

**From derivation to inflection and back:
Variation and change in the Abaza caritive morphology**

Peter Arkadiev, Zürich University, alpgurev@gmail.com

Caritive (privative, abessive) is a construction describing non-involvement of a participant in a situation and expressed as a modifier of another situation (Oskolskaya et al. 2020), e.g. English *a beardless man* or *I came without money*. In many languages caritives are expressed morphologically (Stolz et. al. 2007), and caritive morphemes sometimes show peculiar combinations of properties striding the inflection-derivation divide (e.g. Hamari 2014, Graschenkov 2015). This paper discusses synchronic variation and ongoing change in the morphological behaviour of the caritive marker(s) in Abaza (ISO 630-3 abq, Northwest Caucasian, Russia). The analysis is based on fieldwork data collected in the Karachay-Cherkess Republic in 2019 and 2021 from 16 native speakers.

Abaza has a suffix *-da* forming caritive adjectives from nouns, e.g. *bž'ə* 'voice': *bž'ə-da* 'silent', *zɬ^wa* 'illness': *zɬ^wa-da* 'healthy'. This suffix does not appear to be fully productive and attaches only to common nouns without markers of definiteness or possession. Like other lexical nominals in Abaza, caritive adjectives can occur as predicates and attach verbal morphology (1).

- (1) *šə-zɬ^wa-da-χa-t*
2PL.ABS-illness-CAR-INC-DCL
'Thank you.' (lit. let you become healthy, i.e. without illness)

Caritive adjectives can attach the suffix *-ʔa* (originally locative but no longer productive), resulting in caritive adverbials, e.g. *bž'ə-da-ʔa* 'silently' (the vowel of the caritive suffix is often elided yielding *-d-ʔa*). Such forms are used as modifiers of verbs (2).

- (2) *a-ph^wəspa* *karandaš-d-ʔa* *d-ɬ^w-əj-t*
DEF-girl pencil-CAR-ADV 3SG.H.ABS-write-PRS-DCL
'The girl draws without a pencil'

Remarkably, the distribution of caritive adverbials in Abaza does not mirror that of caritive adjectives. First, the complex marker *-d(a).ʔa* is very productive and attaches to nouns with definiteness or possessive prefixes (*s-h^waspa-d.ʔa* 1SG.PR-knife-CAR.ADV 'without my knife'), proper names (*muradin-d.ʔa* 'without Muradin') and personal pronouns (*awat-d.ʔa* 'without them'). Second, this marker takes phrasal scope over demonstratives (3a), postposed adjectives (3b) and even relative clauses. None of these categories are possible input for the caritive adjectiviser *-da*.

- (3) a. [*arəj* *a-h^waspa*]-*d.ʔa*
 PROX.SG DEF-knife-CAR.ADV
 'without this knife'
 b. [*h^waspa* *çara*]-*d.ʔa*
 knife sharp-CAR
 'without a sharp knife.'

This suggests that caritive adverbs are no longer derivatives of caritive adjectives, but have undergone "affix telescoping" (Haspelmath 1995, Stump 2022) and been reanalyzed as caritive case forms on a par with a handful of other Abaza oblique case markers, e.g. the instrumental *-la* (cf. the treatment of *-d.ʔa* as a case marker in Genko 1955: 118).

The situation described above, with a clear-cut division of labour between the caritive adjectivaliser *-da* and the caritive case marker *-d.ʔa*, is attested in the standard written language and in more conservative speakers. A number of speakers allow the caritive case marker *-d.ʔa* to substitute the caritive adjectivaliser *-da*, and for some other speakers, the latter has become fully unproductive and lexicalised. In the data collected from those

speakers, the suffix *-d.ʔa* has become the only productive caritive marker used both in adverbial (case-like) and adjectival (derivational) contexts, being able to form predicates (4a) and serve as input to verbal derivational morphology (4b).

- (4) a. *sə-nʔarta-d.ʔa-p̄*
 1SG.ABS-work-CAR.ADV-NPST.DCL
 ‘I am jobless.’
- b. *sə-j-rə-nʔarta-d.ʔa-t̄*
 1SG.ABS-3SG.ERG-CAUS-work-CAR.ADV(AOR)-DCL
 ‘He made me jobless.’

Thus, in the synchronic data we observe usage patterns suggestive of three successive diachronic stages: (I) *-da* caritive adjectivaliser vs. *-d.ʔa* caritive case marker; (II) *-d.ʔa* caritive case marker and an allomorph of *-da* in adjectivalising function; (III) *-d.ʔa* general and the only productive caritive marker in all functions.

This illustrates a rather peculiar development of a combination of derivational markers (caritive adjectivaliser *-da* + adverbializer *-ʔa*) first into an inflectional marker with phrasal scope (caritive case *-dʔa*) and then encroachment of the latter into the domain of the original derivational marker (the second process has parallels in Uralic languages, Hamari 2011: 51). As a result, in the system of the innovative speakers, the caritive marker *-dʔa* exhibits a remarkable mix of properties, simultaneously behaving as a case marker with phrasal scope and as a derivational marker with lexical scope. Likewise, the Abaza caritive formations do not fit well into the taxonomy of lexical categories of the language, behaving, on the one hand, as adverbial expressions and, on the other, as derived adjectives able to undergo further predicate formation.

References

- Genko, Anatolij N. (1955), *Abazinskij jazyk. Grammatičeskij očerk narečija tapanta* [*The Abaza Language. A Grammatical Sketch of the Tapanta Dialect*], Moscow: Izdatel'stvo Akademii Nauk SSSR.
- Graschenkov, Pavel (2015), Komitativ i karitiv v tjurkskix i mongol'skix jazykax: funkcii i vozmožnaja èvoljucija [Comitative and caritive in Turkic and Mongolic: functions and possible evolution], *Uralo-altajskie issledovanija* 4(19), 7–16.
- Hamari, Arja (2011), The abessive in the Permic languages, *Journal de la Société Finno-Ougrienne* 93, 37–84.
- Hamari, Arja (2014), Inflection vs. derivation. The function and meaning of the Mordvin abessive, in Fr. Rainer, Fr. Gardani, H. Chr. Luschützky & W. U. Dressler (eds.), *Morphology and Meaning: Selected papers from the 15th International Morphology Meeting, Vienna, February 2012*, Amsterdam / Philadelphia: John Benjamins, 163–176.
- Haspelmath, Martin (1995). The growth of affixes in morphological reanalysis. In: *Yearbook of Morphology 1994*. Dordrecht: Kluwer, 1–29.
- Oskolskaya, Sofia, Natalia Zaika, Sergei Klimenko & Maksim Fedotov (2020), Defining caritive as a comparative concept, *Voprosy Jazykoznanija* 3, 7–25. (In Russian)
- Stolz, Thomas, Cornelia Stroh & Aina Urdze (2007), WITH(OUT): On the markedness relation between comitatives/instrumentals and abessives, *Word* 58, 63–122.
- Stump, Gregory T. (2022). Rule combination, potentiation, affix telescoping. In: A. D. Sims, A. Ussishkin, J. Parker & S. Wray (eds.), *Morphological Diversity and Linguistic Cognition*. Cambridge: Cambridge University Press, 2022, pp. 282–306.

Reconstructing prefixing nominal number marking in East Caucasian

Gilles Authier, EPHE, PSL (gilles.authier@gmail.com)

Most East Caucasian languages have an essentially suffixal morphology. Verbs show gender/number agreement prefixes in all but a few less conservative languages. The agreement system is reconstructed as human masculine (M) *v-*, human feminine (F) *r-*, other animates and semantically associated things and concepts (A) *b-*, and other inanimates (N) *d-*. Plural agreement marking was binary and took up singular markers by using *b-* for humans (HPL) and *d-* for non-humans (NPL). In the nominal domain, attested systems of plural marking are always suffixal, with markers characterized by mostly labial or dental elements, and nouns are not marked for their own gender. However, frozen ‘autogender’ (Nichols 2007) prefixes of the African type have long been recognized within the disproportionate number of third-gender nouns showing a labial onset and of fourth-gender nouns with a dental onset, as well as within exceptional series like Avar *vac / jac / bac* = Dargwa *uzi / ruzi / buzi* for ‘brother’ / ‘sister’ / ‘animal sibling’.

In fact, a considerable number of basic vocabulary nouns retain a gender marker in the onset, e.g. ‘tongue’, ‘eye’ and ‘moon’ for the third gender, and ‘year’, ‘meat’ and ‘heart’ for the fourth gender. Some nouns, like ‘tongue’, always appear with this admittedly gender marking prefix, while others, like ‘brother’ (M), ‘head’ (A) and ‘year’ (N), show it only in some languages. Other nouns, such as ‘moon’ (A) and ‘heart’ (N), show it, in some other languages, in the nominative but not on the oblique stem. There is a notable degree of allomorphy due to branch-specific or areally determined consonant shifts ($b \rightarrow v$; $d \rightarrow r \rightarrow j$) or root-particular nasalization: $b \rightarrow m$; $d \rightarrow n$). Finally, Nichols (ibid.) points out the main crux of this irregular state of affairs: some nouns, like ‘louse’, ‘fingernail’, ‘acorn’ or ‘nettle’, have a labial, i.e. third-gender prefix, in one branch or language, but a dental, i.e. fourth-gender prefix, in another. In some instances, a clear genealogical divide is illustrated, like in the case of Chechen *meza*, Batsbi *mac* ‘louse’ vs . Avar *nac:*, Tsez *noc*, Lak *nac*, Itsari Dargwa *nez*, Aghul *net:*, Khinalug *nimc*, but in other cases, like Avar *mik:* vs Chamalal *niku* ‘oak/acorn’, the languages showing different onsets belong to close genealogical units.

The fact that some languages show different configuration from their closest relatives shows that these prefixes are traces of recently defunct, morphology although it probably has been irregular for a long time (for instance Kryz *mi-til* ‘kid’ vs Budugh *til* ; Kryz etc *medž*, Lezgian *verg* Avar *mič:* ‘nettle’ vs Chamalal *č'i-lu* ; Budugh etc *dix* ‘son’, Dargwa *durħa* vs Lezgian *χwa* ; Dargwar *q'ar*, Rutul *χl'ar* ‘pear’ vs Kryz *džu-χur* ; Rutul etc *mugul* ‘broom’ vs Lezgian *k:ul* ; Tabasaran etc *muxur* ‘breast’ vs Lezgian *χur* ; Tsakhur etc *mič'ri* vs Kryz *džiri*, Lezgian *č:uru*). These clear traces of a consonantal onset alternation are easily explained as number marking opposing overt singulative *b-* (or covert: \emptyset) and (collective) plural (*d-* or \emptyset). They may also reflect an older layer of verb-to-noun derivation morphology.

Focussing on the nouns affected by overt 3rd/4th gender marking and apparent alternation reflected by discrepancies across the languages, we expanded the sample and found out that these represent coherent semantic subclasses of typical mass, group or pair nouns, like ‘cloud/fog’, ‘tear(s)’, ‘grease’, ‘acorn/oak’, ‘pine(cone)’, ‘oats’, ‘fingernail’, ‘honey’, ‘flour/dough’, ‘dove’, as well as ‘dream’, ‘road’, and ‘cave’.

Other nouns are characterized by 3rd/Ø (4th covert) or and 4th/Ø (3rd covert) alternation (including many body parts and elements of natural or artificial entities like ‘head’, ‘eye’, ‘liver/lungs’, ‘nose/nostrils’, ‘hair/beard’, ‘back/spine’, ‘heart’, ‘wool’, ‘ashes’, ‘nettles’, ‘moon/star’, ‘sky/year’, ‘sun/day’, ‘mouse’, ‘flea’, ‘ox’, ‘door’, ‘yoke’, ‘needle/awl’, ‘spindle’, ‘sickle’) and should also be interpreted as reflecting frozen number marking morphology.

Set against the changing Northern Eurasian typological profile, these submerged inflectional (or perhaps originally derivational) features of proto-East Caucasian, as opposed to the current state of affairs in attested languages, bear witness to an important morphological shift, probably driven by long-standing contact with exclusively suffixing languages.

Reference:

Nichols, Johanna. Head gender/source gender in Nakh-Daghestanian: Synchrony and typology. Conference on Languages of the Caucasus, Leipzig, December 8, 2007

Velarization as a way to avoid syncretism in 2nd conjugation Catalan verbs

Manuel Badal

Universitat de València

manuel.badal@uv.es

In Old Catalan, a velar consonant appears in the PyTA morpheme (acronym of *perfecto y tiempos afines*) of some 2nd conjugation verbs (Pérez Saldanya 1998; Wheeler 2011), like *beure* ‘to drink’ (e.g., 3rd BIBUIT > *bec* [bék] ‘s/he drank’), *deure* ‘to owe’ (e.g., 3rd DEBUIIT > *dec* [dék] ‘s/he owed’), *molldre* ‘to grind’ (e.g., 3rd MOLUIT > *molc* [mólk] ‘s/he ground’) or *valdre* ‘to cost’ (e.g., 3rd VALUIT > *valc* [válk] ‘s/he cost’). Initially, in this verb class, the velar was a mark of perfect in the tenses that came from Latin perfect theme. However, the velar progressively extended from PyTA to L-pattern, label that refers to a distribution such that a distinctive form is shared by the whole of the present subjunctive together with the 1st person singular present indicative (Maiden 2018: 84). In this presentation, based on data provided by a corpus comprising Catalan works ranging from the 13th to the 19th century, we prove that the existence of syncretism between the 1st and 3rd persons of the present indicative (e.g., 1st *beu* [béw] ‘I drink’ vs. 3rd *beu* [béw] ‘s/he drinks’) is a factor that accelerates the velarization of L-pattern forms.

Specifically, /g/ spread from PyTA to the 1st person of the present indicative and present subjunctive, following *dir* ‘to say’ verb pattern, in which the velar element was etymological in the L-pattern forms, since it was part of the stem: e.g., 1st DICO > *dic* [dík] ‘I say’; 3rd DICAT > *diga* [díya] ‘that s/he say’. In relation to the verb classes that underwent velarization, it should be noted first that verbs which originally present the velar consonant in PyTA can be divided into two categories: those that had syncretism between the 1st and the 3rd persons of the present indicative, such as verbs of *beure* or *molldre* type (e.g., 1st *mol* [mól] ‘I grind’ vs. 3rd *mol* [mól] ‘s/he grinds’), and the rest, in which regular phonetic evolution from Latin to Catalan had given rise to allomorphic stems in the L-pattern forms, often with a palatal ending, which avoided the aforementioned syncretism (e.g., 1st *deig* [détʃ] ‘I owe’ vs. 3rd *deu* [dév] ‘s/he owes’).

Our data corpus displays that the presence of syncretism between the 1st and the 3rd persons of the present indicative is a factor that accelerates the adoption of /g/: in verb classes represented by *beure* or *molldre*, the velarization of the L-pattern was already consolidated in the 14th century (Badal 2021). In these verbs, the introduction of /g/ helped to solve the homophony problem between the 1st and the 3rd persons of the present indicative: 1st *beu* >> *bec* [bék] vs. 3rd *beu*; 1st *mol* >> *molc* [mólk] vs. 3rd *mol*. From here, in accordance with the class-stability principle (Wurzel 1987), the introduction of the velar in the 1st person of the present indicative entailed the velarization of the whole present subjunctive, due to the implicative relationship established between these forms in Catalan verb inflection. According to this association, when the 1st person of the present indicative presents a specific stem, this allomorph reappears in the whole present subjunctive (Pérez Saldanya *et al.* 2004: 97).

In verbs like *deure* ‘to owe’ or *valdre* ‘to be worth’, that initially presented the velar consonant in PyTA, but without syncretism between the 1st and 3rd persons of the present indicative, the propagation of /g/ in the L-pattern came a little later. In this class, the adoption of the velar consonant in the 1st person of the present indicative was consolidated in the 15th century, while in the present subjunctive the velarization was completed a century later. Since in these verbs the L-pattern palatal-ending stems avoided the syncretism mentioned above (1st *deig* ‘I owe’ vs. 3rd *deu* ‘s/he owes’; 1st *vall* [válʎ] ‘I cost’ vs. 3rd *val* [vál] ‘s/he costs’) the adoption of /g/ was not as necessary as in verbs such as *beure* or *molldre*. In return, in verbs such as *deure* or *valdre*, the adoption of /g/ was favored by the uniformity principle (Mayerthaler 1987): the introduction of /g/ allomorphs from PyTA to L-pattern meant the elimination of palatal allomorphic stems, which homogenized the inflectional paradigm of this verb class.

The velar element, once deprived of its original perfect value, was reanalyzed as a verb class marker, and extended from perfective to imperfective forms. There was therefore a

coalescence process between both morphemes (Maiden 2018: 292), because verbs that initially had /g/ in PyTA also adopted it in the L-pattern; consequently, a new verb class was created within the Catalan 2nd conjugation, characterized by the presence of /g/ in both morphemes. Even in dialects like Valencian, the velar consonant use to undo the syncretism between the 1st and 3rd persons of the present indicative has entailed, in colloquial language, the incorporation into the velarized model of 2nd conjugation pure verbs, such as *perdre* ‘to lose’ (1st *perd* [pért] >> *perc* [pérk]), and of 3rd conjugation pure verbs, such as *dormir* ‘to sleep’ (1st *dorm* [dórm] >> *dorc* [dórk]) or *tossir* ‘to cough’ (1st *tus* [tús] >> *tusc* [túsk]), among others.

References

- Badal, M. (2021). *Morfologia verbal catalana: anàlisi diacrònica dels verbs velaritzats*. Universitat de València.
- Maiden, M. (2018). *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: Oxford University Press.
- Mayerthaler, W. (1987). System-independent morphological naturalness. In W. U. Dressler, W. Mayerthaler, O. Panagl, & W. U. Wurzel (Eds.), *Leitmotifs in Natural Morphology* (pp. 25–58). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.10.13may>
- Pérez Saldanya, M. (1998). *Del llatí al català: morfosintaxi verbal històrica*. Universitat de València.
- Pérez Saldanya, M., Sifre Gómez, M., & Todolí i Cervera, J. (2004). *Morfologia catalana*. Barcelona: Editorial UOC.
- Wheeler, M. W. (2011). The Evolution of a Morpheme in Catalan Verb Inflection. In M. Maiden, J. C. Smith, M. Goldbach, & M.-O. Hinzelin (Eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology* (pp. 182–209). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199589982.003.0010>
- Wurzel, W. U. (1987). System-dependent morphological naturalness in inflection. In W. U. Dressler, W. Mayerthaler, O. Panagl, & W. U. Wurzel (Eds.), *Leitmotifs in Natural Morphology* (pp. 59–96). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.10.22wur>

The distribution of zero marking in inflectional morphology

Laura Becker (University of Freiburg, laura.becker@linguistik.uni-freiburg.de)

Background It has long been known that more frequent forms (i.e. words, constructions, grammatical markers) tend to be shorter than functionally equivalent but less frequent forms (e.g. Zipf 1935, Greenberg 1966, Croft 2003, Hawkins 2004, Haspelmath 2008, 2021, Levshina 2018, Diessel 2019, Stave et al. 2021). The fact that one value of a grammatical feature can be left unexpressed or "zero-marked" in opposition to the other value(s) has also been related to frequency and economy in the typological literature (e.g. Bybee 1985, 2010, Dahl 1985, Greenberg 1966, Croft 2003, Hawkins 2004, Haspelmath 2008). However, two important issues have not received sufficient attention in previous studies. First, grammatical categories and markers often do not occur in a 1-to-1-mapping and can vary across different inflection classes. Second, it is often assumed that zero forms pattern with shorter forms in opposition to longer forms, although this has not yet been tested quantitatively. This is the goal of the present study.

The present study To detect zero markers across languages, I segmented word forms into stems and markers following Beniamine & Guzmán Naranjo (2021) and Guzmán Naranjo & Becker (2021): Stems are defined as the longest sub-string that all forms of a lexeme share, and all additional material of a word form is defined as a marker for a given cell in the paradigm. If the form expressing a given cell of a paradigm corresponds to the stem with no additional marker, I treat it as having a zero marker, since the value of the cell is not overtly encoded. Applying this method to the UniMorph dataset (McCarthy et al. 2020), I extracted the zero forms from the nominal and verbal paradigms of 61 and 104 languages, respectively.

Results The data suggests that zero markers are not generally available in morphological inflection paradigms, which may go against the common expectation in typology that zero forms are the default expression for frequent or "unmarked" values of morphological categories. For nouns, we indeed find that 90% of the languages in the dataset have a zero form in some part of the inflection paradigms. In the verbal paradigms, however, we only find zero forms in 73% of the languages in the dataset. Using a Bayesian logistic regression model with additional controls for the paradigm size and the phylogenetic relations between the languages in the sample confirms this difference between nominal and verbal paradigms: nominal paradigms are more prone to zero forms than verbal ones. However, there are no cells or values that show a robust preference for zero forms. Although cells with absolutive, nominative, singular and indefinite values are more likely to be expressed by zero forms than other cells, the overall estimated probabilities are very low (<0.5) with a high degree of uncertainty. The same holds for cells with imperative, past, present and third person singular values in verbal paradigms.

Consequences The results of this study suggest that zero markers do not necessarily follow the trend we observe between short vs. long markers, and their distribution appears rather language-specific, also varying across lexemes within languages to a large extent. Therefore, efficiency does not seem to account well for the distribution of zero forms in inflectional morphology. The occurrence of zero forms should be viewed as a by-product of other diachronic processes rather than an attractor state of its own. Importantly, these processes can be independent of each other and they need not be coherent or caused by the same factors.

References

- Beniamine, Sacha & Guzmán Naranjo, Matías. 2021. Multiple alignments of inflectional paradigms, *Proceedings of the Society for Computation in Linguistics* 4, Article 21.
- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Bybee, Joan L. 1994. The grammaticization of zero: Asymmetries in tense and aspect systems. In William Pagliuca (ed.), *Perspectives on grammaticalization* (Current Issues in Linguistic Theory 109), 235–254. Amsterdam: John Benjamins.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Cristofaro, Sonia. 2019. Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology*, 25–46. Berlin: Language Science Press.
- Croft, William. 2003. *Typology and universals*, 2nd edn. Cambridge: Cambridge University Press.
- Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.
- Diessel, Holger. 2019. *The grammar network*. Cambridge: Cambridge University Press.
- Greenberg, Joseph H. 1966. *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.
- Guzmán Naranjo, Matías & Laura Becker. 2021. Coding efficiency in nominal inflection: Expectedness and type frequency effects. *Linguistics Vanguard* 7(s3).
- Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19.1, 1–33.
- Haspelmath, Martin. 2019. Can cross-linguistic regularities be explained by constraints on change? In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology*, 1–23. Berlin: Language Science Press.
- Haspelmath, Martin. 2021. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics*, 1–29.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Koch, Harold. 1995. The creation of morphological zeros. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1994*, 31–731. Springer.
- Levshina, Natalia. 2018. *Towards a Theory of Communicative Efficiency in Human Languages*. Habilitationsschrift, Leipzig: Leipzig University.
- Stolz, Thomas & Nataliya Levkovich. 2019. Absence of Material Exponence. *Language Typology and Universals* 72 (3): 373–400.
- Zipf, George Kingsley. 1935. *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

***Selfie-Merkel* and *Villen-Spahn*: Negative evaluation and event-based personal name compounds in German**

Milena Belosevic (Universität Bielefeld), Sabine Arndt-Lappe (Universität Trier)

milena.belosevic@uni-bielefeld.de, arndtlappe@uni-trier.de

Although compounding is one of the most productive word-formation patterns in German (cf. Schlücker 2012), German compounds with a personal name as a constituent, such as *Selfie-Merkel* in example 1) are still an under-researched field. While previous studies have focused on compounds with a proper name as the first constituent with regard to the semantic patterns underlying the relationship between the name and the lexical unit, such as in *Merkel-Besuch* (cf. Schlücker 2020), there are no theoretical and empirical accounts of German PN compounds with a personal name as the second, head constituent (cf. example 1):

- 1) Jetzt steht es fest, Frau Selfie-Merkel!!! 70% „Flüchtlinge“ kamen aus wirtschaftlichen Gründen nach Deutschland. (Twitter)

Now it is official, Ms Selfie-Merkel!!! 70 % of refugees had economic reasons to come to Germany.

The relationship between the constituents *Merkel* and *Selfie* in *Selfie-Merkel* can be paraphrased as *Merkel took a selfie*. The pragmatic function of the compound, i. e., a negative evaluation of the German Chancellor Angela Merkel, cannot be inferred from the paraphrase. Even though compounding is usually not considered the best linguistic candidate for an evaluative function (cf. Dressler & Barbaresi 2020: 423), it has been claimed that German PN compounds bear a negative evaluative function (cf. Fleischer & Barz 2012). However, it is unclear how the negative evaluation arises since the constituents of PN compounds do not have a negative connotation. As shown in Belosevic (2022), German PN compounds fall into the following subtypes: partitive (*CDU-Merkel*), property-based (*Dumm-Merkel*), and event-based PN compounds (*Selfie-Merkel*). The present paper will focus on the latter type of PN compounds, in which the involvement of the name bearer in an extralinguistic social event (for instance, political affairs) are central to their interpretation. On a pragmatic level, the event-based subtype is challenging because its evaluative function must primarily be inferred from the extralinguistic knowledge about the name bearer and the context. Another challenge arises from the fact that, compared to German expressive compounds with lexical units as constituents such as *Arschgesicht* ('ass'- 'face' denoting a swear word, cf. Meibauer 2013), it is unclear why event-based PN compounds convey a negative evaluation since, in German, compounds are held to have an evaluative or expressive function if they either have a (negatively or positively) connotated first constituent or if the first constituent is metaphorical (cf. Meibauer 2013). So the question arises of how the negative evaluation comes about. Furthermore, it is unclear whether the evaluation is always negative and, if so, whether it occurs systematically and can therefore be considered part of this word-formation pattern. We will fill this gap by reporting on a corpus study of 609 types from the German Reference Corpus (W archive) and the Digital Dictionary of the German Language (corpus WebXL3). On the basis of our corpus analysis, we will argue that negative evaluation is a systematic feature of PN compounding in German, which crucially relies on conventional implicature in the sense of Potts (2005). Like the syntactic appositives discussed by Potts, compounding allows for the juxtaposition of constituents, leaving the semantic relation between juxtaposed elements unexpressed. The function of the modifier is that of a supplement, which may be paraphrased by a supplemental relative clause. This is so because the head of the compound is a name, i.e.

a lexical unit with a unique referent. Unlike in the appositive cases discussed by Potts, however, in which the anchor and the appositive are co-referent, eventive PN compounds juxtapose the anchor (the name) and an event in which the referent of the modifier is a salient participant (cf. Belosevic to appear). Crucially, the event evoked by modifiers in our dataset is always an event that deviates from the expectations generated by the head constituent. We show that such supplemental modifiers implicate the negative evaluation of the name bearer and that they fall into the realm of classic criteria used to define evaluative morphology as "assigning a value which is different from that of the standard or default (within the semantic scale to which it pertains) to a concept" (Grandi & Körtvelyessy 2015: 13). For instance, in *Selfie-Merkel* (cf. example 1), the supplemental paraphrase *who took a selfie in Merkel, who took a selfie*, evokes knowledge about German chancellor Merkel's policy regarding the welcoming of refugees in 2015, captured by means of an emblematic image representing this policy, which showed her and a refugee taking a selfie picture. The very juxtaposition of the modifier *Selfie* and *Merkel* indicates that the event evoked is an event that is to be evaluated as a non-default value on a scale relevant to the event (cf. Grandi & Körtvelyessy 2015). The implicature is dissolved by linguistic means in the context. In example 1), the following part of the context *refugees had economic reasons to come to Germany* helps dissolve the implicature comprised by juxtaposing *Selfie* and *Merkel* in the compound. The phrase *economic reasons* is incompatible with knowledge about what motivated the Merkel government's policies in 2015, which was about welcoming war refugees. It is this knowledge that allows readers to understand the contrast, hence the evaluation, with the event evoked by the modifier *Selfie*. In conclusion, the negative evaluation emerges from the interplay between the supplemental modifier that implicates the negative evaluation and the linguistic means from the context. The linguistic means relevant for dissolving the implicature is selected via extralinguistic knowledge about the social event the name bearer participated in.

References

- Belosevic, Milena (2022): *Veggie-Renate* und *Merci-Jens*. Semantik und Pragmatik onymischer Personennamenkomposita. In: *Zeitschrift für germanistische Linguistik* 50/2, 1–31.
- Belosevic, Milena (to appear): Die Semantik ereignisbasierter Personennamenkomposita im Deutschen. To appear in: *Linguistische Berichte*.
- Dressler, Wolfgang U./Barbaresi, Lavinia (2020): Pragmatic explanations in morphology. In: Pirrelli, Vito/Plag, Ingo (ed.): *Word knowledge and word usage. A cross-disciplinary guide to the mental lexicon*, Berlin, 405–452.
- Fleischer, Wolfgang/Bartz, Irmhild (2012): *Wortbildung der deutschen Gegenwartssprache*. Berlin.
- Grandi, Nicola/Körtvelyessy, Lívia (2015): Introduction: Why evaluative morphology? In: Grandi, Nicola/Körtvelyessy, Lívia (ed.): *Edinburgh handbook of evaluative morphology*, Edinburgh, 3–20.
- Meibauer, Jörg (2013): Expressive compounds in German. In: *Word Structure* 6, 21–42.
- Potts, Christopher (2005): *The logic of conventional implicatures*. Oxford.
- Schlücker, Barbara (2020): Von *Donaustrom* zu *Donauwelle*. Die Entwicklung der Eigennamenkomposition von 1600–1900. In: *Zeitschrift für germanistische Linguistik* 48/2, 238–268.
- Schlücker, Barbara (2012): Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In Gaeta, Livio/Schlücker, Barbara (ed.): *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*, Berlin, 1–27.

German plural formation is word class bound

Dagmar Bittner
Leibniz-ZAS, Berlin
bittner@leibniz-zas.de

Most studies on German plural formation have tried to formulate rules on how the plural form of a noun can be generated from the singular form of the noun. In some or the other way each approach was successful in generating the diversity of German plural markers. It remains an open question, however, which of these approaches comes closest to psychological reality of plural formation in the speaker's minds. What all these approaches have in common is the a priori acceptance of a seemingly unmotivated diversity in plural formation. Without explicitly saying so, it seems to be seen a historical accident that leaves us with a complexity lacking any deeper motivation.

In my talk, I will question that the diversity in German plural formation has no deeper motivation. A very first concern with this assumption is that the high complexity and apparent irregularity has been considerably stable for ages. A second concern results from Greenberg's Universal 36 saying that languages marking gender also mark number (Greenberg 1966:94). The universal suggests that gender and number have something in common. In German, gender is related to word class semantics which is obvious in the derivation of nouns: individuativa (-er, -ler, -ling, -eur derivations) tend to be masculine, abstracta (-ung, -heit, -keit, ion derivations) tend to be feminine and collectiva (ge-derivations) tend to be neuter. The gender classes display semantic load. I will argue that the word class semantics of the (derived) singular forms corresponds with different levels of 'pluralizability' and that this is the motivational background for the different plural formations. Best 'pluralizable' are individuativa which are typically masculine. Abstracta are less good 'pluralizable' since their word class semantics implements features of plurality already in singular. Collectiva (and mass nouns) are most badly 'pluralizable' since their word class semantics implements plurality as a core feature. By showing in more detail that the singular and plural forms of German display different constellations of whole-part-relations I will argue that plural formation in German does not function as mere opposition of countability vs. non-countability but is motivated by the semantics of the word classes that are productively related to the three gender classes of German.

A seemingly dark side of the approach is the motivation of the plural forms of the big number of simplicia. I will discuss this issue in more detail too. However, the main aim of the proposed approach is to show that German plural formation is neither irregular nor unmotivated. This is not saying that single nouns or groups of nouns do not have root learned plural forms and, also, that speakers do not apply other features than word class features when forming plurals.

Greenberg, Joseph H. (1966): Some universals of grammar with particular reference to the orders of meaningful elements. In: Greenberg, Joseph H. (ed.): *Universals of Language*. Cambridge, MA/London: MIT Press, 73-113.

**Towards disclosing the mental structure of morphological families:
The case of the so-called non-finites**

Denisa Bordag¹ denisav@uni-leipzig.de;
Andreas Opitz¹ andreas.opitz@uni-leipzig.de;
¹Leipzig University, Germany

We present selected results of a series of priming experiments in which we explore how German word families are organized and structured in the adult human mind, i.e., what are the relations between the word family members and how they affect each other. Previous research has shown that morphologically related word forms typically speed up each other’s recognition indicating that their representations are not independent. At the same time, priming is not equally strong between the members of a word family (e.g., Schriefers et al., 1992) implying that a word family is not a simple list but has a complex structure (cf. van de Vijver & Baer-Henney, 2019). Surprisingly, relatively little research exists that directly addresses the question of word family mental representation (cf. Schmidtke & Kuperman, 2019), though some topics that are subsumed under it have been explored rather extensively, such as the research on inflection and derivation suggesting that the degree to which word family members affect each other depends on the type of morphological relation between them. The results that we plan to present focus on the mental relationships of the so-called non-finites (in particular infinitives, participles, and conversion nouns) between each other and to less extent also on their relationship to other members of their morphological family.

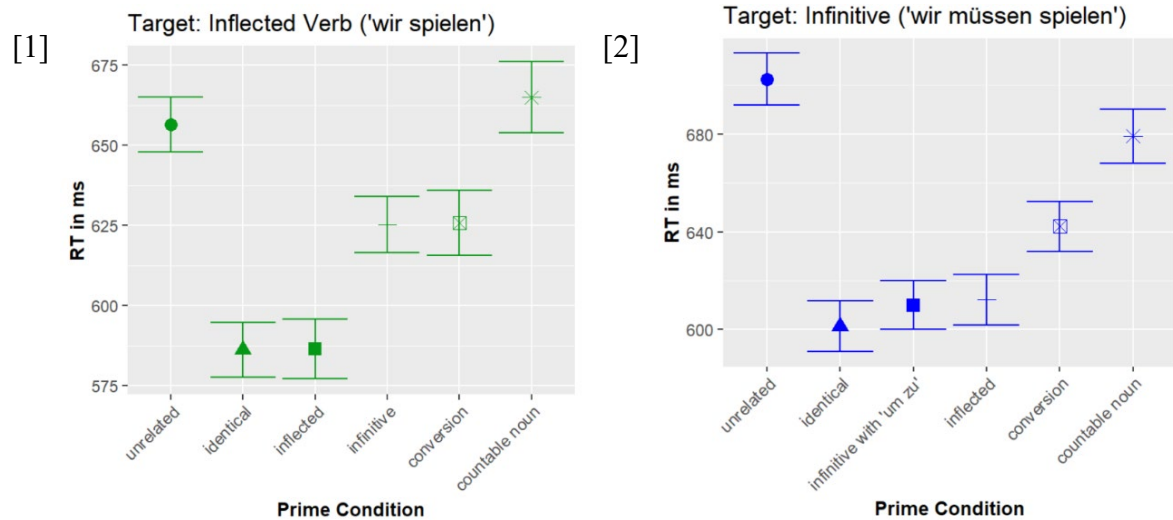
In our unmasked priming experiments, we integrated both primes and targets into short syntactic contexts (presented in a separate step) that enable their morphosyntactic disambiguation in cases, where the forms were overtly homonymous. The relationship between the primes and the targets was manipulated along the way demonstrated in the following table as an example of an experiment, where the target form was an inflected verb form.

Condition	Prime Phrase		Target Phrase	
	Part 1	Part 2	Part 1	Part 2
1 identical	<i>wir</i> ‘we’	<i>SPIELEN</i> ‘play’		
2 inflected	<i>sie</i> ‘they’	<i>SPIELEN</i> ‘play’		
3 infinitive	<i>wir wollen</i> ‘we want to’	<i>SPIELEN</i> ‘play’	<i>wir</i>	<i>SPIELEN</i>
4 conversion	<i>das</i> ‘the’	<i>SPIELEN</i> ‘playing’	‘we’	‘play’
5 countable noun	<i>mit den zwei</i> ‘with the two’	<i>SPIELEN</i> ‘plays’ (pl.)		
6 unrelated	<i>wir</i> ‘we’	<i>WEINEN</i> ‘cry’		

We designed further experiments of this type in which we varied the target form (e.g., infinitive form, conversion noun form) and sometimes the morphological function of the primes. We also performed priming experiments with participles based on the same principle, but with a slightly different design. The results of this series of experiments revealed that inflected forms fully prime (i.e., to the same degree as the identical condition) infinitives and participles, but that participle and infinitive forms only partially prime finite forms. Conversion nouns differed from both infinitives and participles on the one hand and finite forms on the other. They were

partially primed by both the infinite forms and finite forms, and they also only partially primed them when presented as primes. In addition, the priming patterns of the conversion nouns were different when their noun features were highlighted either by orthography (capitalization of nouns in German: *das SPIELEN* vs. *das Spielen*) or by inflection and case-marking (*das SPIELEN* vs. *beim SPIELEN*).

The following two figures show the pattern of results of two of the experiments (targets were inflected (*wir SPIELEN*) [1] and infinitive forms (*wir müssen SPIELEN*) [2]).



The pattern of results allows to draw several conclusions about the representational status of the explored forms. First, the assumption presented in typological literature that there is a class of the so-called non-finites (see Ylikoski, 2003, for an overview) in addition to the classes of verbs and nouns is not fully supported by the psycholinguistic evidence: While infinitives and participles manifest the same properties regarding the representation and processing within their morphological family, conversion nouns diverge from that pattern. However, their priming pattern also differs from that of derived countable nouns indicating that they are more strongly tied to/ closer to the verbal forms in the mental structure of morphological families.

Second, the results indicate that infinitives and participles have the same status in the mental representation of the morphological family. The asymmetrical priming results between them and finite forms indicate a hierarchical structure in which underspecified infinite forms represent (or are linked with) (a) higher node(s) which are also accessed when representations of finite forms are retrieved. Thus, full priming is observed when an infinite form is accessed after the finite form, but only partial priming is observed when the forms are accessed in the reversed order. In our talk we will elaborate on the theoretical implications of the findings and outline the directions of further research.

References:

- Schmidtke, D., & Kuperman, V. (2019). A paradox of apparent brainless behavior: The time-course of compound word recognition. *Cortex*, 116, 250-267.
- Schriefers, H., Friederici, A., & Graetz, P. (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *The Quarterly Journal of Experimental Psychology*, 44(2), 373-390.
- van De Vijver, R., & Baer-Henney, D. (2019). Paradigms in the mental lexicon: Evidence from German. *Frontiers in Communication*, 3, 65.
- Ylikoski, J. (2003). Defining non-finites: action nominals, converbs and infinitives. *SKY Journal of linguistics*, 16(1), 185-237.

Both hyponymy and context concreteness for the second constituent are relevant in the comprehension and production of compound words

Chariton Charitonidis
ORCID iD: 0000-0003-0298-2629
dr.chariton.charitonidis@gmail.com

In recent years, there has been a considerable focus regarding the interface of lexical meaning and emotion (for a review of relevant studies, see Citron et al. 2016 and Yao et al. 2016). By means of online interviews with native speakers of English, Warriner et al. (2013) and Brysbaert et al. (2014) compiled datasets with *representation norms* for English words. These norms referred to the semantic variables ‘valence’ (positivity), ‘arousal’ (excitement, mood-enhancement), and ‘concreteness’. In Sneffjella and Kuperman (2016) the application of representation norms to the 7 billion token USENET corpus (Shaoul & Westbury 2013) resulted in *context norms* for English words. Each context referred to an array of five content words before to five content words after a target word.

The present paper describes the effects of the above mentioned norms on the comprehension (lexical-decision) and production (naming) of compound words. The objects of study were over 2000 concatenated compounds of English taken from the LADEC database (Gagné et al. 2019). In the analysis, the forced-entry method of regression was used. The reaction times for the compounds from the English Lexicon Project (Balota et al. 2007) and the British Lexicon Project (Keuleers et al. 2012) were used as dependent variables. Compound frequency and compound length (in characters) were used as controls.

The results showed that, in lexical decision alone, latency-reducing representation valence for the compound co-occurred with latency-increasing context concreteness for the second constituent. Rerunning the analysis using the hyponymy norms from Gagné et al. (2020) showed that, in both lexical decision and naming, context concreteness for the second constituent and hyponymy were **equally** relevant.

The effects detected are regarded as evidence for dual- and multiple-route models of morphological processing.

REFERENCES

- Balota, D., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., et al. (2007). The English Lexicon Project. *Behavior Research Methods* 39. 445-459.
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* 46. 904-911.
- Citron, F.M.M., Cacciari, C., Kucharski, M., Beck, L., Conrad, M., & Jacobs, A.M. (2016). When emotions are expressed figuratively: Psycholinguistic and Affective Norms of 619 Idioms for German (PANIG). *Behavior Research Methods* 48. 91-111.
- Gagné, C.L., Spalding, T.L., & Schmidtke, D. (2019). LADEC: The large database of English compounds. *Behavior Research Methods* 51(5). 2152-2179.
- Gagné, C.L., Spalding, T.L., Spicer, P., Wong, D., Rubio, B., & Cruz, K.P. (2020). Is buttercup a kind of cup? Hyponymy and semantic transparency in compound words. *Journal of Memory and Language* 113.
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British lexicon project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods* 44. 287-304.
- Shaoul, C. & Westbury, C. (2013). *A reduced redundancy usenet corpus (2005-2011)*. Edmonton, AB: University of Alberta. www.psych.ualberta.ca

- Snefjella, B. & Kuperman, V. (2016). It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition* 156. 135-146.
- Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4). 1191-1207.
- Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2016). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods* 49.

Gender variation in Russian indeclinable nouns: optimality over structuralism

Kirill Chuprinko (Higher School of Economics, Moscow), Varvara Magomedova (independent researcher), Natalia Slioussar (Higher School of Economics, Moscow; Saint Petersburg State University, Saint Petersburg)

churpinko_kirill@mail.ru, varya.magomedova@gmail.com, slioussar@gmail.com

Introduction. In this paper we study the influence of semantic and morphophonological factors on gender assignment to Russian indeclinable nouns. Russian has three genders: masculine (M), feminine (F) and neuter (N). The majority of nouns are inflected for number and case and are divided into 3 inflectional classes (declensions). There is a correlation between gender and declension (e.g. Corbett & Fraser 2000; Kramer 2015; Rice 2006). About 1% nouns are indeclinable. Indeclinable nouns are mainly loanwords. They are mostly N, although it is the least frequent gender otherwise. Often their gender is based on their hypernym: for example, for *saljami* it is *kolbasa*_F ‘sausage’, so *saljami* is F. An overview of the relevant data can be found in (Galbreath 2010; Murphy 2000; Wang 2014). Murphy (2000) notes that more than one gender might be accepted by the speakers, but does not analyze these cases in detail. We focus on them in the present study.

Our study. We analyze data from the General Internet Corpus of Russian, or GICR (LiveJournal subcorpus, see https://github.com/chuprinko-kirill/Russian_indeclinable_nouns_gender) and show that gender variation is massive and determined by both semantic and morphophonological factors. This is difficult to capture in Distributive Morphology and other structural approaches to gender assignment and better fits Optimality Theoretic models.

Dataset. We selected 145 common inanimate indeclinable nouns ending in a vowel (*-a*, *-o*, *-e*, *-i*, *-u*) and searched the GICR using queries with attributives as agreement targets. 36079 corpus instances were collected. We analyzed the influence of the final segment (Table 1), stress position (Table 2, only non-monosyllabic words) and the gender of the hypernym (Table 3, 90 lexemes having a single salient basic level hypernym, yielding 23535 corpus instances).

	Final segment	<i>-a</i> (17)	<i>-o</i> (32)	<i>-e</i> (29)	<i>-u</i> (24)	<i>-i</i> (43)
In parentheses, of the number lexemes is given.	M	341 (20%)	4517 (43%)	1171 (21%)	1126 (22%)	7226 (54%)
	F	300 (18%)	262 (3%)	662 (12%)	1024 (20%)	2018 (15%)
	N	1049 (62%)	5619 (54%)	3617 (66%)	3004 (58%)	4143 (31%)

	Stress	final (57)	penultimate (73)
Only 4 lexemes had other stress positions.	M	2432 (24%)	11419 (49%)
	F	1058 (11%)	2071 (9%)
	N	6454 (65%)	9724 (42%)

	Hypernym	N (15)	M (50)	F (26)
	M	214 (12%)	10893 (66%)	316 (6%)
	F	9 (<1%)	59 (<1%)	1808 (35%)
	N	1537 (87%)	5591 (34%)	3108 (59%)

Statistical analysis. We modeled the data with mixed-effect logistic regressions with random intercepts by noun. We found that the final segment, stress position and hypernym gender are significant predictors. Briefly saying, nouns ending in *-a* are prone to be F, apparently due to their surface similarity with the 1st declension *-a* is not an affix *t* in indeclinable nouns. N is the most frequent in indeclinable nouns, but it becomes even more preferable when nouns end in *-o* or *-e*, like 2nd declension N nouns. Indeclinable nouns with a final stress tend to be N more often than the ones with a penultimate stress. Finally, semantic factors play an important role: in case when an indeclinable noun has a salient declinable hypernym, its gender is assigned significantly more often than otherwise.

Discussion. Theoretical approaches to gender assignment fall into two groups: structural vs. rule or constraint-based. Distributed Morphology is the most widespread family of structural theories. It postulates at least two levels of word-formation: Deep Structure and Phonological

Form, the latter cannot influence the former. As a syntactic feature, gender cannot be affected by phonological factors. Moreover, DM is not geared to deal with variation in general and especially with the interplay of semantic and (morpho)phonological factors.

Some cases of *prima facie* phonological gender assignment can be still explained. For example, Kramer (2020) analyzes loanwords ending in *-a* that receive F gender in Hausa. She suggests that the final *-a* is a realization of the gender feature. To give an example from Russian, the English word *bug* ('computer program error') has two corresponding loanwords in Russian: (*tupoj_M*) *bag_M* and (*tupaja_F*) *baga_F* '(stupid) bug'. In case of indeclinable nouns, the final segment is not an affix, and the variation conditioned by semantic and morphophonological factors affects one lexeme, rather than triggers the emergence of two or three lexemes. This is hardly compatible with the principles of DM.

Our data are more readily compatible with the Optimality Theory framework, in particular with the Optimal Gender Assignment Theory (Rice 2006; Galbreath 2010; Corteen 2018). In the OGAT, semantic, morphological and phonological constraints are equally ranked, allowing for competition, which is crucial for our data. Validating constraints according to the principles suggested by Enger (2009) allows to restrict the power of this model. In the current version of OGAT, a single gender feature is selected as a result of the competition. We are currently working to introduce weighted constraints in it to be able to embrace the full complexity of our data. Taking the significant factors as constraints, for example '*-a* final words cannot be masculine', we can successfully model the data and predict the probabilities of the outputs using the Maximum Entropy approach. We use this kind of analysis anyway as Maximum Entropy is mathematically the same as logistic regression. Moreover, mixed effects regression was shown to predict data better than the original MaxEnt grammar (Zymet 2018), which in its turn was shown to be more precise than Noisy Grammars (Flemming 2021).

Our data may also be interesting for the discussion on gender markedness in Russian and feature markedness in general. At least in functional and OT-based approaches, M is considered to be the default gender in Russian. M is by far the most frequent in declinable nouns, the most resistant to variation. But in the tiny group of indeclinable nouns, N is assumed to be the default. Galbreath (2010) suggests to capture this by reranking constraints for indeclinable nouns in her model. Firstly, we show that it is not true that most indeclinable nouns fall into the N gender class by default — in many cases, other M or F may be selected as a result of the competition. Secondly, N does not behave as unmarked (i.e. something that is selected when no cues point to any other options). It is not the case that indeclinable nouns lack predictors of M and F. Rather, N behaves as a competitively stronger gender. For example, in our dataset, N was assigned to 8699 tokens with M or F hypernyms, i.e. it has overridden strong semantic predictors.

The study was partially supported by the Russian Ministry of Science and Higher Education (project 075-15- 2020-793).

References

- Corbet G.G., Fraser N.M. (2000). Default genders. In: Unterbeck B., Rissanen M. (eds.) *Gender in Grammar and Cognition*. Berlin: Mouton De Gruyter. Pp. 55-97. • Corteen E.C. (2018). *The Assignment of Grammatical Gender in German: Testing Optimal Gender Assignment Theory*. PhD thesis, Trinity Hall. • Flemming E. (2021). Comparing MaxEnt and Noisy Harmonic Grammar. *Glossa*, 6, 141. • Galbreath B.L.E. (2010). *Gender assignment in Contemporary Standard Russian: A comprehensive analysis in Optimality Theory*. PhD dissertation, Washington State University. • Kramer R. (2015). *The Morphosyntax of Gender*. Oxford: OUP. • Murphy D.L. (2000). *The Gender of Inanimate Indeclinable Common Nouns in Modern Russian*. Ph.D. dissertation, University of Ohio. • Rice C. (2006). Optimizing gender. *Lingua*, 116, 1394-1417. • Wang Q. (2014). *Gender Assignment of Russian Indeclinable Nouns*. MA thesis, University of Oregon. • Zymet, J. (2018). *Learning a frequency-matching grammar together with lexical idiosyncrasy: MaxEnt versus mixed-effects logistic regression*. AMP18 talk.

Defectiveness as a product of prescriptivism: a behavioural study

Maria Copot
maria.copot@etu.u-paris.fr

Andrea D. Sims
sims.120@osu.edu

Background

Defectiveness manifests as speakers' rejection of any form of a lexeme that could fill a given paradigmatic cell (Sims, 2015). Work on the causes of defectiveness has so far focused on system-internal factors (uncertainty, homophony, etc). While a number of conditions that correlate with being defective have been established, it has proven challenging to make deterministic predictions about where defectiveness is to be found in a system: for example, uncertainty about the correct form will sometimes lead to defectiveness while at other times to overabundance (Sims, 2006). Similarly, attempts to empirically identify defective forms by seeking words of unexpected low frequency in a corpus have not proven fruitful, since defective forms and lexemes don't seem to have a unique frequency profile (Copot & Bonami, 2020). One underexplored facet of defectiveness and its causes is the impact of extralinguistic factors, such as normative pressures: a societal incentive to "speak correctly" (or even a more explicit knowledge that certain word forms are to be prescriptively avoided in normatively correct speech) can be expected to play a filtering role in deciding which of the words with the necessary prerequisites will be treated as defectives.

To explore this question, we borrow insight and methodology from Vogel (2019). Vogel notes that taboo constructions are the object of a paradox: in order for a construction to be taboo, it must nevertheless exist in language use - speakers will profess that the construction is wrong and does not exist in the language, conflating a belief that it shouldn't exist with an assertion that it doesn't. It should follow from this that in an acceptability judgement task, speakers should rate taboo constructions more variably than ungrammatical controls because 1) an individual's degree of prescriptiveness will determine the extent to which they find the construction unacceptable, 2) more saliently taboo examples of the constructions (those cited as bad by grammar books) will receive worse ratings than less salient ones. Assuming said variability in judgement successfully distinguishes grammatical taboos from ungrammaticality, it provides a sound starting point for investigating whether linguistic prescriptiveness plays a role in inducing defective behaviour.

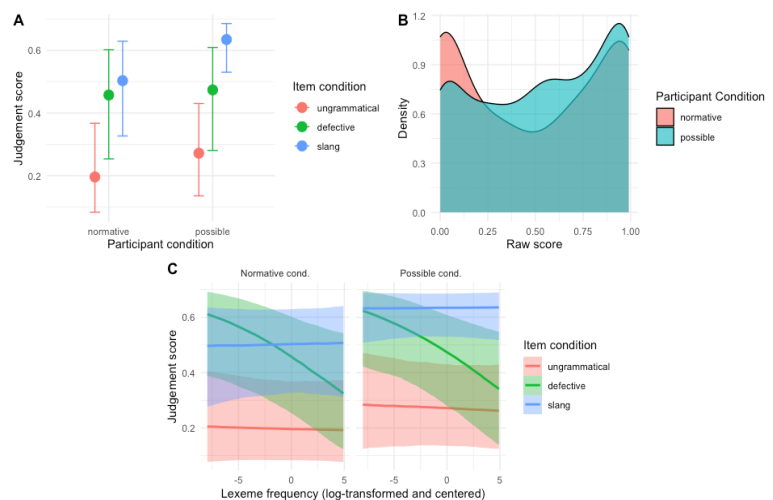
Methodology

The French language has a well-established list of known defective word forms, as well as a strong prescriptive tradition. French speakers were given the task of judging the acceptability of an underlined word in a sentence on an unmarked slider. 60 participants, recruited from Prolific.co, were split in two conditions: one set was asked to give a normative acceptability judgement ("Would this word be marked wrong by a teacher? Could it be found in the dictionary?"), while the other set was asked to give a possibility judgement ("Could you hear someone say this word in casual conversation? Could you catch yourself saying it, even if you might end up correcting yourself?"). The underlined words were all verbs, falling in three categories: 1) known defectives, 2) slang, 3) ungrammatical (agreement errors). The defective verb forms chosen were listed as defective in at least one major French dictionary, and were selected to maximise the likelihood that a speaker would be familiar with the lexeme (so archaic or very infrequent lexemes were excluded). Before the experiment, participants were asked to rate their agreement with statements about linguistic norms and language change, in order to gauge the extent of their prescriptive tendencies. After the judgement task, participants were presented with a list of lexemes in their citation form (the list included all lexemes seen in the judgement task, as well as French pseudoverbs), and they were asked to select all the ones they were not familiar with. Data pertaining to lexemes selected by the participant was excluded from analysis.

A zero- and one-inflated bayesian beta regression with by-participant and by-item random effects was fitted to the scores assigned by participants. Along with the variables characterising the experimental design (participant condition, item condition), and indexical information for each participant, the model includes a prescriptiveness score for the participant, and lexeme frequency for each item.

Results

For both participant conditions, scores for defectives had higher variance than scores for the other two conditions (fig. A, conditional effects plot). Based on Vogel (2019), we had originally predicted that defectives would have higher variance in the normative condition only. Upon closer inspection, the distribution of scores for defectives is in fact bimodal for both participant conditions, though more extremely so in the normative condition (fig. B). This suggests that items that



are labeled as defective in grammars underlyingly belong to two categories which are treated differently by speakers - the difference between the two underlying categories is exacerbated when speakers are asked to think in normative terms.

The effect of lexeme frequency sheds light on the matter (fig. C, conditional effects plot): while it has no effect on slang or ungrammatical items, it has a strong negative effect on defective items: if a lexeme is infrequent, speakers are unlikely to treat its supposedly defective word form as problematic. Defectiveness is meant to manifest as the refusal of all candidate forms that could fill a given cell, but the low ratings that would be expected for defective items on the basis of this are much more likely to be found for high-frequency lexemes. Such a state of affairs is consistent with the proposal that prescriptive pressures play a filtering role in deciding which syntactic words meeting the structural condition for defectiveness get treated as defective: language planning institutions will focus on the most frequently encountered examples of potential defectives, and issue recommendations that they not be used. Assuming that lower-frequency lexemes featuring a syntactic word with the prerequisites to be treated as defective are mentioned less frequently in guidelines from language planning institutions and grammar books, speakers are less likely to be aware that these lower-frequency words are to be avoided.

The current study opens a discussion on the role of extralinguistic factors as causes of defectiveness, and proposes empirical evidence to corroborate the proposal that prescriptiveness has a key filtering role in establishing defective behaviour.

References

- Copot, Maria & Bonami, Olivier. (2020) *Is defectiveness that surprising? The influence of paradigmatic predictability on token frequency*. Presented at the “How to Fill a Cell” Workshop, University of York.
- Sims, Andrea D. (2006) *Minding the gaps: Inflectional defectiveness in paradigmatic morphology*. Ph.D. dissertation: The Ohio State University.
- Sims, Andrea D. (2015) *Inflectional defectiveness*. Cambridge University Press.
- Vogel, Ralf. (2019). Grammatical taboos. *Zeitschrift für Sprachwissenschaft* 38: 37-79.

Product- and source-oriented generalizations: allomorph distribution in Polish locative adjectives

Bartłomiej Czaplicki, University of Warsaw
bczaplicki@uw.edu.pl

An in-depth corpus-based analysis of affix distribution in Polish locative adjectives (2,524 words) provides evidence for (i) generalizations that require reference to construction-specific schemas, both product- and source-oriented and (ii) multiple sublexicons in a grammar. A Polish locative adjective (LocA) is formed using an optional interfix and an obligatory suffix. There are four possible interfixes and two suffixes, as represented in (1). Examples are given in (2).

- (1) root + {∅, ij, iĵ, aŋ, ɛŋ} + {sk, tsk}_{LocA}
- (2) *Noun* *LocA* *Noun* *LocA*
- a. *-∅-sk-* pari_Z-a ‘Paris’ g.sg. pari-sk-i b. *-∅-tsk-* lɛgɲits-a ‘Legnica’ lɛgɲi-tsk-i
- c. *-ij-sk-/-iĵ-sk-* kanad-a ‘Canada’ kanad-ij-sk-i d. *-aŋ-sk-* marək-ɔ ‘Morocco’ marək-aŋ-sk-i
- e. *-ɛŋ-sk-* budapešt ‘Budapest’ budapešt-ɛŋ-sk-i

Table 1 provides the type frequency of the four main ways of forming LocAs in the *plTenTen19* corpus ([*-ij-sk-*] and [*-∅-tsk-*] are analyzed as contextual allomorphs of [*-ij-sk-*] and [*-∅-sk-*], respectively).

Table 1. Type frequency of LocAs in [*-∅-sk-*], [*-ij-sk-*], [*-aŋ-sk-*] and [*-ɛŋ-sk-*]

	<i>-∅-sk-</i>	<i>-ij-sk-</i>	<i>-aŋ-sk-</i>	<i>-ɛŋ-sk-</i>
Type freq.	1865	264	273	122
%	73.9	10.5	10.8	4.8

The pattern without interfixes (*-∅-sk-*) has the highest type frequency, which implies that it should be the most productive of the four (Alegre & Gordon 1999). Productivity measures calculated for the four patterns (ratio of occasionalisms to the token frequency of all the words exhibiting the pattern, Baayen & Lieber 1991) reveal that the patterns with the interfixes are in fact more productive.

A pivotal element of the analysis is the product-oriented schema that defines the preferred segmental frame in LocAs: vowel + sonorant (VR) before the adjectivizing suffix, as shown in (3). A quantitative analysis of the corpus data serves to test the hypothesis that the use of interfixes in LocAs is regulated by the need to satisfy this schema.

- (3) VR-sk-LocA

LocAs that comply with the schema in (3) emerge in three different ways. Adjectives like the one in (4a) are formed by the addition of an interfix, while in the adjectives instantiated in (4b), VR appears at the end of the base. Therefore, the latter adjective does not require an interfix to comply with the schema. VR strings in LocAs like the one in (4c) emerge through phonological operations such as insertion, deletion or mutation.

- (4) *Three ways of forming LocAs that are compliant with the VR-sk- schema*
- a. marək-ɔ ‘Morocco’ marək-aŋ-sk-i interfix

b.	ukrain-a ‘Ukraine’	ukrajn-sk-i	no interfix
c.	gvine-a ‘Guinea’	gvinej-sk-i	phonological operation

Evidence for the relevance of the schema in (3) comes from the comparison of the relative frequency of VR sequences in LocAs and their bases. VR sequences are more common in LocAs than in their base nouns. The proportions of VR sequences are 0.831 for LocAs and 0.709 for their base nouns. A paired-samples proportions test reveals that the proportions difference is statistically significantly different from zero (Mid p adjusted binomial <.001).

A multinomial logistic regression analysis confirms that the presence of VR in the base has an impact on the selection of interfixes ([-ap-sk-] vs. [-∅ -sk-], B=1.220, SE=.310, p<.001, [-ij-sk-] vs. [-∅ -sk-], B=.809, SE=.293, p=.006). Interfixes are significantly more likely to be selected when the base does not end in a VR sequence. Thus, an interfix serves to provide the LocA with the required VR sequence when it is absent in the base. The product-oriented schema in (3) imposes a segmental template on LocAs and, in this way, promotes the selection of interfixes. I search for alternatives to account for this distribution and conclude that the application of the segmental frame is orthogonal to syllable well-formedness. The selection of interfixes does not produce less marked structures.

The multinomial analysis reveals a host of other factors that govern the selection of affixes, such as *an extrasyllabic sonorant* ([-εp-sk-] vs. [-∅ -sk-]: B=-6.744, SE=.469, p<.001), and various base-affix identity constraints such as *base-final [j]* ([-ij-sk-] vs. [-∅ -sk-]: B=-3.372, SE=.281, p<.001). The relevance of these factors confirms the need for source-oriented schemas. For example, an extrasyllabic base-final [j] selects -{ij/ij}-, while an extrasyllabic sonorant (other than [j]) selects [-εp-], as shown in (5).

- (5) a. *j- ↔ -{ij/ij}-sk-_{LocA}
b. *R- ↔ -εp-sk-_{LocA}

Finally, interfixes are more common in foreign than in native LocAs (e.g. [-ij-sk-] vs. [-∅ -sk-], B=-1.290, SE=.231, p<.001), which supports the view that a phonological grammar is composed of multiple sublexicons (cophonologies; Itô and Mester 1999, Czaplicki 2021). This trend is explained by the fact that LocAs in [-∅ -sk-] commonly trigger modifications of the root of the base in the form of consonant mutations or deletions (e.g. n ~ ɲ, v ~ f, z ~ ∅, g ~ ∅, t ~ ∅, d ~ ∅ and k ~ ∅). Mutations or deletions are not found before the interfixes. In response to base identity pressures, foreign and native locative adjectives differ in the preferred morphological composition (frame), as shown in (6), though this is just a strong statistical tendency. On a general level, the gradience of the identified patterns implies that linguistic knowledge is not categorical, but relies on statistical distributions (cf. Becker and Gouskova 2016).

- (6) a. G_{Foreign}: -{ij, ij, ap, εp}-{sk}-_{LocA/Foreign} (Interfix)
b. G_{Native}: -∅ -{sk, tsk}-_{LocA/Native} (No Interfix)

Albright, A. and B. Hayes. 2003. Rules vs. analogy in English past tenses: A computational/ experimental study. *Cognition* 90. 119–161. **Alegre, M. and P. Gordon. 1999.** Rule-based versus associative processes in derivational morphology. *Brain and Language* 68. 347–354. **Baayen, R. H. and R. Lieber. 1991.** Productivity and English derivation: a corpus-based study. *Linguistics* 29. 801–843. **Becker, M. and M. Gouskova. 2016.** Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47(3). 391–425. **Booij, G. 2010.** *Construction Morphology*. Oxford: Oxford University Press. **Bybee, J. 2001.** *Phonology and language use*. Cambridge: Cambridge University Press. **Czaplicki, B. 2021.** The strength of morphophonological schemas: Consonant mutations in Polish. *Glossa: a journal of general linguistics* 6(1): 25. 1–34. **Itô, J. and A. Mester. 1999.** The Phonological Lexicon. In N. Tsujimura (ed.), *The handbook of Japanese linguistics*, 62-100. Oxford: Blackwell.

Sound patterns, frequency and predictability in verbal inflection: a case study from Latin

András Cser

Hungarian Research Centre for Linguistics & Pázmány Péter Catholic University
cser.andras@nytud.hu

The last two decades of morphological research have seen a surge of interest in paradigmatic relations, information content, patterns of predictability, and their relation to frequency (Ackermann & Malouf 2016, Anderson 1992, Blevins 2006 and 2015, Blevins, Ackermann & Malouf 2016, Stump & Finkel 2013, Montermini & Bonami 2013, Sims & Parker 2016). This paper presents a partial analysis of Latin verbal inflection with the goal of exemplifying the interrelatedness of information content and phonological shape as well as correlations between textual frequency and predictability. In particular, we seek to answer the question how much information relevant to the various morphosyntactic categories to which a word form is assigned is encoded in the phonological shape of that word form (bearing in mind that information conveyed by endings can be assigned to various levels (primarily word class and/or particular morphosyntactic categories). On the one hand, part of the answer to this question is trivial in the sense that morphemic analyses have always been based on precisely such correlations of form and content (e.g. all 2Sing verb forms in Latin end in *-s*, hence this is identified as the morpheme/exponent of the value [2Sing] of the morphosyntactic category Person/Number). On the other hand, ambiguities could still arise since, within a verb paradigm, three out of six verb forms altogether end in *-s* (viz. 2Sing, 1Plur and 2Plur), thus this segment is, on first approximation, a non-trivial exponent of [2Sing]. The analysis actually begins here: we demonstrate that ambiguities potentially arising from the non-trivial relation between the phonological form of the ending and the various values of the morphosyntactic category Person/Number (e.g. would the hypothetical form ***upitis* be the 2Sing of ***upito* or the 2Plur of ***upo*?) are absent from the lexicon and the grammatical system. Based on an analysis of the relevant quantitative aspects of the data as well as of diachronic processes it can be argued that the avoidance of such potentially ambiguous forms was an emergent feature of the language.

In a similar vein, we also systematically compare the phonological form of verb endings that are always polysegmental to phonological patterns found inside, or at the end of, verb stems. The case of the cluster [nt], the only cluster functioning as an inflectional suffix in the Latin language, is particularly instructive. As a consonant cluster it is the most frequent CC in the language both in terms of lexical and textual frequency; in addition, there were diachronic morphological processes that had the capacity to actively increase the number of such clusters, particularly in verbs. A large number of verb stems end in a nasal+stop cluster in Classical Latin, but not one ends in [nt]. We argue that this is related to the fact that [nt] itself is also a suffix (that of 3Plur), and the phenomenon itself can be construed as an instantiation the principle of Repetition Avoidance (Walker 2007). While the cluster [nt] is found inside a few verb stems, it is always separated from the ending of the identical form either by two vowels as in *sentiant* 'they feel' or by the vowel of the highest sonority, as in *cantant* 'they sing'. Not unrelated to this is the almost complete absence from verb stems of phonological sequences that are identical to the other two polysegmental verb endings *-mus* (earlier **-mos*) and *-tis* (earlier **-tes*).

The upshot of the argument is that in Latin verbal paradigms, while information pertaining to morphosyntactic properties such as person and number is encoded in the endings, as is well known, their information content and recognisability is effectively enhanced by non-trivial phonological properties of the verb stems.

The general framework of the paper is based on ideas elaborated in the references given in the first paragraph above. The analysis of Latin specifically draws to varying degrees on

Matthews (1974) and Cser (2015), for its quantitative aspects some use has been made of Pellegrini & Passarotti (2018) and Pellegrini (2020), though all calculations in the present paper are our own. The database used for lexical data was Perseus (<http://www.perseus.tufts.edu/hopper/search>), the one used for textual frequencies was the Packard Humanities Institute database of Latin texts (<https://latin.packhum.org/index>).

References

- Ackermann, F. & R. Malouf 2016 Inferential organization in grammar systems. Talk given at the Research Institute for Linguistics, Budapest.
- Anderson, S. R. 1992 *A-Morphous Morphology*. Cambridge: CUP.
- Baayen et al. 2003 Frequency effects in regular inflectional morphology. In: *Morphological Structure in Language Processing*. Berlin: Gruyter.
- Blevins, J. 2006 Word-based morphology. *JLing* 42: 531–573.
- Blevins, J. 2015 The minimal sign. In: *The Cambridge Handbook of Morphology*.
- Blevins, J., F. Ackermann & R. Malouf: 2016 *Word and Paradigm Morphology*. In: *Oxford Handbook of Morphological Theory*.
- Cser, A. 2015 The nature of phonological conditioning in Latin inflectional allomorphy. *Acta Linguistica Hungarica* 62:1–35.
- Matthews, P. H. 1974, *Morphology: An Introduction to the Theory of Word Structure*. Cambridge University Press, Cambridge.
- Montermini, F. & O. Bonami 2013 Stem spaces and predictability in verbal inflection. *Lingue e linguaggio* 2: 171–190.
- Pellegrini, Matteo 2020 Patterns of interpredictability and principal parts in Latin verb paradigms: an entropy-based approach. *JLatLing*. 19(2): 195–229
- Pellegrini, M. & M. Passarotti. 2018 *LatInfLexi: an Inflected Lexicon of Latin Verbs*. CLiC-it. 2018
- Sims, Andrea D & Jeff Parker 2016 How inflection class systems work: On the informativity of implicative structure. *Word Structure* 9/2. 215–239
- Stump, G. & R. A. Finkel 2013 *Morphological Typology: From Word to Paradigm*. Cambridge: CUP.
- Walter, Mary Ann 2007 *Repetition avoidance in human language*. PhD diss., MIT.

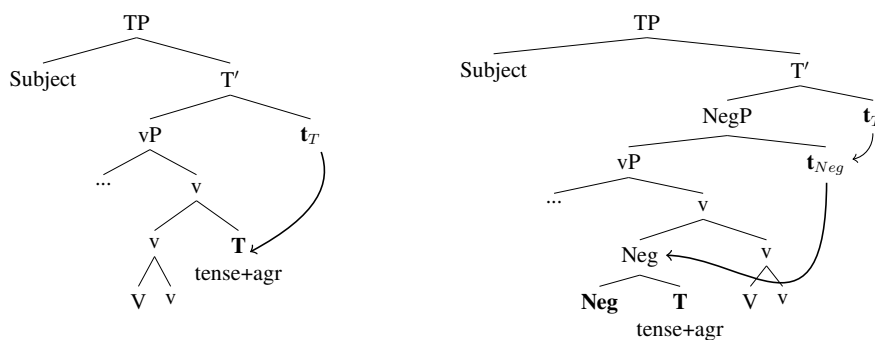
Towards a DM account of verbal morphophonology in Udmurt

Lena Borise & Ekaterina Georgieva
Hungarian Research Centre for Linguistics
lena.borise/ekaterina.georgieva@nytud.hu

Introduction: This paper demonstrates that the stress properties of indicative, imperative, and negated verbs in Udmurt cannot be straightforwardly accounted for by the Distributed Morphology (DM; Hale & Marantz 1993) approaches that derive stress placement from the positioning of category-defining heads like v^0 , n^0 , and a^0 (Embick 2010; Marvin 2003). Instead, we show that the Udmurt data are accounted for if other, non-category-defining functional heads are taken to determine stress placement, in line with Ultra-Massuet & Arregi’s (2005) analysis of Spanish, in which stress placement is determined with reference to T^0 . In our proposal, we rely on instrumental evidence on the prosodic realization of verbs in Udmurt (Borise & Georgieva 2021) and elaborate on the existing syntactic analysis of negated verbs (Georgieva et al. 2021).

Background: According to traditional descriptions, Udmurt has fixed final stress (e.g., *verá* ‘say.PRS.3SG’). There also are several classes of exceptions that have initial stress instead: notably, imperative and negated verbs (e.g., *véra* ‘say.IMP.2SG’, *uz véra* ‘NEG.FUT.3 say.CN.SG’). (Yemelyanov 1927; GSUJa I 1962; Denisov 1980; Winkler 2001). For the syntax of indicative and negated verbs, we adopt Georgieva et al.’s (2021) analysis: functional heads like T^0 and Neg^0 undergo Lowering (Embick & Noyer 2001) to form a complex head with the verb (1)-(2). In non-negative contexts, T^0 is linearized to the right of v^0 , as in (1). In negative contexts, Neg^0 is ‘picked up’ by T^0 , and the resulting complex head Neg^0 is linearized to the left of v^0 , as in (2). This is based on the word order and agreement facts: negation is expressed with a negative auxiliary, specified for tense, followed by the non-tensed connegative form of the lexical verb.

(1) Indicative (affirmative) verbs (2) Negated verbs (Georgieva et al. 2021)



→ Linearization: V-v-T+Agr

→ Linearization: Neg-T+Agr-V-v

Existing DM analyses of stress placement: In DM, lexical verbs are assumed to consist of an acategorial root and a category-defining head (v_{cat}) (represented as ‘V’ in the trees above). The categorizers are taken to be cyclic in the sense of phase theory – i.e., they trigger Spell-Out, while other heads (e.g., T^0) are not (Embick 2010). When a categorizing head is merged, the cyclic domains in its complement are sent to the interfaces. With stress assignment, Marvin (2003) proposes that in e.g., English words like *governmentalese*, which contain several categorizing heads: [[[[[rootP] vP] n₁P] aP] n₂P], Spell-Out is triggered for each phrase. The Main Stress Rule of English (Halle 1998) applies at vP, aP, n₂P and at the next higher phrase, giving rise to *góvèrnméntalése*, in which the stress assigned within previous Spell-Outs is preserved (in the form of secondary stress). In a different strand of DM work, Ultra-Massuet & Arregi (2005) propose that stress assignment in Spanish verbs is determined by the position of T^0 : stress targets the vowel immediately (linearly) preceding the T^0 node. T^0 is preceded by a right bracket that closes the metrical foot to the left of T^0 , and stress is assigned to the rightmost vowel of the foot: ...x) T^0 . Thus, both strands of analyses tie stress assignment to morphosyntactic

structure, but crucially differ with respect to which syntactic heads determine the domains for stress assignment: cyclic heads/categorizers (Embick 2010; Marvin 2003) or non-cyclic heads (Oltra-Massuet & Arregi 2005). We show that the Udmurt stress facts cannot be derived based on the distribution of cyclic heads and make reference to non-cyclic functional heads instead.

Proposal: We adopt Borise & Georgieva's (2021) conclusion that Udmurt has initial and final metrical stress (as opposed to the final, default stress being a non-metrical phrase-edge effect; cf. Jun & Fougeron 1995 for French), which is based on acoustic evidence like vowel quality and alignment with pitch accents. In line with Oltra-Massuet & Arregi (2005), the main tenet of our analysis is that T^0 , the highest functional head in the verbal spine (=HFX⁰), plays the crucial role in stress assignment. The Udmurt T^0 differs from its Spanish counterpart in one respect, though: it can be linearized to the right or to the left of v^0 , as in (1)–(2). Stress placement is derived by the following algorithm (simplified for brevity):

(3) *Stress assignment algorithm for Udmurt:*

- a. Insert a left parenthesis to the right of the HFX⁰ (i.e., T^0 in verbs), indicating the left edge of the metrical foot: T^0 (...
- b. Align stress with the left edge of the foot: T^0 (x...
- c. If no stress-bearing material is available to the right of the HFX⁰, move the left parenthesis one step to the left, then align stress with the left edge of the foot.

❶ **Indicative verbs:** The HFX⁰, T^0 , is linearized to the right of v^0 after Lowering, (1), and the left parenthesis is inserted to the right of T^0 . In verbs that contain overt material to the right of T^0 (e.g., overt agreement morphology), (3b) applies and stress is realized on that morpheme (e.g., *vetl-o-zí* 'GO-FUT-3PL'). Rule (3c) applies in two cases: (i) in verbs that contain an overt exponent of T^0 but no overt / syllabic material to the right of T^0 (e.g., *vetl-í* 'GO-PST[1SG]') and (ii) verbs that contain no overt exponent of T^0 (e.g., *verá* 'say.PRS.3SG'). These two rules derive final stress in indicatives, which is tied to the position of T after linearization.

❷ **Negated verbs:** For negated verbs, we adopt the structure in (2). The HFX⁰, T^0 , is combined with Neg⁰ and linearized to the left of v^0 , yielding Neg-T-V-v. The left bracket is inserted to the right of T^0 , and stress is correctly placed on the syllable following T^0 – i.e., the first syllable of the verb (e.g., *uz véra* 'NEG.FUT.3 say.CN.SG').

❸ **Imperative verbs:** We propose that imperative verbs have the structure [T [Σ_{IMP} [v ...]]]. The (phonologically zero) Σ^0 head (cf. Laka 1990) is also 'picked up' by T^0 and undergoes Lowering, in a parallel fashion to Neg⁰ in (2), yielding Σ -T-V-v. The Σ^0 head, combined with T^0 , is linearized to the left of v^0 . The $\Sigma^0 + T^0$ complex has no overt exponent, the left bracket to the right of T^0 places the left edge of the metrical foot at the left edge of the verb and correctly predicts initial stress placement (e.g., *véra* 'say.IMP.2SG').

Further discussion: In the talk, we will also discuss the implications of our analysis regarding stress assignment in verb clusters containing the (en)clitics *ni* 'already, anymore' and *na* 'still, yet'. These clitics can attach either to the lexical verb or to negation, yielding Neg-cl-V or Neg-V-cl orders (Arkhangelskiy 2014; Georgieva et al. 2021). Finally, we address the variation in stress placement that is found between some Udmurt dialects, and show how it can be captured within our proposal.

Selected references: Embick, D. 2010. *Localism versus Globalism in Morphology and Phonology* • Jun, S. & C. Fougeron. 1995. The accentual phrase and the prosodic structure of French. In: *Proceedings of ICPHS 13* • Georgieva et al. 2021. Negative verb clusters in Mari and Udmurt and why they require postsyntactic top-down word-formation. *NLLT* • Marvin, T. 2003. Is word structure relevant for stress assignment? In: *Distributed Morphology Today* • Oltra-Massuet, I. & K. Arregi. 2005. Stress by structure in Spanish. *LI*

Selectional restrictions of the French suffix *-ance*

Omar Gamboa Gonzalez

University of Toronto

omar.gamboagonzalez@mail.utoronto.ca

This study aimed to determine how the inner aspect and the semantic role of the subject of base-verbs play a role in the selectional properties of the deverbal suffix *-ance* (and its variant *-ence*); e.g., *exigence* ‘requirement’, *espérance* ‘hope’. Our main goal was to pinpoint the selection patterns related to this morphological form. Two questions guided the analysis: 1) what is the Aktionsart of the base-verbs? and 2) is agentivity a key factor in the selection of this suffix?

The research is motivated by several observations. On the one hand, it has been claimed that the morphology of the nominalizations in French is sensitive to the Aktionsart of the base-verb (Heinold 2010, Ferret & Villoing 2012). For instance, the suffix *-age* (*jardinage* ‘gardening’) selects mainly activity verbs while the stem 12 conversion (past participle conjugation, e.g. *percée* ‘an opening’) selects only telic verbs. On the other hand, other authors (Kelling 2001, Uth 2008) have pointed out that the agentivity of the subject also plays a role in the selection of the nominalization’s morphology. The authors claim that *-age* is selected only if the subject of the verb is [+agentive]. We stress the fact that we did not find any studies that address the role of both Aktionsart and agentivity for one single morphological form. We aim at filling this gap by determining if they are both equally relevant.

Methodology: We randomly selected 50 nouns ending in *-ance* that are morphologically associated with a verb from the online dictionary *Wiktionnaire*. We also verified that the noun had an eventive reading (event or state, cf. tests by Godard & Jayez 1996). The Aktionsart of each verb was then assessed via a series of tests (following Dowty 1979, Cann et al. 2009, Olsen 2014). We also conducted a semantic analysis to determine: 1) if the verb presented aspectual alternations related to the nature of the internal argument (singular, plural, or omitted) or due to a (potential) polysemy; 2) if the subject was [\pm animate]; and 3) if the subject was [\pm agentive] (imperative test and embedding under control verbs, cf. Beth & Rappaport Hovav 2005, Grafmiller 2013).

Results and impact of the research: None of the verbs present an aspectual alternation due to the internal argument’s nature, but 11 verbs have different readings according to the animacy of the subject, which results in a different Aktionsart. For instance, the verb *adhérer* with a [–animate] subject means ‘to stick’ (*la colle adhère au mur* ‘the glue sticks to the wall’); with a [+animate] subject it means ‘to join’ (*elle adhère au club* ‘she joins the club’). We treated these differences as two separate items. In total, 61 items were analyzed. Regarding our first research question, this suffix is not specific to one verb class. We found state verbs (*descendre* ‘be descended from’), stage-level states (*endurer* ‘to endure’), activities (*assister* ‘to aid’), achievements (*délivrer* ‘to deliver’), and semelfactives (*exiger* ‘to demand’). However, as opposed to *-age* and the stem 3 conversion (present tense conjugation, e.g. *voyage* ‘trip’), this nominalization may select [–dynamic] verbs: 30 out of the 61 items are [–dynamic] (3 states and 27 stage-level states). Moreover, the suffix is incompatible with accomplishments as there are no verbs of this class. As to our second question, we find both [–] and [+] agentive subjects. However, there are more non-agentive subjects: 39 out of 61. Therefore, contrary to *-age*, that selects only dynamic verbs with [+agentive] subjects, this nominalization concerns mainly (but is not restricted to) [–agentive] subjects, even if the verb is dynamic (e.g., *naissance* ‘birth’). An observation that supports this claim is that we found doublets for 5 verbs (e.g., *adhérence-adhésion*), and in all cases the *-ion* form corresponds to a [+agentive] reading of the subject, with dynamic verbs. This also suggests that the *-ion* suffix is mainly selected with [+agentive] subjects. We take these facts as evidence that both factors (Aktionsart and agentivity) play a role in the selection of the noun’s morphology in the case of *-age*, *-ion* and the converted nominals. We hypothesize that for *-ance*, the irregularities observed are due to the fact that in

some cases, the derivation comes from the adjective, and not from the verb, even if the nominalization accepts the eventive tests.

References

Beth, L. & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge: Cambridge University Press.

Cann, R., Kempson, R. & Gregoromichelaki, E. (2009). *Semantics. An Introduction to Meaning in Language*. Cambridge: Cambridge University Press.

Dowty, D. (1979). *Word Meaning and Montague Grammar. The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: D. Reidel.

Ferret, K. & Villoing, F. (2012). L'aspect grammatical dans les nominalisations en français: les déverbaux en -age et -ée. In R. Marin & F. Villoing (Eds). *Lexique 20. Nominalisations: nouveaux aspects*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

Grafmiller, (2013). *The Semantics of Syntactic Choice: An Analysis of English Emotion Verbs* (Doctoral dissertation, Stanford University, United States).

Godard, D., & Jayez, J. (1996). Types nominaux et anaphores : le cas des objets et des événements. *Cahiers de Chronos 1*, 41–58.

Heinold, S. B. (2010). *Verbal properties of deverbal nominals: an aspectual analysis of French, German and English* (Doctoral dissertation, Universität Stuttgart, Germany).

Kelling, C. (2001). Agentivity and Suffix Selection. *Proceedings of the LFG01 Conference*, 147-162.

Olsen, M.-B. (2014). *A Semantic and Pragmatic Model of Lexical and Grammatical Aspect*. New York: Routledge.

Uth, M. (2008). The division of the causative eventive chain by means of -ment and -age. *Working Papers of the SFB 732 Incremental Specification in Context 01*, 209-234.

Wiktionnaire. *Le dictionnaire libre*. <https://fr.wiktionary.org/>.

Morphology and spelling variation: A case study using spontaneous handwritten data

Stefan Hartmann, University of Düsseldorf, hartmast@hhu.de

Daniel Claeser, University of Bonn, claeser@uni-bonn.de

Kristian Berg, University of Bonn, kberg@uni-bonn.de

The relation between morphology and spelling is an important source of evidence for theories of linguistic processing (see e.g. Kuperman et al. 2009). In particular, spelling errors can help us assess the role of morphological structure in language users' mental representations of words in authentic texts (see e.g. Schmitz et al. 2018; Bar-On & Kuperman 2019; Surkyn et al. 2020). For instance, Gahl & Plag (2019) show that the strength of morphological boundaries affects misspellings of derivationally complex words in English. In a corpus study on non-edited Hebrew texts, Bar-On & Kuperman (2019) show that the occurrence of spelling errors is much less likely if it would disrupt a morphological unit.

These results suggest that some morphological units are more prone to spelling errors than others, partly depending on the degree to which they are perceived as separate units (rather than as parts of the words in which they occur). In this paper, we want to test this hypothesis using German data by exploring graphemic variation in a collection of 1,667 school-exit exams (Author3 et al. 2021). Specifically, we annotate the data for different types of spelling errors, and we code the spelling errors for their morphological structure. Taken together, these data allow for a multifactorial analysis of morphological factors that determine spelling errors.

Take, for example, the following spelling errors that are usually viewed as “slips of the pen”, i.e., errors that depend on general cognitive factors like concentration, attention etc. In other words, the writer knows the correct graphemic form of a word, but is – for some reason or other – not able to produce it.

- (1) a. <Lebewese> (instead of *Lebewesen* ‘living being’)
- b. <Gesellschaf> (instead of *Gesellschaft* ‘society’)
- c. <ermöglich> (instead of *ermöglicht* ‘enabled’)

These misspellings are instances of what may be called *graphemic erosion*: The attested form contains less material than what is expected. In all cases, it is the word ending that is affected (this is what we would expect, see e.g. Wing & Baddely 1980, 2009). All three forms are attested in the corpus, and we would expect the respective writers to be able to rectify them with enough time and concentration.

However, the forms differ regarding one crucial aspect: In (1a), it is the stem that suffers the erosion; in (1b), it is a derivational suffix; in (1c), it is an inflectional suffix. Preliminary analyses show that inflectional suffixes are more prone to graphemic erosion as compared to derivational suffixes and stems. We propose a thorough analysis to test this hypothesis. In particular, confounding factors like word frequency, length etc., but also the frequency distribution of word-final stems, inflectional and derivational suffixes in general have to be taken into account. This calls for the multifactorial analysis mentioned above, which can reveal factors that influence the degree to which different units are prone to erosion, and can as such shed light on the complex relationship between phonology, morphology, and spelling.

References

- Bar-On, Amalia & Victor Kuperman. 2019. Spelling errors respect morphology: a corpus study of Hebrew orthography. *Reading and Writing* 32(5). 1107–1128.
- Gahl, Susanne & Ingo Plag. 2019. Spelling errors in English derivational suffixes reflect morphological boundary strength: A case study. *The Mental Lexicon* 14(1). 1–36.

- Schmitz, Tijn, Robert Chamalaun & Mirjam Ernestus. 2018. The Dutch verb-spelling paradox in social media: A corpus study. *Linguistics in the Netherlands* 35. 111–124.
- Kuperman, Victor, Robert Schreuder, Raymond Bertram & R. Harald Baayen. 2009. Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance* 35(3). 876–895.
- Surkyn, Hanne, Reinhild Vandekerckhove & Dominiek Sandra. 2020. From experiment to real-life data: Social factors determine the rate of spelling errors on rule-governed verb homophones but not the size of the homophone dominance effect. *The Mental Lexicon* 15(3). 422–463.
- Wing, Alan M. & Alan D. Baddeley. 1980. “Spelling errors in handwriting: A corpus and distributional analysis”. In *Cognitive processes in spelling*, Edited by: Frith, U. 251–285. London: Academic Press.
- Wing, Alan M. & Alan D. Baddeley. 2009. Righting errors in writing errors: The Wing and Baddeley (1980) spelling error corpus revisited, *Cognitive Neuropsychology*, 26:2, 223-226, DOI: 10.1080/02643290902823612

**(How much) does morphological change in the Romance verbal paradigm
abide by morphomic (vs extra-morphological) templates?**

Borja Herce
University of Zurich

The autonomy of the morphological component of language from syntax and semantics has been a central debate in the field over the last three decades (see Aronoff 1994, Luís & Bermúdez-Otero 2016, etc.). Although so-called 'morphemes' (i.e. exclusively morphological structures) are also found in many other language families (see Herce 2020), the Romance verb has dominated the discussion on what these structures are like (see e.g. Maiden 1992, 2011, 2018; Esher 2015; Herce 2021; Nevins et al. 2015, etc.). This research has mostly concluded that morphemes can constitute resilient productive templates for the analogical (re)distribution of (allo)morphs in the paradigm. However, due to the overwhelmingly qualitative approach of this research to date, and the focus on morphemes themselves and the attested morphological changes that support their importance, we are still lacking a fair quantitative and replicable assessment of exactly how productive morphomic (vs extra- morphological) templates are as drivers of analogical change.

This paper attempts to quantify the strength of the morphomic templates known as L/U, N, PYTA, and FUÈC in the history of Romance. To do so, the morphological analogical developments surrounding the deletion of stem-final consonants (e.g. Spanish *veo* < Latin VIDEO) will be surveyed in 22 cognate-verb sets (with a single-segment stem coda prone to being lost: AUDIO, BIBO, CADO, CREDO, DEBO, FUGIO, TRAHO, VIDEO, VIVO, etc.) across 65+ Romance varieties (those in the Oxford Online Database of Romance Verb Morphology, Maiden et al. 2010). This concrete phenomenon constitutes an ideal one for assessing the productivity of morphemes because of two factors: a) it involves forms and segments (/d/, /b/, /g/...) unrelated to the ones that are usually involved in L/U, N, PYTA, and FUÈC morphemes, and b) the lenitions and morphophonological changes that introduced alternations in these verbs happened after the split of Proto-Romance into largely independent local varieties. Thus, they provide multiple independent datapoints on the late vitality of morphemes compared to other possible extra-morphological templates.

Surveying these changes, a database of 502 morphophonological innovations was compiled, where the paradigmatic domains of application of each innovation was coded to assess whether it abided by established morphomic templates, or whether it deviated from them, and in which ways. Focusing on the "classical" morphemes, the following scale of productivity was found: N (27) >> PYTA (8) > L (6) > FUÈC (5). This suggests the N-morpheme is the most productive one in Romance, by a comparatively large margin. The raw productivity of morphomic templates, (46, 9.16%), however, seems to be relatively small compared to the much larger number of morphophonological changes (386, 76.89%) giving rise to alternations that do not abide by them, or giving rise to levelled (i.e. non-alternating) paradigms (70, 13.94%).

The exact interpretation of these rich data is, however, a complex matter. One important issue is that raw numbers might be comparatively meaningless compared to a more sophisticated analysis of the expected random probability of established morphomic patterns *vis a vis* other paradigmatic arrangements. In a comparatively large paradigm like the Romance one, it is exceedingly improbable (less likely than winning the lottery) for exponents to adopt one of the 4 established morphomic distributions by chance. Against this background, 9.16% does not seem a negligible percentage. Another important point is that exponents/alternations may abide by the "spirit" but not the "letter" of autonomously morphological templates in that they may respect the stem-spaces (Boyé & Cabredo-Hofherr 2006, Montermini & Bonami 2013, Hecce 2019) that cross-cutting morphemes give rise to in the paradigm. When this possibility (understandable from the perspective of abstractive morphology and conditional-entropy-based research on morphological predictability) is allowed, 192 (38.25%) of the documented innovations (and trivially the 70 that result in a nonalternating paradigm) abide by these morphomic templates. Although these still do not reach half of the alternations, their importance can be thus more clearly appreciated.

The presentation and interpretation of the present results would not be complete without a look at those morphophonological innovations, the majority, which deviate from morphomic templates irreconcilably. The most frequent deviations have been found to involve paradigm cells that are semantically peripheral to their stem-space or morpheme, and/or comparatively frequent as per token frequency (e.g. nonfinite forms like the infinitive (87) or gerund (69), the singular imperative (46), the 1SG.PRS.IND (37), etc.). This reveals that semantics and frequency of use (see Bybee 2007) constitute, on a par with autonomously morphological templates, forces of the utmost importance to predict the occurrence and scope of analogical innovations in the paradigm. The 'autonomy' of morphology, thus, can only be understood as an affirmation that structures that concern just morphology do exist and can play an important role in analogical change, and not as a claim that almost everything else (e.g. semantic natural classes, cf. Blevins forthcoming) is irrelevant in paradigm synchrony and diachrony.

References

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge (MA): MIT press.

Blevins, James P. forthcoming. Two frameworks of morphological analysis. *Linguistic Analysis*.

Boyé, Gilles and Patricia Cabredo Hofherr. 2006. The structure of allomorphy in Spanish verbal inflection. *Cuadernos Lingüística Instituto Universitario Ortega y Gasset* 13: 9-24.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Esher, Louise. 2015. Morphemes and predictability in the history of Romance perfects. *Diachronica* 32, 4: 494-529.

Herce, Borja. 2019. Morpheme interactions. *Morphology* 29.1: 109-132.

Herce, Borja. 2021. Stress and stem allomorphy in the Romance perfectum: emergence, typology, and motivations of a symbiotic relation. *Linguistics*.

Luís, Ana R. & Ricardo Bermúdez-Otero (eds). 2016. *The morpheme debate*. Oxford: Oxford University Press.

Maiden, Martin. 1992. Irregularity as a determinant of morphological change. *Journal of linguistics* 28, 2: 285-312.

Maiden, Martin. 2011. Morphophonological innovation. *The Cambridge history of the Romance languages* 1: 216-267.

Maiden, Martin. 2018. *The Romance Verb: Morphomic Structure and Diachrony*. Oxford: Oxford University Press.

Maiden, Martin; John Charles Smith; Silvio Cruschina; Marc-Olivier Hinzelin and Maria Goldbach. 2010. *Oxford Online Database of Romance Verb Morphology*. University of Oxford. Online at <http://romverbmorph.clp.ox.ac.uk/>

Montermini, Fabio and Olivier Bonami. 2013. Stem spaces and predictability in verbal inflection. *Lingue e linguaggio* 12, 2: 171-190.

Nevins, Andrew, Cilene Rodrigues, and Kevin Tang. 2015. The rise and fall of the L-shaped morpheme: diachronic and experimental studies. *Probus* 27, 1: 101-155.

The syntagmatic properties of agreement morphs and the learning bias against more unnatural paradigmatic patterns

Borja Herce, Carmen Saldana & Balthasar Bickel
University of Zurich

Agreement markers that refer to the same feature or argument tend to be found in the same syntagmatic position; however, little is known about the typology of the exceptions to this trend. In this study, we explore the syntagmatic properties of S, A, and P person-number agreement markers in a phylogenetically diverse sample of 227 languages from 97 different stocks (data from the AUTOTYP database [Bickel et al. 2017] plus the languages in the WALS [Dryer & Haspelmath 2013] 100-language sample). The result is 325 person-number paradigms, whose agreement morphs' syntagmatic order properties were surveyed to explore trends and split patterns. We find that a majority of agreement paradigms only require reference to a single syntagmatic slot, thus obeying the principle of 'category clustering' [Mansfield et al. 2020] and displaying no paradigmatic splits. A sizable minority (128 paradigms, 39.38%), however, require reference to two or more syntagmatic slots and showed paradigmatic splits by which different person-number morphs appear in different syntagmatic slots. The question was then to assess to what degree person-number values sharing the same syntagmatic properties tend to conform to semantically natural classes. The following types were identified based on the degree of feature-value overlap.

	GUMER Imperfective		KOASATI Active		BASQUE Present	
	SG	PL	SG	PL	SG	PL
1	ə-kəft	ni-kəft-inə	há:lo-l	il-há:l	na-bil	ga-bil-tza
2	tɪ-kəft	tɪ-kəft-o	is-há:l	has-há:l	za-bil-tza	za-bil-tza-te
3	yɪ-kəft	tɪ-kəft-o	há:l	há:l	da-bil	da-bil-tza

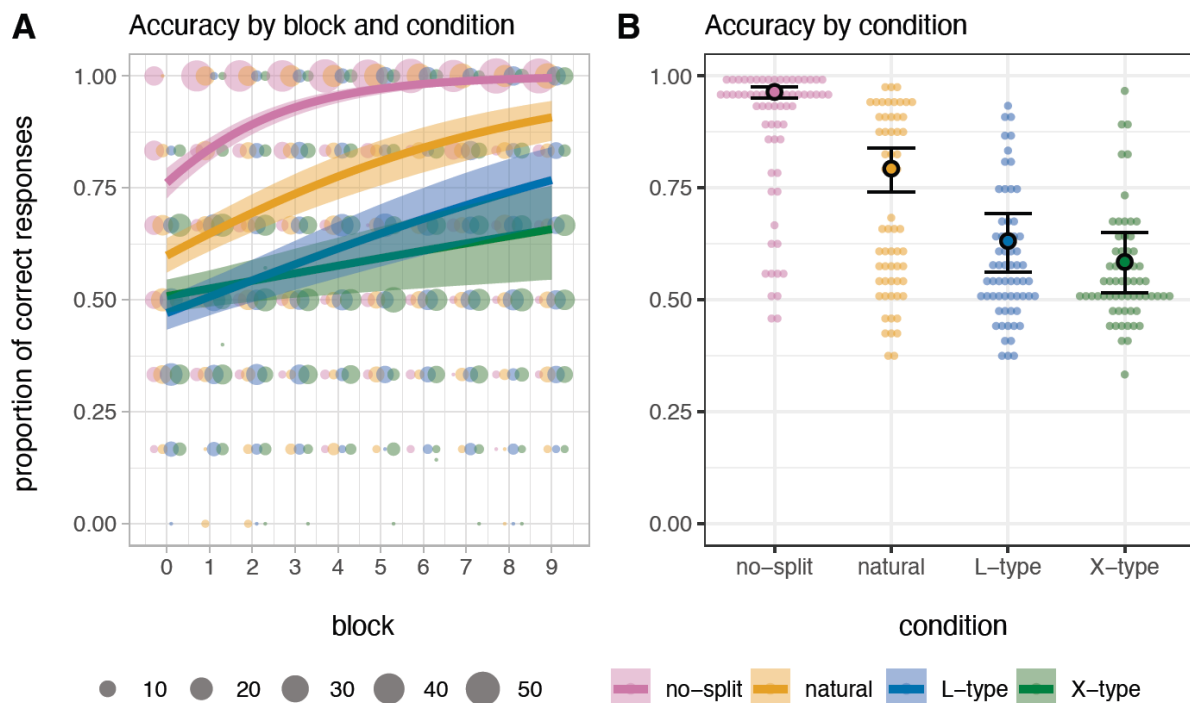
Naturalness types of syntagmatic identity patterns: N, L, X

The Gumer paradigm (verb 'open', Völlmin 2017:122) shows the most natural type (N), in which all cells sharing the same syntagmatic property (circumfixal marking) also share a value (plural). As a result, all circumfixally-marked pairs of cells share 50% of their values. The Koasati paradigm (verb 'hear', Kimball 1985:55) shows a less natural pattern (L), in which cells sharing the same syntagmatic property (prefixal marking) share 33.3% of their values on average. Finally, the Basque paradigm (verb 'walk', Hualde & Ortiz de Urbina 2011:234) shows a more unnatural pattern still (X) whose component cells only share 16.6% of their values on average.

In our cross-linguistic sample, we found 44 Ns, 73 Ls, and 24 Xs. These raw numbers, however, need to be properly interpreted, since each type has a different probability of occurring by chance (note, for example, that there are only 2 logically possible Ns of size three (1PL=2PL=3PL and 1SG=2SG=3SG) but 12 logically possible Ls (e.g. 1SG=2SG=2PL, 1SG=2SG=1PL, 1SG=3SG=1PL, 1SG=3SG=3PL, etc.). Controlling for these asymmetries,

and the non-independence of paradigms from the same language and family, we found, by means of a mixed effects regression model, that the most natural N-patterns are statistically significantly overrepresented in natural languages. The most unnatural X-patterns, in turn, are statistically significantly underrepresented in natural languages, and intermediate-naturalness L-patterns do not occur with a significantly different frequency from their chance level. These findings argue quite forcefully in favour of a preference for naturalness in languages, and for a gradient conception of this dimension.

To verify whether this typological tendency is grounded in a bias toward more natural patterns during learning and transmission, we conducted an artificial language learning experiment (N=552, recruited over Amazon Mechanical Turk). We used an ease-of-learning paradigm where we trained and tested participants on a person-number verbal agreement paradigm with a specific pattern of morphological syntagmatic splits: agreement was marked cumulatively in a single affix that could appear in a different position (e.g., suffixation, prefixation or zero-marking) depending on the person-number feature value bundle. After an initial phase where participants learnt an artificial lexicon of 6 pronouns and 3 verbs, they learned a verbal agreement paradigm with syntagmatic splits conforming to either N, L or X patterns over ten blocks of six trials each. We ran a further control condition where we taught participants a system of person-number agreement where all markers were either suffixes or prefixes.



Results support the greater learnability of natural N-patterns (orange) relative to types L (blue) and X (green). We also found that the increase in accuracy by block was higher in L paradigms than in X, thus matching the natural language tendencies of the different types, and supporting both the cognitive relevance of naturalness for morphological structure learning, as well as its gradient nature. Altogether, our findings suggest, therefore, that we should move beyond dichotomic natural-vs-unnatural and syntagmatic-vs-paradigmatic descriptions and analyses of

morphemic-vs-morphomic distributions. At the same time, our experimental results indicate that cognitive biases in category formation might be (at least partly) driving the patterns observed in natural languages.

References

Bickel, Balthasar & Nichols, Johanna & Zakharko, Taras & Witzlack-Makarevich, Alena & Hildebrandt, Kristine & Rießler, Michael & Bierkandt, Lennart & Zúñiga, Fernando & Lowe, John B. 2017. The autotyp typological databases. Version 0.1. 0. Online: <https://github.com/autotyp/autotyp-data/tree/0.1.0>.

Blevins, James P. forthcoming. Two frameworks of morphological analysis. *Linguistic Analysis*.

Dryer, Matthew S & Haspelmath, Martin. 2013. The world atlas of language structures online. leipzig: Max planck institute for evolutionary anthropology. Online: <http://wals.info>.

Kimball, Geoffrey David. 1985. A descriptive grammar of koasati (louisiana): Tulane University dissertation.

Hualde, José Ignacio & Ortiz de Urbina, Jon. 2011. A grammar of Basque. Walter de Gruyter.

Maiden, Martin. 2016. Morphemes in diachrony. In Ana R. Luís & Ricardo Bermúdez-Otero. (Eds.), *The Morpheme Debate*. Oxford: Oxford University Press.

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

Völlmin, Sascha. 2017. Towards a grammar of gumer-phonology and morphology of a western gurage variety: University of Zurich dissertation.

Periphrastic constructions and inflection internal to derivation

Ágnes Kalivoda¹, Farrell Ackerman², Robert Malouf³

¹ Hungarian Research Centre for Linguistics

² University of California, San Diego

³ San Diego State University

kalivoda.agnes@nytud.hu, fackerman@ucsd.edu, rmalouf@sdsu.edu

Generative tree-theoretic approaches (Bruening 2018) and construction-theoretic approaches (Booij 2002, Loos 2012, Goldberg 2019) argue against standard (and largely superseded) lexicalist conceptions of grammar analysis which assume two components, i.e. a lexical component, conventionally restricted to the association of lexical representations with familiar synthetic exponence of words and a syntactic component responsible for the clausal distribution of these words. Both argue for the sufficiency (and efficiency) of a single component approach, but with very different hypotheses concerning the single supervening component. Bruening provides empirical arguments suggesting the difficulty of identifying the categorical differences entailed by a sharp distinction between syntax and the lexical/morphological component. The desired inference is that all of the relevant phenomena are best treated in a single tree-theoretic component. However, as we will argue, evidence for Hungarian suggests that the single component is best viewed construction-theoretically where grammatical phenomena are organized in terms of patterns of different types participating in networks of relatedness. We argue that once construction-theoretic proposals have been augmented with paradigmatic and horizontal relations between constructions that the purportedly paradoxical behaviors exhibited by Hungarian *pv v* constructions turn out to be artifacts produced by inappropriate theoretical assumptions (Diessel 2004, Diewald 2020, Van de Velde, Maekelberghe, and Fonteyn 2021, Diewald and Politt 2022). They are natural and common constructions that recur in numerous languages which all share similar grammaticalization profiles (see Ackerman, Kalivoda, Malouf workshop presentation).

Morphologists often distinguish between inflection and derivation distinguishable according to numerous criteria (see, e.g., Dressler 1989, Corbett 2010, Štekauer 2015). Researchers have often operated according to a morphotactic hypothesis concerning the relative order of derivational and inflectional affixes with respect to one another and with respect to their closeness to a modified stem (Greenberg 1966:93).

There have been numerous (and often contested) examples where inflectional marking is flanked (either preceding or following) by derivational markers (Bochner 1983, Booij 1993, Rainer 1996, Rice 1985). From a cross-linguistic perspective, this is predicted to occur whenever e.g., a piece of a complex predicate bears inflectional suffixes when functioning as a prefix or bears inflectional prefixes when functioning as a suffix in combination with a verbal stem: $px_{der}\text{-infl } V$ or $V \text{ infl-sfx}_{der}$. In other words, this phenomenon occurs when language change creates a sequence of markers where the inflectional markers are maintained or trapped (Harris and Farlund 2006), rather than externalized (Haspelmath 1993).

We demonstrate that Hungarian periphrastic predicate constructions consisting of separable inflected preverbs and associated verb stems exemplify inflection internal to derivation. Hungarian possesses an enormously productive pattern of periphrastic predicate formation that combines a member from a large inventory of separable preverbs with a verb stem (Kömlösy 1992, Piñón 1991, Kiefer 1995, Kiefer & Ladányi 2000, Forgács 2004, Kalivoda 2021).

We situate complex predicates within a sample of related constructions and argue that the recognition of this family network of constructions provides the resource for native speakers to generate forms of preverb + verb constructions they have never encountered, in fact, don't need to encounter: in effect, this suggests that morphological organization based on analogy solves what otherwise appears to be the poverty of the stimulus status associated with

Hungarian complex predicates. This is an instance of so-called Paradigm Cell Filling Problem from intra-paradigmatic and inter-paradigmatic implicative inflectional relations to implicative relations among interleaving inflectional and derivational constructions, sometimes expressed synthetically and sometimes periphrastically.

The morphotactic distribution of these inflectional markers is typologically noteworthy in at least two respects. First, it exemplifies inflection internal to derivation which occurs irrespective of the synthetic or periphrastic status of complex predicates, and second, it represents an unusual instance of separable preverbs inflecting to satisfy the pronominal status of the argument requirements resulting from the new predicate created by the preverb + verb combination. The analysis of this subtype of pv v combinations in terms of networks of relations among constructions provides further evidence of the value of developing notions of paradigmatic relations within construction theoretic approaches, as inspired by paradigm-based approaches to morphology, and for the explanatory importance, more generally, of paradigmatic organization in contrast to tree-theoretic syntagmatic assumptions.

Morphosyntactic constraints on phonological dominance

Maria Kouneli (U. Leipzig), Paula Fenger (U. Leipzig), Jonathan David Bobaljik (Harvard U.)
 maria.kouneli@uni.leipzig.de, paula.fenger@uni.leipzig.de, bobaljik@fas.harvard.edu

The ‘no dominant prefixes’ generalization It has been observed that in languages with dominant-recessive harmony systems, while either roots or suffixes can be dominant (i.e., they can trigger a phonological change in their surrounding elements), prefixes are always recessive (Clements 2000, Baković 2000, Casali 2003 a.o.). In Kipsigis (Kalenjin; Kenya) (see (1)), a language with [ATR] vowel harmony, a [+ATR] vowel anywhere in the word will cause all [-ATR] vowels to shift to [+ATR]. Either roots (1a-b) or suffixes (1c) can be harmony triggers, but prefixes are always recessive [-ATR] morphemes (1b,c,d) (Hall et al. 1974).

- | | | |
|-----|----------------------------------------------|-----------------------------------------|
| (1) | a. /ŋo:k-ɪ/ → ŋo:gi
dog-DEM | c. /a-tʃam-e/ → aʃame
1SG-love-IPFV |
| | b. /ka-kr-pet/ → kaʒibet
PST-1PL-get.lost | d. /ka-ɔ-tʃam/ → kaɔʃam
PST-2PL-love |

Our main claim: ‘no dominant high affixes’ While the ‘no dominant prefixes’ generalization is acknowledged in most studies of dominant harmony systems, there is, to our knowledge, no systematic explanation for the pattern. Some accounts treat this as an accidental property of the prefix inventory (Baković 2000), others suggest that the juncture between prefixes and the (lexical) root/stem has a different status from that between stem and suffixes (Moskal 2015).

We explore here an alternative, under which the generalization about prefixes is epiphenomenal, a special case (when it holds) of a broader generalization that phonological derivations of morphologically complex words proceed cyclically, where cycles may correspond to syntactic phases.

	INFL high	DERIV low	ROOT	DERIV low	INFL high
no dom pref	✗	✗		✓	✓
no dom high	✗	✓		✓	✗

Following Newell 2008, Fenger 2020 a.o., we propose that elements in the first domain may have unrestricted phonological interactions, including structure-changing operations, but that beyond the first phase/domain, phonological operations may not change material that is fixed on the first cycle. For harmony, this predicts that ‘low’ affixes may be dominant (root-altering) or recessive, but that ‘high’ affixes that participate in harmony (contra Fábregas & Krämer 2020) may only be recessive. Since it is independently established that prefixes tend to represent syntactically higher morphemes than suffixes (Julien 2002), ‘no dominant prefixes’ as a trend falls out as a special case. Table 1 schematizes the differences between our generalization and the no dominant prefix generalization, using inflection vs. derivation as a rough (but imperfect) proxy for syntactic height of affixes.

Support Preliminary support for our proposal comes from a close investigation of the verbal suffixes as well as the prefixes in Chukchi (Chukotko-Kamchatkan), Kipsigis, and Turkana (Nilotic). Replacing ‘no dominant prefixes’ with ‘no dominant high affixes’ offers two advantages: i) we can account for ‘exceptionally’ dominant prefixes and we can also provide an explanation for dominance patterns in suffixes (which are not usually discussed in this context in the literature), ii) following Fenger (2020), we can provide a principled explanation for why morphemes high in the syntactic structure cannot be dominant, and possibly, for why there may be a verbal versus nominal asymmetry among suffixes.

The data Our investigation of verbs in Kipsigis (Toweett 1979 and original fieldwork), Turkana (Dimmendaal 1983) and Chukchi (Skorik 1977, Dunn 1999), three languages from two unrelated families that have been mentioned as lacking dominant prefixes, confirms that the ‘no dominant high’ generalization is more accurate: the most peripheral affixes in all three languages are always recessive. In **Chukchi** the majority of verbal inflectional prefixes are recessive, but so are all of the 20 or so most peripheral (agreement) suffixes. Moreover, many sources (Bogoras 1922, Skorik 1977, Weinstein n.d.) do in fact

report dominant prefixes in Chukchi verbs, however these are all derivational, such as the intensifier in (2), (hence arguably low).

- (2) /kət-γənt-et-rkən-i-tək/ → kət-γənt-at-rkən-e-tək
 NTNS-run-DERIV-ASP-E-2PL ‘Run!’ (Skorik 1977:77)

In **Kipsigis**, there is one dominant inflectional suffix, but importantly, it is an aspectual suffix. Under Fenger’s (2020) proposal (for stress in Turkish and pitch accent in Japanese) the relevant phase boundary delimiting the cut-off between high and low (verbal) affixes is not precisely derivation vs. inflection but instead between Aspect and Tense. Thus the existence of a Kipsigis dominant aspectual suffix conforms to our theory. We also find a dominant number “agreement” morpheme in **Turkana**. This morpheme, however, appears close to the root and precedes voice and aspect morphology. This suggests that it corresponds to a head below Aspect (rather than spelling out agreement higher in the clause), which would also conform to our generalization.

Beyond verbs While our proposed generalization thus seems to provide a better description of the distribution of dominant elements in verbs than the ‘no dominant prefixes’ generalization, it remains a theoretically open question what the prediction is for nouns and adjectives. If Fenger (2020) is correct that the relevant morphophonological domain in verbs is the ‘phase’ boundary between Aspect and Tense, where, if anywhere, is the corresponding boundary in complex nouns? Our investigation of Kipsigis, Turkana and Chukchi suggests that nouns (and adjectives at least in Kipsigis) have no corresponding internal phase boundary. In Chukchi, various case suffixes—the most peripheral affixes in the nouns—may be dominant (schwa may be diacritically dominant in Chukotko-Kamchatkan) and will overwrite the vowel quality of preceding morphemes including the root and prefixes (3). In Kipsigis, an adjectival plural marker is dominant, as shown in (4). Our proposal thus does not, in principle, exclude dominant prefixes in nouns in these languages, although there are very few nominal prefixes (as compared to verbal ones) in these languages. We do note that dominant prefixes are reported for Tunen and Kibudu (Moskal 2015) in both cases within the nominal system.

- (3) /umk-čəku-γtə/ → omk-ə-čəko-γtə/
 bush-INNESS-ALL ‘into the bushes’ (Dunn 1999:283)
- (4) /mʊgʊl-e:n/ → mugule:n
 round-PL

Conclusion and Discussion Fenger (2020), developing ideas in Newell (2008) and elsewhere, proposes that the mapping from morpho-syntactic structure to phonological structure proceeds in ‘phases’. In the verbal system, the first domain includes derivational morphemes (and root compounds in Chukchi) as well as the lowest inflectional morphemes (Aspect). When this domain is spelled out, phonological operations may occur in any direction, and dominant affixes may overwrite the quality of the root vowel. After this step, certain phonological properties of the first domain are fixed and may no longer be overridden, and thus subsequently integrated morphemes may only be recessive. Coupled with the trend for prefixes to be ‘high’ (Julien 2002), the no dominant prefix generalization in verbs, where it holds, is a special case of a broader generalization that also restricts the distribution of dominant elements in suffixes. Finally, if our analysis is on the right track, the ability of a morpheme to trigger harmony could be used as a diagnostic for phase detection in the syntax.

Sel. Refs Baković, E. 2000. *Harmony* PhD Rutgers. • Casali, R. 2003. [ATR] value asymmetries. *Linguistic Typology* • Dimmendaal, G. 1983. *The Turkana Language*. • Dunn, M. 1999 *Chukchi* PhD. Australian Nat’l U. • Fábregas, A. & M. Krämer 2020 In *Evolutionary Linguistic Theory*, 84-111. • Fenger, P. 2020 *Words* PhD UConn. • Julien, M. 2002. *Syntactic Heads* OUP. • Moskal, B. 2015 *Domains* PhD UConn. • Newell, H. 2008 *Aspects* PhD McGill. • Skorik, P. 1961-1977. *Grammatika chukotskogo jazyka* Nauka • Toweett, T. 1979. *Kalenjin linguistics*.

How do penguins differ from kangaroos? Vowel 'dropability' in the pluralization of vowel ending loanwords in Hebrew

Lior Laks

Bar-Ilan University

This study examines variation (and lack thereof) in pluralization of Hebrew loanwords that end with a vowel. Most loanwords take the suffix plural *-im*, as demonstrated for *pingwin-im* 'penguins' (1). In contrast, words that end with vowels other than *a*, demonstrate variation where at least four options can be found. The word *kenguru* 'kangaroo' can take the suffix *-im* (2a), it can take the same suffix with the deletion of the final vowel (2b), it can retain its singular form (2c), and it can surface in its English plural form (2d)

(1) *higi u od šloša pingwin-im*

'three more penguins arrived' (<https://nanitriptonezeland.wordpress.com/2016/11/>)

(2) a. *šloša kenguru-im še hayu ba-kluv barxu*

'three kangaroos that were in the cage escaped' (<http://www.ynet.co.il/articles/0.7340.L-4268082.00.html>)

b. *šney kengur-im ravu al nekeva*

'two kangaroos were fighting for one female' (<https://forum.12p.co.il/index.php?showtopic=24170>)

c. *krav egrofim ben šney kenguru*

'a feast fight between two kangaroos'

(http://holesinthenet.co.il/uncategorized/%D7%A8%D7%A7_%A0%D7%99_%D7%A7%D7%A0%D7%92%D7%95%D7%A8%D7%95_%D7%94%D7%95%D7%9C%D7%9B%D7%99%D7%9D_%D7%9E%D7%9B%D7%95%D7%AA_%D7%91%D7%A8%D7%97-4)

d. *im lahaka šel 34 kanguru-z*

'with a group of 34 kangaroos' (<http://mynetkarsaba.co.il/article/140291?OriginalPostId=140478>)

Why does such variation occur? Hebrew nouns ending with vowels other than *a* are relatively rare. The Hebrew morphological mechanism is not accustomed to pluralizing such words and as a result, speakers apply a variety of strategies. This study focuses on the competition between plural forms in (2a) and (2b), namely variation in deleting and not deleting final vowels. I propose a hierarchy of vowel 'dropability' which predicts which vowels are more or less likely to be dropped. It is based on the interaction of markedness and faithfulness constraints, as well as on paradigm accessibility in word formation.

Hebrew pluralization. Hebrew has two plural suffixes: *-im* (*sir-sir-im* 'pot') is typical of masculine nouns, and *-ot* is typical of feminine nouns (*sira-sirot* 'boat'). However, this system is subject to a great extent of irregularity (Berman 1978, Schwarzwald 1991, Ravid 2004, Ravid & Schiff 2009, Schwarzwald 1991). For example, the masc. noun *šulxan* takes *-ot* (*šulxan-ot*), and the fem. noun *beyca* 'egg' takes *-im* (*beycim*). Loan words demonstrate high regularity. Almost all of them are perceived as masculine and take *-im* (*imeyl-imeyl-im* 'email', *skedyul-skedyul-im* 'schedule'). Words that end with *a* are perceived as feminine and take *-ot* (*pica-pic-ot* 'pizza'), as in native words.

Unlike most loanwords, those that end with a vowel other than *a* demonstrate variation (2). Such variation results from the competition between a faithfulness and a markedness constraint. The faithfulness constraint is Maximality, which penalizes deletion, and as a result, the plural form is more faithful to the base and the structural relations between the two forms is transparent. The markedness constraint is the tendency to avoid sequence of two vowels as *oi* (*avukadoim* 'avocados') or *ui* (*tiramisuim* 'tiramisus'). Such sequences are marked in Hebrew and therefore speakers tend to avoid them.

I propose the following hierarchy of vowel dropability (3), where *a* is systematically deleted, *o* and *u* are rarely deleted, and *i* in an intermediate category, with a relatively stronger tendency to be deleted. *e* ending loanwords are not addressed in this study because they are rare and the few example that were found also involve orthographic matters and a diachronic analysis.

(3) $a \gg i \gg o, u$

I now turn to examine difference between vowel ending loanwords.

a-ending words. Loanwords that end with *a* take the plural suffix *-ot*, similarly to native words. The *a* in native words can be a suffix attached to a masc. base (*xatul-xatula-a* 'cat'), part of a pattern (e.g. *maCCeCa* as in *magrefa* 'rake'), part of a derivational suffix like *-iya* (*guf* 'body'– *gufiya* 'singlet'), and part of a word where no base exists independently (*uga* 'cake'). Regardless of the origin of *a*, words that end with it are mostly feminine, take the suffix *-ot*, and *a* is dropped systematically. This makes the deletion of *a* highly frequent and accessible in many paradigms, and this projects itself onto the behavior of loanwords, e.g. *lazanya-lazanyot*/**lazanya-ot* 'lasagna', *opera-oper-ot*/**opera-ot* 'opera', *pijama-pijam-ot*/**pijama-ot*

'pajamas'. *a* is perceived as an element that can be easily deleted, while the structural relations between singular and plural forms are transparent.

o/u-ending words. Most loanwords that end with *o* or *u* typically do not undergo deletion. This is because native words that end with these vowels are highly rare, and within these rare cases *o* and *u* are not deleted, e.g. *mašehu*– *mašehu-im* 'something'. As a result, *o/u* deletion is not attested in any paradigms and these vowels are perceived as an integral part of words. Deletion would result in low structural transparency between singular and plural forms. However, as shown in (2b), there are cases where *o/u* deletion is possible alongside retaining them. This is partially predictable. *o/u* can be deleted only if the base consists of more than two syllables. Compare *mango* 'mango' and *avukado* 'avocado'. The former is disyllabic and can take only one plural form (4), while the latter has four syllables and can take both plural forms, with (5a) and without (5b) deletion.

(4) ha-ec šelanu heniv hašana kol-kan harbe **mango-im**/***mangim**

'our tree yielded so many mangos this year' https://www.instagram.com/p/CTM_Be3orkJ/

(5) a. laxtox le-xatixot ktanot 5-6 **avukado-im**, limox ve-lehosif melax ve-pilpel

'cut 5-6 avocados into small pieces, squash and add salt and pepper'

<http://www.waterandflour.co.il/%D7%A6%D7%9E%D7%97%D7%95%D7%A0%D7%99-%D7%A7%D7%9C%D7%99%D7%9C-%D7%A1%D7%A0%D7%93%D7%91%D7%95%D7%99%D7%A5-%D7%9E%D7%9C%D7%97%D7%9D-%D7%AA%D7%A4%D7%95%D7%90-%D7%A2%D7%9D-%D7%90%D7%91%D7%95%D7%A7-2/>

b. ani yaxol lenaxeš et ha-matkon, kanire 5-6 **avukad-im**

'I can guess the recipe, probably 5-6' <http://southamericawithoutabike.blogspot.com/2014/04/blog-post.html>

Note that plural forms that demonstrate deletion are far more rare than the ones that preserve the final vowel. In addition, not all loan nouns that exceed two syllables demonstrate variation in pluralization. However, the essential finding here is the contrast between (i) words that systematically retain the last vowel; and (ii) words that demonstrate some variation. Words that consist of less than three syllables systematically retain the last vowel in order to preserve the structure of a minimal word of the stem. Previous studies have demonstrated the role of minimal word in both morphological and phonological processes and specifically in the application of deletion (Prince 1980, Broselow 1982, McCarthy & Prince 1986, Ussishkin 2000, Bat-El 2008). The richer the base is in terms of number of syllables, the easier it is to delete the last vowel. If it has less syllables, vowel deletion would result in low structural transparency between the forms.

i-ending words. Loanwords that end with *i* are an intermediate category. In some such words deletion is obligatory, while in others it is optional. Examine the words *ravyoli* 'ravioli' and *imoji* 'emoji'. The former undergoes pluralization with and without deletion (*ravyol-im/ravioli-im*), and while the latter only has a plural form with deletion (*imoj-im*/**imoji-im*). Similarly to loanwords ending with *a*, *o* and *u*, the behavior of loanwords ending with *i* can be predicted based on plural paradigms of native words. There is a small set of native words that end with *i* and their behavior is unpredictable, e.g. *dli*–*dalyim* 'bucket', *ci*–*ciyim* 'fleet' and *pri*–*perot* 'fruit'. In some cases *i* is deleted, while in others it is retained and there are also morpho-phonological alternations. In addition, Hebrew has native words that end with *-i*, which is a derivational suffix that derives adjectives from nouns, *levanon* 'Lebanon'–*levanon-i* 'Lebanese'. Some of these words, and typically the ones denoting nationalities, can function both as nouns and adjectives. They are pluralized with the suffix *-im*, and the behavior of the final *i* depends on the lexical category. As adjectives, the *i* can be deleted and can be retained, and in the latter case a glide consonant is inserted between the two vowels (*levanon-im/levanoniy-im*(Adj)). In contrast, *i* is systematically deleted in case of nouns (*levanon-im*/**levanoniy-im*(N)). The picture that emerges is that final *i* tends to be deleted in the pluralization of most native nouns. In addition, the suffix *-im* consist of *i*, and therefore it can be reconstructed more easily when deleted. This results in deleting final *i* in loanwords as well, although not systematically as in the case of *a*.

Conclusions. The study proposes a hierarchy that provides predications with respect to final vowel deletion in Hebrew loanwords. The likelihood of a vowel to be deleted is based on the accessibility of deletion in existing pluralization paradigms within native words, in addition to general constraints. Various studies have shown that there is access to entire paradigms during the course of inflection and the application of morpho-phonological processes (Steriade 2000; McCarthy 2005). This study provides further support for the claim that the mechanism of word formation takes into account not only the word itself but also its relations other words in a paradigm (van Marle 1985; Spencer 1988; Corbin 1989; Stump, 2016; Anderson 1992, Bochner 1993; Booij 2008; Blevins 2006; and references therein). The study also sheds light on morphological adaptation of loanwords and provides predication with respect to their degree of integration.

**Competing patterns of loanwords adaptation: Morphological and periphrastic formation in
Palestinian Arabic
Lior Laks & Reem Abu Shah
Bar-Ilan University**

This study examines variation in possession and dual formation of loanwords in Palestinian Arabic (PA). Such variation is manifested in a competition between morphological and periphrastic formation (also known as synthetic vs. analytic), as demonstrated in (1)-(2).

(1) tiftaḥi **il-watsab tabaṣoh** ... bitku:ni mraqbe **watsab-oh**

'open his WhatsApp... and you follow his WhatsApp' (<https://www.facebook.com/admitNajah/posts/3056471841080120>)

(2) a. aṣmelak **layk-e:n**

'I will give you two likes' (<https://ask.fm/Elmado/answers/61212134252>)

b. ma fi: ʔilla **itnen layk:at**

'there are only two likes' (<https://www.youtube.com/watch?v=FQND7X6-JKw>)

Two possession forms of 'WhatsApp' are used in (1): one is periphrastic, using the genitive exponent *tabaṣoh*, and the other is morphological, using the suffix *-oh*. (1). Similarly, the dual form of 'like (Facebook)' can be formed based on either suffixation of *-e:n* (2a) or using *itnen* 'two' before the plural form *layk-a:t* (2b). While 'WhatsApp' and 'like' demonstrate variation can take both forms of possession/dual, some loanwords tend to take one pattern. We show that the selection of either strategy can be partially predicated by morpho-phonological constraints that are responsible for the degree of adaptation of loanwords.

Modern Standard Arabic relies almost exclusively on morphological formation of possession and dual forms, while many colloquial dialects demonstrate variation, e.g. *kta:b-ak* vs. *il-kta:b btaṣtak/taṣak/ta:ṣak/še:tak* 'your (masc. sg.) book' and *kta:b-e:n* vs. *itnen kutub* 'two books' (Blanc 1970, Owens 2002, Sayahi 2015, Poplack et al. 2015, among others). Native nouns tend to take morphological patterns, while loanwords demonstrate variation. The selection between either pattern is also governed by semantic and stylistic factors, e.g. inalienable/alienable possession (Haiman 1983, Haspelmath 2008, and references therein). Such factors are not accounted for in this study, which examines morpho-phonological criteria only. The study is based on initiated web-searches of loanwords. Data is taken from public Facebook and Twitter pages, as well as other websites that demonstrate the usage of colloquial Arabic rather than Modern Standard Arabic. Loanwords were searched with both morphological and periphrastic patterns, and we made sure that the writers were PA speakers. We show that the more 'foreign' loanwords look like, the less likely they are to be integrated into the morphological system and speakers opt for the periphrastic patterns. We discuss four constraints.

Number of syllables. Words that exceed two syllables tend to take periphrastic patterns, e.g. *il-instagram btaṣtak* / **instagram-ak* 'your Instagram', *itnen katalog-a:t* / **katalog-e:n* 'two catalogs'. PA native nouns are typically monosyllabic/disyllabic, unless they consist of derivational affixes (*busta:n-ji* 'gardener', which consists of the agentive suffix *-ji*). Monosyllabic/disyllabic loanwords can be partially perceived as native, and therefore are integrated more easily and take morphological patterns (*imeyl-ak* 'your email', *imeyl-e:n* 'two emails'). Words with more syllables are treated as foreign and the more syllables there are, the more likely they are to take the periphrastic pattern.

Compounds. The tendency to opt for periphrastic patterns is even more dominant in loan compounds. *bleystor* 'Playstore' (3) takes a periphrastic dual form and *ʔiks-boks* 'X-box' (4) takes a periphrastic dual form.

(3) **il-bleystor tabaṣak** / ***bleystor-ak** miš šayya:l

'your Playstore is not working' (<https://www.facebook.com/groups/129512240834627/permalink/368570400262142/>)

(4) baštri **itnen ʔiks-boks** / ***ʔiks-boks-e:n**

'I buy two X-boxes' (<https://twitter.com/moathalkhateeb2/status/684776446745772032>)

Lexical compound formation is relatively unproductive in PA. The combination of two nouns is common in different types of syntactic construct state, e.g. *mudi:r maktab* 'office manager', but it is far less common in lexical constructions. As a result, loan compounds are easily perceived as more non-native elements in comparison to other loan words, and are not integrated into the morphological system.

Vowel-ending words. Words ending with vowels other than *a*, e.g. *selfi* 'selfie' and *biano* 'piano' also tend to take the periphrastic patterns. This results from both faithfulness and markedness constraints. Such words are relatively rare in PA, and attaching possession/dual suffixes either requires phonological alternations that make the output less faithful to the stem, or result in a marked structure. For example,

attaching the *-ak* possession suffix *biano* 'would render either vowel deletion or *h* insertion (**biano-k/bian-ak/*biano-hak*) to avoid the marked *oa* sequence (**biano-ak*). Either way, this renders morphological complexity and speakers tend to select periphrastic patterns (*il-biano btaʃtak*). Similar results in Owen (2002) for Nigerian Arabic and in Sayahi (2015) for Tunisian Arabic.

Non-native affixes. Loanwords with non-native suffixes like English *-er* (*uzer* 'user') and *-s* take periphrastic patterns. Examine the behavior of the loan noun *koment* 'comment'. When it surfaces in its singular form, it can take both morphological (5a) periphrastic possession (5b). However, when it surfaces in its English plural form *koments*, it only takes periphrastic possession (5c). The morphological mechanism is also sensitive to the morphological structure of loan words and when it identifies non-native morphological elements, it tends not to integrate words with such elements.

(5) a. miš ša:yfe **koment-i**

'I don't see my comment' (https://www.facebook.com/groups/508360659336206/?post_id=790469157792020)

b. iħtarmi **il-koment tabaʃi**

'respect my comment' (<https://www.facebook.com/groups/1671359556481629/permalink/1705541389730112/>)

c. illi bišu:f **il-koments taʃu:nak** bifakkirna ša:t'ri:n

'anyone who sees your comments will think that we are smart'

(<https://www.facebook.com/groups/1671359556481629/permalink/1696543433963241/>)

It is important to note that the above constraints represent tendencies. All words can in principle take both forms of possession/dual, and there are counter examples to these constraints. The issue here is the differences between types of loanwords and the type of patterns they are more likely to take.

When selecting possession/dual patterns, the morphological mechanism examines morpho-phonological properties of the base and selects either strategy with some degree of variation. The study adds to previous studies that examine the competition between morphological and periphrastic formation in general (see Kiparsky 2005, Brown et al. 2012, Corbett 2013, Aronoff & Lindsay 2014, Bonami & Samvelian 2015, Ackerman et al. 2011, Haspelmath & Michaelis 2017, Masini 2019, among others), and offers criteria that play a role in selecting either pattern.

References

- Ackerman, F., G.T. Stump & G. Webelhuth. 2011. Lexicalism, periphrasis and implicative morphology'. In R. D. Borsley & K. B. ʔorjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar* (325–358). Wiley-Blackwell.
- Aronoff, M. & M. Lindsay. 2014. Partial organization in languages: la langue est un système où la plupart se tient'. In S. Augendre, G. Couasnon-Torlois, D. Lebon, C. Michard, G. Boyé & F. Montermini (eds.), *Proceedings of Décembrettes 8. Rapport Interne No. 22 de CLLE-ERSS*. 1-14.
- Bonami, O. & P. Samvelian. 2015. The diversity of Inflectional Periphrasis in Persian'. *Journal of Linguistics* 51, 327-382.
- Blanc, H. 1970. Dual and pseudo-dual in the Arabic dialects. *Language*, 42-57.
- Brown, D., M. Chumakina, G.G. Corbett, G. Popova & A. Spencer. 2012. Defining 'periphrasis': key notions. *Morphology*, 22, 233–275.
- Corbett, G.G. 2013. Periphrasis and possible lexemes. In M. Chumakina & G. G. Corbett (eds.), *Periphrasis: The role of syntax and morphology in paradigms*. Oxford: Oxford University Press. 169–189
- Haiman, J. 1983. Iconic and economic motivation. *Language* 59, 781-819
- Haspelmath, M. & S.M. Michaelis. 2017. Analytic and synthetic: Typological change in varieties of European languages. In I. Buchstaller, & B. Siebenhaar (Eds.), *Language variation – European perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8)*, Leipzig, May 2015, 3-22. Amsterdam: John Benjamins.
- Kiparsky, P. 2005. Blocking and periphrasis in inflectional paradigms. In G. Booij & J. v. Marle (eds.), *Yearbook of Morphology 2004*. Dordrecht: Springer. 113–135
- Masini, F. 2019. Competition Between Morphological Words and Multiword Expressions. In F. Rainer et al. (eds.), *Competition in Inflection and Word-Formation, Studies in Morphology* 5, 281-305.
- Owens, J. 2002. Processing the world piece by piece: iconicity, lexical insertion, and possessives in Nigerian Arabic codeswitching', *Language Variation and Change* 14, 173-210
- Poplack, S., L. Sayahi, N. Mourad & N. Dion. 2015. An Exception to the Rule? Lone French Nouns in Tunisian Arabic. *University of Pennsylvania Working Papers in Linguistics* 21:2, Article 20.
- Sayahi, L. 2015. Expression of attributive possession in Tunisian Arabic: The role of language contact. In A. M. Buuts (ed.), *Semitic Languages in Contact* (Brill: Leiden/Boston). 333-347.

Gender agreement processing: evidence from Russian

Natalia Slioussar (Higher School of Economics, Moscow; Saint Petersburg State University, Saint Petersburg), Varvara Magomedova (independent researcher)
slioussar@gmail.com, varya.magomedova@gmail.com

In Russian, nouns come in three genders (M, F, N), and adjectives, participles and verbs (in some forms) exhibit gender agreement. Depending on their set of inflections, nouns are divided into several declensions. We will rely on the most widely accepted system in Table 1 (e.g. Aronoff, 1994; Halle, 1994; Shvedova, ed., 1980), but our conclusions do not depend on it. As Table 1 shows, Nom.Sg inflections differ depending on how typical they are for a particular gender. We present three self-paced reading experiments studying how this typicality (also known as gender-to-ending consistency) and the gender of the noun influence gender agreement processing.

Exp. 1 (160 participants). Materials included 36 target sentence sets and 80 filler sentences. Sentences in one set contained the same six words except for the first one, the subject noun (M nouns with typical endings, F nouns with typical and non-typical endings, balanced in frequency and length) and the second one, the verb form (M, F or N). An example is given in (1). All subject nouns were inanimate. This yielded 9 conditions, 6 of them with a gender agreement error.

Average RTs are shown on Fig.1. In all three experiments, we modeled the data with mixed-effect logistic regressions. Random intercepts and random slopes by participant and by item were included in the model. In the abstract, only significant results are reported ($p < 0.05$).

Firstly, all grammatically correct sentences are processed equally fast. This means that no gender or declension is intrinsically more difficult to process (at least, in a sentential context). Secondly, there are no differences between sentences with the same subjects and different ungrammatical predicates (e.g. an F subject combined with an M or N predicate).

The ending typicality plays a role at a very early stage, and its effect is very short-lived, while the role of gender becomes visible later and its effect is more pronounced. Namely, agreement errors were noticed significantly later with non-typical F nouns than with typical F or M ones. But error-related delay on subsequent words was significantly more pronounced for M subjects than for F ones (typical or non-typical).

The latter result suggests that predictions we make about predicate gender are stronger for M subjects (see also Slioussar & Malko, 2016). The cost of violating these expectations is higher than for F subjects (and we will show in Exp.3 that N subjects can be grouped with F subjects in this respect). The former result can be explained by assuming that the gender of the nouns with non-typical endings is more difficult to retrieve. Notably, since we used predicates of all three genders, we can be sure that this is not due to the fact that 3rd declension F nouns ‘look like’ prototypical M nouns, as it was suggested by Slioussar (2018). In Exp.2, this generalization is extended and confirmed for M nouns with non-typical endings.

Exp. 2 (91 participants). The design and analysis were very similar, but instead of F nouns with non-prototypical endings, we used M nouns with non-prototypical endings. All such nouns are animate, so we also used animate nouns in all other groups. Moreover, in this case, the non-prototypical ending is not a zero inflection in Nom.Sg, which might be important. In this experiment, we used only M and F predicates, because Exp. 1 found no differences between ungrammatical predicates. Thus, we had 36 target sentence sets and 6 conditions.

Although the gender feature is not semantically empty on animate nouns, the results for non-typical M nouns were very similar to those for non-typical inanimate F nouns in Exp. 1 (see Fig. 2). Readers notice agreement errors with them significantly later. Interestingly, on subsequent words, error-related delay is significantly larger in the sentences with typical M subjects than with typical F and non-typical M ones. I.e. only in the former case, the expectation about the gender on the predicate is stronger.

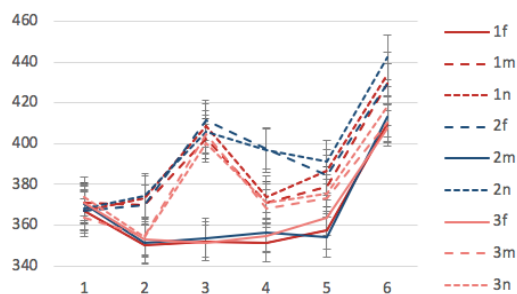
Exp. 3 (91 participants). The design and analysis were very similar, but we introduced N nouns. All nouns had prototypical endings and were inanimate in this experiment, we wanted to assess the role of gender. As Fig. 3 illustrates, we found a difference between M vs. F and N subjects. The former generate much stronger predictions about the gender of the predicate. We had 36 target sentence sets and 6 conditions.

Discussion. Many experimental studies found differences between nouns with more and less typical inflections in a variety of languages. However, these studies usually looked at the processing of isolated nouns. Among the few sentence-processing studies, Caffarra et al. (2015) looked at Italian nouns with more and less typical endings presented in the same sentences in an EEG study. Franck et al. (2008) and Vigliocco and Zilli (1999) demonstrated for Italian, Spanish, and French that heads with regular inflections are more resistant to gender agreement attraction. However, the nature of this advantage is unclear, and our experiments may serve to elucidate it. Apparently, a non-typical ending is less efficient in activating the right feature value if the agreement controller needs to be retrieved and rechecked.

Another major problem discussed in agreement processing literature is associated with asymmetric effects of different features. Russian gender is interesting in this respect because markedness relations in the system are not entirely obvious (see e.g. Kramer 2015; Rice 2006; Slioussar and Malko 2016). Several experimental studies of agreement in comprehension demonstrated that M (the most frequent gender) is different from F and N (e.g. Akhutina et al., 1999, 2001; Romanova & Gor 2017; Slioussar and Malko 2016), although in production, the pattern may be different. Our study confirms this generalization and extends it.

(1) *Xalat / kurtka / šinel' + byl potrepan / byla potrepana / bylo potrepano + ot mnogoletnej noski.*
 robe_{2D-M} / jacket_{1D-F} / overcoat_{3D-F} + was_{M/F/N} shabby_{M/F/N} + from years-long wear

Fig. 1. Average word-by-word RTs (in ms) in Exp. 1.



The study was partially supported by the Russian Ministry of Science and Higher Education (project 075-15-2020-793).

Fig. 2. Average word-by-word RTs (in ms) in Exp. 2.

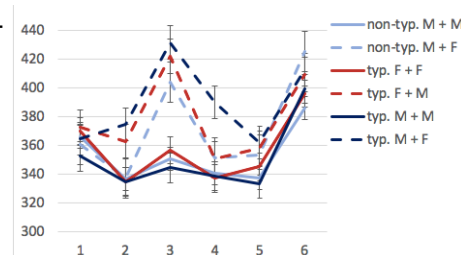
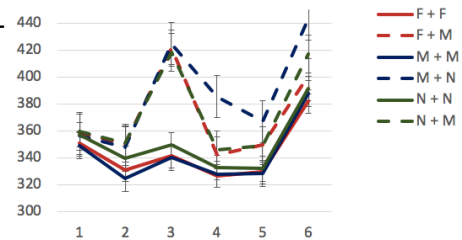


Fig. 3. Average word-by-word RTs (in ms) in Exp. 3



Declension and gender	% in the RNC	Ending in Nom.Sg	Examples
1 st decl. F	29% nouns	end in <i>-a/ja</i>	<i>žena</i> ‘wife’
1 st decl. M	1% nouns	end in <i>-a/ja</i>	<i>djadja</i> ‘uncle’
2 nd decl. M	46% nouns	end in a consonant	<i>syn</i> ‘son’, <i>gel</i> ‘gel’
2 nd decl. N	18% nouns	end in <i>-o/e</i>	<i>pole</i> ‘field’
3 rd decl. F	5% nouns	end in a consonant	<i>mel</i> ‘shallow’
irregular & indeclinable	1% nouns		

Table 1. Gender and declension in the Russian National Corpus (Slioussar & SamoiloVA 2015).

On uniqueness as a criterion of morphomic status
Martin Maiden
Research Centre for Romance Linguistics, University of Oxford

Maiden (2018:2,22, and elsewhere) has claimed that ‘morphomes’ never occur twice in the same way. Distributional patterns in inflexional paradigms which, by virtue of their synchronic irreducibility to any kind of extramorphological conditioning, can be classified as ‘morphomic’ in the tradition established by Aronoff (1994), typically evolve as the consequence of series of phonological, morphological, or semantic changes whose specific nature and historical ordering is unlikely ever to be exactly replicated in the history of more than one language. This is why Maiden develops a ‘uniqueness criterion’ as a typological diagnostic of morphomhood: if some putative morphomic pattern in one language then turns up independently in some unrelated language it is unlikely to be morphomic after all, and probably has some other, more universal, motivation which we have simply failed to grasp.

Recently, Herce (2019:130) has criticized the uniqueness criterion as, no less, ‘an unnecessary footnote to the definition of the morphome that discourages typological approaches to the phenomenon’, saying that ‘there is no reason to give up on typological and comparative research in general because of this’. I shall first argue that this critique is overstated and that the uniqueness criterion remains an important negative diagnostic of morphomhood. A good example of the value of this criterion involves the Romance morphomic pattern which Maiden named the ‘N-pattern’. There is a widespread misapprehension that this pattern simply involves an opposition between first and second person plural forms, on the one hand, and the remaining person and number combinations on the other. If this were so, critics of morphomic concepts (e.g., Anderson 2010; Nielsen 2019) would be justified in pointing out, as they do, that similar patterns occur across a number of unrelated languages. The guarantee that the ‘N-pattern’ really is morphomic, and specific to Romance languages, lies precisely in the too frequently overlooked fact, as I shall show, that it also makes crucial reference to dimensions of ‘tense’ and ‘mood’ in ways never replicated elsewhere. However, I shall also claim that independent emergence of identical morphomic patterns maybe more widespread than Maiden has recognized, and I shall consider at least two ways in which this may come about.

I first critically consider a kind of development identified by Herce (Herce 2020) involving cross-linguistically common grammaticalization pathways. For example, it is sometimes found (e.g., in the wholly unrelated languages Noon and Kven) that the same morphology is used for the expression of the passive voice and third-person plural subject. I shall argue, however, that the difficulty in such cases lies in guaranteeing that the original motivation for such a pattern—presumably the use of third-person plural subjects as a way of demoting the specific identity of the agent—is synchronically extinct.

I then discuss a finding which is unexpected to the point of being downright counterintuitive: that morphomic patterns can be ‘borrowed’ from one language into another. Morphomic patterns are the diametrical opposite of what is involved in the most familiar type of linguistic borrowing, that involving ‘loanwords’ such that a form denoting a clearly defined, and often culturally novel, referent is taken from one language into another. Quite unlike loanwords, morphomic patterns exist independently of the phonological forms which manifest them, and can be assigned no referential or functional meaning. That should make them the most improbable candidates for any kind of ‘borrowing’, and the general impression one gets from the literature seems to be that they are not borrowable (cf. Kossman 2015; Gardani 2018). However, there has emerged in the Ladin dialect spoken in the Val Badia (Italy) a morphomic inflexional pattern such that the same inflexional ending, namely -s, is found in the second person singular present indicative, and in the first, second, and third person singular and the

third person plural of the present subjunctive. This pattern has absolutely no plausible or detectable explanation in the internal history of Badiot morphology. Indeed, I shall give evidence from numerous detailed descriptions of the relevant dialects spanning the nineteenth and twentieth centuries that there was absolutely no sign of this development before 1900 and that it appears almost ‘overnight’ in the twentieth century, being abundantly attested from multiple sources by the 1960s. I shall demonstrate that in the linguistic universe of twentieth century Badiot Ladin there is only one factor that could explain such a development, namely contact with Italian. The incorporation of the South Tyrol into the Italian state at the end of the First World War entailed a massive increase in the importance of standard Italian for Ladin speakers, and it is in Italian that one finds exactly the same inflexional pattern, albeit only in first conjugation verbs (there are internal reasons which conjugational differences would be irrelevant in Badiot, however), and in a different phonological form, as the ending -i. This -i, characteristic of the second person singular present indicative, is also the ending of all the singular forms and the third person plural forms of the present subjunctive. It seems that this morphomic pattern in Italian (itself the result of a complex series of phonological and morphological adjustments in the medieval language), has effectively been ‘calqued’ by Badiot, which has used native morphological matter to replicate a pattern found in an alien (albeit closely related and structurally similar) language.

In conclusion, I will reassess the uniqueness criterion in the light of counterexamples such as shared grammaticalization paths and borrowing. I shall concede that morphemes can indeed occur in the same way in different languages, but only under clearly defined conditions, and I shall continue to maintain that the uniqueness criterion is a valuable typological diagnostic of morphomhood, and one that in no way impedes the broader typological study of morphomic patterns.

References

- Andersen, H. (2010). ‘From morphologization to demorphologization’. In Luraghi, S. and Bubeník, V. (eds) *The Bloomsbury Companion to Historical Linguistics*. London: Continuum, 117-46.
- Aronoff, M. (1994). *Morphology By Itself*. Cambridge, MA: MIT Press.
- Gardani, F. (2018). ‘On morphological borrowing’. *Language and Linguistics Compass* 12. <https://doi.org/10.1111/lnc3.12302>
- Kossmann, M. (2015). ‘Contact-induced change’. In Baerman, Matthew (ed.) *The Oxford Handbook of Inflection*. Oxford: OUP, 251-71
- Herce, B. (2019). ‘Morpheme interactions’. *Morphology* 29:109-32
- Herce, B. (2020). *A typological approach to the morpheme*. PhD Thesis, Univ. Surrey, Surrey, UK/Univ. Basque Country, Leioa, Spain
- Kossmann, M. (2015). ‘Contact-induced change’. In Baerman, Matthew (ed.) *The Oxford Handbook of Inflection*. Oxford: OUP, 251-71
- Maiden, M. (2018). *The Romance Verb. Morphomic Structure in Diachrony*. Oxford: Oxford University Press
- Nielsen, P. J. (2019). ‘Diachronic morphology, indexical function and a critique of the morpheme analysis. The content and expression of Danish *forstå*.’ In Heltoft, L., Igartua, I., Joseph, B., Jeppesen Kragh, K., & Schøsler, L. (eds), *Perspectives on Language Structure and Language Change*. Amsterdam / Philadelphia: Benjamins, 125-50.

When inalienable nouns are inalienable in Spanish? Microvariation in morphological constructions

Bárbara Marqueta
(marqueta@unizar.es)
University of Zaragoza (Spain)

It is a well-known fact about the syntax of languages that some classes of nouns, such as *body parts*, *constitutive parts of objects*, *kinship roles*, or *clothes being worn* (Heine 1997) participate in constructions that codify *inalienable* possession relationships (*grosso modo*, those in which the possessee is conceived of as being inseparable from a possessor, as opposed to unmarked or *alienable* possession structures). The literature on the topic has paid considerable attention to the syntactic features of these structures as well as to the cross-linguistic differences among them (Alexiadou 2003).

Within the domain of morphology, several word formation patterns codifying inalienable possession relationships have been identified –e.g. *brown-eyed* for English (Nevins & Myler 2014); *ostrokanciasty* 'sharp-edged' for Polish (Bisetto & Melloni 2008); *blauäugig* 'blue-eyed' for German (Serrano-Dolader 1996), etc.–. In Spanish, there is a considerable body of literature on the [N-i-Adj] compound class exemplified by *pelirrojo* [lit. *hair+I+red*] 'red-haired', a quite productive inalienable possession compound pattern.

It has been observed that languages differ in which noun classes (body parts, kinship terms, etc.) participate within the inalienable possession structures (Myler 2014). Regarding *pelirrojo* compounds, it has been observed that this type of compound mostly takes as inalienable possessives visible body-parts of animals and objects, with only a few exceptions (Moyna 2011: 88).

Our search into the *Corpus del Español NOW* (<https://bit.ly/3DcgE6W>), however, has revealed that there is a growing number of compounds where the noun term refers to a feature (*mangas* 'sleeves'; *franjas* 'stripes') of sport clothes (1).

1. *manguiverde* [sleeve+I+green] 'green-sleeved' (referring to Oberema football team)
franjablanco [stripe+I+white] 'white-striped' (referring to Manzanares football team)

Moreover, Pinto et al. (2011) shows that *pelirrojo* counterparts in Sardinian dialects take as inalienable possessives internal body parts or parts of objects (*koridoustau* lit. 'heart+I+strong' 'with a strong heart', or *maniyilongu* lit. 'handle+I+long' 'with a long handle'), but examples as such are not attested in Spanish.

The fact that the noun classes treated as inalienable can vary even within the same compound pattern has led us to argue against the customary claim that inalienability is a property *lexically* specified for certain classes of nouns (Vergnaud & Zubizarreta 1992). We hypothesize, instead, that inalienability is a property of specific grammatical constructions, including morphological structures.

Notice that, within the same language, noun and adjectives can be combined to form a compound word with roughly the same properties as *pelirrojo* compounds (right-headed compound adjectives with a noun modifier), only differing because they do not codify inalienable possession (*sinohablante* 'Chinese-speaking', *Pottermaniac* 'Pottermaniac').

In our analysis, we address three different morphological constructions through which Spanish codify inalienable possession, ranging from those with take a more restricted class of nouns as inalienable to those that take a wider range of them (2):

2. a. Only take body parts: *patanegra* (paw+black) 'of good quality';
bocachancla (mouth+slipper) 'bigmouth'
- b. Take body parts and human feelings/abilities: *peludo* [hair+udo] 'hairy';
corajudo [courage+udo] 'very brave', *forzudo* [strength+udo] 'muscular'
- c. Mostly take body parts, but also parts of plants and objects: *cornicorto*
[horn+I+short] 'short-horned' *hojiblanca* [leaf+I+white] 'white-leafed', *puntiagudo*
[point+I+sharp] 'pointed'

The structure of the examples above differs significantly. Those in (2a) are semantically exocentric noun phrases where the noun head takes a modifier (a noun or an adjective with which optionally agree). Examples in (2b) are denominal adjectives in which the noun base is quantified, contrary to *pelirrojo* compounds (2c), where the attributed property is not quantified (in *cornicorto* horns are not necessarily too short).

Following a Nanosyntactic framework (Starke 2009), which, in line with other Neoconstructionist approaches, assumes that complex words, just like phrases, are constructed in the syntax, we develop an analysis for *pelirrojo* structure in which we posit a relational functional category that takes nouns as its complements and turns them into inalienable possessors of an entity external to the compound (Marqueta 2019). The noun possessors are spelled-out by *I*-noun allomorphs in Spanish such as *pele* (via phrasal spell-out). These allomorphs never participate in non-inalienable constructions (*sinohablante*, *Pottermaniacobave*), contrary to other bound forms of Greek or Latin origin.

In conclusion, we suggest that both phrasal and word syntax possess different means of codifying inalienable possession through functional categories with different spell-out and combinatorial properties, which are the ones in charge of delimiting the conceptual classes of nouns that appear in the construction (idiosyncratically or as a matter of convention). The key point is that the nouns will be inalienable only as a consequence of being taken as complements of a Possession head.

References

- Alexiadou, Artemis. 2003. Some notes on the structure of alienable and inalienable possessors. In Martine Coene & Yves D'hulst (eds.), *The expression of possession in Noun Phrases*, 167-188. Amsterdam: John Benjamins.
- Bisetto, Antonietta, Melloni, Chiara. 2008. Parasynthetic Compounding. *Lingue e Linguaggio* 2. 233-259.
- Heine, Bernd. 1997. *Possession: Cognitive Sources, Forces, and Grammaticization*. Cambridge: Cambridge University Press.
- Pinto, Inmacolata, Paulis, Giulio, Putzu, Ignazio. 2011. Sardinian adjectives with the NiA structure. *Lingue e Linguaggio* 11(1). 49-70.
- Marqueta, Bárbara. 2019. The syntactic structure of *pelirrojo* compounds. *Glossa: a journal of general linguistics* 4(1): 40. 1-23.
- Moyna, Irene. 2011. *Compound words in Spanish: Theory and history*. Amsterdam: John Benjamins.
- Myler, Neil. 2016. *Building and interpreting possession sentences*. Cambridge: MIT Press.
- Nevins, Andrew & Neil Myler. 2014. *A brown-eyed girl (UCLA Working Papers in Linguistics: Papers in Honor of Sara van Wageningen)*. Los Angeles, CA. Consulted at: ling.auf.net/lingbuzz/002021.
- Serrano, David. 1996. Del 'Rothaarig' alemán al 'pelirrojo' español: un caso de diversidad morfológica. *Anuario de Estudios Filológicos* 19. 475-484.
- Vergnaud, Jean Roger & María Luisa Zubizarreta. 1992. The definite determiner and the inalienable constructions in French and English. *Linguistic Inquiry* 23. 595-652.

An information-theoretic analysis of the inflectional regular-irregular gradient for optimal processing units

Claudia Marzi¹ and Vito Pirrelli¹

¹Institute for Computational Linguistics, National Research Council, Pisa (Italy) name.surname@ilc.cnr.it

Prediction-driven word processing defines the human ability to anticipate upcoming input words in recognition [5]. From this perspective, input word forms need to be processed as quickly and efficiently as possible. Under the reasonable assumption that spoken words are memorized and processed as word trees (e.g. Marslen-Wilson’s “cohorts”), the larger the size of the cohort of an input word at a certain point in time (and the later its uniqueness point), the harder and slower to process the word is [1]. Regularly and irregularly inflected verb forms have different stem family sizes and different uniqueness points [2]. Using a Recurrent Neural Network (RNN) as a computational model of the human lexical processor, we explore here how their distributional and structural properties may affect (optimal) processing strategies. In a RNN, input words activate chains of processing nodes that are time-aligned by temporal connections. The strength of temporal connections is a function of the frequency with which n -grams of symbols appear in the input [2, 6]. A high-frequency n -gram recruits a specialized temporal chain of processing nodes activated only by that n -gram (entrenchment). Conversely, a low-frequency n -gram tends to activate some of the processing nodes activated by other, partially overlapping n -grams (blending). The processing dynamic of a RNN can be measured by the *pointwise entropy* (pH) of $p(C_{ik} | n_k)$, i.e. the probability for a temporal connection C_{ik} to leave node n_k for node n_i . More entrenched connections score low on pH, reducing processing uncertainty. Blended connections increase pH. Ultimately, levels of connectivity in a RNN reflect levels of entropy in the distribution of input items and affect the way they are processed. We varied RNN input conditions along three dimensions: (i) language; (ii) type frequency of input data; (iii) token frequency of input data. The way an input word is processed is measured by the rate at which the network predicts an upcoming word symbol based on the already processed context, or *prediction rate* (PR). Figure 1a illustrates how well a RNN serially predicts French, Italian and Spanish verb forms administered to the network with a uniform distribution. Forms are grouped by the size of their stem family (0, small, medium and large), a graded measure of inflectional regularity [2]). Note that the nonlinear drop in PR at the stem-suffix boundary (MB = 0) in regulars (large stem family size) diminishes with irregularly inflected forms (smaller stem family size). Variations of PR between the two classes closely mirror the way pH decreases across regulars (R) vs. irregulars (I) (Figure 1b). In fact, regularly inflected forms consist of recurrent patterns that are systematically repeated either within a paradigm (e.g. stems), or across paradigms (affixes). The systematicity of these patterns makes their processing cost low (as shown by the high levels of predictability in Figure 1a). However, their morphological structure causes a word-internal drop in PR at the stem-suffix boundary, where the stem is followed by many different suffixes. As to irregulars, their stem allomorphs are harder to discriminate as they compete for the same processing units. Nonetheless, stems in irregularly inflected forms are more selective about possibly ensuing inflectional endings, and the drop in the network’s PR gets smaller with smaller stem families.

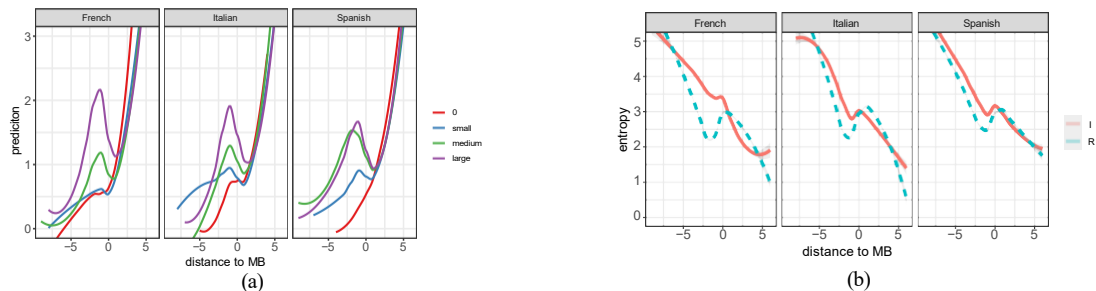


Figure 1: Non-linear regression plots of (a) symbol prediction rates and (b) pointwise entropy fitted by the interaction effects of stem family-size and the distance to stem-suffix boundary (MB=0), for French, Italian and Spanish.

Figure 2a shows how PR varies on average along a word being processed, with pH values being grouped in low, medium and high bins. For low entropic values, prediction scores are high at word onsets, and get

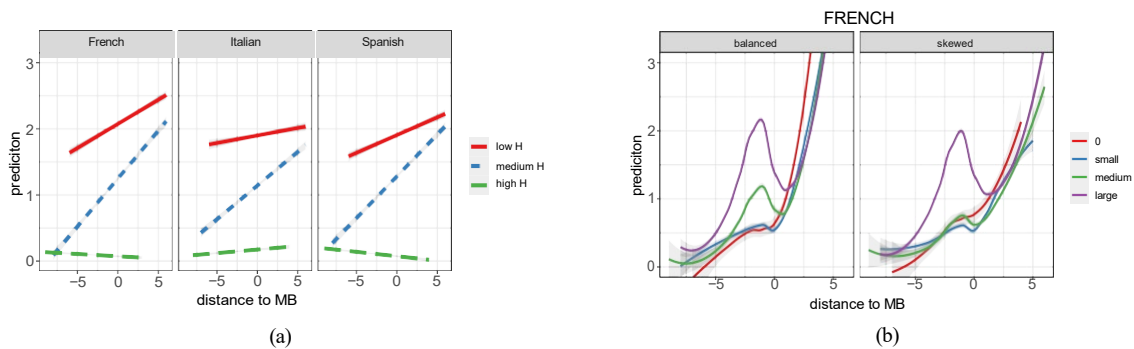


Figure 2: Linear regression plots (a) of symbol prediction rates fitted by the interaction effects of ranges of entropy levels (*low, medium, high H*) and the distance to stem-suffix boundary (MB=0), for French, Italian and Spanish. Non-linear regression plots (b) of symbol prediction rates fitted by the interaction effects of stem family-size (*0, small, medium, large*) and the distance to stem-suffix boundary (MB=0) for French.

higher as more of the word is processed. This corresponds to a situation where we have a small lexicon of highly selected non redundant lexical units. For high entropic values, prediction is nearly impossible: word units are highly combinatorial and the resulting lexicon turns out to be chaotic. Medium entropic values describe the typical situation of real inflection systems, where irregular items mix up with regular ones. Figure 2b shows the impact of corpus-based distributions on PR for French verb forms. A skewed distribution makes high-frequency forms easier to process, and low-frequency forms harder. Forms in larger stem families are penalized (lower PR peaks at MB = 0), while isolated forms tend to be processed more efficiently (higher PR peaks). Frequency has a levelling across the board effect on PR, narrowing down the processing advantage of regulars over irregulars of Figure 1a. In the end, irregulars are processed nearly as effectively as more regular items, due to the combined influence of high frequency [7], short length [4] and lack of morphological structure.

To sum up, optimally functional words must comply with a cluster of information-theoretic properties that affect the ways words are memorized, processed and learned. Accordingly, regulars and irregulars are not categorially distinct. They appear to vary along a gradient of distributional and structural properties that modulate processing uncertainty in different ways. In the end, this is why speakers perceive them differently, and linguistic analyses place considerable emphasis on their difference. Entropy sheds light on their respective role in an inflection system. In fact, although regulars and irregulars score differently along a cluster of quantitative dimensions, their overall entropy must be neither too high nor too low. On the one hand, a too highly entropic communication code is chaotic: this is the case of a system with too many, evenly distributed exceptions, or with too many unpredictable conjugation classes. On the other hand, a too low entropic communication code is closed and unproductive. Optimally, the lexicon of any language should contain a small pool of high-frequency irregulars and a long tail of low-frequency regulars. In the end, this may boil down to a functional explanation of Zipf's law, linking the latter to well-known, discriminative properties of human memory, which have independently been characterized as being well-adapted to how stimuli are distributed in the real world [3].

References

- [1] BALLING, L. W., AND BAAYEN, R. H. Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition* 125, 1 (2012), 80–106.
- [2] MARZI, C., FERRO, M., AND PIRRELLI, V. A processing-oriented investigation of inflectional complexity. *Frontiers in Communication* 4 (2019), 48.
- [3] PIANTADOSI, S. T. Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review* 21, 5 (2014), 1112–1130.
- [4] PIANTADOSI, S. T., TILY, H., AND GIBSON, E. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108, 9 (2011), 3526–3529.
- [5] PICKERING, M. J., AND CLARK, A. Getting ahead: forward models and their place in cognitive architecture. *Trends in cognitive sciences* 18, 9 (2014), 451–456.
- [6] PIRRELLI, V., MARZI, C., FERRO, M., CARDILLO, F. A., BAAYEN, H. R., AND MILIN, P. *Psycho-computational modelling of the mental lexicon. A discriminative learning perspective*. Mouton De Gruyter, 2020, pp. 23–82.
- [7] WU, S., COTTERELL, R., AND O'DONNELL, T. J. Morphological irregularity correlates with frequency. *arXiv preprint arXiv:1906.11483* (2019).

Cognitive linguistic positions in morphological processing and acquisition

Veronika Mattes¹, Sabine Sommer-Lolei², Wolfgang U. Dressler²

¹University of Graz, ²University of Vienna

veronika.mattes@uni-graz.at; sabine.sommer-lolei@univie.ac.at; wolfgang.dressler@univie.ac.at

This paper addresses the question of how to explain the development of morphological structures (i.e. word formation) in terms of cognitive principles of language processing (cf. Onysko & Michel 2010; Booij 2012; Libben, Gagné & Dressler 2020). For this aim, we provide evidence from natural language acquisition data on word formation, mainly from our studies on German, and embed it in a broader typological context in that we compare the results of our studies with that of other, typologically different languages. The data come from the ecologically most valid collection of longitudinal spontaneous speech data of children in interaction with their parental caretakers (Mattes, Sommer-Lolei, Korecky-Kröll & Dressler 2021).

Cognitive processing strategies can explain general word-formation preferences that influence the structures and their developments. These preferences should by their nature hold even more for child speech than for adult speech (Dressler & Kilani-Schoch 2017). Cognitively based preferences in language processing and acquisition are based on frequency, simplicity, iconicity, salience and transparency (Dressler, Libben & Korecky-Kröll 2014).

- The frequency of a (morphological) structure in the language is known to play an essential role in language processing in general and, specifically, in language acquisition (Ambridge, Kidd, Rowland & Theakston 2015). High token frequency of word forms leads to entrenchment and influences their activation, whereas high type or lemma frequency of word-formation patterns enhance their (potential) productivity.
- Simplicity vs. complexity of word formations: Complexity can be formal and/or conceptual (Slobin 1973, 1985). With respect to morphology, formal complexity is caused by allomorphy, morphotactic opacity, combinations of word-formation patterns in one lexeme. Differences regarding formal complexity in languages lead to different rates of acquisition (Clark 2016), at least in terms of number of morphemes and morphotactic opacity. Conceptual complexity is caused by polysemy, morphosemantic opacity, and abstract meanings. Both aspects of complexity come into conflict in conversion, which does not lead to an increase of the number of derivational morphemes, but lacks iconicity.
- Iconicity in the sense of a one-to-one correspondence of formal morphemes and their functions, and iconicity of head-non-head relations, i.e. the formal head of a derivation and compound is also its semantic head, as in *teach-er*, where the suffix is responsible for the whole word being an agent and to which inflectional suffixes also attach.
- Transparency vs. opacity: Morphotactic transparency, i.e. easy formal decomposition of derivations and compounds and morphosemantic transparency, i.e. easy recoverability of the meaning of constituents of derivations and compounds from the holistic meaning of the whole complex word.

Many of the observed developments of word formation in our data of language acquisition refer to various ways of rise of complexity (as assumed since Jakobson 1941), driven by

adaptation and economic principles (Bittner 2014), and provide evidence for the important roles that the factors transparency, iconicity and frequency play, yet to different degrees.

In general, the developmental courses of word formation are rather parallel in different languages, obviously following general cognitive principles listed above. However, there are also obvious differences between courses of development of morphological acquisition between languages, especially for word-formation patterns in which cognitive preferences come into conflict with each other. In these cases, frequency diversity and morphological productivity of word-formation patterns in the individual languages are reflected in the usage of word-formation patterns in child-speech. More generally, differences in the speed of acquisition in each area of morphology can be explained by the important role of the respective morphological richness (Xanthos et al. 2011).

Most of our results are perfectly compatible with usage-based approaches to language acquisition, but not with certain nativistic approaches, whereas our results on blind-alley developments in child language development pose problems both to nativistic and usage-based approaches.

References:

- Ambridge, B., E. Kidd, C.F. Rowland & A. Theakston. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(2), 239–253.
- Bittner, D. 2014. Zum Verhältnis von Sprachwandel und Spracherwerb. In S. Ringmann & J. Siegmüller (eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Jugend- und Erwachsenenalter*. München: Elsevier, Urban & Fischer, 19–37.
- Booij, G. 2012. *The Grammar of Words. An Introduction to Linguistic Morphology*. 3rd ed. Oxford: University Press.
- Clark, E. 2016. *The Lexicon in Acquisition*. 3rd ed. Cambridge: Cambridge University Press.
- Dressler, W.U., G. Libben & K. Korecky-Kröll. 2014. Conflicting vs. convergent vs. interdependent motivations in morphology. In B. MacWhinney, A. Malchukov & E. Moravcsik (eds.), *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press, 181–196.
- Jakobson, R. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.
- Libben, G., C. Gagné & W.U. Dressler. 2020. The representation and processing of compound words. In V. Pirrelli, I. Plag & W.U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*. Berlin: de Gruyter, 336–352.
- Mattes, V., S. Sommer-Lolei, K. Korecky-Kröll & W. U. Dressler (eds.). 2021. *The Acquisition of Derivational Morphology. A cross-linguistic perspective*. Amsterdam: Benjamins.
- Onysko, A. & S. Michel. 2010. Introduction: Unravelling the cognitive in word formation. In A. Onysko & S. Michel (eds.). *Cognitive Perspectives on Word Formation*. Berlin: De Gruyter Mouton, 1–25.
- Slobin, D. 1973. Cognitive Prerequisites for the Development of Grammar. In Ch. Ferguson & D. Slobin (eds.): *Studies of child language development*. New York: Holt, Rinehart and Watson, 175–208.
- Slobin, D. (ed.). 1985. *The crosslinguistic Study of Language Acquisition*. Vol. II: Theoretical Issues. Hillsdale, NJ: Lawrence Erlbaum.
- Xanthos, A., S. Laaha, S. Gillis, U. Stephany, A. Aksu-Koç, A. Christofidou, N. Gagarina, G. Hrzica, F. N. Ketrez, M. Kilani-Schoch, K. Korecky-Kröll, M. Kovačević, K. Laalo, M. Palmović, B. Pfeiler, M. D. Voeikova & W. U. Dressler. 2011. On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31, 461–479.

What licenses event and argument structure in loan nominalisations?

A corpus-based study of Italian nouns borrowed from English

Judith Meinschaefer, Freie Universität Berlin, judith.meinschaefer@fu-berlin.de

Topic – The project described here investigates the borrowing of English deverbal nominalisations with the suffix *-ing* into Italian (in the following termed Italian *ing*-nouns). In current written Italian, there are numerous such nominalisations, which sometimes occur with complex event readings (in the following termed CEN readings) and argument structures. For some of these formations, the base verbs have also been borrowed into Italian; (1). These and all following examples are taken from the *itTenTen16* corpus (5 billion words, Jakubíček et al. 2013).

- (1) a. *diviene importante ... un costante **screening** dello stato di salute (itTenTen16)*
'a constant **screening** of the state of health becomes important'
- b. *3.279 pazienti ... **sono stati screenati** (itTenTen16)*
'3.279 patients ... **have been screened**'

There is currently little to no evidence for morphological borrowing in a strict sense of the suffix *-ing* into Italian, as the suffix does not attach to native verbal bases.

Research Question – Here, we set out to explore the question under which conditions a CEN reading is available for English nouns ending in *-ing* that have been borrowed into Italian. Under the often made – and sometimes contested – assumption (going back to Smith 1972; Grimshaw 1990) that complex event structure and argument structure in nominalisations needs to be licensed by overt affixes, borrowed nouns in general are not expected to dispose of CEN readings – unless morphological borrowing in a strict sense has occurred and the affix *-ing*, as an item of the Italian lexicon, does have the ability to license complex event structure.

Hypotheses – What is it that makes it possible for a given Italian *ing*-noun to have complex event structure and argument structure? Two hypotheses are tested against the data. **H1** An Italian *ing*-noun can have a CEN reading if and only if the Italian lexicon contains a verb sharing its root or part of its root with the Italian *ing*-noun, where the verb can be borrowed from English or pertain to the native lexicon. **H2** Whether an Italian *ing*-noun has a CEN reading or not depends on the size and structure of the derivational paradigm of which it is a part.

Data – The investigation is based on a random sample of 100 Italian *ing*-nouns attested in the *itTenTen* corpus, for which the concordances have been drawn from the corpus and searched for structures that are most likely (i) to realize one or more arguments of the predicate and (ii) to dispose of genuine CEN readings. The selection of Italian *-ing* nouns considered here is a random sample drawn from a larger list of 2.296 Italian *ing*-nouns for which a least 20 occurrences are attested in the corpus. We investigate (i) whether CEN readings (as diagnosed, e.g., by event modifying adjectives, (1a), predicates like *avere luogo* 'take place', realization of arguments and other distributional criteria) are at all attested and how frequent they are, (ii) whether a corresponding verb is found in the corpus, and (iii) the size and morphological structure of the derivational paradigm, as attested in the corpus. As to the derivational paradigm, we consider only a few selected forms, among which verbs, as in [1b], agent nouns (e.g., *twittatore*), event nouns (e.g., *skillaggio*, *shiftamento*) and passive adjectives (e.g., *joinabile*).

Results – The results indicate that roughly half of the Italian *-ing* nouns of the sample studied here are attested in argument-realizing constructions in the corpus, and a third of the forms are attested in genuine CEN contexts. For a couple of these forms, however, no morphologically related verb or any other morphologically related form is attested in the corpus. Neither H1 nor H2 can thus fully account for the data. Rather, it seems that whether an Italian *-ing* noun is attested with argument and event structure in the corpus or not depends on its semantic characteristics. On the one

hand, the data show that Italian *-ing* nouns with a rather general reading (such as *briefing*) often occur in CEN contexts, while Italian *-ing* nouns that pertain to the semantic domain of audio, video and photography (e.g., *microprinting*) often have a result reading, and those pertaining to the semantic domain of sports, art and cultural practices (e.g., *boxing*) often have generic, non-episodic readings and occur without argument and event structure. On the other hand, the meaning of the individual forms appears to determine their usage. To give an example, Italian *-ing* nouns denoting transitive predicates that imply a result (e.g., *undocking*) are often attested in CEN contexts, while this is seldom the case for intransitive activities with only an agent argument (e.g., *wheezing*). Under a more descriptive perspective, the project explores the array of readings which are available for Italian *ing*-nouns, as well as the different degrees of lexical and morphosyntactic integration of these nouns into Italian.

References – Grimshaw, Jane. 1990. *Argument structure*. Cambridge, MA: MIT Press. • Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý & V. Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, 125–127. • Smith, Carlota S. 1972. On Causative Verbs and Derived Nominals in English. *Linguistic Inquiry* 3(1). 136–138.

**an analysis of deverbal adjectivization within an antilexical framework:
the case of *-kanoo* compounding in Japanese**

Junya Morita/ Kinjo Gakuin University/ morita@kinjo-u.ac.jp

1. The interaction between syntax and morphology is a central theme of recent studies in the generative framework. Typical processes include deverbal adjectivization, which derives an adjectival from the relevant verb phrase. In a current theory of antilexicalism, Distributed Morphology (DM), the core properties of word construction are attributed to its syntactic structure while the role of its formal make-up is consigned to the morphological module, due to an economy constraint which requires information available to each stage of computation to be narrowly restricted (Marantz 1996, Embick 2010). The present study proposes a DM-theoretic analysis of deverbal adjectivization in Japanese, focusing on *-kanoo* compounding. Based on a detailed observation of *-kanoo* compounds extracted from a large-scale corpus, their semantic and formal properties are illuminated in §2 and they are accounted for in §3.

2. *-Kanoo* compounds, consisting of the bound morpheme *-kanoo* and a verbal noun, typically involve a property predication, as in (1). Our corpus research has detected 473 word types of *-kanoo* compounds in Balanced Corpus of Contemporary Written Japanese, a 105-million-word corpus. The data analysis identifies their lexical and syntactic characteristics, exposing the two-sidedness of *-kanoo* adjectivals.

- (1) korerano guzai-ga denshirenji-de choori-kanoo-da.
 these ingredients-Nom microwave-by cook-able-be
 ‘These ingredients are capable of being cooked by microwave.’

As illustrated in (1), the cardinal meaning of *-kanoo* adjectives is decomposed into three elements: ‘(a)having the nature of (b)being able to (c)be done VN.’ The semantic element of (c) shows a passive sense in that the subject of *-kanoo* sentence undergoes the action or process expressed by a base VN, a stem that has a predicate function. Element (b) implies that adjectives in *-kanoo* are used in a modalized potential sense. The *-kanoo* construction describes some property to the subject, as derivable from component (a). Morphologically, four points are worth noting. First, *-kanoo* may attach to any VN, regardless of their vocabulary strata. Second, *-kanoo* compounding involves a verb phrase; in example (1), *-kanoo* underlyingly combines with the phrase *denshirenji-de choori*. Importantly, a phrase with which *-kanoo* combines may become a compound; the “instrument” adjunct in (1) is incorporated into the VN *choori* to yield the compound *denshirenji-choori-kanoo(-na)*. Third, the evidence for the lexical status of *-kanoo* words comes from the fact that deadjectival affixes such as *-sei* ‘-ity’ are added to *-kanoo* words, as in *choori-kanoo-sei* ‘cookability.’ And finally, *-kanoo* compounding is very productive; a hapax-based productivity measure shows that when ten different word types are derived by adding *-kanoo* to VNs, three of them are newly created. A crucial syntactic point is that thematic restrictions are imposed on the external argument of *-kanoo* words, which will be analyzed in the next section.

3. Although there are a handful of descriptive studies (Namiki 1988), there has been no systematic analysis of *-kanoo* words. This is a key issue that we will address in this section. The underlying structure is given in (2), which is based on the structure of adjectival passive proposed by Bruening 2014. (Root ($\sqrt{\quad}$) is a bound morpheme that becomes the core of a word; no lexical item actually occurs in syntax.)

- (2) [_{PredP} guzai_i ... [_{aP} OP_i [_{aP} a-kanoo [_{VoiceP} Voice- ϕ [_{vP} v- ϕ [_{vP} [$\sqrt{\quad}$: $\sqrt{\text{VN}}_{\text{choori}}$ t_i][_{pP} p-de n_{denshirenji}]]]]]]]]

The properties of *-kanoo* compounding just presented are derived from this structure. That it involves a verb phrase is evident from the form, its high productivity is attributable to its syntactic computation, and the passive import stems from the functional head Voice. The null

operator, which occupies the internal argument position, is moved to the specifier position of *aP* in an adjectival environment. It is then coindexed with the matrix subject by predication. Evidence for the operator movement is provided by the thematic restrictions on external argument. Empirically, operator movement is possible from an argument *pP*, but not from an adjunct *pP*. Thus, (3a) is acceptable, where the subject is underlyingly related to the Locational argument of the base VN, whereas (3b) is unacceptable because the subject is connected with a Locational adjunct. This is a consequence of the principle of ECP: it is only when the trace left by a null operator is within an argument *pP* that it is theta-coindexed with the head VN (cf. Baker 1988: 39-40). The same argument applies to the difference of acceptability in the operator movement from two types of Instrument *pP*s.

- (3) a. Kono heya-ga shukuhaku-kanoo-da. ‘This room can be stayed in.’
 this room-Nom stay-able-be
 b. ??Tokyo-ga kenshuu-kanoo-da. ‘Tokyo can be trained in.’
 Tokyo-Nom train-able-be

A syntactic output is sent to Morphology on the PF side of the grammar after Spell-Out, where it is required to be constructed into a word form according to Vocabulary insertion and a set of morphological operations. Turning first to Vocabulary insertion, /kanoo:/ is notably inserted in an appropriate syntactic node depending on its formalized lexical entry in (4). The defining features in (i) are associated with the relevant node; the attributive nature of *-kanoo* adjectivals comes from the feature [property] and their modalized potential nature results from the feature [capability]. The license environment in (ii) denotes that *-kanoo* connects to Voice phrase whose lexical head is a dynamic transitive VN and hence stative VNs like *icchi* ‘correspond (to)’ are ruled out as the base of *-kanoo* (cf. **icchi-kanoo(-na)*).

- (4) (i) [A][property][capability], (ii) +<Voice, VN [transitive][dynamic]>

In connection with structure (2), the complex compound *denshirenji-choori-kanoo(-na)* has the underlying structure: [_{aP} a-kanoo [_{VoiceP} Voice- \varnothing [_{vP} v- \varnothing [_{√P} [_{√V'} $\sqrt{\text{VN}}_{\text{choori}}$ t_i][_{pP} p- \varnothing n_{denshirenji}]]]]]]. Formation of the compound is carried out in a purely mechanical way. First, the ‘boundness’ of the zero-morpheme \varnothing in the underlined *pP* triggers the merger of *n* and *p* to generate ‘... [_{pP} [_p n_{denshirenji} p- \varnothing]].’ Merger is a morphological process of combining linearly adjacent zero-level categories into a zero-level category (Marantz 1996: 24). Subsequent merger is enforced three times to yield the structure: [_{Voice} [_v [_{√VN} [_p n_{denshirenji} p- \varnothing] $\sqrt{\text{VN}}_{\text{choori}}$] v- \varnothing] Voice- \varnothing]. Then, the removal of a null-exponent node is demanded in this structure to produce [_{√VN} n_{denshirenji} $\sqrt{\text{VN}}_{\text{choori}}$]. The removal of \varnothing is part of impoverishment, which deletes morphosyntactic features irrelevant to morphology (Noyer 1995: 23-24). Finally, $\sqrt{\text{VN}}$ -*a* merger is required to form the word structure: [_a [_{√VN} n_{denshirenji} $\sqrt{\text{VN}}_{\text{choori}}$] a-kanoo].

References

- Baker, Mark C. (1988) *Incorporation: A Theory of Grammatical Function Changing*, University of Chicago Press, Chicago.
- Bruening, Benjamin (2014) ‘Word Formation Is Syntactic: Adjectival Passives in English,’ *Natural Language & Linguistic Theory* 32, 363-422.
- Embick, David (2010) *Localism versus Globalism in Morphology and Phonology*, MIT Press, Cambridge, MA.
- Marantz, Alec (1996) ‘‘Cat’ as a Phrasal Idiom: Consequences of Late Insertion in Distributed Morphology,’’ ms., MIT.
- Namiki, Takayasu (1988) ‘A Contrastive Analysis of Japanese Compounds and English Derivatives: Word + *kanoo* and Word + *-able*,’ *Bulletin of Ibaraki University (Studies in Humanities)* 37, 53-75.
- Noyer, Rolf (1995) ‘Impoverishment Theory and Morphosyntactic Markedness,’’ ms., Brown University.

Inflectional complexity from a cross-linguistic perspective

Matías Guzmán Naranjo (Tübingen University - mguzmann89@gmail.com)

Olivier Bonami (Université de Paris, LLF - olivier.bonami@linguist.univ-paris-diderot.fr)

Ackerman and Malouf (2013) propose the distinction between Enumerative (E) complexity and Integrative (I) complexity. The first is the complexity in morphosyntactic distinctions and the way languages encode them, while the second one is the difficulty a paradigm poses to speakers in terms of implicative relations.

E-complexity has received considerable attention in the typological literature (Baerman, Brown, and Corbett, 2015; Dressler, 2011, for an overview), but so far little work has gone towards developing implementations that can automatically and systematically quantify the E-complexity of a language. The large majority of work on E-complexity relies on handcrafted linguistic analysis, which are not always commensurable across researchers and languages. At the same time, while there are multiple computational proposals for capturing I-complexity (Ackerman and Malouf, 2013; Bonami and Beniamine, 2016; Cotterell et al., 2019; Guzmán Naranjo, 2020; Marzi et al., 2018), most studies have looked at a relatively small samples and the emphasis has not been on cross-linguistic comparison (although see Cotterell et al., 2019). This means that we still do not have a picture of how I-complexity varies across languages and systems. For example, one still open question is how verb and noun paradigms compare crosslinguistically in terms of I-complexity.

While recent studies have proposed well formalized methods for exploring I-complexity (Bonami and Beniamine, 2016; Guzmán Naranjo, 2020), these implementations are very costly in terms of dataset size and quality as well as computational time, which makes them unsuitable for large scale typological studies. In this talk we present a new hybrid method. We focus pairwise analogies as proposed by Bonami and Beniamine (2016), but instead of calculating the conditional entropy we fit a classifier to calculate predictive accuracy as suggested by Guzmán Naranjo (2020).

Our approach consists of two steps. First, we find all optimal pairwise alternations for all pairs of cells in a given paradigm. Second, we perform analogical classification to predict, from the form in a predictor cell, which form is expected in a predicted cell. To do this, we first compute an edit distance matrix between the forms of all lexemes in the predictor cell. For a given predictor form, we predict as an output the result of applying to it the compatible alternation that is the most prevalent among its five nearest neighbors in the distance matrix. We check how frequently that prediction is correct, giving us an accuracy score for analogical prediction. We average these accuracies over all ordered pairs of cells to quantify the I-complexity of the system: the higher the accuracy, the less complexity the system has.

For E-complexity we present two approaches. First, we use Morfessor (Virpioja et al., 2013) to extract morphemes from corpora automatically. We use this automatically segmented corpus to calculate the average number of morphemes per word of a language, similar to the index of synthesis by Greenberg (1960)). Second, we develop a new metric we call fragmentation which instead of being morpheme-based, is coached on a Word and Paradigm perspective. The fragmentation of an alternation is the number of variables in it. This captures the idea that the alternation $Xa \square X_o$ is simpler than the alternation $XaY_i \square X_oY_i$.

We applied these methods to a combination of the Unimorph dataset (Kirov et al., 2018) as of the 5th of January of 2021, and another 12 datasets, for a total of 110 datasets (including nouns, verbs and adjectives). To control for the fact that different datasets have different sizes, for each

dataset, we repeated the process for subsets of up to 200, 500, 1000, 2000, and 5000 lexemes. To train the Morfessor we used bible corpora (Mayer and Cysouw, 2014) in order to have comparable results across all languages.

Table 1 shows a subset of mean I-complexity for some of the languages in our dataset. A consistent result is that as datasets increase in size, I-complexity rapidly decreases. For example, while the mean predictability of the 200 lexeme Hungarian noun dataset was of 0.87, the mean predictability of the 2000 lexeme dataset was of 0.94. Including more nouns increases the number of different inflection classes but it also makes predictions easier because the classifiers have more reliable neighbors. If there are enough items in a dataset, a large number of classes does not increase the I-complexity of the system for speakers. From a practical perspective this shows that we cannot reliably estimate the complexity of an inflectional system based on a small tables of inflection classes. Regarding E-complexity, we find that there is a mild correlation between I-complexity and fragmentation. Cells with high fragmentation tend to be harder to predict than cells with low fragmentation. Contrary to Cotterell et al. (2019), we also found that the number of cells in a paradigm did not correlate with I-complexity.

lexeme	cell 1	cell 2	proportion		
lexeme 1	pata	pato	Xa	□	Xo
lexeme 2	mata	mato	Xa	□	Xo
lexeme 3	kis	ki	Xs	□	X

Table 1: Example of analogical proportions

References

- Ackerman, Farrell and Robert Malouf (2013). “Morphological Organization: The Low Conditional Entropy Conjecture”. In: *Language* 89.3, pp. 429–464.
- Baerman, Matthew, Dunstan Brown, and Greville G. Corbett, eds. (2015). *Understanding and Measuring Morphological Complexity*. 1st ed. Oxford University Press.
- Bonami, Olivier and Sacha Beniamine (2016). “Joint Predictiveness in Inflectional Paradigms”. In: *Word Structure* 9.2, pp. 156–182.
- Cotterell, Ryan et al. (2019). “On the Complexity and Typology of Inflectional Morphological Systems”. In: *Transactions of the Association for Computational Linguistics* 7, pp. 327–342.
- Dressler, Wolfgang U (2011). “The Rise of Complexity in Inflectional Morphology”. In: *Poznań Studies in Contemporary Linguistics* 47.2, p. 159.
- Greenberg, Joseph H (1960). “A quantitative approach to the morphological typology of language”. In: *International journal of American linguistics* 26.3, pp. 178–194.
- Guzmán Naranjo, Matías (2020). “Analogy, Complexity and Predictability in the Russian Nominal Inflection System”. In: *Morphology* 30, pp. 219–262.
- Kirov, Christo et al. (2018). “UniMorph 2.0: Universal Morphology”. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. Paris, France: European Language Resources Association (ELRA).
- Marzi, Claudia et al. (May 2018). “Evaluating Inflectional Complexity Crosslinguistically: A Processing Perspective”. In: *Proceedings of the 11th Language Resources and Evaluation Conference (Miyazaki, Japan)*. European Language Resource Association.
- Mayer, Thomas and Michael Cysouw (2014). “Creating a massively parallel bible corpus”. In: *Oceania* 135.273, p. 40.
- Virpioja, Sami et al. (2013). *Morfessor 2.0: Python implementation and extensions for Morfessor Baseline*. Tech. rep. Aalto University

Reduced agreement: a system that some elements ignore

Tania Paciaroni¹, Sebastian Fedden^{1,2,3} & Greville G. Corbett³

¹Ludwig-Maximilians-Universität, University of Zurich; Université Sorbonne Nouvelle;
University of Surrey

tania.paciaroni@romanistik.uni-muenchen.de; sebastian.fedden@sorbonne-nouvelle.fr;
g.corbett@surrey.ac.uk

Recent years have shown a growing interest in agreement, especially in non-canonical agreement (Fedden, Audring & Corbett 2018). A challenging area is reduced agreement, the situation where fewer morphosyntactic distinctions are made than the system makes available. For example, German attributive adjectives have more than one set of inflectional forms, depending on the syntactic context. The ‘strong’ form in (1) is a dedicated neuter form:

(1)	ein	neu-es	Buch
	INDF.NOM.N.SG	new-NOM.N.SG	book(N)[NOM.SG]
	'a new book'		

The ‘weak’ form in (2), required after the definite article, fails to distinguish neuter gender:

(2)	das	neu-e	Buch
	DEF.NOM.N.SG	new-NOM.SG	book(N)[NOM.SG]
	'the new book'		

This phenomenon was presented as ‘reduced agreement’ (RA) in Corbett (2006: 93-97). Since then it has received little consideration from typologists. Yet fascinating new systems keep being found, especially in works on Romance (see e.g. Pomino & Stark 2009, Bonet 2018, Cappellaro 2018).

It is time, therefore, to offer an overall typology of RA. We take a canonical perspective: canonically, all lexemes agree fully and consistently, i.e. they can show all values of all relevant features and they always do so (Corbett 2013: 57). This idealized baseline leads us to distinguish three types of limitation (which can co-occur): 1. lexemes; 2. features; 3. conditions. We focus on specific reductions, both *partial* and *complete*, from a wide range of languages.

1. On the level of the **lexeme**, *partial* reduction is found in Latin where adjectives like *bonus*(M)/*bona*(F)/*bonum*(N) ‘good’ distinguish three genders, whereas adjectives like *fortis*(M/F)/*forte*(N) ‘strong’ distinguish only two. *Complete* reduction is found in some Papuan languages (Fedden 2011, 2020): in Mian about one third of transitive verbs agree with their object, e.g. *-nâ* ‘hit’, while the rest do not, e.g. *bou* ‘swat’. A similar phenomenon of agreeing and non-agreeing lexemes within a part of speech are observed in the Nakh-Dagestanian languages Archi (Chumakina & Corbett 2016) and Tsez (Polinsky & Comrie 1999).
2. On the **feature** level, *partial* and *complete* reduction are found in superclassing, where some but not all possible distinctions are made. The Australian language Jingulu (Pensalfini 2003) has masculine, feminine, vegetable and neuter genders. Adjective agreement can be canonical or partially reduced, where masculine can be used for masculine and feminine controllers and neuter for vegetable and neuter controllers. Alternatively, masculine can be used as an ultimate default with any of the four genders (completely RA).
3. Reduction on the feature level with an added syntactic **condition** is found in German adjectives (2). And in the Italo-Romance variety of Ripatransone, items belonging to

almost all parts of speech have two sets of inflectional forms with the same feature specifications, a full and a reduced set, whose selection depends on conditions at different linguistic levels (Paciaroni & Loporcaro 2018).

In conclusion, we

1. offer a new, fuller typology of RA systems;
2. show that they are more prevalent than previously realized;
3. review the findings of Nichols (2018) and Fedden (in press), for the insights RA can offer on the function of agreement.

References

- Bonet, Eulàlia. 2018. Missing inflectional features and missing exponents in DP-internal agreement asymmetries. *Glossa: a journal of general linguistics* 3(1), p.79. [doi: <https://doi.org/10.5334/gjgl.579>]
- Cappellaro, Chiara. 2018. Why is Ladin lazy agreement a feminine gender issue? *UCD Working Papers in Italian Studies* 1, 4-20.
- Chumakina, Marina & Corbett, Greville G. 2015. Gender number marking in Archi: Small is complex. In Baerman, Matthew; Brown, Dunstan & Corbett, Greville G. (eds), *Understanding and measuring morphological complexity*, 93-116. Oxford: Oxford University Press
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013. Canonical morphosyntactic features. In Brown, Dunstan; Chumakina, Marina & Corbett, Greville G. *Canonical Morphology and Syntax*, 48-65. Oxford: Oxford University Press.
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: De Gruyter Mouton.
- Fedden, Sebastian. 2020. Morphology in Trans New Guinea languages. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. [doi:10.1093/acrefore/9780199384655.013.633]
- Fedden, Sebastian (in press). Agreement and argument realization in Mian discourse. In Tania Paciaroni, Alice Idone and Michele Loporcaro (eds.), *The Many Facets of Agreement*. Special Issue of *Word Structure*.
- Fedden, Sebastian, Audring, Jenny & Corbett, Greville G. (eds). 2018. *Non-canonical gender systems*, Oxford: Oxford University Press.
- Nichols, Johanna. 2018. Agreement with overt and null arguments in Ingush. *Linguistics* 56(4), 845-863.
- Paciaroni, Tania & Loporcaro, Michele. 2018. Overt gender marking depending on syntactic context in Ripano. In Fedden, Sebastian; Audring, Jenny & Corbett, Greville G. (eds), *Non-canonical gender systems*, 147-75. Oxford: Oxford University Press.
- Pensalfini, Robert. 2003. *A Grammar of Jingulu: An Aboriginal Language of the Northern Territory* (Pacific Linguistics 536). Canberra: Pacific Linguistics.
- Polinsky, Maria & Comrie, Bernard. 1999. Agreement in Tsez. In Corbett, Greville G. (ed). *Agreement*. (Special Issue of *Folia Linguistica* 33/2), 109-130. Berlin: Mouton de Gruyter.
- Pomino, Natascha & Stark, Elisabeth. 2009. *Adnominal Adjectives in Romance. Where Morphology seemingly meets Semantics*. In: Espinal, M. Teresa; Leonetti, Manuel & McNally, Louise (eds). *Proceedings of the IV Nereus International Workshop: Definiteness and DP Structure in Romance Languages*, 113-135. Konstanz: Fachbereich Sprachwissenschaft der Universität Konstanz.

A usage-based approach to the opacity and alternation of locatives in Hungarian

László Palágyi
Eötvös Loránd University
palagyi.laszlo@btk.elte.hu

There has been little discussion on two problems of Hungarian locative expressions: neither the issue of opacity (see examples (1) and (2)), nor alternation (see (2) and (3)) has been dealt with in depth. Kracht considers opaque instantiations of Hungarian spatial cases to be idiosyncrasies motivated by the peculiar shape of a location and concludes that „the choice of localizers is not completely random but at the same time largely unpredictable” (2005: 151).

- | | | |
|-----------------|---------------------|------------------------------------------------------------|
| (1) a hajó-n | ‘on the ship.SUPE’ | (It might denote the board as well as the cabin.) |
| (2) a metró-n | ‘on the metro.SUPE’ | (It denotes the interior of the vehicle.) |
| (3) a metró-ban | ‘in the metro.INE’ | (It denotes the interior of the vehicle, and the station.) |

This study focuses on the IN-configuration and ON-configuration markers (see Creissels 2008) of the word forms in question to explore systemic motivation behind them. To handle their apparent idiomatic nature, a constructional model is adopted in which the spatial cases are treated as word-level schematic form-meaning pairings (see Booij 2010a) that are generalisations over the lexicon. Constructions are not fully independent of their instantiations (see Booij 2010b), productivity of a construction rests upon its exemplar cluster (Bybee 2010), exemplars provide associative links to the potential bases via their own that can then serve as a reference point for analogical extensions. Thus, it is assumed that the characteristics of lexical organisation are detectable in inflection as well (cf. Gaeta 2008). Accordingly, we hypothesise that Hungarian constructions that are ON-configuration markers (expressing location, source, and destination as ‘sister schemas’) are associated with the {PUBLIC VEHICLE} lexical-semantic class. The study aims at demonstrating and interpreting this associative relationship with a corpus-based morphosemantic investigation.

As for lexical-semantic classes, they are considered dynamic representations. Semantic classification is a context-dependent categorising process that correlates to the chosen way of conceptualisation known as construal (Langacker 1987). A noun invokes several abstract properties arranged in line with a variety of cognitive domains of which “only a limited number can be activated on a given occasion” (Langacker 1987: 57). Thus, the same lexical item might direct one’s attention to a transportation device (~CANONICAL FUNCTION domain) or to a bounded object with doors and windows (~STRUCTURE domain) depending on the foregrounded semantic sub-structure, which partially explains the alternation of locatives (see again (2) and (3)). However, class membership can relatively be fixed in usage patterns, lexical meaning can be characterised by likelihood of activation because certain aspects of the encyclopaedic knowledge associated with a word are more salient than others.

Through the use of corpora, it is possible to demonstrate the correlations between lexical-semantic classes and spatial cases, to analyse the alternation of locatives, and to explore the central cognitive domain of the nouns. It is the examination of relative frequency (distribution of a construction in a noun’s inflectional pattern) that enables us to draw consequences concerning to which extent a noun associates a construction. The method relies upon the assumption that if a construction is associated with a lexical-semantic group, members of the latter bear resemblance in the distribution of the construction. Examining in Hungarian National Corpus (HNC) (Oravecz et. al) the relative frequency of ON-configuration and IN-configuration marker cases concerning twenty nouns naming vehicles and buildings we found by *k*-means clustering that the nouns can be divided into two clusters according to their lexical-semantic properties (i.e., vehicles and buildings).

Evidence for the associative relationship can also be found by looking at the verbs with which the spatial cases co-occur, because verbs profile different aspects of nominal meaning. Even in the case of alternation, our hypotheses predict that the distribution of localisers differs according to the cognitive domains the verbs evoke. Examining eleven Hungarian nouns naming vehicles in the corpus of HNC we found that, in each noun, the token frequency of the superessive and sublative was higher in the context of the verbs that directly activate the CANONICAL FUNCTION domain, i.e., *utazik* ‘travel’ and *(fel/be)száll* ‘get on’, ‘get in’ (the preverbs express direction, and they are coordinated with the case), than in the context of those which might primarily evoke the STRUCTURE domain, i.e., *(fel/be)ül* ‘sit on(to)’, ‘sit in(to)’. However, as hypothesised, there was a major distinction between the public vehicles and the other sub-classes. The former occurred with an ON-configurational construction much more frequently in both types of contexts (comparing to the overall distribution of the spatial cases within the argument structure of the verbs in VAB, Sass 2009), which suggest that they are distinctly classified according to their CANONICAL FUNCTION. On the other hand, as for other vehicles, IN-configuration markers function as default, which might be traced back to the activation of the STRUCTURE domain that is central to their conceptual matrix. Nevertheless, it is still possible (via analogy) to foreground the telic semantic aspects and background the structural properties of any vehicles by integrating the superessive with a semantically correspondent verb (e.g., *autó-n utazik* ‘travel by car.SUPE’ occurs as well).

As far as the experiential basis of the locatives is concerned, it is obvious that perceptual experiences are crucial to the establishment of a morpholexical cluster (see Kothencz 2012; Szilágyi N. 1996; Tolcsvai Nagy 2013), for instance, busses, trams, etc. usually had steps that passengers had to ascend, which might explain the associative relationship between ON-configuration and public vehicles. However, our data suggest that speakers tend to overlook kinaesthetic perceptual experiences when an already established exemplar cluster gains a new member by analogical extension mainly based on culturally accessible associative information, as formally summarised in (4).

(4) *busz* {PUBLIC VEHICLE} : *busz-ra* [SUBLATIVE] ‘SURFACE and GOAL’ ~ *metró* {PUBLIC VEHICLE} : *x* (= *metró-ra*)

References • BOOIJ, G. 2010a. Construction Morphology. Oxford: Oxford University Press. • BOOIJ, G. 2010b. Compound construction: schema or analogy? A Construction Morphology perspective. In: *Cross-disciplinary studies in compounding*. Scalise, S. and Vogel, I. (eds.) Amsterdam/Philadelphia: John Benjamins. 93–108. • BYBEE, J. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press. • CREISSELS, D. 2008. Spatial cases. In: Malchukov, A. and Spencer, A. (eds.) *The Oxford Handbook of case*. Oxford: Oxford University Press. 609–625. • GAETA, L. 2008. On the double nature of productivity in inflectional morphology. *Morphology* 17: 181–205. • ORAVECZ, Cs. & VÁRADI, T. & SASS, B. *The Hungarian Gigaword Corpus*. In: *LREC 2014 Proceedings*. Reykjavik: ELRA. 1719-1723. • KOTHENCZ, G. Sz. 2012. Képi sémák és határozóragok. In: Tolcsvai Nagy, G. & Tátrai, Sz. (eds.) *Konstrukció és jelentés. Tanulmányok a magyar nyelv funkcionális leírására*. Budapest: Eötvös Loránd Tudományegyetem. • KRACHT, M. 2005. The semantics of locatives in Uralic languages. In: Jocelyne Fernandez-Vest, M. M. (ed.) *The Uralic Languages Today. A Linguistic and Cognitive Approach*. Paris: Honoré Champion. 145–158. • LANGACKER, R. W. 1987. *Foundations of Cognitive Grammar*. Stanford: Stanford University Press. • SASS, B. 2008. The Verb Argument Browser. In: Sojka, P. et al. (eds.) *Proceedings of TSD, 2008*, Brno. LNCS 5246. 187–192. • SZILÁGYI, N. S. 1998. *Hogyan teremtsünk világot?* Kolozsvár. • TOLCSVAI NAGY, G. 2013. *Bevezetés a kognitív nyelvészetbe*. Budapest: Osiris.

Inflectional class interactions and valency changes in Matlatzinca

Enrique L. Palancar, *enrique.palancar@cnr.fr*, CNRS, France
 Leonardo Carranza, *leo.carranza.mtz@gmail.com*, UIEM, Mexico

We show a case of how changes in the inflection class membership of verbs allow for changes in their valency and their meaning. We show the case in Matlatzinca (Oto-Pamean, Oto-Manguéan, Mexico). Verbs in Matlatzinca often display multiple class membership, in such a way that a given verb may be inflected as transitive in one class but as intransitive in another. When this happens, the verbs in question display slight differences in meaning which are of the type observed in valence changing mechanisms in other languages. For example, the verb CHUNTA is a transitive verb meaning ‘waking sb. up’ when inflected as a transitive verb, but as a member of the intransitive classes, it is an intransitive verb conveying the resulting state of ‘being awake’ or a verb that conveys the spontaneous event of ‘waking up’. This phenomenon is typologically and theoretically interesting, because it raises the question as to whether such meanings instantiate three different lexical entries, or different aspects of the same verb (Spencer 2013).

Matlatzincan verbs are inflected in TAM and in person/number (P/N) of subject by means of an intricate series of grammatical markers that precede the stem, which are prosodically and syntactically independent of the stem (i.e., they are not prefixes), e.g. the complex *ri+’o’+rú* in (1) realizing second person singular (2SG) plus completive irrealis (CPL.IRR). This situation means that Matlatzincan verbal inflection is periphrastic in composition. Adverbials indicating notions of associated motion (among others) appear between these markers and the verbal stem, such as the ventive from a higher place (VEN) *nen* in (1). The stem of some verbs is preceded by a stem formative (STF) (e.g. *tú*, which encliticizes phonologically to the preceding word).

- (1) *ri+’o’+rú* *nen=tú* *huhti*
 CPL.IRR.2SG VEN=STF blow.up[O3]
 ‘You’ll blow it (the balloon) up as you’ll come down’

Verbs are divided into five inflection classes: three of them (I-III) comprise only transitive verbs, while all intransitive verbs belong to the two remaining classes (A and B). Inflection class distinctions revolve around a combination of selecting different grammatical markers and stem formatives. The formal contrast between the five classes is shown in Table 1 for 2SG forms only. In the completive realis (CPL.REALIS), verbs of classes I-III receive the marker ‘*o*’, while class A and B verbs receive ‘*i*’ and ‘*é*’, respectively. The contrast between classes I-III is observed in the CPL.IRR, where class I receives *ri+’o’* and classes II and III *ri+’o’+rú*; the contrast between classes II and III is based on the fact that class III verbs require the stem formative *tú*, while class II verbs do not.

2SG		TAM+P/N	VEN	STEM		
CPL.REALIS	I	’o’	nen	táni	‘you bought it as you came down’	
	II	’o’	nen	habi	‘you left it as you came down’	
	III	’o’	nen	huhti	‘you blew it up as you came down’	
	A	’i’	nen	chapi	‘you jumped as you came down’	
	B	’é’	nen	té	bati	‘you got lost as you came down’
CPL.IRR	I	ri+ ’o’	nen	táni	‘you’ll buy it as you’ll come down’	
	II	ri+ ’o’ +rú	nen	habi	‘you’ll leave it as you’ll come down’	
	III	ri+ ’o’ +rú	nen	tú	huhti	‘you’ll blow it up as you’ll come down’
	A	ri+ ’o’	nen	chapi	‘you’ll jump as you’ll come down’	
	B	ri+ ’o’	nen	té	bati	‘you’ll get lost as you’ll come down’

Table 1. Inflection class distinctions

We show that verbs in Matlatzinca may and often display multiple class membership. By multiple class membership we refer to a situation whereby verbs that share the same lexical stem and have identical lexical semantics can be inflected as members of different inflection classes. For descriptive purposes, we talk about this phenomenon in terms of ‘verb pairings’. In such pairings, the verbs involved often contrast in transitivity value (although not always) and display differences in grammatical meaning of the type observed in valence changing mechanisms in other languages, commonly attained through affixation or stem changes. For example transitive verbs (classes I-III) frequently have an intransitive correlate in classes A or B. Consider the examples in Table 2. The verbs in (a) convey accomplishments when transitive, but activities when intransitive of class A. The transitive verbs in (b) have middle voice semantics in their intransitive pairings in class B.

		Tr	Action	Intr	Activity
a.	TANI	I	‘buy sth.’	A	‘go shopping’
	’IWI	II	‘make sb. go to sleep’	A	‘sleep’
	CHANA	III	‘shout at sb.’	A	‘scream’
b.	KWANCHI	I	‘hit sth./sb.’	B	‘hit each other’
	TOCHI	I	‘break sth.’	B	‘get broken’
	HATI	II	‘cover sth./sb.’	B	‘cover up (e.g. with blanket)’

Table 2. Examples of verb pairings

In our paper, we study a sample of 500 different verbal lexical stems that we have collected through elicitation or from natural texts. We have registered the instances where these stems represent verbs that can only be inflected in one class (unpaired verbs) or in more than one. The number of paired verbs is necessarily higher than the number of stems, especially because there are instances where a given stem can be found to represent up to three different, but lexically related, verbs (we have not found more than three pairings).

We establish the semantic differences involving the pairings and we point to typologically similar (but not yet free of controversy) phenomena occurring in languages like Latin, Greek and French, as discussed in the literature. We raise awareness of the theoretical problem involving multiple class membership by suggesting different ways that we can talk about it. One possible way is to think that a given lexical entry would come with an associated network of different paired senses bearing different inflectional properties. As a consequence of the latter, the verbs would then be inflected differently depending on the sense in which they are used. But it would mean that for every verb we would have to postulate two or three paradigms. As there are many unpaired verbs, we would be forced to say that such verbs have some defective inflection. Another alternative view is to treat the diathetic changes observed as involving derivation. This implies that multiple class membership is accounted for if we think of the paradigms that constitute the different inflection classes as if they were invested with derivational power; that is, as if the classes had a semantic core associated to them. This would be particularly the case of the intransitive conjugations in Matlatzinca, where class A holds activity verbs proper of semantic antipassives (“buy sth.” > “go shopping”), but more particularly so of class B, which we treat as a mediopassive paradigm with verbs that convey typical middle voice semantics (“scratch sth.” > “scratch (oneself)”).

References

Spencer, Andrew. 2013. *Lexical Relatedness*. Oxford: Oxford University Press.

Overabundance in English verbs: Canonicity and distribution

Jeff Parker, Brigham Young University, jeff_parker@byu.edu

Overabundance is when two forms exist for the same paradigm cell. For example, both *strove* and *strived* are well-attested past tense forms of STRIVE in English. From a Canonical Typology perspective (Corbett 2007), overabundance is a deviation from canonical inflection because it violates the criterion of being “univocal: every cell contains one form” (Thornton 2011). Examples of overabundance, which are themselves non-canonical inflection, can be investigated in canonical terms. Looking at data from a variety of languages, Thornton (2019) identifies six ways of being overabundant and four criteria that define canonical overabundance. In this paper, I provide a (near) comprehensive corpus-based look at overabundant English verbs, highlighting the ways English verbs are overabundant, ways in which they deviate from canonicity and how their distribution in the lexicon compares to previous investigations of overabundance in other languages. Despite having comparatively poor inflection, English verbs provide insight into the nature and distribution of overabundance.

To investigate overabundance in English verbs, all overabundant verbs among the 60,000 most frequent lexemes in COCA (Davies 2012) were identified. The resulting set consists of 74 English verbs lexemes that are overabundant in the past tense (15), the past participle (27) or both (32), which is comparable in size to corpus studies on overabundance in highly inflecting languages like Czech (Bermel and Knittl 2012) and Croatian (Botica and Hržica 2016). Among these examples, four of the six ways of being overabundant are attested, including suppletive stems (e.g., *spoke.PST* vs. *spake.PST*), phonologically identical stems belonging to different inflection classes (e.g., *disproved.PPT* vs. *disproven.PPT*), suppletive stems belonging to different inflection classes (e.g., *saw.PST* vs. *seen.PST*), and single vs. double marking (e.g., *dreamed.PST* (affix) vs. *dreamt.PST* (stem change and affix)). Of the four criteria for canonical overabundance, English verbs show deviations in all four criteria: overabundance can be found in more than one cell; overabundance is present in more than one lexeme; the frequency ratio between forms varies greatly (see discussion below); and, various type of conditioning are present.

In addition to the various ways that overabundance is represented in English, I show that some distributional aspects of overabundance in English are informative compared to overabundance in other languages. The frequency proportion between overabundant forms, i.e., how much more frequent one form is than the other, covers the full possible range (.5 to 1 in Figure 1) in a skewed distribution, not dissimilar from the range covered by Czech genitives (Bermel and Knittl 2012). This suggests a cross-linguistic similarity in one distributional aspect of overabundance. This frequency proportion is positively correlated with lemma frequency ($F(1, 104) = 76.64$, Adj. $R^2 = 0.418$, $p < 0.001$): the more frequent a lemma, the more likely it is to have a form that is much more frequent than its counterpart (see Figure 2 below). At the same time, the lemma frequency of overabundant English verbs is limited in that overabundance does not occur in the lowest range of the frequency spectrum (See Figure 3 below). Of the 5764 verbs represented in Figure 3, over 53% are less frequent than the least frequent overabundant verb. Furthermore, seven of the 25 most frequent verbs are overabundant, suggesting that overabundance exists among items that are not very low frequency (at least in English). This is surprising given previous research in other languages noting that the “persistence [of overabundance-JP] correlates positively with *low frequency* and *later acquisition*” (Cappellaro 2018:1). Unlike the distribution of frequency ratios between forms, this cross-linguistic difference suggests that factors that promote overabundance are influenced by language-specific traits like the number, types and productivity of inflectional patterns.

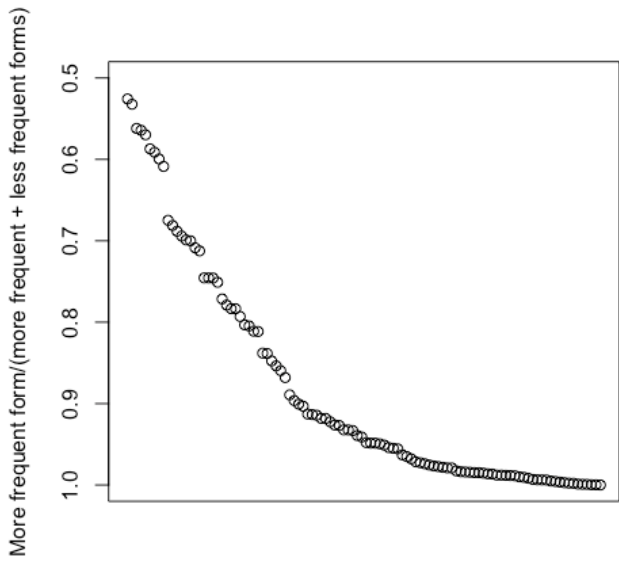


Figure 1. Distribution of frequency proportion of overabundant forms

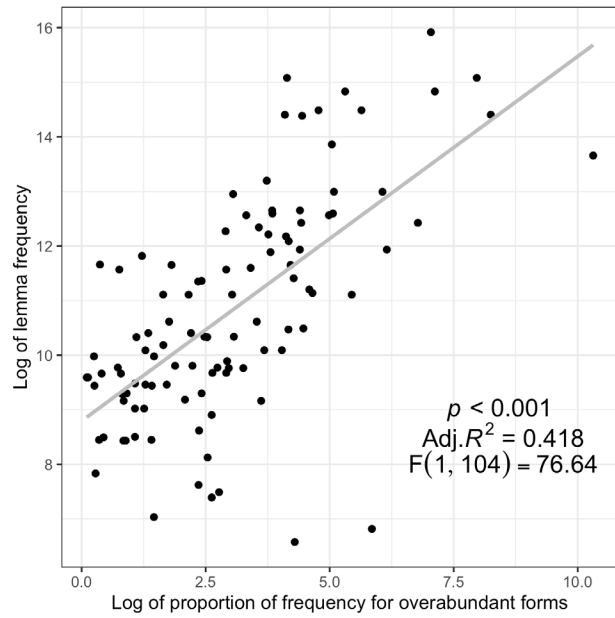


Figure 2. Correlation between lemma frequency and frequency proportion of overabundant forms

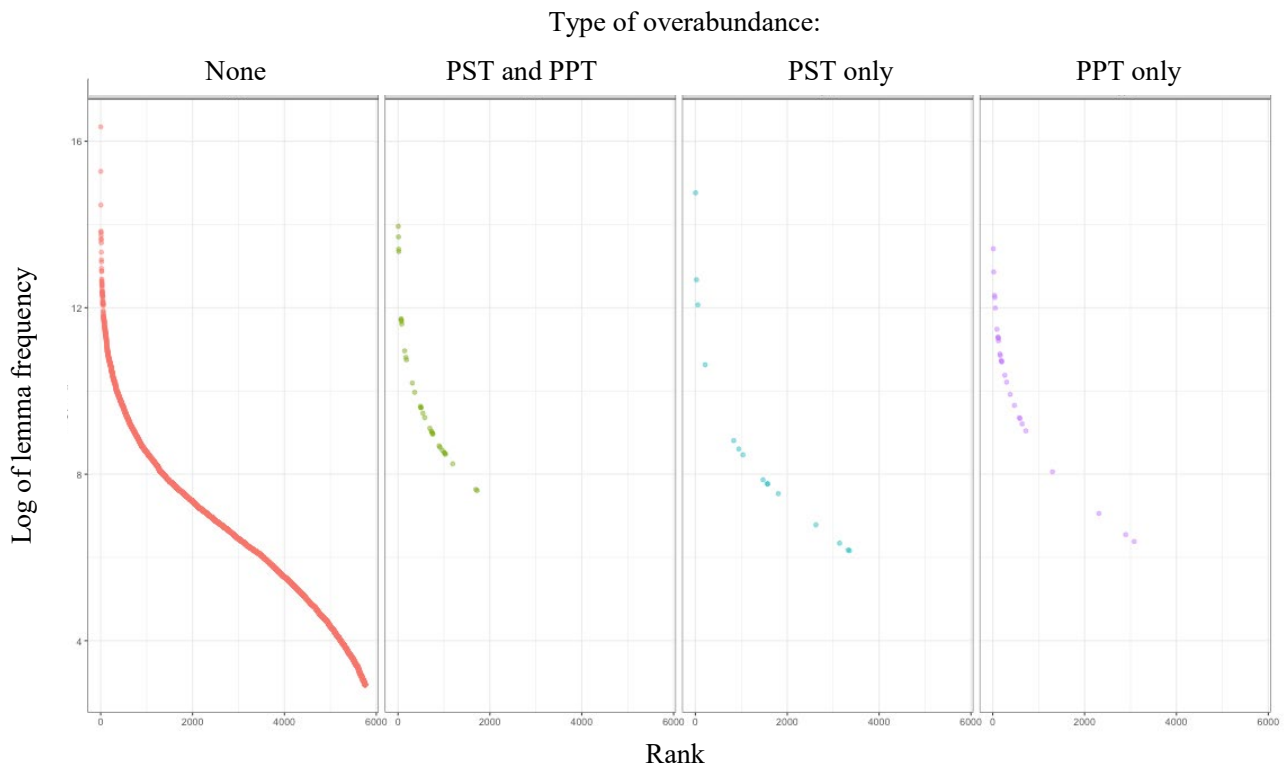


Figure 3. Distribution of lemma frequency across overabundant and non-overabundant English verbs

Modeling German singular and plural recognition in aphasia with discriminative learning

Ingo Plag, Maria Heitmeier & Frank Domahs

Heinrich-Heine-Universität Düsseldorf, Universität Tübingen & Universität Erfurt

German noun plural formation took center stage in the debates about the architecture of the lexicon, morphology and grammar in the 1990s and early 2000s. More recently, computational studies using deep neural networks have started new attempts to solve the problem of irregularity and semi-regularity characterizing the German noun system. From a linguistic point of view, however, deep learning networks are not very interesting as they act as black boxes. In this paper we employ an alternative to deep learning networks, discriminative learning networks (e.g. Baayen et al. 2019), to test whether measures gleaned from discriminative networks can be used to predict grammatical number of German word-forms, and the number decisions and errors of an aphasic patient confronted with German plural and singular noun phrases.

The data for the present study are taken from Domahs et al. (2017), who tested grammatical number decisions of a patient with Primary Progressive Aphasia. They found that the accuracy of the patient's decisions was influenced by the degree to which stimuli conformed to the prototypical, phonologically defined, schema for plural nouns (Köpcke 1988, 1993).

The present study implements a radically different approach, naive discriminative learning (NDL), and linear discriminative learning (LDL). In the NDL analyses, we use discriminative association measures between the phonological representations and the traditional phonological variables on the one hand, and number on the other, in order to predict number and the patient's responses and errors. In the LDL models, following Heitmeier et al. (2021), we used word2vec vectors as semantic representations, and diphones and the traditional phonological variables as form representations, and created two-layer networks that map form representations directly onto semantic representations (modeling comprehension), and that map semantic representations onto form representations (modeling production). We then used the association measures between form and meaning to predict the grammatical number of a given word form, and to predict the number decisions and the errors by the patient when confronted with a given singular or plural form.

The NDL models reveal that the patient has similar association weights for the different phonological cues, with a general tendency towards lower association weights in general. Three differences emerge between the patient and the German system: an insensitivity towards stress other than penult stress, an insensitivity towards determiners other than *die*, and a different association pattern for neuter gender. The association weights are also predictive of the patient's errors. Those plural word-forms that have weaker associations with plural, i.e. are not easy to discriminate from singulars, are more error-prone. The analogous effect holds for singulars: For a singular form, the weaker the support for singular, the more likely the singular is judged erroneously to be a plural.

The LDL measures very successfully distinguished between singular and plural forms, as shown in Figure 1 (see next page). The t-SNE plot reduces the dimensions of the matrix with the LDL measures to two dimensions. As shown by the distribution of the dots (for singulars) and triangles (for plurals), singulars and plurals cluster separately.

In logistic regression models with LDL measures as predictors for the number decision of the aphasic patient, the LDL measures also turned out to be highly predictive. Plural decisions are favored for word forms that live in a denser semantic network, and whose articulatory trajectory has a stronger association with the word's semantic vector.

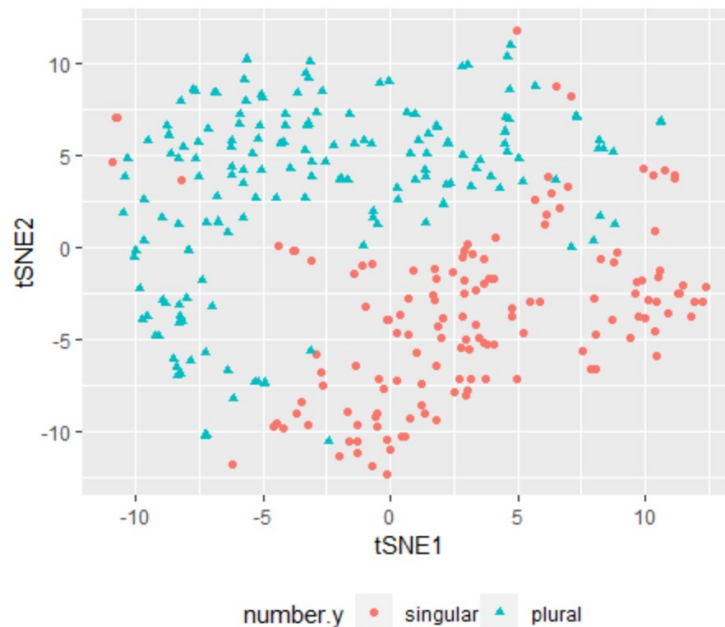


Figure 1: tSNE plot of LDL measures by number

These results have important theoretical implications. First, a discriminative learning approach can help to better understand the problems of an aphasic patient. Second, semantics also plays an important role in the marking of German plural. Third, discriminative learning is a viable alternative to linguistically uninformative neural networks. Finally, the results provide further support for the idea that behavioral patterns can be understood as emerging from the distributional properties of words on the one hand, and basic principles of human learning on the other.

References

- Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (De)composition but in linear discriminative learning. *Complexity* 2019, 1–39. doi: 10.1155/2019/4895891
- Domahs, Frank, Lisa Bartha-Doering, Ulrike Domahs & Margarete Delazer. 2017. Wie muss ein 'guter' deutscher Plural klingen? In Nanna Fuhrhop, Renata Szczepaniak & Karsten Schmidt (eds.), *Sichtbare und hörbare Morphologie* (Linguistische Arbeiten Band 565), 205–237. Berlin, Boston: De Gruyter.
- Heitmeier, Maria, Yu-Ying Chuang & R. H. Baayen. 2021. Modeling morphology with Linear Discriminative Learning: Considerations and design choices. *Frontiers in Psychology* (Language Sciences) 12, <https://doi.org/10.3389/fpsyg.2021.720713>.
- Köpcke, Klaus-Michael. 1988. Schemas in German plural formation. *Lingua* 74. 303–335.
- Köpcke, Klaus-Michael. 1993. *Schemata bei der Pluralbildung im Deutschen. Versuch einer kognitiven Morphologie*. Tübingen: Narr.

Why theme vowels matter: Some insights from Romance

Natascha Pomino, University of Wuppertal, pomino@uni-wuppertal.de &

Eva-Maria Remberger, University of Vienna, eva-maria.remberger@univie.ac.at

Theme vowels (ThV) are considered by many linguists (e.g. Oltra-Massuet 1999) to be mere ornamental elements without any effect on syntax and semantics. We agree on this assumption, in principle, rejecting the idea that ThVs are the realizations of the verbalizer v° (cf. e.g. Fábregas 2017), but we would like to argue for the view that ThVs are not as ornamental as one may think: ThV have an impact on the (ir)regularity of verbal forms (cf. Calabrese 2015). Work on the systematicity of stem allomorphy has shown that irregularity of inflected elements is tightly related to a reduced number of affixes of these forms, cf. Vanden Wyngaerd (2018). His *Suppletion Generalization* (cf. 1a) is illustrated by the Italian past participles (PP) (cf. 1b): The regular PP are formed by adding the endings *-t-* and *-o* to a verbal stem that consist of a verbal root extended by a ThV, i.e. regular PP are thematic verbal forms. In contrast, the irregular PP (with or without stem allomorphy) show allomorphy in the ending; and, what is more, they are athematic. The root is not followed by a ThV, a fact that leads to shorter forms and allomorphy.

- (1) a. *Suppletion Generalization*: If there is irregularity in the form of either the root or the suffixes, the number of suffixes gets reduced. (Vanden Wyngaerd 2018: 1)
- b. Italian past participles (based on Calabrese 2015)

regular				irregular				
root	ThV	ending		root		ending		
am-	-a-	-t-	-o	'loved'	per-	∅	-s- -o	'lost'
batt-	-u-	-t-	-o	'beaten'	cor-	∅	-s- -o	'run'

In our talk, we show that the link between athematicity and irregularity is attested also in other verbal forms and that it holds for French as well, a language for which the assumption of ThV is controversially discussed in the literature. We first discuss the conjugation class (CC) system of French and propose that it has two thematic and several athematic CCs. The athematic CCs are exactly the ones with the (most) irregular verbs. Finally, following Vocabulary Insertion-Only model proposed by Haugen & Siddiqi (2016), within the framework of Distributed Morphology, we propose an analysis for the mentioned link between athematicity and irregularity based on *Spanning*. Since the exact conditions and rules for *Spanning* have not been investigated in detail yet, our proposal and subsequent analyses will contribute to a clarification of this issue. The core idea is that the spanning size of the vocabulary item realizing a root depends on its CC-features and that it is conditioned by the (a)thematicity of the roots. Our analysis is predominantly based on the general process of Vocabulary Insertion in order to correctly derive the Romance forms. There is no need for processes like Fusion, Pruning or Impoverishment, and root suppletion is instead explained in essence via Vocabulary Insertion only.

For French, two well-known cases of root allomorphy will serve as illustration: consonant-zero-alternation and suppletion. The alternation in (2a) shows that the root final consonant of the verb is deleted in some cases (see *vivre*), but maintained in others (see *arriver*). In line with Schane (1966) and others, we will associate this type of allomorphy with the absence of a ThV in those cases where the root final consonant is deleted. The second example to be considered is suppletion with French *aller* 'go', an extreme case of allomorphy. We will see that in the present tense the indicative 1st and 2nd person plural are thematic verbal forms and have the default realization for the root (i.e. *all-* based forms), whereas more marked Vocabulary Items for the root GO (e.g. *va-* based forms) are clearly athematic and show root allomorphy.

- (2) (a) C/∅-alternation: e.g. *vivons* [vivɔ̃] vs. *vis* [vi] / *[viv] (athematic)
but *arrivons* [aʁivɔ̃] vs. *arrives* [aʁiv] / *[aʁi] (thematic)
- (b) Suppletion: e.g. *allons* [alɔ̃] vs. *vas* [va]

Based on this observation, we will argue that the allomorphy *vais* /vɛ/ ‘I go’, *vas* /va/ ‘you_{sg} go’ and *vont* /vɔ̃/ ‘they go’ is triggered by the agreement features in T° that appear span adjacent to the root due to the athematicity of the latter one. Note that this is not the case for the *all-*forms (cf. Fig. 1): Here the ThV, which does not surface as general rule, avoids span adjacency and no influence of the following context on allomorphy is expected, i.e. the default is inserted.

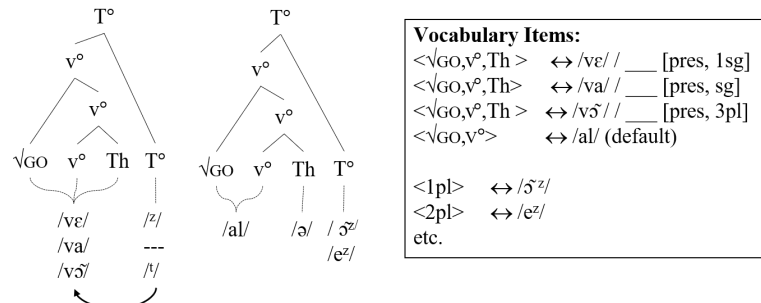


Figure 1: Spanning approach to French GO-suppletion

We will strengthen our analysis with further examples of Romance verbal inflection: For Spanish, we will argue, for example, that the highly irregular forms of *ir* ‘to go’ in the *indefinido* can be analyzed in a similar vein.

(3) Spanish: The *indefinido* for GO and regular verbs

	1st sg	2nd sg	3rd sg	1st pl	2nd pl	3rd pl
GO	<i>fui</i>	<i>fui-ste</i>	<i>fue</i>	<i>fui-mos</i>	<i>fui-steis</i>	<i>fue-ron</i>
CANTAR	<i>cant-é</i>	<i>cant-a-ste</i>	<i>cant-ó</i>	<i>cant-a-mos</i>	<i>cant-a-steis</i>	<i>cant-a-ron</i>

Both verbs in (3) lack an exponent for T° and have a set of exponents for φ that is different from the present tense. We take this as an indicator that, independent of the realization of the root and following elements, T°, Th and φ are realized as one span in the *indefinido*. This explains the exceptional shortness of these forms. In addition, the verb GO shows root alternation between *fue-* and *fui-* which we attribute, again, to the athematicity of the root (cf. Fig. 2).

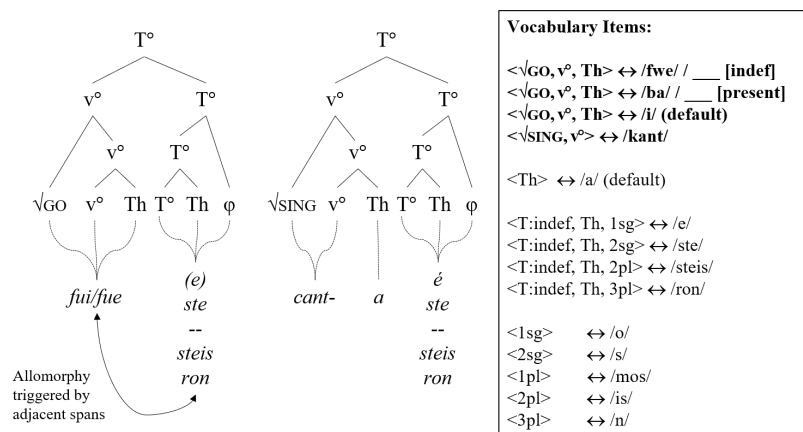


Figure 2: The *indefinido* of Spanish GO

This brings us back to the role of ThVs: In our view, ThVs do not have a significant semantic role, but they are morphological elements whose presence matters in terms of regularity, and whose absence or vanishing (as in language change) triggers allomorphy.

References: Oltra Massuet, I. 1999. On the constituent structure of Catalan verbs. *MIT Working Papers in Linguistics* 33. 279–322. / Fábregas, A. 2017. Theme vowels are verbs. In Caha, DeClercq & Vanden Wyngaerd (eds.), *The unpublished manuscript*. A collection of Lingbuzz papers to celebrate Michal Starke’s 50th birthday. 51–62. / Calabrese, Andrea. 2015a. Irregular Morphology and Athematic verbs in Italo-Romance. *Isogloss, Special Issue on Italo-Romance morphosyntax*: 69–102. / Vanden Wyngaerd, Guido. 2018. Suppletion and affix reduction. Accessed October 29, 2021. <https://lingbuzz.net/lingbuzz/004095>. / Schane, Sanford A. 1966. The morphophonemics of the French verb. *Language* 42/4: 746–758.

Zero-derived formations in Italian (of the type *abbandono*, *sosta*, and *verifica*) in a diachronic perspective: A Google Books based research

Jan Radimský (University of South Bohemia, radimsky@ff.jcu.cz) & Pavel Štichauer
(Charles University, Prague, pavel.stichauer@ff.cuni.cz)

In this paper, we address an issue deliberately left aside in a recent article devoted to the competition of two Italian deverbal suffixes *-mento/-zione* over the past two centuries (cf. Radimský & Štichauer 2021: 44), namely the fact that in the overall derivational paradigm of Italian action nouns a (presumably large) number of formations are of the “zero-derived” type *abbandono*, *sosta*, or *verifica*. However, such deverbal nouns represent a heterogeneous group both from the morphological point of view (arising from simple conversion or from a process of suffix clipping) and from the diachronic point of view if taken as direct rivals of suffixed formations with *-mento*, *-zione* or *-aggio*.

In this research, we thus deal with this heterogeneous group of what we (innocuously) call zero-derived (or zero-suffixed) formations within the time span from 1800 up to 2009, based on a large dataset drawn from Google books. We first briefly describe the basic properties of zero-derived nouns turning then to a presentation of the dataset along with some examples. We conclude by discussing, in a very preliminary way, possible outcomes.

Following Thornton (1990; 2004), a distinction is usually posited between deverbal nouns in *-o* (such as *abbandono*, *ripristino*) and feminine nouns in *-a* (such as *sosta*, *firma*, *conquista*, etc.). Furthermore, as Thornton (2004), Montermini & Thornton (2014), and Štichauer (2006; 2018) emphasize, there is a further difference between the deverbal nouns in *-a* which are the result of genuine conversion and the superficially identical deverbal nouns in *-a* such as *bonifica*, *condanna*, *conferma* which in turn are the result of a suffix clipping (see Montermini & Thornton 2014: 187 ff.; Štichauer 2018). The converted nouns in *-a* are usually taken to be no longer productive, whereas those in *-o* (along with clipped formations such as *verifica*) display a modest degree of productivity (cf. Thornton 2004: 517).

In order to obtain a possibly complete list of such formations, we used the latest version of Italian Google N-Grams (“GNs”) published in 2020 (“Version 3”), which is based on an updated version of the underlying Google Books corpus that contains 3 times more text data compared to Version 2 (120*10⁹ tokens vs. 40*10⁹ in Version 2). Besides the *-mento/-zione/-aggio* formations, identified according to a similar procedure as the one described in Radimský & Štichauer 2021: Sect. 3, zero-derived formations have been added to the dataset, obtained as automatically generated outputs in *-o/-a* from all infinitives ending in *-are* attested in GNs). Such an automatic procedure yields a wide range of forms which not only look unrealistic as possible nouns (such as *atterro/atterra* from *atterrare*), but they also mostly correspond to inflectional verbal forms (1st and 3rd pers. sg.). This pervasive homonymy is a major obstacle to any serious corpus-based investigation. Therefore, we applied a contextual constraint on all GN data under investigation and we filtered out only word forms that followed a determiner or a preposition. In this way we have limited the possible homonymy of zero-derived nouns and verbs only to cases where a zero-derived noun is homonymous with a transitive verb and, at the same time, the preceding definite article is homonymous with an accusative pronoun, such as *l'arresto*, analysable either as [*l'*]_{DET}[*arresto*]_N or [*l'*]_{PRON}[*arresto*]_V (experimental queries from synchronic corpora however show that the actual proportion of verbs in these cases tends to be very low). The same filter was also applied (on both suffixed forms and zero-derivations) when the number of occurrences in each time period was calculated, so that the frequency counts for all competing word forms are comparable.

Since the main premise is that deverbal zero-derived formations often represent a (recent) rival noun with respect to an already suffixed counterpart (such as *utilizzo* vs *utilizzazione* or *noleggio* vs *noleggiamento*), we first worked on a maximally exhaustive level

where all five cells are realized with both deverbals in *-o/-a* and with *-mento/-zione* and also *-aggio*. This first level obviously yields just a limited number of cases (the truly interesting ones being just *decollo* and *porto* from *decollare* and *portare*). Much more extended lists of formations were obtained where only the *-mento/-zione* formations are included, and, subsequently, where either *-mento* or *-zione* are considered. Finally, we also worked on a subset of verbs such as those ending with *-eggiare*, *-izzare*, and *-ificare*. The outcome of this gradual procedure is a list of more than 250 zero-derived formations which crucially possess an attested suffixed counterpart. This is a somewhat limited number with respect to what is usually claimed in the literature (cf. Thornton 2004: 516 who also points out that the number must certainly be overestimated). Therefore, a final list will also contain those cases where no suffixed formations are attested, but only the presumably zero-derived ones (this seems to be the case of *sosta*).

Apart from a detailed presentation of the data, our aim is to discuss a group of selected couples whose diachronic behaviour is interesting and whose properties might also reveal something about the productivity of such conversions (or clippings). Therefore, we intend to discuss not only the diachronic evolution of couples such as *rilascio* – *rilasciamento*, *indenizzo* – *indennizzazione*, *noleggio* – *noleggiamento*, *giustifica* – *giustificazione*, but we also set out to consider possible constraints on the processes involved which make it impossible to have, for instance, **ondeggio* – *ondeggiamento* (while the apparently unacceptable *festeggio* is actually attested in the 19th century), and **stratifica* – *stratificazione*, **mummifica* – *mummificazione* (while *giusitifica* has been attested since the 19th cent.). We thus also wish to pursue a qualitative investigation of the large amount of data that we have elaborated on the basis of the Google books corpus.

References

- Montermini, F. – Thornton, A. M. (2014). Su alcune differenze tra deverbali formati per conversione e deverbali formati per riduzione in italiano e francese. In Florence Villoing, Sophie David & Sarah Leroy (eds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 179-203. Paris: Presses universitaires de Paris Ouest.
- Radimský, J. – Štichauer, P. (2021). *Nomina Actionis* in the diachrony of Italian: a paradigm-based model of competition. *Lingue e Linguaggio* XX/1, 33-55.
- Štichauer, P. (2006). Codifica versus codificazione, condanna versus condannazione: riflessioni su alcuni sostantivi italiani nati per troncamento di *-zione*. *Linguistica Pragmatis* XVI/1, 28-35.
- Štichauer, P. (2018). Some remarks on clipping of deverbal nouns in French and Italian. In Olivier Bonami, Gilles Boyé, Hélène Giraud & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 159-172. Berlin: Language Science Press.
- Thornton, A. M. (1990). Vocali tematiche, suffissi zero e “cani senza coda” nella morfologia dell’italiano contemporaneo. In Monica Berretta, Piera Molinelli & Ada Valentini (eds.), *Parallela 4, Morfologia/Morphologie*, 43-52. Tübingen: Gunter Narr Verlag.
- Thornton, A. M. (2004). Femminili in *-a* nati per troncamento di suffisso (Troncamento di *-zione*: il tipo *revoca*). In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 518-520. Tübingen: Niemeyer.

Explorations of Morphotactic and Morphosemantic Transparency in Croatian – the case of suffixation

Ida Raffaelli, Matea Filko, Daniela Katunar, Marina Grubišić, Janja Čulig Suknaić

Faculty of Humanities and Social Sciences, University of Zagreb

ida.raffaelli@ffzg.hr, matea.filko@ffzg.hr, dkatunar@ffzg.hr, mgrubisic@ffzg.hr,

jaculig@ffzg.hr

Morphotactic (MT) and morphosemantic (MS) transparency were not typically at the forefront of investigations of paradigmatic relations structuring the lexicon, the role usually taken by synonymy, semantic fields etc. As to our knowledge, the research of MT and MS transparency was usually conducted on compounds (Mattiello and Dressler 2018) and restricted sets of derivatives, e.g. diminutives (Schwaiger et al. 2018), thus leaving the role of MT and MS transparency in the general lexicon more or less aside. However, MT and MS transparency can be expected to provide an explanation of the way speakers form novel words and shape their mental lexicon, especially in morphologically rich languages, such as Croatian. Nevertheless, research of this type has not been undertaken for Croatian so far, especially within an experimental methodological framework. Therefore, this study is the first attempt of this kind. Morphotactic transparency refers to the identification of morphemic components of motivated words (e.g., *lavica* ‘lioness’ < *lav* ‘lion’ is morphotactically more transparent than *narudžba* ‘order’ < *naručiti* ‘to order’, which has a sound change at the morpheme boundary (see Talamo et al. 2016)), while morphosemantic transparency refers to the identification of a semantic connection between the source word and the derivative, i.e., the realization of the common meaning of a stem and affixes and their role in the formation of the meaning of morphologically complex words (e.g., *kućica* ‘house (diminutive)’ < *kuća* ‘house’ is morphosemantically more transparent than *maslina* ‘olive’ < *maslo* ‘butter oil’ (also see Dressler 2005)).

In order to investigate the way speakers process motivated in relation to unmotivated words, and whether they recognize the differences in MT and MS transparency of motivated words, we constructed a study using the Survey Monkey platform. An online questionnaire was chosen since it provided the greatest accessibility to speakers, especially during the ongoing pandemic. The study consisted of a series of tasks in which participants were required to identify morphemes or segment words with differing levels of MT and MS transparency, and to provide judgments on the relatedness of words based on morphological or semantic criteria. All the motivated words from the study were produced by suffixation and belong to the general lexicon

of Croatian. 101 adult native speakers of Croatian participated in the study, with an average age of 31,98 years.

The study was designed to investigate several research questions, namely a) whether a speaker's educational level has an impact on morphological awareness, i.e. the ability of speakers to understand morphological relations between words and manipulate their morphological structure (cf. Apel 2014), b) do speakers favor semantic or morphological criteria in assessing relatedness between words, c) does MT transparency facilitate suffix and stem identification, and d) does MT or MS transparency facilitate the identification of word-formation relatedness between words.

The results show that there is no difference in the accuracy of identification of morphological units between participants with a moderate (high school) and high (university) educational level. Furthermore, participants tend to prefer synonymy and morphological motivation when assessing relatedness between words (rather than antonymy or hyperonymy/hyponymy). Also, morphotactically transparent words facilitate the identification of suffixes ($\chi^2=234,48$; $df=1$; $p < .01$) and stems ($\chi^2=166.35$; $df=1$; $p < .01$). Finally, MS transparency had a facilitating effect in identifying word-formation relatedness between words, while MT transparency did not show the same effect.

As to our knowledge, these results are the first reported to show the extent of influence that MT and MS transparency have in structuring the Croatian lexicon with regards to other similar relational phenomena and open new avenues for future research.

References:

- Apel, Kenn. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, 34(3), 197–209.
- Dressler, Wolfgang U. (2005). Word-formation in Natural Morphology. In Pavol Štekauer and Rochelle Lieber (eds.) *Handbook of Word-Formation, Studies in Natural Language and Linguistics Theory*. Berlin: Springer, 267–284.
- Mattiello, Elisa and Wolfgang U. Dressler (2018) The Morphological Transparency/Opacity of Novel English Analogical Compounds and Compound Families. *Studia Anglica Posnaniensia* 53(2018), 67–114.
- Schwaiger, Sonja et al. (2017) Scaling morphosemantic transparency/opacity: A corpus-linguistic and acquisitionist study of German diminutives. *Yearbook of the Poznań Linguistic Meeting* 3(2017), 141–153.
- Talamo, Luigi, Chiara Celata and Pier Marco Bertinetto (2016). DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure* 9(1):72–102.

Neutralization of contrast in affixes: the relevance of prosodic organization

Renate Raffelsiefen

Leibniz-Institut für Deutsche Sprache

It is widely accepted that the contrastive potential of roots exceeds that of affixes, expressed in the universally fixed ranking among input-output correspondence constraints in (1) (Root-Affix Faithfulness Meta-constraint (McCarthy & Prince 1995:365)):

(1) IO-CORR_{ROOT} >> IO-CORR_{AFFIX}

Given the ranking in (1), the reduced potential for contrast in affixes is captured by ordering phonological markedness constraints between these IO-CORR-constraints. For instance, the ranking in (2) accounts for the presence of pharyngeals in Arabic roots, as opposed to their absence in affixes (McCarthy & Prince 1995:366). It also illustrates the necessarily lower markedness of the structure of affixes compared to that of roots.

(2) Arabic: IO-CORR_{ROOT} >> *[pharyngeal] >> IO-CORR_{AFFIX}

Based on English and German data I will demonstrate the need to take into account the prosodic organization of affixes to adequately delineate domains of neutralization. Cohering affixes, which integrate into the phonological word of the stem (Dixon 1977), may in fact exhibit IO-CORR-effects not seen in simplexes (see A). Non-cohering affixes, which do not integrate into the phonological word of the stem, support the meta-constraint in (1) (see B).

A. Potential for contrast in cohering affixes

Cohering affixes in English or German consist of two types: affixes lacking a phoneme which could serve as nucleus (e.g. English /θ/ '-th' as in (truθ)_ω 'truth') and suffixes whose initial phoneme forms a nucleus (i.e. vowel-initial suffixes (e.g. English /ɪʃ/ '-ish' as in (ʃɪpɪʃ)_ω). (All transcriptions here are phonemic, those concerning English aim at General American.)

Evidence for enhanced contrastive potential in cohering affixes concerns violations of markedness not seen in simplexes. English simplexes exhibit constraints against voiced /z/ in final consonant clusters (only /s/, (3a) ((maɪz)_ω 'Mars', (brɔnz)_ω 'bronze' are rare exceptions) or against /ɪz/ in final unstressed syllables (only /ɪs/, (3c)). German simplexes do not allow a lax high vowel before word-final {rln} in non-main-stressed syllables (4a) nor tense mid vowels before a word-final consonant in non-main-stressed syllables (4c). Significantly, all of these constraints are violated in the suffixations where suffixes are marked in boldface: plural /z/ in (3b), /ɪz/ denoting disease in (3d), (very productive) /ɪn/ denoting female gender in (4b), /es/ denoting plurals in (4d).

(3)a. (fɪs)_ω 'fierce' b. (fɪ**z**)_ω 'fears' c. ('ɑfɪs)_ω 'office' d. ('ɪeɪ**bɪz**)_ω 'rabies'
(fɔls)_ω 'false' (fɔ**l**z)_ω 'falls' ('pɹɪə**mɪs**)_ω 'promise' ('hɪpɪ**z**)_ω 'herpes'
(sɪns)_ω 'since' (sɪ**n**z)_ω 'sins' ('tɛnɪ**s**)_ω 'tennis' ('kɛɪ**ɪz**)_ω 'caries'

(4)a. ('plætɪn)_ω 'platinum' b. ('ʃɛ**fɪn**)_ω 'fem. boss'
(mu'ɛt**ɪn**)_ω 'muezzin' (ko'le**gɪn**)_ω 'fem. colleague'
('harle**ɪ**kin)_ω 'harlequin' ('mɪtə**ɹɪn**)_ω 'fem. tenant'

c. ('hædɪs)_ω 'Hades' d. ('fætsɪs)_ω 'feces'
('lɪmɪs)_ω 'limes' ('mɛnzɪs)_ω 'menses'
(he'rodɪs)_ω 'Herod' ('ɪndɪ**t**sɪs)_ω 'indices'

Many additional violations of markedness in cohering suffixes concern stress, in particular final main stress as in Engl. (trʌs'tɪ)_ω 'trustee', linked to the suffix /i/ -ee, contrasting with stresslessness of final /i/ in simplexes organized as trochees such as ('kɑfɪ)_ω 'coffee'. All of these effects indicate the ranking in (5), contradicting the meta-ranking stated in (1):

(5) IO-CORR_{COHERING--AFFIX} >> IO-CORR_{ROOT}

The evidence from German and English for the ranking in (5) considered here concerns prosodic organization, in particular unusual stress patterns, the distribution of voiced versus voiceless obstruents, and tense versus lax vowels deviating from regular restrictions pertaining to syllable position. As a counter-argument one might note that there are

phonemes which occur in roots but not in cohering affixes, such as pharyngeals in Arabic (see (2)) or /ð/ in English (cf. /suð/ 'soothe'). Such cases appear to concern consonants which are overall strongly underrepresented in the respective languages, making their absence in the rather small inventory of cohering affixes all but expected. By contrast, the absence of relevant patterns in the simplexes in (3a,c), (4a,c) is hardly coincidental, indicating the ranking of specific markedness constraints between the IO-CORR-constraints in (5).

B. Potential for contrast in non-cohering affixes

Non-cohering affixes in German and English must contain a syllabic vowel, where non-cohering suffixes require an initial onset as well. Neutralization patterns in English non-cohering suffixes are illustrated in (6). Apart from the suffix (hʊd)_Σ 'hood', which forms a foot due to the necessary alignment of /h/ with left foot boundaries in English, all suffixes form single syllables characterized by drastic neutralization. Only schwa (6a) and high front vowels occur (6b,c), the distribution among these determined by the following segmental context (as opposed to the full set of vowels in cohering affixes (e.g. (pɪktʃə'ɪɛsk)_ω 'picturesque', (fɔɪəs'teɪfən)_ω 'forestation', (mə'saʒ)_ω 'massage', (mə'sʊs)_ω 'masseur'). Marked phonemes such as voiced fricatives or affricates are altogether absent in (6), again in contrast to cohering suffixes such as ('dadaɪzəm)_ω 'Dadaism', (ə'biʊsɪv)_ω 'abusive', ('ʃɔɪtɪdʒ)_ω 'shortage'.

- (6)a. (ləs)_σ '-less', (nəs)_σ '-ness', (lət)_σ '-let', (dəm)_σ '-dom', (səm)_σ '-some', (mən)_σ '-man',
 (fəl)_σ '-ful', (wəɪd)_σ '-ward', (mənt)_σ '-ment'
 b. (ʃɪp)_σ '-ship', (lɪŋ)_σ '-ling', (nɪk)_σ '-nik'
 c. (li)_σ '-ly', (ti)_σ '-ty' (e.g. *certainty*), (si)_σ '-cy' (e.g. *chieftaincy*)

Historical evidence supports the systematicness of the neutralization patterns, as suffixes which deviated from those patterns either fossilized (e.g. *wedlock*, *knowledge*) or adjusted in ways violating regular sound change (cf. the closed syllables with high lax vowels in /ʃɪp/ -*ship* or /hʊd/ -*hood*, which according to regular sound change ought to have diphthongs instead). These data then motivate the meta-ranking in (7) (s. Raffelsiefen (in-press)).

(7) IO-CORR_{COHERING-AFFIX} >> IO-CORR_{ROOT} >> IO-CORR_{NON-COHERING-AFFIX}

The recognition of the truly drastic neutralization (cf. (6)), which motivates the ordering of numerous phonological markedness constraints between the last two faithfulness constraints in (7), presupposes a strict definition of the category "affix". Specifically, it is necessary to exclude many items sometimes categorized as affixes based on the properties that they cannot stand alone and recur (e.g. *-phobe* in *Anglophobe*, *agoraphobe*, *-eyed* in *one-eyed*, *blue-eyed*). Here considerations pertaining to meaning are relevant. This will be further explored with German data, where in fact two types of non-cohering affixes must be distinguished to account for neutralization patterns: type 1 affixes, which exhibit very similar restrictions to the English non-cohering suffixes shown in (6), including monosyllabicity and a restriction to high vowels or schwa (e.g. /çən/ -*chen*, /lɪç/ -*lich*, /nis/ -*nis*), and type 2 affixes, which are restricted to back vowels occupying the nucleus (e.g. /bar/ -*bar*, /haft/ -*haft*, /hait/ -*heit*).

References:

- Dixon, R. M. W. 1977. "Some phonological rules in Yidin^y." *Linguistic Inquiry*, 8(1): 1–34.
 McCarthy, John, and Alan Prince. 1995. "Faithfulness and reduplicative identity." *Papers in Optimality Theory* (University of Massachusetts occasional papers, 18): 249–384.
 Raffelsiefen, Renate (in press). Morpho-phonological asymmetries in affixation. In: Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet, and Antonio Fábregas. *The Wiley Blackwell Companion to Morphology*. John Wiley & Sons.

Feast beyond morphology

Kristian Roncero

University of Sheffield / MPI-EVA (Leipzig)

kristian.roncero@gmail.com

In this paper, I argue that the sociolinguistic dimension can play a determining role in the creation and preservation of FEAST (OVERABUNDANCE), based on data from minoritised Slavic languages. Phenomena around inflectional morphology are often explained by pure morphology, morphophonological processes and/or diachrony; especially, OVERABUNDANCE and DEFECTIVITY. For example, Thornton (2019) argues that OVERABUNDANCE is caused by either HETEROCLISIS or SUPPLETION; or Baerman (2011) shows that HOMOPHONY avoidance may play an important role in the development of DEFECTIVITY in many languages. These analyses are plausible and have enlightened our understanding of inflectional morphology. However, as I explain, they fall short explaining other language external causes that can contribute to the creation or preservation of OVERABUNDANCE and DEFECTIVITY.

For this paper, I want to focus on the sociolinguistics-inflectional morphology interface. In Dorian's (2010) study of fisherfolk Scots Gaelic varieties, she showed how certain sociolinguistic settings could favour the existence or preservation of OVERABUNDANCE. Among the most defining factors would be being a small-speech community, economically deprived (and thus, in which everyone belongs to the same social class) and a lack of awareness of a standard form of their variety, which Milroy (1999) had already pointed out. In this paper, I try to replicate the experiment with several minoritised Slavic varieties/languages, namely West Polesian, Podlasian and Rusyn. These languages are in similar settings similar to the ones described by Dorian (2010), with the addendum of a multi-side and more pronounced language contact with other closely related and more standardised Slavic varieties.

For this study, I have used data from my own fieldwork as well as existing corpora from other colleagues. I have selected 30 inflectional paradigms, in which I look at instances of OVERABUNDANCE. After this, I test whether there are more or fewer of these in the minoritized varieties than in their cognates in the more prestigious and standardised varieties. Not surprisingly, minoritised varieties score higher in OVERABUNDANCE, since as Dorian (2010) argued, languages in such settings are far less likely to judge it negatively. Finally, I look at OVERABUNDANCE across varieties, considering the etymology of each form. I show how language contact (i.e. in minoritized varieties borrowing from neighbouring standardised varieties) can also contribute to having more forms per inflectional cell.

In sum, I focus on phenomena around OVERABUNDANCE to test some hypotheses about the correlation between language contact, sociolinguistic status and inflectional morphology. The results point to the fact that future studies of inflectional morphology should be more considerate of language-external factors.

Keywords: *corpora, minoritised, overabundance, Slavic, sociolinguistics*

References

- Baerman, Matthew, 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 7, pp. 1-29.
- Dorian, Nancy C., 2010. *Investigating variation. The effects of social organization and social setting*. New York/Oxford: Oxford University Press.
- Milroy, James, 1999. The consequences of standardization in descriptive linguistics. In: Bex, T. and R.J. Watts eds., *Standard English: The widening debate*. London: Routledge, pp. 16-35.
- Thornton, Anna M., 2019. Overabundance: a Canonical typology. In: Rainer, F., Gardani, F., Luschützky, H. and Dressler, W. U. eds., *Competition in inflection and word-formation*. Dordrecht: Springer, pp. 189-224.

hierarchical aspect inflection in bangla
adrita dutta roy
jawaharlal nehru university
duttaroyadrita@gmail.com

There are four language families in India, namely, Indo-Aryan, Dravidian, Austroasiatic, and Sino-Tibetan, apart from a few other families with fewer population of speakers. Bangla, the endonym for Bengali, is a highly inflectional language hailing from the Indo-Aryan family. It is the third largest language in India and the seventh largest in the world in terms of number of speakers. Bangla is an SOV type language that shows influence from both Tibeto-Burman and Austro-Asiatic languages, not only in lexical items but also grammatical features. This congruence of grammaticality that we see in the language makes for a warranted enquiry.

Unlike most other Indo-Aryan languages, Bangla shows no agreement for gender. Bengali verbs only show person and number agreement along with other verbal features such as tense, aspect, modality, case, and more. In my paper, I have delved into only a minute section of aspect morphology in Bangla, i.e., in connection with personal pronouns. This paper attempts to give a detailed analysis of how aspectual inflection in Bangla is in a hierarchical order, comparing it with studies on other languages as well. While the traditional notion is that there are two separate aspect morphemes in Bangla, my study will show that they are not merely so, they are in fact working at two levels. There is a particular order in which aspectual inflection is executed in this language, where the realization of the perfective is only upon the imperfective. I will also shed light on negation to show how the perfective is indeed treated differently in Bangla, just like the literature suggests. The paper will finally conclude with a formal description of aspect morphology in Bangla.

Reference

- Bhat, D.N. (1999). *The prominence of tense, aspect, and mood*. John Benjamins Publishing.
- Comrie, B. (1976). *Aspect: an introduction to the study of verbal aspect and related problems*. Cambridge University Press.
- Croft, W. (2010). *Verbs: aspect and argument structure*. University of New Mexico.
- Dahl, Ö. (1985). *Tense and aspect systems*. Oxford: Blackwell.
- Das, P.K. (2006). *Grammatical agreement in Hindi-Urdu and its major varieties*. LINCOLM Publishers.

Deo, A. (2012). The imperfective-perfective contrast in Middle Indo-Aryan. *JSAL Vol.5*. Yale University.

Katamba, F. (1993). *Morphology: modern linguistics series*. Palgrave Macmillan.

Leipzig. (2015). *The Leipzig glossing rules: conventions of interlinear morpheme-by-morpheme glosses*. Editors of Linguistics Journals.

Word formation with *nicht* in German

Katinka Rózsa

University of Szeged

rozsa.katinka@yahoo.com

As Joshi (2020: 87) points out, affixal negation is a “relatively underexplored and understudied subdomain of linguistic negation”. It is especially observable in the case of the German particle *nicht* in adjectives such as *nichtdeutsch* ‘not German’ or in nouns such as *Nichtmensch* ‘not human’. Primarily *nicht* is used as a particle in syntactical negation. Its categorisation and its exact function in word-formation is poorly researched and therefore not clear in the literature.

In many handbooks, word-formation with *nicht* is only briefly mentioned. Adjectives and nouns are mentioned as its possible bases, e. g. *nichtchristlich* ‘not Christian’, *Nicht-Akademiker* ‘not graduate’. In many cases, it is referred to as a negative word-formation unit and is compared to syntactic negation and the prefix *un-*. It is described as a prefix (cf. Lenz 1995: 132; Zimmer 1964: 53) or as a compound member (cf. Kühnhold et al. 1978: 176; Motsch 2004: 293). However, the decision is not motivated and seems rather arbitrary in many cases.

The aim of the research presented in this talk is to categorize *nicht* as a word-formation unit and to describe its word-formation pattern(s) based on corpus data. The word formation theory *Pattern-and-Restriction Theory* [PR] (Nolda 2012; 2018) serves as the theoretical framework of the research presented here.

PR itself is cast in terms of the language theory of *Integrational Linguistics* [IL] (Lieb 1983). Within the framework of IL particles such as *nicht* ‘not’, *nur* ‘only’ or *auch* ‘also’ and affixes such as *un-*, *-schaft* or *-lich* are minimal lexical units with an *empty lexical meaning* (cf. Lieb 1983: 171; 2013: 65). The meaning of syntactic particles is expressed via syntactical semantics, and the meaning of affixes via morphological semantics (the semantic means of PR’s word-formation patterns).

The first question that has to be addressed is whether *nicht* can be treated as a prefix in the domain of word-formation. If there are no theoretical or empirical reasons to the contrary, it could remain a syntactic particle with an empty lexical meaning, which is not excluded by the theoretical framework. Another option is to assume that it is a morphological stem with an empty or a non-empty lexical meaning – in these three cases the *nicht*-products would be treated as compounds. However, if it is motivated, *nicht* can be analysed as an affix with an empty lexical meaning.

It is often discussed which criteria can be used to differentiate between roots and affixes (cf. Mugdan 2015: 256) – along these criteria it can be motivated to treat *nicht* as a prefix. One of the most frequent criteria is that roots are free – a root “can occur as a word by itself” (ibid.) – whereas affixes are bound morphemes. From a semantic aspect, the distinction between lexical and grammatical meaning is also a frequent criterion (cf. Mugdan 2015: 257).

However, with respect to *nicht* as a word-formation unit these criteria do not seem to be helpful – it occurs freely as a syntactic particle and has an empty lexical meaning such as affixes. How does it contribute to the meaning of the composition if its lexical meaning is empty and how can it be treated as a prefix if it occurs freely?

To investigate the option of treating *nicht-* as a prefix, word-formations with other particles have to be analysed as well. Could word-formations such as *Nurhausfrau* ‘only housewife’ or *Noch-Student* ‘still student’ be analysed as derivatives?

In our corpus data no formal or semantic change can be observed. Thus, it is not justified to assume that particles such as *nur* or *noch* can be treated as prefixes. This result prompts the question if in these cases the empty lexical meaning permits to talk about composition –

composition members have a contribution to the lexical meaning of the composition; this is not possible with an empty lexical meaning. Can we talk about stems with a nonempty lexical meaning in case of *nicht* or *nur*? This talk is trying to answer this question as well. An option is to assume apart from the free particle *nicht* with an empty lexical meaning a bound stem *nicht* with a nonempty lexical meaning corresponding to the syntactic semantics of the particle that can serve as a basis in compounding.

However, in certain *nicht*-products such as *Nichtbuch* in the following example, there is a change in meaning that can be observed:

(1) Neulich hat er gelesen, dass die Autorin von SHADES OF GREY 95 Millionen Dollar mit ihren [sic!] Nichtbuch verdient hat. (Guido Rohms gestammelte Notizen – accessed: 11.06.2021)

In this context *nicht* has the same meaning as the prefix *un-* in *un*-nouns like *Unmensch* ‘brute’ – it expresses a meaning that can be described as a ‘deviation from a norm’. The property of being a human in *Unmensch* or of being a book in *Nichtbuch* is not negated; rather, it is expressed that they deviate from a norm or a concept related to humans and books. This change in meaning can justify the discussion of *nicht* as a prefix in these particular cases of word-formation as it does not emerge in syntactic negation but in case of affixes such as *un-*.

References

- Joshi, S. (2020). Affixal Negation. In V. Déprez M. T. Espinal (Eds.), *The Oxford Handbook of Negation* (pp. 74–88). Oxford University Press.
- Kühnhold, I., Putzer, O., Wellmann, H. (1978). *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Dritter Hauptteil: Das Adjektiv.* de Gruyter.
- Lenz, B. (1995). *Un-Affigierung. Unrealisierbare Argumente, unausweichliche Fragen, nicht unplausible Antworten.* Narr.
- Lieb, H.-H. (1983). *Integrational Linguistics. Vol. I: General Outline.* Amsterdam; Philadelphia: Benjamins. (= *Current Issues in Linguistic Theory* 17).
- Lieb, H.-H. (2013). *Towards a general theory of word formation: the Process Model.* Berlin: Freie Universität Berlin. (= *Refubium – Repositorium der Freien Universität Berlin*).
- Motsch, W. (2004). *Deutsche Wortbildung in Grundzügen* (2nd ed.). de Gruyter.
- Mugdan, J. (2015). Units of word-formation. In P. O. Müller (Ed.), *Word-Formation. An International Handbook of the Languages of Europe* (Vol. 1, pp. 235–301). de Gruyter Mouton.
- Nolda, A. (2012). *Konversion im Deutschen – Muster und Beschränkungen: mit einem Grundriss einer allgemeinen Theorie der Wortbildung.* Habilitation thesis, Humboldt-Universität zu Berlin, 2013.
- Nolda, A. (2018). "Explaining linguistic facts in a realist theory of word formation". In: Christina Behme, and Martin Neef (eds). *Essays on Linguistic Realism.* Amsterdam; Philadelphia: Benjamins. (= *Studies in Language Companion Series* 196). 203–233.
- Zimmer, K. E. (1964). Affixal Negation in English and Other Languages: An Investigation of Restricted Productivity. *Word* (Supplement), 20.

**The acquisition of the German nominal plural system:
A comparison of the productions of children with typical hearing and with cochlear implants**

Marcel Schlechtweg, Bénédicte Grandon, & Esther Ruigendijk
Carl von Ossietzky Universität Oldenburg, Germany
marcel.schlechtweg@uol.de, benedicte.grandon@uol.de, & esther.ruigendijk@uol.de

The complexity of the German plural system is often presented only in parts, concentrating primarily on orthographic, affixal, and vocalic (umlaut) properties and missing additional, more subtle distinctions between singulars and plurals, as the alternation between voiceless and voiced stem-final consonants in *Zug_{SG}* /tsu:k/, *Züge_{PL}* /'tsy:gə/ ‘train(s)’. Although previous studies suggest that various plural categories are acquired with a different degree of difficulty, by both children with typical hearing (TH) and children with cochlear implant (CI), they either do not apply the more nuanced plural categories or do not address the effects of different plurality markers in combination (see, e.g., Lahaa et al. 2015; Szagun 2001). We aim at examining children’s productions of German plural nouns in a more fine-grained way.

Thirty children with TH (Mean age: 8;4(2;0), Range: 5;1–11;10), and fifteen children with CI (Mean age: 8;1(1;10), Range: 5;8–11;1) participated in an elicitation experiment: they were first presented simultaneously with an audio recording and a picture of the singular form of a word, and then were asked to produce the plural form of this word when presented with this same item depicted twice. We used ten word categories (see Table 1):

Table 1. Categories

Change from singular to plural	Example
1 = No change [NoCh]	<i>Teller_{SG}</i> , <i>Teller_{PL}</i> ‘plate(s)’
2 = Stem umlaut [Uml]	<i>Apfel_{SG}</i> , <i>Äpfel_{PL}</i> ‘apple(s)’
3 = Affixal <i>s</i> [S]	<i>Auto_{SG}</i> , <i>Autos_{PL}</i> ‘car(s)’
4 = Affixal schwa [Schwa]	<i>Tisch_{SG}</i> , <i>Tische_{PL}</i> ‘table(s)’
5 = Affixal <i>n</i> [N]	<i>Katze_{SG}</i> , <i>Katzen_{PL}</i> ‘cat(s)’
6 = Stem umlaut + affixal schwa [Uml+ Schwa]	<i>Ball_{SG}</i> , <i>Bälle_{PL}</i> ‘ball(s)’
7 = Stem umlaut + affixal a schwa [Uml+ ASchwa]	<i>Schloss_{SG}</i> , <i>Schlösser_{PL}</i> ‘castle(s)’
8 = Stem-final consonant voicing + affixal schwa [Voi+ Schwa]	<i>Hund_{SG}</i> , <i>Hunde_{PL}</i> ‘dogs(s)’
9 = Stem-final consonant voicing + affixal a schwa [Voi+ ASchwa]	<i>Schild_{SG}</i> , <i>Schilder_{PL}</i> ‘sign(s)’
10 = Stem umlaut + stem-final consonant voicing + affixal (a) schwa [Uml+Voi+ (A)Schwa]	<i>Zug_{SG}</i> , <i>Züge_{PL}</i> ‘train(s)’

Children with TH responded overall more accurately (TH: 93 %; CI: 84 %; $z = 2.602$, $\text{Pr}(> |z|) = 0.09271^{**}$), as did children with increasing age ($z = 5.481$, $\text{Pr}(> |z|) = 4.23e-08^{***}$) (based on a binomial logistic regression model in R (R Core Team 2021), using the lme4 (Bates et al. 2015) and the lmerTest package (Kuznetsova 2020)). Category had an effect, too, and crossing the Group (TH, CI) with the Category (1 to 10) resulted in the picture in Figure 1 (no significant interaction).

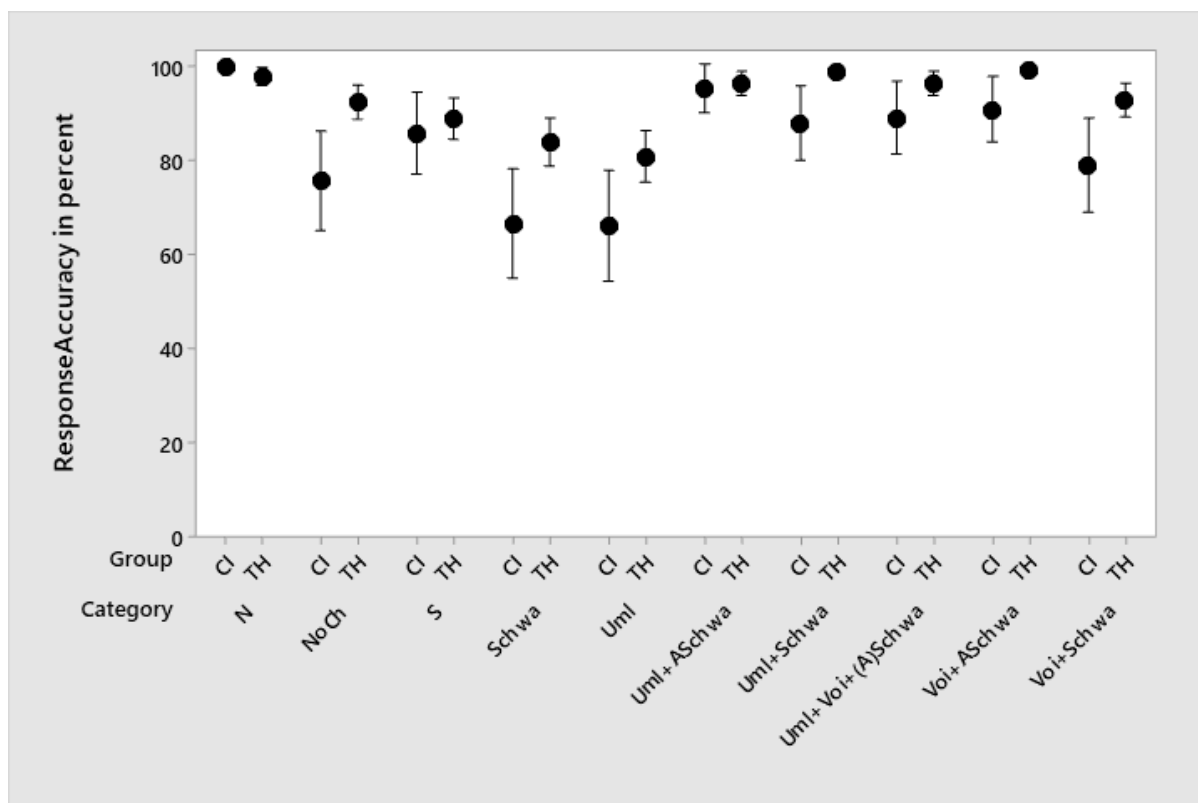


Figure 1. Response accuracy in the two groups and across the plural categories (total of 2,660 responses).

In the talk, we will discuss further variables from our analyses, such as the frequency or length of the items. Further, we will report on the types of errors children made in the individual plural categories to see if the two groups of children differ with regard to the error patterns. We interpret our findings against the background of the interplay of phonetic, phonological, and morphological aspects, which is relevant to understand the challenges for children (with / without hearing impairment) during acquisition.

References

- Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. Version 1.1.27.1. *Journal of Statistical Software* 67 (1), 1–48.
- Kuznetsova, Alexandra, Per Bruun Brockhoff, Rune Haubo Bojesen Christensen & Sofie Pødenphant Jensen. 2020. Package lmerTest. Tests in linear mixed effects models. Version 3.1.3. <https://github.com/runehaubol/lmerTestR/issues>.
- Lahaa, Sabine, Blineder, Michaela & Gillis, Steven. 2015. Noun plural production in preschoolers with early cochlear implantation: An experimental study of Dutch and German. *International Journal of Pediatric Otorhinolaryngology* 79. 561–569.
- R Core Team. 2021. *R: a language and environment for statistical computing*. R version 4.0.5. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>.
- Szagun, Gisela. 2001. Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language* 21. 109–141.

Ubiquity of metaphor in language and cognition – Analytic dichotomies and social-cognitive continuums in understanding linguistic mappings

Zsuzsanna, Schnell

**Department of Cultural Theory and Applied Communication Studies
University of Pécs, Hungary**

schnell.zsuzsanna@pte.hu

Introduction

Metaphor has long been present in the philosophical tradition. It plays a central role in the linguistic typologies of analytic perspectives and is also prevalent in cognitive views and pragmatic approaches aiming to identify the boundaries of literal and non-literal language. Linguistic typologies offer descriptive and clear-cut categories of idiomatic phenomena whereas the social-cognitive views hypothesize a continuum between the domains. We believe the differentiation of irony and metaphor is important as in linguistics they are often seen as different categories, even dichotomies (cf. tropes vs. figures of speech), whereas in the cognitive approach they are seen alongside a continuum of interpretation of first-order and second-order mindreading prerequisites (Happé 1993).

Background

Literal processing contributes to the understanding of some metaphorical constructions, therefore, a morphosyntactic analysis is a cornerstone in the process of deciphering meaning. In this process, however, to decipher the intended meaning mentalization, our ability to understand and predict others' intentions and behavior is of key importance.

Method

The study sets out to identify differences in metaphor and irony as two forms of polysemy and looks behind the interpretative conceptual strategies that operate in their understanding in a developmental perspective. The experimental results of testing preschoolers through their social-cognitive skills shed light on the morphological and interpretative differences of irony and metaphor. Thus we propose a novel typology where conceptual and interpretative features clear the picture of the opaque and often fuzzy categories offered by traditional linguistic approaches.

Results

Our results confirm that irony is apparently easier for preschoolers than metaphor, due to its distinctive features (ironical tone of voice, mocking attitude, normative bias (Wilson 2013) which function as ostensive cues (Csibra 2010) and trigger the shortcut strategy (Györi et al. 2002) based on a heuristics of taking the opposite meaning (Fig. 1.).

Significant differences were found in the performance of the two groups in the irony task, confirming the key role of mentalization (Theory of Mind, ToM) in irony processing. The group with fully fledged mentalization skills was significantly more successful in the irony task than the group not able to rely on mentalization at the required level (Mann Whitney U=115,5; $p < 0,01$). In the irony with surface cue condition no significant difference was found between the performance of the two groups (MWU=150; $p = .104$); Contextual cues made the mental state of speaker explicit, thus, semantic interpretation suffices (Csibra 2010). In the control condition, the No-ToM group was equally successful as the ToM group (MWU=186; $p = .664$) since in these scenarios no human agents participated and thus children did not have to

mentalize about intentions and intended meanings, to decode the intended meaning, a pure semantic interpretation suffices.

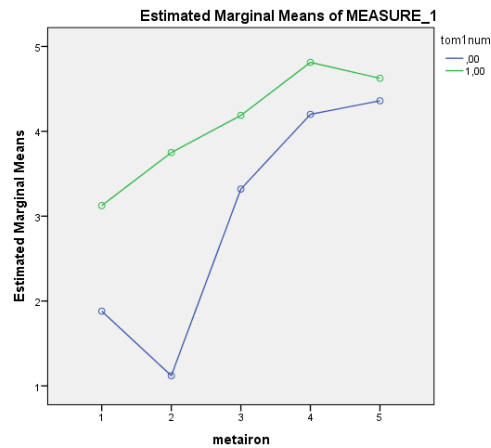


Fig. 1.

Metaphor vs. irony processing in view of mentalization

Lower line (00): NoToM, Upper line (1,00) ToM group

1- Decontextualized metaphor; 2- Contextualized metaphor; 3- Irony, 4- Irony with surface cue; 5- Control task

Results indicate that metaphor is more complex, probably due to the emergent properties involved in their interpretation, where comparison is based on optimal relevance-based traits of similarity, making it a more graded conceptual phenomenon (Colston-Gibbs 2002, Wilson – Carston 2006). Metaphor and irony performance of the ToM and NoToM group differ significantly. Metaphor proved to be more difficult than irony ($p < 0,01$; mean: 2,503; (F) 4,156=44,305 Mixed Way ANOVA). Context was found to facilitate comprehension at several levels. In the group capable of fully fledged mentalization contextual stimuli was decoded with more success; which supports the Cognitive Congruence Hypothesis (Zigler et al. 1966).

Discussion and conclusions

Our ‘irony with linguistic surface cue’ trial confirmed findings of the facilitating effect of contextual cues in interpretation and suggests that mental-state specific context facilitates the deciphering of implicit idiomatic meaning to a significantly greater extent. Hence we suggest a context-aware interpretation that is in place already in preschool years (Schnell 2021, Schnell in press).

The ubiquity of metaphor sheds light on the dichotomic typologies of interpretation, but it also offers a framework for the understanding of its multifaceted interpretational patterns where metarepresentation and metacognition serve as a basis for deciphering meaning. Importantly, metaphor serves social purposes guided by optimal relevance, and is capable of changing frames of reference and thus of understanding (Schnell-Ervas 2021) in a new form and perspective.

On co-plurals: Cross-linguistic evidence of competing pluralization strategies in the domain of nouns

Thomas Stolz & Nataliya Levkovich
(Universität Bremen)
stolz@uni-bremen.de; levkov@uni-bremen.de

In a still undetermined number of languages, at least some nouns allow for the coexistence of two or more plurals (= COPL[ural]s). Within the framework of Canonical Morphology (Corbett 2005), Thornton (2010) describes COPLs of the type SG *braccio* ‘arm’ → PL1 *braccia* ~ PL2 *bracci* in Italian as instances of noncanonicity which mostly correspond to overabundance (i.e. one and the same cell of a given paradigm hosts two or more synonymous occupants). Overabundance has mostly been investigated in the domain of verbal categories (Thornton 2012a–b, 2013) whereas other word-classes are touched upon less systematically (Thornton 2019a–b). We argue that overabundance is by no means a marginal phenomenon in the grammar of nouns either. Similar to the above Italian case, a sizable but still undetermined number of languages of different genealogical affiliation as well as different typological and geographic backgrounds attests to COPLs. We claim that taking stock of the data is of general interest to morphological theories which want to account for patterns of variation. The investigation of COPLs is not the least of importance for the typology and theory of number categories (Corbett 2000).

The talk investigates the recurrence of COPLs across a convenience sample of twenty languages from the Americas, Asia, Europe, and Oceania representing seven different language families (Afro-Asiatic, Austro-Asiatic, Austronesian, Chadic, Indo-European, Uralic, Uto-Aztecan) to show that the phenomenon is widespread and that certain recurrent patterns emerge. In some languages, it affects numerous nouns as e.g. in Maltese SG *werqa* ‘leaf’ → PL1 *werqat* ~ PL2 *werqiet*, SG *giddieb* ‘liar’ → PL1 *giddieba* ~ PL2 *giddiebin*, SG *čekk* ‘cheque’ → PL1 *čekks* ~ PL2 *čekkijiet*, etc. (Borg/Azzopardi-Alexander 1997: 186–187) without any discernible semantic differentiation between the members of a given pair of COPLs. Style and register are often decisive factors for the choice of plural as in the case of Polish masculine nouns ending in *-un* like SG *opiekun* ‘legal guardian’ → PL1 *opiekunowie* ~ PL2 *opiekuni* whose PL1 is judged to be more ceremonial (Laskowski 1972: 42). In many other languages, COPLs may also be used to differentiate meanings of otherwise polysemous or semantically differently specified nouns. Albanian and Romanian give ample evidence of this pattern with cases like SG *bar* ‘weed; herb’ → PL1 *barëra* ‘weeds’ ~ PL2 *barna* ‘medicines’ (Buchholz/Fiedler 1987: 257) and SG *bucată* ‘piece’ → PL1 *bucăți* ‘morsels’ ~ PL2 *bucate* ‘meals’ (Beyrer/Bochmann/Bronsart 1987: 82), respectively.

As to the cases of semantically fully equivalent COPLs, we determine the token frequencies of the competing forms for those languages of our sample which are invested with suitable digitized corpora. In this way, it is possible to show that more often than not the COPLs differ considerably in terms of their attestations in the databases. COPLs which involve semantic differentiation are studied qualitatively, i.e. we test the working hypothesis according to which similar meanings are involved across the sample languages. Where diachronic information is available, we try to determine whether COPLs arise from similar scenarios such as merger of erstwhile distinct singulars or metonymically/ metaphorically motivated semantic change. The impact of contact-induced processes on the development of COPLs is addressed too. In the conclusions, the results of our study are evaluated typologically on the basis of canonical approach (Corbett 2005).

References

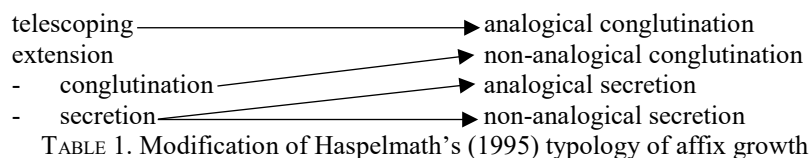
- Beyrer, Arthur, Klaus Bochmann & Siegfried Bronsert. 1987. *Grammatik der rumänischen Sprache der Gegenwart*. Leipzig: Enzyklopädie.
- Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese*. London & New York: Routledge.
- Buchholz, Oda & Wilfried Fiedler. 1987. *Albanische Grammatik*. Leipzig: Enzyklopädie.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic Diversity and Language Theories*, 25–49. Amsterdam & Philadelphia: Benjamins.
- Laskowski, Roman. 1972. *Polnische Grammatik*. Leipzig: Enzyklopädie.
- Thornton, Anna M. 2010. La non canonicità del tipo it. *braccio // braccia / bracci*: sovrabbondanza, diffetività o iperdifferenziazione. *Studi di grammatica italiana* 29/30. 419–477.
- Thornton, Anna M. 2012a. Overabundance in Italian verb morphology and its interactions with other non-canonical phenomena. In Thomas Stolz et al. (eds.), *Irregularity in Morphology (and beyond)*, 251–270. Berlin: Akademie.
- Thornton, Anna M. 2012b. La sovrabbondanza nei paradigmi verbali dell'italiano contemporaneo. In Patricia Bianchi, Nicola De Blasi, Chiara de Caprio & Francesco Montuori (eds.), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali, Atti dell'XI Congresso SILFI*, 445–456. Firenze: Cesati.
- Thornton, Anna M. 2013. Compagni di cella in una gabbia dorata: sull'uso di *vo* vs. *vado* nell'italiano contemporaneo. In Emili Casanova & Cesáreo Calvo Rigua (eds.), *Actes du XXVIe Congrès International de Linguistique et de Philologie Romanes (València, 6–11 de septembre de 2010)*, 447–458. Berlin: de Gruyter.
- Thornton, Anna M. 2019a. Overabundance: a canonical typology. In Franz Rainer et al. (eds.), *Competition in inflection and word-formation*, 223–258. Berlin: Springer.
- Thornton, Anna M. 2019b. Overabundance in morphology. *Oxford Research Encyclopedia of Linguistics* (April 2019). [DOI: 10.1093/acrefore/9780199384655.013.554]

Affix changes due to reanalysis: typology and directionality

Rafał Szeptyński, Institute of Polish Language PAS, rafal.szeptynski@ijp.pan.pl

The aim of this paper is to elaborate on the typology of affix changes (i.e., growth and reduction) resulting from reanalysis, proposed by Haspelmath (1995) in his early article on affix growth. I will argue that the original account can be improved in the following points: the division of affix growth (§1), the general approach as well as some more specific issues related to affix reduction (§2), and the inclusion of an additional type of change (§3). These modifications should contribute to our understanding of the directionality of affix change due to reanalysis (§4).

(1) Haspelmath divides affix growth into two main types, viz. affix telescoping and affix extension (cf. Table 1). While the latter term is not defined explicitly, its scope might be inferred from the definitions of the related subtypes, i.e., affix conglutination and affix secretion. CONGLUTINATION, like TELESCOPING, consists in extending an affix by another one; the difference between the two phenomena lies in the role played in actualization by proportional analogy: telescoping is based on it, conglutination is not. SECRETION differs from both in that it extends an affix with a part of the root and not with another affix; its actualization may or may not be based on proportional analogy. Thus, arguably, there is no feature that links conglutination and secretion together, as subtypes of extension, in a consistent way. Accordingly, I propose to abandon the term *affix extension* as well as *affix telescoping* and tentatively divide the two remaining terms, i.e., *conglutination* and *secretion*, into *analogical* and *non-analogical* subtypes, cf. Table 1:



The distinction between analogical and non-analogical reanalysis requires further discussion. While this issue remains beyond the scope of this paper, the question must be raised (see below on the reasons) whether indisputably non-analogical instances of affix growth are in fact attested, especially spontaneous shifts of morpheme boundaries within secretion (note that conglutination may be treated as inherently analogical in purely formal terms). The answer depends, of course, on the definition of analogy (cf., e.g., Fertig 2013: 46–47).

(2) Haspelmath (1995; 1998, endnote 8) suggests that affix reduction does not occur and thus affix change (i.e., growth) due to reanalysis is irreversible, just as, basically, grammaticalization and any other cyclic language change. As is well known, views on the unidirectionality of grammaticalization have generally become less radical in the course of time (Norde 2009; cf. Haspelmath 2004), so it should not come as a surprise that counterevidence also to the negative statement on affix reduction due to reanalysis may be identified. I will argue that some types of reduction predicted but rejected by Haspelmath (cf. Table 2) are in fact attested, namely those consisting in a false identification of a morpheme already existing in the system, viz. either a shorter affix (DISGLUTINATION) or an extended allomorph of the root (a subtype of INCRETION). Both phenomena seem to require proportional analogy for actualization. Relevant examples from Slavic will be discussed. On the other hand, it is hardly probable (*pace* Langacker 1977) that a spontaneous creation or a shift of a morpheme boundary can occur, yielding two new morphemes, i.e., beside the reduced affix, either another new reduced affix (SPLITTING) or a new extended allomorph of the root (the other subtype of INCRETION). Arguably, the examples of splitting and incretion proposed by Melissaropoulou and Manolessou (2010) require reexamination. I tentatively propose the following terminological changes for affix reduction:

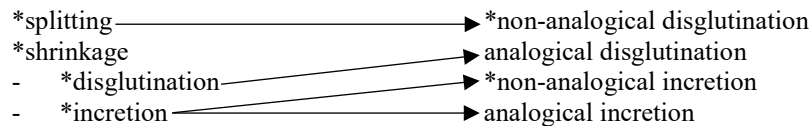


TABLE 2. Modification of Haspelmath's (1995) typology of affix reduction

(3) The typology can be supplemented by yet further types of affix change, most notably resegmentation of two adjacent affixes. Its analogical subtypes involve a preexisting affix which can be either shorter or longer than one of the source affixes *-ab-cd-*, e.g., *-a-* or *-abc-* respectively. The result pertaining to the other affix is growth in the former case (*-bcd-*) and reduction in the latter (*-d-*). A relevant example from Slavic will be discussed. Spontaneous, non-analogical resegmentation, which would result in the growth of one affix and the reduction of the other, is probably subject to the constraint mentioned in §2. Some other (sub)types of affix change, which do not yield brand new affixes, are less central to the typology and will not be discussed.

(4) In conclusion, contrary to what has been claimed in previous literature, the main types of affix growth turn out to have the reductive counterparts attested (apart from the overwhelming difference in frequency). On the other hand, at least some non-analogical subtypes of affix change seem in fact absent. If the non-spontaneity constraint applied exclusively to affix reduction, it could be hypothesized as a common feature of so-called counterdirectional morphological phenomena (cf. Kiparsky 2012 on the possible factors behind “apparent degrammaticalization”). Further research is, however, needed concerning the role of analogy in affix growth.

REFERENCES

- FERTIG, DAVID. 2013. *Analogy and Morphological Change*. Edinburgh: Edinburgh University Press.
- HASPELMATH, MARTIN. 1995. The growth of affixes in morphological reanalysis. *Yearbook of Morphology 1994*, ed. by Geert Booij and Jaap van Marle, 1–29. Dordrecht: Kluwer.
- HASPELMATH, MARTIN. 1998. Does grammaticalization need reanalysis? *Studies in Language* 22(2), 315–351.
- HASPELMATH, MARTIN. 2004. On directionality in language change with particular reference to grammaticalization. *Up and down the Cline – The Nature of Grammaticalization*, ed. by Olga Fischer, Muriel Norde, and Harry Perridon, 17–44. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- KIPARSKY, PAUL. 2012. Grammaticalization as optimization. *Grammatical Change. Origins, Nature, Outcomes*, ed. by Dianne Jonas, John Whitman, and Andrew Garrett, 15–51. Oxford/New York: Oxford University Press.
- LANGACKER, RONALD W. 1977. Syntactic reanalysis. *Mechanisms of Syntactic Change*, ed. by Charles N. Li, 57–139. Austin/London: University of Texas Press.
- MELISSAROPOULOU, DIMITRA, and IO MANOLESSOU. 2010. Theoretical and diachronic aspects of augmentation: Evidence from Greek. *Morphology and Diachrony. On-line Proceedings of the Seventh Mediterranean Morphology Meeting (MMM7)*, ed. by Angela Ralli, Geert Booij, Sergio Scalise, and Athanasios Karasimos, 63–77. [Patras]: University of Patras.
- NORDE, MURIEL. 2009. *Degrammaticalization*. Oxford/New York: Oxford University Press.

The comparative in Ukrainian: suffixes, augments, and root sizes
 Anastasiia Vyshnevskva, KU Leuven, anastasiia.vyshnevskva@kuleuven.be

1. outline I propose that comparative suffixes are composed of two separate heads. I also suggest that there are classes of roots that differ in terms of their structural size within a nanosyntactic approach of spell-out. This explains the morphosyntactic forms of the comparative suffix in Ukrainian across different adjectival classes.

2. the data Ukrainian comparative adjectives are formed by adding a comparative suffix, which follows the root and precedes the agreement marker *-yj*, to the positive degree. This suffix can be either *š*, as in *star-yj–star-š-yj* ‘old’–‘older’, or *-iš*, as in *čyst-yj–čyst-iš-yj* ‘clean’–‘cleaner’. There is a very limited set of adjectives with the apparent *-č* and *-šč* suffixes.

	POS	CMPR	gloss
(1) a.	dzvin- k -yj	dzvin- k -iš-yj	voiced
b.	žorst- ok -yj	žorst- ok -iš-yj	cruel
c.	korot- k -yj	korot-š-yj	short
d.	šyr- ok -yj	šyr-š-yj	wide

I follow Bevzenko (1960) and Plušč (2010) in analysing them as allomorphs of the *-š* suffix. A number of comparative adjectives have the so-called augments (AUG) *-k* or *-ok* in the positive and in the comparative (1a-b). There are other adjectives that have augments in the positive and lose them in the comparative (1c-d).

Theoretically there are six possible patterns for the comparative formation: three possible augment situations (no augment, augment retention, augment drop) for two suffixes (*-š* and

	POS	CMPR	example	gloss	<i>-iš</i>)
(2) a.	root	root-š	star-yj–star-š-yj	old	summarized in (2). The last two patterns are unattested. When an adjective has an augment in the positive form and loses it in the comparative it cannot be followed by the suffix <i>-iš</i> (2e). When it has an augment in the positive form and keeps it in the comparative it cannot be followed by <i>-š</i> (2f).
b.	root	root-iš	čyst-yj–čyst-iš-yj	clean	
c.	root-AUG	root-š	korot- k -yj–korot-š-yj	short	
d.	root-AUG	root-AUG-iš	dzvin- k -yj–dzvin- k -iš-yj	voiced	
e.	root-AUG	root-iš	<i>unattested</i>	<i>unattested</i>	
f.	root-AUG	root-AUG-š	<i>unattested</i>	<i>unattested</i>	

3. DM analysis Bobaljik (2012) proposes that the comparative form of an adjective contains the positive. Under this assumption there is one comparative head that attaches to the positive degree. However, certain Slavic languages have two comparative suffixes where one is a

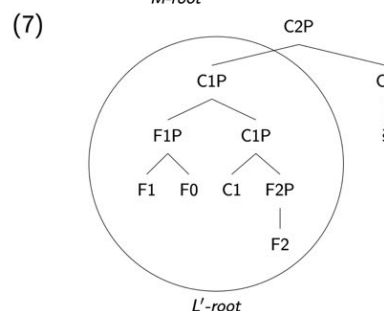
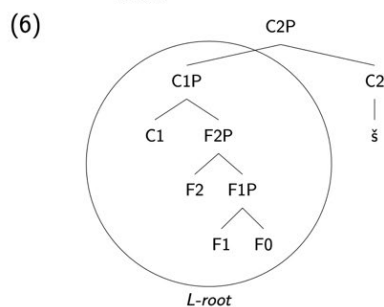
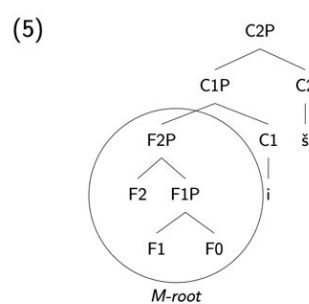
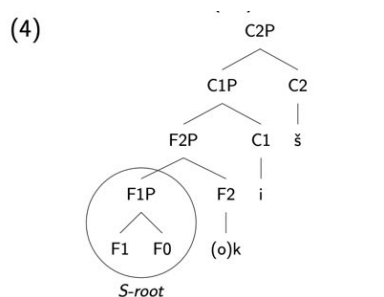
	POS	CMPR <i>-š</i>	CMPR <i>-iš</i>	gloss
(3)	dešev-yj	dešev-š-yj	*dešev-iš-yj	cheap
	važlyv-yj	*važlyv-š-yj	važlyv-iš-yj	important

phonological subset of the other, as *-iš/-š* in Ukrainian. The table in (3) has two bisyllabic adjectival roots, both ending in labiodental [v], but the first one takes the *-š* suffix, while the second one can only take *-iš*. Meanwhile, certain adjectives can take both suffixes *-š* and *-iš*, as *bahat-yj–bahat-š-yj/bahat-iš-yj* ‘rich’–‘richer’. Both of these observations suggest that this alternation is not phonological. Thus, there has to be more to the structure of the comparative instead of one comparative head. Secondly, these languages have augments in the positive that sometimes disappear in the comparative (1c-d), violating the containment relation.

4. Nanosyntactic analysis I propose a nanosyntactic analysis (Starke 2009) for the Ukrainian data. It has two key notions that are relevant for this analysis: submorphemic syntax and phrasal spell-out. The first one is built on the observation that there are more featural distinctions than there are morphemes available, e.g. the Ukrainian agreement marker *-yj* stands for Masculine, Singular, and Nominative. The second one assumes that if multiple heads make up a single

morph, “then it must be possible for spellout to target phrases (XPs) and not just heads” (Baunaz and Lander 2018:16). Caha et al. (2019) propose that the comparative morpheme is represented in syntax not by one but by two functional heads: C1 and C2. I propose that *-iš* is to be decomposed into two morphemes *-i* and *-š*, where *-i* is a spellout of C1 and *-š* is a spellout of C2. I will be also using such labels as F1, F2 (feature 1, 2) for the trees in (4-7). I propose that there are four different root types in Ukrainian comparatives, each differing in the amount of structure that they spell out. Vanden Wyngaerd et al. (2020) divide roots of adjectives into three types: ‘small’, or S-roots, ‘medium’, or M-roots, and ‘extra-large’, or XL-roots. I will be using a similar notation for the root sizes in Ukrainian.

5. deriving the empirical patterns S-roots (4) are ‘small’ as they spell out only F1P. They need an augment to spell out F2P, and both C1P and C2P to spell out the comparative. Such adjectives show the pattern in (2d). It also explains the impossibility of (2f), where an AUG is followed by *-š*: F2P is spelled out by an augment which is then followed by C1P *-i*, and then by C2P *-š*; a C1P has to be lexicalized. M-roots (5) are ‘bigger’ as they spell out more structure, not only F1P, but also F2P. As a result, they do not take augments, but need C1P and C2P in



the comparative; this explains the pattern in (2b). L-roots (6) spell out C1P in addition to F1P and F2P. Such adjectives only need C2P in the comparative; this explains the pattern in (2a). The last type of the roots is what I call L'-roots (7). They spell out the same amount of structure as L-roots in the

comparative, but their positive degree is different. L'-roots need an augment in the positive, but lose it in the comparative, as in (2c). I will show how this pattern can be derived using Movement Containing Trees (inspired by Blix 2021).

6. conclusions The difference in comparative morphology and the distribution of augments is the result of the difference in adjectival root sizes: the bigger the root, the less morphology it needs. The nanosyntactic approach presented above not only helps to explain the mechanism behind allomorphy in Ukrainian, but also explains patterns of augment drop, augment retention, and augment absence.

references Baunaz, L. and Lander, E. 2018. Nanosyntax: the basics. In Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring nanosyntax*, 3–56. Oxford: Oxford University Press. Bevzenko, S.P. 1960. *Istorična morfolohia ukrajn'skoji movy*. ‘The historical morphology of the Ukrainian language’. Uzhhorod: Zakarpats'ke Oblastne Vydavnytstvo. Blix, H. 2021. Phrasal Spellout and Partial Overwrite: On an alternative to backtracking. *Glossa: a journal of general linguistics* 6(1):62. 1–17. DOI: <https://doi.org/10.5334/gjgl.1614>. Bobaljik, J. 2012. *Universals In Comparative Morphology*. Cambridge, MA: MIT Press. Caha, P., et al. 2019. The fine structure of the comparative. *Studia Linguistica* 73(3). 470–521. Plušč, M. 2010. *Hramatyka Ukrajin'skoji Movy*. ‘The Grammar of the Ukrainian Language’. Kyiv: Slovo. Starke, M. 2009. *Nanosyntax: A short primer to a new approach to language*. Nordlyd 36. 1–6. Vanden Wyngaerd, G., et al. 2020. How to be positive. *Glossa: a journal of general linguistics* 5(1): 23.1–34.

The emergence of relational adjectives in the history of German and the role of argument structure

Martina Werner (ACDH-CH, ÖAW & University of Vienna), martina.werner@univie.ac.at

Despite their long-standing description in grammatical theory (seminal Bally 1944: 96f), relational adjectives (RAs, also: pseudo-adjectives, classificatory or classifying adjectives; on terminological issues see e.g. Schlücker 2014) such as *golden plate* or *senatorial leadership* are a frequent phenomenon of many languages and also of present-day German (see 1; cf. e.g. Gunkel/Zifonun 2008, Zifonun 2011, ten Hacken 2019). In contrast to qualitative adjectives (QA), i.e., typical ‘property words’, the emergence, the morphosyntactic development, and basic motivation of RAs has never been comprehensively studied for the history of German, even though isolated synchronic studies are available (cf. e.g. Hotzenköcherle 1968).

(1) German

- a. *morgige Veranstaltung* ‘tomorrow’s event’
- b. *richterliche Entscheidung* ‘judicial decision’
- c. *studentisches Lernen* ‘student’s learning’

Crosslinguistically, RAs are subject to characteristic formal restrictions (Fábregas 2007, Schlücker 2020: 51ff, and the references therein). These include especially the impossibility–glossed here with # – of use as a predicative (2a), as a graduated form (1b), and as the basis for nominalization (1c), where semantically only QAs can be the source of interpretation in each case. In the contexts like in (2), only denominal QA-readings with the form ‘like a N / typical for a N’ are available (see e.g. Holzer 1996, Motsch 2004: 196f; Trost 2006: 15f, Mravlag 2013, among others).

(2) German

- a. #*Die Entscheidung ist richterlich.* # ‘The decision is judicial’
- b. #*richterlichere Entscheidung* # ‘more judicial decision’
- c. #*Richterlichkeit* # ‘judicialness’

Recent literature on RAs especially emphasized the role of nominalizations of the corresponding NP in the sense that there is an underlying relation based on argument structure (AS) between the adjective and noun (as in 1b-c; Gunkel & Zifonun 2008, typologically Fábregas 2007) which provides a more fine-grained differentiation between AS-based (cf. 1b-c) and non-AS-based relational adjectives (cf. 1a). For the history of German, this raises the question of what was first (AS-based or non-AS-based RAs) or if one RA-type could have served as the basis for the development of the other RA-type in the light of determining a potential language change. This is especially since literature makes very heterogeneous claims about the first occurrences of RA-patterns in the history of German (first attestations are dated in Old High German cf. Wilmanns 1896: 476f on *-lich*, on Middle High German, cf. Grimm 1826:179ff, and on Early New High German, cf. Hotzenköcherle 1968).

For answering this question, the corpus-based talk presents for the first time more comprehensive data from Old High German (corpus ReA), Middle High German (corpus ReM), Early New High German (corpus ReF), and from New High German (GermanC corpus). Based on generalizations drawn from the corpus data, the talk shows that the differentiation of AS-based and non-AS-based can indeed be confirmed: Old and Middle High German only possesses traces of non-AS-based RAs, and only token-based examples of AS-based RAs, while patterns of AS-based RAs develop from early New High German. So, different patterns of RAs and the emergence of AS-based RAs are identified including a theoretical discussion of reasons for this language change.

Cited references:

- Bally, C. (1944): *Linguistique générale et linguistique française*. Berne: A. Francke.
- Fábregas ,A. (2007): The internal structure of relational adjectives. In: *Probus* 19/1, 1-36.
- Grimm, J. (1826): *Deutsche Grammatik*. Vol. 2. Göttingen: Dieterichsche Buchhandlung.
- Gunkel,L. & Zifonun,G.(2008): Constraints on relational-adjective noun constructions. A comparative view on English, German and French. In: *Zeitschrift für Anglistik und Amerikanistik* 56.3, 283–302.
- Holzer, P. (1996): *Das Relationsadjektiv in der spanischen und deutschen Gegenwartssprache*. Wilhelmsfeld: Egert.
- Hotzenköcherle, R. (1968): Gegenwartsprobleme im deutschen Adjektivsystem. In: *Neuphilologische Mitteilungen* 69, 1–28.
- Marchis, M. (2010): On the morpho-syntactic properties of relational adjectives in Romanian and Spanish. In: *Bucharest Working Papers in Linguistics* XII.1, 77–92.
- Marchis Moreno, M. (2015): Relational adjectives as interfaces. In: *Studia Linguistica* 69.3, 304–332.
- Motsch, W. (2004): *Deutsche Wortbildung in Grundzügen*. 2nd ed. Berlin: de Gruyter.
- Mravlag, H. (2013): *Relationsadjektive im Deutschen, Französischen und Russischen*. Innsbruck: University Press.
- Schlücker, B. (2014): *Grammatik im Lexikon. Adjektiv-Nomen-Verbindungen im Deutschen und Niederländischen*. (Linguistische Arbeiten 553). Berlin/Boston: de Gruyter.
- Schlücker, B. (2020): Between word-formation and syntax: Cross-linguistic perspectives on an ongoing debate. In: *ZWJW* 2020/2,2, 26-74.
- Ten Hacken, P. (2019). Relational adjectives between syntax and morphology. In: *SKASE Journal of Theoretical Linguistics* 16(1), 77–92.
- Trost, I. (2006): *Das deutsche Adjektiv. Untersuchungen zur Semantik, Komparation, Wortbildung und Syntax*. Hamburg: Buske.
- Wilmanns, W. (1896): *Deutsche Grammatik. Gotisch, Alt-, Mittel- und Neuhochdeutsch*. Vol. 2: *Wortbildung*. Straßburg: Trübner.
- Zifonun, G. (2011): Relationale Adjektive – ein „klassisches“ Muster im europäischen Vergleich. In: *Deutsche Sprache* 9.2, 98–112.

Sources:

GermanC Corpus. URL:

<https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2544> (22/01/14)

ReA *Reference corpus Old German*. URL: <https://korpling.german.hu-berlin.de/annis3/ddd> (2022/01/14)

ReM *Reference corpus Middle High German*. URL: <https://annis.linguistics.rub.de/REM/> (2022/01/14)

ReF *Reference corpus Early New High German*. URL: <https://annis.linguistics.rub.de/REF/> (2022/01/14)

A diachronic approach to the distributional differences between infixes and circumfixes

Tim Zingler
University of Innsbruck
tim.zingler@uibk.ac.at

While almost all introductory textbooks on morphology mention the existence of infixes and circumfixes, statements about their distribution often only refer to the fact that they are cross-linguistically rare compared to prefixes and suffixes. Yet, since circumfixation has barely been explored at all (cf. Harris 2010) and infixation has mostly been investigated from a phonological perspective, even some of the most fundamental properties of these two phenomena are unknown. Among the open issues is the question of whether there are certain semantic domains in which infixes and/or circumfixes are particularly (un)common, or whether they cluster in specific languages or families. In order to address these questions, this presentation will aggregate the results of two studies on the cross-linguistic distribution of infixes and circumfixes. The infix study, Zingler (2022a), covers 61 infixes from 48 languages of the Americas, which belong to 32 families (including isolates). Meanwhile, the circumfix study, Zingler (2022b), generalizes over a database of 83 circumfixes from 59 languages, which belong to 31 families (including isolates), spread across the six global macro-areas. Even though the two affixation types show different distributions, I will argue that the patterns can be understood once their (likely) diachronic trajectories are taken into account.

My study shows that infixes occur in a large proportion of American language families (and seem to be especially frequent in Mesoamerica), but many of the relevant languages only have a single infix, which often alternates with prefixal and/or suffixal “elsewhere” allomorphs. With regard to their semantics, the most salient result of my study is that infixes almost exhaustively distribute into the categories of number, person, voice/valency, and imperfective aspect. Both of these findings differ sharply from the results of my worldwide circumfix study. That is, circumfixes are fairly abundant in Austronesian, Chukotko-Kamchatkan, Kartvelian, and Pano-Tacanan languages, but they are entirely absent from the vast majority of language families. Furthermore, circumfixes encode the whole range of inflectional and derivational functions, with slight preferences for negation and also nominalization.

Overall, the broader semantic range of circumfixes aligns with the idea that a circumfix arises from the reanalysis of a prefix and a suffix as a single affix (e.g., Harris 2010). That is, since prefixes and suffixes show a wide variety of functions cross-linguistically, the circumfixes resulting from their fusion are also unconstrained as far as their semantics is concerned. The fact that the relative majority of circumfixes encode negation might then be a diachronic consequence of the negative cycle (e.g., van Gelderen 2008), such that two separate but mutually reinforcing negators both fuse to the verb. Crucially, the functional patterns of the infixes also find a diachronic explanation. Specifically, their tendency to encode number and imperfective aspects ties in with the fact that many infixes derive from reduplication constructions, which often express these two functions (Yu 2007). According to Yu (2007), the other two pathways that commonly lead to infixes are “entrapment” between a bipartite stem and metathesis. On the assumption that voice/valency markers occur close to the verbal stem (Bybee 1985), metathesis would seem to be the most plausible scenario by which these markers end up inside the stem. By contrast, Bybee (1985) shows that person markers are distant from the stem and loosely bound to it. This suggests that they will be “trapped” once their former stem has fused with another stem and that they become infixes once the two stems can no longer be semantically decomposed. That said, the synchronic distributions of the two affixation types suggest that there must also be differences between the ways in which they arise, and this will hopefully inspire further morphological, typological, and diachronic research. First and foremost, the fact that circumfixes show strong ties to specific language families argues for the idea that they are largely inherited family features, whereas the processes that bring about infixes must happen more frequently, even though they are still sporadic.

References

- Börstell, Carl. 2021. GlottoshinyR. <https://linguistr.shinyapps.io/glottoshinyr/>
- Bybee, Joan. 1985. *Morphology*. Amsterdam: Benjamins.
- Harris, Alice. 2010. Explaining typologically unusual structures: The role of probability. In Wohlgenuth, Jan & Cysouw, Michael (eds.), *Rethinking universals*, 91–103. Berlin: De Gruyter.
- Van Gelderen, Elly. 2008. Negative cycles. *Linguistic Typology* 12. 195–243.
- Yu, Alan. 2007. *A natural history of infixation*. Oxford: Oxford University Press.
- Zingler, Tim. 2022a. Infixation in the Americas: A cross-linguistic survey. *International Journal of American Linguistics* 88, 171-228.
- Zingler, Tim. 2022b. Circumfixation: A semasiological approach. *Word Structure* 15, 55-113.

Maps

Map 1. Geographical distribution of the languages in the infix study.



Map 2. Geographical distribution of the languages in the circumfix study (created with the GlottoshinyR feature; Börstell 2021).



Data: glottolog.org; Libraries: {ggnewscale, mapproj, maps, randomcolor, shiny, tidyverse}