. *ʔ-adhaša*   *r-rajul-u*   *l-walad-a*
     caus-astonish.past   the-man-nom the-child-acc
     'The man astonished the child'.
     b. *sallaa*   *r-rajul-u*   *l-walad-a*
     caus-amuse.past   the-man-nom   the-child-acc
     'The man amused the child'.

(8) a. *ʔ-idhaaš-u*   *r-rajul-i*   *l-walad-a*
     caus-astonishment-nom   the-man-gen   the-child-acc
     'The man's astonishment of the child'.
     b. *tasliyat-u*   *r-rajul-i*   *li-l-walad-i*
     caus.amusement-nom   the-man-gen   of-the-child-gen
     'The man's amusement of the child'.

The difference in behavior between (8) and (4) can be traced, we argue, to the distinct nature of the source of the causative/transitive. Thus, while the source of both verbal and nominal contructions in (7) and (8) is a *complex root*, combing a lexical root and a causative root (Lowenstamm 2014), the apophonic affix in (3) is *templatic*, i.e. limited to the verbal category (hence excluding nominals in (4)). Consequently, all psych nominals turn out to be derived from their simple or complex roots, and sharing their argument structure, while only the psych verb derivation involves templatic (in addition to root) affixation (for more complexity Marantz 2022, Harley 2014, Borer 2014, Hale & Keyser. 2002). The analysis is implemented in a minimalist distributed morphology and templatic syntax model (Chomsky 1995, Halle & Marantz 1993, Fassi Fehri 2023).

References

Belletti A. & Rizzi L. 1988. Psych verbs and θ-Theory. *Natural Language and Linguistic Theory* 6: 291-352.

Borer, H. 2014. Derived nominals and the domain of content. *Lingua* 141: 71-96.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.

Fassi Fehri, A. 1987. Anti-Causatives in Arabic, Causativity, and Affectedness. *Lexicon Project Working Paper* 15, 1-56. Cambridge MA: MIT Center for Cognitive Science.

Fassi Fehri, A. 2012. *Key features and parameters in Arabic grammar*. Amsterdam: John Benjamins.

Fassi Fehri, A. 2023. Psych Construction Types in Arabic as Root-based. Paper presented at SLE 56. Athens: Univ. of Athens. *Under review*.

Hale, K. & Keyser, S. J. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge, MA: The MIT Press.

Halle, M. & Marantz, A. 1993. Distributed morphology and the pieces of inflection. In Hale,

K. & Keyser, S.J. eds*. The view from building 20.* Cambridge MA: The MIT Press, 111- 176.

Harley, H. 2014. On the identity of roots. *Theoretical Linguistics* 40.3-4: 225-276.

Iordachioaia, G. 2019. The root derivation of psych nominals: Implications for competing overt and zero nominalizers. *Bucharest Working Papers in Linguistics* XXI.2: 57-79.

Landau, I. 2010. *The Locative Syntax of Experiencers.* Cambridge, MA: The MIT Press.

Lowenstamm, J. 2014. Derivational affixes as roots. In Alexiadou A. et al eds.*The syntax of roots and the roots of Syntax*. Oxford: Oxford University Press, 230-259.

Marantz, A. 2022. Rethinking the syntactic role of word formation. In <u>Boneh</u>, N. et al. eds. *Construire sur les décombres de Babel*. Paris: Presses universitaires de Vincennes, 293- 316.

Pesetsky, D. 1995. *Zero syntax*. Cambridge, MA: The MIT Press.

WS20 The concept of possibility and its morphological, syntactic and pragmatic realizations in natural language

Irrealis as non-assertion: Evidence from the Biblical Hebrew imperfect

Ariel Cohen
(Ben-Gurion University of the Negev)

The nature of the realis/irrealis distinction is debated: some, like Palmer (1986), have proposed that it indicates the dichotomy factual/non-factual; but later, Palmer (2001) proposed that the relevant distinction is, in fact, assertion/non-assertion.

Evidence from Biblical Hebrew can help decide between these two views. Based on comparisons with irrealis forms in several unrelated languages, Grasso (2017) demonstrates that the imperfect in Biblical Hebrew is irrealis (cf. Hatav 1997; Joosten 2012). As predicted by the non-factual theory, the imperfect is often used in non-factual contexts. For example, it can refer to future events as in (1a), to express modality, as in (1b), or habituality, as in (1c).

(1) a.   Ləmāḥār   **yihyeh** hāʾōṯ   hazzeh. (Exodus 8:19)
         for-tomorrow **be-IMPF** the-sign this
         `This sign shall appear tomorrow'

   b.   ʾim yûḵal ʾîš   limnôṯ ʾeṯ ʿăp̄ar   hāʾāreṣ
         if   can   man to-count ECM the-dust-of the-earth
         gam zarʿăḵā   **yimmāneh** (Genesis 13:16)
         also your-seed **counted-IMPF**
         `If one can count the dust of the earth, your offspring also can be counted.'

   c.   waʾănî śənēʾṯîw kî   lōʾ **yiṯnabbēʾ**   ʿālai
         and-I   hate-him because not **prophesy-IMPF** about-me
         ṭôḇ   kî-ʾim   rāʿ (1 Kings 22:8)
         good but-rather bad
         `And I hate him, because he doesn't prophesy good things about me but rather bad things.'

However, the imperfect also occurs in clearly factual contexts. For example:

(2) wayyimṣāʾēhû ʾîš   wəhinnēh ṯōʿeh baśśāḏeh
    found-him   man and-there lost   in-the-field
    wayyišʾālēhû hāʾîš   lēʾmōr   mah **təḇaqqēš**? (Genesis 37:15)
    and-asked-him the-man as-follows what **you-look-for-IMPF**
    `A man found him (Joseph) in the field looking lost. And the man asked him: "What are you looking for?"'

It is clear that Joseph is, in fact, looking for someone, and the man is aware of this.

Factual uses of the imperfect are not restricted to questions, though:

(3)   Ḥîrām meleḵ   ṣōr niśśāʾ   ʾeṯ   šəlōmōh baʿăṣêʾ   ărāzîm
      Hiram king-of Tyre supplied ECM Solomon with-trees-of cedars
      ûḇaʿăṣê   bərôšîm ûḇazzāhāḇ   ləḵol ḥepṣô.
      and-with-trees-of cypresses and-with-gold for-all his-desire.
      ʾāz **yittēn**   hammeleḵ šəlōmōh ləḥîrām ʿeśrîm ʿîr   bəʾereṣ
      then **give-IMPG** the-king   Solomon to-Hiram twenty town in-the-land-of
      haggālîl (1 Kings 9:11–13)
      Galilee
      `Hiram, King of Tyre, supplied Solomon with cedar trees and cypress trees and gold, as much as he desired. Then King Solomon gave Hiram twenty towns in Galilee'

Solomon's giving the towns to Hiram is clearly described as factual.

Such examples are problematic for the non-factual theory of irrealis. One would be forced either to assume an ambiguous imperfect, or to extend the notion of non-factuality implausibly. But the non-assertion theory of irrealis fares much better.

Regarding (2), note that the question does not assert that Joseph is looking for someone; it only presupposes it. Thus, although the imperfect refers to a factual event, it is not asserted. In (3), the imperfect is preceded by ʾāz. In Biblical Hebrew, "p ʾāz q", usually translated "p, then q", can also mean "p, so q". Strawson (1952) explains the meaning of *so* as follows: "If one statement is a ground for another and we believe the first statement to be true, we are justified in saying something of the form `p, so q'." Thus, (3) is actually saying that Hiram's giving Solomon precious gifts is grounds for Solomon's giving him twenty towns, and that Hiram did give Solomon these gifts. A trivial bit of inference allows us to conclude that Solomon did give Hiram twenty towns; but, crucially, although factual, this statement is not actually asserted, but rather entailed.

Hence, the reason why the imperfect is acceptable in (2) and (3) is that, although it describes a factual event, this event is not asserted: it is presupposed in (2) and entailed in (3). Thus, the Biblical Hebrew imperfect provides evidence for the view of irrealis as non-assertion rather than non-factuality.

References
Grasso, Kevin (2017), Yiqtol as an Irrealis-Imperfective form. *Annual Meeting for the Society of Biblical Literature*, Boston, Massachusetts.
Hatav, Galia (1997), *The Semantics of Aspect and Modality: Evidence from English and Biblical Hebrew*. Philadelphia: John Benjamins.
Joosten, Jan (2012), *The Verbal System of Biblical Hebrew: A New Synthesis Elaborated on the Basis of Classical Prose.* Jerusalem: Simor Publishing
Palmer, Frank R. (1986), *Mood and Modality*, Cambridge: Cambridge University Press.
Palmer, Frank R. (2001), *Mood and Modality*, second edn, Cambridge: Cambridge University Press.
Strawson, Peter F. (1952), *Introduction to Logical Theory*. London: Methuen.

Subjective and objective possibility in Ainu, Japanese, and Korean

Katsunobu Izutsu, Mitsuko Izutsu & Yongtaek Kim
(Hokkaido University of Education, Fuji Women's University & Georgia Institute of Technology)

Subjective and objective modality have been a widely accepted dichotomy, for which diagnostic tests have also been proposed (e.g., Hengeveld 1988, Halliday and Matthiessen 2004, Larm 2014). However, the subjectivity and objectivity in question have not been well-articulated. What does the "subjective/objective" distinction mean in the concept of possibility? Are there some more features related to the distinction and the concept?

This paper makes four points drawing on Ainu-Japanese parallel texts (ILT 1984-1989) and their Korean translations. First, and most importantly, we propose a newly defined distinction between subjective and objective possibility, in which possibility judgment is grounded on the "subject and object of conception" (Langacker 2008: 468), respectively. In a lexical contrast like *It may*[*can*] *be bitter cold in Carlyle*, the possibility of the event or state referred to ('getting chilly') is attributed to the speaker's weak anticipation of the event/state, as in *may*, and to some characteristic of a participant in the event/state (the climate of *Carlyle*), as in *can*. The subjective/objective distinction can also manifest as a constructional contrast, as in *It is possible that (Carlyle is...)* and *It is possible for (Carlyle) to (be...)* (Leech 1987: 81).

Secondly, we show that expressions of subjective possibility outnumber those of objective possibility and that the latter exhibit a strong tendency to develop into the former. The three Asian languages have diverse expressions of possibility: five constructions (tentatively called ontological, nominal, suspicious, emotive, deontic), two verb types (modal, sensory), and two adverb types (interjectional conditional, degree), and final particles. But we find only a few expressions of objective possibility in Japanese and Korean and none in Ainu. Ainu *nankor*, seemingly having initially/literally had an objective meaning ('have the appearance of'), currently has only a subjective meaning ('may/will').

Thirdly, we demonstrate that subjective possibility can be conceptualized in terms of "desirability" (Akatsuka 1999: 196) and that undesirable rather than desirable possibility is dominant in the three languages in varying degrees. The Japanese and Korean modal verbs for 'may' derive from negative verb phrases: *-kamosirenai* '(*lit.*) (it) is not known whether...' and *-ljidomolla* '(*lit.*) (I) don't know whether....' Japanese nominal construction allows for *osore* 'fear,' but not *nozomi/mikomi* 'hope.' Ainu selects *sitoma* 'be afraid of' in an emotive construction, *ikiya* 'so as not to do' as part of the interjectional conditional adverb, and *yak wen* '(It) is bad if...' in a deontic construction.

Finally, we argue that the three languages can vary in linguistic systems and categories of possibility conceptualization. Our observation reveals that they have developed different pragmatic functions in formally comparable expressions. Ainu would appear to be a language in which the category of low probability (glossed felicitously as 'may') is underdeveloped, but the discourse pragmatic function of the category can be somewhat covered by the well-developed system of evidentiality. The variation accounts for some important facts: e.g., Japanese *-kamo-yo* is often, Ainu *nankor* is not, and Korean *-jido* could be employed for warning though they are all final particles (or modal verbs) glossed as 'may.'

References

Akatsuka, Noriko. 1999. Towards a theory of desirability in conditional reasoning. In Akio Kamio and Ken-ichi Takami (eds.), *Function and Structure: In honor of Susumu Kuno*, 195-213. Amsterdam: John Benjamins.

Halliday, M.A.K., and Christian M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar.* London: Hodder Arnold.

Hengeveld, K. 1988. Illocution, mood, and modality in a functional grammar of Spanish. *Journal of Semantics* 6: 227-269.

Institute of Language Teaching (ILT) (ed). 1984-1989. *Ainugo Onsei Shiryo* (Ainu language phonetic materials), Vols.1-6. Tokyo: Waseda University.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction.* Oxford: Oxford University Press.

Larm, Lars. 2014. Modality Packaging in Japanese: The Encoding of Modal Meanings and Subjectivity. *Studies in Pragmatics* 16: 20-46.

Leech, Geoffrey N. 1987. *Meaning and the English Verbs.* 2nd edn. London: Longman.

The expression of epistemic possibility in English, German, Dutch and French: A case study

Tanja Mortelmans
(University of Antwerp)

Keywords: epistemic possibility, parallel corpora, modal verbs, modal adverbs, grammaticalization

It is well-known that English, German, Dutch and French – all members of the Standard Average European *Sprachbund* – feature modal verb systems with different degrees of formal grammaticalization and semantic/pragmatic (inter)subjectification. The English modal verb system stands out as being most strongly grammaticalized, on the one hand, and most strongly geared towards the expression of epistemic modality, which qualifies as inherently strongly subjectified, on the other (see for instance Kranich & Gast 2015). With respect to German, Dutch and French, it has been shown that their modal verbs systems are less grammaticalized (Mortelmans et al. 2009, Cornillie et al. 2009) and seem to specialize more in expressing non-epistemic values. Epistemic uses make up a small minority of all modal verb uses in German, Dutch and French (Diewald 1999, Baumann 2017, Nuyts/ Byloo/ Diepeveen 2010, Cornillie et al. 2009).

If we zoom in at the expression of (subjective) epistemic possibility in English, three verb forms are regularly used to express this notion: *may*, *might* and *could* (Usonienė & Soliene 2010), whereby *could* often expresses objective non-epistemic possibility as well. The other languages only have one modal verb regularly expressing epistemic possibility at their disposal: *pouvoir*, *kunnen*, *können* (Dutch *mogen* and German *mögen* also feature epistemic uses, but with a strong concessive flavour). For French and German, elements of the well-preserved mood system must be considered in this respect as well: both *pouvoir* and *können* can occur either in the indicative or in the *conditionnel* (*pourrait*) or past subjunctive (*könnte*), whereby the latter forms even seem to specialize in expressing epistemic uses (see Diewald 1999).

In my talk, I will present a case study – based on an analysis of a self-compiled parallel English-German-Dutch-French corpus – of how these languages actually deal with the expression of epistemic possibility – in comparison to the elaborated and highly grammaticalized system in English. More concretely, I will compare the use of epistemic *could*, *may* and *might* with their German, Dutch and French translation equivalents. The case study will address the conditions under which correspondences between modal verbs expressing epistemic possibility are found as well the interplay of modal verb use with adverbs of epistemic possibility (like *maybe* and *perhaps* or Dutch *misschien*). One of the preliminary results of this study is the observation that epistemic *might* is not typically rendered by a modal verb, but most often corresponds to an epistemic adverb, as in the following extract from the UK-crime novel 'The Girl on the Train' and its translations in Dutch, German and French.

ENG   The boyfriend's name is not given […]. He **might** not even exist.
DUT   **Misschien** bestaat hij niet eens.
GER   **Vielleicht** gibt es ihn aber auch gar nicht.
FR   **Ou alors** il n'existe pas.

The fact that epistemic possibility adverbs in German, Dutch and French outnumber the corresponding modal verbs expressing epistemic possibility aligns with this. The opposite tendency is found in

English. In my talk, I will thus demonstrate how the four languages have different preferences expressing epistemic possibility and try to account for (some of) the differences found.

References

Baumann, Carolin (2017), *Bedeutung und Gebrauch der deutschen Modalverben. Lexikalische Einheit als Basis kontextueller Vielheit*. Berlin / New York: de Gruyter.

Cornillie, Bert, Walter de Mulder, Tine Van Hecke and Dieter Vermandere (2009), Modals in the Romance languages, in B. Hansen, and F. de Haan (eds.), (2009), *Modals in the Languages of Europe: A Reference Work*, Berlin / New York: De Gruyter Mouton, 107-138.

Diewald, Gabriele (1999), *Die Modalverben im Deutschen*. Tübingen: Niemeyer.

Kranich, Svenja and Volker Gast (2015), Explicitness of epistemic modal marking: Recent changes in British and American English, in J.R. Zamorano-Mansilla, C. Maíz, E. Domínguez, and M. V. Martín de la Rosa (eds), *Thinking Modally: English and Contrastive Studies on Modality,* Cambridge: Cambridge Scholars Publishing, 1-22.

Mortelmans, Tanja, Kasper Boye and Johan van der Auwera (2009), Modals in the Germanic languages, in B. Hansen, and F. de Haan (eds.)*,* (2009), *Modals in the Languages of Europe: A Reference Work*, Berlin / New York: De Gruyter Mouton, 11-70.

Nuyts, Jan, Pieter Byloo and Janneke Diepeveen (2010), On deontic modality, directivity, and mood: The case of Dutch *mogen* and *moeten*, *Journal of Pragmatics* 42, 16–34.

Usonienė, Aurelia and Audrone Soliene (2010). Choice of strategies in realizations of epistemic possibility in English and Lithuanian, *International Journal of Corpus Linguistics* 15(2), 291-316.

From tolerance to commitment: A force-dynamic account of Finnish implicative verbs of ability

Arnaud Godet & Rea Peltola

(University of Caen Normandy, CRISCO & University of Caen Normandy, CRISCO)

Keywords: ability, dynamic possibility, modality, Finnish, Force dynamics

This paper explores the meaning construction of expressions of dynamic possibility from a cognitive-semantic perspective. The focus is on a set of Finnish semi-lexical implicative verbs expressing ability in terms of tolerance (e.g. *sietää*), energy (*jaksaa*), daring (*tohtia*), sensitivity (*ra[a]skia*), and commitment (*malttaa*). They all indicate that it is possible for a state of affairs to take place "given the relevant circumstances" (Kratzer 1991), and despite a counteracting force (Gärdenfors 2024). They are all more or less specific concerning these circumstances (Flint 1980). Moreover, they all display affinity to negative contexts (Kiuru 1977). The paper first tracks the specific force-dynamic pattern underlying the semantics of these verbs (Talmy 1988, 2000). The circumstantial modal base can be of physical (e.g. temperature, space), subjective (attention, emotion), or intersubjective (social norms and relations) type, depending on the lexical component incorporated in the verbs' semantic structure. Some of the verbs cover more than one of these types (*sietää* 'tolerate [physical/ subjective/intersubjective]), some display a more limited lexical scope (*tarjeta* 'tolerate cold'). The paper sheds light on the interconnections between the conceptual forces of different order that underlie the types of ability, from tolerance to commitment, from physical to intersubjective. What does the semantics of these verbs tell us about the position of the schematic structure of ability within the overall ecology of force-dynamic patterns of modality? How can Force dynamics deal with the negative affinity of these verbs? By describing the elaborate system of verbs of dynamic possibility in Finnish and the modal semantics of tolerance, daring and related concepts, the paper contributes to developing "a more sophisticated taxonomy of forces" (Copley 2019). In doing so, it addresses the status of dynamic possibility among modal possibilities. The analysis is based on a corpus of contemporary language use in media and literature, as well as transcriptions of dialectal interviews, available in the Language Bank of Finland (Kielipankki), https://www.kielipankki.fi/language-bank/.

References

Copley, Bridget (2019), Force Dynamics, in R. Truswell (ed.), (2019), *Oxford Handbook of Event Structure*, Oxford: Oxford University Press, 103–149.

Flint, Aili, (1980), *Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency*, Helsinki: Finnish Literature Society.

Gärdenfors, Peter, (2024), Event structure, force dynamics and verb semantics, *Language Sciences* 102, 101610.

Kiuru, Silva, (1977), *Suomen kielen kieltohakuiset verbit*, Helsinki: Finnish Literature Society.

Kratzer, Angelika, (1991), Modality. In A. von Stechow, and D. Wunderlich (eds), (1991), *Handbuch Semantik/Handbook Semantics*, Berlin: de Gruyter, 639–650.

Talmy, Leonard, (1988), Force Dynamics in Language and Cognition, *Cognitive Science* 12, 49–100.

Talmy, Leonard, (2000), *Toward a Cognitive Semantics, Vol 1,* Cambridge, MA: The MIT Press.

Possibility versus irrealis in Maa

Doris Payne
(University of Oregon; SIL International)

Keywords: adverb, diachrony, epistemic, irrealis, Nilotic

"Irrealis" has been defined as that which is "purely in the realm of thought" (Mithun 1999:173). The Maa (Eastern Nilotic) adverb *náají* 'possible, might' may derive from the verb root *jo* 'say', which has additional senses of 'think [something to be the case]/hold an opinion, try'. The 'possible' meaning of *náají* may extend from the 'think' sense via a slight shift in what meaning is profiled, to '(think) something is/could be possibly the case'. If *jo* is the source of *náají,* and if "irrealis" concerns what is only in the realm of thought, it suggests that *náají* is an irrealis operator. However, *náají* need not correlate with (other) irrealis grammar, motivating questions about how coherent "irrealis" is as a domain of natural language.

The frozen adverb form *náají* has functions of a (weak) epistemic operator, as in (1)—(2). In (3) it clearly correlates with something in the realm of thought. It is sometimes translated as 'for example', as in (4), to indicate a hypothetical or generic possibility (rather than a particular referential exemplar).

Structurally, *náají* looks like an impersonal relative clause form of *jo*. A relative form occurs in (5) where the "speech" sense is evident. The impersonal ending *-î* on the relativized verb in (5) developed diachronically from a plural subject pronominal element (Greenberg 1959; the slight tone shift between *náájî* and *náájí* is due to another plural falling-tone morpheme). As (5) shows, relative clauses follow their head noun.

Though *náají* may have developed from relativization of *jo* 'say', in most instances of use *náají* has no relation to an act of saying. Also differently from the relative use, *náají* can follow verbs, as in (1)—(3); this is characteristic of other adverbs.

Under the definition of "irrealis" given above, and if *náají* 'possible' developed from the 'think' sense of *jo*, we might expect it to co-occur with irrealis grammar. The basic Maa irrealis marker is a bipartite 'subjunctive' form *tV-…-a(k)*, seen in the inflected verb *ɛ-**ta**-yíéwū-**ā*** (3-SBJV-want-SBJV) 'he/she wants it'. This subjunctive also occurs in singular imperatives (Tucker & Mpaayei 1955), on verbs in 'until' adverbial clauses, in complements of matrix verbs like 'be able' in (2), 'refuse', 'fail', 'like (to do)', etc.

However, corpus study shows that *náají* co-occurs with both subjunctive-irrealis and non-irrealis verb forms: in (2), *índɪm* 'you can' requires the complement verb *a**ta**rríayu* be in a 'subjunctive' irrealis form; but in (1), *nélo* 's/he goes' and *nɛya* 's/he dies' are both declarative realis verbs. Also, *náají* almost never occurs in counterfactual, negative, or imperative contexts.

Altogether, the data suggest that Maa "possibility" is not grammatically coextensive with the putative "irrealis" domain, calling into question both the definition of "irrealis" and its coherence. The study contributes to our typological knowledge of what 'possible' adverbs can derive from diachronically, in this case, the domain of speech, thought (and perhaps attempt).

Examples

(1) Né-lo **náají** nɛ́-yɛ.
    3-go **possible** 3-die
    'And he might go and die./'It's possible he will go and die.'

(2) Kóre tɛ-ní-réʊ̌ ɛmbáísikil ínó áŋ í-ndɪm **náají** a-**ta**-rría-yu.
    now when-2-drive bike you home 2-can **possible** INF.SG-**SBJV**-fall-INCHOATIV
    'When you ride your bike home, you might/it's possible you will fall.'

(3) Á-yíéú ní-tɛ́m **náají** a-dʊmʊ̌ tan-áa k-é-lotú ɔlápâ ɛnkɔ́p.
    1SG-want 2-consider **possible** INF.SG-come IF-be COMP-3-come moon ground
    'I want you to imagine if the moon will come to the ground.'

(4) peê e-tumokí ɪlɛ́wâ **náají** áa-ɪnɔs-á
    so CONVERB.3-get.opportunity men.NOM **possible** INF.PL-eat-SBJV
    'so men for example can eat [in a house to be cleansed after a woman gives birth]'

(5) inkíshu **náa-j-î** ɪsayíɛ́ta
    cows REL.FEM.PL-say-PL/IMPERSONAL bride.price
    'cows that are called/referred to as (lit. that which is said) "bride-price"'

References

Greenberg, Joseph (1959), The Origin of the Masai Passive, *Africa* 29: 171–176.

Mithun, Marianne (1999), *The Languages of Native North America*. New York: Cambridge Univ. Press.

Tucker, Archibald N., and John T. Ole Mpaayei (1955), *Maasai Grammar, with Vocabulary*, London: Longmans, Green & Co.

Measuring the consistency of the "irrealis" domain using semantic maps

Wesley Kuhron Jones
(University of Oregon)

This paper applies semantic map methodology to the domain of irrealis. Using a sample of 144 languages, I quantify the extent to which emic irrealis categories vary across languages in terms of which functions they encode. I show that irrealis can be empirically characterized as a broad semantic domain with low consistency in terms of encoding by linguistic forms. This contrasts with the high consistency of its better-established sub-domains, such as negative or future. While it is possible to create a notionally consistent cross-linguistic definition of irrealis meaning, as many authors have (Mithun 1999; von Prince, Krajinović & Krifka 2019, 2022), none of these definitions reflects a category that languages prioritize in their TAM systems.

I conclude that any etic definition of reality status is not very reliably encoded by linguistic forms. This is because of the wide variety of patterns among languages regarding how semantic functions are categorized as realis, irrealis, or something else. This finding does not deny that, in some languages, an emic formal distinction exists that can well be called "reality status". For those languages, "irrealis" should remain a valid descriptive shorthand if no more specific label suffices to characterize a form's functions succinctly.

As for etic notions of irrealis, a category being broad in conceptual space should not necessarily be counted as a mark against its validity. The consistency measure (silhouette score), however, is more telling about the nature of how irreality is marked in languages. Related domains such as future and imperative have very high consistency, reflecting the fact that linguistic forms tend to distinguish these more clearly from other TAM categories. This does not mean that "irrealis" is wholly invalid as a descriptive or typological notion, as Bybee (1998) concludes. However, cross-linguistically, "irrealis" has lower compactness and lower consistency than certain other functional notions, when measured by how reliably it is encoded by forms cross-linguistically. A semantic category being etically notionally consistent is not enough to guarantee that languages will encode it reliably.

Descriptive and typological studies should be aware of the amount of compactness and consistency of the definitions they use. Descriptive work should avoid overly vague labels such as "irrealis" as a primary way to describe an emic category. The label is fine as a shorthand, but the emic category's component parts must first be clearly specified with examples which show fine-grained semantic distinctions among the functions it encodes. Typological studies should be aware of the variation among emic irrealis categories, the suitability of other labels for very similar categories, and the fuzziness of how irrealis-like a given semantic function is. Any typological study which seeks to make statements that a language which has an irrealis category will do X (even if this statement is probabilistic) is misguided. Irrealis categories in different languages cannot be equated; they should instead be correlated using quantitative methods.

References

Bybee, Joan L. 1998. "Irrealis" as a Grammatical Category. *Anthropological Linguistics* 40(2). 257–271.

Mithun, Marianne. 1999. The Languages of Native North America. Cambridge University Press.

von Prince, Kilu, Ana Krajinović & Manfred Krifka. 2019. Irrealis in branching time. In *Austronesian Formal Linguistics Association*. London, Canada.

von Prince, Kilu, Ana Krajinović & Manfred Krifka. 2022. Irrealis is real. *Language* 98(2). 221–249.

Translating Greek Modalising Adverbs into French:
Exploring Modal Shifts and the Notion of Possibility in Assertion

Fryni Kakoyianni-Doa
(University of Cyprus)

Keywords: Greek, modal, adverbs, possibility, assertion

The primary objective of this research is to elucidate the distinct modal markers inherent in the French language when confronted with the task of translating two Greek adverbs which fall under the category of modals (Molinier & Levrier, 2000: 91), or modalising adverbs of assertion (Borillo, 1976). The specific adverbs in question are φυσικά [fisika – évidemment - naturellement - of course- naturally] and σίγουρα [sigoura – asurément, certainement - certainly - surely]. These adverbs were chosen for two reasons: firstly, they denote the visual access to information to different degrees and, secondly, they are used frequently in the two languages.

Semantically, their role resides in the evaluation of truth or the degree of certainty, inherently positioned on a positive scale, in relation to the proposition to which they are appended. Accordingly, they function as modalisers of assertion, aligning unmistakably with the alethic or epistemic modalities of logic. Adopting the nomenclature of "assertive adverbs" in line with Borillo (1976), Molinier and Levrier (2000) and Kakoyianni-Doa (2008, 2022), we underscore the particular intrigue surrounding these adverbs, as our study reveals that, despite overarching morphosyntactic similarities in both French and Greek, a substantial typological disparity emerge during their translation from Greek into French. This study is based on data which were obtained from the Sketch Engine parallel corpora.

Intriguingly, while these adverbs are non-parenthetical in both languages, they become thoroughly integrated into the semantic fabric of the proposition they accompany and express possibility as well given their close association with verbal forms such as, "il est tout à fait possible", "il est sans doute possible". As Haßler (2018) says "evidentiality is considered to be a structural dimension of grammar, the values of which are expressed by types of constructions that code the source of information which a speaker imparts. Drawing a boundary between speaker's stance, epistemic modality and evidentiality presents difficult problems in many European languages" (2018: 18).

In light of these observations, we postulate the hypothesis that these adverbs can be employed parenthetically, and it is plausible to consider the notion of possibility in the context of expressing the modalisation of assertion. Concurrently, we endeavour to pinpoint the underlying factors contributing to this disparity. We inquire whether it is a matter of enunciative rhetorical strategy on the part of the speaker or a rhetorical embellishment characteristic of the proficient interpreter.

References

Borillo, Andrée (1976), Les adverbes et la modalisation de l'assertion. *Langue française* 30, 74-89.

Haßler, Gerda (2018), Evidential and epistemic sentence adverbs in Romance languages. *Linguistik online* 92 (5), 87-109.

Kakoyianni-Doa, Fryni (2008), *Adverbes de phrases français et grecs : étude contrastive et perspectives didactiques* [Thèse de doctorat].

Kakoyianni-Doa, Fryni (2022), Formal and functional features of modal adverbs in French and Modern Greek, in O. Duplâtre and P-Y. Modicom (eds), (2022), *Adverbs and Adverbials: Categorial Issues*, Berlin, Boston: De Gruyter Mouton, 167-194.

Molinier, Christian & Lévrier, Françoise (2000), *Grammaire des adverbes. Description des formes en -ment*. Genève: Librairie Droz.

Argumentative anchoring of possibility in a corpus linguistics perspective

Corinne Rossari & Claudia Ricci

(University of Neuchâtel & University of Neuchâtel)

Keywords: possibility, argumentation, corpus linguistics, statistical association, necessity

Possibility is generally associated with the notion of necessity (cf. Auwera & Plungian 1998, Le Querler 2001, and Vetters 2004, 2012). Either the two are opposed or they are considered as being two poles of a *continuum*. In an argumentative perspective, they are also conceived together: Toulmin (1958) associates the different phases of argumentation processes with different degrees of modal force; he points out that putting forward hypotheses is characterized by the modality of possibility, drawing conclusions by that of necessity.

We will investigate possibility in relation to necessity considering that the two modalities are two poles on a scale of modal strength, in argumentative environments identifiable as such by means of the presence of connectives, in order to systematically extract such environments within large corpora. Taking, for instance, a concessive structure as the one set up by French *mais* ('but'), we want to understand if the content preceding *mais* (expressing concession) and the content following it (presented as being endorsed by the speaker) differ in terms of their specific associations with linguistic indications leaning towards possibility or necessity. We will take different connectors in corpora representing various discourse genres (press, political discourse, encyclopedia) and we will test the preferred argumentative structures for the use of several modal forms covering the spectrum from possibility to necessity in two languages, French and Italian. Investigating different languages and discourse genres will allow us to see whether common trends can be highlighted according to the argumentative structure; for instance, we will see if in both languages and in the different corpora considered a specific association is at play between a conclusive movement and indicators of strong possibility up to necessity, or between sequences introducing concession and markers of weak possibility.

Regarding our methodology, we will use the TXM platform to extract argumentative structures by systematically searching for connectives in corpora covering the three genres mentioned, and we will identify the modal forms showing preferential associations with these patterns, by means of different methods and statistical calculations (specificity scores, AFC). After checking the relevance of the association 'connective + modal form' by examining the actual occurrences in the KWIC concordance, we'll establish correlations between argumentative movements instantiated by connectives and the type and sub-types of modalities specifically associated with them. This will allow us to see to what extent the 'modality of possibility + connective' patterns cut across the genres and languages considered.

References

Auwera, Johan van der and Plungian, Vladimir A. (1998), "Modality's semantic map", Linguistic Typology 2, 79-124.

Le Querler, Nicole (2001), « La place du verbe modal pouvoir dans une typologie des modalités », Cahiers Chronos 8, 17-32.

Toulmin, Stephen E. (1958) *The uses of argument*, Cambridge: Cambridge University Press.

Vetters Carl (2004), Les verbes modaux pouvoir et devoir en français, *Revue Belge de Philologie et d'Histoire* 82, 657-671.

Vetters Carl (2012), Modalité et évidentialité dans pouvoir et devoir : typologie et discussions, *Langue française* 173, 31-47.

Possibility as (ir)relevance:
Conditional possibility expressions with epistemic *mag*

Katja Politt

(Leibniz University Hannover)

Keywords: German, factuality, irrelevance, conditionality, modal verbs

The epistemic German modal *mag* provides a non-factuality judgement of a proposition with an inherent concessive and adversative reading. This is a distinguishing function of *mag* in contrast to the other epistemic modals in German (cf. Politt 2022, Rossari & Smirnova 2022).

> (1) *Die Freiheit der Forschung* **mag** *für viele Professoren oberste Maxime* **sein***, für die Wissenschaftler der Fraunhofer-Gesellschaft ist sie zweitrangig*. (DWDS; o.A. 2000 [1999])
> 'Freedom of research may be the highest maxim for many professors, for the scientists of the Fraunhofer-Gesellschaft it is of secondary importance.'

Epistemic *mag* suggests that the truth value of p_1, *freedom of research is the highest maxim for many professors*, is irrelevant in regard to the truth value of the second proposition p_2, *it is of secondary importance for the scientists of the Fraunhofer-Gesellschaft*. This can be paraphrased as *Even though it is possibly true that $p_1 \vee \neg p_1$, p_2 still holds.* Its contrastive possibility judgement is often combined with additional markers of irrelevance (cf. Haspelmath & König 1998, Leuschner 2006) in the sentence containing p_1, further strengthening the notion of p_2 being true regardless of the truth conditions of p_1.

Based on 1391 sentences from the DWDS core corpus of the 20th century (Geyken 2007) containing both epistemic *mag* and a marker of irrelevance, this study suggests that epistemic *mag* is close to the constructional family of concessive conditionals (Politt 2023, Vander Haegen et al. 2022). This rests on the analysis of co(n)text features such as the type of irrelevance markers, negation, and the presence of connectors. Results show preferences for non-stative verbal complements, impersonal subjects, as well as scalar and universal irrelevance expressions. The inherent twofold (ir)relevance assessment of *mag* is strengthened by these quantification strategies. Most importantly, this reading is still present even if no additional strengthening co(n)text features are.

References

Geyken, Alexander (2007), The DWDS corpus: A reference corpus for the German language of the 20th century. In *Collocations and idioms: Linguistic, lexicographic, and computational aspects*. *Corpus and discourse. Research in corpus and discourse*, Christiane Fellbaum (Hg.), 23–41. London, New York: Continuum.

Haspelmath, Martin & Ekkehard König (1998), Concessive conditionals in the languages of Europe. In *Adverbial constructions in the languages of Europe*. *Empirical approaches to language typology. Band 3*, Johan van der Auwera (Hg.), 563–640. Berlin: Mouton de Gruyter.

Leuschner, Torsten (2006), *Hypotaxis as building-site*. *The emergence and grammaticalization of concessive conditionals in English, German and Dutch*. Teilw. zugl.: Berlin, Freie Univ., Diss., 2003. *LINCOM studies in Germanic linguistics, LSGL*. *Vol. 24.* München: Lincom Europa.

Politt, Katja (2023), Epistemic *mag* revisited: Assessing the (ir)relevance of propositions. Studia Germanica Gedanensia 49, 29–39. https://doi.org/10.26881/sgg.2023.49.02

Politt, Katja (2022), *Formen und Funktionen von Paradigmen*. Berlin: Peter Lang. (= Sprache – System und Tätigkeit 75). DOI: 10.3726/b19526

Rossari, Corinne & Elena Smirnova (2022), Post-modal Concessive Meanings: A Contrastive Corpus Study of French and German Modal Verbs. In *Aspects of tenses, modality, and evidentiality*. *Cahiers chronos*. *volume31*, Laura Baranzini & Louis d. Saussure (Hgg.), 234–261. Leiden, Boston: Brill.

Vander Haegen, Flor, Tom Bossuyt & Torsten Leuschner (2022), Emerging into your family of constructions. German [IRR was] 'no matter what'. *Constructions and Frames* 14 (1): 150–180.

Possibility modal expressions and negation in Javanese

Jozina Vander Klok
(Humboldt Universität zu Berlin)

Javanese modals lexically express possibility modal force using distinct lexical items that also referentially specify the modal flavour (under a Kratzerian analysis whereby the modal flavour is a presupposition on the type of possible worlds that the modal can quantify over) (Vander Klok 2012, 2013). In the dialect spoken in Paciran, East Java, the possibility modals are *paleng* 'maybe', which is lexically specified for epistemic modal flavour; *iso* 'can', specified for circumstantial and teleological, and *oleh* 'may', specified for deontic. In addition, the Indonesian borrowing *mungkin* (itself an Arabic borrowing) is used in Javanese, which expresses epistemic possibility modality like *paleng*. Other modals are used to express necessity (*mesthi* 'EPIS.NEC', *kudu* 'ROOT.NEC'). In previous work on VP-ellipsis with modals in Javanese (Kendal, Central Java), Sato (2013) proposes the structure in (1), where epistemic modals structurally dominate root modals and where possibility and necessity modals are located in the same projection:

(1)   CP > TP > NegP > ModEpisP > ModRootP > VP   (Sato 2013:192)

In Sato's proposal, modals can undergo head movement to T, which structurally dominates negation (NegP), allowing for negation to scope above and below *any* modal.

This paper takes a closer look at modals expressing possibility in Paciran Javanese and their interaction with negation, arguing for a different clausal structure. I show that of the two modals expressing epistemic possibility (*mungkin, paleng*), both adverbs, only *mungkin* allows to be embedded under negation *ora/gak*. I argue that it is constituent negation, yielding the interpretation 'impossible'. *Paleng* only allows for negation to scope below the modal, negating the prejacent (the proposition without the modal). I explore whether *paleng* is a positive polarity item (cf. Iatridou & Zeijlstra 2012) and/or that it is paradigmatic blocking with *ora/gak mungkin* 'impossible'. Concerning the two modals expressing root possibility, *iso* and *oleh* (both auxiliaries), I show that both allow for negation to scope above and below, where clausal negation scopes above the modal: **Neg** > ModRoot > VP. With ModRoot > **Neg** > VP word order, only the verb phrase is negated, interpreted either as an assertion or a polar question. This analysis predicts that these modals should allow double negation, as borne out in (2):

(2) Yu   Siti   **gak   iso   gak**   ngerasa-ni   mbek Yu   Dur
    sister Siti NEG CIRC.POS NEG   AV.feel-APPL   with   sister Dur
    'Siti can't not gossip with Dur.' (10may2012.003)

This richer empirical picture suggests that negation along the spine can occur in multiple positions below epistemic modal projections, as proposed in (3), and which also supports a structural distinction for different possibility modals, separate from necessity ones (cf. Cinque 1999; contra Sato 2013). Where negation seemingly occurs above ModEpisP (*gak mungkin* 'impossible'), this is constituent negation. Moreover, I propose that—contra Sato (2013)— syntactic movement of a modal is constrained to yielding a question interpretation, supported also by subject inversion in this case.

(3)   CP > ModEpisP (> NegP) > TP > (NegP >)

> ModRootNecP > (NegP >) ModRootPossP > (NegP >) VP

References

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads. A Cross-linguistic Perspective*. New York: Oxford University Press.

Iatridou, Sabine and Hedde Zeijlstra. 2013. Negation, Polarity, and Deontic Modals. *Linguistic Inquiry* 44 (4): 529–568. doi: https://doi.org/10.1162/LING_a_00138

Sato, Yosuke. 2013. Cyclic Spell-Out and modal complement ellipsis in Javanese. *Linguistic Analysis* 38 (3,4): 183-206.

Vander Klok, Jozina. 2012. 'Tense, aspect, and modal markers in Paciran Javanese.' Ph.D. dissertation, McGill University.

Vander Klok, Jozina. 2013. 'Pure possibility and pure necessity modals in Paciran Javanese'. Oceanic Linguistics 52.2:341-374.

WS21 The Determinism Assumption in Morphology

Panini's Principle leaks: When and why

Vsevolod Kapatsinski
(University of Oregon)

Keywords: morphological productivity, frequency effects, regularity, Elsewhere Condition, Proper Inclusion Precedence

Morphological choice, like other form choice, is often non-deterministic. It is often assumed that this nondeterminism is constrained in a specific way, known as Panini's Principle (Baerman, 2005) and also referred to as Elsewhere Condition, Blocking, (Statistical) Preemption, Proper Inclusion Precedence, and the Subset Principle. The Principle states that if the context of application for Rule A is a superset of the features of the context of application of Rule B, then Rule A will always take precedence over Rule B when both are applicable. The present paper discusses cases that violate the Principle, and aims to elucidate the reasons for such violations.

Leaks involve the emergence of exceptions to A as a result of the overapplication of B. For example, exceptions to velar palatalization in Russian (Kapatsinski, 2010) arise where there is a competing more general rule of higher type frequency that should have not applied in the present context according to Panini's Principle. For example, because the verbal stem extension +i usually attaches to non-velars, a general 'just add -i' rule develops high productivity, and is then overextended to velar inputs. Accordingly, palatalization often fails before +i. In contrast, the diminutive masculine suffixes +ok and +ek tend to attach to velars. As a result, velar palatalization does not tend to fail before +ek or +ok. The most frequent diminutive with non-velars, +ik, almost never attaches to velars in established vocabulary, but is now extended to velars, which it often fails to palatalize.

These kinds of results are consistent with several approaches to language in which competing generalizations are activated in parallel, based on partial overlap with the current context. These theories vary in whether selection of a generalization is based on inference that a generalization is applicable to the current context (e.g., Albright & Hayes, 2003; Suttle & Goldberg, 2011), or on the cues present in the current context activating a form that is usually used in other contexts that is then extended to the present context (e.g., the diminutive +ik in Russian being overextended to velar inputs; see Barthdal, 2003; Caballero & Kapatsinski, 2022). The former class of theories predicts that likelihood of extension is determined primarily by type frequency, rather than token frequency, and (if anything) token frequency hurts productivity. The latter class of theories suggests that token frequency helps a form to be extended by increasing its resting activation level and therefore its accessibility (Kapatsinski, 2023). We are currently testing this difference using a miniature artificial language modeled on Harmon & Kapatsinski (2017) but varying token and type frequency orthogonally.Please do not fully capitalise the title; if there is a subtitle, it should start with a capital letter.

References

Albright, Adam, & Hayes, Bruce (2003), Rules vs. Analogy in English Past Tenses: A Computational/Experimental study. *Cognition* 90(2), 119-161.

Baerman, Matthew (2005). Directionality and (un)natural classes in syncretism. *Language* 80(4), 807-827.

Barthdal, Johanna I. (2003), *Case in Icelandic: A synchronic, diachronic and comparative approach*. Amsterdam: John Benjamins.

Harmon, Zara, and Kapatsinski, Vsevolod (2017), Putting Old Tools to Novel Uses: The Role of Form Accessibility in Semantic Extension. *Cognitive Psychology* 98, 22-44.

Kapatsinski, Vsevolod (2010), Velar Palatalization in Russian and Artificial Grammar: Constraints on Models of Morphophonology. *Laboratory Phonology* 1(2), 361-393.

Kapatsinski, Vsevolod (2023), Understanding the roles of type and token frequency in usage-based linguistics, in M. Diaz-Campos & S. Balasz (eds.), *The Handbook of Usage-Based Linguistics*. Malden, MA: Wiley.

Suttle, Laura, & Goldberg, Adele E. (2011), The Partial Productivity of Constructions as Induction. *Linguistics* 49(6), 1237-1269.

Diatopic Dimension of Overabundance:
The Case of the Croatian Genitive Plural

Tomislava Bošnjak Botica, Jurica Polančec, Gordana Hržica, Sara Košutar and Perina Vukša Nahod
(Institute of Croatian Language and Linguistics, University of Zagreb, University of Zagreb, The Arctic University of Norway, Department of Dialectology)

When non-deterministic outcomes occur in standard languages characterized by prescriptive tendencies, this can represent an intriguing phenomenon. A notable example of this is seen in the overabundant genitive plural (GPL) forms of Croatian nouns where the stem ends in a consonant cluster. Words like *kocka* 'cube' are an example of this, where *-ck-* is the stem-ending consonant cluster. Here, three distinct endings are theoretically possible: *-ā* (*kockā*), *-i* (*kocki*), and the *-ā* ending with an alternate stem where an epenthetic *ā* breaks up the stem-final consonant cluster (*kocākā*) (Babić et al. 2007). While the *-i* ending is more recent, having been first observed in the early 20th century (Iveković 1905), the *-ā* endings are older, dating back to the 14th century (Jurišić 1992).

Our preliminary research based on corpus data has shown that in a sample of 239 nouns from this category, the *-i* ending predominates, regardless of the language register and despite the normative recommendations, which favor the *-ā* ending (with or without epenthesis). However, the corpus findings do not fully address a relevant aspect of overabundance within this category, namely its geographic (diatopic) variability. The literature on Croatian dialects indicates that in Southern Croatia, the adoption of the *-i* ending is limited (Kapović 2018).

We aim to investigate the distributional patterns of overabundance in GPL forms of Croatian nouns to provide empirical evidence of a shift towards the usage of the *-i* ending in Croatian, accompanied by diatopic variability. We anticipate that the entrenchment of the *-i* ending is influenced by the impact of larger urban centers, notably Zagreb (the capital), where the *-i* ending prevails. Within this urban setting, speakers from diverse dialectal backgrounds are likely to adopt the *-i* ending even if it is not present in their regional varieties.

To this end, an experimental study with two groups of participants will be conducted. An elicited productin task was developped for this purpose, where participants will be given a noun in NSG and asked to produce the GPL form (prompted by: "This is xxx. The rabbit sees many ____."). The stimuli include 30 target nouns, 30 nouns without consonant clusters (thus lacking overabundance), and 60 fillers. Data will be collected from 100 young speakers of Southern Croatian dialects, half of which are high school students (16–18 years old) who live in their birthplace, and the other half are college students (21–24 years old) who moved to Zagreb a few years ago.

We hypothesize that: 1. Speakers from the first group use the *-a* ending; 2. Speakers from the second group, when exposed to it (i.e., after living in Zagreb for a few years), switch to the *-i* ending to varying degrees.

This study will allow us to investigate two important questions:

a) Whether currently overabundance in this category mainly depends on inter-speaker diatopic variation, with only limited intra-speaker variability?

b) Whether the entrenchment of the -*i* ending is facilitated by the trend of the Croatian population moving towards larger urban centers (especially Zagreb)?

References

Babić, Stjepan, Dalibor Brozović, Ivo Škarić & Stjepko Težak. 2007. *Glasovi i oblici hrvatskoga književnoga jezika*. Zagreb: Nakladni zavod Globus.

Iveković, Franjo. 1905. Drugi padež množine s nastavkom î u imenicama ženskih, koje imaju u prvom padežu jednine nastavak a. *Rad JAZU* 162. 186–191.

Jurišić, Blaž. 1992 [1944]. *Nacrt hrvatske slovnice. 1, Glasovi i oblici u poviestnom razvoju*. (Ed.) Dalibor Brozović. Zagreb: Matica hrvatska.

Kapović, Mate. 2018. Širenje nastavka –i u genitivu množine e–deklinacije u suvremenoj štokavštini. *Suvremena lingvistika* 44(85). 39–72.

Beyond the Study of Defectiveness: A Comprehensive Analysis of Unbalanced Paradigms in Czech Nouns, Adjectives, and Verbs

Dominika Kovarikova
(Charles University)

Recent advances in linguistics have deepened our understanding of paradigm defectiveness, where certain grammatical forms are missing (e.g. Sims 2015; Nikolaev and Bermel 2022; Bermel et al. 2023). This includes inherent defectiveness, identified by native speaker judgment, and contingent defectiveness, noted through corpus evidence (Brown and Evans 2021; Brown et al. 2023); the latter type is the subject of this research. This study broadens the scope to examine 'paradigm balance' in Czech nouns, adjectives, and verbs, encompassing not only the absence but also the overrepresentation and underrepresentation of certain forms, thus revealing a more intricate phenomenon than defectiveness alone. A dataset of 35,000 frequent lemmas is analyzed by the GramatiKat application (Kováříková and Kovářík 2021) employing boxplot analytics – a statistical tool adept at visualizing distributions and identifying outliers. Thus, we establish benchmarks for a balanced paradigm and identify disproportionate frequency of certain forms.

This research shows that individual morphological paradigms frequently diverge from established benchmarks—expected behaviors in a broad set of lemmas—where one form or a set of forms appears with a frequency that significantly exceeds the norm, thereby challenging the presumption of uniformity. The study shows how semantic, syntactic, and collocational features of lemmas, such as the high frequency of first person in sense verbs or the predominance of the feminine form of 'rychlovarný' (fast-boiling), due to its collocation with 'konvice' (kettle), contribute to patterns of usage that defy deterministic predictions.

Preliminary findings suggest that unbalanced paradigms are characterized not only by complete defectiveness but also by subtler forms, such as deviations from expected frequency distributions. By highlighting these non-deterministic elements, the study contributes to a nuanced understanding of Czech morphological paradigms, suggesting that the structure and usage of language are influenced by a range of unpredictable and fluctuating linguistic factors.

References

Bermel, N., Knittl, L., Nikolaev, A. (2023): Uncertainty in the production of Czech noun and verb forms. Word Structure 16(2-3), pp. 258–283.

Brown, D., Evans, R. (2021): Two kinds of nothing: Exploring Corpus-based Approaches to Defectiveness in Russian. Presentation, Feast & Famine Workshop.

Brown, D., Bermel, N. & D. Kováříková (2023): What we know, don't know, and want to know about defectivity and overabundance. Presentation, SLE Conference Athens 2023.

Kováříková, Dominika & Kovářík, Oleg. 2021. GramatiKat. Prague: FF UK. Available at WWW: <http://www.korpus.cz/gramatikat>.

Nikolaev, A., & Bermel, N. (2022). Explaining uncertainty and defectivity of inflectional paradigms. Cognitive Linguistics, 33(3), 585–621. https://doi.org/10.1515/cog-2021-0041

Sims, A. D. (2015). Inflectional defectiveness. Cambridge University Press.

Acquisition of non-deterministic input: The case of overabundance and defectivity in Croatian and Estonian

Virve Vihman, Gordana Hrzica, Sara Košutar & Mari Aigro
(University of Tartu, University of Zagreb, The Artic University of Norway & University of Tartu)

The Determinism Assumption prevalent in theoretical morphology also underlies research on morphosyntactic development, in which an implicit assumption is that the target language includes one accurate form for every lexical meaning and grammatical function. This approach informs discussion of gaps in children's knowledge and overgeneralisation, seen as a developmentally important step in acquiring morphological productivity (e.g. Marchman 1997). Explanation of the retreat from overgeneralisation relies on either the "uniqueness principle", according to which children assume a single lexical form expressing each grammatical function (e.g. Marcus et al. 1992, Dressler 2011), or a more probabilistic approach, by which children gradually acquire the target forms encountered in the input (e.g. Ambridge et al. 2013). Comparing theoretical models to data from children who are faced with non-determinism in the input is fruitful for understanding both children's morphosyntactic acquisition and the cognitive status of non-deterministic paradigm cells.

We discuss corpus and experimental evidence from two languages, Croatian and Estonian, on the acquisition of non-deterministic input, including overabundance (the situation where multiple forms can fill a single paradigm cell) and defectivity (or paradigm gaps) in the inflectional systems. Croatian has plentiful overabundance (e.g. okva-NOM.SG, lokava ~ lokva ~ lokvi for GENITIVE.PL) within a morphological class, while Estonian has overabundance (e.g. loik-NOM.SG, loikusid ~ loike for PARTITIVE.PL) cutting across inflectional classes (see Kaalep 2010). The nature of Child-Directed Speech, however, is such that individual children are less likely to encounter overabundance for any single lexeme. Although children's own productions are variable and include overgeneralisation, our experimental results suggest that overabundant forms hinder acquisition.

Morphological defectivity, on the other hand, as discussed in reference to the target language, is difficult to relate to the acquisition trajectory. Children's exposure to, knowledge of, and access to lexical paradigms is necessarily incomplete. Hence, a certain kind of defectiveness is inherent to the input they receive. To acquire language, children need to overcome the limited input and often produce unencountered forms. We discuss how evidence from overabundance in the input is qualitatively different and has different implications from defectivity. This is theoretically informative for theories of both acquisition and morphology.

References

Ambridge, B., Pine, J.M., Rowland, C.F., Chang, F. & Bidgood, A. 2013. The retreat from overgeneralization in child language acquisition: word learning, morphology, and verb argument structure. *WIREs Cogn Sci*, 4: 47-62.

Dressler, W. (2011). The rise of complexity in inflectional morphology. *Poznań Studies in Contemporary Linguistics* 47(2), 159.

Kaalep, H.-J. 2010. Mitmuse osastav eesti keele käändesüsteemis [Partitive plural in the Estonian case system]. *Keel Ja Kirjandus*, *2*, 94–111.

Marchman, V. A. 1997. Children's productivity in the English past tense: The role of frequency, phonology, and neighbourhood structures. *Cognitive Science* 21, 283–304.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F. & Clahsen, H. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development* 57(4), i–178.

Network-based quantification of morphological irregularity

Maria Copot & Andrea Sims
(The Ohio State University & The Ohio State University)

Keywords: network theory, morphology, exponence, irregularity, word-and-paradigm theory

Nondeterministic morphological behaviour is rooted in the lexicon as a complex whole. A lexeme's behaviour is influenced by that of lexemes similar to it in relevant dimensions, such as exponence patterning and phonological similarity. Lexemes often receive contrasting influences from different parts of the lexicon, generating probabilistic optionality in their inflectional behaviour (Ackerman & Malouf 2016): if a lexeme's exponents are a mix of those of class A and B, nondeterminism is introduced in the realisation of unknown wordforms of the lexeme, since these could pattern with either class. We extend graph-theoretic approaches to morphology (Beniamine 2021, and Sims 2020), which mathematically represent the lexicon as a set of items engaged in relationships with each other, to propose a computationally-implemented, network-based methodology for describing and modelling these contrasting influences.

To illustrate, we show how different types of irregularity and their differing contributions to competition within morphological systems can be defined and quantified in network terms. A conventional distinction separates lexemes into 'regular' and 'irregular': the former behave in accordance with the rules of the inflectional system, which the latter fall outside of. However, when viewing the lexicon as an ensemble of structured relationships, irregular lexemes aren't necessarily isolated from the regular subsystem: they may share exponence with more regular classes or combine regular exponents into unusual (irregular) sets. Different types of irregulars will contribute differently to nondeterminism in the morphological system. We model the French and Croatian lexicons as bipartite networks (Gross & Yellen 2006) where nodes in the two partitions are lexemes and exponents, with edges connecting each lexeme to its exponents. We use the resulting structure to illustrate a quantitative typology of the ways in which lexemes can be 'irregular' in their exponent patterns and how they contribute to nondeterminism in the morphological system.

We focus on two types of irregularity: the first are lexemes having exponents peripheral to the system, unique to the lexeme (French ÊTRE 'to be'), a dimension of irregularity that can be captured by the network-theoretic measure of *degree* (number of outgoing edges from an exponent node: if an exponent features in few words, it will have low degree). Such lexemes introduce new exponents into the system, potentially creating competition. Conversely, heteroclite behaviour (Croatian OKO 'eye' follows different inflectional classes in the plural and singular) is characterised as connecting two otherwise separate sets of exponents, a property can be captured in quantitative, gradient terms by the network-based measure of *betweenness-centrality* (quantifying how many shortest paths between two nodes in the network pass through a node: if a lexeme borrows exponents from two different morphological classes, it will act as a bridge between the classes in network-theoretic terms). This type of irregularity fosters competition for all members of the two inflectional classes it shares exponence with by creating a bridge between them.

Modelling the lexicon as a network of lexemes and exponents, thus adopting precise mathematical definitions of lexical relatedness, provides a method for making precise quantitative predictions about where nondeterminism might arise in the lexicon.

References

Ackerman, Farrell & Robert Malouf (2016), Implicative relations in word-based morphological systems. In Andrew Hippisley & Gregory Stump (eds), *The Cambridge handbook of morphology*, Cambridge: Cambridge University Press, 297-328.

Beniamine, Sacha (2021), One lexeme, many classes: Inflection class systems as lattices. In Berthold Crysmann & Manfred Sailer (eds), (2021), *One-to-many relations in morphology, syntax and semantics*, Berlin: Language Science Press, 23-51.

Gross, Jonathan L. & Jay Yellen (2006), *Graph theory and its applications*. Boca Raton, FL: Chapman & Hall.

Sims, Andrea D. (2020), Inflectional networks: Graph-theoretic tools for inflectional typology. *Proceedings of the Society for Computation in Linguistics* 3: 88-98.

Sociolinguistically-driven non-deterministic outcomes:
A comparative case study of the Slavic family

Kristian Roncero, Martin Beneš, Tomislava Bošnjak Botica
(LMU, Czech Language Institute & Institute of Croatian Language)

Overabundance (i.e. having more than one form available for a given cell) poses a challenge to traditional deterministic frameworks in morphology. Yet, most frequently, this has been studied from a purely language-internal perspective. In this paper, we argue the relevance of the sociolinguistic dimension for understanding non-deterministic outcomes.

All the languages for our case studies belong to the same family (Slavic). Despite having a similar morphophonological material, we show that the fate of overabundance for the GEN.PL has been very heterogeneous, which we argue, it is strongly related to their unique sociolinguistic settings, the role of language authorities and speakers' metalinguistic awareness.

On the non-diglossic side of the spectrum, we look at Russian, where there is a strong prescriptive tradition. Authorities have attempted to limit well-known cases of "doublets" (present in other East Slavic varieties) that are a fruit of the uncertainty about the forms; e.g. *+pástorov ~ °pastoróv* 'pastors' (Brown & Hippiseley, Mahota 1993). However, given the high level of speakers' metalinguistic (and normative) awareness, overabundance has a less limited reach than in the other cases. In settings where this awareness is – despite having a generally well-established codification – absent or not as strong among all speakers, the prescribed standard can significantly diverge from common usage (Czech, Croatian). In some cases, overabundance has been artificially created by the academies on the level of description (based on the traditional deterministic approach), e.g. Croatian *bitka* 'battle' >[GEN.PL] *+bitaka~bitka~bitki*; cent 'cent' >[GEN.PL] *centa~centi*, see Birtić et al. (2012); Czech *louka* 'meadow' >[GEN.PL] *luk~+louk*, but only *luk* is in real usage. Conversely, in other cases, academies have proscribed some forms which are very frequently used; e.g. Croatian *prsa* 'breast' >[GEN.PL] *prsa~°prsiju*. Or the Czech Institute only approves *pět +eur* 'five euro >[GEN.PL]' (2,292 corpus hits on SYN v11), whilst *pět °euro* is common colloquially (409 hits) (for normative description of Czech FEM. GEN.PLS see MČ2 (1983: 322–332), for a corpus-based description see Štícha (2021: 486–505)).

On the other end of the spectrum, we assess West Polesian. Speaking West Polesian is heavily stigmatised and only used with trusted members of the community; whilst, at the same time, speakers are exposed to up to four highly standardised varieties. Thus, this comprises an extreme case of extended diglossia, without a prescriptive authority and little awareness of a norm. As predicted by Dorian (2010), speakers are comfortable using multiple forms for the same cell, without any added sociolinguistic or discursive meaning. We present corpus examples we have documented of individual speakers using up to five GEN.PL forms for the noun 'year' (out of six possible forms recorded in the group: *ɦod~ɦaˈduʋ~rik~ˈrokɪʊ~roˈkɪʊ~lʲɪt*) (Roncero et al [Forthcoming]).

In sum, we depart from examples of the use of the GEN.PL in different Slavic languages and how the role of language authorities can shape their outcome.

+Prescribed form - °Proscribed form

References

Birtić, M., Blagus Bartolec, G., Hudeček, L., Jojić, Lj., Kovačević, B., Lewis, K., Matas Ivanković, I., Mihaljević, M., Miloš, I., Ramadanović, E., Vidović, D. (2012), *Školski rječnik hrvatskoga jezika*. Zagreb: IHJJ, ŠK.

Brown, D. P., & Hippisley, A. R. (1994), Conflict in Russian Genitive Plural Assignment: A Solution Represented in DATR. *Journal of Slavic Linguistics*, *2*(1), 48–76.

Dorian, N. C. (2010), *Investigating variation: The effects of social organization and social setting*. Oxford: Oxford University Press.

Mahota, W. (1993), The genitive plural endings in the East Slavic languages. *Journal of Slavic Linguistics*, 325-342.

MČ2: Komárek M., Kořenský, J., Petr, J., & Veselková, J. (eds) (1983), *Mluvnice češtiny: 2, Tvarosloví* [Grammar of Czech: 2, Morphology]. Prague: Academia.

Roncero, Kristian; Bermel, Neil; Russell, Jean; Accad, Grace [Forthcoming], A statistical approach to Overabundance, inter-and intra-speaker variation and suppletion.

Štícha, F. (ed.) (2021), *Velká akademická gramatika spisovné češtiny: II, Morfologie: Morfologické kategorie / Flexe: Část 2* [Great Academic Grammar of Standard Czech: II, Morphology: Morphological Categories / Inflection: Part 2]. Prague: Academia.

Overabundance as a result of defectiveness?

Emilija Tribocka
(University of Oslo)

This contribution is concerned with a case of coexistence of defectiveness and overabundance found in the paradigm of modal verbs in Norwegian.

A prototypical verb paradigm in Norwegian consists of four paradigm cells – the infinitive, the present, the preterit, and the supine. However, the modal verbs are seldom used in the infinite forms – the infinitive and the supine (cf. Torp, 2003, p. 237). These forms are simply missing for many speakers. For instance, the supine of the verb *skulle* 'shall' is not attested in the ancestor language Old Norse, and the speakers commonly leave out the infinitive and the supine when inflecting this verb (and other modal verbs) (Venås, 1974, p. 340). This development is connected to decategorization which is a part of the grammaticalization process. When lexical items are changing into more grammatical, they lose typical morphological traits (Hopper & Traugott, 2003, p. 107). In this case, the modal verbs lose typical, but less "verby", members of the paradigm – the infinitive and the supine. As a result, the paradigm of the modal verbs is defective.

Defectiveness does not necessarily equal complete absence of forms – they may be attested in a low number of tokens (Sims, 2015). The attested "missing" forms may be understood as the speakers' attempt to fill in the gaps or solve The Paradigm Cell Filling Problem (PCFP) (Ackerman et al. 2009). Interestingly, when Norwegian speakers are faced with the task to produce the infinitive or the supine – the missing cells of the modal verb paradigm – they come up with multiple different forms (Thornton, 2011). For *skulle* 'shall', following supine forms are attested: *skulla, skulli, skulja, skult* and *silt*. In other words, there is a supine form corresponding to each possible inflection class, as well as one form identical with the infinitive. These forms are collected from several dialects (not just one speaker), but such variety of forms associated with one cell is not found in any other verbs in Norwegian, besides the modals.

The observed overabundance may arise from the irregularity of the modals leading to the lack of reliable inferences about the outcome based on the non-defective forms (cf. Blevins et al., 2018, p. 283). The irregularity of the modal verbs makes it hard to find parallels which can serve as reliable basis for analogy in the inflectional system. Nevertheless, I argue that irregularity is not the only source of overabundance, and that defectiveness also plays a role. Irregular but frequent forms are memorized instead of being deduced each time. These forms are stable and less exposed to analogical innovations. Irregular but defective forms, on the other hand, are not frequent enough and must be deduced, resulting in the lack of stability and a variety of different outcomes. In this way, defectiveness may call on overabundance.

References

Ackerman, Farrell, James P. Blevins, Robert Malouf (2009), Parts and wholes: implicative patterns in inflectional paradigms. In J. P. Blevins, and J. Blevins (eds), (2009), *Analogy in Grammar,* Oxford: Oxford University Press, 54-82.

Blevins, James, P., Farrell Ackerman, Robert Malouf (2018), Word and Paradigm Morphology. In J. Audring, and F. Masini (eds), (2018), *The Oxford Handbook of Morphological Theory*, Oxford: Oxford University Press, 265-284.

Hopper, Paul, J., Elizabeth C. Traugott (2003), *Grammaticalization.* Cambridge: Cambridge University Press.

Sims, Andrea (2015), *Inflectional Defectiveness*. Cambridge: Cambridge University Press.

Thornton, Anna, M (2011), Overabundance (Multiple Forms Realizing the Same Cell): A Non-canonical Phenomenon in Italian Verb Morphology. In M. Maiden, J. C. Smith, M. Goldbach, and M. O. Hinzelin (eds), (2011), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, Oxford: Oxford University Press, 358-381.

Torp, Arne (2003), Frekvens, trykkletthet, reduksjon. In J. T. Faarlund (ed), (2003), *Språk i endring: Indre norsk språkhistorie*. Oslo: Novus forlag, 219-254.

Venås, Kjell (1974), *Linne verb i norske målføre: morfologiske studiar*. Oslo: Universitetsforlaget.

The Interplay of Morphotactics and Underspecification:
Overabundance in Athpare

Berthold Crysmann & Baptiste Loreau Unger
(CNRS & ENS Paris Saclay)

Among the factors conditioning overabundance, flexible morphotactics have been identified as a potential source, alongside alternations in exponence proper. While it is now well-established that morphotactic constraints being too lax may lead to alternate morph orders in the same cell, as in Chintang (Bickel et al., 2007) or Mari (Luutonen, 1997), we shall investigate the opposite situation where overly strict morphotactics leads to competition for overt expression of two distinct morphosyntactic properties.

Athpare (Ebert, 1997) verbs show agreement with both A(gent) and P(atient) roles, typically expressed by discrete person and number markers. As witnessed by the 2>1 subparadigm in Table 1a, we find two discrete markers for person, prefixal a- for second person and suffixal -ŋa for first person exclusive. Within the 2>1 paradigm, however, expression of number is special in that at most one exponent for number may occur. This contrasts with other parts of the Athpare verbal paradigm, where overt exponents of Agent and Patient number do cooccur, e.g. in the 2>3 subparadigm (Table 1a). Number marking in the 2>1 corner displays two characteristic properties: first, there is a morphotactic restriction to only have a single overt exponent of number in these cells. Second, the shapes of dual or plural markers in the 2>1 subparadigm distinguish neither role nor person, meaning that number marking is syncretic for A and P roles. The morphotactic restriction therefore confronts speakers with a forced choice as to which number specification to express overtly, giving rise to overabundance whenever two distinct non-singular values are involved.

Given morphotactic competition and the way it is resolved, overt expression of dual/plural

↓ A \ P →	1SG	1DU	1PL	...	3SG	3NSG
2SG	a- -ŋa	a- -ci -ŋa	a- -i -ŋa		a- -u	a- -u-ci
2DU	a- -ci -ŋa	a- -ci -ŋa	a- -ci -ŋa / a- -i -ŋa		a- -c-u	a- -c-u / a- -c-u-ci
2PL	a- -i -ŋa	a- -i -ŋa / a- -ci -ŋa	a- -i -ŋa		a- -u-m	a- -u-m-ci-m

(a) Athpare

↓ A \ P →	1SG	1DU	1PL
2SG	kɛ- -aŋ	a-kɛ-	a-kɛ-
2DU	a-kɛ-	a-kɛ-	a-kɛ-
2PL	a-kɛ-	a-kɛ-	a-kɛ-

(b) Limbu

↓ A \ P →	1SG	1DU	1PL
2SG	-ŋri	-ttsiki	-tiki
2DU	-ŋritsi	-ttsiki	-tikitsi
2PL	-ŋrini	-ttsikini	-tikini

(c) Thulung

Table 1: Number marking in 2>1 cells

number for any single role must be considered optional in the 2>1 cells. While optionality is a known source of overabundance, the case at hand is peculiar: competition not only limits the number of non-singular exponents but also requires there to be at least one. The intricate combination of underspecification, optionality and competition thus gives rise to a novel pattern of overabundance, where optionality is morphotactically conditioned.

The aims of this paper are threefold: first, we want to add to the typology of non-determinism in morphology (Thornton, 2019). Second, we shall compare the Athpare pattern to those found in related Kiranti languages, showing how Athpare 2>1 sits halfway between number neutralisation for these cells in Limbu (van Driem, 1987) and almost full combination in Thulung (Lahaussois, 2020). Finally, we shall present a formal analysis within Information-based Morphology (Crysmann & Bonami, 2016) that combines underspecification with positional competition.

References

Bickel, Balthasar, Goma Banjade, Martin Gaenzle, Elena Lieven, Netra Prasad Paudya, Ichchha Purna Rai, Rai Manoj, Novel Kishore Rai & Sabine Stoll. 2007. Free prefix ordering in Chintang. *Language* 83(1). 43–73.

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

van Driem, George. 1987. *A Grammar of Limbu*. Mouton grammar library.

Ebert, Karen H. 1997. *A Grammar of Athpare*. Mouton de Gruyter. Lincom Europa.

Lahaussois, Aimée. 2020. Descriptive and methodological issues in Kiranti grammar(s): Université de Paris / Université Paris Diderot (Paris 7) Habilitation à diriger des recherches. https://halshs.archives-ouvertes.fr/tel-03030562.

Luutonen, Jorma. 1997. *The variation of morpheme order in Mari declension*. Helsinki: Suomalais-ugrilainen Seura.

Mansfield, John, Sabine Stoll & Balthasar Bickel. 2020. Category clustering: A probabilistic bias in the morphology of verbal agreement marking. *Language* 96(2). 255–293.

Thornton, Anna M. 2019. Overabundance: A canonical typology. In Francesco Gardani, Hans Christian Luschützky & Franz Rainer (eds.), *Competition in morphology*, Springer. U. L'Acquila.

The typology of internal and external non-determinism

Greville G. Corbett

(University of Surrey; Max Planck Institute for Evolutionary Anthropology)

Taking consistent paradigms with uniquely filled cells as a baseline has proved a successful research strategy. It covers substantial swathes of data, and it highlights significant phenomena, including heteroclisis (Kaye 2015), uninflectability (Spencer 2020), and notably overabundance (Thornton 2019, Pellegrini 2023). For each non-canonical phenomenon, we can isolate simple instances. Yet they interact, and many such interactions remain to be specified, (but see Beniamine 2021, Guzmán Naranjo & Bonami 2021). Furthermore, we also find *external* non-determinism, with intriguing relations to *internal* (non-)determinism.

Consider *gazda* 'landlord, boss' in Serbo-Croat (BCMS, Corbett & Browne 2018, Bugarski 2019). This noun is *internally* fully determinate, each cell is uniquely specified (and inflectionally unexceptional). *Externally*, however, it is a remarkable hybrid. When SINGULAR it is MASCULINE. But when PLURAL it is MASCULINE *or* FEMININE. The external choice is non-deterministic: moving up the Agreement Hierarchy (attributive modifier - predicate - relative pronoun - personal pronoun), MASCULINE agreement becomes increasingly likely, and is obligatory for the personal pronoun (Leko 2010: 99):

(1) *propal-e /* *propal-i*   *gazd-e ...*   *on-i*   */ *on-e*
     ruined-PL.F   ruined-PL.M   landlord-PL   3-PL.M / 3-PL.F
     'ruined landlords ... they'

This example illustrates the dimensions of the canonical typology (Round & Corbett 2020). *Internally*, *gazda* 'landlord' is fully deterministic (each cell has a unique form), unlike items with overabundant cells. Externally, however, this internal specification underdetermines its requirements (two gender values are possible). Thus far it resembles Latin *dies* 'day', which also has alternative gender possibilities. The difference is that the gender value controlled by *gazda* 'landlord' varies according to its agreement target (it induces Agreement Hierarchy effects). It belongs in cell 4 of our typology:

		external		non-deterministic wrt internal spec.	
		deterministic			
		simple deterministic	deterministic wrt internal spec.	deterministic wrt target	non-deterministic wrt target (hybrid)
internal	deterministic	1. SC *soba* 'room'	2. SC *oko* 'eye'	3. Latin *dies* 'day'	4. SC *gazda* 'landlord'
	non-deterministic (overabundant)	5. Russian second locative	6. SC *dokument* 'document'	7. SC *doba* 'time'	8. Polish *ręka* 'hand'

This typology will be motivated with a sketch of each cell and references to full analyses.

Internal deterministic (fixed inflectional morphology)

Lexemes in this row (1-4) have paradigms with deterministic specification, but they show increasingly interesting external effects. Serbo-Croat *soba* 'room' is a canonical noun (one of thousands): internally its paradigm is determined by its inflection class specification, and externally its agreements are therefore consistently FEMININE. At the other extreme, we have *gazda* 'landlord' as analysed above.

Internal non-deterministic (inflectional morphology includes overabundance)

When we allow for overabundance, the same range of external behaviours can be found, from no external effect (5. Russian second locative: *v ètom cexu / cexe* 'in this workshop'), up to 8. Polish *ręka* 'hand', where a single paradigm cell, the LOCATIVE SINGULAR, shows overabundance, and induces alternative agreements (with Agreement Hierarchy effects).

Conclusions
1. Our typology takes us a step towards understanding (non-)determinism in inflection.
2. We have further evidence that internal and external feature specifications need not coincide.
3. Since we can specify these distinctions independently, we can use external (non-)determinism to better understand inflectional (internal) (non-)determinism.

References

Beniamine, Sacha (2021), One lexeme, many classes: Inflection class systems as lattices. In Berthold Crysmann & Manfred Sailer (eds.), *One-to-many relations in morphology, syntax, and semantics*, 23–51. Berlin: Language Science Press. doi: 10.5281/zenodo.4729789.

Bugarski, Ranko (2019), Past and current developments involving pluricentric Serbo-Croatian and its official heirs. In Tanneke Schoonheim & Johan Van Hoorde (eds) *Language variation: A factor of increasing complexity and a challenge for language policy within Europe: Contributions to the EFNIL Conference 2018 in Amsterdam*, 105-114. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences. https://efnil.org/documents/publications/.

Corbett, Greville G. & Wayles Browne (2018), Serbo-Croat: Bosnian, Croatian, Montenegrin, Serbian. In Bernard Comrie (ed.) *The World's Major Languages,* 3rd edition. 339-356. London: Routledge.

Guzmán Naranjo, Matías & Olivier Bonami (2021), Overabundance and inflectional classification: Quantitative evidence from Czech. *Glossa: a journal of general linguistics* 6(1): 88. 1–31. doi: 10.5334/gjgl.1626.

Kaye, Steven J. (2015), *Conjugation class from Latin to Romance: heteroclisis in diachrony and synchrony*. PhD thesis, University of Oxford.

Leko, Nedžad (2010), *Sintaksa imeničkih sintagmi u bosanskom i engleskom jeziku* (Lincom Studies in Slavic Linguistics 37). Munich: Lincom Europa.

Pellegrini, Matteo (2023), Flexemes in theory and in practice. *Morphology* 33. 361–395. Round, Erich R. & Greville G. Corbett (2020), Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24. 489–525. doi: 10.1515/lingty-2020-2060.

Spencer, Andrew (2020), Uninflectedness: Uninflecting, uninflectable and uninflected words, or the complexity of the simplex. In Lívia Körtvélyessy & Pavol Štekauer (eds) *Complex Words: Advances in Morphology*, 142–158. Cambridge: Cambridge University Press. doi:10.1017/9781108780643.009.

Thornton, Anna M. (2019), Overabundance: A Canonical Typology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds) *Competition in Inflection and Word-Formation* (Studies in Morphology, 5), 223-258. Cham: Springer. doi: 10.1007/978-3-030-02550-2_9.

Are lexically derived words more subject to overabundance?
Evidence from Hebrew non-concatenative morphology

Lior Laks
(Bar Ilan University)

This study examines the relations between overabundance and the component of the grammar where word formation takes place. Underlying this study is the assumption that the lexicon is an active component of the grammar, where word formation can take place. I make a distinction between three types of words with respect to formation and storage: basic non-derived words, e.g. *piano*; words derived and stored in the lexicon,
e.g. *pianist*; and words that are derived in the syntax, e.g. *pianos*. I will argue that words that are derived in the lexicon are more subject to morphological change, resulting in overabundance. In contrast, basic entries are stable, and words that are derived in the syntax are typically more predictable, and therefore overabundance is less likely to occur.

Semitic morphology relies highly on non-concatenative morphology, namely the combination of root and pattern (Berman 1978; Schwarzwald 1981, Bolozky 1978, Ravid 1990, Bat-El 1994, Aronoff 1994). Patterns indicate the prosodic structure of words, their vocalic patterns and their affixes (if any), e.g. *hitraxec* (*hitCaCeC*) 'wash oneself'. The study examines two case studies of doublets within Hebrew patterns.

Inchoative Verbs

Some verbs have doublets in other patterns, e.g. *nirtav* (*niCCaC*) ~ *hitratev* (*hitCaCeC*) 'get wet'. Doublet formation is mainly attested in *CaCaC* and *niCCaC* verbs, which take an additional form in *hitCaCeC*, due to morpho-phonological factors (Schwarzwald 1996, Bat-El 2002). Most verbs with doublets are inchoative verbs, which are assumed to be the output of derivational operations. I assume that such operations apply in the lexicon (Reinhart & Siloni 2005). For example, *yavaš* (*CaCaC*) 'become dry' and *nirtav* (*niCCaC*) 'become wet' have *hitCaCeC* doublets (*hityabeš* and *hitratev*), while *yarak* (*CaCaC*) 'spit' and *nizhar* (*niCCaC*) 'be careful' have no doublets. This is because the former are assumed to be derived entries, and the latter are basic entries. Moreover, *niCCaC* also hosts passive verbs, which are assumed to be derived in the syntax (Horvath & Siloni 2008), and have no doublets. For example, the passive verb *nimdad* 'be measured' has no *hitCaCeC* doublet like **hitmaded*.

Location nouns

Hebrew location nouns are formed in the *miCCaCa* pattern, e.g. *mispara* 'hairdresser shop' and *mitpara* 'sewing workshop'. Some of them have doublets in *maCCeCa* (Bolozky 1999, 2003), e.g. *mispara~maspera*. This takes place only in case there is a derivational relation between location nouns and verbs. For example, *mispara~maspera* are semantically related to **siper** 'cut hair'. I assume that such location nouns are derived from the verbal counterparts, as they refer to the location where the action that the verb denotes takes place. In contrast, location nouns that are not derived from verbs have no doublets, e.g. *mixlala~*maxlela* and *midšaʔa~* madšeʔa* 'lawn'. Such nouns are assumed to be stored as basic entries and therefore are not subject to doublet formation.

Assuming that outputs of derivational operations are listed in the lexicon differently than basic entries, the morphological component in the lexicon is sensitive to such differences. This predicts that certain types of morphological operations are more likely to involve non-determinism.

References

Aronoff, M. 1994. Morphology by Itself. Cambridge: MIT Press.

Bat-El, O. 1994. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory* 12, 572-596.

Bat-El, O. 2002. Semitic verb structure within a universal perspective. In J. Shimron (ed.), *Languages Processing and Acquisition in Languages of Semitic, Root-based, Morphology*. Amsterdam: John Benjamins. 29-59.

Berman, R.A. 1978. Modern Hebrew structure. Tel Aviv: University Publishing Projects. Bolozky, S. 1978. Word formation strategies in Modern Hebrew verb system: denominative

Verbs. *Afroasiatic Linguistics* 5, 1-26.

Bolozky, S. 1999. *Measuring productivity in word formation: The case of Israeli Hebrew*. Leiden: Brill.

Bolozky, S. 2003. Phonological and morphological variations in spoken Hebrew. In B.H. Hary (ed.), *Corpus Linguistics and Modern Hebrew*. Tel Aviv: Rosenberg School of Jewish Studies. 119-156.

Horvath, J. & T. Siloni. 2008. Active lexicon: Adjectival and verbal passives. In S. Armon- Lotem, G. Danon, and S. Rothstein (eds.), *Current Issues in Generative Hebrew Linguistics*. Amsterdam/ Philadelphia: John Benjamins. 105-134.

Ravid, D. 1990. Internal structure constraints on new-word formation devices in Modern Hebrew. *Folia Linguistica* 24, 289-347.

Reinhart, T. & T. Siloni. 2005. The Lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry* 36(3), 389-436.

Schwarzwald, O.R. 1996. Syllable structure, alternations and verb complexity: The Modern Hebrew verb patterns reexamined. *Israel Oriental Studies* 16, 99-112.

Turning surprise into expectation: Karakalpak valency-changing morphology

Matthew Zaslansky, Farrell Ackerman & Robert Malouf
UC San Diego, UC San Diego & San Diego State University)

Assumptions about biunique mappings between meaning and form reflect conceptions of parsimony within concatenative morpheme-based models of morphology (Siddiqi 2019) and define canonical mappings in Canonical Typology (Bond 2019); biuniqueness is conceptualized as a deterministic mapping (Determinism Assumption) from feature to exponent. The quantity and quality of surprising violations of the Determinism Assumption are instructively prodigious; among them, some individual languages display overabundance (Thornton 2019), multiple exponence (Harris 2017), variable affix order (Manova 2022) or periphrastic morphological expression (Ackerman and Stump 2004). This talk describes how all of them (and more) co-occur in the synthetic and periphrastic valency-changing morphology of Karakalpak, a Turkic language (Da'wletov et al. 2010) spoken in Uzbekistan and Kazakhstan. One finding of this description is that morphotactic restrictions on the ordering of various valency-changing affixes (particularly causative and passive) which occur in synthesis also occur in periphrasis. A second, more interesting finding of the description is that the Karakalpak passive and other valency-decreasing suffixes have been recruited in periphrastic constructions to participate in patterns of overabundance and multiple exponence typically associated mostly with the valency-increasing (causative) suffixes in other Turkic languages (Kulikov 1999).

Causatives are exponed by freely varying alternative suffixes appearing on synthetic verbforms (1a-1b) and participate in multiple exponence (1c), which creates the possibility for variable affix order (1d).

(1)   a.   *mʲen*   *wonɪ*   *tur-**dɪr**-dɪ-m*
           1SG.NOM   3SG.ACC   stand-**CAUS₁**-PST-1SG
           'I got him/her up.'

      b.   *mʲen*   *wonɪ*   *tur-**ʁɪz**-dɪ-m*
           1SG.NOM   3SG.ACC stand-**CAUS₂**-PST-1SG
           =1a.

      c.   *mʲen*   *wonɪ*   *tur-**ʁɪz-dɪr**-dɪ-m*
           1SG.NOM   3SG.ACC stand-**CAUS₂-CAUS₁**-PST-1SG
           =1a.

      d.   *mʲen*   *wonɪ*   *tur-**dɪr-ʁɪz**-dɪ-m*
           1SG.NOM   3SG.ACC stand-**CAUS₁-CAUS₂**-PST-1SG
           =1a.

Benefactives are one of many periphrastic converb constructions (consisting of a gerundial main verb plus an auxiliary) in which causative affixes occur on lexical stems (2a) or are distributed across the periphrase (2b).

(2)   a.   *Anel Berik-ke jaz-**dır**-**t**-**qız**-ıp ber-di*
              Anel Berik-DAT write-**CAUS$_1$**-**CAUS$_4$**-**CAUS$_2$**-CVB.B BEN-3.PST
              'Anel made Berik write it down [for someone].'

    b.   *Anel Berik-ke jaz-**dır**-**d**-ıp ber-**giz**-di*
              Anel Berik-DAT write-**CAUS$_1$**-**CAUS$_4$**-CVB.B BEN-**CAUS$_2$**-3.PST
              =2a.

Like the synthetic causative constructions in (1), the periphrastic causatives in (2) allow multiple exponence and variable affix ordering. Furthermore, periphrastic constructions allow the extension of these violations of determinism to passives:

(3)   a.   *Mağan usınıs   xat-ı jaz-ıp ber-**il**-di*
              1SG.DAT recommendation letter-3SG.POSS write-CVB.B BEN-**PASS**-CVB.B-3.PST
              'A letter of recommendation was written for me.'

    b.   *Mağan usınıs xat-ı jaz-**il**-ıp ber-**il**-di*
              1SG.DAT recommendation letter-3SG.POSS write-**PASS**-CVB.B BEN-**PASS**-3.PST
              =3a.

The passive (unlike the causative) has only one allomorph, so the kind of multiple exponence seen in (1c) and (1d) is impossible outside a periphrastic converb construction.

        The examples (1)–(3) exhibit mappings from one meaning to multiple forms. Periphrasis also creates the possibility for one form to map to multiple meanings. In a stacked converb construction, the dative *Aygerim-ge* can be interpreted as either the causee or the beneficiary and this is true whether the causative is marked in the gerund or either of the converbs:

(4)   a.   Anel   Aygerim-ge   jaz-dır-t-qız-dır-ıp   ber-ip   ket-ti
              Anel   Aigerim-DAT   write-CAUS-CAUS-CAUS-CAUS-CVB.B give-CVB.B   go-3.PST
              i. 'Anel has made someone write it down for Aigerim.'
              ii. 'Anel has made Aigerim write it down for someone.'

    b.   *Anel   Aygerim-ge   jaz-dır-t-qız-ıp*   *ber-dır-ıp*   *ket-ti*
              Anel   Aigerim-DAT   write-CAUS-CAUS-CAUS-CVB.B   give-CAUS-CVB.B go-3.PST

    c.   ?*Anel   Aygerim-ge   jaz-dır-t-qız-ıp*   *ber-ip*   *ket-tir-ti*
              Anel   Aigerim-DAT   write-CAUS-CAUS-CAUS-CVB.B   give-CVB.B   go-CAUS-3.PST

These apparently surprising violations of determinism in Karakalpak are interpretable as reflexes of the diachronic development of paradigmatic contrasts; comparison with other Turkic languages reveals the co-occurrence of overabundant allomorphs and their participation in multiple exponence (Schönig 1999), which in Karakalpak uniquely develops into variable affix orders in synthesis and periphrasis. By analogy, the contrasting passive markers are recruited into a system of similar distributions in synchrony.

References

Ackerman, Farrell & Gregory Stump. 2004. Paradigms and periphrastic expression: A study in realization-based lexicalism. In Louisa Sadler & Andrew Spencer (eds.), *Projecting morphology*, 111-157. Stanford, CA: CSLI.

Bond, Oliver. 2019. Canonical Typology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 409-431. Oxford: Oxford University Press.

Da'wletov, Abatbay, Ma'denbay Da'wletov, & Ma'mbetkerim Qudaybergenov. 2010. *Ha'zirgi Qaraqalpaq A'debiy Tili* [The Modern Literary Karakalpak Language]. Nukus: Bilim.

Harris, Alice C. 2017. *Multiple exponence*. Oxford: Oxford University Press.

Kulikov, Leonid I. 1999. Remarks on double causatives in Tuvan and other Turkic languages. *Aikakauskirja Journal de la Société Finno-Ougrienne* 88. 49-58.

Manova, Stela. 2022. Ordering restrictions between affixes. To appear in P. Ackema, S. Bendjaballah, E. Bonet & A. Fábregas (eds.), *The Wiley Blackwell Companion to Morphology*. Blackwell. Accessed via https://ling.auf.net/lingbuzz/006219.

Schönig, Claus. 1999. Die türkischen Kausativsuffixe mit anlautendem *G~K und/oder auslautendem *-z (I). *Ural-Altaische Jahrbücher*. 183-201.

Siddiqi, Daniel. 2019. Distributed Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 143-165. Oxford: Oxford University Press.

Thornton, Anna M. 2019. Overabundance: A canonical typology, in F. Rainer, F. Gardani, W. U. Dressler, and H. C. Luschützky (eds.), *Competition in Inflection and Word-Formation*. Cham: Springer, 223–58.

Underdetermined morphology in Daakaka and Dalkalaen (Oceanic, Vanuatu)

Cat Butz & Kilu von Prince
(HHU Düsseldorf & HHU Düsseldorf)

Keywords: Oceanic, vowel harmony, underdetermination, probabilistic morphophonology

In this talk, we present novel data from two endangered Oceanic languages which was collected during fieldwork in Vanuatu. Daakaka and Dalkalaen exhibit morphological variation which seems to be, at least in part, probabilistic rather than determined by strict rules. TAM markers can take a range of different shapes, and to some extent, these realizations are reliably determined by the context. For example, the **Daakaka** potential marker will always cliticize to a preceding subject marker:

> *Da=p   lyung   vyan   pyan!*
> 1D.IN=POT bathe   go   under
> "Let's dive!"

However, in some contexts, different realizations of the same sequence are possible. For example, both the following pair of examples contain the sequence of a realis marker and the verb *esi* 'see', but in one case, the realis marker is realized as mono-syllabic *ma*, while in the other case, it cliticizes to the verb.

> *ma   esi   na   ya=m   du   ngapngap.*
> REAL   see   COMP   3PL=REAL CONT   rest
> "he saw that they were sleeping"

> *Mw=esi   na   mwe   tung myaek   na   mwe   tung   myaek   ten*
> REAL=SEE COMP REAL   dark night   COMP   REAL   dark   night   very
> "he saw that he was completely black"

Another type of variation concerns the vowels of monosyllabic TAM markers, which can be vowel-disharmonic or vowel-harmonic. This choice, too, may be specific to a certain sequence of items rather than explainable through a wider generalization. In the following examples, the realis marker is disharmonic, but the distal marker harmonizes with the subsequent verb, even though it's the same verb in both cases:

> *ma   minyes*
> REAL   be.different

> *na   ti=minyes*
> COMP   DIST=be.different
> "which is different"

Likewise, the allomorphic manifestations of vowels in TMA markers in **Dalkalaen** can be grouped into five phonological categories: disharmonic, front/back harmonic, fully harmonic, schwa, and zero. Their conditioning is also probabilistic, as illustrated by the sequence of a realis marker and the verb *keene* 'want'. The surface realization of the marker is underdetermined, with all five categories being acceptable and contributing a certain percentage of the overall set of occurrences:

> *lo-g*   *Ø-ma/mi/me/mə/m-keene*   *Ø-na-kala*
> inside-1 3-REAL-want   POT-1-say
> "I want to speak"

The most clearly recognizable patterns in the allomorph distribution are all lexically conditioned. For example, the realis marker is realized consistently with a zero vowel before a copula:

> *ro-m-do-bang~bang*   *Ø-m-ii*   *ta*
> 2DU-REAL-CONT-play~play 3-REAL-COP   DEM:PROX
> "they were playing like this"

However, the above pattern is an exception in its reliability. In almost all other environments, the surface realization of the realis marker can only be predicted with partial accuracy.

We propose that this variation is better modelled as a function of preferences regarding word shapes and sequence frequencies than as generated by a fixed set of general rules.

Verbal imperfections: a comparative study of Ukrainian and Russian

Kateryna Krykoniuk & Neil Bermel
(University of Sheffield & University of Sheffield)

Tyschenko (2012) identified 12 morphological features of Ukrainian verbs that distinguish them from Russian. These features include, for example, the presence of a synthetic future tense in the imperfective aspect of verbs (e.g., *знаймо, писатимеш*), alongside the more conventional analytic future tense (e.g., *будемо знати, будеш писати*), providing an indication that usage in Ukrainian may be more open to overabundance as a type of non- deterministic outcome than Russian is. Defectivity, on the other hand, is a well-documented phenomenon in Russian in the 1sg. nonpast (Baerman 2008, Sims 2015), but is said to be absent in analogous contexts in Ukrainian; yet no studies, to the best of our knowledge, have investigated the nuances of verbal defectivity in Ukrainian and Russian, and how and why this phenomenon differs between these languages.

Differences in the verbal morphology in Ukrainian and Russian, given their intertwined historical development, provide linguists with a unique opportunity to gain a deeper understanding of these differing non-deterministic outcomes in paradigm cells and thus, in general, shed light on determinism assumptions in morphology. To pinpoint these differences, we compare 100 Russian verbal paradigms with defective cells to their Ukrainian counterparts, first focusing on phonological features (i.e., syllable patterns, consonant clusters, palatalization, and vowel description) and prosodic elements (i.e., stress patterns). The analysis also includes grammatical features, encompassing a morphological structure of verbs and cell status (i.e., the number of morphological forms per cell—for a biunique mapping between form and function of a cell, it is 1; for defective cells, it is 0, and for overabundant cells, it is 2+), as well as etymological features, namely verb origin and origin match (i.e., whether the origin of verb matches in both languages). Additionally, the relative frequency of morphological forms (extracted from ukTenTen and ruTenTen, SketchEngine, (Jakubíček et al. 2013)) is considered in the analysis. All these linguistic features are collected, using formalisation techniques tailor- made for our research objectives.

Subsequently, we apply k-prototypes clustering—a method adept at handling mixed variables (e.g., Szepannek 2018)—to map, compare and analyse the patterns within these features. This method allows us to identify trends linked to the morphological defectivity and overabundance of cells and provide insights into their emerging environment. As a result, each morphological paradigm of a verb in the sample forms a distinct cluster of data points, which is further mapped against other morphological paradigms. In mapping the clusters of morphological paradigms, we also incorporate the classification of verbs based on lexical aspects such as states, accomplishments, achievements, and activities. Finally, we perform a comparative analysis of these cluster spaces, examining the differences between Ukrainian and Russian verbs.

These findings have been further contextualized within a broader picture. Applying the principles of usage-based linguistics (Bybee 2023), the study explores how language standardisation and social experience with language in Ukrainian and Russian societies influence the grammatical aspects of language (cf. Author XXXX). The degree of linguistic prescriptivism in these two societies has been investigated to assess its impact on grammar norms and usage.

References

Krykoniuk, K. & Bermel, N.

Baerman, M. (2008). Historical observations on defectiveness: the first singular non-past. Russian linguistics, 32(1), 81-97.

Bybee, J. (2023). 1 What Is Usage-Based Linguistics?. The Handbook of Usage-Based Linguistics, 9.

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013, July). The TenTen corpus family. In 7th International Corpus Linguistics Conference CL (pp. 125-127).

Sims, A. D. (2015). Inflectional Defectiveness (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107053854

Szepannek, G. (2018): clustMixType: User-Friendly Clustering of Mixed-Type Data in R, The R Journal 10/2, 200-208, doi:10.32614/RJ2018048.

Тищенко, К. (2012). Правда про походження української мови. Український тиждень, 39(256), 21-64.

Procedural knowledge and memory structure in non-deterministic morphological behavior: A(nother) computational study of English past tense

Micha Elsner & Andrea D. Sims

(The Ohio State University & The Ohio State University)

Wug tests reveal extensive gradience in the acceptability of inflected forms (Bybee & Moder 1983). In English past tense verbs, this gradience stems from competition between memorized forms and procedural knowledge. Yet computational models tend to focus on one or the other sort of knowledge. Historically, connectionist/neural network models of inflection (Kirov & Cotterell 2018) and some probabilistic grammar models (Albright & Hayes 2003) have extracted procedural knowledge from the lexicon, without accessing specific lexical items at prediction time. Such systems have been criticized (Kodner & Khalifa 2022, Corkery et al. 2019, Xu & Pinker 1995); some over-predict regularity, some under-predict, and some fail in other ways. In contrast, analogical/exemplar models (Skousen 1989) have historically focused on capturing memorization (e.g. neighborhood-similarity) effects, but the need to specify dimensions of similarity by hand limits their ability to model procedural knowledge. We present a model that seeks to take advantage of the best of both types of model.

We start from the idea that memorized forms and procedural knowledge are closely integrated, with memory filling in underspecified or variable information in production. We introduce a neural model with lexical access during production as well as learning. It implements four-part analogy: the input to an inflection task includes the base form of a lexeme and target cell information and also an analogical exemplar---a pair of wordforms of a different lexeme in the same cells. The model is able to produce wordforms procedurally based on the grammatical properties of the wordform, *and/or also* to copy from the exemplar. At test time, we make predictions with varied exemplars, generating gradient predictions of the well-formedness of two candidate output forms each for English wug verbs from the SIGMORPHON 2021 shared task (Kodner et al. 2022). We compare these predictions to human ratings from the shared task.

Figure 1 shows that in each word pair the form preferred by humans was also generally preferred by the model (blue lines). Comparing our model output to others (Table 1), our model performs comparably to the Minimal Generalization Learner (Albright & Hayes 2003) as reimplemented by Wilson & Li (2021), the shared task winner, which tries to find phonological environments in which rules can be reliably applied. The best neural learner from the shared task, Calderone et al. (2021), performs less well. We show that the exemplar component allows our model to better predict low-type-frequency patterns, while reducing (though not eliminating) under-regularization errors. Broadly consistent with results of previous analogical models (Skousen 1989), affinity for a close neighbor can override the influence of a large number of more distantly-related forms within the model (Table 2). Unlike in pre-neural exemplar models, however, the definition of "close" is induced by the network, not predefined by the linguist.

By integrating memorized knowledge and procedural knowledge into a single model, our model represents a step in the direction of more cognitively plausible modeling of non-deterministic morphological behavior.
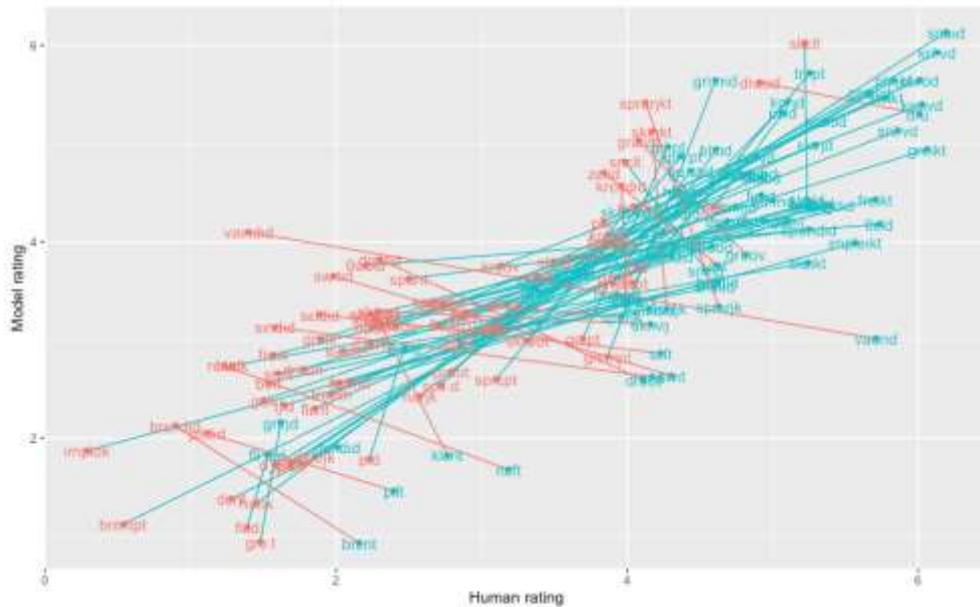
Figure 1: Model predictions (with lexeme as fixed effect) vs. human ratings on English development data. Each lexeme has two alternative past tense forms. Human-preferred wordforms in each pair are shown in blue; dispreferred ones are in red. Lines connect paired wordforms of the same lexeme. Ones in which the model preference matches human preference are shown with blue lines; ones with a mismatch are shown with red lines.

Model	AIC	Mismatch %
Ours	-106.1	**24**
Wilson & Li (2021)	**-113.3**	27
Calderone et al. (2021)	-43.0	NA

Table 1: Wug goodness-of-fit metrics on English development data: AIC (lower = better) and percentage of mismatch pairs (lower = better). Mismatch pairs are ones in which the model-predicted wordform is different than the human-preferred one.

Feature	R^2 alone (in %)	R^2 ablated (in %)
Gang consistency measure 1	54.7	**62.2**
Gang consistency measure 2	37.2	65.0
Nearest neighbor	**56.4**	65.1
Type frequency of eliciting set	**55.5**	66.6
Most frequent word in eliciting set	46.1	65.2
Non-candidate	35.8	66.7
Full statistical model		66.9

Table 2: Relative importance of various lexical neighborhood measures as predictors of human acceptability ratings. The "eliciting set" is the number of exemplar lexemes in the training set that would elicit the target outcome. "Non-candidate" means that the model never produces a target form, regardless of exemplar. R^2 alone gives the variance explained by a linear fit using a single predictor (higher = better); R^2 ablated gives the variance explained by the full model with a single predictor removed (lower = better).

References

Albright, Adam and Bruce Hayes (2003), Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 119-161.

Bybee, Joan and Carol Moder (1983), Morphological classes as natural categories. *Language* 59(2), 251-270.

Calderone, Basilio, Nabil Hathout, and Olivier Bonami (2021), Not quite there yet: Combining analogical patterns and encoder-decoder networks for cognitively plausible inflection. *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 274-282. Association for Computational Linguistics.

Corkery, Maria, Yevgen Matusevych, and Sharon Goldwater (2019), Are we there yet? Encoder-decoder neural networks as cognitive models of English past tense inflection. In A. Korhonen, D. Traum, and L. Màrquez (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3868-3877. Association for Computational Linguistics.

Kirov, Christo and Ryan Cotterell (2018), Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics* 6, 651-665.

Kodner, Jordan and Salam Khalifa (2022), SIGMORPHON-UniMorph 2022 Shared Task 0: Modeling inflection in language acquisition. In G. Nicolai & E. Chodroff (eds.), *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 157-175. Association for Computational Linguistics.

Kodner, Jordan, Salam Khalifa, Khuyagbaatar Batsuren, Hossep Dolatian, Ryan Cotterell, Faruk Akkuş, Antonios Anastasopoulos, Tasas Andrushko, Aryaman Arora, Nona Atanelov, Gábor Bella, Elena Budianskaya, Yustinus Ghanggo Ate, Omer Goldman, David Guriel, Simon Guriel, Silvia Guriel-Agiashvili, Witold Kieraś, Andrew Krizhanovsky, Natalia Krizhanovsky, Igor Marchenko, Magdalena Markowska, Polina Mashkovtseva, Maria Nepomniashchaya, Daria Rodionova, Karina Sheifer, Alexandra Serova, Anastasia Yemelina, Jeremiah Young, and Ekaterina Vylomova (2022), SIGMORPHON-UniMorph 2022 Shared Task 0: Generalization and typologically diverse morphological inflection. In G. Nicolai and E. Chodroff (eds.), *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 176-203. Association for Computational Linguistics.

Skousen, Royal (1989), *Analogical modeling of language*. Dordrecht: Kluwer.

Wilson, Colin and Jane S. Y. Li (2021), Were we there already? Applying minimal generalization to the SIGMORPHON-UniMorph shared task on cognitively plausible morphological inflection. *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 283-291. Association for Computational Linguistics.

Xu, Fei and Steven Pinker (1995), Weird past tense forms. *Journal of Child Language* 22(3), 531-556.

Predicting non-deterministic outcomes in overabundant paradigms

Martin Alldrick

(University of Surrey)

Keywords: non-determinism; morphology; overabundance; predictive factor; Slovak

Morphological overabundance (Thornton, 2011) poses a challenge to the deterministic hypothesis in that it allows for two forms to express the same meaning (Divjak et al., 2021). Assuming that it is indeed possible for the same meaning to be conveyed by two distinct forms, it raises questions whether a speaker's choice of variant form is random or is driven by large scale rules.

Predictive (predisposing) factors have been used to predict outcomes across various scientific fields. By applying the notion of predictive factors to linguistics, it should be possible to examine whether a speaker's choice of variant form can be predicted or not, and by extension, whether choice of variant form is governed by large scale rules or determined on a lexeme-by-lexeme level.

In this talk, I will examine to what extent overabundant paradigm cells in Slovak are non-deterministic; and whether the choice of variant form is predictable based on idiolectal, demographic, and grammatical factors. Working with Sokolová's (2007) categorisation of declensional paradigms in Slovak, two overabundant paradigmatic cells in the masculine inanimate paradigm were investigated: genitive singular and instrumental plural. The acceptability of variant forms among native speakers was tested using a forced choice survey similar to Bermel and Knittl (2012) and demographic information (age, gender, education, dialectal region) was noted.

GENDER	Masculine inanimate			
PARADIGM	'dub'		'stroj'	
GLOSS	oak		machine	
NUMBER	sg.	pl.	sg.	pl.
Nom.	Ø	y	Ø	e
Gen.	a / u	ov / Ø	a / u	ov / Ø
Inst.	om	ami / mi	om	ami / mi

Table 1: excerpt of Slovak noun paradigms

The results showed that it is impossible to predict a speaker's choice of inflectional form from their previous behaviour. There is no link between how a speaker inflects two different words. Furthermore, speakers who use one or other form exclusively are equally statistically normal. In contrast to the assumption that overabundance is a transitional state resulting from the coexistence of multiple grammars in a speech community (e.g. Cummins, 1995), these findings confirm (e.g. Dorian, 2010) that overabundance can also occur due to persistent unpredictability in the grammar of individual speakers.

The only demographic variable that was found to have an effect was the speaker's gender, and then only in the genitive case. Similar to the Czech prepositional case (Bermel, 1993), prepositional government in the Slovak genitive syntactically triggers -u inflectional suffix. Men analogised this rule

across all forms of government, whilst for women, instances of overabundant noun forms not under prepositional government remained in free variation.

In summary, preference for one form or other cannot be considered as a predictive factor for the selection of a variant form. This suggests that the selection of a variant form is made each time a speaker is confronted with a variant noun rather than on the basis of large-scale licensing rules. The only demographic variable found to be a predicting factor in morphological variation was gender suggesting that women may be more tolerant of non-deterministic outcomes in their mental grammars that men.

References

Bermel, Neil (1993), Sémantické rozdíly v tvarech českého lokálu, *Naše řeč*, 76(4), 92-98.

Bermel, Neil and Knittl, Luděk (2012), Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech, *Corpus Linguistics and Linguistic Theory*, 8(2), 241-275.

Cummins, George (1995), Locative in Czech: -u or -e: Choosing locative singular endings in Czech nouns, *Slavic and East European Journal*, 39, 241-260.

Divjak, Dagmar; Milin, Petar; Ez-Zizi, Adnane; Józefowski, Jarosław and Adam, Christian (2021) What is learned from exposure: an error-driven approach to productivity in language, *Language, Cognition and Neuroscience*, 36(1), 60-83.

Dorian, Nancy (2010). Investigating variation: The effects of social organization and social setting: Oxford University Press.

Morfológia slovenského jazyka (1966), Ružička, Jozef (ed.), Bratislava: Vydavateľstvo SAV.

Sokolová, Miroslava (2007), Nový deklinačný systém slovenských substantív, Prešov: Universitatis Prešoviensis.

Thornton, Anna (2011), Overabundance (multiple forms realizing the same cell): A non-canonical phenomenon in Italian verb morphology. *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358-381.

WS22 The Limits of the Comparative Method: Innovative Approaches to Understanding Orphan Languages

Timucua, A language isolate of Florida:
New evidence for a connection to Muskogean

George Aaron Broadwell
(University of Florida)

Keywords: Muskogean, isolates, switch-reference, borrowing, ethnographic evidence, orphans

Background: Timucua (ISO: tjm) is a language isolate formerly spoken in Northern Florida and adjacent parts of Georgia. In the period from 1612-1635, seven religious books were published in Timucua, including bilingual Spanish-Timucua catechisms and confessionals. A brief *Arte* and two secular letters also survive. The last known speakers of Timucua lived in the 18th century. However, despite the lack of speakers, there is a large corpus of available material.

Possible genetic connections: Crawford (1988) proposed connections to the Muskogean family, but on the basis of limited evidence. This paper seeks to provide additional evidence for the proposal of a relationship between Timucua and Muskogean by citing new lexical, morphological, and ethnographic evidence.

New approaches in lexical evidence: At the time of Crawford's proposal of a Muskogean connection, linguists lacked 1) a comprehensive set of Muskogean cognate sets and 2) a reliable dictionary of Timucua. The last three decades have seen substantial improvement in both areas.

This has allowed the identification of about forty-five sets of lexical comparanda, due either to genetic relationship or to borrowing between the two language families.

New approaches in morphological evidence: The development of a more comprehensive and reliable treatment of Timucua morphology has allowed the identification of the inflectional morphology associated with nominal possession, clause chaining, and switch-reference. All three areas show similarity to the comparable systems in Muskogean and strengthen the argument for a genetic relationship between Timucua and Muskogean.

New approaches in ethnographic evidence: Better translations of portions of the Timucua corpus dealing with ritual and taboo have allowed the identification of many cultural practices shared with the Seminole of Florida, a contemporary Muskogean people. These shared practices point to prolonged contact and interaction between Timucua and Muskogean people.

References

Crawford, James M. 1988. "On the Relationship of Timucua to Muskogean." In *In Honor of Mary Haas*, edited by William Shipley, 157–64. Berlin: De Gruyter Mouton

How to leave the linguistic orphanage: Ket and its Siberian family

Eugen Hill, Simon Fries, Laura Günther, Natalie Korobzow & Svenja Bonmann
(Universität zu Köln, Universität zu Köln, Universität zu Köln, Julius-Maximilians-Universität Würzburg & Universität zu Köln)

Keywords: Yeniseian, Ket, Yugh, sound correspondences, historical-comparative method

While living in exile during the second world war, Andrej P. Dul'zon developed an interest in Ketology, i.e., the scientific study of the Ket language. Ket was, and still is, spoken in a remote region of Inner Asia, alongside the middle reaches of the Yenisei River. Dul'zon became interested in this linguistic orphan – Ket had no known living relatives – and started to study its different dialects: Northern, Central and Southern Ket and a differing variety known as Sym-Ket. Over the course of the 1950s, Dul'zon gradually realized that what was known as Sym-Ket actually was a different, yet genetically related language which is nowadays known as Yugh, and Dul'zon started to search for other relatives that had potentially died out (Dul'zon 1961, 1963). What brought him to the realization that Sym-Ket/Yugh was a different, yet related language was the detection of recurring, i.e. systematic sound correspondences between Northern, Central and Southern Ket on the one hand and Sym-Ket/Yugh on the other. In other words: Dul'zon made use of the Historical-Comparative Method.

Nowadays it is an established fact that the Yeniseian (also Yeniseic, abbreviated: Yen.) language family once consisted of six members: Ket, Yugh, Pumpokol, Arin, Assan and Kott. Apart from Ket, which has very few competent speakers by now, all Yeniseian languages are extinct, notably including Yugh which died out in the second half of the 20th century. Ket, Yugh and Kott are reasonably well-attested (cf. Werner 1997a,b, 1998; Vajda 2004; Georg 2007; Kotorova & Nefedov 2015), whereas our limited knowledge of Arin, Assan and Pumpokol is based on fragmentary word-lists from the 18th and 19th centuries, also discovered by Dul'zon in Russian archives and early printed works on Siberian history and geography (cf. Werner 2005). Lexical correspondences and striking similarities in grammar (cf. I. Verner 1969; G. Verner 1988; G. Starostin 1995) must be explained as inheritance from a common ancestor, Proto-Yeniseian (PY), which may have been spoken some 2500-3000 years ago (cf. e.g., Werner 2005: 15, Fortescue & Vajda 2022: 238-240, 277) somewhere near the headwaters of the Yenisey, the Ob or the Irtyš rivers (to judge by toponymic data, see Dul'zon 1959a,b; Maloletko 1992).

In this talk, we want to stress the importance of an approach to so-called linguistic orphans which is both open-minded and methodologically rigorous, giving examples from Ket and other Yen. languages. Reputed linguistic isolates can lose their status, if – and only if – a meticulous comparison of their phonological data with those of other supposedly unrelated languages reveals recurring sound correspondences which indicate a genetic relation. This can then result in the well-substantiated postulation of a hypothetical new family and the reconstruction of grammatical aspects of the last common ancestor of the members of this family. Ket was a perceived orphan in the mid-20th century, historical-comparative linguistics then led to the identification of a living relative, and in 2024, Ket is again an orphan – yet this time actually. Similar results may be expected for other prima facie linguistic orphans.

References

Dul'zon, Andrej P. (1959a), Geografičeskie nazvaniija Zapodnoj Sibiri kak istočnik eë drevnej istorii [Toponyms of Western Siberia as a source on its ancient history], *Nekotorye voprosy drevnej istorii Zapodnoj Sibiri*. Tomsk: Tomskij oblastnoj kraevedčeskij muzej, 24-28.

Dul'zon, A. P. (1959b), Ketskie toponimy Zapodnoj Sibiri [Ket toponyms of Western Siberia], *Učёnye zapiski, Tomskij gos. ped. institut* 18, 91-111.

Dul'zon, Andrej P. (1961), Slovarnye materialy XVIII v. po enisejskim narečijam [18th century lexical materials of the Yeniseian dialects]. *Učёnye zapiski Tomskogo pedagogičeskogo instituta*, 19 (2), 152–189.

Dul'zon, Andrej P. (1963), Ketskie narečija pervoj poloviny XVIII veka [Ket dialects of the first half of the eighteenth century]. *Trudy Tomskogo oblastnogo kraevedčeskogo muzeja* 6 (2), 38–44.

Fortescue, Michael and Edward Vajda (2022), *Mid-Holocene Language Connections between Asia and North America*. Leiden-Boston: Brill.

Georg, Stefan (2007), *A Descriptive Grammar of Ket (Yenisei-Ostyak). Part 1: Introduction, Phonology, Morphology*, Folkestone: Global Oriental.

Kotorova, Elizaveta and Andrej Nefedov (eds.) (2015), *Comprehensive Dictionary of Ket with Russian, German, and English translations*, 2 volumes, München: Lincom Europa.

Maloletko, Aleksej M. (1992), *Paleotoponimika*, Tomsk: Tomskij gos. universitet.

Starostin, Georgij S. (1995). Rekonstrukcija praenisejskoj glagol'noj sistemy [Reconstruction of the Proto-Yeniseian verb system], *Moskovskij lingvisticheskij žurnal* 1, 129-173.

Vajda, Edward J. (2004), *Ket*, München: Lincom.

Verner, Izol'da G. (1969), Voprosy imennoj klassifikacii v sovremennyx enisejskix dialektax [Questions of noun classification in modern Yeniseian dialects], *Voprosy filologii* (*Učёnye zapiski, Omskij gos. ped. institut* 52), 170-179.

Verner, Genrix K. [= Werner, Heinrich] (1988), Opyt rekonstrukcii obščeenisejskoj deklinacionnoj modeli [Reconstructing the Common Yeniseian declension system], *Voprosy jazykoznanija* 5, 84-95.

Werner, Heinrich [= Verner, G. K.] (1997a), *Das Jugische (Sym-Ketische)*, Wiesbaden: Harrassowitz.

Werner, Heinrich (1997b), *Die ketische Sprache*, Wiesbaden: Harrassowitz.

Werner, Heinrich (1998), *Abriß der kottischen Grammatik*, Wiesbaden: Harrassowitz.

Werner, Heinrich (2005), *Die Jenissej-Sprachen des 18. Jahrhunderts*, Wiesbaden: Harrassowitz.

Internal reconstruction of body part terms in a language isolate

Jeremy Coburn & Bonny Sands
(Indiana University & Northern Arizona University)

Keywords: isolate, internal reconstruction, body parts, Hadza, grammaticalisation

We investigate body part terms in Hadza, a language isolate spoken in Tanzania, using a two-pronged approach. First, in a bottom-up approach, we identify frequently-recurring syllables in Hadza terms as possible evidence for frozen morphology (cf. Miller 2023), e.g., a 3rd person possessive /*ʔá-/ prefix on certain body part terms (and kinship terms) (cf. Miller 2023, Miller 2016) and a /*-kʷʰa/ 1st person singular suffix (cf. Sands et al. forth.). We evaluate these reconstructions in part by examining verbs which appear to include a fossilized morpheme, e.g. /*-**kʷa**/ derived from /ʔú**kʷʰá**/ 'hand' (involving loss of a laryngeal contrast), e.g., /*t͡ʃʼú-kʷà/ 'dig', /*wat͡ʃiː-kwa/ 'to flick water off of hands', /*!ʼo-!ʼoː-kwa/ 'pluck'. A similar loss of aspiration occurs in nouns apparently derived from the root /*ʔá**kʷʰá-**/ 'eye', e.g. /*lémbé-**kwa**/ 'eyebrow'. The term 'tears' /*ɦáts'á-**pʰú**kʷá-/ 'tears', however, appears to have retained the aspirated consonant of /ʔu**pʰu**kʷá-/ 'leg, foot' (i.e. 'track, trail'). We diagram body part terms, indicating the ones with the proposed fossilized morphemes and discuss what is known about their context of use (e.g. humans vs. animals, 1st vs. 3rd person), e.g. for /ʔátʰàmá-/ 'blood'.

Second, in a top-down approach, we identify semantic changes that are cross-linguistically common involving a select number of body part terms. These include words referring to spatial semantics connected to body part terms (cf. Kuteva et al. 2020) and semantic shifts based on cross-linguistically common polysemies (cf. List et al. 2018). We find little support for the reconstruction of spatial terms as body parts but do find one botanical source, i.e. /t͡ʃiki-/ 'inside' (< /t͡ʃiki-/ 'thicket'), also occurring in the verb /t͡ʃi-t͡ʃiki-ʔe-/ 'interfere' (i.e. to be in the thick of things). The spatial term /puɦu-/ 'on' likely appears on the noun /*pú-ǀʰé/ 'spleen' together with an unknown root (possibly with an original meaning of 'stomach', the organ to which the spleen generally sits atop). This methodology must take into account cultural differences, e.g. body part terms are often used in descriptions of emotions cross-linguistically, we do not find this to be the case in Hadza. Instead, at least one emotion word (ŋǀúkúsúnà- 'be sad') may be internally reconstructed as being connected to a verb related to eating (ŋǀùʔà- 'be astringent (of food)'). Emotion words and phrases in Hadza are quite infrequent and more than one-third are loans. While no emotion terms in Hadza appear to derive from organs such as 'heart' or 'liver' that are cross-linguistically common in such expressions, terms connected to loans for 'pimple' and 'pus' do occur. The identification of loans is a crucial step in distinguishing rare patterns that occur in loan strata versus those that reflect earlier, inherited grammatical patterns. We discuss the role of phonotactics in identifying inherited vs. loan items (cf. Elderkin 1978, Miller 2023, Sands et al. forth.) as well as the distribution of lexical doublets and triplets.

References

Elderkin, E. D. (1978), *Loans in Hadza: Internal evidence from consonants*. (Occasional Paper, Dept. of Foreign Languages, 3). Dar es Salaam: University of Dar es Salaam.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long & Heiko Narrog (2020), *World Lexicon of Grammaticalization*, 2nd rev. edn. Cambridge: Cambridge University Press.

List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi & Robert Forkel (2018), CLICS2: An improved database of cross-linguistic colexifications assembling lexical data with the help of cross-linguistic data formats. *Linguistic Typology* 22(2), 277-306.

Miller, Kirk (2016), Hadza kinship terms. in S. Shah & M. Brenzinger (eds), *Khoisan Languages and Linguistics: Proceedings of the 5th International Symposium, July 13-17, 2014, Riezlern/Kleinwalsertal,* (Quellen zur Khoisan-Forschung, 34), Cologne: Rüdiger Köppe, 277-333.

Miller, Kirk (ed.) in collaboration with Mariam Anyawire, G.G. Bala & Bonny Sands (2023), *A Hadza Lexicon and Etymological Dictionary*. Mang'ola, Tanzania. Unpublished manuscript.

Sands, Bonny, Andrew Harvey, Maarten Mous & Mauro Tosco (2023), Why Hadza is (probably) not Afroasiatic: a discussion of Militarev's "Hadza as Afrasian?" *Journal of Language Relationship* 21(2), 317-328.

Linguistic Palaeontology for *Hard to Study* Histories:
Mapping Morphology to Meaning & People to Places

Fabian Zuk
(CNRS-LLACAN)

Linguistic palaeontology is the practice of drawing inferences about the material culture, social structures, mythologies, beliefs and the geographic location of speakers based on evidence of often ancient, generally reconstructed, language (Mallory 2020). Because orphan languages like ancient ones are a punctual remnant of a once larger language community, and both are characterised by an extreme poverty of horizontal *comparanda,* we propose that cross linguistic comparison and traces left behind in neighbouring languages may bring clarity to the earlier history of language orphans.

To demonstrate this process, we take the case study of the early Frankish kings. Hidden among fragmented tellings of an origin story linking the royal family with the legendary city of Troy (Yavuz 2015), evidence of a more likely migration of the early Germanic Franks into Gaul has been preserved in the Latin written tradition, often hidden behind a veneer of language contact (Zuk, 2017). This talk focuses on a passage from the 6th century scholar Gregory of Tours, who, recounting the contemporary legend of the Merovingian rulers of his time, writes:

> *Ferunt etiam tunc Chlogionem utilem ac nobilissimum in gente sua regem fuisse Francorum qui apud Dispargum castrum habitabat, quod est in terminum Thoringorum*
>
> (MGH SS res. Merov. Krusch 1969: 58)

> 'They still recount that thereupon Clodio, suitable and truly noble by his birth, was made king of the Franks, and he dwelled near **Dispargum** Castrum which lies at the boundary of the Thuringians'

Gregory's account presents several linguistic puzzles, among which the identity of said <Dispargum castrum>. Looking within Latin alone, the attestation is meaningless; <Dispargum> appears on no maps, nor is it an obvious representation of any known place in any neighbouring language. However, by considering the synchronic vocabulary of Germanic through the lens of language contact, the passage becomes interpretable: <**parg**> represents P(roto)-G(ermanic) **bergaz* 'refuge' mirrored in the Latin *castrum* 'fortress', while <**dis**>**,** we argue**,** is a Late Latin representation of PG **Þēođ(ō)z* 'the people.GEN.PL.' with the regular substitution of Germanic /Þ/ by Late Latin /d/ and of the Germanic [đz] cluster by Latin /s/. The Latin form likewise reflects the opening of PG **e → Frk **æ, and final vowel apocope within West Germanic.

(1) a. Late Latin   ***Di-s***   ***parg-um***   ***castr-um***
         ?-?   ?- ACC.SG   fortress-ACC.SG

     b. Germanic   ***Þḗođ-(ō)z***   ***berg-a^n***
         people-GEN.PL   refuge-ACC.SG

On etymological and archaeological grounds, we connect <Dispargum> (1a) with the Roman *castrum Deutonis* (Encyclopedia Britannica, Editors 2023; Everett-Heath 2019), where *Ꝺeutonis* is a Late antique representation of classical *Teutones* 'the Teutons, i.e. the people', the change of <t> → <d>, shedding light on the nature of the Germanic source language itself. We hereby place imperial *castrum Deutonis*, early-medieval *Dispargum* and modern *Duisberg* in a chronologically and phonologically derived relation, refining earlier hypotheses by Everette-Heath (2000: 120; 2019) hereby shedding light not only on the name, *Θēoḍzberg* (1b), but on the origin of the Franks and their language. We conclude this talk with notes on those linguistic features of both Germanic and Late Latin which can be inferred from palaeolinguistic practice and suggest potential application to the study of modern orphan languages.

References

Zuk, Fabian. 2017. Finding Troy in the Rhineland: Phonology, Fredegar and the Frankish Founding Myth. In *Temas y Perspectivas de la Historia*. Salamanca: Universidad de Salamanca.

Encyclopedia Britannica, Editors. 2023. Duisburg. *Encyclopedia Britannica*. //www.britannica.com/place/Duisburg. (19 October, 2023).

Everett-Heath, J. 2000. *Place Names of the World - Europe: Historical Context, Meanings and Changes*. Springer.

Everett-Heath, John. 2019. Duisburg. In *The Concise Oxford Dictionary of World Place Names*. Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780191882913.001.0001/acref-9780191882913-e-1936. (19 April, 2022).

Krusch, Bruno (ed.). 1969. *Munumenta Germaniae Historica Scriptores Rerum Merovingicarum*. 2nd edn. Vol. 1. Hannover.

Mallory, Fintan. 2020. The Case Against Linguistic Palaeontology. *Topoi* 40(1). 273–284. https://doi.org/10.1007/s11245-020-09691-5.

Yavuz, Nurgül Kıvılcım. 2015. *Transmission and Adaptation of the Trojan Narrative in Frankish History between the Sixth and Tenth Centuries*. University of Leeds PhD. https://etheses.whiterose.ac.uk/12687/. (19 October, 2023).

Triangulating language contact across families:
Linguistic and ethnomusicological data

Bernat Bardagil
(Ghent University)

Keywords: syntax, paleolinguistics, Amazonian languages, ethnomusicology

This talk discusses the first results of an ongoing research project that seeks to gain insight into the historical contact, mobility and settlement of the Guaporé-Mamoré area in southern Amazonia, by comparing three layers of high-resolution data: the morphosyntax of argument marking, the presence or absence of shared vocabulary, and the characteristics of flute-playing traditions.

The Guaporé-Mamoré area is a region of a very high linguistic diversity, with approximately 50 different languages belonging to at least 18 different genealogical units. These include larger South American language families (Arawakan, Tupian, Macro-Jê, Panoan), smaller families attested exclusively in this area (Chapacuran, Nambikwara, Tacanan), plus eleven isolated languages (Crevels 2002; Crevels & Van der Voort 2008). This area has been considered a "residual zone," defined as a region where linguistic diversity has accumulated as the languages spoken by politically and economically more powerful communities moved in, pushing aside the previous communities and their languages (Nichols 1992).

Given the shallow information depth that can be provided with more classical approaches, such as tracing loan words, alternative approaches are necessary. My current attempt to identify potential traces of contact patterns in the Guaporé-Mamoré area consists of a comparison of three different orally-transmitted systems: morphosyntax, lexicon, and instrumental music:

1. Differential argument marking. An identification of different patterns of argument marking (non-canonical subjects, DOM) in order to use them to inform us on the contact history of the peoples that inhabit a specific region.
2. Anti-Swadesh lexical list. A comparison of lexical items for highly specific semantic domains, such as types of creek, types of sand, specific medicinal plants.
3. Ethnomusicology. The distribution of rich flute playing traditions, the morphology of the instruments, the lexical items for instruments and musical styles.

This research seeks to explore the potential of tracing convergences across linguistic orphans, small language families and isolated languages, by adopting an interdisciplinary approach to combine grammatical, lexical and ethnomusicological comparative elements. Given the amount of languages spoken in the targeted area, the Juruena valley, on the eastern margin of the area, is used as a more focused and in-depth window for some aspects of this methodology.

References

Crevels, Mily & Hein van der Voort. 2008. The Guaporé-Mamoré region as a linguistic area. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 151–179. Amsterdam/Philadelphia: John Benjamins.

Crevels, Mily. 2002. *Itonama o Sihnipadara, Lengua no Clasificada de la Amazonía Boliviana*. Alicante: University of Alicante.

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

Using lexical and grammatical data to automatically affiliate language isolates and orphans

Frederic Blum
(Max-Planck Institute for Evolutionary Anthropology)

Keywords: Language classification; Computational methods; Lexibank; Comparative method; Language orphans

Language isolates and orphans defy the traditional techniques for historical language comparison as defined by the comparative method (Osthoff and Brugmann 1878). This method uses lexical and grammatical features to assign individual languages to one of the language families that have been proposed so far (Nichols 1996, Campbell 2017). It is usually carried out manually and compares the language to related languages, which, in the case of language orphans, are not available for comparison. To better understand the history of those orphan languages, we should thus aim to develop new methods that help us analyzing their linguistic past and establishing their linguistic affiliation. With standardized collections of linguistic data having become available on a large scale, we can design workflows for automatic language affiliation that complement the comparative method with a quantitative perspective.

We provide a computational model that affiliates languages to families by making use of large-scale lexical (List et al. 2022) and grammatical (Skirgård et al. 2023) databases, both featuring more than 1000 languages. Based on a neural network model, we train a supervised classifier for each of the datasets, as well as a model which combines lexical and grammatical data, and compare those results to previous classification attempts (Holman et al. 2008). We explicitly address the affiliation to small language families through our model architecture, which balances the model output with respect to the size of all language families present in the data. Overall, the models provide very good results for classification (>90%), with the lexical models providing better results than the models based on grammatical data. This method contributes directly to the study of orphan languages by providing a quantitative perspective on their affiliation. In some cases, the model creates plausible hypotheses for possible genealogical relationships that should be analyzed further. In other cases, the parallel analysis of lexical and grammatical data permits us to detect evidence for mixed structures present in languages, or confirms the absence of similar languages.

In the talk, we will discuss the affiliation of several language orphans from all over the world. We will discuss two case studies specifically. First, we will discuss the affiliation of around 20 language orphans from Northern Peru and surrounding areas. For this case study, we make use of a newly created lexical dataset (Barrientos et al. 2024). In our second case study, we target two language orphans from different parts of the world: Bangime and Basque. Our model compares each language with hundreds of other languages and classifies them either with an established language family or as isolate. The results provide us with information about lexical and grammatical affinity to other language families. Two different cases arise: Those languages which show a deep lexical or grammatical connection to other language families, pointing towards contact or shared (undetected) genealogical relations, and the case of true isolates, which show no affiliation to other established language families. We will also show in our talk

that contact phenomena seem to provoke more change in the grammatical than in the lexical domain.

References

Barrientos, C., Blum, F., List, J.-M., Peña, J., & Gonzales, R. (2024). NorthernPeru - A Lexical dataset of small language families and isolates from northern Peru.

Campbell, L. (2017). How to Show Languages are Related: Methods for Distant Genetic Relationship. In B. D. Joseph & R. D. Janda (Eds.), The Handbook of Historical Linguistics (pp. 262–282). Blackwell Publishing. https://doi.org/10.1002/9781405166201.ch4

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008). Explorations in automated language classification. Folia Linguistica, 42(3–4). https://doi.org/10.1515/flin.2008.331

List, J.-M., Forkel, R., Greenhill, S. J., Rzymski, C., Englisch, J., & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. Scientific Data, 9(1), 1–16. https://doi.org/10.1038/s41597-022-01432-0

Nichols, J. (1996). The comparative method as heuristic. In M. Durie & M. Ross (Eds.), The comparative method reviewed (pp. 39–71). Oxford University Press.

Osthoff, H., & Brugmann, K. (1878). Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen (Vol. 1). Hirzel. Osthoff, H., & Brugmann, K. (1878). Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen (Vol. 1). Hirzel.

Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latarche, J. J., Lesage, J., Weber, T., Witzlack-Makarevich, A., Passmore, S., Chira, A., Maurits, L., Dinnage, R., Dunn, M., Reesink, G., Singer, R., Bowern, C., Epps, P., Hill, J., et al. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. Science Advances, 9(16). https://doi.org/10.1126/sciadv.adg6175

Basque, Iberian and the comparative method

Mikel Martínez-Areta
(University of the Basque Country)

Keywords: Basque, Iberian, comparative method, language isolate, population genetics

Basque is a language isolate, with a textual history from the 16th century onwards. Today it is commonly assumed that the dialects stretching from Biscay to Zuberoa stem from an Old Common Basque (= OCB) spoken in the basin of Pamplona in about the 4th-6th centuries. In the Early Empire Period, its ancestor Aquitanian-Vasconic (= AV) is attested at least in Aquitania and modern Navarre in about 400 anthroponyms and 70 theonyms. These being the data, Basque historical linguistics has traditionally resorted to two methods: making inferences from the behaviour of borrowings from Latin and Romance, and internal reconstruction from OCB (itself a projection reached by means of a "light" comparative method among dialects) backwards.

The picture gets complicated by the fact that there is no consensus about whether AV (hence Basque) and Iberian are genetically related or may have even belonged to a single linguistic *continuum* at the arrival of the Romans. To the traditionally common features (many nominals are common to both, they resemble each other in both typology and phonotactics…), recently the question of numerals has been added. One would expect that the increasing data would allow some kind of application of the comparative method. And yet long Iberian texts simply do not help understand any structural problem of the pre-history of Basque grammar. Moreover, whereas internal reconstruction has led several independent Bascologists to postulate a VO stage of Basque at some point before AV (along with the features associated with it, such as NG, NRel, etc.), Iberian is certainly an OV language. This causes us a dilemma, as either the VO stage postulated for Basque is valid for their common ancestor, or they are not genetically related in spite of those resemblances.

Another element which in recent years has arisen as a potential tool is genetics. However, the criteria necessary for this to be reliable are not met in the case of Basque. In the Indo-European problem, a nearly certain correlation between languages and DNA (the IE languages and the haplogroup R1b) can be established because (a) the IE expansion entailed large-scale population movements, and (b) data come from sequencing the DNA of an increasing number of skeletal remains of the 3rd, 2nd and 1st millenia BC, which at least in Europe enables us to draw a line between populations previous (without R1b) and subsequent (with R1b) to that expansion. In the Basque-Iberian issue, the linguistic groups liable to belong to the same clade are geographically too close to each other, and, although the DNA of some skeletons of Iberian archaeological sites of the Late Iron Age has been sequenced with promising results (they all share an haplogroup R1b-DF27), on the Basque side no old AV skeleton has been found. Instead, the Basque data employed belong to 21th century population, which renders any alleged correlation suspicious.

The purpose of this presentation would be to put forward the main elements of the question and propose some clues which could cast light on the conundrum it poses.

References

Ferrer, Joan (2009), El sistema de numerales ibérico: avances en su conocimiento, *Palaeohispanica* 9, 451-479.

Orduña, Eduardo (2005, "Sobre algunos posibles numerales en textos ibéricos". *Palaeohispanica* 5, 491-506.

Reich, David (2018), *Who we are and how we got here. Ancient DNA and the new science of the human past*, New York: Vintage Books.

Trask, Robert Lawrence (1997), *The History of Basque*, New York: Routledge.

Velaza, Javier (2019), Iberian writing and language, in: A. G. Sinner & J. Velaza (eds), (2019), *Palaeohispanic languages & epigraphies*, Oxford University Press, 160-197.

Linguistic endemism

Rik van Gijn & Sietze Norder
(Leiden University & Utrecht University)

Understanding the uneven distribution of linguistic diversity over the globe is a central issue in linguistics, and the subject of a growing body of studies (see Gavin et al. 2013 for an overview). Most studies measure diversity in terms of richness: the number of linguistic units (languages, language families) in a given geographical area. Richness has proven to be an insightful measure, having given rise to generalizations about ecological contexts and circumstances that facilitate the development and maintenance of linguistic diversity (see Greenhill 2014 for an overview).

Richness does not take into account, however, to what extent the languages or language families in a high-richness area are *exclusive* to that area. This seems particularly important when it comes to genealogical diversity: richness equates areas with languages belonging to wide-spread families with areas with a high number of isolates and small language families, whereas the histories of these areas may be quite different. Moreover, richness-based metrics overlook areas with a relatively low number of languages, but a high proportion of narrow-ranged language families, which can be very informative for understanding which areas have given rise to language expansions and which have not.

Therefore, we introduce a new approach to quantifying patterns of diversity, linguistic endemism. This measure quantifies an overlooked aspect of linguistic diversity: the degree of *uniqueness* of a given geographical unit in terms of its languages, language families or linguistic structures. We focus on genealogical endemism, where resulting patterns indicate those areas that are home to a disproportionate share of the global genealogical diversity, i.e. areas with a high degree of small language families and isolates.

Based on both richness and endemism, we propose a typology of diversity zones, which identifies three types of diversity zones (Fig. 1):

1. endemism enclaves: low in richness, high in endemism (e.g. Tasmania, the Sudan-Ethiopia-Eritrea border area, the Southern Cone of South America);
2. accretion zones (terminology from Nichols 1997): high in richness, high in endemism. (New Guinea, lowland Bolivia, the US west coast)
3. Invasion zones (terminology from Kaufman 1990): high in richness, low in endemism (Upper Xingu in Brazil and south-west Africa).
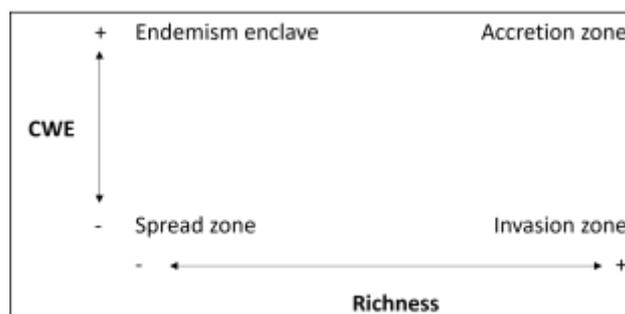

Figure 1

A next step is to combine the proposed typology of genealogical diversity zones with the distribution of linguistic features. For high-richness zones, (Nichols 1992, 248–49) observes that these zones tend to behave as microcosms, reflecting patterns also found on a global level. She hypothesizes that this "standard profile" reflects the social situations in high-density contact situations, where multilingualism is the norm. We hypothesize that, for zones with a high endemism, social isolation may have played a bigger role, and that we therefore would expect more "rare" features in these areas, leading to predictions in Fig. 2.
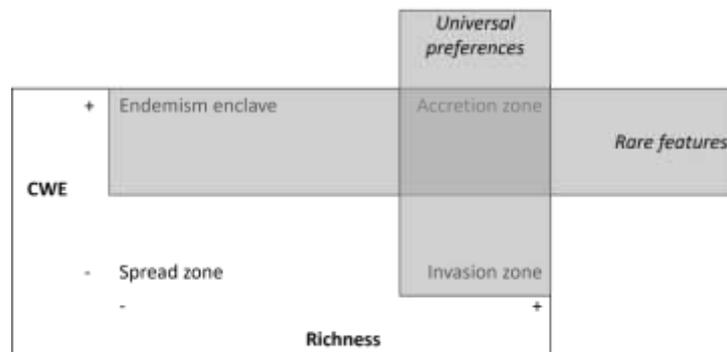

Figure 2

We test these predictions with the presence or absence of rare phonemes, based on Phoible (Moran and McCloy 2019), and zoom in on exemplars of the three diversity zones in South America: the southern cone (endemism enclave), lowland Bolivia (accretion zone), and the Upper Xingu (invasion zone).

References

Comrie, Bernard. 2016. "Measuring Language Typicality, with Special Reference to the Americas." In *Studies in Language Companion Series*, edited by Andrea L. Berez-Kroeker, Diane M. Hintz, and Carmen Jany, 173:363–84. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/slcs.173.16com.

Cysouw, Michael. 2011. "Quantitative Explorations of the Worldwide Distribution of Rare Characteristics, or: The Exceptionality of Northwestern European Languages." In *Expecting the Unexpected: Exceptions in Grammar*, edited by Horst J. Simon and Heike Wiese, 411–32. DE GRUYTER MOUTON. https://doi.org/10.1515/9783110219098.411.

Gavin, Michael C., Carlos A. Botero, Claire Bowern, Robert K. Colwell, Michael Dunn, Robert R. Dunn, Russell D. Gray, et al. 2013. "Toward a Mechanistic Understanding of Linguistic Diversity." *BioScience* 63 (7): 524–35. https://doi.org/10.1525/bio.2013.63.7.6.

Greenhill, Simon. 2014. "Demographic Correlates of Language Diversity." In *The Routledge Handbook of Historical Linguistics*, edited by Claire Bowern and Bethwyn Evans. London: Routledge.

Kaufman, Terrence. 1990. "Language History in South America: What We Know and How to Know More." In *Amazonian Linguistics : Studies in Lowland South American Languages*, edited by Doris L. Payne, 13–74. Berlín: Mouton de Gruyter.

Moran, Steven, and Daniel McCloy, eds. 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History. http://phoible.org.

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

Nichols, Johanna. 1997. "Modeling Ancient Population Structures and Movement in Linguistics." *Annual Review of Anthropology* 26: 359–84.

Genetic affiliation of Bangime: A view from clausal architecture

Vadim Diachkov

(Institute of Linguistics, Russian Academy of Sciences)

In my talk, I deal with the possible genetic affiliation of the West African isolate Bangime, bringing syntactic evidence to light. The data come from my own fieldwork as well as from a grammatical description of the language (Heath & Hantgan 2018).

Comparing Bangime to other West African languages is justified by its striking lexical similarities to neighboring languages. Dogon loanwords constitute a significant part of the vocabulary of Bangime (Hantgan & List 2022). However, Bangime also exhibits unusual lexical affinities with Gur languages such as Natioro[1], spoken in southwest Burkina Faso, whose speakers do not have (and presumably never had) contacts with those of Bangime. Furthermore, there are certain grammatical features of Bangime that can be traced to a Gur source yet not through contact. Although Gur languages exhibit a high degree of diversity, I claim that the general organization of the clausal architecture among Gur languages – which differs from that found among Dogon languages – suggests that Bangime is unlikely to be a Dogon language yet it shares these features with the Gur group. This claim is based on the following basic principles of clausal architecture:

1. In many Dogon languages, TAM suffixes can be easily traced back to lexical verbs occupying the second position in V1 + V2 sequences, which is typical of SOV languages. In contrast, Gur languages frequently have systems of the type S-V1-O-V2 where V1 is grammaticalized. In this respect, Bangime patterns with Gur, Mande and Kwa groups but not with Dogon languages. Another Dogon strategy is to grammaticalize preverbal adverbials, see, e. g. (Heath 2017), but this is also absent from Bangime. At the same time, perfective morphology in Bangime resembles that of many Gur languages, indicating a possible common source.
2. The Bangime negation markers resemble those reconstructed for Gur (and Kwa) languages, cf. (Winkelmann & Miehe 2009), and occupy the same slot in the clausal structure.
3. The Bangime definite marker resembles those found in Gur languages (including those that are unlikely to have had contacts with Bangime). In my talk, I propose that the former can go back to pronouns that were grammaticalized in Gur languages both as class markers and as determiners.
4. Unlike Dogon languages, Bangime uses poly-predicative structures where the dependent clause follows the main clause headed by subjunctive-like forms. The same strategy is widely attested among Gur languages. This feature is due to right branching but can be an additional argument in favor of the common source hypothesis.

In my presentation, I will consider the probability of each scenario of syntactic developments for Bangime, concluding that the scenario whereby Bangime and Dogon languages go back to a common origin much less probable than a scenario whereby Bangime inherits basic syntactic features from a Proto-Gur/Bangime source.

---

[1] The Natioro data discussed in my talk come from my own fieldwork data as well.

References

Hantgan, Abie and List, Johann Mattis (2022). Bangime: secret language, language isolate, or language island? A computer-assisted case study. *Papers in Historical Phonology* (7) 1-43. https://doi.org/10.2218/pihph.7.2022.7328.

Heath, Jeffrey (2017). A grammar of Bunoge (Dogon, Mali). doi:10.17617/2.2417511

Heath, Jeffrey and Hantgan, Abbie (2018). *A Grammar of Bangime*, Berlin, Boston: De Gruyter Mouton.

Winkelmann, Kerstin and Miehe, Gudrun (2019). Negation in Gur: Genetic, areal and unique features. In: Norbert Cyffer, Erwin Ebermann and Georg Ziegelmeyer (eds.). *Negation Patterns in West African Languages and Beyond* [Typological Studies in Language 87]. Amsterdam: John Benjamins, 2009. Pp. 167–204.

WS23 The semantic transparency of morphologically complex words

The role of semantic transparency in morphological change:
Evidence from doublet formation in Palestinian Arabic

Asma Taha & Lior Laks

(Bar-Ilan University & Bar-Ilan University)

This study examines doublet formation of instrument nouns (INs) in Palestinian Arabic (PA), as demonstrated in the online examples below.

(1) šu: siʕr il-**midbase** ?

'how much is the **stapler**?'

(Facebookتم بحمد... - المعدات الصناعية |

🔥 🔥 🔥 الاكثرررر طلباااا 🔥 🔥)

(2) biddi **dabba:se** dabbu:s seħri.

'I need a **stapler** with the magical staple'

(https://m.facebook.com/102974454820106/photos/a.108114064306145/114595163658035/?type=3&_rdr)

Both INs in (1)-(2) denote 'stapler', share the consonantal root *d-b-s*, and are formed in two patterns, *miCCaCe* (*midbase*) (1) and *CaCCa:Ce* (*dabba:se*) (2). The study examined the distribution of such doublets and accounts the criteria that trigger their formation and lack thereof. Data collection is based on three sources: dictionary search (Elihay 2005), online web searches and an experiment ,where 30 participants were asked to name pictures of INs.

The results show that INs that are formed in the *miCCaCe* pattern, receive an additional form in the *CaCCa:Ce* pattern, which has become more productive and frequent. Why are such doublets formed? We show that doublet formation can be predicted based on both morphological and semantic criteria. Doublet formation triggers both structural and semantic transparency between INs and verbs, from which they are assumed to be derived. In this talk we will focus on the role of semantic transparency. INs with doublets are related to verbs, e.g. *midbase/dabba:se* 'stapler' - *dabbas* 'stapled'. INs that are derived from verbs correspond to the argument structure of the verbs and their thematic roles (Grimshaw 1990, Rappaport-Hovav & Levin 1992, Alexiadou & Schafer 2008, Schafer 2008, among others). The IN has to be agentive in order to be semantically transparent and undergo a morphological change. In cases like *dabbas - midbase/dabba:se* the semantic relation is highly transparent, as the IN corresponds to the action that the verb denotes and can function as its subject (*il- midbase/dabba:se dabbasat* 'the stapler stapled'). Similarly, the IN *miqšare* 'peeler' has the *qašša:re* doublet, as it is semantically related to the verb *qaššar* 'peeled'. Semantic transparency triggers doublet formations, so that the structural relations are also transparent. The formation of *CaCCa:Ce* INs based on verbs requires less changes in comparison to the formation of *miCCaCe* INs.

In case the semantic relation is not transparent, doublet formation is blocked. For example, *milʕaqe* 'spoon' has no *CaCCa:Ce* doublet like **laʕʕa:qe*. The semantic relation between *milʕaqe* and *laʕʕaq* 'licked' is not transparent, as the IN does not perform the action of the verb (**il-milʕaqe laʕʕqat '*the spoon licked'), but can only facilitate this action. It is important to note that the proposed analysis reflects only tendencies. There are INs with high semantic transparency between them and the related verbs, that have no doublets for no apparent reason. However, doublet formation is systematic in the sense that it only happens when there is a transparent semantic relation between the IN and the verbal counterpart, and not in other cases.

The study sheds light on the relations between the verbal and nominal systems and the important role of semantic transparency in morphological change.

References

Alexiadou, Artemis & Florian Schäfer. Instrumental -er Nominals Revisited. In K. Ryan (ed.), *Online-Proceedings of WCCFL XXVII* (Poster Session), 10-19. UCLA Working Papers in Linguistics.

Elihay, Yohanan. 2005. *The Olive Tree Dictionary*. Jerusalem: Minerva Instruction and Consultation.

Grimshaw, Jane. 1990. *Argument Structure*. MIT Press.

Rappaport Hovav, Malka & Beth Levin. 1992. *–er* nominals: Implications for the theory of argument structure. *Syntax and Semantics* 26, 127-153.

Schäfer, Florian. 2008. Event denoting -er nominals in German. In F. Schäfer (ed.) *SinSpec 1 (Working Papers of the SFB 732)*, 173-187 (http://elib.unistuttgart.de/opus/volltexte/2008/3554/).

The semantic transparency of newly-coined compounds

Fritz Günther
(Humboldt-Universität zu Berlin)

Based on pragmatic principles, one would expect that new compounds (such as *airzoom* or *lowcoin*) are created transparent: Unlike existing compounds with lexicalized (transparent or opaque) meanings that are likely known by most recipients, a speaker cannot expect that a novel compound they create is known by others. To achieve communicative success, the meanings of such novel compounds should be clearly related to and predictable from their constituents (and hence semantically transparent). The present study tests this expectation empirically.

Rather than examining compounds newly appearing in a representative language corpus at a given point in time, we investigate newly-coined compounds that are produced "on the fly" by speakers in a controlled experimental setting. These words were obtained from a study employing the taboo game paradigm (Pugacheva & Günther, 2024). Here, speakers are presented with a target word (e.g., "dragon") and have to produce a single word that expresses the meaning of the target without using this word, so that another speaker would later be able to identify the original target (e.g., "serpent" or "firelizard"). Participants were allowed (Experiment 1) or explicitly instructed to (Experiment 2) produce novel compounds in order to convey the intended meaning (and produced a total of 2,127 and 3,493 novel compounds, respectively).

If we assume that compounds are "born" transparent and can become more semantically opaque over time as a consequence of lexicalization and meaning shift, we should expect these newly-coined compounds to be more semantically transparent than a representative sample of 4,638 existing compounds obtained from a corpus (which also includes opaque compounds). To measure the semantic transparency of (novel) compounds in a comparable manner, we used compositional distributional models to obtain a compositional meaning representation for both existing and novel compounds (Günther & Marelli, 2019, 2020). In distributional semantic models, each word meaning is represented as a high-dimensional vector that captures the word's distribution in a large corpus of natural language. Since by definition novel compounds do not occur in such corpora, their meanings are instead estimated compositionally, usually as a function of their constituent meanings (here, a linear combination; Marelli, Gagné & Spalding, 2017). This allows us to calculate the semantic transparency of novel and existing compounds as the similarity between this compositional meaning and either the modifier or the head (Günther & Marelli, 2019, 2020). Using these metrics, we however observed the opposite of the expected pattern: The newly-coined compounds are significantly *less* transparent both in their modifier and their head constituent than the existing compounds.

This difference is smaller and ultimately disappears for novel compounds that are more similar (according to our distributional models) to their original target (i.e., which express the intended meaning better). It is also smaller for novel compounds for *other* participants are more likely to guess the original target (i.e., that better allow others to recover the intended meaning). However, the

differences never reverse in the expected direction. Potential methodological and theoretical explanations and implications of these results will be discussed.

References

Günther, Fritz, & Marco Marelli (2019). Enter sand-man: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1872–1882.

Günther, Fritz, & Marco Marelli (2020). Trying to make it work: Compositional effects in the processing of compound "nonwords". *Quarterly Journal of Experimental Psychology, 73,* 1082-1091.

Marelli, Marco, Christina Gagné, & Thomas Spalding (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition, 166,* 207-224.

Pugacheva, Vasilisa, and Fritz Günther (2024). Lexical choice and word formation in a taboo game paradigm. *Journal of Memory and Language, 135,* 104477.

Polyfunctionality and Semantic Transparency in English

Martha Booker Johnson, Micha Elsner & Andrea D. Sims
(The Ohio State University, The Ohio State University & The Ohio State University)

At one level semantic transparency is intuitive: *busyness* is semantically composed of *busy* and *-ness* whereas *business* (etymologically *busy* and *-ness*) is not. However, polyfunctionality of roots/affixes complicates base-derivative relations (Lapesa et al. 2018, Salvadori & Huyghe 2023) and it is not clear how polyfunctionality affects speaker judgments (e.g. Hay 2001, Günther et al. 2020) or quantitative distributional semantic measures (e.g. Günther et al. 2020, Varvara et al. 2021) of semantic transparency. We present an empirical study of English derivation showing that polyfunctionality affects judgment data and distributional semantic measures differently, reflecting different (implicit) understandings of semantic transparency.

We present two analyses. In both we implement semantic transparency as cosine similarity of a derived word's vector and its predicted vector; the latter is computed from base and affix vectors. The data come from pre-trained, 300-dimensional vectors from Fares et al.'s (2017) English model, trained on English Wikipedia (February 2017).

The first analysis compares cosine similarity to human judgments (N=24) of 109 English base-derivative pairs from McKenzie (2019). As shown in Figure 1, when cosine similarity is high, English speakers also rate the derivative as highly semantically similar to its base. However, when cosine similarity is low, speaker ratings are more variable. This mismatch is an effect of task and polyfunctionality. Figure 2 shows that high ratings primarily correspond to situations in which both a derivative and its base have a low number of senses; derivative and base polyfunctionality both tend to decrease judgments. This suggests that speakers assigned high ratings if a transparent base-derived relationship could be located, but that either derivative or base polyfunctionality led to variable success in this endeavor. In contrast, cosine similarity inherently measures *all* meanings of bases/derivatives in aggregate.

The second analysis seeks to further pin down the nature of the mismatch between human judgments and cosine measures by examining the token distribution of different senses of derivatives and bases. Tokens of the same 109 bases and derivatives were drawn from the training set of Tensorflow Wiki40b dataset (Guo et al. 2020). We use hierarchical clustering over the words' token vectors to automatically identify senses and then match base meanings and derivative meanings by hand. We ask how much of the variation in Figure 1 can be accounted for by the *relative frequency* with which different senses of bases/derivatives are used. We expect to find that polyfunctionality plays a role in the mismatch primarily when the most frequent derivative meaning is not transparent based on the most frequent base meaning, even if a transparently related base meaning is available.

Showing that polyfunctionality affects human and cosine-based measures differently, this study points to the value of a multimethod approach to measuring semantic transparency.

Figure 1. Comparison of vector-based cosine similarity measure (x-axis) to human judgments of base-derivative relationship (y-axis, normalized as z-scores with mean z value shown).

Figure 2. Model-predicted response values for human judgments. Number of derivative senses (color series) and number of base senses (panels) show the centered mean values and values one standard deviation above and below the mean.

References

Fares, Murhaf, Andrey Kutuzov, Stephan Oepen, and Erik Velldal (2017), Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In J. Tiedemann (ed.), *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, 271–276. Association for Computational Linguistics.

Günther, Fritz, Marco Marelli, and Jens Bölte (2020), Semantic transparency effects in German compounds: A large dataset and multiple-task investigation. *Behavior Research Methods 52*(3), 1208–1224.

Guo, Mandy, Zihang Dai, Denny Vrandečić, and Rami Al-Rfou (2020), Wiki-40B: Multilingual language model dataset. In N. Calzolari, et al. (eds.), *Proceedings of LREC 2020: 12th International Conference on Language Resources and Evaluation*, 2440–2452. European Language Resources Association.

Hay, Jennifer (2001), Lexical frequency in morphology: Is everything relative? *Linguistics 39*(6): 1041–1070.

Lapesa, Gabriella, Lea Kawaletz, Ingo Plag, Marios Andreou, Max Kisselew, & Sebastian Padó (2018), Disambiguation of newly derived nominalizations in context: A Distributional Semantics approach. *Word Structure 11*(3), 277–312.

McKenzie, Michelle (2019), *Effects of relative frequency on morphological processing in Russian and English*. B.A. Thesis, The Ohio State University.

Salvadori, Justine and Richard Huyghe (2023), Affix polyfunctionality in French deverbal nominalisations. *Morphology* 33, 1-39.

Varvara, Rossella, Gabriella Lapesa, & Sebastian Padó (2021), Grounding semantic transparency in context: A distributional semantic study on German event nominalizations. *Morphology 31*, 409–446.

Conceptualizing semantic transparency within an associative account of morphological complexity

Nikolaos Ntagkas
(Aristotle University of Thessaloniki)

Research question: Considering that the meaning of a morphologically complex word comprises a more formal (semantic-grammatical) part and a less formal (semantic-pragmatic) part (Corbin 1987/1991, Dressler 2005, and Lieber 2004), how could we deal with the notion of semantic transparency?

Approach: In the literature, the semantic transparency of morphologically complex words has been discussed in many different terms (e.g., compositionality, predictability, relatedness, overlap, similarity, literality, etc.; see, Auch, Gagné & Spalding 2020, Bourque 2014, and Schäfer 2018), usually irrespective of an integrated theory about the relationship between morphology and meaning. Hence, in this study we approach the notion of semantic transparency within the framework of associative morphology (Corbin 1987/1991, 1989, 1991, and Anastassiadis-Symeonidis 1992, 1998, 2016, 2023), focusing on the distinction among the attested, the predictable, and the conventional meaning of a morphologically complex word. We, namely, treat semantic transparency as the distance (convergence/divergence) between the attested and the predictable meaning of a morphologically complex word, mediated by its conventional meaning (i.e., the part of meaning grounded in pragmatic knowledge that helps predictable meaning adapt to extra-linguistic reality; Anastassiadis-Symeonidis 1992, 1998).

Method: Our proposal was experimentally tested in the research field of morphological processing (Crepaldi 2023) and the linguistic domain of derivational morphology (Lieber & Štekauer 2014), by examining suffixed derivatives (i.e., denominal verbs and adjectives) in Modern Greek (Anastassiadis-Symeonidis 2023, Efthymiou 2023, and Ralli 2022). Two hundred and forty typical adult native speakers participated in three metalinguistic judgment tasks, rating the degree of semantic transparency for 566 derivative–base pairs by means of a seven-point Likert-type scale (Gonnerman, Seidenberg & Andersen 2007, and Xu & Taft 2015). Furthermore, we selected 90 critical derivative–base pairs and conducted two unmasked priming lexical decision tasks with 96 other participants from the same population, employing the overt visual priming paradigm (Feldman et al. 2004, and Smolka, Preller & Eulitz 2014) and the immediate cross-modal priming paradigm (Gonnerman, Seidenberg & Andersen 2007, and Marslen-Wilson et al. 1994).

Data: Metalinguistic ratings captured the gradation between the meaning of the derivative and the meaning of its noun-base, featuring three ordinal categories of semantic transparency: opaque derivatives, e.g., /xoˈnevo/ – /xoˈni/ 'to digest' – 'funnel'; semitransparent derivatives, e.g., /xalˈkevo/ – /xalˈkos/ 'to forge' – 'copper'; transparent derivatives, e.g., /stoˈxevo/ – /ˈstoxos/ 'to aim' – 'target'. On the other hand, reaction times revealed statistically significant and equivalent priming between the derivative (prime) and its noun-base (target) in the transparent and semitransparent category, yet null priming in the opaque category.

Results: The data indicate that offline semantic transparency ratings (opaque < semitransparent < transparent) are mainly based on the attested meaning of the derivatives, while online morphological priming (opaque < semitransparent = transparent) is crucially driven by the predictable meaning of the derivatives. Our findings underpin the theoretical distinction among the attested, the predictable, and the conventional meaning of a morphologically complex word, as well as the proposed conceptualization of semantic transparency, suggesting that the gradation of semantic transparency results from the relative contribution of the predictable meaning and the conventional meaning to the attested meaning (see, Figure 1).

– Semantic transparency   **+ Semantic transparency**

◄──►

Opaque   **Semitransparent**   **Transparent**
complex words   complex words   complex words

[– Predictable meaning]   [+ Predictable meaning]   [+ Predictable meaning]

[+ Conventional meaning]   [+ Conventional meaning]   [– Conventional meaning]

Figure 1. The semantic transparency continuum of morphologically complex words

References

Anastassiadis-Symeonidis, Anna (1992), I neoelliniki paragogi kata to montelo tis D. Corbin [Modern Greek derivation within D. Corbin's model], *Studies in Greek Linguistics* 13, 505–526.

Anastassiadis-Symeonidis, Anna (1998), On Modern Greek denominal adjectives, in G. E. Booij, A. Ralli, and S. Scalise (eds.), (1998), *Proceedings of the 1st Mediterranean Morphology Meeting (MMM 1): Allomorphy, Compounding, Inflection*, Patras: University of Patras, 29–40. https://doi.org/10.26220/mmm.2341

Anastassiadis-Symeonidis, Anna (2016), Suffixation and the expression of time and space in Modern Greek, *Lingue e Linguaggio* 15(1), 59–84. https://www.rivisteweb.it/doi/10.1418/83654

Anastassiadis-Symeonidis, Anna (2023), Arches morfologikis analysis me efarmogi stin epithimatopoiisi [Principles of morphological analysis applying to suffixation], *Studies in Greek Linguistics* 42, 13–19. http://ins.web.auth.gr/images/MEG_PLIRI/MEG_42_13_19.pdf

Auch, Leah, Christina L. Gagné, and Thomas L. Spalding (2020), Conceptualizing semantic transparency: A systematic analysis of semantic transparency measures in English compound words, *Methods in Psychology* 3, Article 100030. https://doi.org/10.1016/j.metip.2020.100030

Bourque, Yves Stephen (2014), *Toward a Typology of Semantic Transparency: The Case of French Compounds*, doctoral dissertation, University of Toronto. https://hdl.handle.net/1807/68190

Corbin, Danielle (1987/1991), *Morphologie dérivationnelle et structuration du lexique*, Tübingen / Villeneuve d'Ascq: Max Niemeyer Verlag / Presses Universitaires de Lille. https://books.openedition.org/septentrion/124160

Corbin, Danielle (1989), Form, structure and meaning of constructed words in an associative and stratified lexical component, in G. E. Booij, and J. van Marle (eds.), (1989), *Yearbook of Morphology*, vol.2, Dordrecht: Foris, 31–54. https://doi.org/10.1515/9783112420560-004

Corbin, Danielle (1991), Introduction. La formation des mots: Structures et interpretations, *Lexique* 10, 7–30.

Crepaldi, Davide (ed.), (2023), *Linguistic Morphology in the Mind and Brain*. London: Routledge. https://doi.org/10.4324/9781003159759

Dressler, Wolfgang U. (2005), Word-Formation in natural morphology, in P. Štekauer, and R. Lieber (eds.), (2005), *Handbook of Word-Formation*, Dordrecht: Springer, 267–284. https://doi.org/10.1007/1-4020-3596-9_11

Efthymiou, Angeliki (2023), Paragogi [Derivation], in D. Papadopoulou, and A. Revithiadou (eds.), (2023), *Eisagogi sti morfologia: Theoria kai peiramatikes efarmoges* [Introduction to morphology: Theory and experimental applications], Thessaloniki: Institute of Modern Greek Studies, 199–234.

Feldman, Laurie Beth, Emily G. Soltano, Matthew J. Pastizzo, and Sarah E. Francis (2004), What do graded effects of semantic transparency reveal about morphological processing?, *Brain and Language* 90(1-3), 17–30. https://doi.org/10.1016/S0093-934X(03)00416-4

Gonnerman, Laura M., Mark S. Seidenberg, and Elaine S. Andersen (2007), Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology, *Journal of Experimental Psychology: General* 136(2), 323–345. https://psycnet.apa.org/doi/10.1037/0096-3445.136.2.323

Lieber, Rochelle (2004), *Morphology and Lexical Semantics*, Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486296

Lieber, Rochelle, and Pavol Štekauer (eds.), (2014), *The Oxford Handbook of Derivational Morphology*, Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641642.001.0001

Marslen-Wilson, William D., Lorraine K. Tyler, Rachelle Waksler, and Lianne Older (1994), Morphology and meaning in the English mental lexicon, *Psychological Review* 101(1), 3–33. https://psycnet.apa.org/doi/10.1037/0033-295X.101.1.3

Ralli, Angela (2022), *Morfologia: Nea, anatheorimeni ekdosi* [Morphology: New, revised edition], Athens: Patakis.

Schäfer, Martin (2018), *The Semantic Transparency of English Compound Nouns*, Berlin: Language Science Press. https://zenodo.org/doi/10.5281/zenodo.1134594

Smolka, Eva, Katrin H. Preller, and Carsten Eulitz (2014), '*Verstehen*' ('understand') primes '*stehen*' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs, *Journal of Memory and Language* 72, 16–36. https://doi.org/10.1016/j.jml.2013.12.002

Xu, Joe, and Marcus Taft (2015), The effects of semantic transparency and base frequency on the recognition of English complex words, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(3), 904–910. https://psycnet.apa.org/doi/10.1037/xlm0000052

Compounds and pseudo-compounds as primes in masked LDT: Do lexical status and degree of transparency matter?

Linda Sandström & Maria Rosenberg
(Umeå University & Umeå University)

Keywords: morphological decomposition, pseudo stems, priming effects, semantic transparency, masked lexical decision task

Introduction: Masked Lexical Decision Task (LDT) experiments have provided strong evidence that experienced readers automatically decompose monomorphemic words as long as the decomposed parts conform to existing morphemes (i.e., a pseudo-complex word, like *corner*, is decomposed into a pseudo-stem *corn* and a pseudo-affix *-er*). Whether the same type of "garden path"-decomposition is evident for pseudo-compounds is less clear (Auch et al. 2023).

Pseudo-compounds are here defined as monomorphemic words with a surface compound form (e.g., *carpet*). Decomposition of a pseudo-compound would thus result in two word stems (*car* and *pet*), each with its own lexical representations with no semantic relation to the monomorphemic pseudo-compound (*carpet*).

In this study, we used the unique characteristics of pseudo-compounds to further explore whether morphological decomposition is automatic in early word processing, as has been hypothesized in models of word recognition (e.g., Beyersmann & Grainger 2023). If so, we expect to find comparable priming effects from pseudo-compound and compound primes on (pseudo-)constituent targets. On the other hand, if semantic relations between primes and targets affect early processing, as has been argued by others (e.g., Chee & Yap 2022), we expect graded priming effects due to semantic transparency between the prime and the target (i.e., transparent compound priming > opaque compound priming > pseudo-compound priming).

Research questions: To what extent do pseudo-compounds prime their pseudo-constituents?
To what extent do degrees of transparency affect the early stages of compound processing?

Method: We conducted online masked LDT experiments (Angele et al. 2023) to investigate how experienced readers of Swedish process pseudo-compounds and compared priming effects of pseudo-compounds (*carpet*) to that of true compounds that vary in transparency (e.g., *newspaper*, *mushroom*), and non-compounds (e.g., *sandwich*). Importantly, whereas previous investigations (e.g., Gagné et al. 2018, and Melvie et al. 2023) have included some pseudo-compound primes that could be considered morphologically ambiguous, like "crowding" (decomposable into either two stems *crow + ding*, or stem + affix *crowd + ing*), we included only Swedish pseudo-compounds with no morphological ambiguity to avoid confounding the result. Furthermore, we have controlled that the pseudo-compound primes and their pseudo-stem targets do not share meaning and that left-over letter strings in non-compound primes do not function as real suffixes/stems.

Compound transparency was measured from human ratings and categorized according to Libben et al. (2003).

Data analyses: Response times and accuracy rates will be analyzed as dependent measures using linear mixed-effects models. The following variables will be considered: condition (compounds vs. pseudo-compounds vs. non-compounds), prime type (related vs. unrelated), semantic transparency, prime/target frequency, prime/target length, and target morphological family size.

Interpretation of results: Results showing priming effects of comparable magnitude between conditions could be considered supportive of form-then-meaning accounts, in which the initial stages of word processing are assumed to be "semantically blind". Conversely, results showing varying priming effects may indicate semantical influence during the early stages of word processing.

In conclusion, this study contributes evidence of value to word recognition models of how the lexical status of words affects early word recognition and processing.

References

Angele, Bernhard, Ana Baciero, Pablo Gómez and Manuel Perea (2023), Does online masked priming pass the test? The effects of prime exposure duration on masked identity priming, *Behavior research methods* 55(1), 151–167.

Auch, Leah, Karen Pérez Cruz, Christina L. Gagné and Thomas L. Spalding (2023), LaDEP: A large database of English pseudo-compounds, *Behavior Research Methods* 1–17.

Beyersmann, Elisabeth and Jonathan Grainger (2023), The role of embedded words and morphemes in reading, in D. Crepaldi (ed), (2023), *Linguistic morphology in the mind and brain*, Abingdon, Oxon: Routledge, 26–49.

Chee, Qian Wen and Melvin J. Yap (2022), Are there task-specific effects in morphological processing? Examining semantic transparency effects in semantic categorisation and lexical decision, *Quarterly Journal of Experimental Psychology* 75(11), 2073–2086.

Gagné, Christina L., Thomas L. Spalding, Kelly A. Nisbet and Caitrin Armstrong (2018), Pseudo-morphemic structure inhibits, but morphemic structure facilitates, processing of a repeated free morpheme, *Language, Cognition and Neuroscience* 33(10), 1252–1274.

Libben, Gary, Martha Gibson, Yeo Bom Yoon and Dominiek Sandra (2003), Compound fracture: The role of semantic transparency and morphological headedness, *Brain and Language* 84, 50–64.

Melvie, Taylor, Alexander Taikh, Christina L. Gagné and Thomas L. Spalding (2023), Constituent processing in compound and pseudocompound words, *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 77(2), 98–114.

Exploring the relationship between word embeddings and human classification systems for interpretation patterns of German noun-noun compounds

Annelen Brunner & Anna Pohle
(Leibniz Institute for the German Language & Leibniz Institute for the German Language)

One facet of interpretability is recognizing the relationship between the components of a compound in order to derive its meaning. Scientific endeavours to capture these relations have yielded diverse classification systems ranging from lists predominantly reliant on the semantic roles of the constituents (e.g. Downing 1977, Ortner et al. 1991) to approaches rooted in underlying essential verbs and arguments (e.g. Ó'Séaghdha 2008, Levin et al. 2019), some of those including paraphrasal prepositions (eg. Girju 2005, Telljohann et al. 2017/ Dima 2019). In a parallel discourse within the domain of distributional semantics, it has been suggested that vector-based representations may be able to also capture relational information (Marelli et al. 2017, Günther & Marelli 2022).

Our study aims to explore the relationship between human-made classification systems and the information captured from context by word embeddings. We prepared a dataset of 400 German noun-noun compounds with competing annotations by two different relational classification systems both of which have already been applied to large sets of words. The first, by Ortner et al. (1991), endeavours a comprehensive mapping of semantic relations in German nominal compounds relying heavily on semantic roles. The system by Telljohann et al. (2017) as applied by Dima (2019), narrows the focus to concrete nouns, but additionally includes paraphrasal prepositions in its annotations. As much as possible, we used category assignments attested by the original authors.

Our dataset contains only unique heads and modifiers to minimize lexical bias. The 400 compounds were encoded as regular as well as compositional word vectors according to the method described by Marelli et al. (2017). We tested two different semantic spaces for the word embeddings, one based on the deWaC corpus as used by Günther et al. (2020) and one based on the German reference corpus (DeReKoVecs, c.f. Fankhauser & Kupietz 2019, 2022), in order to see how much the underlying corpus affects the results. Using clustering techniques and other similarity measures, we tested whether groups of relational patterns are detectable based on the vectors. For a first estimation, we clustered the vectors based on cosine similarity and measured the decrease in entropy with respect to category labels. An above chance decrease was measurable and we found that generally – but not reliably – it was stronger for compositional vectors. For a close-up examination, we determined the most similar partners for each compound and calculated how many of those were of the same category as the compound itself, thus measuring the cohesion of each category within our dataset. We found that in both category systems, some relational patterns gave much better results than others. The workshop presentation will include discussion how the two annotation systems relate to each other, which factors determine which human-made categories are better aligned with vector similarities and what this tells us about the capabilities of word-embeddings – especially compositional embeddings – to capture relational information.

References

Dima, Corina (2019), *Composition models for the representation and semantic interpretation of nominal compounds* (Doctoral Dissertation), Universität Tübingen. https://doi.org/10.15496/publikation-28485

Downing, Pamela (1977), On the creation and use of English compounds. *Language*, *53*(4), 810–842.

Fankhauser, Peter and Kupietz, Marc (2022), Count-Based and Predictive Language Models for Exploring DeReKo, in P. Banski, A. Barbaresi, S. Clematide, M. Kupietz, and H. Lüngen (eds), (2009), *Proceedings of the LREC 2022 Workshop on Challenges in the Management of Large Corpora (CMLC-10 2022),* Paris/Marseille: ELRA, 27-31.

Fankhauser, Peter and Kupietz, Marc (2019), Analyzing domain specific word embeddings for a large corpus of contemporary German, International Corpus Linguistics Conference, Cardiff, Wales, UK, July 22-26, 2019.

Girju, Roxana; Moldovan, Dan; Tatu, Marta and Antohe, Daniel (2005), On the semantics of noun compounds. *Computer Speech & Language*, *19*(4), 479–496. https://doi.org/10.1016/j.csl.2005.02.006

Günther, Fritz; Marelli, Marco and Bölte, Jens (2020), Semantic transparency effects in German compounds: A large dataset and multiple-task investigation, *Behavior Research* 52, 1208–1224. https://doi.org/10.3758/s13428-019-01311-4

Günther, Fritz and Marelli, Marco (2022), Patterns in CAOSS: Distributed representations predict variation in relational interpretations for familiar and novel compound words, *Cognitive Psychology 134*, 101471. https://doi.org/10.1016/j.cogpsych.2022.101471

Levin, Beth; Glass, Lelia and Jurafsky, Dan (2019), Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds, *Linguistics*, *57*(3), 429–471. https://doi.org/10.1515/ling-2019-0013

Marelli, Marco; Gagné, Christina and Spalding, Thomas (2017), Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model, *Cognition 166*, 207–224. https://doi.org/10.1016/j.cognition.2017.05.026

Ortner, Lorelies; Müller-Bollhagen, Elgin; Ortner, Hans-Peter; Wellmann, Hans; Pümpel-Mader, Maria and Gärtner, Hildegard (1991), *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Eine Bestandsaufnahme des Instituts für Deutsche Sprache, Forschungsstelle Innsbruck*: De Gruyter.

Ó Séaghdha, Diarmuid (2008), Learning compound noun semantics (Doctoral dissertation), University of Cambridge, Computer Laboratory, Published as University of Cambridge Computer Laboratory Technical Report 735. https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-735.pdf

Telljohann, Heike; Hoppermann, Christina; Dima, Corina; Hinrichs, Ehrhard; Henrich, Verena and Versley, Yannick (2017), *Stylebook für die Annotation deutscher Nominalkomposita.* Seminar für Sprachwissenschaft, Universität Tübingen.

Semantic transparency across languages – Comparing French and Italian

Irene Fally
(University of Vienna)

Semantic transparency has a strong influence on lexical processing, facilitating word retrieval. However, there is some evidence that sensitivity to semantic transparency in lexical access can differ between languages, calling into question generalized models of lexical access (Creemers et al., 2020; De Grauwe et al., 2019; Smolka et al., 2018).

This study sets out to examine how semantic transparency affects processing of complex words in French and Italian, two related languages that share the same word formation mechanisms (Rainer, 2008). To test the impact of semantic transparency on word retrieval, two parallel primed lexical decision experiments with 50 native speakers of French and Italian respectively were conducted. In these experiments the effect of prior presentation of prefixed verbs to base verbs on decision latencies of the base verbs was measured, as exemplified in (1).

(1) a. (fr.) *surnommer* 'to nickname' – *nommer* 'to name'
    b. (it.) *richiedere* 'to ask again' – *chiedere* 'to ask'

For each language base-verbs were chosen and paired up with three prefixed verbs that have a transparent, opaque and no semantic relationship with the base verb (cf. 2 for example items from French). To ensure comparability between the two experiments the same types of base-verbs and prefixes were chosen from the *Lexique* 3.0 (New et al., 2004) and Subtlex-It (Crepaldi et al., 2015) databases. Thus the items for both experiments stem from subtitle corpora, further assuring comparability and allowing to control for frequency within and between experiments.

(2) a. (fr.) *remettre* 'to put back' – *mettre* 'to put'
    b. (fr.) *commettre* 'to commit' – *mettre* 'to put'
    c. (fr.) *déformer* 'to distort' – *mettre* 'to put'

The experiments are set up to determine whether speakers decompose the prefixed verbs and to what extent this is facilitated by semantic transparency. This includes the use of two measures of semantic transparency, to take into account potential differences in the underlying constructs (cf. Auch et al., 2020): one measure is derived from human ratings, where L1-speakers were asked to rate word pairs on a transparency scale, the other is based on vector semantic models (Grave et al., 2018; Marelli & Baroni, 2015).

Preliminary results show that French speakers are more sensitive to semantic information than speakers of Italian, where facilitation effects for semantically related items are generally weaker, regardless of which measure of semantic transparency is applied. These results suggest a robust difference in how semantic transparency affects processing in the two languages respectively, thus supporting claims that lexical access may be language specific. The aim of this talk is to explore repercussions of these differences within one language family and their implications for models of lexical access.

References

Auch, L., Gagné, C. L., & Spalding, T. L. (2020). Conceptualizing semantic transparency: A systematic analysis of semantic transparency measures in English compound words. *Methods in Psychology*, *3*, 100030. https://doi.org/10.1016/j.metip.2020.100030

Creemers, A., Goodwin Davies, A., Wilder, R. J., Tamminga, M., & Embick, D. (2020). Opacity, transparency, and morphological priming: A study of prefixed verbs in Dutch. *Journal of Memory and Language*, *110*, 104055. https://doi.org/10.1016/j.jml.2019.104055

Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., & Brysbaert, M. (2015, October). *SUBTLEX-IT. Subtitle-based word frequency estimates for Italian* [Talk]. Annual Meeting of the Italian Association for Experimental Psychology, Rovereto, 10-12 September 2015, Rovereto. http://crr.ugent.be/subtlex-it/

De Grauwe, S., Lemhofer, K., & Schriefers, H. (2019). Processing derived verbs: The role of motor-relatedness and type of morphological priming. *Language, Cognition and Neuroscience*, *34*(8), 973–990. https://doi.org/10.1080/23273798.2019.1599129

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018, March 28). Learning Word Vectors for 157 Languages. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*. LREC 2018. https://doi.org/10.48550/arXiv.1802.06893

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515. https://doi.org/10.1037/a0039267

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516–524. https://doi.org/10.3758/BF03195598

Rainer, F. (2008). Konvergenz- und Divergenzphänomene in der Romania: Wortbildung. In *Romanische Sprachgeschichte. Ein internationales Handbuch zur Geschichte der romanischen Sprachen* (pp. 3293–3307). De Gruyter.

Smolka, E., Libben, G., & Dressler, W. U. (2018). When morphological structure overrides meaning: Evidence from German prefix and particle verbs. *Language, Cognition and Neuroscience*, *34*(5), 599–614. https://doi.org/10.1080/23273798.2018.1552006

Shedding Light on Linguistic Shadows: A Cross-Lingual Experiment on the Transparency of Nominal Compounds in Large Language Models

Mădălina Zgreabăn, Alex Stasica, Cristina Reguera-Gómez & Denis Paperno
(Utrecht University, Utrecht University, Utrecht University & Utrecht University)

Prior research has demonstrated that the perceived compositional nature of compounds is influenced by their frequency (Bell & Schäfer, 2013, 2016). Despite the remarkable capabilities of large language models (LLMs), their behavior wrt semantic compositionality is poorly understood. Notably, no previous studies have explored whether LLMs exhibit a preference for more or less compositional compound forms.

We introduce a novel prompting methodology to investigate whether context modulates human and LLM preferences for expressions displaying differing levels of compositionality in similar ways. We hypothesize that humans and LLMs default to fully compositional compounds (Guz, 2012), but that contextual cues (eg. degree of figurativeness of the context) can influence their choices. In our task, the respondent (LLM or human) is given a definition of an existing compound and asked to express their preference between a compound and two plausible alternatives with different degrees of compositionality if they had to use it in a given context. For each compound, we created two contexts, one rich in metaphorical language and one literal. We include 20 compounds in three different languages (English, French, Romanian) representing fully compositional (C), partially compositional (PC), and non-compositional (NC) categories (Reddy et al., 2011). The NC are real compounds taken from the NCTTI dataset (Garcia et al., 2021) and EuroParl (Ziering & van der Plas, 2014), the PC and C compounds are plausible synonyms for NC compounds coined by us.

We evaluated the perplexity of several LLMs across three languages: English, Romanian, and French, and we selected the lowest perplexity value for each stimulus to signify the chosen compound form of the models. The models assessed included two sizes of Bloom (Scao et al., 2022), and two sizes of GPT-Neo (Gao et al., 2020) models. We also evaluated ChatGPT 3.5 (OpenAI, 2022) prompting it three times to choose amongst three compounds for each prompt for the three above languages.

Six multinomial logistic regression models were conducted to predict the categorical response variable (C, PC, NC) based on several predictors: literal/figurative context, language, and size of the LLM. These models concerned: each language with human participants, all LLMs, all languages for human participants, both LLMs and human participants.

Global results indicate a significant interaction between NCs and the metaphorical context: the more figurative a context is, the lower the NCs' perplexity. Analysis of combined human data across all languages revealed that the figurative context raises the likelihood of choosing NC or PC over C, with a strong effect observed in Romanian. Similarly, when combining data from all LLMs, the figurative context raises the likelihood of choosing NCs over Cs. The statistical model combining human and LLM data suggests that figurative context increases the likelihood of choosing NCs or PCs over Cs, impacting humans and LLMs similarly, as shown by no significant interaction between the type of respondent (human/LLMs) and the context.

References

Bell, M. J., & Schäler, M. (2013, March). Semantic transparency: Challenges for distributional semantics. In *Proceedings of the IWCS 2013 workshop towards a formal distributional semantics* (pp. 1-10).

Bell, M. J., & Schäfer, M. (2016). Modelling semantic transparency. *Morphology*, *26*, 157-199.

Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., ... & Leahy, C. (2020). The pile: An 800gb dataset of diverse text for language modeling. arXiv preprint arXiv:2101.00027.

Garcia, M., Vieira, T. K., Scarton, C., Idiart, M., & Villavicencio, A. (2021, August). Assessing the representations of idiomaticity in vector models with a noun compound dataset labeled at type and token levels. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 2730-2741).

Guz, W. (2012). Are nonce words really deviant, context-dependent, and unlexicalizable. *Sound structure and sense: studies in memory of Edmund Gussmann*, 223-238.

Oh, B. D., & Schuler, W. (2023). Transformer-based LM surprisal predicts human reading times best with about two billion training tokens. *arXiv preprint arXiv:2304.11389*.

OpenAI. (2022). ChatGPT 3.5. OpenAI Blog. https://openai.com/blog/chatgpt.

Reddy, S., McCarthy, D., & Manandhar, S. (2011, November). An empirical study on compositionality in compound nouns. In *Proceedings of 5th international joint conference on natural language processing* (pp. 210-218).

Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... & Al-Shaibani, M. S. (2023). Bloom: A 176b-parameter open-access multilingual language model.

Ziering, P., & Van der Plas, L. (2014, August). What good are 'nominalkomposita'for 'noun compounds': multilingual extraction and structure analysis of nominal compositions using linguistic restrictors. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 1047-1058).

Semantic transparency in formation of denominal verbs in Czech

Hana Hledíková & Magda Ševčíková

(Charles University & Charles University)

Keywords: nouns, verbs, conversion, prefixation, Czech

The paper compares the formation of denominal verbs in Czech by prefixation and by conversion in terms of semantic transparency. Prefixation is the most productive verb formation process in Czech, but, as in other Slavic languages, it is employed as the word-class-maintaining process of forming verbs from verbs (Körtvélyessy 2016). In the formation of verbs from nouns, which is the focus of this paper, prefixation is used less frequently than conversion. Whereas individual prefixes add specific, mostly transparent meanings to the input nouns, conversion shows a wide range of meanings, which are more difficult to predict (Štekauer 2006). In these respects, Czech is similar to English and some other languages, despite the differences in the languages' morphological structure (e.g. Plag 1999, Bauer et al. 2013, Valera 2023).

The study is based on a dataset that was extracted from a 100-million corpus of written Czech (Křen et al. 2015) and includes potentially all pairs made up of a verb and its motivating noun, namely 847 prefixless verbs converted from nouns and 230 verbs that were derived from nouns by prefixes. The meaning of the verb with respect to the motivating noun was determined by manual annotation, partially performed in parallel with an excellent inter-rater agreement.

Converted verbs are capable of expressing multiple semantic relations to the base noun, cf. ex. (1). With a total of 962 meanings assigned to the converted verbs, the data document an average polysemy rate of 1.14 meanings per verb. In contrast, a verb formed via a prefix has a single semantic relation to the noun; e.g. (2a) and (2b) (the unprefixed counterpart *lodit is not attested in the data). However, it is not the case that the meanings of prefixed denominal verbs are a subset of semantic categories attested in conversion. For instance, the PERFORMATIVE and AGENTIVE categories ((3) and (4), respectively) do not appear in denominal verbs formed by prefixation, while the PRIVATIVE category (5) is rare with converted verbs in Czech.

(1)   *slin-a*   →   *slin-i-t*
       saliva- NOM.SG   saliva-IPFV-INF
       'saliva'   'to salivate' or 'to wet with saliva'

(2)   a.   *loď*   →   *na-loď-i-t*
           ship   PREF-ship-PFV-INF
           'ship'   'to ship'
      b.   *loď*   →   *vy-loď-i-t*
           ship   PREF-ship-PFV-INF
           'ship'   'to debark'

(3)   *boj*   →   *boj-ova-t*
       fight   fight-IPFV-INF
       'fight'   'to fight'

(4)   *král* 'king' →   *kral-ova-t*
       king   king-IPFV-INF
       'king'   'to reign'

(5)   *vir-us*   →   *od-vir-ova-t*
        virus- NOM.SG   PREF-virus-PFV-INF
        'virus'   'to remove virus'

In addition, an analysis is carried out of the potential of converted verbs to enter into prefixation in order to express more subtle semantic nuances, always on top of one of the meanings of a polysemous verb (cf. both (6a) and (6b) related to the second meaning of the unprefixed verb in (1)). Although certain semantic categories have a tendency to be associated with certain prefixes, there is no one-to-one relationship.

(6)   a.   *slin-i-t*   →   *na-slin-i-t*
                saliva-IPFV-INF   PREF-saliva-PFV-INF
                'to wet with saliva'   'to cover with saliva'
        b.   *slin-i-t*   →   *pro-slin-i-t*
                saliva-IPFV-INF   PREF-saliva-PFV-INF
                'to wet with saliva'   'to soak with saliva'

References

Bauer, Laurie, Lieber, Rochelle, and Plag, Ingo (2013), *The Oxford reference guide to English morphology*, Oxford: Oxford University Press.

Körtvélyessy, Lívia (2016), Word-formation in Slavic languages, *Poznań Studies in Contemporary Linguistics* 52(3), 455–501.

Křen, Michal, Cvrček, Václav, Čapka, Tomáš, Čermáková, Anna, Hnátková, Milena, Chlumská, Lucie, Jelínek, Tomáš, Kováříková, Dominika, Petkevič, Vladimír, Procházka, Pavel, Skoumalová, Hana, Škrabal, Michal, Truneček, Petr, Vondřička, Pavel, and Zasina, Adrian J. (2015), *SYN2015: A Representative Corpus of Written Czech*, Prague: Charles University. http://www.korpus.cz

Plag, Ingo (1999), *Morphological productivity: Structural constraints in English derivation*, Berlin: Mouton de Gruyter.

Štekauer, Pavol (2006), On the meaning predictability of novel context-free converted naming units, *Linguistics* 44(3), 489–539.

Valera, Salvador (2023), The semantics of noun-to-verb zero-derivation in English and Spanish, *Zeitschrift für Sprachwissenschaft* 42(1), 153–180.

WS24 Typological approaches to non-canonicity in demonstratives

Demonstrative frames of reference in Norton Sound Kotlik Yup'ik

Nicholas Bunderson-Toler
(University of Alberta)

Keywords: Demonstratives, Frames of Reference, Space-Time-Discourse, Inuit-Yupik-Unangan

Central Alaskan Yup'ik is purported to have one of the world's most complex demonstrative systems with several dozen reflexes, which can surface as nominals, particles, and verbs (Miyaoka, 2012). However, previous descriptions based on elicited data posited an idealized and widely adopted paradigm, focusing on fitting the system into a canonical distance-based typology for spatial location (Diessel, 1999). Effectively, these accounts overlay a two-way distance opposition with features for object shape, accessibility, and riverine directionals (Jacobson & Jacobson, 1995). The resulting system is semantically imprecise and restricted in its descriptive power. Utilizing a corpus of spontaneous speech, I examine demonstrative usage in the under-documented Norton Sound Kotlik dialect and argue that their abundant semantics influence their use in spatial, temporal, and discourse reference and in discourse cohesion. My analysis of connected discourse presents an account of the 30-plus demonstrative reflexes that reduces the importance of distance oppositions while emphasizing frame of reference models (FoR), which index objects of attention from different perspectivizing grounds. Three distinct demonstrative FoR models emerge in the Yup'ik system: an intrinsic egocentric frame, a relative allocentric frame, and an absolute geocentric frame that contains a distance opposition (Diessel, 2014; Levinson, 2006). Within these FoR models, demonstratives emphasize the highly variable ground landmark, the figure's shape, and the speaker's perceptual space. As nominals, they bring objects into spatial and temporal joint attention or place narrative focus on objects, and as particles, they bring thematic cohesion to discourse. Demonstratives also cluster in topic constructions to help the speaker ground the textual world within the real world. Additionally, each FoR exhibits different grammaticalization patterns.

I showcase a few demonstratives functioning in each of these contexts in (1)-(2). In (1), a series of demonstrative reflexes, one from each FoR, are used in a topic cluster to index the topic of discussion temporally and ground it within the timeframe of the narrative. The first demonstrative base, *aw-*, indexes the figure by establishing an absolute FoR; the second, *uk-*, uses an intrinsic FoR; and the third, *tama(t)-*, uses a relative FoR. These three demonstratives are used together to triangulate and index the (temporal) location of the object of attention within a narrative. In (2), a single relative FoR demonstrative base, *tau(at)-*, is employed in several distinct functions and contextual distributions within a comment construction. The first usage is as one of the demonstrative particles used in discourse cohesion to link two clauses. The second is as a pronoun in a prenominal position serving to index and bring deictic joint attention to the figure. The third, by contrast, is in a postnominal position, which serves to place narrative focus on the object of attention, which in (2) is the main character indexed by the preceding demonstrative. My analysis of NSKY demonstratives—which make up around 30% of the NSKY Elder's lexical tokens in my corpus—applies three simpler but interconnected models functioning across multiple

domains of use rather than restricting the focus of Yup'ik demonstrative use within a single complex model functioning only spatially.

(1) **_awani_**$_i$ **_ukaqvarni_**$_j$ **_tamaani_**$_k$ _1980_
'**back then**$_i$, **right after this**$_j$, **during this**$_k$ **time** in 1980'

NSKY Corpus (MH 2016:90)

(2) **_Tua_**$_i$**_=llu=guq_**, **_tuana_**$_j$ _aqumgaluni, Kaikvayak=guq_—**_tauna_**$_k$—_qanertuq ..._
'**And then**$_i$ (**it is said**), **this one**$_j$ who is sitting, Kaikvayak—**this one**$_k$—he said ...'

NSKY Corpus (MP 2016:6.1)

References

Diessel, Holger. (1999). *Demonstratives form, function, and grammaticalization*. J. Benjamins.

Diessel, Holger. (2014). Demonstratives, frames of reference, and semantic universals of space. *Language and Linguistics Compass, 8 (3)*, 116–132. https://doi.org/10.1111/lnc3.12066

Jacobson, Steven A. & Anna W. Jacobson. (1995). *A practical grammar of the Central Alaskan Yup'ik Eskimo language*. Alaska Native Language Center and Program, University of Alaska.

Levinson, Stephen C. (2006). *Space in language and cognition: Explorations in cognitive diversity.* Cambridge University Press.

Miyaoka, Osahito. (2012). *A grammar of Central Alaskan Yupik (CAY)*. De Gruyter Mouton.

Noncanonical uses of similative demonstrative *gam2* in conversational Cantonese and beyond

Ryan Ka Yau La & Lu Liu
(University of California)

Keywords: similative demonstratives, manner deixis, grammaticalisation, discourse markers, Cantonese

Similative or manner-quality-degree demonstratives, translating roughly to *like this*, *thus* or *such* in English, have received considerable typological attention in previous years. Prior studies of Cantonese similative demonstrative 噉(樣) *gam2(joeng2)* (e.g. Chén 1985, Dèng and Tāng 1988, Sio & Tang 2007) have noted 'canonical' functions, i.e. anaphoric and exophoric reference to manners, qualities and propositions, and a clause-level connective use. Drawing from the Hong Kong Cantonese Corpus (Luke & Wong 2015), we argue that it has several other noncanonical functions, and situate them within current typological knowledge.

We first examine instances where *gam2(joeng2)* points to same-clause, immediately preceding preceding material. We find such uses as quotatives, general extender, approximative and hedge marker, widely attested in Eurasia (e.g. König 2020, Teptiuk 2020, Diessel & Breunesse 2020, Iwasaki & Dechapratumwan 2022):

(1)   佢   話   「Hi！阿琪！」   **噉**   喎,
        keoi5   waa6   haai1   aa3kei4   **gam2**   wo5
        3sg   say   Hi   A.Kei   **GAM**   FP
        They said, "Hi, Ah Kei!", like that,'

We show that these uses derive from *gam2*'s ability to close and/or mark as approximate the immediately preceding stretch of discourse (similar to Tibetan *'di-dra/de-'dra*), rather than the cataphoric and recognitional deixis often documented in Europe (König 2012, 2015, 2020).

Secondly, revisiting the connective function, we find that *gam2* not only expresses clausal relations like cause-effect, inference, and temporal progression, as extensively documented typologically, but also connect scenes to events and events to subevents (as in (2)), connect different arguments for the same point, preface actions warranted by previous talk, etc:

(2)   1   我   上年   呢,
           ngo5   soeng6nin2   ne1
           1sg   last.year   TOP
           'I last year,'

→   2   **噉**   上年   書展   我   都   有   去。
           gam2   soeng6nin2   syu1zin2   ngo5   dou1   jau5   heoi3
           GAM   last.year   book.fair   1sg   FOC   AUX   go
           '**gam2** Last year's book fair I also went.'

→   3   **噉**   我   喺度   買-咗   啲   字典   喎。
           gam2   ngo5   hai2dou6   maai5-zo2   di1   zi6din2   wo3
           GAM   1sg   PROG   buy-PFV   CL   dictionary   FP
           '**gam2** I was there and bought a dictionary.'

→   4   **噉**   我   覺得   「啊！

gam2	ngo5	gok3dak1		aa3
GAM	1sg	feel		ah
'**gam2** I thought 'Ah!'				

5	都	會	幾	好。」
	dou1	wui5	gei2	hou2
	FOC	IRR	quite	good
	'That would be quite nice.'			

We argue that *gam2* here expresses progression within an ongoing communicative project, within or across turns. This is evident in sequences where *gam2 resumes* a communicative project after progressivity is impaired (e.g. by repair):

(3)	1	C:	嗰度	對面	呢	有	間, //	誒,
			go2dou6	deoi3min6	ne1	jau5	gaan1	e6
			there	opposite	TOP	EXST	CL	er
			'Opposite of there, there's a, // uh,'					
	2		燒臘	舖	嘅	。		
			siu1laap6	pou2	ge3			
			roast.meat	shop	FP			
			'Roast meat shop.'					
	3	A:	係。					
			hai6					
			COP					
			'Right.'					
	4		哦	麵線	吖嗎,			
			o5	min6sin3	aa1maa3			
			oh	misua	FP			
			'Oh that's *misua* (a Fujianese noodle),'					
	5		係	唔	係	啊？		
			hai6	m4	hai6	aa3		
			COP	NEG	COP	FP		
			'Right?'					
	6	C:	係	嘞。				
			hai6	laa3				
			COP	FP				
			'Yeah'.					
→	7		**嗽**	呢	佢	就	有	一	隻
			gam2	ne1	keoi5	zau6	jau5	jat1	zek3
			GAM	TOP	3sg	FOC	have	one	CL

			非常之		利害		嘅	酸辣		米線。
			fei1soeng4zi1		lei6hoi6		ge3	syun1laat6		mai5sin3
			very		awesome		ASSOC	sour.spicy		mixian
			'**Gam2** they have very good sour and spicy *mixian* (a Yunnanese noodle).'							

Finally, *gam2* also prefaces responsive turns that either a) express mitigated disaffiliation with a previous action (as in (4)) or b) defends one's previous action after a challenge from the interlocutor:

(4)   [Context: A commented that it was 'cursed' for dogs to wear clothes.]

	1	B:	唔	得	呀？			
			m4	dak1	aa4			
			NEG	can	FP			
			'Is it not allowed?'					
→	2	A:	**噉**	又	冇	話	唔	得	嘅
			gam2	jau6	mou5	waa6	m4	dak1	ge2
			GAM	PRT	NEG	say	NEG	can	FP
			'Gam2 it's not that it's not allowed.'						

Likely grammaticalised directly from canonical anaphoric uses in turn-initial position, these uses have not been described for similative demonstratives in other languages; instead they partly overlap with words like English *well* and Spanish *pues*.

We argue that conversational data plays an invaluable role both for describing noncanonical uses of similative demonstratives and for uncovering grammaticalisation pathways of better-attested uses.

References

陈慧英 [Chén Huìyīng]. 1985. 广州话的"噉"和"咁" [*Gam2* and *gam3* in Cantonese]. 方言 [Fāngyán] (04). 297–304.

邓大荣 [Dèng Dàróng] & 汤燕珍 [Tāng Yànzhēn]. 1988. 广州话的"咁" [*Gam2/3* in Cantonese]. 广州研究 [Guǎngzhōu Yánjiù] (10). 48–51.

Diessel, Holger & Merlijn Breunesse. 2020. A typology of demonstrative clause linkers. (Ed.) Åshild Næss, Anna Margetts & Yvonne Treis. *Demonstratives in discourse*. Language Science Press 6. 305.

Iwasaki, Shoichi & Parada Dechapratumwan. 2022. Creating versatility in Thai demonstratives. *Studies in Language* 46(3). 517–558.

König, Ekkehard. 2012. Le rôle des déictiques de manière dans le cadre d'une typologie de la deixis. *Bulletin de la Société de Linguistique de Paris* 107(1). 11–42.

König, Ekkehard. 2015. Manner deixis as source of grammatical markers in Indo-European languages. In Carlotta Viti (ed.), *Perspectives on historical syntax*, vol. 33, 60. Amsterdam: John Benjamins.

König, Ekkehard. 2020. Beyond exophoric and endophoric uses: Additional discourse functions of demonstratives. In Åshild Næss, Anna Margetts & Yvonne Treis (eds.), *Demonstratives in discourse*, vol. 6, 21. Berlin: Language Science Press.

Luke, Kang Kwong & May LY Wong. 2015. The Hong Kong Cantonese corpus: design and uses. In Zhou Bin, Simon Smith & Michael Hoey (eds.), *Linguistic corpus and corpus linguistics in the Chinese context* (Journal of Chinese Linguistics Monograph Series 43), 312–333. Berkeley.

Sio, Joanna Ut-Seong & Sze-Wing Tang. 2007. The indexical expressions *gam2* and *gam3* in Cantonese. *Studies in Cantonese linguistics* 2. 55–73.

Teptiuk, Denys. 2020. Manner deictics in quotative indexes of Finno-Ugric. In Åshild Næss, Anna Margetts & Yvonne Treis (eds.), *Demonstratives in discourse*, vol. 6, 273. Language Science Press.

The conceptual domain of Sakhalin Ainu demonstratives

Elia Dal Corso

(Ca' Foscari University of Venice)

Keywords: Ainu, Sakhalin, demonstratives, semantics, anaphora

In this presentation, we will discuss four determiners (namely *ta*, *nea/neya*, *nean*, and *ne*) of Sakhalin Ainu, an SOV head-marking Ainuic language of Russia, which have been traditionally described in reference grammars as fulfilling the function of demonstratives (Murasaki 1979: 86). Most commonly, these four items syntactically precede either an independent noun, thus behaving as demonstrative determiners (1), or a dependent noun, such as *-p(e)* 'thing', thus behaving as demonstrative pronouns (2).

(1)   ***Ta***   *mahneku.*
         this   young.woman
         'This young woman.' (Piłsudski 1912: 86)

(2)   ***Ta*-*p***   *oka-ke-ta.*
         this-thing   after-PTV-LOC/LAT
         'After this' (Piłsudski 1912: 239)

Nea/neya and *nean*, which are employed chiefly for discourse deixis (Murasaki 1979: 86, Dal Corso 2021: 65-67), show no significant variation with regards to their syntax and semantics, and overall can be said to behave as canonical demonstratives (Diessel 2012).
On the contrary, *ta* exhibits a number of non-canonical functions, along with the canonical function of indicating spatial and discourse deixis, which are evident from morphosyntax. Specifically, *ta* or *tap*:
-   can be found in a post-verbal, pre-copula position to convey emphasis (3),
-   it is most often used as a place holder by elderly speakers,
-   it hosts possessive morphology (normally reserved for nouns and verbs) in what appears to be an anaphoric and topicalizing function (4),
-   it is used as the locative adverb 'here'.

(3)   *An-nukara*   *kusu*   *an-e-ko-asur*   *an-hi*
         PRM.A-3P.O/see CAU.FIN PRM.A-2S.O-APPL-spread.news   IPFV-PK
         ***ta*-*p*_**   *ne.*
         this-thing(EPH)   COP
         'Because I saw them, here I am reporting the news to you!' (Piłsudski 1912: 235)

(4)   ***Ta*-*ha***   *eci-nukara*   *kane*   *eci-yap_*   *kusu*   *neyke …*
         this-POSS   2P.A-3S.O/see   ADV   2P.S-go.ashore.PL   CAU.FIN   TOP
         'If you go towards land while seeing this [very thing] …' (Piłsudski 1912: 89)

Also, as noted by Tamura (2000: 263), the demonstrative *ne* has its origin in the copula *ne*. The copula is also the diachronic source of *nea* and *nean*, where it combines with the verbs *a* 'sit' and *an* 'exist'.

Due to their stative semantics, these two verbs developed the aspectual function of indicating the telicity features of a preceding verb, and were on the path of grammaticalization into auxiliaries (Satō 2006, Bugaeva & Nakagawa 2013, Dal Corso 2022). The semantic correlation between aspectual auxiliaries and demonstrative is non-trivial and suggest how *a*, *an*, and also the copula may in fact be regarded as a kind of demonstrative verbs (Guérin 2015).

Sakhalin Ainu demonstratives seem to cut across multiple word classes, which ultimately suggests that a chiefly semantic categorization of demonstratives in terms of conceptual (sub-)domains is more desirable than a syntactic one. Drawing from analogous instances of non-canonical demonstratives attested in other languages, we will try to draw the boundaries of the conceptual domain subsumed by the Sakhalin Ainu forms under scrutiny, which helps account more systematically for their specific pragmatic uses, and reason on whether this semantic approach can prompt their syntactic (re-)categorization.

References

Bugaeva, Anna and Hiroshi Nakagawa (2013), *V-V complexes in Ainu*. Presented at NINJAL International Symposium "Mysteries of Verb-Verb complexes in Asian languages", Tokyo, 14-15 December 2013.

Dal Corso, Elia (2021), *The language and folklore of West Sakhalin Ainu – a re-edition of Kyōko Murasaki's "Karafuto Ainugo" with translation and grammatical notes*. München: Lincom Publishing.

Dal Corso, Elia (2022), "Some insights on the clause linker -teh of Sakhalin Ainu through a consideration of aspect". Elisabetta Ragagnin and Bayarma Khabtagaeva (eds.) *Endangered languages of Northern Asia – Languages of Asia Series 28*. Leiden/Boston: Brill. 57-80.

Diessel, Holger (2012), "Deixis and demonstratives". *An International Handbook of Natural Language Meaning*, vol. 3. Berlin: Mouton de Gruyter. 2407–2431.

Guérin, Valérie. 2015. "Demonstrative verbs: a typology of verbal manner deixis". *Linguistic Typology* 19(2). 141–199.

Murasaki Kyōko (1979), *Karafuto ainugo: Bunpō hen* [Karafuto Ainu: grammar]. Tokyo: Kokushokan Kōkai.

Piłsudski, Bronisław (1912), *Materials for the Study of the Ainu Language and Folklore*. Cracow: Imperial Academy of Sciences.

Satō Tomomi (2006), "Ainugo Chitose hōgen no asupekuto kor an, wa an o chūshin toshite" [Focusing on aspectuals kor an, wa an in the Chitose dialect of Ainu]. *Hokkaidōritsu ainu minzoku bunka kenkyū sentā kenkyū kiyō 12*. Sapporo: Center for the Study of the Ainu People. 43-67.

Tamura Suzuko (2000), *The Ainu language*. Tokyo: Sanseido Co.

Towards a typology of deictic articles

Max Wahlström & Don Killian
(University of Helsinki & University of Helsinki)

Keywords: demonstratives, deixis, definite articles, definiteness, grammaticalization

Several unrelated languages of the world are reported to have a set of articles pertaining to the definite domain which have been characterized as deictic, such as Macedonian (Slavic) and Wolof (Atlantic). Curiously, however, grammaticalization accounts of definite articles often deem the loss of deictic meaning as fundamental for the process. The goal here is theoretical and methodological: through individual case studies we map the challenges this categorial overlap poses for the typology of deictic articles. We argue for a set of distributional criteria that help to distinguish deictic articles, particularly from demonstrative determiners.

Robert (2022: 19) states of Wolof that "definiteness obligatorily goes with the localization of the referent (as close vs. remote) with regard to the speaker… combining the consonant class marker with the proximal (-*i*) or distal (-*a*) spatial deictic."

Becher (2001) contrasts Wolof articles using *buur* 'king':

buur bi the king (here); the king who currently reigns
buur ba the king (there); the king of another country, or of times gone by

Table 1 shows a subset of Wolof demonstrative classes, using the *b-* noun class as an example.

Table 1. Wolof demonstrative classes

Class	Proximal	Distal
definite article or relativizer	*bi*	*ba*
deictic demonstrative determiner	*bii, bile*	*bee, bale*
adverb	*fi* (place), *ni* (manner)	*fa* (place), *na* (manner)
non-verbal predicator	*a-ng-i*	*a-ng-a*
adposition	*ci*	*ca*
*demonstrative pronoun	same as determiner	

Deictic definite articles can be thus viewed as a rare but distinct category of the umbrella category of *demonstrative* (following Diessel 1999), showing a systematic morphological and semantic relationship to other lexical categories of demonstratives (Table 2), albeit one that has never seen any coverage typologically.

Table 2. Classes of demonstratives

Part of Speech	Function	Example
determiner	used in apposition to a noun	I like *this* book.
adverb	marks adverbial deictic location, manner, etc.	He read the book *here*.
pronoun	replaces a noun	I like *this*.
non-verbal predicator	used in non-verbal predicates	*Here-is/this-is* John.
verb	used for verbal heads of predicates	The book *is-here.*
adposition	used to relate an NP to the clause	I bought soda *in-here* the store.
article	indicates definiteness of the referent	I saw *the-here* dog.

South Slavic Rhodopian and Pomak varieties and Standard Macedonian display a three-way definite article, proximal–neutral–distal; parallel distinctions are also found with the demonstrative pronouns. Some accounts question whether the proximal and distal articles primarily signal a deictic meaning in Macedonian (Sonnenhauser 2009), or whether they should be called articles in the first place (Topolińska 2006). Regarding Pomak, other specialized functions have been suggested (Adamou 2011). However, no study examines whether the less frequent articles pattern like the neutral article in expressing definiteness related meanings in the first place.

We discuss criteria suitable for the evaluation of "article-hood" for deictic articles. Following Becker (2021), we argue that to be called definite articles, they must systematically denote situationally unique referents, contextually unique referents, or bridging referents. With regards to Macedonian, based on a distributional analysis, we also call to question whether the deictic articles are definite articles in the first place.

References

Adamou, Evangelia, (2011), Temporal uses of definite articles and demonstratives in Pomak, *Lingua* 121(5), 871–889.

Becher, Jutta (1991*), Untersuchungen zu den Verbalweiterungen des Wolof*, MA thesis, Universität Hamburg.

Becker, Laura (2021), *Articles in the World's Languages*, Berlin / Boston: De Gruyter.

Diessel, Holger (1999), *Demonstratives: form, function and grammaticalization* (Oxford Studies in Typology and Linguistic Theory), Amsterdam: Benjamins.

Robert, Stéphane (2022), (preprint version) Wolof: a grammatical sketch, in: F. Lüpke (ed), (2009) *The Oxford guide to the Atlantic languages of West Africa*, Oxford: Oxford University Press.

Sonnenhauser, Barbara (2009), The Macedonian tripartite article: a discourse-oriented account. *Makedonski jazik* 60, 123–136.

Topolin'ska, Zuzana (2006), Trojniot člen – da ili ne? *Južnoslovenski filolog* 62, 7–1.

Demonstrative sentence connectives in Nivkh

Ekaterina Gruzdeva
(University of Helsinki)

Keywords: demonstratives, sentence connectives, clause linkers, discourse, Nivkh

In Nivkh (Amuric; Russian Far East) demonstratives form a closed class of forms that belong to six morphosyntactic categories: pronouns, determiners, verbs, adverbs, identifiers, and sentence connectives (Gruzdeva 2020). Sentence connectives are used to signal cross-clause or cross-sentence relations, typically summarizing the information expressed in the preceding discourse. In (Diessel & Breuness 2020), such forms are described as conjunctive adverbs, which represent one of the types of clause linkers. In Nivkh, these grammatical words, which can be conventionally called 'quasidemonstratives', are never used in the exophoric function, but they are employed in discourse functions typical of canonical demonstratives, which they also resemble structurally. Being very common in narratives, they can appear only sentence- or clause-initially and cannot move around in the sentence or clause, see (1–2):

(1) *nuknuko*   *cʰ-o:la*   *paɲd-r*   *pil-ŋan*   *cʰi*
     wait   2SG-child   grow_up-CVB_NAR.3SG be_big-CVB_ANT/SIM   2SG

     həjm-ɟ=ra.   **ha-ŋan**   *ɲi*   *cʰəŋ+ɟuz+ɲi-nə-ɟ=ra.*
     know-IND=FOC   do_like_that-CVB_ANT/SIM   1SG   2PL+meat+eat-FUT-IND=FOC
     'Wait, when your child grows up, you will know. Then I shall eat your meat.' (Panfilov 1965: 228)

(2) **hoʁo-ŋan**   *hə+uɣŋa*   *j-erq+cʰe-r*   *j-ajma-ɟ.*
     be_like_that-CVB_ANT/SIM   that+loon   3SG-side+turn-CVB_NAR.3SG   3SG-look-IND
     hoʁo-ror   *hə+ŋa*   *it-c.*
     be_like_that-CVB_ANT.3SG   that+beast   say-IND
     'Then that loon turning at his side, looked. After that that beast said.' (Panfilov 1965: 225)

Sentence connectives can express various semantic relations, such as relations of narration, time, condition, and concession. Derivationally, they represent different converbal forms of the quasidemonstrative verb *ha-* 'do/be like that' and the 'non-proximal' manner demonstrative verb *hoʁo-/hoʁa-* 'be like that', see the table.

ha- 'do/be like that'	*hoʁo-/hoʁa-* 'be like that'	Semantic relationships between clauses
ha-r : *ha-t*, *ha-gu-r* : *ha-gu-t*	*hoʁo-r* : *hoʁo-t*, *hoʁa-r* : *hoʁa-t*, *ho-r* : *ho-t*, *ho(ŋ)-gu-r* : *ho(ŋ)-gu-t*	narration
ha-rot/ha-ror : *ha-tot*	*hoʁo-rot/hoʁo-ror* : *hoʁo-tot*	temporality: anteriority
ha-ŋan	*hoʁo-ŋan*, *hoʁa-ŋan*	temporality: anteriority or simultaneity
ha-ɟke	*ho(ŋ)-ke*, *huŋ-ge*	temporality: anteriority
ha-ʁaj	*hoʁo-ʁaj*	condition
–	*hoʁa-gin*, *hoʁa-girn*	concession

Sentence connectives ending in -r(or/t) : t(ot) demonstrate an untrivial system of agreement, which has the same pattern but a different target of agreement, as compared to the corresponding converbs. In both kind of forms the consonant *r* is attested in the 2sg and 3sg forms, whereas the consonant *t* appears in all other persons and numbers. The converbs agree with the grammatical subject, cf. *paɲd-r* 'when grows up', which agrees with its subject *cʰ-o:la* 'your child' in (1). In contrast, as discourse markers, sentence connectives agree not with the subject, but with the topic, cf. *hoʁa-r* 'then' in (3), which agrees with the topic *if* 'he' that functions as the subject of the preceding clause:

(3)　　*if*　　*j-ama-ɟ*　　　*hur+mi*　　　*təf*　　*pil-kar*　　　*ha-ɟ.*
　　　　3SG　　3SG-look-IND　　here+inside　　house　be_big-AUG　　do_like_that-IND
　　　　hoʁa-r　　　　　　　　*puks-kir*　　*j-up-tot*　　　　　*qʰas+tʰov-ɟ.*
　　　　be_like_that-CVB_NAR.3SG　　belt-INSTR　　3SG-tie-CVB_ANT.3PL　　post+tie-IND
　　　　'He saw that the house was very big inside. Then (they) tied him to the post with the belt.'
　　　　(Panfilov 1965: 222)

The paper discusses the semantic and functional properties of sentence connectives in comparison with more canonical demonstratives. The analysis is based on the small corpus of four texts (altogether 1230 sentences) which have been published in (Panfilov 1965) and on my own field data.

References

Gruzdeva, Ekaterina (2020), Demonstratives in Nivkh: A semantic and pragmatic analysis, *Studia Orientalia Electronica* 8 (1), 1–60.

Diessel, Holger & Merlijn Breunesse (2020), A typology of demonstrative clause linkers, in Å. Næss, A. Margetts & Y. Treis (eds), *Demonstratives in discourse*, Berlin: Language Science Press, 305–340. (Topics at the Grammar-Discourse Interface 6)

Panfilov, Vladimir Z (1965), *Grammatika nivxskogo jazyka* [A grammar of Nivkh] 2, Moskva, Leningrad: Nauka.

The role of referent accessibility in demonstrative choice: Evidence from Vietnamese discourse particles

Huy Phan
(University of California)

Keywords: demonstrative, accessibility, Vietnamese, discourse particle, grammaticalization.

Beyond their basic functions of identifying and tracking language-external/-internal referents, demonstratives often acquire new roles of discourse organizing, information-structural marking, or stance-marking (cf. Næss et al. 2020). Current literature proposes that non-canonical functions emerge from semantic weakening since these usages no longer convey spatial distinction (i.e. near versus far) (e.g. König 2020, Iwasaki & Dechapratumwan 2022).

This grammaticalization pathway operates on the traditional assumption of distance being the semantic core of demonstratives (e.g. Diessel 1999). However, this assumption has been challenged by multiple studies, which claim that demonstratives serve mainly social and interactive purposes. For instance, Burenhult (2003) suggests that the semantic core is cognitive accessibility, conveying whether referents are accessible in the common ground. Moreover, Enfield (2003) suggests that this core locates a referent within the elastic, interactional 'here-space' that is formed during communication.

If the primary semantic meaning does not concern distance, then the lack of spatial distinction in grammaticalized terms does not automatically correspond to semantic weakening. The grammaticalization of demonstratives could involve a very different mechanism, one that has yet to be discussed but is addressed in the current study on Vietnamese demonstratives. Data for the study come from 98 minutes of YouTube videos of spontaneous conversations from 17 speakers.

The study focuses on the demonstrative subcategory 'demonstrative particles.' Compared to their standard counterparts, these particles are found only at the end of a sentence/utterance, have different phonological realizations, and do not convey physical distance. Instead, Bui (2014) proposes that they express psychological distance, a semantic extension of their physical sense, to indicate speakers' familiarity with a proposition (i.e. proximals for 'familiar,' distals for 'unfamiliar').

Category	Standard demonstratives	Demonstrative particles
Proximal	*Này* ('this')	*Nè*
Distal	*Kia* ('that/there')	*Kìa*
	Đó ('that')	*Á, ớ*

Comparison in the form of standard demonstratives and demonstrative particles

Spoken discourse data does not support this notion of psychological distance. Instead, speakers' nonverbal communication and discourse contexts suggest that particle choices reflect REFERENT ACCESSIBILITY, including both physical (i.e. visibility, reachability) and mental accessibility (e.g. Burenhult 2003, Visser 2020). Proximals are reserved for referents that are assumed to be or have become accessible, and distals for those that are inaccessible. Moreover, considering the collaborative nature of deictic referencing activities (cf. Clark & Wilkes-Gibbs 1986, Diessel 2006) further reveals an intersubjective aspect

of accessibility. In this sense, proximals actually correlate with referents that are accessible to *all* interlocutors while distals are for those that are accessible *only* to the speaker. Thus, non-canonical usages of demonstratives come not from semantic weakening nor semantic extension of spatial distance, but from a manipulation of the shared/non-shared status of accessibility. For instance, distal particles may be used strategically to subvert a shared referent to speaker-only, allowing speakers to promote their epistemic authority and establish their propositions as newsworthy, as seen in (1).

(1)   (OF and WM were talking about a homeless man next to them, who OF claimed to be her long lost relative. WM questioned OF's claim. OF responded to WM by pointing to the homeless man's teeth first then to hers while saying)

OF: *Hai   cái   răng   cửa   bự   như   chị   **đó***
       Two   CLF   teeth   door   big   like   3SG.F   DEM.PAR.DIST
       'His front teeth are big like mine ***đó***.' ('**As you may not have known**, his front teeth are big like mine.')

In summary, the current study posits accessibility as the primary semantic core of demonstratives in Vietnamese and perhaps in other languages, supporting Enfield (2003)'s suggestion that spatial distinction is but a secondary meaning that has developed out of a more intrinsic property of referents. Additionally, how accessibility shapes the emergence of Vietnamese demonstrative particles offers a new way to explain cross-linguistic similarities in the development of demonstratives into devices for information structure and discourse organizations that have been discussed in previous literature.

References:

Bui, L.T. (2014). *Vietnamese demonstratives: A spatially-based polysemy network.* The University of Queensland (PhD Thesis). doi:10.14264/uql.2014.430.

Burenhult, Niclas (2003): Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics*. International Pragmatics Association (IPrA). 13(3). 363–379. doi:10.1075/prag.13.3.01bur.

Clark, Herbert H. & Deanna Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition* 22(1). 1–39. doi:10.1016/0010-0277(86)90010-7.

Coventry, K. R., Valdés, B., Castillo, A., and Guijarro-Fuentes, P. (2008). Language within your reach: near- far perceptual space and spatial demonstratives. *Cognition* 108, 889–895. doi: 10.1016/j.cognition.2008.06.010

Diessel, Holger (1999). *Demonstratives: form, function and grammaticalization* (Typological Studies in Language 42). Amsterdam: Benjamins.

Diessel, Holger (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics* 17(4) doi:10.1515/COG.2006.015.

Enfield, N. J. (2003). Demonstratives in space and Interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 79(1), 82–117.

Jungbluth, K. (2003). "Deictics in the conversational dyad," in Friedrich Lenz (ed), *Deictic Conceptualizations of Space, Time and Person*, 13–40. Amsterdam: John Benjamins. doi: 10.1075/pbns.112.04jun

Iwasaki, Shoichi & Parada Dechapratumwan (2022). Creating versatility in Thai demonstratives. *Studies in Language* 46(3). 517–558. doi:10.1075/sl.20083.iwa.

König, Ekkehard (2020). Beyond exophoric and endophoric uses: Additional discourse functions of demonstratives. In Åshild Næss, Yvonne Treis & Anna Margetts (eds.), *Demonstratives in discourse*, 21-42. Language Science Press. doi:10.5281/ZENODO.4054814.

Næss, Åshild, Treis, Y. & Anna Margetts (2020). *Demonstratives in discourse*. Language Science Press. doi:10.5281/ZENODO.4054814.

Visser, Eline (2020). Tracking and recognitional use of Kalamang *opa*: Demonstrative of cognitive accessibility. In Åshild Næss, Yvonne Treis & Anna Margetts (eds.), *Demonstratives in discourse*, 21-42. Language Science Press. doi:10.5281/ZENODO.4054814.

Is the Estonian demonstrative *too* grammaticalizing into a personal pronoun?

Maria Reile, Piia Taremaa, Helen Hint & Renate Pajusalu

(University of Tartu, University of Tartu, University of Tartu & University of Tartu)

Keywords: reference, demonstratives, person, anaphora, Estonian

Third person pronouns and demonstratives are overlapping categories in many languages (Bhat 2004). Moreover, demonstratives often serve as the starting point for various grammaticalization paths, including the development into personal pronouns or, at the very least, into pronouns with personal reference (Heine, Kuteva 2006, see also Yurayong 2020: 204). Using demonstrative pronouns for personal reference is a common feature in Finnic languages, such as *se* in Finnish (Seppänen 1998) and *tuu* in Võro (Pajusalu 2015). In Estonian, both the proximal (*see*) and distal demonstrative pronoun (*too*) can refer to animate referents. Whereas the use of *see* has been well-studied, little is known about the potential grammaticalization of the distal demonstrative *too* into a personal pronoun.

Research on the exophoric use of demonstratives has shown that the Estonian *too* is used – similarly to distal demonstratives in other, typologically diverse, languages – to indicate that the referent is outside the speaker's grasping distance (Coventry et al. 2023). This provides evidence that Estonian *too* shares the same functions as distal demonstratives in other languages. However, unlike the typical trajectory of distal demonstratives that lead to grammaticalization into definite articles (Greenberg 1978: 61; Kuteva et al. 2019: 137-139), *too* shows no indication of evolving into a definite article. Instead, one of its main functions is to track the referent throughout the discourse, which tends to be a characteristic associated more with proximal demonstratives (Himmelmann 1996: 226)

Previous findings on anaphoric use indicate that the distal demonstrative *too* might be on its way to grammaticalizing into a personal pronoun. In anaphoric pronominal use, *too* frequently refers to human referents (Pajusalu 2006). In its adnominal form, *too* occurs frequently in time expressions. Additionally, the overall use of *too* is infrequent compared to the proximal/distance-neutral demonstrative *see*, and some speakers have reported that *too* implies a negative evaluation. This indicates the specialization of *too* to certain usage contexts.

Our research questions are as follows:

- What are the common referential functions of the Estonian demonstrative *too* when used as a bare pronoun in contemporary written Estonian?
- To what kind of human referents does *too* refer? Are there any indications that *too* is used as a (negative)-evaluating referential device?
- What are the usage patterns of *too* in terms of its morphosyntactic behavior?

Our data come from the most recent corpus of the Estonian language, Estonian National Corpus 2023, using the lexicological research tool Sketch Engine. The data consist of a random selection of 500 sentences where *too* is a bare pronoun. In the multifactorial data analysis, we include the referent's semantic, morphosyntactic, and referential features (Diessel 1999; Gundel, Hedberg & Zacharski 1993). In addition to quantitative methods, qualitative methods will also be used.

We expect that *too* refers most frequently to an activated (Gundel, Hegberg & Zacharski 1993) human referent with contrastive focus (see, e.g. Frajzyngier 2023: 374) and does not necessarily imply a

negative stance towards the referent. We also expect that some syntactic constructions (e.g. presentational clauses) are not typical contexts for *too.*

References

Bhat, D.N.S. (2004), *Pronouns*, Oxford: Oxford University Press.

Coventry, Kenny R.; Gudde, Harmen B.; Diessel, Holger; Collier, Jacqueline; Guijarro-Fuentes, Pedro; Vulchanova, Mila; Vulchanov, Valentin; Todisco, Emanuela; Reile, Maria; Breunesse, Merlijn; Plado, Helen; d ... Incel, Ozlem Durmaz (2023), Spatial communication systems across languages reflect universal action constraints, Nature Human Behaviour, 7, 2099–2110.

Diessel, Holger (1999), *Demonstratives. Form, Function and Grammaticalization,* Amsterdam—Philadelphia.

Frajzyngier, Zygmunt (2023), *A Typology of Reference System*, Oxford: Oxford University Press.

Greenberg, Joseph (1978), How does a language acquire gender markers?, in J. Greenberg et al. (eds.), *Universals of human language, vol. 3: Word structure*, Stanford: Stanford University Press, 47–82.

Gundel, Jeanette, Nancy Hedberg, Ron Zacharski (1993), Cognitive Status and the Form of Referring Expressions in Discourse, *Language* 69, 274–307.

Kuteva, Tania; Heine, Bernd; Hong, Bo; Long, Haiping; Narrog, Heiko; Rhee, Seongha (2019), World Lexicon of Grammaticalization. Second Edition, Cambridge: Cambridge University Press.

Heine, Bernd & Tania Kuteva (2006), *The Changing Languages of Europe,* Cambridge: Cambridge University Press.

Himmelmann, Nikolaus (1996), Demonstratives in Narrative Discourse: A Taxonomy of Universal Uses, in Fox, B. (ed.) *Studies in Anaphora*. Amsterdam/Philadelphia: John Benjamins

Pajusalu, Renate (2006), Death of a Demonstrative: Person and Time. The Case of Estonian too, *Linguistica Uralica* 42 (4), 241–253.

Pajusalu, Renate (2015), Võro demonstratives: changing or disappearing?, *Journal of Estonian and Finno-Ugric Linguistics* 6 (2), 167–190.

Seppänen, Eeva-Leena (1998), *Läsnäolon pronominit: Tämä, tuo, se ja hän viittaamassa keskustelun osallistujaan*, PhD Dissertation, Helsinki: Helsingin yliopisto.

Yurayong, Chinduang (2020), *Postposed demonstratives in Finnic and North Russian dialects*, Helsinki: University of Helsinki.

Finnish demonstrative *se* in the clause-second position:
Can it still be classified as demonstrative in a canonical sense?

Chingduang Yurayong & Seppo Kittilä
(University of Helsinki & University of Helsinki)

In spoken Finnish, the demonstrative *se*, classified spatially as medial or conversationally addressee-oriented (see e.g. Larjavaara 1990; Laury 1997), can be used in the clause-second position, following a topical unit of an utterance, as shown in (1). Importantly, such use is not possible for the other demonstratives: *tämä* [proximal] and *tuo* [distal].

(1)   *Meijän poika* **se** *anto-i.*
        [our   boy   **SE**]   give-3SG.PAST
        'Our son gave up his wish for his wife's sake!' (Suomi24, 2005)

From a comparative Finnic perspective, previous studies have shown that the discussed phenomenon is observed across contemporary Finnic languages. Similarly to Finnish, only the Proto-Finnic form **śej*, which is the only demonstrative mutually retained in all modern Finnic languages, can be used in this morphosyntactic context (see e.g. Larjavaara 1986). A broad distribution may suggest a relatively early emergence of this construction in Finnic, although clause-second demonstratives are not cross-linguistically rare, observed also in other language families (see Yurayong 2020).

Despite being relatively common in spoken language, the phenomenon has not been described extensively. Instead, only its partitive form *sitä* has been discussed as a discourse particle denoting the speaker's own experience and affectedness, and addressee-related inclusivity (see Hakulinen 1975; Hakulinen et al. 2004: §827). These previous observations serve as our starting point for further investigation of whether similar pragmatic functions also apply to the nominative form *se* and demonstrate several non-canonical aspects for demonstratives in the morphosyntactic behavior of *se* in the clause-second position. In this study, we use a questionnaire survey with native and near-native speakers of Finnish. Based on their responses, we discuss various formal and functional aspects, viewed from the perspective of epistemicity, evidentiality and engagement (cf. Bergqvist & Kittilä 2020).

Our observation and analysis result in the following generalisation. Morphosyntactically, *se* more frequently occurs after a nominative constituent, whereas the partitive form *sitä* is used after a constituent in other cases as well with similar pragmatic functions. Functionally, *se* no longer exclusively serves a referential function as a resumptive pronoun as understood in some previous studies (cf. Larjavaara 1986). We also observe instances of *se* co-occurring with another preceding demonstrative which is rather primarily responsible for the expression of referentiality of the whole constituent unit. Subsequently, this may entail that a functional scope of the clause-second *se* can control more broadly pragmatic nuances of the entire clause, expressing the speaker's evaluation of the state-of-affairs concerned. Usage-wise, a wide range of contexts are noticed, particularly contrastive reading, direct and concrete evidence, addressee-centrality, and the speaker's affectedness such as sarcasm and irony. The mentioned contexts demonstrate that the functions of

clause-second *se* have drifted away and extended beyond its canonical referential uses towards information-structural, evidential, and evaluative uses (see also Yurayong & Kittilä 2024).

Our study raises a more general question whether this type of non-canonical usage must also be ontologically considered as part of the overall functional range and complexity of demonstratives. Discussing this matter can ultimately contribute to the improvement of a theoretical basis for both language-specific and cross-linguistic research of demonstratives.

References

Bergqvist, Henrik, & Seppo Kittilä (eds.). 2020. *Evidentiality, egophoricity and engagement*. Berlin: Language Science Press.

Hakulinen, Auli. 1975. Suomen *sitä*: Pragmatiikan heijastuma syntaksissa [The Finnish *sitä*: Reflection of pragmatics in syntax]. *Sananjalka* 17. 25–41.

Hakulinen, Auli, Riitta Korhonen, Maria Vilkuna, & Vesa Koivisto. 2004. *Iso suomen kielioppi* [The big Finnish grammar]. Helsinki: Finnish Literature Society.

Larjavaara, Matti. 1986. *Itämerensuomen demonstratiivit, 1, Karjala, aunus, lyydi ja vepsä* [Finnic demonstratives, 1, Karelian, Olonets Karelian, Lude and Veps]. Helsinki: Finnish Literature Society.

Larjavaara, Matti. 1990. *Suomen deiksis* [The Finnish deixis]. Helsinki: Finnish Literature Society.

Laury, Ritva. 1997. *Demonstratives in Interaction*. Amsterdam: John Benjamins Publishing Company.

Suomi24 = City Digital Group. 2021. *The Suomi24 Sentences Corpus 2001–2020, Korp version* [text corpus]. Helsinki: Language Bank of Finland. http://urn.fi/urn:nbn:fi:lb-2021101525.

Yurayong, Chingduang. 2020. *Postposed demonstratives in Finnic and North Russian dialects*. Doctoral dissertation. University of Helsinki.

Yurayong, Chingduang & Seppo Kittilä. 2024. A typological approach to intersubjective uses of the Finnish clitic markers *=hAn* and *=se* from the perspectives of engagement and their interrelations with subject person. *Open Linguistics* 10 (Special Issue: Subjectivity and Intersubjectivity in Language - Part II). 20240006.

Demonstrative forms vs. demonstrative systems:
Patterns of overlap and cooccurrence in Äiwoo demonstratives

Åshild Næss
(University of Oslo)

Typologising over demonstratives requires not only descriptions of the functions of individual forms, but also an adequate understanding of the larger system that they enter into. This may appear trivial for demonstratives in their canonical spatial functions, cf. the traditional distinction between distance-based and person-based demonstratives and the number of distinctions involved in each case; but it becomes challenging when demonstrative forms show a wide range of functions at the level of discourse and information structure, and where multiple sets of demonstratives interact in complex constructions.

This paper will discuss this challenge as it applies to the Oceanic language Äiwoo. Äiwoo shows a range of demonstrative forms all encoding the same basic proximal-distal distinction, and these interact and cooccur in complex ways. Äiwoo uses demonstrative particles and enclitics to delimit interpretationally relevant 'chunks' of speech and link them to preceding discourse (Næss 2023). In addition, the two sets of forms *kele/kâlâ* and *ile/ilâ* show a complex behaviour which includes both spatial and discourse-related functions. They differ in their basic function in that the former is locational, the latter identificational, as suggested by these examples:

(1) a. Butete kele.
         sweet.potato here
         'Here there are potatoes/The potatoes are here.'

    b. Butete ile.
         sweet.potato this
         'These are potatoes.'

In reference to discourse entities, *kâlâ* tends to be cataphoric while *ilâ* is anaphoric. However, the forms also overlap in a range of contexts, and in many of these they cooccur with other demonstrative forms. This is illustrated in (2-3), where in (2), *kâlâ* and *ilâ* both precede a posture verb, and in (3) *kele* and *ile* premodify the head of a noun phrase; in all cases another demonstrative form occurs at the end of the phrase, and it is far from clear what each demonstrative form contributes to the overall interpretation:

(2) a. Doo=le nugo-nââ   le   ki-amoli-wâ-no=nge
         what=PROX leaf-sago   PROX   IPFV-see.UV-DIR:2-1MIN=PROX
         kâlâ   **ki-so=dä=nâ**?
         there   IPFV-stand=some=DIST
         'Is that sago I see standing there?'

    b. Mo   mi-tu-mä-ji   **ile**
         but   one-bring-DIR:1-1+2MIN   PROX

> **ki-tokoli-lâ-wâ**   **ngä**   **nyibä=ke.**
> IPFV-sit-go.out-DIR:2   LOC   basket=PROX
> 'But the one we brought is sitting here in the basket.'

(3)   a.   **Kele**   **nye-liluwaio**   **enge**ₗ
           here   NMLZ-be.small.PL   DEM:PROX
           nâ-lotaläi-to-mä-i=to   ngä   bepo.
           IRR-prepare-go.in-DIR:1-3AUG=now   LOC   k.o.basket
           'These small ones, let them put them in your basket'

       b.   **Ile**   **lopâ=enge**   i-kää   mi-li-dâu.
           PROX   talk=DEM:PROX   PFV-know   one-3AUG-many
           'This story is known by a lot of people.'

Teasing apart the functions of each form thus involves not just accounting for the distribution of each form individually, but also identifying the contributions of each demonstrative in complex constructions. I will discuss these challenges as they apply descriptively to the case of Äiwoo, but also the broader challenge involved in taking a typological approach to such complex interactions between demonstrative forms: they require us to typologise not just over sets of individual forms with distinct functions, but over complex patterns of cooccurrence and interaction.

Non-standard abbreviations
AUG augmented number, DIR directional, LOC locative preposition, MIN minimal number, PROX proximal, UV undergoer voice

References
Næss, Åshild (2023), Demonstratives taking over discourse: the grammaticalisation of deictic clitics in Äiwoo. Paper presented at ICHL26, Heidelberg, September 2023.

Abui demonstratives as interactional regulators

František Kratochvíl
(Palacký University Olomouc)

Demonstratives have long been considered as primarily spatial devices and in consequence the non-spatial uses were considered as non-canonical even though they are widespread and perhaps even more salient. There are voices suggesting moving beyond spatial understanding of demonstratives and to consider seriously their interactional and social-cognitive properties and to reframe the account of demonstratives from that perspective. It is time to move away from an egocentric perspective focusing on the spatial to empirically supported accounts which acknowledge the joint action that demonstrative reference constitutes (cf. Peeters and Özyürek 2016).

This paper recapitulates some major findings in research on demonstratives from a social cognition perspective, from spatio-logophoric deictic elements to socio-interpersonal devices. Demonstratives demarcate the speaker's sphere of interest or influence (e.g. Enfield 2003, González Pérez 2023), and manage interpersonal relationships in individual languages (e.g. Heritage 2012) and cross-linguistically (Evans et al. 2018a,b). Demonstratives are deployed as markers of speaker's 'territory of information' (Kamio 1994) and as devices 'for managing interlocutor alignments' (Du Bois 2007).

We show that in Abui (Timor-Alor-Pantar family) demonstratives manage joint attention and shared knowledge, marking the resets and updates thereof (cf. Hanks 1992, Tomasello et al. 2007, Kratochvíl 2011, Kratochvíl and Delpada 2015). We show that Abui demonstrative roots are semantically quite empty except of the viewpoint and distance information which serves as a prompt for a search in the relevant knowledge domain(s) (spatial, temporal, social). However, the domain is not indicated and has to be deduced from the context, similar to Phola, as argued by González Pérez (2023). The viewpoint alternation and engagement are an integral part of the Abui deictic and grammatical system, regardless of whether the addressee-based forms are used or not as shown below, where the speaker takes the addressee's perspective in selecting the distal adverbial demonstrative *ya* (usually used with invisible reference) to refer to her own position.

(12)   context: speaker inside the house, not visible to addressee, whose perspective is chosen
         na   ***ya***   *e-afu*   *walangai*   *h-iel*
         1SG.AGT   DIST   2SG.AL-fish   fresh   3.PAT-roast.IPFV
         'I am roasting you a fish here (lit. over there (invisible for ADDR), i.e. in the kitchen)'

Because the degree of shared knowledge differs across speech acts (conversations among friends, interview with healers, performing experimental tasks) the interpretation of the demonstratives is sensitive to discourse actions (directives, explanations, invitations, and question-answer pairs). The interpretation is revealed in the subsequent discourse (acceptance; agreement; repetition; (re)negotiation; etc.).

References

Enfield, Nick J. (2003), Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis, *Language* 79(1), 82–117.

Evans, Nicholas, Henrik Bergqvist & Lila San Roque (2018a), The Grammar of Engagement I: Framework and initial exemplification, *Language and Cognition* 21, 1– 31.

Evans, Nicholas, Henrik Bergqvist & Lila San Roque (2018b), The Grammar of Engagement II: Typology and diachrony, *Language and Cognition*, 22, 1–30.

González Pérez, Manuel David (2023), Spheres of interest: Space and social cognition in Phola deixis, *Open Linguistics* 9(1), 20220215.

Hanks, William (1992), The indexical ground of deictic reference, in Alessandro Duranti & Charles Goodwin (eds.), (1992), *Rethinking context: language as an interactive phenomenon*, Cambridge: Cambridge University Press, 43–76.

Heritage, John (2012), Epistemics in action: Action formation and territories of knowledge, *Research on language & social interaction* 45(1), 1–29.

Kamio, Akio (1994), The theory of territory of information: The case of Japanese, *Journal of Pragmatics* 21(1), 67–100.

Kratochvíl, František (2011), Discourse-structuring functions of Abui demonstratives, in Foong Ha Yap, Karen Grunow-Hårsta & Janick Wrona (eds.), (2011), *Nominalization in Asian languages. Diachronic and typological perspectives*, Amsterdam: John Benjamins, 757–788.

Kratochvíl, František, and Benidiktus Delpada (2015), Definiteness and Specificity in Abui, in *Proceedings of the Second International Workshop on Information Structure of Austronesian Languages*, Tokyo: Research Institute for Languages and Cultures of Asia and Africa Tokyo University of Foreign Studies, 179–208.

Peeters, David, and Aslı Özyürek (2016), *This* and *that* revisited: A social and multimodal approach to spatial demonstratives, *Frontiers in Psychology* 7, 222.

Tomasello, Michael, Malinda Carpenter & Ulf Liszkowski (2007), A new look at infant pointing, *Child Development* 78(3), 705 –722.

The postpositive demonstrative clitic *-to* and relativization:
A case of South Slavic Torlak

Bojana Damnjanović
(University of Helsinki)

I this paper, I will address questions regarding relativization and use of post-positive demonstrative clitics *-to* attached to relativizers in Torlak dialects. Relative clause is seen as a clause modifying the noun phrase (Givon 2001: 175). Only restrictive relative clauses will be taken into analysis. I aim to answer the question: what can a relative clause with the clitics *-to* modify? By answering this question, my objective is to advance our knowledge on the various properties of demonstratives, their syntactic multifunctionality, and their semantic specification.

Gendered demonstrative pronouns *taj, ta, to* 'this' in standard Serbian precede the noun they modify and are accentuated. In Torlak dialects, both demonstrative pronouns *taj, ta, to* and post-positive articles *-at, -ta, -to,* derived from them, exist. Post-positive articles in Torlak are not accentuated and they are attached to the noun they modify. This is the most salient feature, that sets Torlak apart from standard Serbian and other Serbian dialects, and brings it into the Balkan Slavic group, together with Macedonian and Bulgarian. In standard Bulgarian and Macedonian, post-positive demonstratives are fully grammaticalized as definite articles, whereas in Torlak, they appear less frequently used and have not been thoroughly studied (Vuković 2022: 231). The example (1a) illustrates demonstrative pronoun and (1b) article attested in Torlak (adapted from Vuković 2022: 232):

(1)   a. tá žéna   b. žená-ta

         that.F.SG.NOM woman.F.SG.NOM   woman.F.SG.NOM-DEM.F.SG.NOM

I this paper, I aim to analyze the formation of relative clauses using the two most frequent relativizers: 1) *koji* and *koji-to* (meaning 'who' or 'which'), a variable relativizer; 2) and the pronoun *što* and *što-to* (meaning 'that'), an invariable relativizer. I wish to determine the syntactical-semantical surrounding in which relativizers would be found with the attached *-to* clitics and the form of demonstrative meaning.

Here are some preliminarily examples:

(2)   sin   što-to   nestade

         son   that-DET   disappear.AOR.3SG

 'a son who has disappeared'

(3)   a   mi koj-to   smo   ležali   na slamu-tu

         and   we who-DET   aux.1PL   lay.PCPT on hay-DET

'and we who were lying on the hay'

My focus area is the Torlak speech communities in Serbia. I will examine the semantics of the head noun. Specifically, I will examine how animacy can shape speakers' choice in the usage of post-positive demonstrative clitics *-to*.

Data will be extracted from the corpora of Torlak, which consists of semi-structured interviews with the Torlak speakers. Interviews were conducted during the period between 2015 and 2018, following the questionnaire by Plotnikova (1996).

I expect to find attached element *-to* in noun and pronoun relativization when the head noun is [+human]. Since it is not fully grammaticalized as article, I expect to find attested examples of case forms of demonstrative pronoun, such as *-tu* 'this'OBL.F.Sg. or *-toga* 'this'OBL.M.Sg. The findings will contribute not only to the understanding of Slavic dialectal syntax but also to the variation in other non-standard linguistic varieties of Europe.

References

Arsenijević, B. (2018). Atypical demonstratives in an articleless language. *Atypical demonstratives: Syntax, semantics, and pragmatics*, *568*, 161–196.

Mišeska Tomić, O. (2006). Balkan Sprachbund Morpho-Syntactic Features. *Studies in Natural Language and Linguistics Theory*, 67, Dordrecht: Springer.

Murelli, A. (2011). Relative Constructions in European Non-Standard Varieties. Berlin Boston: Mouton de Gruyter.

Plotnikova, A. A. (1996). Materialy dlja etnolinggvisticheskogo izuchenija balkanoslavjanskogo areala. Institut slavjanovedenija RAN.

Shagal, K. (2016). Factors regulating variation in Macedonian relative clauses. In: M. Makartsev, M. Wahlström eds., *Slavica Helsingiensia*, Vol. 49, Helsinki: University of Helsinki, 163–177.

Vuković, T. (2020). Spoken Torlak dialect corpus 1.0. CLARIN.SI. https://www.clarin.si/repository/xmlui/handle/11356/1281 .

Vuković, T., Escher, A., & Sonnenhauser, B. (2022). Degrees of non-standardness: Feature-based analysis of variation in a Torlak dialect corpus. *International Journal of Corpus Linguistics*, *27*(2), 220-247.

WS25 What is egophoricity in Tibetic, and beyond?

Copulas in Kutang (Northern Gorkha, Nepal): "caught between descriptive systems" – and evidence for the unity of these systems

Marius Zemp
(University of Berne)

Until recently, all we knew about Kutang language stemmed from a socio-linguistic survey of Northern Gorkha conducted by Jeff Webster in 1992, in which 240 words were elicited by Warren Glover in around a dozen related 'Ghale' varieties. After Dhakal et al. (2012) presented a first sketch of the Kutang past tense distinctions, Donohue and Gautam (2019) observed that the distinctions between two existential (*jaŋ* and *goŋ*) and two equative (*na* and *noŋ*) copulas could be accounted for in terms of either evidentiality or egophoricity (or 'conjunct/disjunct'), if not mirativity.

Based on interviews conveyed in Kutang after the 2015 earthquake (supervised by Kristine Hildebrandt, https://av.mandala.library.virginia.edu/collection/kutang-reflections-2015-nepal-earthquakes) as well as more than twenty stories I recorded there in 2023, I will show that the existential and equative copulas of this language form two evidential contrasts that are fully in line with those found in all modern Tibetic varieties as well as a number of languages that have been influenced by the latter. These evidential contrasts all share the following characteristics:

i. The contrasting verb forms are defined against each other (and hence, the contrasts are mostly binary).

ii. While the forms have contrasting implications as to how one knows what is being profiled in a statement (e.g. direct evidence vs. plain knowledge, new vs. old knowledge, high vs. low personal involvement, or privileged vs. non-privileged access), one of them always implies longer and/or more direct personal involvement than the other. Accordingly, contrasting evidentials qualify as ego- and allophoric, respectively.

iii. Contrasting evidentials have a shared aspecto-temporal value (i.e. the *tertium comparationis* regarding which they have contrasting evidential implications).

iv. Contrasting forms reflect the perspective of the same (evidential) origo, which corresponds to the speaker in statements, the addressee in questions, and the source in reported speech clauses.

The contrast between *goŋ* and *jaŋ* resembles that between the two existential copulas *ŋja* and *ɲipa* in Kaike (Watters 2006) and that between the suffixes *-li* and *-yek* in Wutun (Sandman 2018), which both indicate whether a statement about a present state is based on what the interlocutors currently see before them (new knowledge) or on the speaker's personal experience up to the moment of speech (old knowledge). Unlike Kaike *ŋja* and Wutun *-li*, however, Kutang *goŋ* may also refer to past states (e.g. *lemu goŋ* '(it) is/was tasty' may refer to something the speaker is currently eating or something s/he tasted on the previous day). Hence, it seems to even more closely resemble 'direct evidential' Purik Tibetan *duk* (Zemp 2017). While the existential copula *duk* derives from a Proto-Tibetan verb form meaning *'stayed, was there', the cognate full verb mostly means 'sit' in modern Purik. As this is also what *goŋ* means in some varieties of Tamang (Mazaudon 1994: 86), it seems likely that Kutang *goŋ* derives from a form of this verb which meant 'sat, was there', and – like the Tibetic *duk* – in contrast to a neutral existential copula came to imply that the speaker directly witnessed the thing referred to.

References

Dhakal, Dubi Nanda, Mark Donohue, Bhojraj Gautam & Naijing Liu. 2012. Diagnosing a contact history for Tsum. *Nepalese Linguistics* 27: 14–20.

Donohue, Mark & Bhojraj Gautam. 2019. Truth, person, and personal truth: Kuke copulas, a construction caught between descriptive systems. *Studies in Language* 43(2): 444–58.

Mazaudon, Martine. 1994. Problèmes de comparatisme et de reconstruction dans quelques langues de la famille tibéto-birmane. Thèse d'Etat, Université de la Sorbonne Nouvelle.

Sandman, Erika. 2018. Egophoricity in Wutun. In Simeon Floyd, Elisabeth Norcliffe & Lila San Roque (eds.), *Egophoricity*, 173–96. Amsterdam: Benjamins.

Watters, David. 2006. The conjunct-disjunct distinction in Kaike. *Nepalese Linguistics* 22: 300–19.

Webster, Jeff. 1992. A sociolinguistic survey of the Tibeto-Burman dialects of North Gorkha District, Nepal. Kathmandu: The Summer Institute of Linguistics Nepal. https://www.sil.org/resources/archives/70341.

Zemp, Marius. 2017. The origin and evolution of the opposition between testimonial and factual evidentials in Purik and other varieties of Tibetan. *Open Linguistics* 3(1): 613–37.

A corpus study of egophoric markers in spoken Lhasa Tibetan

Jan Nuyts & Eric Mélac
(University of Antwerp & Université Paul Valéry Montpellier 3)

This paper aims to test the main assumptions formulated in the cross-linguistic literature regarding the nature and properties of egophoric markers (see notably San Roque et al. 2018) on the basis of a corpus of spoken Lhasa Tibetan (the language for which the notion was first proposed in Tournadre 1992).

The corpus involves 4 hours of semi-guided conversations with 8 native speakers of Lhasa Tibetan (Tibet Student Corpus; 26,000 words, collected in 2010-2011; Mélac 2014). We have extracted all occurrences of the three egophoric markers in the language (413 instances in total): *yod*, *yin*, and *byung* (and their negative alloforms). For *yod*, we have considered the use of the form as a copula, an existential-possessive-locative verb, and a suffix (both perfective *-yod* and imperfective *-gi yod*). For *yin*, we have looked at the use as a copula and as a suffix (both perfective *-pa yin* and future *-gi yin*). For *byung*, too, we have extracted both the copula and the suffix (perfective 'receptive' *-byung*).

We have examined a range of features, including, as the most important ones:

(i)   To what extent does the use of the markers concern the speaker's personal experience of his/her own volitional action? More generally, what is the type of event in the scope of the marker (dynamicity, control) and what are its aspectual properties (perfectivity, habituality, etc.)?

(ii)   To what extent is the use of the markers in line with the principle of person sensitivity (first person in declarative clauses, second person in interrogative clauses), and when it is not, what are the possible reasons?

One of the main findings is that the different markers behave very differently in terms of these features, thus questioning the status of egophoricity as a semantically homogeneous notion. For example, control and volition turn out to be irrelevant for copular *yin* and *byung*, as well as for existential-possessive-locative *yod*, while the suffixes *-pa yin* and *-gi yin* only appear with controllable verbs. Contrary to certain claims, the notion of egophoricity as a whole does not seem to be associated with specific aspectual configurations, since *yod*, *yin*, and *byung* may be used with all types of events (depending on their morpho-syntactic categories). Person sensitivity, that is, the association of egophorics with first person subjects in declarative clauses and second person subjects in interrogative clauses, is on the whole statistically confirmed. However, our corpus data show, for instance, that in declarative clauses suffixal *yin* (both *-pa yin* and *-gi yin*) is used with first person subjects 100% of the time, while this is true for only a small minority of copular uses of *byung*.

References

Mélac, Eric (2014), *L'évidentialité en Anglais: Approche contrastive à partir d'un corpus Anglais-Tibétain*, PhD dissertation, Université Sorbonne Nouvelle – Paris 3.

San Roque, Lila, Simeon Floyd & Elisabeth Norcliffe (2018), Egophoricity: An introduction, in S. Floyd, E. Norcliffe & L. San Roque (eds.), (2018), Amsterdam: John Benjamins, *Egophoricity*, 1–78.

Tournadre, Nicolas (1992), La déixis en tibétain, in M.-A. Morel & L. Danon-Boileau (eds.), (2018), *La deixis*, Paris: PUF, 197–208.

The Emergence of Egophoric Marking: A Comparative Diachronic Analysis Egophoric Marking in Tibetan and Newar

Marieke Meelen, Nathan Hill, Alexander O'Neill & Christian Faggionato
(University of Cambridge, Trinity College, Musashina University & University of Cambridge)

In this talk, we will focus on how languages develop egophoric marking in different ways. Newar and Tibetan offer an excellent starting point to answer this nearly unexplored question because of their long written histories. In Modern Lhasa Tibetan egophoric markers like yin can be found in a wide range of examples, but always with reference to the perspective of the speaker, e.g. *nga em- chi yin'* I'm a doctor', *'di nga'î bu-mo yin'* This is my daughter', and *'di khyed-rang-gi gsol-ja yin* 'This is your tea [that I have made for you]'. Classical Tibetan has the verb yin to, but it is a normal copula that does not exhibit egophoricity (Tournadre & Jiatso, 2001). On the Newar side, Modern Kathmandu Newar exhibits egophoricity in inflectional ("conjunct/disjunct") morphology (Hale & Shrestha, 2006), temperature predicts, simple/causative change in the auxiliary *dhun-* "finish," (Hargreaves, 2015) and in dependence upon formality (Shrestha, 2023). However, Classical Newar exhibits greater variety, for example, using the same forms in both conjunct/disjunct cases.

First, we show how Lhasa Tibetan *byung* with a recipient egophoric function (*khog-gis gnang- byung'* He gave it to me') differs from its Dzardzongke Tibetan cognate, which has kept its original lexical meaning only, not conveying any kind of egophoric perspective even today (cf. Meelen & Ramble 2022). Our deeply-annotated diachronic corpus allows us to find instances of *byung* in both its morphosyntactic and information-structural context, thus tracking different types of arguments, e.g. animate vs inanimate subjects, and, in particular, whether *byung* can be combined with datives providing an alternative way of indicating the recipient.

Second, we trace the modern Newar morphological forms for egophoric markers back to Classical Newar, paying specific attention to the context and information-structural features that can reveal how the compulsory marking of speaker perspective arose. In Modern Kathmandu Newar, for verbs expressing controllable, intentional, and subject-evidential actions, the egophoric conjunct form is used (such as in first-person declarative and second-person interrogative situations, e.g. *jiṃ rājeś nāpalānā* "I met Rajesh," vs. *vaṃ rājeś nāpalāna"* He met Rajesh."). In Classical Newar, however, this distinction appears not to be consistently applied, and we see the ancestor of the modern egophoric marker or "conjunct" cases, whereas, in Modern Newar, we would expect the disjunct form. For example, Classical Newar has examples of "egophoric" morphology (< perf.ptc.-ā) in contexts that have nothing to do with speaker perspective *"thana māhānanaṃ inālapā, bho mahārāja,"* "then the official said (perfective participle): O great king…" (Or. 10 Br.M., fol. 7a). In Modern Newar varieties, these contexts require a non-egophoric ("disjunct") ending -*aṃ* or -*a* derived from the perfective past or impf. stative instead.

Finally, although there are similarities between the egophoric marking in both languages, we will also zoom in on how and why the binary morphological system ('conjunct' vs 'disjunct') found in Newar differs so much from the Tibetan system, which sees egophoric markers integrated into a complex system of evidential and epistemic markers.

References

Hargreaves, D. (2015). "Am I Blue," in Egophoricity, eds., S. Floyd, E. Norcliffe, & L. San Roque. Amsterdam/Philadelphia: John Benjamins.

Hale, A. & K. Shrestha (2006). Newar (Nepāl Bhāṣā). Munich: Lincom Europa.

Meelen, Marieke & Ramble, Charles. 2022. An Audio-Visual Archive of Dzardzongke (South Mustang Tibetan). Endangered Languages Archive. Handle: http://hdl.handle.net/2196/aa07e8d9-de4a-4820-af20-a34054068b91 . Accessed on [15 Jan. 2024].

NAK (National Archives of Nepal) B 18-23, Acc. No. NAK 1/1583 (12th to 14th C.). Gopālarājavaṃśāvalī. Or. 10 Br. M. (18th C.). Batīsaputrikākathā. British Library, Oriental Collection (formerly at British Museum).

Otter, F. (2021). A Course in Reading Classical Newari: Selections from the Vetālapañcaviṃśati. Heidleberg: CrossAsia-eBooks.

Shrestha, S. S. (August 2023). Egophoricity & Formality in Kathmandu Newā(r). ACL SIGEL.

Egophoricity and Evidentiality as Distinct Categories in Golog Tibetan

Jiahong Wang & Lawrence Yam-Leung Cheung

(ETH Zürich & University of Hong Kong)

Whether egophoricity is a sub-category of evidentiality or an independent category remains a controversial issue. Garrett (2001) suggested egophoric markers in Lhasa Tibetan mark the speaker's knowledge as the source of information, thus subsuming egophoricity under evidentiality, just like direct evidentials marking direct perceptual evidence as the source. Further, egophoricity and evidentiality do not co-occur in Lhasa Tibetan (Garrett 2001) (same in Golog), making the analysis seem plausible. Nevertheless, this study argues they are distinct because they differ in distribution constraints and co-referential function.

I. Distribution constraints

1.1 Copular support

In Golog, egophoric ([EGO]) and non-egophoric ([N.EGO]) values are encoded in bound morphemes attached to verbs (Wang & Cheung 2023). These EGO/N.EGO morphemes impose a strict requirement on their hosting verbs, requiring copulas for phonological realization (1a & 1b). Constructions lacking copulas/copular auxiliaries (e.g. simple past, perfective, and modal constructions) never license egophoricity (2a & 2b). Conversely, evidential markers are free morphemes and can realize without a copular host (1c & 1d; 2c & 2d).

(1) *khir.sges*   *las.bya*   *'bri.gi.**yod***   (a)∅/(b)*ku*/(c)*thal*/(d)*zig*
     he.ERG   homework   write.IMPF   (a)EGO/(b)N.EGO/(c)EVI$_{sensory}$/(d)EVI$_{inferential}$

   'He's doing the homework.' (***yod*** is the existential copula.)

(2) *khir.sges*   *las.bya*   *bris.dang*   (a)*EGO/(b)*N.EGO/(c)*thal*/(d)*zig*
     he.ERG   homework   write.PFV   (a)*EGO/(b)*N.EGO/(c)EVI$_{sensory}$/(d)EVI$_{inferential}$

   'He has done the homework.'

1.2 Volition sensitivity and animacy sensitivity

In Golog, the choice of evidentials is purely pragmatically driven, reflecting the information source. Interestingly, the choice between EGO/N.EGO markers in Golog not only depends pragmatically on whether the *origo* (cf. Garrett 2001) has full knowledge of an event/state, but is also subject to non-pragmatic factors.

First, egophoricity in Golog exhibits sensitivity to VOLITIONALITY. EGO is prohibited if the subject is non-volitional. This constraint is grammatical and cannot be pragmatically accommodated. In (3a), even if the speaker has complete certainty of his own falling, EGO is impossible. Evidential markers are not subject to comparable constraint (3b).

(3) *nga*   *log.yod*   (a)*∅/(b)*thal*

    I.ABS   fall.PERF   (a)*EGO/(b)EVI$_{sensory}$ 'I've fallen.'

Second, egophoricity displays sensitivity to ANIMACY in factual constructions. In factual constructions with animate subjects, egophoricity strictly adheres to the *canonical distribution* (San Roque et al. 2018): EGO is used in first-person declarative and second-person interrogative sentences, while NON-EGO is used elsewhere (4). Conversely, factual constructions with non- animate subjects are more flexible. In (5), N.EGO emphasizes that the proposition is an objective fact, while EGO conveys the nuance that the origo considers it a fact without sufficient evidence. In contrast, evidential markers remain insensitive to animacy across all constructions.

(4) *nga'i*   *zhi.lus* *las.bya*   *bras*   *ni-red/*ni-yin*

    I.POSS   son.ERG homework   write.PAST   FACT-N.EGO/*FACT-EGO

    'My son wrote the homework.'

(5) *ltag.ga.na*   *char*   *'bab*   *gol*   *ni-red/ni-yin*

    outside   rain(N)   rain(V)   IMPF   FACT-N.EGO/FACT-EGO

    'It's raining outside.'

II. Co-referentiality in embedded clauses

In embedded clauses, egophoricity and evidentiality serve different functions. EGO unambiguously marks matrix subject-embedded subject co-referentiality, while N.EGO marks matrix subject-embedded subject non-co-referentiality. Conversely, evidentiality cannot indicate such (non-)co-referentiality. In (6), 'he' may refer to either *Tashi* or another individual.

(6) *bkra.shis.gis* *khir.sge* *log*   *thal*   *zer*

    Tashi.ERG   he   fall.PAST   EVI$_{sensory}$ say

    'Tashi$_i$ said he$_{i/j}$ fell.'

To conclude, egophoricity's sensitivity to grammaticality makes it different from evidentiality morpho-syntactically.

References

Garrett, E. J. (2001). *Evidentiality and assertion in Tibetan* (Doctoral dissertation)*.* Los Angeles: University of California.

San Roque, L., Floyd, S., & Norcliffe, E. (2018). Egophoricity: An introduction. In S. Floyd, E. Norcliffe, & L. San Roque (Eds.), *Egophoricity* (pp. 1–77). Amsterdam and Philadelphia: John Benjamins.

Tournadre, N., & LaPolla, R. (2014). Towards a new approach to evidentiality: issues and directions for research. *Linguistics of the Tibeto-Burman Area, 37*(2), 240–262.

Wang, J., & Cheung, L. Y. (2023 September). Morphosyntactic analysis of egophoricity in Golog Tibetan. Paper presented at 26th Himalayan Language Symposium (HLS-26), Paris: INALCO.

What if the 'ego-phoric' also *bears* upon the non-ego – and the non-'ego-phoric' on the ego?

Bettina Zeisler
(Universität Tübingen)

Keywords: Ladakhi, egophoric marking, pragmatic flexibility, territory of information

So-called 'egophoric' marking is part of the Tibetic attitudinal-evidential-epistemic systems. Although the *experiential* markers *byuṅ* and *myoṅ* are counted as 'egophoric', because they can only refer to the 'ego' or epistemic origo (speaker in statements, addressee in information-seeking questions), the crucial opposition is between non-experiential *yin* and *yod*, on the one hand, and all other markers, on the other.

    yin and *yod* are prototypically associated with the origo as (active) subject or possessor. What may appear – and be misinterpreted – as an odd person-marking system (cf. Aikhenvald 2004) is highly flexible to a degree not observed elsewhere. Tournadre & LaPolla (2014) describe it as "access to information (not just source), plus the subjective strategy or perspective of the speaker". I should define it as *representation of the subjective perspective* of the origo, based on her *knowledge access* and the socio-pragmatically conditioned *rights* to claim exclusive personal knowledge. The degree of flexibility varies among Tibetic languages, as does the use of specific markers. I shall illustrate this flexibility with data from Ladakhi.

    yin and *yod* are evidential *only in so far as* they indicate intimate acquaintance and exclusive personal knowledge and stand in opposition to other evidential markers, such as the visual and non-visual markers, (1). However, they also signal the origo's involvement as being related or responsible, (1), sometimes the origo's authoritative stance of indignation, (2). They may *refer* to non-ego subjects, while the origo may *refer* to herself or items in her personal sphere with non-'egophoric' markers, indicating non-identification, (3), non-responsibility, or non-relatedness, (5)–(6).

(1)   Kenhat-Ladakhi (Leh dialect)
    tharmos-inaŋa   ʧa daruŋ dug-a?   / rag-a?   –
    thermos.flask-PPOS tea still exist.VIS-QM   exist.NVIS-QM
    *duk   / rak   / **jod**   le.*
    exist.VIS   exist.NVIS   exist.'EGO' hon
    'Is there still tea in the thermos flask (**3P**, **VISUAL**: have a look / **NON-VISUAL**: feel or hear by shaking)?'
    – 'Yes, there is (**3P**, **VISUAL**: I had a look / **NON-VISUAL**: I shook it / '**EGO**': I know, **as I just filled it up**).'

(2)   Shamskat-Ladakhi (dialect of Eastern Sham)
    *l̥tos-aŋ!   tsamʃik   kha   rdaŋ-**et**!*
    look.IMP-DIR how.much mouth open.wide-'EGO'=PRS
    'Look, how [you/s/he] (**2P/3P!**) are/is yawning ('**EGO**': **it's annoying**)!'

(3)   Kenhat-Ladakhi (dialect of Central Zanskar)
    ŋaʒa   / **ŋa** *ɲalbaʧan **(ɦi)noʔ**.*
    we.excl   I   poor   be.**NON-'EGO'**
    '**We** are / **I** am just poor (other than you).' (**NON-'EGO'**: **feeling inferior**.)

The 'egophoric' markers are non-evidential *in so far as* they do not specify source or access of information and their use is pragmatically conditioned and/or independent of the origo's knowledge base, (4)–(6).

    With inclusive plural reference, 'egophoric' markers are avoided, since the speaker cannot claim privileged access:

(4) a. Kenhat-Ladakhi (dialect of Central Zanskar)

taʔsa	*haɣo*	*leha*	*ʧø-in-**joteno**ʔ;*		*habɣot*	*ma-taŋ!*
now	we.incl	work	do-CNT-**NON-'EGO'**=CNT.PRS	laughter	NG-give=PRHB

'**We** (<u>including you</u>) are working now (**NON-'EGO': you know as well as I**); stop laughing!'

b. Kenhat-Ladakhi (dialect of Central Zanskar)

taʔsa	*ŋaʒa*	*leha*	*ʧø-in-**jot**;*	*ϑiŋna*	*ʃoʔ!*
now	we.excl	work	do-CNT-'EGO'=CNT.PRS	afterwards	come.IMP

'**We** (<u>other than you</u>) are working now ('**EGO**': as you did't know.); [please] come later!'

While the knowledge base remains the same, Ladakhi speakers tune down their epistemic authority when pointing to the addressee's (non-)relationship to their own item as if lying beyond their personal sphere:

(5) Kenhat-Ladakhi (Dialect of Lalok/Changthang)

ʂiŋ	*ʂiŋ!*	*te*	*ʤola*	*ɲiri*	**mạndak.**	/	**mạn.*
wait=IMP	wait=IMP	that	bag	fam.you.GEN	NG.be.**NON-'EGO'**		*NG.be.'EGO'

ŋ˳e	**hin.**	/	**hindak.*
I+GEN	be.'**EGO**'		*be.NON-'EGO'

'Wait, wait! That bag (**1P-possession!**) *might* not be (**NON-'EGO'**) / *is definitely not yours. [It] (**1P-possession**) is ('**EGO'**) definitely / *might be mine.'

Similar shifts occur when contrasting items inside and outside one's personal sphere, depending on which one is focused as contrastee. The knowledge base is the same for each item and the difference as such:

(6) a. Kenhat-Ladakhi (dialect of Upper Indus)

CONTRASTEE		STANDARD				
çar˳e	*gonpa*	*lē*	*lākaŋ*	*som˳esaŋ*	*ʧhuŋ-a-ʑik*	**hot.**
Çara+GEN	monastery	Leh	temple	new+CNTR	be.small-NLS-LQ	be.'**EGO**'

'The Çara monastery (**3P**) is small in contrast to the New Temple of Leh.' ('**EGO**': Focusing on one's home monastery, **knowing it well**, **feeling related**, the other building was experienced only briefly.)

b. Kenhat-Ladakhi (dialect of Upper Indus)

CONTRASTEE		STANDARD				
lē	*lākaŋ*	*soma*	*ŋaʑ˳e*	*gonp˳esaŋ*	*ʧhe-a-ʑik*	**duk.**
Leh	temple	new	we.excl+GEN	monastery+CNTR	be.big-NLS-LQ	be.**VIS**

'The New Temple of Leh (**3P**) is large in contrast to our monastery.' (**VISUAL**: having **experienced the focused building only briefly**, **not feelig related**, while knowing one's monastery well.)

The Tibetic 'egophoric' markers *yin* and *yod* specify a proposition as belonging to the origo's subjectively evaluated and pragmatically conditioned *territory of information* (Kamio 1997). Propositions falling outside, take evidential or epistemic markers: NON-'EGO', (3)–(6), or VISUAL, (7).

The term 'egophoric', picked up crossliguistically, was coined prematurely for the Tibetic languages when subjective and pragmatic factors were little known. Redefined as, e.g., *ego-involved perspective*, it may still serve as a convenient label for most Tibetic languages.

References

Aikhenvald, Alexandra Yurievna (2004), *Evidentiality*. Oxford: Oxford University Press.

Kamio, Akio (1997), *Territory of information*. Amsterdam, Phil.: Benjamins.

Tournadre, Nicolas, and Randy LaPolla (2014), Towards a new approach to evidentiality. Issues and directions for research. *Linguistics of the Tibeto-Burman Area* 37(2), 240–262.

Posters

West Uralic: problems with taxonomy

Juha Kuokkala & Sampsa Holopainen
(University of Helsinki & University of Helsinki)

Keywords: Uralic languages, historical phonology, genealogical taxonomy

The internal structure of the Uralic language family is contested, and different kinds of taxonomical models have been presented (see Aikio 2022: 3–4). The idea that the three westernmost branches of the family (Saami, Finnic, Mordvin) form a genealogical node, presented by Häkkinen (2007) who called this node West Uralic, has received some acceptance in recent literature (Grünthal 2022, Lang 2020). However, there are actually very few arguments in favor of a West Uralic unity: Häkkinen's phonological arguments have since become mostly obsolete due to Aikio's (2012, 2015) studies on Uralic historical phonology and there seem to be little grounds for postulating common innovations in the field of lexicon (cf. Syrjänen et al. 2013: 337) or morphology (see Ylikoski 2016 for remarks on the West Uralic *s-cases, which however also have some kind of reflexes in Mari).

There are very few potential innovations that are common and specific to Saami, Finnic and Mordvin. There are possible Finno-Mordvinic, Finno-Saamic, as well as Saami-Mordvinic innovations, but it is uncertain whether any innovations can be reconstructed to a Finno-Saami-Mordvinic level. As the traditional Finno-Volgaic node (Finnic, Saami, Mordvin and Mari), largely rejected in modern studies, seems to be almost equally well grounded, the recent success of West Uralic as a taxonomical node appears premature.

In our talk we will concentrate on the putative innovations in the field of historical phonology, where some potential evidence remains and requires further scrutiny. Some innovations have been considered common to these three branches already in older literature. The innovations we will analyze are the following: the merger of PU *ẹ and *a in *a-stems; the emergence of *o in non-initial syllables (briefly argued by Zhivlov 2023: 121 to represent a West Uralic innovation: PU *waja- > WU *wajo- > Fi vajo- 'to sink'); the loss of word-initial *w- before labial vowels. We aim to determine whether the innovations in question can be shown to be independent of each other or whether some of them can be secondary, areally dispersed features. We analyze the possible relative chronology of sound changes and also pay attention to the typology of sound change to assess the possibility of parallel innovations instead of shared ones. In each case, it appears that the evidence is either scarce or the changes are typologically trivial. The merger of *ẹ and *a is also chronologically problematic as a West Uralic innovation, because it has been argued by Zhivlov (2014: 115–117, 2023: 127) that this is preceded by certain Saami-Mordvinic vowel changes.

Our research points to the conclusion that Proto-West Uralic cannot be reconstructed on the basis of phonological innovations. The unclarities in the reconstruction of West Uralic show that even in the best researched part of a well-studied language family there are disputed phonological innovations which are not sufficiently well understood and that strict methodology is in order to establish taxonomical groupings. This makes our presentation also interesting for scholars of genealogical taxonomy of languages and the historical linguistics more generally.

References

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2012. On Finnic long vowels, Samoyed vowel sequences, and Proto-Uralic *x. In Hyytiäinen, Tiina et al. (eds.), *Per Urales ad Orientem: Iter polyphonicum multilinguae: Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februar 2012* (Mémoires de la Société Finno-Ougrienne 264), 227–250. Helsinki.

Aikio, Ante. 2015. The Finnic 'secondary *e*-stems' and Proto-Uralic vocalism. *Journal de la Société Finno-Ougrienne* 95. 25–66. https://doi.org/10.33340/susa.82642

Aikio, Ante (Luobbal Sámmol Sámmol Ánte). 2022. Proto-Uralic. In Bakró-Nagy, Marianne & Laakso, Johanna & Skribnik, Elena (eds.), *The Oxford guide to the Uralic languages*, 3–27. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198767664.003.0001

Grünthal, Riho. 2022. The rise and dispersal of Western Uralic. (Presentation, Methodological perspectives on research into Proto-Uralic and West Uralic, University of Tartu, 14 June 2022.)

Häkkinen, Jaakko. 2007. *Kantauralin murteutuminen vokaalivastaavuuksien valossa*. Helsinki. (MA thesis.) http://urn.fi/URN:NBN:fi-fe20071746

Lang, Valter. 2020. *Homo fennicus: Itämerensuomalaisten etnohistoria*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Syrjänen, Kai & Honkola, Terhi & Korhonen, Kalle & Lehtinen, Jyri & Vesakoski, Outi & Wahlberg, Niklas. 2013. Shedding more light on language classification using basic vocabularies and phylogenetic methods. A case study of Uralic, *Diachronica* 30. 323–352. https://doi.org/10.1075/dia.30.3.02syr

Ylikoski, Jussi. 2016. The origins of the western Uralic *s*-cases revisited: historiographical, functional, typological and Samoyedic perspectives. *Finnisch-Ugrische Forschungen* 63. 6–78. https://doi.org/10.33339/fuf.86120

Zhivlov, Mikhail. 2014. Studies in Uralic vocalism III. *Journal of Language Relationship* 12. 113–148. https://doi.org/10.31826/jlr-2015-120109

Zhivlov, Mikhail. 2023. Reconstruction of Proto-Uralic. In Abondolo, Daniel & Valijärvi, Riitta-Liisa (eds.), *The Uralic languages*. 2nd edition. London: Routledge. https://doi.org/10.4324/9781315625096-3

Inflectional Asymmetries in Kagayanen

Thomas E. Payne & Carol J. Pebley
(University of Oregon; SIL International & SIL Philippines)

Kagayanen (CGC) is a typical Austronesian Philippine-type language spoken in the Central Philippines. In CGC, Grammatical transitivity and modality are the major dimensions expressed by paradigmatic verbal affixes. We name this group of affixes "inflectional" (following Wolff 1970). In this paper, we concentrate on one distinction in the system, which we identify as "general" vs. "limited" values. This distinction has relevance in transitive and intransitive, realis and irrealis verbal predicates. The general categories are more frequent in discourse, and express a wider range of meanings than the corresponding limited categories. Furthermore, the general categories for the most part allow the meanings associated with the limited ones, while the limited ones enforce them. For example, repetitive events may be expressed with the general inflections (1), while a certain kind of repetitive meaning is enforced by the limited expression of the same value (2). In other words, the general inflection, does not mean "non-repetitive," while the repetitive inflection specifically asserts repetition.

(1) ***Naka-iling*** *kay* *ta* *Cawili.*
     HAP.R-go   1pEXCL.ABS   NABS   Cawili
     'We have gone to Cawili.' (Once, or several times).
     'We were able to go to Cawili.'
(2) ***Ag-iling*** *kay* *ta* *Cawili.*
     HAP.REP.R-go 1pEXCL.ABS   NABS   Cawili
     'We have gone at times to Cawili.' (Several times)

Similarly, while abilitative semantics may be expressed by the general happenstantial forms (1), a prefix *ka-* zeros in on a specific type of abilitative sense, which we describe as "external motivation" or "enablement" (3).

(3) Dayon   a   ga-kamang ta   kaoy na sungkod   para **ka-panaw**   a.
     right.away 1sABS I.R-get   NABS   wood LK cane   for EXM-go/walk   1sABS
     'Then I got a wooden cane so that I could walk.'

In (3), the speaker had just survived an accident, and needed a wooden cane to enable him to walk. This example does not refer to the actor's innate ability to walk, but specifically the ability imparted by an external factor, in this case the cane. Like "repetitive" in example (2), this "external motivation" feature is not a dimension in the inflectional system, but is idiosyncratic to a particular form. Nevertheless, these "limited" values participate in paradigms with the other inflectional affixes. Therefore, inflection itself is not simply a matter of "slot filling" in a grid.

This phenomenon raises questions regarding the nature of inflectional morphology. Similar cases are among those mentioned by Haspelmath (2023) as contributing to his questioning of "inflection" and "derivation" as useful comparative concepts in linguistic analysis. Our conclusion is that, while inflection and derivation may not be well-defined universal comparative concepts, the Kagayanen system does exhibit most of the properties often claimed to hold for inflection in general (namely obligatoriness and paradigmaticity). These properties clearly distinguish "inflection-in-

Kagayanen" from another large group of morphological processes, which we term "stem-forming" processes.

Abbreviations: HAP.R = happenstantial, realis modality; EXCL = 1st person plural exclusive; ABS = absolutive case; NABS = non-absolutive case; HAP.REP.R = happenstantial, repetitive, realis; NEG.R = negative, realis; 2s = second person singular; INC = incompletive particle; EXM = external motivation modality.

References

Haspelmath, M. (2023), Inflection and derivation as traditional comparative concepts. *Linguistics*. https://doi.org/10.1515/ling-2022-0086.

Wolff, John U. (1970), The classification of Cebuano verbs. *Philippine journal of linguistics* 1(1), 74–91.

How dendrophilic is clause linkage?

Carlo Y. Meloni, Chundra Cathcart, Jessica Ivani, Erik Ringen & Balthasar Bickel
(University of Zurich, University of Zurich; NCCR; ISLE, University of Zurich, University of Zurich; LiRI, University of Zurich; NCCR; ISLE)

Clause linkage varies in the extent to which it forms hierarchical structures subject to syntactic constraints, such as constraints on co-reference, finiteness, or TAM scope. While meaning has been argued to explain some of this variation (Givón 1980; Van Valin et al. 1984; Cristofaro 2003; Van Valin 2005), an unresolved question is whether the evolution of clause linkage syntax is also driven by overall biases, above and beyond differences in meaning, individual constructional properties, and/or historical contingencies. Under one hypothesis, clause linkage syntax is overall driven by a neuro-cognitive bias towards (supra-regular) hierarchy ("dendrophilia", Fitch 2014), as known, for example, for noun phrase syntax in Indo-European (Widmer et al. 2017). Alternatively, clause linkage syntax is overall driven by a bias towards simplicity (Chater and Vitányi 2003), as known for neural computations more generally (Friston 2010).

Here, we test these opposing hypotheses by quantifying hierarchy in a given clause linkage construction through the syntactic constraints it is subject to. Exhaustively coding constructions for the presence/absence of 18 constraints in 59 languages from Indo-European, Sino-Tibetan, and Tupi-Guarani (with resolved phylogenies from Heggarty et al. 2023, Sagart et al. 2019, and Ferraz Gerardi et al. 2023 respectively), we first use Bayesian two-state Continous-Time Markov (CTM) models on three selected constructions in each language ('and'-coordination, 'because'-subordination, and transitive-subject relative clause constructions) to show that the stationary probabilities of the constraints vary decisively more between families than between constructions, consistent with earlier findings that challenged the cross-linguistic validity of broad constructional distinctions (Bickel 2010). This justifies, in a second step, pooling all constructions in a single evolutionary model for each family. This model captures how the overall probability of constraints (and, therefore, syntactic hierarchy) has most likely evolved, allowing for random variations within constraints and languages through a hierarchical ("mixed-effects") design. We find that in all three phylogenies, the best-fitting model is one where the (logistically transformed) probabilities evolve according to an Ornstein-Uhlenbeck process (Butler and A. King 2004), with a long-term attraction towards probabilities near 0.3 (posterior median $\theta_{IE} = 0.27$, $\theta_{ST} = 0.34$, $\theta_{TG} = 0.37$) and, outside Indo-European, with relatively strong variance ($\sigma_{IE} = 0.18$, $\sigma_{ST} = 0.72$, $\sigma_{TG} = 0.37$).

We conclude that the evolution of clause linkage syntax is not driven by dendrophilia but by a moderate bias toward simplicity (fewer constraints) and substantial lineage-specific variation.

References

Bickel, Balthasar (2010). Capturing particulars and universals in clause linkage. In: *Clause Linking and Clause Hierarchy: Syntax and Pragmatics*. Amsterdam: Benjamins, pp. 51–101.

Butler, Marguerite and Aaron King (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. In: *The American Naturalist* 164.6, pp. 683–695.

Chater, Nick and Paul Vitányii (2003). Simplicity: a unifying principle in cognitive science? In: *Trends in cognitive sciences* 7.1, pp. 19–22.

Cristofaro, Sonia (2003). *Subordination*. OUP Oxford.

Ferraz Gerardi, Fabricio et al. (2023). Lexical phylogenetics of the Tupi-Guarani family: Language, archaeology, and the problem of chronology. In: *Plos one* 18.6, e0272226.

Fitch, W Tecumseh (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. In: *Physics of life reviews* 11.3, pp. 329–364.

Friston, Karl (2010). The free-energy principle: a unified brain theory? In: *Nature Reviews Neuroscience* 11.2, pp. 127–138.

Givón, Talmy (1980). The binding hierarchy and the typology of complements. In: *Studies in Language* 4.3, pp. 333–377.

Heggarty, Paul et al. (2023). Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. In: *Science* 381.6656, eabg0818.

Sagart, Laurent et al. (2019). Dated language phylogenies shed light on the ancestry of Sino-Tibetan. In: *Proceedings of the National Academy of Sciences* 116.21, pp. 10317–10322.

Van Valin, Robert D et al. (1984). A typology of syntactic relations in clause linkage. In: *Annual meeting of the Berkeley Linguistics Society*. Vol. 10, pp. 542–558.

Van Valin, Robert D (2005). *Exploring the syntax-semantics interface*. Cambridge University Press.

Widmer, Manuel et al. (2017). NP recursion over time: Evidence from Indo-European. In: *Language*, pp. 799–826.

Expansion of differential object marking beyond animacy in Early Modern Ukrainian and Belarusian (Ruthenian)

Dmitri Sitchinava
(Potsdam University)

Slavic languages are known to feature an animacy-driven differential object marking (DOM, cf. Witzlack-Makarevich, Seržant 2018 on foci of variation) system where animate objects are marked by a genitive form (see Eckhoff 2015 on an early stage of this difference in Slavic). There is a more recent development in some "northern" Slavic languages to include into this system also some inanimate object types, such as nouns referring to plants, mushrooms, small objects, body parts or abstract notions:

(1) Ukrainian

Ja	vyby-v	zub-a.
1SG	knock.out-PST.M	tooth-GEN/ACC.SG

'I knocked a tooth out.'

(2) Polish

Widzę	papiros-a.
see.1SG	cigarette-GEN/ACC.SG

'I see a cigarette.'

The phenomenon is widespread in the northern Slavic area, including, to different extent, modern Ukrainian, Belarusian, Polish, Czech, Slovak, and Sorbian (Skwarska 2018). In many languages, including Polish and Ukrainian, this pattern is productive (Kosek 2022; Shvedova 2018). In Russian the genitive marking is restricted to "physically or functionally anthropomorphic or zoomorphic objects" (Krys'ko 1994).

Using the medieval and Early Modern Ukrainian/Belarusian texts (a written language known as *prosta mova* or Ruthenian) I research in what extent the expansion of DOM in the East Slavic lects depends on inherent and non-inherent object properties. The research questions concern diachronical sources of this expansion in East Slavic, the types of contexts where the expansion is attested first, and typological context of the change. There is no readily available corpus of Ruthenian. Editions of legal texts from the Belarusian territory (such as the Lithuanian statute) and from the Ukrainian territory (documents of different city archives, the Monuments of Ukrainian language series) were digitized and analyzed.

Two classes of DOM instances with *o*-masculina names are prominent in Ruthenian texts: borrowed names for artifacts such as *hak* 'arquebuse' or *taljar* 'Taler, dollar' and 'landmarks' such as *dub* 'oak'. The landmark contexts are marked by specificity and high prominence that is relevant in many versions of the Extended animacy hierarchy (cf. Croft 2003). Borrowed names for artifacts in the Ruthenian contexts behave as a kind of proper names (cf. *mercedes* 'Mercedes' and other trademarks as DOM-triggering semantic class of its own right in Polish, as well as recent borrowings from Polish to English like *smartfon* 'smartphone'). Proper names were described as an element of the Extended animacy hierarchy by Croft (2003: 10), but proposed for removal as typologically irrelevant (Helmbrecht et al. 2018); they feature prominently as a DOM parameter in Romance (Reina 2020). Individual animal names are a relevant DOM parameter in medieval East Slavic, especially in the Old Novgorod dialect (Gippius 2020).

In García García 2018 it is claimed that in Spanish animate-like marking of inanimate objects is fully predicate-triggered. Modern Ukrainian (and to a lesser extent Belarusian) retained more classes of verbs with Genitive object (or with Gen/Acc variativity) than Russian. Our finding is that for some argument classes like body parts, featuring in Modern Ukrainian, the animate-like marking was originally predicate-triggered (cf. Ruthenian legalese *urizaty vux-a* 'cut (from) ear' with partitive genitive).

References

Croft, William (2003) Typology and universals. 2nd edn. Cambridge: Cambridge University Press Gippius, Alexei A. (2020) Berestjanye gramoty iz raskopok 2019 goda v Novgorode i Staroj Russe. Voprosy jazykoznanija, No 5 (2020), 22-37.

Eckhoff, Hanne Martine (2015): "Animacy and differential object marking in Old Church Slavonic". Russian Linguistics 39:2.

García García, Marco (2018). Nominal and verbal parameters in the diachrony of differential object marking in Spanish. In: Seržant, Ilja A. & Witzlack-Makarevich, Alena (eds.) Diachrony of differential argument marking. (Studies in Diversity Linguistics 19), 209-242. Berlin: Language Science Press.

Helmbrecht, Johannes, Lukas Denk, Sarah Thanner, Ilenia Tonetti (2018). Morphosyntactic coding of proper names and its implications for the Animacy Hierarchy. In: Sonia Cristofaro. Fernando Zúñiga (eds.) Typological Hierarchies in Synchrony and Diachrony, 377-402/ Amsterdam: Benjamins.

Kosek, I. Rodzaj męskożywotny a typ paradygmatu. O tendencjach rozwojowych we współczesnej deklinacji polskiej (na przykładzie anglicyzmów). Język polski, Rocznik CII, z. 1 (2022), 57-69 Laca, Brenda. 2006. El objeto directo. La marcación preposicional. In Concepción Company Company (ed.), Sintaxis historica del español. Primera parte: La frase verbal, vol. 1, 423–475. Mexico: Universidad Nacional Autónoma de México.

Reina, Javier Carlos (2020). Differential object marking with proper names in Romance languages. In: Luise Kempf, Damaris Nübling & Mirjam Schmuck (eds.), Linguistik der Eigennamen, 225-258. Berlin: de Gruyter.

Shvedova, Maria (2018). Variatyvnist' form imennykiv u znakhidnomu vidminku odnyny v ukrajins'kij movi, Studia philologica [Kyiv]. Vol. 10, 19-28.

Skwarska, K. „Zapnul už jsi toho mobila?" O akuzativu singuláru neživotných maskulin ve slovanských jazycích. Slavistika. Roč. 22, č. 1 (2018), s. 166-177

Witzlack-Makarevich, Alena, Ilja A. Seržant (2018). Differential argument marking: Patterns of variation. In: Seržant, Ilja A. & Witzlack-Makarevich, Alena (eds.) Diachrony of differential argument marking. (Studies in Diversity Linguistics 19), 1-40. Berlin: Language Science Press.

The category of causation in West Minyag (Qiangic)

Tamdrin Lhamo & Zoe Tribur
(Nankai University & Nankai University)

Keywords: Qiangic, Causitivization, Historical linguistics, morphological typology

This paper describes three types of formal mechanisms of causation in West Minyag, a Qiangic (Sino-Tibetan) language spoken in western Sichuan, China. These three mechanisms are lexical causatives, morphological causatives and analytic causative. This study contributes to our understanding of the development of causation as an important valence-changing process in Qiangic. Although individual languages like Minyag have diverse formal ways of expressing causation, it is clear that many Qiangic causatives are quite old with reflexes found in even very distant Sino-Tibetan languages (Sun 1999).

This paper also serves as a case study for exploring certain typological claims made about causation. Among these, we find that Minyag conforms to Comrie's claim (1985:342) regarding a general tendency of the causee argument in a causative clause to occupy the leftmost or highest unoccupied position in the following hierarchy: the subject—direct object—indirect object. When intransitive verbs are causativized, in the resulting clause the patient slot is unoccupied so the causee is relegated to patient. In other clauses, the patient slot is occupied, so the causee with its dative marker acts as a non-core argument.

There are two kinds of lexical causative: (a) single lexemes express both causative and non-causative meaning: (b) causative and non-causative meanings are expressed by different verbs.

Inflectional morphology is a common way to express causation in Minyag. For intransitive verbs, we find three types of morphological causative: (a) Prefix vowel fronts to i or e; (b) Initial consonants of the verb root change from voiced to voiceless; (c) both changes are combined. Morphological causation is relatively rare in transitive verbs, being restricted to a few action verbs and cognition verbs such as 'drink', 'wear', 'carry', 'hold', 'understand', etc. Among these, 'drink' and 'wear' are subdivided into different subjunctive semantics.

We find two cases of analytic causative: (a) verbalizing an adjective with nɵ-və 'make'; (b) adding the verb tɕhə 'let' to the VP and having it take all person agreement marking, etc.. This second type is the most productive causative expression in Minyag.

Examples of the above-described causative types are given below.

Type	non-causative	causative	non-causative	causative
Lexical causative A	*i-kɵ* 'flood'	*i-kɵ* 'make flood'	*i-ɣə* 'become wide'	*i-ɣə* 'make wide'
Lexical causative B	*thɵ-sə* 'die'	*nɵ-sɵ* 'kill'	*khə-tɕɵʐi* 'look'	*ngi-ti* 'show'
Morphological causative A	*na-ndʑa* 'rain'	*ne-ndʑa* 'make rain'	*na-ngə* 'break''	*ne-ngə* 'break'
Morphological causative B	*khə-nə* 'drink milk'	*khi-nə* 'feed milk'	*tə-ngə* 'wear clothes'	*ti-ngə* 'put on clothes'
Analytic causative A	*tʂhɵtʂhø* 'white'	*tʂhɵtʂhø nɵvə* 'make white'	*sivə* 'good'	*sivə nɵvə* 'make good'
Analytic causative B	*thɵ-sə* 'die'	*thɵ-sə tɕhə* 'make die'	*tə-ngə* 'wear clothes'	*tə-ngə tɕhə* 'make wear clothes'

References

Sun, Hongkai. 1999. The category of causative verbs in Tibeto-Burman languages. Linguistics of the Tibeto-Burman Area. 22(1), 183-199.

Comrie, Bernard. 1985. Causative verb formation and other verb-deriving morphology. Language typology and syntactic description，III. ed. by Timothy Shopen. Cambridge:Cambridge University Press. 309–348.

Dixon，R.M.W.2012.Basic Linguistic Theory(Vol. III). Oxford: Oxford University Press. 239-293.

Payne, Thomas E.1997. Describing Morphosyntax:A guide for field linguists. Cambridge University Press.169-222.

"It's Just a Joke!":
Critical Discourse Analysis on the Use of Humour in Racist Discourse

Ellen Verenieks
(Aston University)

There has been extensive research into the ways in which power is routinely exercised through language, but little that has explored the relationship between hatred and humour.

This project takes a step to close the gap in research thus far by presenting an analysis of the use of humour in racist discourse, comparing 'joke' pages on white-nationalist websites.[1]

Previously, Billig (2001) looked at three KKK-linked US sites, examining how humour is used in order to propagate hatred. The white nationalist type of humour is extreme and displays somewhat different characteristics from other types of humour (Billig, 2001): for example, the rhetoric of taboo is missing, signalling a certain enjoyment. The integral links between the extreme hatred and the dehumanising strategies on display in these violent jokes (Simpson & Mayr, 2010) require a distinct analysis, more focused than previously seen.

The primary research questions:
1. What are the lexical choices used in humour in racist discourse?
2. What are the parallels or differences between these features on similar US sites over two decades?

For this analysis of the use of humour in racist discourse, Critical Discourse Analysis (CDA) was the preferred methodology, as it reflects the role discourse plays in society. CDA is used to foreground aspects of ideology and power, and examines the ways in which discourse reinforces and reproduces power relationships (Simpson & Mayr, 2010). In order to mitigate subjectivity from the perspective of the analysist, corpus-assisted CDA was used in looking for recurrent linguistic
patterns in hate 'humour'.

Data was pulled from two white nationalist/KKK-affiliated websites originally flagged by Billig in his 2001 study, as well as by various hate-watch groups: WhitePower.com and WhitesOnly.net/Tightrope.cc. For the purposes of this research, 136 joke entries were pulled from each: WhitePower from 1997 and WhitesOnly/Tightrope from 2017. The racist and race-emphasis word frequency was used to refine search parameters within the preliminary Sketch Engine search. This provided an important foundation of patterns and concordance, as linguistic repetition reinforces the imposed image of the target and the ideology claimed by the joke-teller (Simpson & Mayr, 2010).

Lexical choices and patterns were consistent across the two datasets, showing little difference in the rhetoric used. There was marked use of modifiers alongside racist terminology in order to associate meanings: for example, modifiers referencing violence such as *kill*, *whip*, *shoot*.

---

[1] The use of "humour" and "joke" does not reflect any aesthetic judgement nor that these jokes are funny (Billig, 2001). The claim to humour is made by the joke-tellers themselves.

The full findings suggest that the use of racist humour creates a reality in getting an agenda out into the wider societal sphere, and this is accomplished on these white-nationalist sites through soft power: 'control' or influence over minds of others while in a comfortable and jovial state. Recurrent linguistic patterns serve to intensify the meaning of the select words, with this repetition unifying those who participate in the joke. The rhetoric throughout is motivated by the intent to influence and mould a particular line of thinking or ideology.

A better understanding of the framework of 'joking' in racist discourse allows for improved tools to identify and prevent this more subversive promotion of ideology.

References

Billig, M. (2001) "Humour and hatred: the racist jokes of the Ku Klux Klan," *Discourse & Society*, 12(3), pp. 267–289. Available at: https://doi.org/10.1177/0957926501012003001.

Billig, M. (2005) *Laughter and ridicule: towards a social critique of humour*. London: Sage Publications.

Davies, C. (1990) *Ethnic humour around the world*. Bloomington, IN: Indiana University Press. Fairclough, N. (1992) *Discourse and social change*. London: Polity Press.

McFarland, K. (2006) *Strengthening critical discourse analysis: The Baby Book revisited*. Queen's University, Belfast. Available at: https://s3-eu-west-1.amazonaws.com/s3-euw1-ap-pe-ws4-cws-documents.ri-prod/9781138569232/KateMcFarlandArticle.pdf.

Simpson, P. and Mayr, A. (2010) *Language and power*. London: Routledge.

van Dijk, T.A. (1992) "Discourse and the denial of racism," *Discourse & Society*, 3(1), pp. 87–118. Available at: https://doi.org/10.1177/0957926592003001005.

van Dijk, T.A. (1993) *Elite discourse and racism*. Newbury Park, CA: Sage Publications.

Weaver, S. (2016) *The rhetoric of racist humour: US, UK and global race joking*. Oxon: Routledge.

The Mountains of Affixes: A phonological and typological approach to affixation in the Caucasus

Filip Larsson
(Lund University)

Introduction: The Caucasus contains some of the world's most elaborate affixation patterns (Daniel & Ganenkov 2009: 671), which is paired with some of the world's largest consonant inventories (Beguš 2020: 699). The languages of the Caucasus therefore constitute an interesting case to examine correlations between morphology and phonology. Bybee (2005) conducted a similar study of only TAM affixes in 23 languages from around the world, but it only included Abkhaz from the Caucasus. Is there thus a significant correlation between the number of consonant phonemes and morphological complexity in the Caucasus? Which grammatical functions are expressed by affixation across the Caucasus?

Data: The study is based on a dataset of almost 11,000 nominal and verbal affixes collected from grammatical descriptions of 56 languages from five language families spoken in the Caucasus, i.e. Kartvelian, Nakh-Dagestanian, Northwest Caucasian, Turkic and Indo-European. The affixes were coded for grammatical function and phonological form. The phonological form of each affix was coded according to its syllable structure, and thereafter its consonants and vowels. The grammatical functions were divided into 23 grammatical variables, i.e. case, TAM, preverbs, converbs, etc.

Results: The phonological results shows a significant positive correlation between phoneme inventory and number of grammatical functions expressed by affixation in the Caucasus (Pearson's $r(54) = .61$ $p < .001$), which corroborates similar results found world-wide for syllable complexity by Easterday et al (2021). A Wilcoxon sign-rank test also indicated a significant difference between the number of consonant phonemes in the affix data (Mdn = 23) and the full consonant inventories for each language (Mdn = 38.5), $z = 1.46$, $p < .001$, which partly contradicts the previous results by Bybee (2005).

The most interesting typological finds relate to preverbs and converbs. Spatial preverbs are found in all language families of the Caucasus except Turkic (Boeder 2005, Ganenkov & Maisak 2020, Arkadiev & Lander 2020, Belyaev 2020, Johanson & Csató 2022), and the Nakh-Dagestanian languages share numerous preverb orientations with the Northwest Caucasian languages (Kumaxov 2013, Fenwick 2011, Rogava & Keraševa 1966, Nichols 2011, Čokaev 1970, Sumbatova 2020, Haspelmath 1993, Authier 2009, Magometov 1970). Kartvelian preverbs generally express deictic functions while the Megrelian and Laz spatial preverb inventories are almost as complex as the Northwest Caucasian systems (Boeder 2005, Reseck 2015, Öztürk & Pöchtrager 2011).

Converbs constitute a category of non-finite verb forms whose main function is to mark adverbial subordination (Haspelmath 1995: 3) and they are found in all Nakh-Dagestanian, Turkic and Northwest Caucasian languages of the Caucasus (Ganenkov & Maisak 2020, Johanson & Csató 2022, Hewitt 2010, Kumaxov 2013, Fenwick 2011). Converb-like constructions are also present in Laz (Lacroix 2009) and

Juhuri (Authier 2012), and 14 specialized converbs were generally found regardless of language family. The widespread presence of immediate anterior converbs is surprising since an immediate posterior converb ('just before') has only been found in Northern Akhvakh (Creissels 2010: 128) and Chechen (Yakovlev 1940: 268).

References

Arkadiev, Peter & Lander, Yury (2020), 'The Northwest Caucasian Languages', Maria Polinsky (ed.), The Oxford Handbook of Languages of the Caucasus. Oxford: Oxford University Press, 369-446.

Authier, Gilles (2009), Grammaire kryz: Langue caucasique d'Azerbaïdjan, dialecte d'Alik. Collection linguistique de la Société de linguistique de Paris. Paris: Peeters.

Authier, Gilles (2012), Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est. Wiesbaden: Reichert.

Beguš, Gašper (2020), 'Segmental Phonetics and Phonology in Caucasian Languages', in Maria Polinsky (ed.) The Oxford Handbook of Languages of the Caucasus. Oxford: Oxford University Press, 689-728.

Belyaev, Oleg (2020), 'Indo-European languages of the Caucasus', Maria Polinsky (ed.), The Oxford Handbook of Languages of the Caucasus. Oxford: Oxford University Press, 574-639.

Boeder, Winfried (2005), 'The South Caucasian Languages'. Lingua, volume 115, Issues 1–2, 5-89.

Bybee, Joan (2005). 'Restrictions on phonemes in affixes: A crosslinguistic test of a popular hypothesis'. Linguistic Typology, 9, 165–222.

Čokaev, K. Z. (1970). Affikatsiya. Morfologiya Čečenskogo Jazyka, II. Groznyj: Čečeno-Ingušskoe Knižnoe Izdatek'stvo.

Creissels, Denis (2010). 'Specialized converbs and adverbial subordination in Axaxdərə Akhvakh', in I. Bril (ed.) Clause linking and clause hierarchy: syntax and pragmatics. John Benjamins, 105-142.

Daniel, Michael & Ganenkov, Dmitry (2009). 'Case Marking in Daghestanian', in Andrej Malchukov & Andrew Spencer (eds.), The Oxford Handbook of Case. Oxford: Oxford University Press, 668-685.

Easterday, Shelece, Stave, Matthew, Allassonnière-Tang, Marc & Seifart, Frank (2021). 'Syllable Complexity and Morphological Synthesis: A Well-Motivated Positive Complexity Correlation Across Subdomains'. Frontiers in Psychology, 12. doi: 10.3389/fpsyg.2021.638659

Fenwick, Rohan S.H. (2011). A Grammar of Ubykh. LINCOM Studies in Caucasian Linguistics. München: LINCOM Europa.

Ganenkov, Dmitry & Maisak, Timur (2020). 'Nakh-Dagestanian Languages', in Maria Polinsky (ed.) The Oxford Handbook of Languages of the Caucasus. Oxford: Oxford University Press, 87-146.

Haspelmath, Martin (1993). A Grammar of Lezgian. Berlin: Mouton de Gruyter.

Haspelmath, Martin (1995). 'The converb as a cross-linguistically valid category', Martin Haspelmath & Ekkehard König (eds.), Converbs in cross-linguistic perspective: structure and meaning of adverbial verb forms – adverbial participles, gerunds. Berlin: Mouton de Gruyter, 1-56.

Hewitt, B. G. (2010). Abkhaz: A Comprehensive Self-tutor. München: LINCOM Europa.

Johanson, Lars & Csató, Éva Ágnes (eds.) (2022). The Turkic languages. 2nd ed. London: Routledge.

Kumaxov, M. A. (2013). Кабардино-черкесский язык. Moskva: Inst. Jazykoznanija, RAN.

Lacroix, René (2009). Description du dialecte laze d'Arhavi (caucasique du sud, Turquie): Grammaire et textes. Diss. Lyon: Université Lumière Lyon 2.

Magometov, Aleksandr A. (1970). Agul'skij jazyk: Issledovanija i teksty. Tbilisi: Izdatel'stvo "Mecniereba".

Nichols, Johanna (2011). Ingush Grammar. Berkeley: University of California Press.

Öztürk, Balkız & Pöchtrager, Markus A. (eds.) (2011). Pazar Laz. Languages of the World/Materials, 484. München: LINCOM Europa.

Reseck, Tamar (2015). Präverbien im Megrelischen. Diversitas Linguarum, vol. 37. Bochum: Universitätsverlag Dr. N. Brockmeyer.

Rogava, G. V. & Keraševa, Z. I. (1966). Grammatika adygejskogo jazyka. Majkop: Adygejskoje otdelenije Krasnodarskoje Knižnoje Izdatel'stvo.

Sumbatova, Nina (2020). 'Dargwa', Maria Polinsky (ed.), Handbook of the Languages of the Caucasus. Oxford: Oxford University Press, 147-200.

Yakovlev, N. F. (1940). Sintaksis Čečenskogo literaturnogo jazyka. Moskva: Akademia Nauk SSSR.

Do mosquitoes hibernate at all? The particle *üldse* 'at all' in multifunctional interrogatives in Estonian everyday interaction

Andriela Rääbis
(University of Tartu)

The paper aims to examine the use of the Estonian particle *üldse* 'at all; ever; absolutely (not)' in interrogatives that convey several social actions simultaneously. The particle *üldse* and its counterparts in other languages (e.g. *any*, *ever*, *at all* in English, *ollenkaan*, *lainkaan*, *yhtään* in Finnish) have been described as negative polarity markers (Metslang 2017: 523; Heritage 2010: 51; Heritage & Raymond 2021: 42; Couper-Kuhlen et al. 2023; VISK § 1634, § 1641, § 1722) and as intensifiers and extreme case formulations (Erelt 2017: 185–186; Hennoste 2000: 1805–1806; Rääbis et al. 2019; Pomerantz 1986; Edwards 2000). The usage of *üldse* has not been previously studied in spoken Estonian.

Our research questions are:
1. What are the functions of the particle *üldse* in interrogatives?
2. In which syntactic type of interrogatives does *üldse* occur and which social actions are conveyed by these interrogatives?

The data comes from the Corpus of Spoken Estonian of the University of Tartu and consists of 360 everyday face-to-face and telephone conversations. The particle *üldse* occurred in 122 interrogatives: 54 information-seeking questions (e. g. *millega te `tulite üldse*. 'How did you come at all?') and 68 multifunctional interrogatives, examined in this study. The analysis draws on the methodological framework of interactional linguistics.

The analyzed interrogatives perform several social actions at the same time (see also Laanesoo 2018). The question form allows the recipient to treat them as questions. In addition, they are used to convey doubts, challenges, statements, judgements and reproaches. The primary function of any given interrogative is made clear by the reaction of the interlocutor, whether he/she responds by providing information, justifying his/her position or doesn't react to the question.

Syntactically, there were two types of interrogatives in our data: 24 content questions and 44 polar questions; 47 in affirmative and 21 in negative form. In affirmative questions, *üldse* functions as a negative polarity marker, suggesting an expectation for a negative response (on reversed polarity questions, see Koshik 2005). For example, the question *õpid sa `üldse seal koolis midagi=ve.* 'Do you learn anything at all in that school?' expresses the opinion that the interlocutor is not learning anything. In negative questions, *üldse* emphasizes the extreme end of the scale and is used to highlight someone's norm-violating behavior, e.g. the interrogative *sa ei=ole Viljandis `üldse tapeeti vaatamas käind=vä.* 'Haven't you looked for the wallpaper in Viljandi at all?' is used for reproaching.

The primary context for the particle *üldse* is reacting to some non-preferred, unexpected, or surprising information, simultaneously expressing one's attitude and asking for more information. *üldse* is

affective, increases intensity and adds evaluativeness. To sum up, *üldse* is commonly backward-looking and its usage is often associated with non-preference, disagreement, or contradiction.

References

Couper-Kuhlen, Elizabeth, Sandra A. Thompson, and Barbara A. Fox (2023), Do English affirmative polar interrogatives with *any* favor negative responses? In G. B. Bolden, J. Heritage, and M.-L. Sorjonen (eds), *Responding to polar questions across languages and contexts* (Studies in Language and Social Interaction 35). Amsterdam, Philadelphia: John Benjamins, 350–376.

Edwards, Derek (2000), Extreme case formulations: softeners, investment, and doing nonliteral. *Research on Language and Social Interaction* 33(4), 347–373.

Erelt, Mati (2017), Öeldis [Predicate]. In M. Erelt, and H. Metslang (eds), *Eesti keele süntaks* [Syntax of the Estonian language] (Eesti keele varamu III). Tartu: Tartu Ülikooli Kirjastus, 93–239.

Hennoste, Tiit (2000), Sissejuhatus suulisesse eesti keelde IV. Suulise kõne erisõnavara 3. Partiklid [Introduction to Spoken Estonian IV. Vocabulary of Spoken Estonian 3. Particles]. *Akadeemia* 8, 1773–1806.

Heritage, John (2010), Questioning in medicine. In A. F. Freed, and S. Ehrlich (eds), *Why do you ask? The function of questions in institutional discourse*. New York, Oxford: Oxford University Press, 42–68.

Heritage, John, and Chase Wesley Raymond (2021), Preference and polarity: Epistemic stance in question design. *Research on Language and Social Interaction* 54(1), 39–59.

Koshik, Irene (2005), *Beyond rhetorical questions: Assertive questions in everyday interaction*. Amsterdam, Philadelphia: John Benjamins.

Laanesoo, Kirsi (2018), *Polüfunktsionaalsed küsilaused eesti argivestluses* [Multifunctional interrogatives in Estonian everyday interaction] (Dissertationes linguisticae Universitatis Tartuensis 33). Tartu: Tartu Ülikooli Kirjastus.

Metslang, Helle (2017), Kommunikatiivsed lausetüübid [Communicative types of sentences]. In M. Erelt, and H. Metslang (eds), *Eesti keele süntaks* [Syntax of the Estonian language] (Eesti keele varamu III). Tartu: Tartu Ülikooli Kirjastus, 515–536.

Pomerantz, Anita (1986), Extreme case formulations: A way of legitimizing claims. *Human Studies* 9(2/3), 219–229.

Rääbis, Andriela, Tiit Hennoste, Andra Rumm, and Kirsi Laanesoo (2019), *They are so stupid, so stupid*. Emotional affect in Estonian school-related complaints. *Journal of Pragmatics* 153, 20–33.

VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, and Irja Alho (2004), *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio, 1.11.2008. http://scripta.kotus.fi/visk

(Anti-)defective phenomena in Greek toponymic inflection

Michail Marinis

(University of Ioannina)

Some recent studies, mainly in morphosyntax, argue that toponyms' grammatical status differs from common names. For example, according to Stolz & Levkovych (2019: 45), the most common cross-linguistic difference between toponyms and common names is that toponyms involved in spatial-adverbial affixes do not always have a preposition in languages for which in corresponding structures, the preposition is obligatory for all other word classes. Therefore, Stolz, Levkovych & Urdze (2017) refer to the need for a *Special Toponymic Grammar* for which "currently our knowledge is very limited."

This paper proposes arguments supporting the idea that toponyms do have a special grammatical status that, interestingly, is not restricted to the level of morphosyntax. The analysis is based on the hitherto unexplored area of the paradigmatic inflectional morphology of toponyms. It examines synchronic and diachronic data from Greek since the language, due to its rich morphology, lends itself to the study of paradigmatic inflection.

The analysis shows that the inflection morphology of toponyms differs from common nouns and other categories of proper names, like anthroponyms. In Table 1, the comparison of the inflectional paradigms of the common name ˈ*furnos* 'oven', of the homonymous name ˈ*furni* 'the name of a Greek island', of the name ˈ*korinthos* 'Korinthos', and the anthroponym ˈ*janis* 'John', shows that a critical property of toponyms is that they have a mandatory tantum inflectional paradigm. Diachronic data also reinforce this observation. For example, the present capital of Greece has presented a fascinating circular history from the time of Homer to the present. Today it is a well-known singularia tantum, while there was a well-thought-out pluralia tantum case in the classical era. However, example (1) shows that the name probably started as a singularia tantum (*Ἀθῆνα*), became pluralia tantum (*Ἀθῆναι*), and again Singularia Tantum (*Ἀθήνα*), maintaining its compulsory tantum substance (Table 2).

(1)   […] *ἵκετο δ᾽ ἐς Μαραθῶνα καὶ εὐρυάγυιαν **Ἀθήνην*** […]
        'He reached Marathon and broad-road Athens'.   Homer's Odyssey 7.80

In addition, the inflectional paradigm of the Greek toponyms presents an impressive phenomenon, for which I propose the term *anti-defectiveness*. The Greek words in *-aki* always have an inflectional paradigm defective for the cell of the genitive. For example, the type **kolona*ˈ*ciou* 'little pilar.gen.sing.' is ungrammatical. However, the homonymous toponym does not present this restraint, restoring the initially defective inflectional paradigm. Thus, the genitive *kolona*ˈ*ciu* 'a famous area in the center of Athens' is grammatical (2).

(2)   ˈ*ta*   ˈ*steca*   ˈ*tu*   ***kolona*ˈ*ciu***
        the.NOM.PL.   haunt.NOM.PL.   the.GEN.SING.   Kolonaki.GEN.SING.
        'the haunts of Kolonaki'

The present study supports the need for a Special Toponym Grammar. It shows that additionally to morphosyntax, the toponyms are differentiated at the level of paradigmatic inflectional morphology. Crucially, their status differs not only from that of (homonymous) common names but also from that of other proper names, such as anthroponyms. <u>Specifically, the toponyms</u>:

    i. must have a tantum inflectional paradigm (either pluralia or singularia)
    ii. it is possible for them to exhibit anti-defective phenomena when their inflectional paradigm is compared with that of homonymous common nouns.

References

Stolz, Thomas & Nataliya Levkovych. (2019). Toponomastics meets linguistic typology: Glimpses of Special Toponymic Grammar from Aromanian and sundry languages. *Onomastica Uralica (11)*, 43-61.

Stolz, Thomas, Levkovych, Nataliya & Urdze, Aina. (2017). When zero is just enough… In support of a *Special Toponymic Grammar* in Maltese. In: Achermann, Tanja–Schlücker, Barbara eds. 2017. *The morphosyntax of proper names.* Special issue of Folia Linguistica 51/2. Berlin–Boston, De Gruyter Mouton. 453–482.

	ˈjanis (anthroponym)	ˈkorinϑos (toponym)	*ˈfurnos (toponym)	ˈfurnos (common noun)
NOM.SING.	ˈjanis	ˈkorinϑos		ˈfurnos
GEN.SING.	ˈjani	koˈrinϑu		ˈfurnu
ACC.SING.	ˈjani	ˈkorinϑo		ˈfurno
VOC.SING.	ˈjani			ˈfurne
VOC.PL.	ˈjaniðes			ˈfurni
NOM.PL.	ˈjaniðes		ˈfurni	ˈfurni
ACC.PL.	ˈjaniðes		ˈfurni	ˈfurni
GEN.PL.	ˈjaniðon		ˈfurnon	ˈfurnon

Table 1: Comparison between the inflecDonal paradigm of the anthroponym ˈjanis (first column), the toponym ˈkorinϑos (second column) and the noun ˈfurnos, as a toponym (third column) and also as a common noun (forth column).

	Before (?)	Homer (8th century b.C.)	Classical era (~5th century b.C.)	Present
NOM.SING.	Ἀθῆνα*	Ἀθῆνα		Αθήνα
GEN.SING.	Ἀθήνας*	Ἀθήνας		Αθήνας
DAT.SING.	Ἀθήνᾳ*	Ἀθήνᾳ		Αθήνα
ACC.SING.	Ἀθῆναν*	Ἀθῆναν		Αθήνα
VOC.SING.				
VOC.PL.				
NOM.PL.		Ἀθῆναι		
ACC.PL.		Ἀθήνας		
DAT.PL.		Ἀθήναις		
GEN.PL.		Ἀθηνῶν		

Table 2: The diachrony of the inflectional paradigm of Αθήνα.

Attitudinal resources in academic talks:
A corpus-based analysis across languages

Yuxiang Duan & Liesbeth Degand
(Université Catholique de Louvain & Louvain la Neuve)

Attitudes are integral to human cognition, shaping perceptions and evaluations of the world. They play a crucial role in identity formation and self-presentation. As a sort of linguistic resource, expressions of attitudes manifest predominantly through attitudinal words (e.g., verbs, adverbs, adjectives), attitudinal bundles and clauses (e.g., it is adj. (that/to- clause)). The study analyses how scholars from diverse language backgrounds, including L1 English, English as a Lingua Franca (ELF), and L1 Chinese, articulate and convey their viewpoints through the use of attitudinal markers (ATMs) in academic talks. Three self-compiled corpora, comprising a total of 90 individual presentations, were collected based on accessible open-source academic video materials. Two key questions are addressed: (1) How do scholars manage attitudinal resources grammatically and functionally in academic talks; and (2) how do linguistic expressions differ in fulfilling the ATMs model across different languages/varieties.

The investigation refines and narrows the scope of ATMs as distinct linguistic elements and formulates an ATMs framework, drawing insights from earlier taxonomies proposed by scholars such as Vande Kopple (1985), Crismore et al. (1993), Hyland (2002a, 2002b, 2005), and Martin & White (2005). Enriched with additional categories and adjustments derived from the study materials, the model systematically classifies ATMs into three primary groups based on pragmatic functions: person's Affect, value-assessments, and language force and rhetorical tools. Acknowledging the inherent subjectivity of annotation, the target items were identified with the aid of computer-assisted programs, with two annotation rounds for reasons of validity and reliability, alongside the development of a rigorous coding scheme.

Findings reveal that, whether in Germanic English or Sino-Tibetan Chinese, scholars predominantly use positive value-assessment resources, with widespread use of intensification elements. Value-assessment cues are more frequent than Affect markers, aligning with the genre characteristics of academic discourse. Grammatically, adjectives and verbs with personal Affect and evaluation functions are predominant. This tendency reflects some peculiarities of the speakers' cultural or ideological backgrounds, although the degree of subjectivity varies depending on the evaluation parameter of the marker. A difference lies in the heightened density of attitudinal evaluation underscores the enhanced significance that the English data places on the manifestation of academic stances. In contrast, Chinese scholars emphasize the constructive guiding influences of both positive and negative emotions on individuals and society. Additionally, ELF scholars in the international academic community strive to align their attitudes expressions with those of L1 English speakers to mitigate potential miscommunication from ideological differences. Language force and rhetorical strategies are notable examples, with ELF scholars using them most frequently, followed by L1 English speakers and L1 Chinese speakers. These tactics enhance speakers' ability to accommodate varied communicative demands and expectations.

The implementation of contrastive (corpus) analysis, as exemplified in this study, contributes to an enhanced comprehension of identifying specific attitude cues and elucidates the role of language as a vehicle in conveying individual perspectives, particularly within the academic discourse domain.

The proposed labelling scheme for attitudinal resources is featured by objectivity and precision, making it applicable to other research practices of stance and (inter)subjectivity.

References

Crismore, Avon, Markkanen, Raija and Steffensen, Margaret S (1993), Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10, 39-71.

Hyland, Ken (2002a), Authority and invisibility: Authorial Identity in academic writing. *Journal of Pragmatics*, 34, 1091-1112.

Hyland, Ken (2002b), Activity and evaluation: Reporting practices in academic writing. In: J. Flowerdew (Ed.), *Academic discourse*, 115-130. New York: Longman.

Hyland, Ken (2005), Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7, 173-192.

Martin, James Robert and White, Peter Robert Rupert (2005), *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.

Vande Kopple William J (1985), Some explanatory discourse on metadiscourse. *College Composition and Communication*, 36, 82-93.

A corpus-based approach to consonant acquisition in Chinese-speaking preschool children

Shu-Chuan Tseng

(Institute of Linguistics, Academia Sinica)

Keywords: normative database, speech corpus, wordlist, consonant acquisition, Chinese

Phonological development data are crucial for speech acquisition research (Sander 1972). Likewise, phonetically labelled speech data are indispensable for acoustic-phonetic analysis and modelling (Usha & Alex 2023). However, large-scale datasets of both kinds are rare. This paper presents consonant acquisition results based on a speech corpus of 798 Taiwan Mandarin-speaking preschool children (Tseng 2024). 70 multisyllabic words with a balanced design of phonemes and tones were recorded, totalling 55,860 words. A two-stage transcription procedure was adopted to reduce manpower, but ensure high transcription quality. After annotating acceptability at the levels of word, syllable, and tone, words with no mispronounced or omitted segments or tones were automatically converted into canonical form. The rest of 18,694 words were manually transcribed.

There are four lexical tones and one unstressed tone in Mandarin Chinese. They are acquired early around three (Tseng 2024). The onset consonants in Mandarin Chinese consist of six plosives /p pʰ t tʰ k kʰ/, six fricatives /f s ʂ ɕ x ʐ/, six affricates /ts tsʰ tʂ tʂʰ tɕ, tɕʰ/, two nasals /m n/ and one lateral /l/, or it can be vacant. Only /n/ or /ŋ/ is allowed in the coda position (Duanmu 2007). Due to differences in subjects, data size, and language variety, developmental data vary even within the same language (Lin & Lin 1994 and Hua & Dodd 2000). Table 1 summarizes the developmental results of consonant acquisition with a PCC rate higher than 90%. Plosives and nasals are acquired early. But depending on the syllable position, /n/ is acquired much later in the coda than in the onset position. Mandarin Chinese has a larger number of affricates than the other classes. They are acquired at different paces. Among them, /tɕ/ is acquired early as also reported in the previous literature.

Table 1: Developmental data on consonant acquisition by manner and place of articulation

Manner of articulation	AGE 3-3;6	3;6-4	4-4;6	4;6-5	5-5;6	5;6-6	6-6;6	6;6-7	>7
Plosive	p t	pʰ tʰ	k kʰ						
Nasal	m n					n (coda)		ŋ	
Lateral					l				
Fricative			f		s ɕ				ʂ x ʐ
Affricate	tɕ			tɕʰ	ts tsʰ				tʂ tsʰ

Place of articulation	AGE 3-3;6	3;6-4	4-4;6	4;6-5	5-5;6	5;6-6	6-6;6	6;6-7	>7
Bilabial	m p	pʰ							
Labiodental				f					
Dental-alveolar	n t	tʰ			l	n (coda)			
Alveolar					s ts tsʰ				
Alveolar-palatal	tɕ			tɕʰ	ɕ				
Velar			k kʰ					ŋ	x
Retroflex									ʂ ʐ tʂ tʂʰ

In terms of place of articulation, bilabials are acquired the earliest, while retroflex sounds are the hardest for children to produce. Unaspirated consonants are acquired earlier than their aspirated counterparts. For /l/ and /x/, coarticulation due to neighbouring speech sounds affects pronunciation. More varieties of syllable structure are needed to examine acquisition details. Moreover, aspiration

and velarization are the two most frequent error types. It includes pronouncing unaspirated sounds as aspirated and vice versa. This applies to velar sounds, too. But it may be due to progressive and retrogressive contextual influences and requires further research.

To conclude, higher linguistic levels such as syllable structure, word position, and accompanying tones should also be considered to provide sophisticated phonological details of consonant acquisition. Using this corpus, acoustic analyses of speech sounds and learning models for phoneme recognition are currently being performed to obtain prospective insights into speech acquisition research and speech assessment applications.

References

Duanmu, San (2007), *The phonology of standard Chinese*. OUP Oxford.

Hua, Zhu and Barbara Dodd (2000), The phonological acquisition of Putonghua (modern standard Chinese), *Journal of Child Language*, 27(1), 3-42.

Lin, Grace Bao-Guey and Mei-hsiu Lin (1994), A study on the development of language abilities in preschool children in Taiwan, R.O.C., *Bulletin of Special Education*, 10, 259-281. (In Chinese)

Sander, Eric K. (1972), When are speech sounds learned? *Journal of Speech and Hearing Disorders*, 37, 55-63.

Tseng, Shu-Chuan (2024), Corpus-based research on speech acquisition and automatic assessment of Taiwan Mandarin-speaking children aged 3 to 6, in *Linguistic Diversity, but Unity in Research: Celebrating the Twentieth Anniversary of the Institute of Linguistics, Academia Sinica*, S.-C. Tseng and E. Zeitoun (Eds.) Taipei: Institute of Linguistics, Academia Sinica, pp. 475-502. (In Chinese)

Tseng, Shu-Chuan (2024), Tone acquisition in Chinese-speaking children: Developmental data of tone acceptability and contour pattern. Speech Prosody 2024. July 2-5. Leiden.

Usha, Gowri Prasood and John Sahaya Rani Alex (2023), Speech assessment tool methods for speech impaired children: a systematic literature review on the state-of-the-art in speech impairment analysis, *Multimed Tools Applications*, pp. 1-38.

Iterativity and verb classes

Petr Biskup
(Universität Leipzig)

Keywords: iterativity, verb classes, participial adjectives, motion verbs, impersonal constructions

1. Introduction: This paper deals with the interaction between iterativity (and aspect generally) and verb classes. It discusses participial adjectives, motion verbs and impersonal constructions, mainly from Czech. The analysis is couched in the minimalist morphosyntactic framework. It argues that the specific verbal theme -*a*- realizes iterativity and the agentive or expletive voice. Thus, -*a*- is responsible for changing the verb class behavior.

2. Slavic uses two specific types of participial adjectives: *past active*, with -*l*-, and *past passive*, with -*n/t*- (Schoorlemmer 1995, Kosta & Frasek 2004, Veselovská & Karlík 2004). Standardly, *l*-participles are derived from unaccusative stems and -*n/t*-participles from transitives. However, there are certain unaccusative stems co-occurring with -*n/t*-; consider (1). The stem has the ITER(ative) interpretation, as shown by the sg./pl. contrast.

(1) vy-pad-a-n-é   vlas-y   / # vy-pad-a-n-ý   vlas
     out-fall-**ITER**-n/t-pl   hair-pl   out-fall-**ITER**-n/t-sg hair.sg
     'fallen out hair'

3. It has been argued for Russian that directed motion verbs are unaccusative, whereas non-directed motion verbs are unergative. An analogous *na*- test for Czech shows that the directed verb (2a) behaves unaccusatively and the ungrammatical non-directed verb ((2b), with -*a*-) unergatively:

(2) Tolik   lidí   tam   a. na-běh-l-o   /   b.* na-běh-a-l-o!
     so.many   people   there   on-run-l-sg.n   on-run-**ITER**-l-sg.n
     'So many people ran there!'

While the directed *běží* has an progressive reading, the non-directed *běhá* has the iterative or generic meaning:

(3) Jirka a. běž-í   / b. běh-á   do školy.
     Jirka   run-TH   run-**ITER** to school
     'Jirka is running/runs to school.'

4. In contrast to transitive and unergative verbs, the formation of impersonal constructions from unaccusatives is very restricted; compare the (im)perfective (4a), (4b) with (5a) (see also Fehrmann *et al*. 2010).

(4) a. Už   se   (do-)četlo.   b. (Do-)pracovalo se.
         already self   to-read.l.sg.n   to-work.l.sg.n   self
         'People already finished reading.'   'People finished working.'

(5) a. * Umře-l-o se   na covid.   b. Umír-a-lo   se   na covid.
           at.die-l-sg.n self   on covid   at.die-**ITER**-l-sg.n self   on covid
                                                        'People died of covid.'
                                                  b'.# Ale nikdo nezemřel.
                                                        'But nobody died.'

Contrary to the perfective/quantized (5a), the imperfective (5b, with -*a*-) is grammatical. (5b) can only have the iterative or habitual interpretation (but not progressive), as shown by the contradictory (5b').

5. Analysis: All three phenomena (2,3,4) display an unexpected argument structure behavior and have an iterative interpretation. The theme *-a-* brings about iterativity (see (1),(3b),(5b)) and marks changes in the verb class behavior. For the iterative head, I assume (6), with the iteration set *E*:

(6) $[[\text{ITER}]] = \lambda P \lambda E \exists e.P(e) \wedge e \in E \wedge |E| > 1 \wedge \forall e'.e' \in E \rightarrow P(e') \wedge \neg \tau(e') \supset\subset \tau(e)$

The specific *-a-* cannot be just a secondary imperfective or progressive marker; see (3b) and (5b,b'). *-a-* licenses *se*, introduced by Voice (which is expletive in the non-agentive (5b), e.g. Alexiadou *et al*. 2015). *-a-* also verbalizes roots; compare *běh* 'the run' with (3b). Hence, *-a-* spells out more heads (cf. Ramchand 2008): verbalizing *v*, Iter(ative) and Voice (agentive (3b) or expletive (1)). Participle head is spelled out as *-n/t-* if it merges with a complement containing Voice (expletive (1) or agentive), otherwise as *-l-*.

References

Alexiadou, Artemis, Elena Anagnostopoulou and Florian Schäfer (2015), *External Arguments in Transitivity Alternations: A Layering Approach*, Oxford: Oxford University Press.

Fehrmann, Dorothee, Uwe Junghanns and Denisa Lenertová (2010), Two reflexive markers in Slavic, *Russian Linguistics* 34, 203–238.

Kosta, Peter and Jens Frasek (2004), Neakuzativita (ergativita) vs. neergativita v češtině, polštině a jiných slovanských jazycích na rozhraní morfologie a syntaxe, in Z. Hladká, and P. Karlík (eds.), (2004), *Čeština - univerzália a specifika 5,* Praha: Nakladatelství Lidové noviny, 189–212.

Ramchand, Gillian C. (2008), *Verb Meaning and the Lexicon: A First Phase Syntax,* Cambridge: Cambridge University Press.

Schoorlemmer, Maaike (1995), *Participial Passive and Aspect in Russian*, Utrecht: Ots dissertation series.

Veselovská, Ludmila and Petr Karlík (2004), Analytic passives in Czech, *Zeitschrift für Slawistik* 49, 163–235.

Lifting the stepchild out of poverty: Open Text Collections as a complement to grammars and dictionaries

Christian Döhler, Sebastian Nordhoff and Mandana Seyfeddinipur
(Berlin-Brandenburg Academy of Sciences and Humanities, Freie Universität Berlin & Berlin-Brandenburg Academy of Sciences and Humanities)

Franz Boas established the "Boasian Trilogy" in language documentation and description (Himmelmann 1998), consisting of a grammatical description, a dictionary, and a text collection. All three are necessary to get a comprehensive overview of a language, and they complement each other. While we have good outlets for grammars (eg Comprehensive Grammar Library) and dictionaries (eg Dictionaria), such is not the case for text collections. This means that only few of them are published, and even fewer follow the FAIR principles of findability, accessibility, interoperability, and reusability (Wilkinson 2016).

The project Open Text Collections (http://opentextcollections.org) will remedy this by making high quality text collections from endangered lan- guages available in an open interoperable format. Next to providing pdfs or printed books to the communities themselves, this setup will also provide the data in CLDF format (Forkel et al. 2018) for downstream use in NLP applica- tions.

Most reference grammars published today are the result of a language docu- mentation project, often part of authors' dissertation projects. These grammars should be data-driven and accompanied by a corpus in order to facilitate the verification or falsification of the analysis (Nordhoff 2008, Mosel 2012). While countless hours are invested into the structuring and glossing of texts, in many cases, however, these texts are not made available in a reusable way. Linguists tend to have them somewhere on their hard drive, or uploaded to an archive but there is no generally established way of publishing them, at least not in a format which would feed further research downstream (e.g. linguistic typology, corpus-based language description, or NLP). This means that these valuable results of language documentation often fail to be discovered.

Open Text Collections will provide a quality venue for publishing text col- lections, following the setup established by Language Science Press. The plat- form is community-driven and aims at being attractive to both data producers (ie language documenters) as well as data users (language communities, ty- pologists, NLP practitioners). For data producers, the platform will provide rigorous peer review, quality control, and top-notch publishing (pdf and print- on-demand), making sure that the time invested in a text collection will not harm job prospects. For data consumers, different outlets will be available to suit different needs: printed books will be available for communities; a search interface (prototype available at https://imtvault.org) will be available for typologists, and all data will be available as CLDF dump for NLP practitioners. By making reuse easy, the research will spread more widely, which in turn is very attractive for the data producers.

As of today, there are 5 regional boards and 40 proposed text collections. This poster will showcase the platform, its motivations, and its benefits for data producers and consumers.

References

Forkel, Robert, Johann-Mattis List, Simon Greenhill, Christoph Rzymski, Se- bastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping & Russell D. Gray. 2018. Cross-Linguistic Data Formats, advancing data sharing and re-use in comparative linguistics. *Sci Data* 5. DOI: 10.1038/sdata.2018.205.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Lin- guistics* 36. 161–195.

Mosel, Ulrike. 2012. Advances in the accountability of grammatical analysis and description by using regular expressions. *Language Documentation & Conservation Special Publication* 4. 235– 250.

Nordhoff, Sebastian. 2008. Electronic reference grammars for typology: Chal- lenges and solutions. *Language Documentation & Conservation* 2(2). 296– 324.

Wilkinson, M. et al. 2016. The FAIR guiding principles for scientific data man- agement and stewardship. *Sci Data* 3. 160018. DOI: 10.1038/sdata.2016. 18.

Variation in Cappadocian Greek relative clauses:
Pattern replication and diatopy

Eline Daveloose
(Ghent University)

Cappadocian Greek (CG) is a critically endangered Greek dialect that was spoken in the Turkish region of Cappadocia until the Greek-Turkish Population Exchange of 1923-24 (Janse 2020: 47-8). Consequently, this dialect was thoroughly influenced by the surrounding Turkish language on all levels of its grammar and lexicon (Thomason & Kaufman 1988: 215-22; Winford 2003: 83-4). One of the numerous instances of contact-induced language change is the CG relative clause (RC), which shows, among other things, clear Turkish influence in its word order, as can be seen in the following examples:

(1) CG	*tu=puliδj-ú*	*t=óima*	*t=áxsen*	***don=dópo***
	the=bird-GEN	the=blood	REL=fall.PFV.PST.3SG	the=place
(2) Tr.	*kuş-un*	*kan-ı-nın*	*düş-tüğ-ü*	***yer-de***
	bird-GEN	blood-POSS.3SG-GEN	fall-PTCP-POSS.3SG	place-LOC
(3) ModGr.	***s-ton=dópo***	*pu*	*épese*	*to=éma*	*tu=puli-ú*
	on-the=place	REL	fall.PFV.PST.3SG	the=blood	the=bird-GEN

'in the place where the bird's blood fell' (Dawkins 1916: 312)

Though briefly addressed by Janse (1999) and mentioned within a broader analysis of relativizers in Modern Greek dialects (Liosis & Kriki 2013), a comprehensive study of the CG RC, with emphasis on Turkish influence, is still absent from current scholarship. Therefore, this paper provides an in-depth analysis of the CG RC, addressing this construction as an example of pattern replication (or 'PAT'; Matras 2020: 257-64) from the Turkish RC. This process resulted in a complex hybrid composition of, on the one hand, features that remained Greek (e.g., the use of a relativizer and a finite verb), and, on the other hand, elements that were replicated from Turkish (e.g., the possibility to have the relative clause in prenominal position). This construction corresponds to what Keskin (2023) calls "X-clauses", i.e., clauses that consist of a mixture of features of a Turkic and an Indo-European model, resulting in a patchwork of different clause types.

Focusing on variation in the CG RC, the objectives of the present study are twofold. First, this paper will investigate what PAT precisely entails in this case, taking into account a third category of features that may exhibit either a Greek or a Turkish version, and "oscillate" (Keskin 2023: 175-6) between these two versions. For example, many tokens show a prenominal RC (cf. Turkish), whereas the RC follows the head noun (cf. Greek) in another part of the tokens. Second, this paper will also analyse diatopic variation in this construction, as CG consists of multiple subdialects, which show notable regional diversification regarding their Turkish influence.

This paper addresses these two research objectives through the analysis of a text corpus including the transcriptions of 58 orally transmitted folktales (ca. 50,000 words) in 11 CG subdialects.

In doing so, the present study contributes to the understanding and description of understudied, though frequent syntactic features in the underdocumented and severely endangered CG dialect.

References

Dawkins, R.M. (1916). *Modern Greek in Asia Minor*. Cambridge University Press.

Janse, M. (1999). 'Greek, Turkish, and Cappadocian Relatives Revis(It)Ed', *Greek Linguistics 97*, 453–462.

–. (2020). 'Ἑλληνιστὶ γινώσκεις; (Acts 21:37): The survival of Cappadocian Greek', *Cursor: Zeitschrift für Freunde der Lateinischen Sprache und europäischen Kultur 16*: 49-57.

Keskin, C. (2023). 'Transient subordinate clauses in Balkan Turkic in its shift to Standard Average European subordination. Dialectal and historical evidence', *Folia Linguistica Historica 44*(1): 155–197.

Liosis, N., & E. Kriki (2013). Towards a typology of relative clauses in Modern Greek dialects. In M. Janse, B. Joseph et al. (Eds.), *Proceedings of the Fifth International Conference of Modern Greek Dialects and Linguistic Theory (Ghent, 20–22 September 2012)*. University of Patras.

Matras, Y. (2020). *Language Contact. 2nd ed*. Cambridge University Press.

Thomason, S.G., & T. Kaufman (1988). *Language Contact, Creolization, and Genetic Linguistics*. University of California Press.

Winford, D. (2003). *An introduction to contact linguistics*. Blackwell.

Exceptions to animacy in Czech

Karolina Wieserová
(Paris Lodron University Salzburg)

Keywords: animacy, syncretism, Czech, corpus linguistics, differential object marking

My poster shows the results of my corpus linguistic study on exceptions to the grammatical category of animacy in Czech. Different factors can influence whether inanimate masculine nouns are treated as masculine animate or masculine inanimate.

The grammatical category of animacy, which distinguishes between animate and inanimate masculine nouns, can be found in all Slavic languages except for Bulgarian and Macedonian. My study focuses on (ontologically) inanimate objects in Czech that have a genitive-accusative syncretism in the singular which should, as expected, be reserved for (ontologically) animate objects. However, West Slavic languages tend not to make this distinction in the accusative case of certain nouns and mark inanimate objects with the genitive ending *-a*, which is the accusative ending for masculine animates and at the same time homonymous with the genitive (cf. Kosta 2003). In the example *Nastavil(a) jsem budík* ('I set the alarm clock'), a nominative-accusative syncretism i.e., *budík*, should be used according to the Czech written standard variety. However, Czech native speakers tend to use the ending *-a*. *Nastavil(a) jsem budíka* is prefered. We can see the genitive ending *-a* in the accusative case.

The distinction between animate and inanimate can also be found in other languages of the world, and we also find exceptions there. García García (2014), for example, shows in his corpus study how different features, such definiteness, object position and verb semantics have an impact on the marking of animate objects in Spanish.

I have chosen a corpus-based study that helps to reveal trends in the use of genitive-accusative syncretism with inanimate objects. The corpus is analysed using Sketch Engine. The evidence found is analysed for features such as agentivity, definiteness, etymology, verb semantics or object position. The factors were selected on the basis of previous studies on other languages than Czech. For example, the most relevant criterion for animacy exceptions in Polish is the origin of the nouns: Inanimate borrowed nouns tend statistically more often towards a genitive accusative syncretism than inanimate inherited nouns do (cf. Kiklewicz 2023). In Spanish, on the other hand, verb semantics plays a central role (cf. García García 2014). From the perspective of differential object marking (DOM), definiteness can be furthermore responsible for irregularities in the marking of nouns in the accusative case (cf. Aissen 2003). This corresponds to the concept of DOM, which is why it is used as the theoretical basis for the study.

References

Aissen, Judith (2003), Differential object marking: Iconity vs. Economy, *Natural Language & Linguistic Theory* 21, 435–483.

García García (2014), *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*. Berlin / Boston: De Gruyter (= Linguistische Arbeiten, 545).

Kiklewicz, Aleksander (2023): „Niesforny" biernik rzeczowników męskonieżywotnych w liczbie pojedynczej: o ekspansji tego zjawiska we współczesnym języku polskim i jego czynnikach, *Język Polski* 103(2), 5–25.

Kosta, Peter (2003), The New Animacy Category in Slavic Languages: Open Questions of Syntax, Semantics and Morphology, *Germanoslavica: Zeitschrift für germano-slawische Studien* 9, 179–198.

The Idioms of Serbs in Metohia in Contact with Albanian Language

Andrey N. Sobolev
(St. Petersburg)

Keywords: dialects of Serbian, dialects of Albanian, Metohia, language contact, Balkan Sprachbund

In this conference presentation, from the perspective of comparative-historical Slavic linguistics and areal-typological Balkan linguistics, innovations in the territorial dialects of the Orthodox Serbian population in Metohia (Alb. Rrafshi i Dukagjinit 'Dukagjini Plain') in the Autonomous Province of Kosovo and Metohia within the Republic of Serbia (resp. partially recognized state of the Republik of Kosovo) are examined. These innovations are likely induced by prolonged contact with the Albanian language, including its local territorial dialects, at all levels of linguistic structure: phonetic-phonological, morphonological, morphological and morphosyntactic, syntactic, lexical, and phraseological.

A crucial theoretical question is raised regarding the possibility of distinguishing, through strictly linguistic methods, the results of language contact from developments influenced by internal factors. This includes innovative changes in the inventory and distribution rules of phonemes, grammatical units (markers), categories, lexical, and phraseological composition of the language. The potential role of contact in preserving archaisms in language structure is also considered.

As direct and indirect evidence of contact-induced linguistic innovations, the following circumstances are proposed to be primarily considered: the absence of a phenomenon observed in the presumed recipient language in its proto-state preceding language contact (e.g., in Proto-Slavic, Proto-Serbian); the absence of a phenomenon in closely related languages and dialects that were not and are not in similar contact situations (e.g., in North and West Slavic languages, in Serbian dialects outside the zones of interaction with the core languages of the Balkan Sprachbund); borrowing of linguistic material from the presumed donor language (in this case, Albanian); the presence of regular, systemic correspondences in the functions of linguistic units and categories of the presumed recipient and donor languages. The methodology for collecting such evidence, limitations in terms of their evidential strength, and their evidential hierarchy are discussed.

Against the backdrop of the reconstructed Common Serbian state and the situation in other territorial dialects of the Serbian language, innovations in the Metohia region, such as the labialized vowel of the front row of the upper lift /y/ (dygmen͡tse, gyrbet, dyʃsek), the opposition of narrow and wide vowels in the middle row, trembling /rr/, neutralization of the opposition in palatality of sonants (kral', l'epo), and affricate (rod͡ʒak), borrowed word-formational elements, word order in attributive phrases, pronominal doubling of the indirect and direct object (briga mu je n'emu 'he doesn't care'), clitic order (ne mu treba nemu 'he doesn't need'), interrogative particles (Alb. a), lexical and phraseological borrowings and calques, and others, are examined in detail.

The main attention is given to the local idioms of the Kosovo-Resava dialect, such as the dialects of Northern Metohia [Bukumirić 2003]), and the local idioms of the Prizren-South Morava dialect, such as the dialects of Podrima (town of Orahovac, village of Velika Hoča), town of Đakovica [Stevanović 1950], town of Prizren [Remetić 1996]. Data on the dialects of the Kosovo-Metohia region of the Northern Šar-Planina and Kosovo plain are also brought in for comparison. Currently, all these dialects do not form a single linguistic continuum.

Against the backdrop of historical-linguistic information about the Serbian language in the region (data from comparative-historical Slavic linguistics, Serbian dialectology, toponymy, historical anthroponymy, medieval monuments of Slavic literacy, etc.) and the latest sociolinguistic information, linguistic situations of contact in which the studied idioms were and are currently located are

systematized. Local features of historically developed contact situations in rural and urban settlements of Northern and Southern Metohia are characterized, and conclusions are drawn about the causal relationship between the conditions of linguistic contact and its observed results, taking into account ethnic and confessional differences between the interacting ethnic groups.

The material of Serbian Metohian (Prizren-South Moravian and Kosovo-Resava) dialects, located at the westernmost edge of the Serbian language space, is a valuable source of information about contact-induced language processes characteristic of idioms that have been in conditions of prolonged ethnic and cultural separation.

References

Bukumirić, Mileta. *The Dialects of Northern Metohia*. Belgrade: SANU, 2003.

Remetić, Slobodan. *The Serbian Prizren Dialect*. Belgrade: SANU, 1996.

Stevanović, Mihajlo. *Đakovica Dialect*. Belgrade: SANU, 1950.

Sobolev, Andrey N. *Between Separation and Symbiosis: South Eastern European Languages and Cultures in Contact*. Berlin; Boston, 2021.

A quantitative approach to the semantic classification of English blends with token-based semantic vector space modeling

Qingnan Meng, Martin Hilpert & Weihua Luo
(Dalian Maritime University, University of Neuchâtel & Dalian Maritime University)

This research analyzes the semantic headedness of English blends with distributional semantics methods. The semantic head of a blend is the source word that transfers its semantic information to the blend as a whole. For example, a sitcom is a kind of comedy. But is *FedEx* a kind of *express*, and is *wi-fi* a kind of *fidelity*? We use corpus data and token-based semantic vector space modeling (Hilpert & Correia Saavedra, 2020) in order to address these questions. Specifically, we investigate whether Plag's (2018) ternary division of endocentric, exocentric and coordinative compounds based on semantic headedness can also be applied to English blends, and whether the general tendency of semantic right-headedness can be observed for all the three subtypes.

In order to address this issue, we examine whether the three categories display different patterns of semantic overlap with the two source words. We operationalized this question by creating token-based semantic vector spaces for 55 English blends and their respective source words on the basis of the Corpus of Contemporary American English (COCA) and the English Web Corpus 2021 (TenTen corpus). Each word was represented by an equal number of 200 hits. We used metric multidimensional scaling (MDS) to visualize the results in two-dimensional graphs. The degree of semantic similarity between blend and its two source words was measured by the Euclidean distance between medoids.

The results show that the general tendency of right-headedness holds true for most endocentric blends. For coordinative blends, a weak semantic right-headedness is also observed in most cases, which may serve as a *post-hoc* explanation for the sequence of their component parts. For exocentric blends, few examples of complete semantic independence between the blend and its two components were found, and *FedEx*, *fortran* and *bionic* may be the only three cases in point.

On the basis of these findings, we argue for a prototype-based gradient approach to the semantic classification of blends. We conclude that Plag's (2018) classification offers a useful point of departure for the semantic analysis of blends and that distributional semantics methods can provide new insights into their semantic behavior.

References
Hilpert, Martin and Correia Saavedra, David. (2020). Using token-based semantic vector spaces for corpus-linguistic analyses: From practical applications to tests of theoretical claims. *Corpus Linguistics and Linguistic Theory*, 1–32.
Plag, Ingo. (2018). *Word-Formation in English (2nd edn.)*. Cambridge: Cambridge University Press.

Southern Finnic polar questions in their areal context

Karl Pajusalu, Helle Metslang, Miina Norvik, Eva Saar & Uldis Balodis
(University of Tartu, University of Tartu, University of Tartu, University of Tartu & University of Tartu;
University of Latvia Livonian Institute)

Keywords: Finnic languages, polar questions, language contact, typology, areal linguistics

Asking questions is one of the central actions of communication; specialised tools for forming questions are likely found in all languages (König & Siemund 2007). One of the main question types is the polar question, which concerns the truth content of a sentence. Various means have developed for expressing polar questions, of which special word order and clause-initial interrogative particles are the most common in European languages; in the Circum-Baltic area, particles are used (Grambank; Koptjevskaja-Tamm & Wälchli 2001).

Our presentation deals with the polar question formation strategies of the southern Finnic (SF) languages and beyond. This study is part of a broader study combining areal and intragenetic typology to describe the typological uniqueness of this language group, which is distant from the eastern Finno-Ugric languages (cf. Miestamo 2011). We have chosen polar questions as an indicator phenomenon, because these are found in all languages, but their form varies areally.

We seek to answer research questions such as: (a) what are the main strategies for forming polar questions in SF languages and dialects? (b) what are the shared and unique features of the forms of polar questions in SF languages – both among SF languages as well as in comparison with related and contact languages, and what might be the reason for them? (c) what is the distribution of polar question devices in the Central Baltic Area in the context of European and Uralic languages?

Our research method is a microareal functional typological analysis, which considers theories involving grammaticalisation as well as different types of borrowing (Sakel 2007). The languages and dialects examined are Estonian and its varieties, the South Estonian language islands in Latvia (Lutsi, Leivu) and Russia (Kraasna), Ingrian, Votian, Courland and Salaca Livonian, and the contact languages Latvian, Latgalian, and Russian. We use data from text collections, corpora, and typological databases (Grambank, UraTyp).

Polar question markers are typically formed from coordination, subordination, or negation markers (Metslang et al. 2017, and Kuteva et al. 2019). Preliminary results show that in the SF languages, and also in the Baltic languages, the main formative tool of the polar question is particles originating from disjunctive ('or' > VÕI-type) or conjunctive ('also' > KAS-type) coordination markers, the development paths of which are intertwined with MAT-borrowing (e.g., Latvian *vai* 'or' is a Finnic loan), language-internal grammaticalisation (Latvian *vai* 'or'> 'question'), and PAT-borrowing (e.g., Latvian-influenced polysemy 'or'/'question' of *või* in Livonian). The distribution of question particles in the studied area allows for distinguishing between two main groups: (i) languages containing clause-initial VÕI (Livonian and Leivu, marginally also Votian and Ingrian), ex. (1), (2); (ii) languages that make use of the clause-initial KAS or clause-final VÕI (Estonian varieties, Lutsi), ex.(3), (4). SF languages are different from northern Finnic languages, Russian, and the Mari language, where a clause-internal particle is found. Clause-final particles are found in some distantly related languages, but the emergence of the

final VÕI in variants of Estonian is a recent development. The inversion typical of European languages is rare in SF languages.

Examples

(1) ***Või*** *sa* *tuoid* *ka* *leibõ?* (Livonian)
     Q 2SG bring:PST:2SG also bread.PRTV
     'Did you also bring some bread?'

(2) ***Võis*** *teil* *is* *ole* *külm* *ši* *kevaja?* (Leivu)
     Q 2PL:ADE NEG.PST be.CNG cold that spring
     'Weren't you cold that spring?'

(3) ***kas*** *ma* *jyvva* *sinno* *ravidaq?* (Lutsi)
     Q 1SG be.able:PRS.1SG 2SG:PRTV feed:INF
     'Can I feed you?'

(4) ***ka*** *maha* *putti* ***võh****?* (Estonian Mulgi dialect)
     Q down fall:PST.3SG Q
     'Did it fall down?'

References

Grambank, https://grambank.clld.org/ (27.05.24)

Koptjevskaja-Tamm, Maria & Bernhard Wälchli (2001), The Circum-Baltic Languages: An Areal-Typological Approach, in Östen Dahl, and Maria Koptjevskaja-Tamm (eds), (2001), *The Circum-Baltic languages: typology and contact. Volume 2: Grammar and Typology*, Studies in Language Companion Series 55, Amsterdam / Philadelphia: John Benjamins Publishing Company, 615–750.

König, Ekkehard & Peter Siemund (2007), Speech act distinctions in grammar, in Timothy Shopen (ed), (2007), *Language typology and syntactic description*, Cambridge: Cambridge University Press, 276–324.

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

Metslang, Helle, Külli Habicht & Karl Pajusalu (2017), Where do question particles come from? *STUF – Language Typology and Universals* 70(3), 105–137.

Miestamo, Matti (2011), Polar interrogatives in Uralic languages: A typological perspective, *Linguistica Uralica* 47(1), 1–21.

Sakel, Jeanette (2007), Types of loan: Matter and pattern, in Yaron Matras, and Jeanette Sakel (eds), (2007), *Grammatical Borrowing in Cross-Linguistic Perspective*, Berlin / New York: Mouton de Gruyter, 15–29.

UraTyp, https://uralic.clld.org/ (27.05.24)

Transitivity prominence in Indo-European and beyond

Tania Ahmad & Eystein Dahl
(Adam Mickiewicz University & Adam Mickiewicz University)

This paper explores to what extent transitivity prominence represents a stable feature across the Indo-European family. Transitivity prominence is a measure for comparing the relative pervasiveness of transitive encoding in the verbal lexicon across languages, transitive encoding being understood as the morphosyntactic encoding characteristic of so-called core transitive verbs like *break*, *build* or *kill* (Haspelmath 2015). One would expect that languages with typologically similar alignment and argument realization systems show roughly the same degree of transitivity prominence, and that this also applies to genetically related languages. However, Seržant et. al (2022) suggest that the areal factors sometimes impact this dimension of grammar. It remains to be established how genetic and areal factors interact in determining how widespread transitive encoding is in the verbal lexicon of a language. This research question motivates this contribution. Drawing on Creissels (2018), we utilize a sample of 30 verb meanings as the basis for extracting verbal lexemes and/or idioms from Hittite, Vedic Sanskrit, Homeric Greek, Early Latin, Old Irish and Old Lithuanian, six Indo-European languages which differ in the manner and extent to which they have been subject to historical contact with other languages. The lexemes are classified according to three morphosyntactic parameters:

- Active voice or not
- Canonical/nominative subject case marking and verb agreement
- Canonical/accusative object case marking

Predicates selecting for active voice, nominative subject and consistently accusative object are classified as canonically transitive, whereas predicates differing from the transitive prototype regarding one or more parameters are classified as having non-canonical voice and/or argument marking. Preliminary results of comparison of some of the pertinent data from the six languages are given in Table 1 and their relative proportions are visualized in Figure 1. These figures are indicative of some fluctuation between the languages. However, a chi-squared test of the observed frequencies did not yield a statistically significant result. The chi-squared test yielded a p-value of 0.32, a chi-square value of 16.971 and 15 degrees of freedom ($\chi2(15) = 16.971$). (These results were obtained by using the chisqu.test() function in the standard package of R, cf. R Core Team 2023). At face value, this finding indicates that the languages under scrutiny show roughly corresponding patterns of transitivity prominence, suggesting that language contact has limited effect upon the relative transitivity prominence of a language and that genetic affiliation plays a more important role in this respect. On the other hand, it cannot be excluded that this result is due to the somewhat limited sample size, which represents the necessary minimum. Another, related problem concerns the extent to which Creissels' (2018) sample of verb meanings is sufficiently representative to capture relevant divergences at a more fine-grained level of analysis. Our presentation will deal with these and related problems.

Table 1: Transitivity prominence across six branches of Indo-European

	Hittite	Vedic	Homeric Greek	Early Latin	Old Irish	Lithuanian
Canonically transitive morphosyntax	14	14	15	19	13	12
Middle voice marking	7	7,5	6	3,5	7,5	4
Non-canonical subject marking	0.5	0	1	0	0	1
Non-canonical object marking	9	4,5	12	5,5	8	17
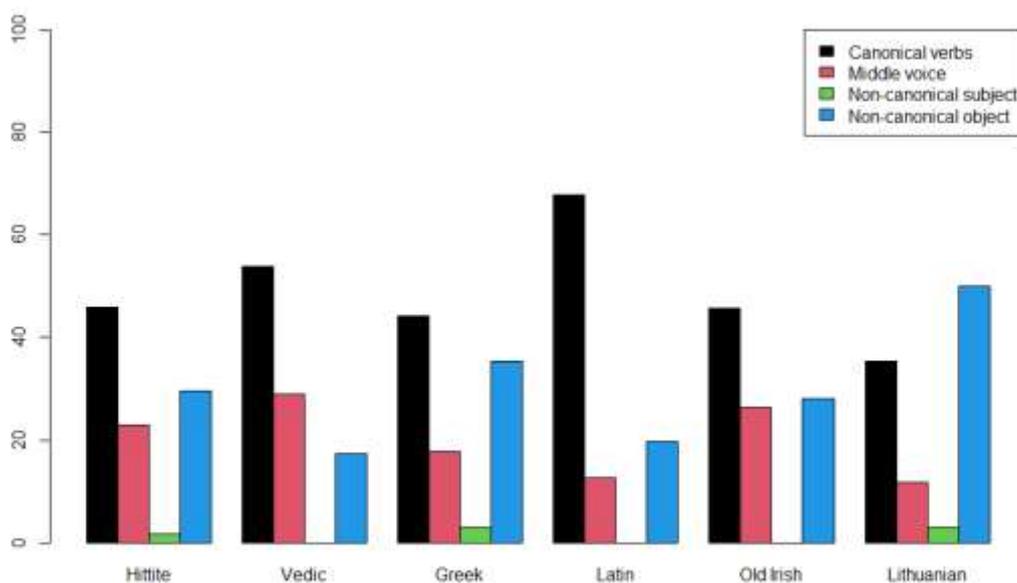
Figure 1: Relative proportions of canonically transitive predicates, middles/deponents and non-canonical argument marking

References:

Creissels, Denis. 2018. Transitivity prominence in typological perspective: the case of Basque. In Joseba A. Lakarra and Blanca Urgell (eds.) *Studia Philologica et Diachronica in honorem Joakin Gorrotxategi Vasconica et Aquitanica*. Special issue of Anuario del Seminario de Filología Vasca «Julio de Urquijo» International Journal of Basque Linguistics and Philology LII: 1-2 (2018), 175-187.

Haspelmath, Martin. 2015. 'Transitivity Prominence', in Andrej L. Malchukov and Bernard Comrie (eds.), *Valency Classes in the World's Languages*, vol. 1: *Introducing the Framework, and Case Studies from Africa and Eurasia* (Comparative Handbooks of Linguistics 1/1). Berlin: de Gruyter Mouton, 131–147.

Janda, Laura A., Anna Endresen, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nesset, and Svetlana Sokolova. 2013. *Why Russian aspectual prefixes aren't empty. Prefixes as Verb Classifiers*. Bloomington, IN: Slavica Publishers.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Seržant, Ilja A., Björn Wiemer, Eleni Bužarovska, Martina Ivanová, Maxim Makartsev, Stefan Savić, Dmitri Sitchinava, Karolína Skwarska and Mladen Uhlik. 2022. Areal and diachronic trends in argument flagging across Slavic. In: Eystein Dahl (ed.), *Alignment and Alignment Change in the Indo-European Family*. Oxford: OUP.

Elements of BE-perfect grammaticalization in the person-based auxiliation system of Barese

Danguolė Kotryna Kapkan
(Vilnius University)

Keywords: perfect, grammaticalization, person-based auxiliation, Italo-Romance

The Barese dialect belongs to the group of Italo-Romance dialects of Central and Southern Italy, featuring typologically rare person-based auxiliation systems in the perfect (Loporcaro 2007, 2022, Andriani 2017, 2018). While Barese is documented to adhere to a pattern of BE in 1SG, 2SG, 1PL, 2PL and HAVE in 3SG, 3PL, my study, which draws on a collection of published written Barese texts, shows that some degree of auxiliary variation occurs in all persons, except for 2SG. In my talk, I propose an account of this variation based on a grammaticalization cline specific to perfects formed with a BE auxiliary and a participle (Kapkan (2021, forthcoming) on Bulgarian and Lithuanian, also applicable to some developments of the Old Church Slavonic (Plungian & Urmanchieva 2017, 2018)).

Cennamo (2008) demonstrates how a comparable variation in certain Campanian varieties with person-based auxiliation systems depends on lexical classes of verbs in the perfect and reflects a gradual spread of BE into the domain of HAVE, which proceeds along the lines of the Auxiliary Selection Hierarchy (ASH) (Sorace 2000, 2011). In part, the developments observed in Barese are similar to those described by Cennamo (2008), but the BE auxiliary also appears in other contexts unforeseen by the ASH or the person-based pattern. I argue that this group of tokens can be defined as ascriptive copular constructions with adjectivized participles, termed statives and directly comparable to equivalent constructions in Lithuanian and Bulgarian, despite the different diathesis of the participles (passive diachronically in Romance and active in Baltic and Slavic).

Statives represent the first stage of a BE-perfect grammaticalization cline modelled on the 'X is Y' schema (Heine 1993, Kuteva 2004). This cline proceeds via ascriptive copular constructions with adjectivized participles (statives, Stage 1), to subject-oriented resultatives (Nedjalkov & Jaxontov 1988), defined as perfects with change of state and change of location predicates (Stage 2), and on to other cross-linguistically common perfect values, such as experientials.

My study draws on an annotated database of 675 perfect tokens, created by extracting all perfects from a collection of publicly available written Barese texts. The data analysis reveals that the BE auxiliary in 3rd person perfects appears in precisely the contexts aligning with the semantic values defined as the first two steps of the grammaticalization cline proposed for Lithuanian and Bulgarian BE perfects (Kapkan (forthcoming)):

1. With statives, that do not clearly presuppose a prior event, but rather assign a quality to the subject, the BE auxiliary is used almost exclusively with all persons (97%):

Acquànne	*la*	*carne*	*iè*	*ccotte*
when	DEF.SG.F	meat.SG.F	be.PRS.3SG	cook.PP.F

[*se lève e se mètte a ttàuue.*]
'When the meat is cooked/ready, you take it off and put it on the table.'

2. With subject-oriented resultatives which not only rank higher in ASH but also constitute the prototypical semantic value for BE perfects, BE appears in ~30% of 3SG tokens and ~20% of 3PL tokens.

Chèdda	*giacchètte,*	*addò*	*è*	*ssciùte?*	*[A la uèrre?]*
DIST.SG.F	jacket.SG.F	where	be.PRS.3SG	go.PP	

'That jacket, where has it been? [To war?]'

Such an account of the auxiliary variation in Barese not only allows to explain all contexts where the BE auxiliary appears in violation of the person-based pattern, but also relates this variation, which only occurs in the perfect, to the perfect category as such and to its development.

References

Andriani, Luigi. 2017. *The syntax of the dialect of Bari*. Homerton College, University of Cambridge PhD dissertation. https://doi.org/10.17863/CAM.15625. (27 October, 2022).

Andriani, Luigi. 2018. Instability and Change: A Parametric Approach to Barese Auxiliary Selection. In Gabriela Pană Dindelegan, Adina Dragomirescu, Irina Nicula, Alexandru Nicolae (ed.), *Comparative and Diachronic Perspectives on Romance Syntax*, 361–400. Cambridge Scholars Publishing.

Cennamo, Michela. 2008. The rise and development of analytic perfects in Italo-Romance. In Thórhallur Eythórsson (ed.), *Linguistik Aktuell/Linguistics Today*, vol. 113, 115–142. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.113.05cen.

Heine, Bernd. 1993. *Auxiliaries: cognitive forces and grammaticalization*. New York: Oxford University Press.

Kapkan, Danguolė Kotryna. 2021. Perfect in Lithuanian: A case study based on data from Facebook comments. *Baltic Linguistics* 12(12). 21–71. https://doi.org/10.32798/bl.921.

Kapkan, Danguolė Kotryna. forthcoming. BE perfects and grammaticalization in Bulgarian and Lithuanian: A study based on data from Facebook comments.

Kuteva, Tania. 2004. *Auxiliation: An Enquiry Into the Nature of Grammaticalization*. Oxford University Press.

Lindstedt, Jouko. 2000. The perfect – aspectual, temporal and evidential. In *Tense and Aspect in the Languages of Europe*, 365–384. De Gruyter Mouton. https://doi.org/10.1515/9783110197099.3.365.

Loporcaro, Michele. 2007. On triple auxiliation in Romance. *Linguistics* 45(1). https://doi.org/10.1515/LING.2007.005.

Loporcaro, Michele. 2022. The morphological nature of person-driven auxiliation: Evidence from shape conditions. In *Periphrasis and Inflexion in Diachrony*, 213–238. Oxford University Press. https://doi.org/10.1093/oso/9780198870807.003.0009.

Nedjalkov, Vladimir P. & Sergej Je. Jaxontov. 1988. *The Typology of Resultative Constructions*. (Ed.) Vladimir P. Nedjalkov & Bernard Comrie. Vol. 12. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.12.06ned.

Plungian, Vladimir A. & Anna Yu. Urmanchieva. 2017. Перфект в старославянском: был ли он результативным? / The Perfect in Old Church Slavonic: Was It Resultative? *Slavistična revija* 2. 13–56.

Plungian, Vladimir A. & Anna Yu. Urmanchieva. 2018. К типологии нерезультативного перфекта (на материале старославянского языка). *Slovene* (letnik 66/2018). 421–440.

Sorace, Antonella. 2000. Gradients in Auxiliary Selection with Intransitive Verbs. *Language*. Linguistic
    Society of America 76(4). 859–890. https://doi.org/10.2307/417202.

Sorace, Antonella. 2011. Gradience in Split Intransitivity : The End of the Unaccusative Hypothesis?
    Archivio glottologico italiano : XCVI, 1, 2011. Le Monnier 67–86.
    https://doi.org/10.1400/179059.

The perceptivity and productivity of tonal change triggered by tone compounds in the Pingding variety of Chinese

Pingping Jia
(Freie Universität Berlin)

Keywords: tone sandhi, Chinese, productivity, perceptivity, tone compound

This study explores tone sandhi, a phenomenon in tone languages where tones undergo systematic changes when adjacent to certain other tones (Chen 2000). Existing studies have highlighted that in the Pingding Chinese, certain two-tone sequences, a low checked tone followed by a falling tone (Lq-TF), can be reanalyzed as Tone Compounds (TC, TCs), inducing tonal changes in a preceding tone, where neither of the tones alone has such an effect (Jia & Meinschaefer, to appear). All three possible TCs present this effect, and each is equivalent in its sandhi-triggering effect to one of the three falling tones existing in the Pingding tone inventory: high-falling tone (HF), middle-falling tone (MF), and low-falling tone (LF). Hence, each TC undergoes reanalysis as one of the falling tones, presenting evidence for a chain shift (Łubowicz 2011). Unlike the usual "tone circle" described in the literature (Barrie 2006, Chen 1987), in Pingding, it is not the surface realization that undergoes the chain shift. Instead, the tones LF, MF, and HF – the second elements of the sequence Lq-TF – are, in their sandhi effects, reanalyzed as equivalent to one of the single tones MF, HF, and LF.

This study centres on two crucial questions for a deeper understanding of TCs: whether native speakers can perceive these changes, indicating the psychological mechanisms underlying tonal alternation; and whether these tone-changing patterns are still productive, indicating if they are tone sandhi processes or lexicalized tonal templates, like in Turkish or Japanese (Beckman & Pierrehumbert 1986, Levi 2005).

To address these questions, a wug test and a perception test were conducted with native speakers. The wug test involved generating existing words, novel words, and pseudowords, all featuring tonal sequences with TCs (T-[Lq-TF]). The perception test involved participants identifying differences in monosyllabic citation tones, tones produced before a TC, and tones preceding a single falling tone.

In Pingding, two sandhi processes are observed: one driven by co-articulation, also known as natural tone sandhi, and the other induced by OCP-constraint, representing a less-natural tone sandhi (Zhang 2022). In the wug test, co-articulation occurs in about 97.30% of relevant contexts, and less-natural sandhi in about 92.86%. In the perception test, 68.06% of tonal changes in less-natural sandhi are recognized, slightly higher than the 67.29% for co-articulation. The percentages of TCs closely align with results from sandhi processes triggered by corresponding falling tones. These results indicate that tonal changes preceding TCs are sandhi processes rather than lexicalized patterns.

Studying TCs, the active triggers for sandhi processes, facilitates a thorough exploration of tonal alternations. Meanwhile, despite the unclear metrical structure of Pingding TCs, their behaviour resembling corresponding falling tones in sandhi processes suggests a potential correspondence between a Pingding three-tone sequence TF-[Lq-TF] and a trisyllabic layered foot (Breteler & Kager 2022).

References

Barrie, Michael (2006), Tone circles and contrast preservation. *Linguistic Inquiry* 37(1), 131–141.

Beckman, Mary E & Janet B Pierrehumbert (1986), Intonational structure in Japanese and English. *Phonology yearbook*, vol. 1986, Cambridge / New York: Cambridge University Press, 255–309.

Breteler, Jeroen & René Kager (2022), Layered feet and syllable-integrity violations: The case of Copperbelt Bemba bounded tone spread. *Natural Language & Linguistic Theory* 40(3), 703–740.

Chen, Matthew (2000), *Tone sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

Chen, Matthew (1987), The syntax of Xiamen tone sandhi. *Phonology Yearbook*, vol. 4, Cambridge University Press, 109–149.

Jia, Pingping & Judith Meinschaefer (to appear), Reanalysis of two-tone sequences as tone compounds in the Pingding dialect of Chinese. *Investigationes Linguisticae* XLVII.

Levi, Susannah (2005), Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association*. Cambridge University Press. 35(1), 73–97.

Łubowicz, Anna (2011), Chain Shifts. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (Hrsg.), *The Blackwell Companion to Phonology*, John Wiley & Sons, 1717–1735.

Zhang, Jie (2022), Tonal Processes Defined as Tone Sandhi. *The Cambridge handbook of Chinese linguistics*, Cambridge University Press, 291–312.

Czech (expletive) negation

Anne-Li Demonie
(Masaryk University)

This paper investigates the Czech negator *ne-* and compares its 'normal' instantiations to those inadverbial clauses headed by *dokud* ('until'). *Ne-* generally gives rise to real negation (NEG), (1a), and licenses negative concord items (NCI), (1b).

(1)   a.   **Ne**-jsem šťastný / Jsem **ne**-šťastný
                Neg-am happy / am   Neg-happy
                'I'm not happy /I'm unhappy'
      b. **Ni**kdo **ni**koho **ne**viděl.
           nobody nobody Neg-saw
           'Nobody saw anyone'

In *until*-clauses *ne-* does neither, even though its presence is obligatory (2a-b). Such instances of negation are referred to as expletive negation (EN) (Dočekal 2012, Makri 2013, Cepéda 2018, Mari & Tahar 2020, Delfitto 2020 a.o.)

(2)   a. Petr   čekal,   dokud Pavel ***(ne)**-přišel.

          Peter waited.PAST.imPF   until Pavel Neg-came.PAST.PeRF
          'Peter was waiting until Pavel came.'
       b. Petr   čekal,   dokud **někdo**   ***(ne)**-přišel.

          Peter waited.PAST.imPF   until someone   Neg-came.PAST.PeRF
          'Peter waited until someone came.'

Previous accounts of NEG and EN provide two views. One argues that EN is semantically vacuous(Espinal 1992, 2000, 2007 a.o.) and contributes nothing. The other argues that EN corresponds either to NEG which is cancelled out (Abels 2005 a.o.), or to a modal (Makri 2013) or pragmatic (Delfitto et al. 2019, Delfitto 2020 a.o.) element which has an epistemic or evaluative function with regards to the proposition.

Given the morphological identity between both instances of *ne-*, their different syntactic be- haviour, and the obligatory presence of EN in *until*-clauses, my proposal follows the second type of analyses above. However, it deviates from them by arguing that *ne-* is a single exponent which can lexicalise NEG as well as EN, depending on the context. The analysis thus treats the identity between NEG and EN as a case of syncretism.

I formalise the analysis in Nanosyntax (Starke 2009), a late-insertion, morphological theory. In Nanosyntax, lexical items are links between phonology and a tree with multiple features. In late-insertion theories, morphemes may be inserted in various environments, corresponding to subconstituents in the item.
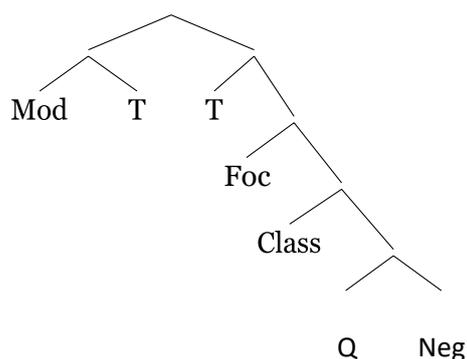
For ne- I adopt the features from De Clercq (2020) who argues for the hierarchy: Tense > Focus > Class > Q > Neg, (3). These features allow ne- to engage in constituent (CN) and sentential negation (SN) (cf. (1a)).

(3) ne-
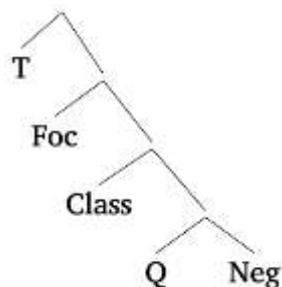
For EN, I use the possibility to have different shapes of lexical items, as in Blix (2021, 2022). I assume that the tree in (3) has a separate branch, which I call Modality (Mod) for now, (4).

(4)

The negator can lexicalise parts of its structure, depending on its function. The right branch takes care of CN and SN, the structure as a whole creates modal negation and the left branch corresponds to EN. The feature Mod is essentially a marker of epistemicity/undesirability/unlikelihood (Makri 2015, Delfitto et al. 2019), but a more fine-grained structure is still in order. Since EN does not realise Neg, it cannot license NCIs (cf. (2b)). A question which still needs answering is why *until*-clauses specifically prevents the right branch from lexicalising.

References

Abels, Klaus. (2005). Expletive Negation" in Russian: A Conspiracy Theory. *Journal of Slavic Linguistics 13*(1), 5–74. http://www.jstor.org/stable/24599547.

Blix, Haagen. 2022. Interface Legibility and Nominal Classification: A Nanosyntactic Account of Kipisigis Singulatives. *Glossa: a journal of general linguistics* 7(1). doi: https://doi.org/10.16995/glossa.582.

Blix, Haagen. 2021. Phrasal Spellout and Partial Overwrite: On an alternative to backtracking. Glossa: a journal of general linguistics. 6(1), 62. doi: https://doi.org/10.5334/gjgl.1614.

Cépeda, Paola. 2018. *Negation and Time. Against expletive negation in temporal clauses* [PhD dissertation]. *New York: Stony Brook University*.

De Clercq, Karen. 2020. *The Morphosyntax of Negative Markers: A Nanosyntactic Account.* Oxford: Oxford University Press.

Delfitto, Denis. 2020. Expletive Negation. In Viviane Depréz & M. Teresa Espinal (eds.), *The Oxford Handbook of Negation*, 255–268. Oxford: Oxford University Press.

Delfitto, Denis, Chiara Melloni & Maria Vender. 2019. The (en)rich(ed) meaning of expletive negation. *Evolutionary Linguistic Theory 1*(1), 57–89.

Dočekal, Mojmír. 2012. Czech *dokud* and telicity. In Ana Aguilar Guevara, Anna Chernilovskaya & Rick Nouwen (eds.), *Proceedings of Sinn und Bedeutung 16. Boston, MA: MIT Working Papers in Linguistics*, 183-197.

Espinal, M. Teresa. 2007. Li- censing expletive negation and negative concord in Catalan and Spanish. In Franck Floricic (ed.),*La négation dans les langues romanes. Linguisticae Investigationes Supplementa 26. vol. 1*, 49-74. John Benjamins Publishing Company.

Espinal, M. Teresa. 2000. Expletive negation, negativeconcord and feature checking. *Catalan Working Papers in Linguistics 8*. 47-69.

Espinal, M. Teresa. 1992. Expletive Negation and Logical Absorption. *The Linguistic Reviewm 9*(4), 333-358. https://doi.org/10.1515/tlir.1992.9.4.333.

Makri, Maria-Margarita. 2013. *Expletive Negation beyond Romance: Clausal Complementation and Epistemic Modality* [MA dissertation]. York: Uni- versity of York.

Mari, Alda & Chloé Tahar. 2020. On prohibitive and expletive negations. HAL (halshs-02865961).

Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Tarald Taraldsen (eds.), *Nordlyd 36: special issue on Nanosyntax*, 1–6. Tromsø: University of Tromsø. lingbuzz/001230.

Split intransitivity in Shughni

Daria Chistiakova
(HSE University, Moscow)

Keywords: split intransitivity, telicity, agentivity, actionality, unaccusativity diagnostics

This paper investigates split intransitivity (SI) in Shughni (Pamir group, Eastern Iranian languages). I follow Levin and Rappoport Hovav (1995) in the idea that SI is syntactically represented but semantically determined. In my work, I explore syntactic manifestations of SI in Shughni and demonstrate that SI correlates with the actionality of verbs. The data were obtained through elicitation from 23 informants during three expeditions. In compiling lists of unergative and unaccusative verbs and in examining actionality (in collaboration with Maksim Malenchenko), I asked informants to translate stimuli from Russian to Shughni using a targeted verb. In investigating morphological causative formation, I used information from the online-dictionary (Makarov et al. 2022).

The basis for postulating SI in Shughni is the distribution of the second-position 3rd singular subject enclitic *=i* in past tenses, noted in Edelman (1990): it is used with transitive (1) and some intransitive (unergative) verbs (2), while it is not used with the majority of intransitive (unaccusative) verbs (3).

(1) *yu=**yi**   mu   wīn-č*
     D3.M.SG=**3SG**   1SG.O   see-PF
     'He saw me'

(2) *yu=**yi**   lap   paloys-t*
     D3.M.SG=**3SG**   much   work-PST
     'He worked a lot'

(3) *yā(*=**yi**)   vo   sifā-d   pi   kū*
     D3.F.SG   again   climb.F-PST   UP   mountains
     'She climbed the mountains again'

SI has been studied in Shughni by Parker (2023) using four unaccusativity diagnostics. I show that one of them, morphological causative formation, is unreliable for Shughni, as it includes some unergative verbs and misses some unaccusatives.

Within the semantic approach, two main parameters are distinguished cross-linguistically: agentivity and lexical aspect (Van Valin 1990, Dowty 1991). Agentive and atelic verbs tend to be unergative, while non-agentive telic ones tend to be unaccusative. Other verbs' distribution depends on the more salient parameter in the language (see Table 1).

Table 1. Dowty 1991:607

	Atelic	Telic
Agentive	definitely unergative	?
Non-agentive	?	definitely unaccusative

I argue that in Shughni, SI is determined by lexical aspect: non-agentive atelic verbs are unergative, agentive telic verbs are unaccusative.

To confirm the correlation with telicity, I analyze the aspectual characteristics of 29 predicates within Tatevosov's (2016) framework, which defines lexical aspect in terms of possible interpretations for perfective and imperfective forms. In Shughni, there is no grammatical aspect, thus past and present tense are used in appropriate contexts, unambiguously representing perfective or imperfective aspects. Examples 4-5 show that *šīntow* 'laugh' is classified as atelic, with only a processual interpretation available.

(4)   PST

      uz=um ***šīn-t***

      1SG=1SG   laugh-PST

      'I was laughing <when my mom entered>'

      # '<I heard a funny joke> and laughed'

(5)   PRS

      uz   *ik=šīč*   ***šānd-um***

      1SG   EMPH=now   laugh-1SG

      'I am laughing now'

I demonstrate that unergative verbs encompass atelic and multiplicative ones, while telic, stative verbs, and verbs of position are unaccusative; I also analyze the distribution of predicates with unique aspectual characteristics. In the talk, I will show how this data corresponds to the linking rules proposed in Levin and Rappoport Hovav (2000) and developed in Arkadiev (2008).

I will also discuss additional aspects of SI: I analyze the behavior of complex verbs, demonstrating that unergatives are never used as light verbs.

References

Arkadiev, Peter (2008), Thematic roles, event structure, and argument encoding in semantically aligned languages, in *The typology of semantic alignment*, ed. Mark Donohue and Søren Wichmann, 101–120, New York: Oxford University Press.

Dowty, David (1991), Thematic Proto-Roles and Argument Selection, *Language*, *67*(3), 547–619.

Edelman, Dzhoy (1990), *Sravnitelnaya grammatika vostochnoiranskih yazykov. Morfologiya. Elementy sintaksisa* [Comparative grammar of Eastern Iranian languages. Morphology. Syntax elements], Moscow: Nauka.

Levin, Beth, & Rappaport Hovav, Malka (1995), *Unaccusativity: At the syntax-lexical semantics interface*, Cambridge, MA: MIT Press.

Levin, Beth, & Rappaport Hovav, Malka (2000), Classifying Single Argument Verbs, in P. Coopmans, M. Everaert, and J. Grimshaw, eds., *Lexical Specification and Insertion,* John Benjamins, Amsterdam, 269-304.

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 Language Resources and Evaluation Conference*, 61–64, (pamiri.online), 2020–2023.

Parker, Clinton (2023), *A Grammar of the Shughni Language*, PhD Thesis, McGill University.

Tatevosov, Sergei (2016), *Glagolnye klassy i tipologiya aktsionalnosti* [Verb Classes and the Typology of Aktionsart] Moscow: Yazyki slavyanskoi kultury.

Van Valin, Robert (1990), Semantic Parameters of Split Intransitivity, *Language*, Vol. 66 (2), 221-260.

Cross-linguistic differences in pragmatic inference: A comparison of Chinese and German anaphora resolution in discourse comprehension

Linlin Sun

(Huazhong Agricultural University)

Compared to many European languages, Chinese is recognized as a language whose production and comprehension are highly dependent on semantic and contextual information rather than strict syntactic rules/cues (e.g., LaPolla 1995). It has been argued that a distinctive feature of languages like Chinese is the involvement of pragmatics in linguistic domains that are governed by syntax in languages such as English, German or French (Huang 1994).

This study investigated the pragmatic inference of native Chinese speakers and native German speakers by comparing their anaphora resolution during discourse comprehension. We chose the Chinese zero anaphora and the German personal pronoun *er* 'he' for our experiments, given that they are functionally comparable (Huang 2003), their referents have the same cognitive status (Gundel et al. 1993), and that they are the most reduced/unmarked and inference-triggering anaphoric devices in their respective languages (Neeleman & Szendrői 2007) and thus best reflect the extent to which the languages allow the use of pragmatic inference. We tested the hypothesis that the different degrees of reliance/openness to pragmatics, as opposed to syntax, in Chinese versus German grammars result in their speakers using pragmatic inference to resolve anaphora differently during discourse comprehension (cf. Tao & Hearly 2005).

Using the same 112 story paragraphs in a Chinese and a German version, our parallel experiments examined the anaphora resolution in the subject position of two consecutive sentences (the second and third sentences) in a story paragraph, respectively. These story paragraphs followed Labov and Waletzky's (1967) narrative structure, where the first sentence introduces the setting of a story with two participants (exposition), followed by two sentences describing the major actions of the story that change it from one state to another (from complication to resolution). The test conditions were subject-continuation versus subject-shift in the two consecutive sentences (2 x 2 design). Participants were asked to complete a judgment task quickly and accurately after reading the two critical sentences of each paragraph, respectively. The experiments collected participants' reaction times and accuracy on both tasks. Data were analyzed using generalized linear mixed-effects models.

The results confirmed our hypothesis and indicated cross-linguistic differences in how speakers use pragmatics-based and syntax-based strategies. The resolution of Chinese zero suggested a strategy built on story development, where the anaphora assignment was biased towards the component that is most likely to drive the storyline. This pattern of inference is based on world knowledge, semantic and contextual cues that override syntactic or structural influences such as the first-mention effect or subject preference (pragmatics/discourse > syntax). In contrast, the German pronoun resolution showed that syntax was stronger than pragmatics/discourse in the first critical sentence (syntax > pragmatics/discourse), where the anaphora assignment was biased towards the first-mentioned, subject/agent antecedent (e.g., Bouma & Hopp 2007), and that this sensitivity of native speakers to syntactic and structural rules in turn affected their inference based on world knowledge, semantic and

contextual cues.

References

Bouma, Gerlof and Holger Hopp (2007), Coreference preferences for personal pronouns in German, *ZAS Papers in Linguistics 48*, 53–74.

Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski (1993), Cognitive status and the form of referring expressions in discourse, *Language 69*, 274–307.

Huang, Yan (1994), *The Syntax and Pragmatics of Anaphora*, Cambridge: Cambridge University Press.

Huang, Yan (2003), *Anaphora: a cross-linguistic study*, Oxford: Oxford University Press.

Labov, William and Joshua Waletzky (1967), Narrative analysis: Oral versions of personal experience, in June Helm (ed), (1967), *Essays on the Verbal and Visual Arts*, Seattle: University of Washington Press, 12–44.

LaPolla, Randy J. (1995), Pragmatic relations and word order in Chinese, in Pamela A. Downing and Michael Noonan (eds), (1995), *Word Order in Discourse*, Amsterdam / Philadelphia: John Benjamins, 297–329.

Neeleman, Ad and Kriszta Szendrői (2007), Radical pro drop and the morphology of pronouns, *Linguistic Inquiry 38*, 671–714.

Tao, Liang and Alice F. Healy (2005), Zero anaphora: transfer of reference tracking strategies from Chinese to English, *Journal of Psycholinguistic Research 34*, 99–131.

Cultural variation in the instantiation of the conceptual prototype of lying

Alejandro Erut

(University of Texas at Austin)

Keywords: Lying, prototypes, cross-cultural, intent, deception

The centrality of concepts like "deception," "lying," "bullshit," and "hoax" in the public discourse of the last decades, as well as in the infodemic associated with COVID-19, motivated a fresh interest in the scientific community to determine the semantics, pragmatics, and intuitions that lie behind these concepts. Using the structure of the "Geography of Philosophy Project," I investigated lying intuitions in 14 cultural and linguistic communities: Afrikaans, Chinese Mandarin, Ecuadorian Spanish, Hindi, isiZulu, Japanese, Korean, Moroccan Arabic, Peruvian Spanish, Sepedi, Slovak, Tamil, Ukrainian, and US English. The participants (N=1,556) were presented with eight scenarios based on the work of Coleman and Kay (1981) in which they were asked to judge if the main character lied ("Did x lie?"). The structure of the stimuli was based on a full-factorial design that articulates a) the factual component of the speaker's assertion, b) the epistemic status of the beliefs of the speaker, and c) the intent of the speaker (e.g., to deceive the addressee). The results show that the conceptual prototype of lying is, to some degree, stable across cultures. However, only some linguistic communities showed a hierarchical organization of the subcomponents of the prototype. In addition, communities in which the prototype subcomponents are hierarchically organized differ in how this hierarchy is instantiated.

References

Coleman, L., and Kay, P. (1981). Prototype Semantics: The English Word Lie. *Language*, *57*(1), 26–44

Elucidating complex inflectional analogies
via a holistic, diachronic and geolinguistic approach

Louise Esher

(CNRS; LLACAN)

Keywords: inflectional analogy, historical linguistics, diachrony, dialectology, Romance linguistics

In Foissenc Occitan (Gallo-Romance), the preterite desinences of all verbs are known to develop from mediaeval 'weak' (non-root-stressed) preterites (Table 1). Existing accounts, based on comparison of modern dialect forms with mediaeval standard forms, assume that the Foissenc preterite results from an analogical innovation in which the third person singular form in –[ˈɛk] acts as a model for all other person forms (Ronjat 1937, Esher 2021). Bybee & Brewer (1980) notably interpret the Foissenc development as offering empirical support for a theory of inflectional relationships in which the third person is cognitively more basic and influential than the other persons.

	mediaeval 'weak'	modern Foissenc
1SG	kanˈtjɛj	kanˈtɛgi
2SG	kanˈtjɛst	kanˈtɛges
3SG	kanˈtɛk	kanˈtɛk
1PL	kanˈtem	kanˈtɛgen
2PL	kanˈtɛts	kanˈtɛgets
3PL	kanˈtɛren	kanˈtɛgen

Table 1. Preterite forms of the majority inflectional class (illustrated for *cantar* 'sing') in mediaeval standard Occitan and modern Foissenc Occitan (Ronjat 1937, Skårup 1997).

	aver 'have'	*voler* 'want'	*poder* 'be able'
1SG	*aˈgui*	*volˈgui*	*poˈgui*
2SG	*aˈguist*	*volˈguist*	*poˈguist*
3SG	*ac*	*volc*	*poc*
1PL	*aˈguem*	*volˈguem*	*poˈguem*
2PL	*aˈguetz*	*volˈguetz*	*poˈguetz*
3PL	*ˈagren*	*ˈvolgren*	*ˈpogren*

Table 2. Examples of strong preterites with a root-final velar in mediaeval Occitan (Skårup 1997).

A particularity of the Foissenc preterite, not predicted by the traditional analogical account, is the alternation between -[ɛk] in the third person singular preterite and -[ɛg]- in all other persons: intervocalic [k] is licit in Occitan, yet there are no attestations of preterites in *-[ˈɛki] 'I sang', *-[ˈɛkes], etc.

By combining historical, inflectional and geolinguistic data, the models and mechanisms involved can be precisely identified. The modern Foissenc preterite desinences become established from the mid-fourteenth century onwards (Anglade 1919). At this early period, there exists a small inflectional class of high-frequency lexemes characterised by 'strong' (partially root-stressed) preterite forms. Following perceptual labial-to-velar change (Lief 2004), multiple 'strong' preterites show alternation between [k] in the third person singular and [g] in all other persons, parallel with the alternation found in the modern preterite desinences (Table 2).

A more complex analogical relationship drawing on rich networks of inflectional relationships between the preterite forms of different persons and different inflectional classes can thus be inferred. The final [k] shared by third person singular forms in almost all lexemes (weak and strong alike) acts not as a model but as a pivot, on the basis of which speakers spread the [k]~[g] alternation characteristic of velar strong preterites into the corresponding person forms of weak preterites.

The joint role of the strong preterites and the third person singular is confirmed by the absence of velar preterite desinences from other Occitan varieties, which historically had third person singular preterite form in -[ɛt]: such forms could never have mediated the introduction of a [k]~[g] alternation, and no alternations involving [t] were available for introduction.

In this way, careful attention to longitudinal study of the historical textual record, and to the inflectional system as a complex whole, improves the accuracy and detail of analyses of inflectional analogy, at both an empirical and a theoretical level.

References

Anglade, Joseph (ed.), (1919). *Las Leys d'Amors*, Toulouse: Privat.

Bybee, Joan & Mary Brewer (1980), Explanation in morphophonemics: Changes in Provençal and Spanish preterite forms, *Lingua* 52, 201–242.

Esher, Louise (2021), Hypercorrection and velar-to-labial change in Occitan preterites. *Zeitschrift für romanische Philologie* 137, 61–98.

Lief, Eric (2004), Proto-Romance *[w] and the velar preterites. In J. Auger, J. C. Clements, and B. Vance (eds.), *Contemporary approaches to Romance linguistics*, Amsterdam: Benjamins, 257–274.

Ronjat, Jules (1937), *Grammaire istorique des parlers provençaux modernes*, vol. 3, Montpellier: Société des Langues Romanes.

Skårup, Povl (1997), *Morphologie élémentaire de l'ancien occitan*, Copenhagen: Museum Tusculanum Press.

(Co)-Subordination Strategies in Sa

Emma Mazzuchi
(University of Oregon)

This paper discusses the different subject (DS) marker *-ndna* in Sa [pup], a Rai Coast (< Madang < Trans New Guinea) language from Papua New Guinea (Pawley & Hammarström 2018). As is common in Papuan languages (Roberts 1997), Sa exhibits clause chaining and switch-reference. Switch-reference indexes the identity or non-identity of the marked verb's subject with that of the following verb (van Gijn 2016). While Sa same subject (SS) marking is straightforward, different subject (DS) has two separate forms, here termed 'coordinate,' and 'subordinate,' marked respectively with the suffixes *-nde* and *-ndna*. I have termed *-ndna* 'subordinate' because it marks the type of background information one might expect to find in a subordinate clause (Cristofaro 2003, Reesink 2014) and because I believe that it arose as a subordinate clause. However, this paper argues that *-ndna* has completed the majority of a grammaticalization path from topicalization, through subordination, to switch-reference.

The main function that separates *-ndna* from *-nde* is that *-ndna* often marks backgrounded information. Combining switch-reference with a foreground/background distinction is uncommon; in fact, much of the literature on Papuan-style SR indicates that clauses with background information would likely be "skipped" by the SR system, either going unmarked or receiving whatever SR marking their corresponding main event line verb took (Reesink 2014, van Gijn 2016). While a few other descriptions of Papuan languages have noted that some information is relegated to 'subordinate medial' clauses, (Pennington 2016, MacDonald 1988), the type of information in these clauses tends to be more limited, (eg temporal information, weather verbs). Sa follows neither of these systems, having special SR marking for a much wider variety of possible backgrounded information.

For example, (1) comes from a story wherein a spirit kills a person. The killing was mentioned in the previous sentence, then detailed in (1). The clauses providing the extra, background information are marked with *-ndna*

1. | *Hai* | *si* | *imba* | *punu-ndna* | | *ia* | *no* | *gama* | *tuma-ndna* | *kuma-ndu* |
    |-------|------|--------|-------------|---|------|------|--------|-------------|------------|
    | Stone | cook | finish | give.to.3SG-DS.SUB | | eat | 3SG | lung | cook-DS.SUB | die-SG.FPST |

    'She cooked a stone and fed it to him and his lungs cooked and he died.'

The etymology of *-ndna* is fairly transparent: coordinate DS marking + the topicalizer *na*, which, at least historically, formed a subordinate clause. The question that follows from this etymology is whether *-ndna* still marks syntactic subordination as its source construction did. I will therefore also present data from a variety of syntactic tests that show that *-ndna* clauses do not always behave like subordinate clauses. For example, *-ndna* clauses do not always agree in negation with their corresponding final verb, but they do agree in mood. I will compare these data to other work done in the family on subordinate clauses (eg MacDonald 1988). Finally, since the data presented here are from a pilot trip, I will also illustrate gaps in the data set and discuss the kinds of syntactic questions I plan to ask when I next return to the field to further examine the syntactic status of *-ndna* clauses.

References

Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.

Van Gijn, R. (2016). Switch reference. *Typological Studies in Language*, 1-54.

Haiman, John & Pamela Munro. 1983. Introduction. In John Haiman & Pamela Munro (eds.), Switch reference and Universal Grammar, ix–xv. Amsterdam: John Benjamins.

MacDonald, L. (1988). Subordination in Tauya. *Clause combining in grammar and discourse*, 227-246.

Pawley, Andrew & Harald Hammarström. 2018. The Trans New Guinea family. In Bill Palmer (ed.), The languages and linguistics of the New Guinea area: A comprehensive guide, 21–195. Berlin: De Gruyter Mouton.

Pennington, R. (2016). *A Grammar of Ma Manda a Papuan Language of Morobe Province, Papua New Guinea* (Doctoral dissertation, James Cook University).

Reesink, G. (2014). Topic management and clause combination in the Papuan language Usan. *Information structure and reference tracking in complex sentences*, *105*, 231.

Roberts, John R. 1997. Switch-reference in Papua New Guinea: A preliminary survey. In Andrew Pawley (ed.), Papers in Papuan linguistics no. 3 (Pacific Linguistics A 87), 101–241. Canberra: Pacific Linguistics.

Duality of pragmatic markers in linguistic borrowing: The pragmatic borrowing of *sir/madam* as vocative markers among Hong Kong police

Chuang Liu & Jueyu Hou
(Leiden University & Leiden University)

Vocative markers fulfil the pragmatic function of addressing without affecting the propositional contents of what is said (Fraser, 2009; 2013). Their discursive structures and pragmatic functions can be integrated from the source language to the target language in language contact (Matras, 2009). The present study unpacks how English words *sir* and *madam* as vocative markers are integrated into the Hong Kong Cantonese (HKC) vocative terms used among police officers and answers: 1) How are they integrated into the discursive structures of vocative markers of HKC? 2) What is the division of labour of the loan vocatives sir and madam and other inherited vocative markers?

The study collected 116 contextualised utterances from in total six relevant Hong Kong films and TV dramas, which were all written and spoken by native HKC speakers, and identified 88 with *sir/madam-* vocative markers. We first analysed the structures of *sir/madam-* vocative markers, and then their uses according to their division of labour with inherited vocative markers in contexts with different risk of face threat (Brown & Levinson, 1987).

The analyses revealed that, first, *sir* and *madam* are asymmetrically integrated in complex vocative markers in combination with family names. Whereas *sir*-vocatives are perfectly integrated into the HKC family name-first discursive structure (cf. Ex.1 a—b), the integration of *madam*-vocatives is not complete (cf. Ex.1 c—d), which can be attributed to the fewer number and shorter history of female officers in HK police.

(1) a   程 sir   *Cing4 [family name]-sir*   'officer Ing'
     b   罗护士   *Lo4 [family name]-wu6si6*   'nurse Lo'
     c   madam 方   *madam-Fong1 [family name]*   'officer Fong'
     d   Mr. [Title] Holmes [Family name]

Second, the integrated vocative markers under study are more commonly used in circumstances with greater risk of face threat, but different contextual variables (e.g., social distance, absolute ranking) have different predictability in their uses (Table 1). Moreover, if and only if the addressee has a management position, rather than an ordinary basic-level police sergeant, can *sir/madam-* vocative markers definitely be used.

Hypotheses; Contextual variables	Positive cases (with example)	Negative cases (with example)
Sir/madam- vocative markers are used when S is at an inferior status; Absolute ranking/[+R]	56 (Ex. 2)	32 (Ex. 3)

Sir/madam- vocative markers are used when there is a larger social distance between S and H, or a larger social distance is required; Social distance/[+D(H,S)]	53 (Ex. 4)	35 (Ex. 5)
Sir/madam- vocative markers are used when H is at a management position (i.e., not an ordinary police sergeant)	88 (Ex. 2–5)	0 (N/A)

Table 1. Predictability of different contextual variables in the uses of *sir/madam*- vocative markers

(2)   [Sub-unit commander, S, is reporting the progress of her work to her boss, the director of the anti-drug dealership department, H]

S: '巩 **sir**, 你怀疑邓国彬已经变节？'

　　　　Gung2-sir, nei5 waai4ji4 Dang6Gwok3ban1 ji5ging1 bin3zit3?

　　　　Gung-sir, 2SG suspect Dang Gwokban already betray?

'Are you suspecting that Dang Gwokban has already betrayed us, **Gung-sir**?'

H: [Looking at S, sighs]

S: '那你打算怎么做?'

　　　　Naa5 nei5 daa2syun3 zam2mo1 zou6?

　　　　Then 2SG plan   how   do?

'So how will you deal with it?'

H: '你是 handler, 你认为怎么做就怎么做.'

　　　　Nei5 si6 handler, nei5 jing6wai4 zam2mo1 zou6 zau6 zam2mo1 zou6.

　　　　2SG be handler, 2SG think   how   do   then how   do.

'You are the handler, and you can decide.'

S: 'Yes **sir**, 有什么进展再同您汇报.'

　　　　Yes sir, jau5 sam6mo1 zeon3zin2 zoi3 tung4 nei5   wui6bou3.

　　　　Yes sir, have what   progress then to   2SG-HON report.

'Yes, **sir**. I will come back to you once I make some progress.'

　　　　　　　　　　　　　　　　　　　　　　　　　(*Lives of Omissions*, Episode 12)

(3)   [When the chief inspector of anti-bomb squad, S, is sending files to another chief inspector of anti-terrorism squad, H]

S: '**Madam**, 有东西给你.'

　　　　Madam, jau5 dung1sai1 kap1 nei5.

　　　　Madam, have thing   give   2SG.

'**Madam**, here's something for you.'

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　(*Shockwave II*)

(4)   [An inspector, S, is talking with a close friend, H1, when a phone call from another inspector, H2, working in the same office cuts in, reminding S to fetch him some food for dinner.]

S [to H1]: '**Ocean** [H1's nickname], 你冇事吧?'

Ocean, nei5 mou5 si6   baa1?
Ocean, 2SG NEG   matter PART?

'**Ocean** [H1's nickname], are you still all right?'

[A phone call cuts in]

S [to H2]: '喂, 程 **sir**, 外卖已经到了? 我马上去拿.'

Wai3, Cing4-sir, ngoi6maai6 ji5ging1 dou3 liu5,
Hello, Cing-sir, food   already   arrive PERF?
ngo5 maa5soeng6 heoi3 naa4.
1SG   right-away   go   fetch.

'Yes? **Cing-sir**, your dinner has just arrived? OK, I'm fetching it for you right away.'

(*The Invisibles*, Episode 1)

(5)   [In the party for an inspector, H, after he received an award, his friends, inspectors from other departments, S1, S2, S3, are congratulating H]

S1   '章 **sir**, coffee 就没有了, 就喝这个吧.'

Zoeng1-sir, coffee zau6   mut6 jau5 liu5, zau6 hot3 ze5 go3 baa6.
Zoeng-sir, coffee already NEG   have PERF, then drink this CL PART.

'**Zoeng-sir**, they have run out of coffee. Have a glass of water then!'

H   [Takes the glass of water]

S2   '恭喜啊, 章 **sir**!'

Gung1hei2   aa1,   Zoeng1-sir!
Congratulation PART,   Zoeng-sir!

'Congratulations, **Zoeng-sir**!'

H   [Nods, smiling]

S3   '来干杯, 章 **sir**!'

Loi4   gon1bui1, Zoeng1-sir!
Come   cheers,   Zoeng-sir!

'Cheers, **Zoeng-sir**!'

(*Shockwave I*)

The above observations may indicate a duality of *sir/madam-* vocative markers. On one hand, sir/madam- vocative markers bracket out the propositional contents of what is said (Fraser, 2009), and are therefore pragmatically detachable (Matras, 1998; 2009) from the rest of the utterance, so following Matras's (2009) suggestions, the integration of their pragmatic functions into HKC, the target language, is worthy probing into, rather than the concept they encode. On the other, although *sir/madam-* vocative markers are not referential, and therefore, according to Fraser (2009), have no content meaning, they do encode concepts, and the concepts they encode are held accountable for their division of labour with their inherited counterparts.

This duality implies an overlap (Figure 1) between lexical borrowing (Haspelmath, 2009) and pragmatic borrowing (Andersen, 2014; Matras, 1998; 2009). Lexical borrowing focuses on the concepts loanwords encode, and pragmatic borrowing, on the contrary, focuses more on the non-referential pragmatic functions, but these two focuses are not complementary: there are pragmatic markers that

neither encode a content meaning, nor "gesture-like" (Matras, 2009: p. 193), with no conceptual meaning. This overlap should also be accounted for by taking into consideration the duality of these borrowed items.
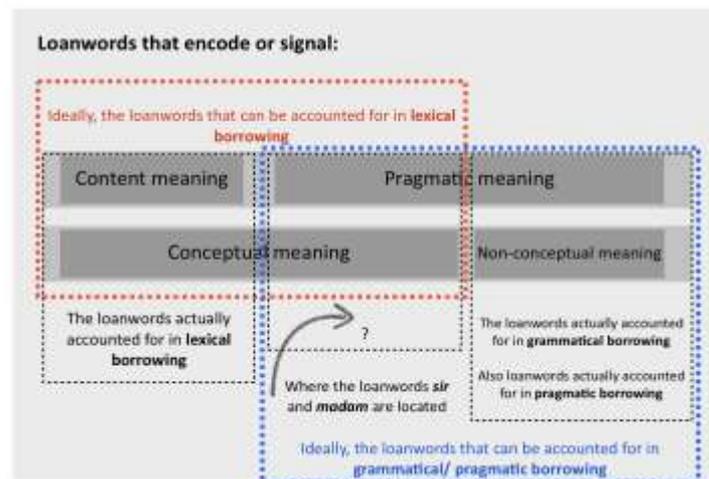

Figure 1. The landscape of lexical borrowing and pragmatic borrowing with regard to lexical items.

References

Andersen, Gisle, Furiassi, Cristiano, & Ilić, Biljana M. (2017). The pragmatic turn in studies of linguistic borrowing. *Journal of Pragmatics, 113*, 71—76. https://doi.org/10.1016/j.pragma.2017.03.010

Andersen, Gisle. (2014). Pragmatic borrowing. *Journal of Pragmatics, 67*, 17–33. https://doi.org/10.1016/j.pragma.2014.03.005

Brown, Penelope, & Levinson, Stephen. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Fraser, Bruce. (2009). An account of discourse markers. *International Review of Pragmatics, 1*(2), 293–320. https://doi.org/10.1163/187730909X12538045489818

Fraser, Bruce. (2013). Combinations of contrastive discourse markers in English. *International Review of Pragmatics, 5*(2), 318–340. https://doi.org/10.1163/18773109-13050209

Haspelmath, Martin. (2009). Lexical borrowing: Concepts and issues. In: M. Haspelmath, & U. Tadmor (eds.), *Loanwords in the World's Languages: A Comparative Handbook* (pp. 35—54). Berlin: De Gruyter Mouton.

Matras, Yaron. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics, 36*, 281— 331.

Matras, Yaron. (2009). *Language contact.* Cambridge: Cambridge University Press.

Interacting semantic preferences: The belief-intent alternation in Basque

Silvie Strauß

(Universität Leipzig)

Keywords: complement clauses, attitude predicates, complement alternations, Basque, morphosyntax-semantics interface

The fact that the type of complement clause (CC) depends on the semantics of the clause-embedding predicate (CEP) has been observed by several authors (e.g. Givón 1990; Cristofaro 2003; Schmidtke-Bode 2014; Wurmbrandt/Lohninger 2023). When the same verb can combine with two different types of CCs yielding different meanings, it becomes even more visible how the meanings of the CEP and the CC interact in order to determine the meaning of the whole construction.

Basque, like English (Jackendoff 1985; Grano 2019), has a class of CEPs which can combine either with a CC expressing a belief or a CC expressing an intention. Typically, finite indicative clauses express beliefs, while verbal nouns (VN) express intentions, but in principle both indicative clauses and VNs in the absolutive or instrumental can express both.

A study of the patterns of eight such alternating verbs (*gogorarazi* 'remind', *pentsatu* 'think', *ukatu* 'deny', *konbentzitu* 'convince', *bururatu* 'occur', *zin egin* 'swear', *erabaki* 'decide' and *proposatu* 'propose') based on data from the Basque *Corpus of Contemporary Texts* (ETC) as well as speaker judgements reveals that the possibility of atypical form-meaning pairs, namely VNs expressing belief and indicative clauses expressing intent, depends on the CEP. Some CEPs like *erabaki* 'decide' below are "intent-prominent", they allow intent readings for both VN and indicative clauses (1), and belief readings only for indicative clauses (2):

(1) a. *[hiri-a-n   gera-tze-a-Ø]   **erabaki-Ø**   d-u-Ø*
       city-SG-INE stay-NMLZ-SG-ABS **decide-PFV** 3SG.ABS-AUX.TR-3SG.ERG (ETC: Berria, 2017-03-15)

    b. *[hiri-a-n   gera-tu-ko   d-ela]   **erabaki-Ø**   d-u-Ø*
       city-SG-INE stay-PFV-FUT AUX.ITR.3SG-COMP **decide-PFV** 3SG.ABS-AUX.TR-3SG.ERG
       'S/he decided to stay in the city.'

(2) a. *[etxe   horr-etan […]   gizonezko-a-Ø   bizi   d-ela]   **erabaki-Ø***
       house   DEM.MED-INE   male-SG-ABS   live   AUX.ITR.3SG-COMP   **decide-PFV**
       d-u-Ø
       3SG.ABS-AUX.TR-3SG.ERG   (ETC: *Eulien bazka*, Hasier Etxeberria)

    b. *#[etxe   horr-etan   gizonezko-a-Ø   bizi-tze-a-Ø]   **erabaki-Ø**   d-u-Ø*
        house   DEM.MED-INE   male-SG-ABS   live-NMLZ-SG-ABS   **decide-PFV** 3SG.ABS-AUX.TR-3SG.ERG

       'She decided that it is a man […] who lives in that house.'

"Belief-prominent" verbs, on the other hand, are the exact mirror image, allowing belief readings for both and intent readings only for VN clauses, as summarised in Table 1.

CC \ CEP	belief-prominent	intent-prominent
finite	B	B/I
non-finite	B/I	I

Table 1: Readings resulting from combinations of CEP and CC.

Belief- or intent-prominence of a CEP is directly correlated with the higher frequency of the respective reading relative to the other in the corpus. Thus, when a CEP that is more likely to have a belief complement combines with an indicative CC, which is also more likely to express belief, the only possible interpretation is belief. If it combines, however, with a non-finite CC, which is more likely to express intent, it is either the preference of the CEP that "wins", yielding a belief reading, or the preference of the CC, yielding an intent reading. These two possible outcomes depend to a certain extent on the additional factors tense, aktionsart and coreference of the embedded agent with a matrix argument, which may render the intent reading incompatible or at least less plausible.

The patterns of these alternating verbs thus show how the choice of CC type is not simply a consequence of the meaning of the CEP, but both the CEP and the CC have semantic profiles whose interaction determines the meaning of the whole construction, which in turn can be further influenced by the semantic content of the CC.

References

Cristofaro, Sonia (2003): *Subordination*. Oxford: Oxford University Press.

Egungo Testuen Corpusa (= ETC) [Corpus of Contemporary Texts] (2021): https://www.ehu.eus/etc/.

Givón, Talmy (1990): *Syntax: A functional-typological introduction.* Vol. II. Amsterdam: John Benjamins.

Grano, Thomas (2019): "Belief, intention, and the grammar of persuasion." *Proceedings of the Fifty-fourth Annual Meeting of the Chicago Linguistic Society*: 125-136.

Jackendoff, Ray (1985): "Believing and intending: Two sides of the same coin." *Linguistic Inquiry* 16. 445-460.

Schmidtke-Bode, Karsten (2014): *Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization.* University of Jena, doctoral dissertation.

Wurmbrandt, Susanne / Lohninger, Magdalena (2023): "An implicational universal in complementation – theoretical insights and empirical progress." In Hartmann, Jutta M. / Wöllstein, Angelika (eds.): *Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues*. Tübingen: Narr Francke Attempto. 183-229.

Factual information as a window into Balkan evidentials

Vesela Simeonova, Izabela Jordanoska, Orkida Backus Borshi & Seckin Arslan
University of Graz, Max Planck Institute for Psycholinguistics, Czechia Charles University &
Université Côte d'Azur; CNRS)

Evidentiality - linguistic reference to information source - has been subject to a great deal of interest from a number of theoretical perspectives, mostly based on spoken data focusing on what a speaker utters. We examine how evidential forms are used in written factual information, such as reporting of news (journalism), historical facts, and police reports, with novel data from Bulgarian and Macedonian (Slavic), and Turkish (Turkic), all of which belong to the Balkan Sprachbund. The evidential system in these languages exhibits the typical split of the so-called Eurasian areal evidential belt broadly differentiating between witnessed and non-witnessed events (Aikhenvald 2004, Friedman 2003). Balkan evidentials are obligatorily marked on past verbs using similar linguistic strategies.

We would expect that evidentials of the same type would have similar properties and distribution in identical communicative genres across the Balkan languages, where the type of evidence and the degree of certainty are the same. Importantly, because we have chosen communicative genres with non-witnessed sources of information but high degree of certainty. This comparative analysis has implications for the theoretical models of evidentiality and its interaction with epistemic modality signaling degree of certainty. We propose that marking of evidentiality has different degrees of interaction with degree of certainty exposed by the communicative genres (e.g., when discussing historical facts about the Roman empire, the present speaker does not have direct evidence but makes different evidential choices in Bulgarian, Macedonian or Turkish). We show that the data turn out to be much more complex and the distribution of evidentials across or even within the languages of this study is not uniform.

We chose three environments to compare across the languages with sample data from naturally occurring sources: (1) reporting on accidents in local (i.e., not translated) news outlets in the respective language; (2) Historical facts (taken from Wikipedia for comparability); (3) police reports.

We find a great deal of diversity among the strategies used by the different languages, summarized in Table 1. We find that in identical environments, the languages in this study used different evidential strategies even despite the superficial morphological and areal similarity among their evidential systems. This suggests that different languages with evidentiality partition the epistemic space of certainty differently and that there is no one-to-one mapping between the type of evidential source and the reliability of the information. Our findings can inform various theories of evidentiality and epistemic modality (Aikhenvald 2004, de Haan 2001, van der Auwera and Plungian 1998). These findings will also serve as a starting point for better understanding evidentiality in complex environments in other languages as well.

Table 1. Summary of findings

Genre/language	Bulgarian	Macedonian	Turkish
News reports on accidents	Indirect	Indirect	Direct
Historical facts	Historical present	Indirect	Direct or Epistemic / Habitual form
Police reports	Indirect	Indirect	Direct

References

Aikhenvald, Alexandra Y. (2004), *Evidentiality*. OUP: Oxford.

De Haan, Ferdinand (2001), The relation between modality and evidentiality, *Linguistische Berichte*, 9, 201-16.

Friedman, Victor A. (2003), Evidentiality in the Balkans with special attention to Macedonian and Albanian. In R.M.W Dixon and A. Y. Aikhenvald (eds), *Studies in evidentiality*, Amsterdam: John Benjamins, 189-218.

Van der Auwera, Johan, and Plungian, Vladimir A. (1998), Modality's semantic map. *Linguistic Typology* 2(1), 79-124,

Productivity in contrast to creativity or productivity in accordance with creativity?

Pavel Stekauer & Livia Kortvelyessy
(PreŠov University & Šafárik University)

In traditional accounts, productivity and creativity are viewed in opposition: productivity is grammatical, and its degree can be computed either as a proportion between the actual and the potential words (Aronoff 1976) or through hapax legomena (Baayen and Lieber, and a series of Baayen's works in the following years); creativity is extra-grammatical (Dressler & Merlini Barbaresi 1994, Dressler 2005; Mattiello 2013), it is based on aberration (Berg 2019) and therefore the meaning of creative formations is unpredictable. The latest contribution to this field is the conception of extravagant morphology (Ettelmann and Haumann 2022) that captures word-formation processes characterized by constraint violations, interface phenomena as well as borderline phenomena.

In contrast to the indicated opposition between productivity and creativity, the present paper offers a different view in which productivity and creativity are "reconciled" within a cognitively founded onomasiological model (Štekauer 2005). The model encompasses both the 'grammatical' productivity and 'extra-grammatical' creativity.

Regarding productivity, this model does away with the limitations of the traditional computation models of productivity that are restricted to affixation processes. Productivity, viewed from the cognitive perspective, is conceived as a *competition* between any and all word- formation processes or rules that can be employed to form a new complex word to express a particular semantic category, such as Agent, Instrument, Manner, Location, etc. By comparing various word-formation processes or rules, it is possible to objectify their respective role and, therefore, the degree of productivity as their respective share of all the complex words that fall within a particular semantic category.

Regarding creativity, the onomasiological model conceives of each naming act, irrespective of its degree of productivity, as a creative act for two reasons: (i) decisions at each level of the naming act (conceptual, semantic, onomasiological, and onomatological) depend on a coiner's creative performance preconditioned by their creative potential (Körtvélyessy, Štekauer and Kačmár (2022); (ii) new complex words meet all psychologically defined principles of creativity: novelty, originality, appropriateness, relevance, usefulness, quality, and effectiveness – all this as a result of deliberate cognitive activity. Word-formation creativity manifests the universal, biologically preconditioned feature of human beings (D'Agostino 1984), thanks to which any and all speakers of a language can produce a new word. Moreover, if we accept the view that the product of creativity has to be something "different, new, or innovative" (Kaufman and Sternberg 2019), each new complex word meets these criteria because each such new coinage is different from the existing actual words and, by definition, it is innovative with regard to the naming needs of a speech community. The criterion of quality is guaranteed by the acceptance of a new complex word by a speech community and its use for communication purposes.

The paper thus provides a *unified theoretical account* of productivity and creativity within the *framework of an onomasiological model* of word-formation.

References

Aronoff, Mark. 1976. *Word-formation in generative grammar*. Cambridge/London: The MIT Press.

Baayen, Harald R.. and Rochelle Lieber. 1991. Productivity and English derivation: a corpus- based study. *Linguistics* 29: 801-843.

Berg, Alexander. 2019. What, If Anything, Is Linguistic Creativity? *Gestalt Theory* 41(2): 173- 184.

D'Agostino, Fred. 1984. Chomsky on Creativity. *Synthese* 58: 85-117.

Dressler, Wolfgang U. 2005. Word-formation in Natural Morphology. In Pavol Štekauer and Rochelle Lieber (eds.), *Handbook of Word-Formation*, 267-284. New York: Springer.

Dressler, Wolfgang U. and Merlini Brabaresi, Lavinia. 1994. *Morphopragmatics*. Berlin: Mouton de Gruyter.

Ettelmann, Matthias and Dagmar Haumann. 2022. *Extravagant morphology.* Amsterdam/Philadelphia: John Benjamins.

Kaufman, James C. and Robert J. Sternberg. 2019.Preface." In J. C. Kaufman and R. J. Sternberg (eds.), *The Cambridge Handbook of Creativity*. Cambridge: Cambridge University Press.

Körtvélyessy, Lívia, Pavol Štekauer and Pavol Kačmár. 2022. Creativity in word-formation and word- interpretation. Creative potential and creative performance. Cambridge: Cambridge University Press

Mattiello, Elisa. 2013. *Extra-grammatical morphology in English. Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin: Mouton de Gruyter.

Štekauer, Pavol. 2005. Onomasiological approach to word-formation. In Pavol Štekauer and Rochelle Lieber (eds.), *Handbook of Word-Formation*, 207-232. Dordrecht: Springer.

Investigating the semantic contribution of complement clauses in Hindi verbs of experience

Lucrezia Carnesale
(University of Pavia)

Keywords: Hindi, complementation, perception, cognition, syntax-semantics interface

This paper is a corpus-based study of complement clauses of Hindi verbs of cognition (such as *socnā* "think" and *jānnā* "know") and perception (such as *dekhnā* "see" and *sunnā* "hear"). These two verb classes typically show a tendency to occur with Stimuli instantiated by complement clauses rather than NPs. Building upon previous studies that propose that the choice of complement types is semantically motivated (Bolinger 1968, Dik and Hengeveld 1991, Dixon 1984, Goddard 2007, Luraghi 2020, and Wierzbicka 1988), I argue that complement clauses in Hindi have their own semantics and that, as a corollary, they show a distributional preference for verbs with meanings closely related to that of the construction.

For the data retrieval in this study I rely on the hiTenTen corpus, available on SketchEngine[1]. I extract a random sample of 100 occurrences for 20 different verbs belonging to the two classes and I manually scrutinize each occurrence. Relying on a constructional approach (Goldberg 1995, 2006; Barðdal 2008), I assess the number of different verbs that occur with a given construction to determine the type-frequency of the construction. I then analyze the semantic properties of the verbs occurring with a given construction to establish its semantic coherence.

The results show that verbs belonging to these two classes exhibit some similarities in their complement type distribution. For example, they seem to occur with a comparatively similar frequency with finite complement clauses (1). However, the analysis also shows some significant differences. For example, Hindi verbs of (visual) perception exhibit a high frequency of occurrence with predicative participles (2), while verbs of cognition rarely allow for such constructions. On the other hand, verbs of cognition often allow non-finite constructions (3), which are never found with perception verbs.

Additionally, I demonstrate that even verbs with different meanings can still occur with a given construction if there is a way of integrating the semantics of the verb into the constructional semantics (Perek 2015). In such instances, the construction contributes its meaning and changes the interpretation of the verb. For example, the predicative participle construction is typically used to express direct perception of states of affairs, due to its ability to represent simultaneity between two events (Noonan 2007). Some Hindi verbs of cognition, however, may sometimes occur with a predicative participle clause: in such cases, they result in a different interpretation, which stands somewhere between cognition and perception (4).

1. *maiṁ*   *kabhī-kabhī* *soc-t-ī*   *hūṁ,*   *ki*   *tum*
   1SG.NOM   sometimes   think-IPRF-F.SG   be.1SG.PRS   that   2PL.NOM
   yahāṁ na   *ā-te,*   *to*   *acch-ā* *ho-t-ā.*
   here   not   come-IPRF-M.PL   then   good-M.SG   be-IPRF-MSG
   "I sometimes think it would have been better if you had not come here".

---

[1] https://www.sketchengine.eu/hitenten-hindi-corpus/

2. *mujhe   apne   ghar=meṁ=se   ek   aurat   bāhar*

    1SG.DAT   REFL   house=in=INS   one   woman(F.SG.NOM)   out

   nikal-t-ī   hu-ī   dikhāī   d-ī

    come_out-IPRF-F.SG   be.PRF-F.SG   seeing(F)   give-PRF.F.SG

    "I saw a woman coming out of my house."

3. *tum   bahut   jaldī   bambaī   choṛ=kar   dillī   ā-n-e=kī*

    2PL.NOM   very   soon   Bombay   leave=CP   Delhi   come-INF-OBL=GEN-F

   soc   rah-e

    think   PRGR-PRF.M.PL

    "You are thinking to leave Bombay very soon and come to Delhi."

4. *jālpā=ko   in   śabd-oṁ=meṁ sneh   aur*

    jalpa=DAT   these.OBL   word(F)-PL.OBL=in   affection(M.SG.OBL)   and

   sahānubhūti=kā   ek   sāgar   umaṛ-t-ā

    sympathy(F.SG.OBL)=GEN   one   ocean(M.SG.NOM)   rise-IPRF-M.SG

   hu-ā   jān   paṛ-ā.

    be.PRF-M.SG   knowledge(M.SG.NOM)   fall-PRF.M.SG

    "Jalpa felt an ocean of affection and sympathy rising in these words."

References

Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.

Bolinger, Dwight L. 1968. Entailment and the meaning of structures. *Glossa* 2.2: pp. 119-127.

Dik, Simon Cornelis and Kees Hengeveld. 1991. The Hierarchical Structure of the Clause and the Typology of Perception-verb Complements. *Linguistics* 29: pp. 231–259.

Dixon, Robert M.W. 1984. The semantic basis of syntactic properties. *Berkeley Linguistics Society, Proceedings* 10: pp. 583-595.

Goddard, Cliff. 2007. A 'lexicographic portrait' of forgetting. In Mengistu Amberber (ed.) *The language of memory in crosslinguistic perspective*, pp. 119-138. Amsterdam: John Benjamins.

Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Luraghi, Silvia. 2020. *Experiential Verbs in Homeric Greek*. Leiden, The Netherlands: Brill.

Perek, Florent. 2015. *Argument Structure in Usage-Based Construction Grammar. Experimental and corpus-based perspectives*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Noonan, Michael. 2007. Complementation. In Shopen T. (ed.) *Language Typology and Syntactic Description,* pp. 52–150. Cambridge: Cambridge University Press.

Wierzbicka, Anna. 1988. The Semantics of Grammar. Amsterdam: John Benjamins Publishing.Zalizniak.

Existentials: A Standard Average European Feature?

Eyal Liron Dolev & Guido Seiler
(University of Zurich & University of Zurich)

Introduction: The notion that European languages share many traits that are typologically relatively rare can be dated as early as Whorf (1941), who coined the term Standard Average European (SAE). Haspelmath (2001) proposes a list of 26 features (e.g., definite/indefinite articles, relative pronouns) to define the major linguistic characteristics and the core area of an SAE *Sprachbund*. However, Haspelmath's feature list does not include existential sentences. An existential is a "non-canonical construction which expresses a proposition about the existence or the presence of someone or something" (McNally, 2011), e.g., English *there* + be constructions (t*here's a cat at the window*).

Research Question: We examine whether there is a specific type of existential sentence that is prototypical for SAE languages.

Approach & Method: Our central claim is that the SAE-ness of existentials becomes visible only under the appropriate definition of the feature. Previous work has noted the syntactic similarity of existentials across Germanic languages, cf. Platzack (1983), or the use of expletive pronouns in some Romance languages as a commonality with Germanic, cf. Faarlund (1990) and McNally (2016). Creissels (2014), the most extensive typological study, classifies existentials according to the semantics of lexical items recruited for existential constructions (e.g., locative, possessive, etc.). However, the striking similarities in existentials of Germanic and Romance core SAE languages are not necessarily found in the semantics of the involved lexical items at their pre-recruitment stage, but rather in a specific abstract constructional scheme, namely the combination of an expletive element with a desemanticized verb.

Data & Results: We analyze existential constructions in eight Germanic languages (Danish, Dutch, English, German, Icelandic, Swedish, Swiss German, Yiddish) as well as in four Romance (Italian, Romansh, French, Spanish) and two Slavic (Czech, Russian) languages. We show that, despite variation in lexical sources, the syntactic structure of existentials (expletive + bleached verb) is identical across all the Germanic and three of the investigated Romance languages, as well as partly Czech (while Spanish and Russian – but notably also Finnish – make use of other strategies than the SAE expletive + bleached verb construction). In Examples (1)-(2), *es* and *gl'* (citation form *i*, "it") are the expletives; *het* (infinitive *haa* "have") and *ò* (infinitive *avair/aveir*, "have") are the bleached verbs, respectively.

(1)   Swiss German
      Es   het   vil   lüt   do.
      it   has   many   people   there
      "There are a lot of people there."

(2)   Romansh
        Gl'-ò   vegn   ainten   la   fretsgera.
        3SG.INDF-has   wine   in   the   fridge
        "There is wine in the fridge."

Discussion: Based on the hitherto limited data set we tentatively conclude that a specific constructional scheme (but not the lexical sources of its ingredients) is indeed a hot candidate for yet another SAE feature: existential = expletive + bleached verb. It is likely that the feature is shared by a contiguous area in (Western) Europe, crossing the Germanic-Romance boundary, while worldwide a completely different strategy is most widespread (a dedicated existential predicator, cf. Creissels (2014), which is also found in European languages like Spanish or Russian).

Glossary
3   third person   SG   singular
INDF   indefinite

References

Creissels, Denis (2014), *Existential predication in typological perspective*. http://www.deniscreissels.fr/public/Creissels-Exist.Pred.pdf

Faarlund, Jan T. (1990), *Syntactic Change: Toward a Theory of Historical Syntax*, De Gruyter Mouton. https://doi.org/10.1515/9783110854947

Haspelmath, Martin (2001), The European linguistic area: Standard Average European, In *Language Typology and Language Universals* (pp. 1492–1510), De Gruyter. https://doi.org/10.1515/9783110194265-044

McNally, Louise (2011), Existential sentences, In Klaus Von Heusinger, Claudia Maienborn, & Paul Portner (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science* (pp. 1829–1848), De Gruyter. https://doi.org/10.1515/9783110255072.1829

McNally, Louise (2016), Existential Sentences Crosslinguistically: Variations in Form and Meaning, *Annual Review of Linguistics*, *2*(1), 211–231. https://doi.org/10.1146/annurev-linguistics-011415-040837

Platzack, Christer (1983), *Existential Sentences in English, German, Icelandic and Swedish: Vol. I* (Karlsson, Fred, Ed.; pp. 80–100).

Whorf, Benjamin Lee (1941), The Relation of Habitual Thought and Behavior to Language, In Leslie Spier (Ed.), *Language, Culture and Personality. Essays in Memory of Edward Sapir* (pp. 75–93), Sapir Memorial Publication Fund.

Quotative constructions with *ba* and *så här* in spoken Swedish

Michał Piosik

(Adam Mickiewicz University)

Keywords: quotative markers, spoken Swedish, Interactional Construction Grammar

In (spoken) Swedish, a number of new quotative markers have emerged during the last few decades, a development found also across other languages (Buchstaller/Van Alphen 2012: XI-XIII). The new Swedish signals of direct speech have been recruited from various semantic and lexical sources such as comparatives (*liksom* 'like'), demonstratives (*så här* 'such like this') and quantifiers (*ba(ra)* 'only, just'). Previous studies investigating these innovations (e.g. Eriksson 1995, Jönsson 2005 and Svensson 2009) were almost entirely based on corpora of adolescents' speech. Therefore, the question whether and how the markers listed above are used in the language of (young) adults has hitherto not been answered in a satisfactory manner. These aspects form the starting point of the present contribution.

On the basis of a corpus consisting of 13 YouTube-videos in Swedish (format: story time, total duration: 3:44:50, content creators aged between 19 and 30 years, annotation conventions and software: GAT2, FOLKER) and using the framework of the ICxG (*Interactional Construction Grammar*, Imo 2007, 2015), I will focus on formal and functional characteristics of two selected quotative markers in the speech of (young) Swedish adults, namely *ba(ra)* (1) and *s(åh)är,* as well as their combinations (3).

(1)
```
0137   å ja SÄger- NÄ- ja !VILL! verkligen inte;=
```
and I say – no, I really don't want to
```
0138 → HAN_ba- JO_JO_JO-
```
he just- yes, you want

(2)
```
0154   (-) °h ja hade plaNErat de allså;=
```
so I had planned det
```
0155 → =bara för att ja KÄNde s_är- JA kommer få en paNIKattack;
```
just because I felt like (this)- I'm going to get a panic attack

(3)
```
0085 → =jag SA inte s_är ba- Å- MEN den här KILLEN skulle ja kunna TÄNka mej
       å !DEJ!ta;
```
I didn't say like (this) just- well, I could imagine myself dating this guy

The analysis reveals that 54,9% of all 422 quotations were introduced with *ba(ra)* or a combination of quotative markers involving *ba(ra),* whereas *s(åh)är* has a much lower frequency in the material (11,4%). Moreover, the quantitative analysis indicates yet another significant difference between the markers in question. The majority of *ba*-instances (54,3%) is used without any quotative verb: [[Sub] [*ba*] [quotation]]. In contrast, *s(åh)är* occurs almost always in a clausal structure with a finite verb: [[Sub] [V_{fin}] [*s(åh)är*] [quotation]], which per se is evidence of syntactic and pragmatic dependence of *s(åh)är*.

In this context, I argue that both *ba* and *s(åh)är* have, despite some differences, become fully-fledged and highly frequent constructions in spoken Swedish. Furthermore, against the assumption of Eriksson (1995: 45), there are no signs that *ba* as a quotative marker should disappear in the near future. Instead, it can be hypothesized that *ba*, due to its general high frequency and advanced pragmaticalization will spread to other speaker groups and contexts.

References

Buchstaller, Isabelle/Van Alphen, Ingrid (2012), Introductory remarks on new and old quotatives, in I. Buchstaller, and I. Van Alphen (eds), (2012), *Quotatives. Cross-linguistic and cross-disciplinary perspectives.* Amsterdam / Philadelphia: John Benjamins, XI-XXX.

Eriksson, Mats (1995), A case of grammaticalization in modern Swedish: The use of *ba* in adolescent speech, *Language Sciences.* Vol. 17, 19-48.

Imo, Wolfgang (2015), Interactional Construction Grammar, in *Linguistics Vanguard* 2015; 1(1), 69-77.

Imo, Wolfgang (2007), *Construction Grammar und Gesprochene-Sprache-Forschung. Konstruktionen mit zehn matrixfähigen Verben im gesprochenen Deutsch*. Tübingen: Niemeyer.

Jönsson, Linda (2005), Tal i tal. Grammatiska konstruktioner av anföring i tonårsflickors samtal, in J. Anward, and B. Nordberg, Bengt (eds), (2005), *Samtal och grammatik. Studier i svenskt samtalsspråk*. Lund: Studentlitteratur, 89-108.

Svensson, Gudrun (2009), *Diskurspartiklar hos ungdomar i mångspråkiga miljöer i Malmö*. Lund: Lunds Universitet.