



# UNRAVELING LINGUISTIC PRODUCTIVITY

Insights into usage,  
processing and variability



21-23  
MAY 2024

[languageproductivity.ugent.be](http://languageproductivity.ugent.be)



St Peter's Abbey  
Ghent, Belgium



# **Book of abstracts**

**In alphabetical order  
by the surname of the (first) author**

## **Historical and psycholinguistic perspectives on morphological productivity: A sketch of an integrative approach**

---

***Harald Baayen & Kristian Berg***

*University of Tübingen; University of Bonn*

In our presentation, we approach morphological productivity from two perspectives: the perspective of an ideal individual speaker-listener, and the perspective of an actual historical speaker in their historical context. For developing the first perspective, we make use of a cognitive computational model of the mental lexicon, the discriminative lexicon model (DLM). For developing the second perspective, we study how the intake and output of one prolific writer changes over time.

The DLM model implements mappings between numeric representations for words' forms, and numeric representations for their meanings, using embeddings from distributional semantics. The weights on the connections in networks that implement these mappings are learned from usage data. Model mappings can be evaluated both on how precise their predictions are not only for the data on which they have been trained (mappings as memory systems), but also how well they predict novel, unseen data. Accuracy on held-out data provides a much more fine-grained measure of productivity than corpus-based measures assessing the probability of unseen forms.

However, with the level of detail provided by empirical embeddings, it is becoming increasingly clear to us that it is impossible to predict the full richness of the meanings of even inflected forms, due to world knowledge being absorbed in embeddings along with lexical co-occurrence information (which we believe is a good thing). As a consequence, how for instance a regular English plural noun is understood, depends on both listeners' experience with the language and their domain knowledge. As the domain knowledge of a community of speakers is much richer and diversified than the domain knowledge of any individual speaker, being able to get the gist of novel words is especially important for comprehension. Speakers are potentially more likely to hear a new word as opposed to producing one. If this is true, productivity is more important for comprehension as opposed to production, where our experience with the DLM suggests that much more depends on past experience in comprehension.

In our presentation, we complement this perspective on the individual with a socio-historical perspective. We illustrate the limitations of productivity for production for an individual speaker as member of society by means of a case study of the language use of Thomas Mann, comparing what he is known to have read (community input) with his own writings (personal output). We will also show, using embeddings, how over time more successful derivations undergo shrinkage in semantic space, which we see as a form of semantic regularization necessitated by the communicative needs of language users with different domain knowledge.

Our ideal of a computational model that brings together current knowledge about the mental lexicons of individual speakers, and our knowledge of how innovations spread in communities is a simulation model in which agents, equipped with DLM lexicons that are continuously updated with ongoing learning, and domain-specific knowledge that also changes over time, interact with each other over their lifetimes.

## (Eye-)tracking syntactic productivity: the case of the Spanish inchoative

**Mariia Baltais & Robert Hartsuiker**

*Ghent University*

So far, productivity has mostly been investigated on usage data attested in language corpora (Zeldes, 2012; Perek, 2015). Existing experimental studies on syntactic productivity in comprehension focused on acceptability ratings. For example, instantiations of constructions with a higher type-token ratio (TTR) were rated as more acceptable than those of constructions with a lower TTR (Suttle & Goldberg, 2011; Baltais & Hartsuiker, 2024). However, acceptability ratings are an offline metalinguistic task, so it is still unclear whether and how corpus-based productivity measures match online sentence processing. Here, we used eye-tracking during reading, a sensitive and naturalistic experimental technique, to investigate the relationship between TTR and speakers' reading behavior.

We recorded eye movements of 66 native speakers of Spanish while they read sentences with the Spanish inchoative construction, which expresses the onset of an event (e.g., *romper a llorar*, lit. 'break to cry'). It is strikingly productive: first, a wide range of verbs can fill the inchoative verb slot; second, inchoative verbs vary in their productivity regarding the infinitive slot (Enghels & Van Hulle, 2018; Van Hulle et al., 2024). Importantly, inchoatives with a lower TTR are mainly attested in combination with few highly frequent, conventionalized infinitive types, which makes these infinitives predictable and potentially expected from a reader's perspective. We hypothesized that encountering an infrequent and thus unpredictable infinitive after a less productive inchoative verb would violate readers' expectations and lead to processing difficulties, as opposed to encountering it after a more productive inchoative.

To develop materials, we used a dataset derived from the Spanish Web corpus (Van Hulle & Enghels, 2021). Each of 32 critical sentences appeared in two versions: in the *higher productivity* condition (1a), we employed inchoative verbs with a higher TTR (mean = 0.66, SD = 0.07), while the *lower productivity* condition (1b) featured inchoative verbs with a lower TTR (mean = 0.22, SD = 0.12). We used 8 different inchoatives (4 per condition). In both conditions, the inchoative-infinitive combinations had equally low average token frequency in the corpus dataset, and a pre-test cloze task confirmed that the infinitives were equally unpredictable. Critical sentences were distributed across two presentation lists and mixed with 198 filler sentences.

(1a)	<i>Francisco</i>	<b><i>se metió a</i></b>	<i>hablar</i>	<i>de política</i>	<i>con sus amigos.</i>
(1b)	<i>Francisco</i>	<b><i>rompió a</i></b>	<i>hablar</i>	<i>de política</i>	<i>con sus amigos.</i>
		[pre-target]	[target]	[post-target]	
	Francisco	started to	talk	about politics	with his friends.

Contrary to our expectations, we found no increase of the processing cost in the lower productivity condition neither in the target nor in the post-target region (see Table 1). However, we observed significantly longer total reading times for the lower productivity condition in the pre-target region (corrected  $p$ -value = .001). This effect is unlikely to be caused by uncontrolled differences between the conditions, as we matched the verbs for the most relevant variables known to influence reading measures, including length and lemma frequency in the Spanish Web corpus. Our data suggest that lower syntactic productivity of inchoative verbs does not hinder processing of infrequent infinitive types despite the existence of highly predictable alternatives. However, readers experienced difficulties integrating and interpreting sentences that contained inchoatives with a lower TTR, which shows that lower syntactic productivity does affect the way readers process the construction.



	Pre-target (e.g., <i>se metió a / rompió a</i> )		Target (e.g., <i>hablar</i> )		Post-target (e.g., <i>de política</i> )	
	HP	LP	HP	LP	HP	LP
Skipping (probability)	-	-	0.06 (0.07)	0.06 (0.07)	-	-
First fixation duration (ms)	208 (28)	212 (32)	223 (36)	225 (37)	215 (33)	216 (32)
Gaze duration (ms)	311 (83)	326 (93)	268 (61)	271 (65)	348 (91)	346 (78)
Regression-path duration (ms)	386 (130)	386 (115)	312 (117)	319 (98)	401 (120)	406 (122)
First-pass regressions-out (probability)	-	-	0.10 (0.13)	0.11 (0.14)	0.09 (0.08)	0.09 (0.09)
Regressions-in (probability)	0.15 (0.15)	0.18 (0.17)	-	-	-	-
<b>Total reading time (ms)</b>	<b>409 (156)</b>	<b>439 (173)</b>	326 (102)	336 (106)	434 (142)	449 (137)

*Table 1.* Descriptive statistics of eye movement measures comparing two conditions: higher productivity (HP) and lower productivity (LP). Means and SDs are provided only for those regions for which the corresponding generalized linear mixed-effects models were fitted (with either binomial or Gamma-distribution). After applying the Bonferroni correction for multiple testing, we observed a significant difference between conditions in the total reading time for the pre-target region ( $p = .001$ ; highlighted in bold).

## References

- Baltais, M., & Hartsuiker, R. 2024. *Productivity of Spanish inchoatives: From corpus to acceptability ratings* [Manuscript in preparation]. Department of Experimental Psychology, Ghent University.
- Enghels, R., & Van Hulle, S. 2018. El desarrollo de perífrasis incoativas cuasi-sinónimas: Entre construccionalización y lexicalización [The development of nearly-synonymous inchoative periphrases: Between constructionalization and lexicalization]. *ELUA*, 32, 91–110. <https://doi.org/10.14198/ELUA2018.32.4>
- Perek, F. 2015. *Argument Structure in Usage-Based Construction Grammar: Experimental and corpus-based perspectives*. John Benjamins. <https://doi.org/10.1075/cal.17>
- Suttle, L., & Goldberg, A. 2011. The partial productivity of constructions as induction. *Linguistics*, 49(6), 1237–1269. <https://doi.org/10.1515/ling.2011.035>
- Van Hulle, S., & Enghels, R. 2021. De Spaanse inchoatiefconstructie in beeld: Clusteranalyse als antwoord op het quasi-synonymie vraagstuk [A panorama of inchoative constructions in Spanish: Cluster analysis as an answer to the near-synonymy puzzle]. *Handelingen (KZM)*, 75(1), 277–305. <https://doi.org/10.21825/kzm.87036>
- Van Hulle, S., Enghels, R., & Lauwers, P. 2024. *The many guises of productivity: A case-study of Spanish inchoative constructions* [Manuscript submitted for publication]. Department of Linguistics, Ghent University.
- Zeldes, A. 2012. Productivity in Argument Selection: From Morphology to Syntax. In *Productivity in Argument Selection*. De Gruyter Mouton. <https://doi.org/10.1515/9783110303919>

## Productivity and the paradigm of preposed adverbial clause connection: On the development of *so*-resumption since Early New High German

**Barthe Bloom**

Universität Potsdam

What has been called the resumptive use of *so* (e.g. Catasso, 2021; Axel-Tober, 2023) – illustrated in (1) – could combine with a considerably wider range of adverbial clauses in Early New High German than in Present-Day German. In addition to co-occurrence with concessive and conditionals (1a), *so* collocated with temporal (1b), locative, causal, comparative, and final adverbial clauses (Thim-Mabrey, 1987; Bloom, fthc). It therefore seems intuitive that this construction has become more ‘picky’.

- (1) a. *Ist es sin wille so sterben wir*  
is it his wish so die we  
‘If it is his wish, we will die’ (Pontus, 4rb)
- b. *Do nun der dritte morgen vergieng. so kommt die scho<sup>o</sup>ne iunkfraw in*  
then now the third morning passed so comes the beautiful lady in  
*einem grünen klejde*

Besides surfacing in this *so*-resumption construction, preposed adverbial clauses were also regularly integrated (2a), resumed by a *da* ‘then’ (2b), or juxtaposed to their host (2c).

- (2) a. *vnd als er geessen het rú<sup>o</sup>fft er Lüpoldo*  
and when he eaten had called he Lüpoldo  
‘And when he had eaten, he called Lüpoldo.’ (Fortunatus, 442)
- b. *vnd da alle ding bereyt waren da gieng sie zu<sup>o</sup> dem Peter*  
and when all things ready were then went they to the Peter  
‘And when all things were ready, she went to Peter.’ (Magelone, 670)
- c. *Da der swartz ritter das gewar wart Er greiff yne mit dem helm*  
when the black knight that aware became he grabbed him with the helmet  
‘When the black knight noticed that, he grabbed him by the helmet.’ (Pontus, 45rb)

These constructions constitute a paradigm of constructions that connect preposed adverbial clauses to their host (henceforth: *PACC-paradigm*). The restriction of *so*-resumption to concessive and conditionals did not happen in isolation, but the PACC-paradigm as a whole reconfigures, with integration of adverbial clauses becoming the dominant pattern in Early New High German (e.g., Axel, 2004).

The study considers two issues in light of the development of *so*-resumption: 1) how to analyze the productivity of such highly schematic, syntactic constructions, and 2) how the larger network may be involved in the decrease of a construction’s productivity. This is based on a study of ca. 1500 preposed adverbial clauses extracted from the Roko.UP (Bloom et al., 2023) and supported by frequencies extracted from the DTA (Deutsches Textarchiv).

Measures of productivity have been primarily calculated based on hapax legomena and type frequencies (e.g., Norde and van Goethem, 2014). Both approaches cannot be straightforwardly applied to clausal-constructions due to uncertainty as to what constitutes a type of a hapax. Counting constructs as types or hapaxes is not a fruitful endeavour, as this typically equals or is very close to token frequency. Instead, this depends on the level of abstraction and the dimension of analysis (e.g., propositional meaning, semantic role in the host).

From the perspective that productivity is based on extensibility (Barðdal, 2006), it is, in hindsight, quite puzzling that the most productive construction of the PACC-paradigm – *so*-resumption – becomes increasingly more semantically coherent and less productive. To understand this development, the reconfiguration of the

PACC-paradigm needs to be taken into account. The functional overlap between *da*-resumption and integration in Early New High German attracts non-conditional and non-concessive adverbial clauses. As a consequence, *so*-resumption becomes picky.

### **References**

- Axel, K. (2004). The syntactic integration of preposed adverbial clauses on the German left periphery: A diachronic perspective. In H. Lohnstein and S. Trissler (Eds.), *The Syntax and Semantics of the Left Periphery*, pp. 23–58. Berlin/New York: Mouton de Gruyter.
- Axel-Tober, K. (2023). Adverbial resumption in German from a synchronic and diachronic perspective. In K. De Clercq, L. Haegeman, T. Lohndal, and C. Meklenborg (Eds.), *Adverbial Resumption in Verb Second Languages*, pp. 167–194. Oxford: Oxford University Press.
- Barðdal, J. (2006). Predicting the Productivity of Argument Structure Constructions. *Annual Meeting of the Berkeley Linguistics Society* 32(1), 467–478.
- Bloom, B. (fthc.). Adverbial V3 in Early New High German? Construction(s) with *So*. *Journal of Germanic Linguistics*.
- Bloom, B., M. Reetz, and U. Demske (2023, November). Romankorpus Früh- neuhochdeutsch (Roko.UP). <https://zenodo.org/records/10210706>.
- Catasso, N. (2021). Generalized and specialized adverbial resumption in Middle High German and beyond. *Journal of Historical Syntax* 5(1-13), 1–38.
- Norde, M. and K. van Goethem (2014). Bleaching, productivity and debonding of prefixoids. *Linguisticæ investigationes* 37 (2), 256–274.
- Thim-Mabrey, C. (1987). Adverbiale + *so*. Distribution und Funktion des Korrelats *so*. *Sprachwissenschaft* 12, 180–219.

## Semantic Factoring Lies Behind Productivity: A Distributional Approach to Investigate two Synonymous Quasi-affixes in Mandarin

**Jing Chen**

*The Hong Kong Polytechnic University*

A discernible trend has emerged over the last few decades in the proliferation of new quasi-affixes to coin new words/expressions in Mandarin Chinese. Yet, quasi-affixes, i.e., affix-like morphemes with certain productivity but not fully abstracted lexical meaning [1, 2, 3, 4], have sparked the debate in terms of their varying degrees of productivity and the bleaching semantics, and in turn call for an empirical investigation. As a case study, we statistically examine the productivity of two synonymous quasi-affixes, 团 ('*tuan*') and 族 ('*zu*'), both signifying 'a group of people' from a diachronic view and further investigate semantic factoring in productivity with word embeddings over time [5, 6]. We frame the base words with the quasi-affixes as partially filled constructions [7], namely *X tuan/zu*.

The dataset exploited in this study derives from *People's daily*, one of the most popular newspaper in China, with over 293 million words spanning from the 1950s to 2000s. We first extract all slice 'types' having target characters with a window size of  $\pm 4$  and manually filtered those instances denoting 'ethnicity' for '*zu*' and 'military' for '*tuan*' agreed by two native speakers. The first attestation for '*zu*' was in 1993, but '*tuan*' got continuous attestation across the investigated period. The overall type frequency for '*zu*' and '*tuan*' are 110 and 503, respectively.

By calculating the occurrences in each year, our data reveals a consistent trend in which both *zu* and *tuan* progressively incorporate a greater number of established types, along with introducing new types, into the *X* slots throughout the examined timeframe. However, when evaluating the potential productivity – a measure that gauges the likelihood of creating new terms in unseen data by assigning weights to less frequent types [8, 9, 10] – it becomes evident that *X tuan* demonstrates a more moderate growth pattern within the underlying corpus, while *X zu* are more productive. We further observed two robust positive correlations, between type frequency and hapax occurrences ( $\rho_{zu} = 0.778$ ,  $\rho_{tuan} = 0.94$ ,  $p < 0.05$ ), and between type frequency and semantic similarity ( $\rho \approx 0.7$ ,  $p < 0.05$ ) obtained by averaging and normalizing word embeddings trained by *Tencent AI Lab* [11] for all attestances. It indicated that the increasing number of word types does not lead to more semantic sparsity for both two cases.

However, the statistics suggests different tendencies regarding the correlation between averaged similarity and potential productivity ( $\rho_{zu} = -0.04$ ,  $p > 0.05$ ;  $\rho_{tuan} = 0.48$ ,  $p < 0.05$ ). We speculate that the results might be influenced by the different spans of available data. An alternative hypothesis could be that prototypical semantic groups for *X tuan* consistently attracts new members over time, while *X zu* welcomes a more diverse array of members.

To validate our hypothesis, we first inspected the cosine similarity among *X* after regressing out frequency, and hapax counts from a dynamic view, and it shows that the introduction of new types does not exert a significant influence on cosine similarity for *X tuan* but a more pronounced fluctuations for *X zu*. We then scrutinize semantic preferences of *X* slots from both synchronic and diachronic perspectives using the t-SNE algorithm. Both two perspectives echo with our hypothesis that *X zu* has a more spread-out distribution while *X tuan* got more clearly-cut dense clusters. The diachronic view also provides insight into the evolving trajectory of *X zu/tuan*.

Our case-driven study statistically indicated that semantic factoring lies behind productivity, and the embedding-based method, as an efficient tool, further visualizes the differences among semantic preferences for the slot of base words, with both synchronic and diachronic viewpoints.

**References**

- [1] Y. Chao, *A Grammar of Spoken Chinese*, University of California Press, 1985. URL: <https://books.google.nl/books?id=JlcKngEACAAJ>.
- [2] C.-R. Huang, D. Shi, *A Reference Grammar of Chinese*, Reference Grammars, Cambridge University Press, 2016. doi:10.1017/CB09781139028462.
- [3] C.-R. Huang, S.-K. Hsieh, K.-j. Chen, *Mandarin Chinese Words and Parts of Speech: A Corpus-based Study*, 2017. doi:10.4324/9781315669014.
- [4] C.-R. Huang, K.-j. Chen, F.-y. Chen, W.-j. Wei, L. Chang, 《资讯处理用中文分词规范》设计理念及规范内容 design criteria and content of segmentation standard for chinese information processing, *语言文字应用 (Applied Linguistics)* 1003-5397 1997 (1997) 92–100.
- [5] F. Perek, Recent change in the productivity and schematicity of the *way*-construction: A distributional semantic analysis, *Corpus Linguistics and Linguistic Theory* 14 (2018) 65–97. doi:10.1515/cllt-2016-0014.
- [6] F. Perek, Using distributional semantics to study syntactic productivity in diachrony: A case study, *Linguistics* 54 (2016). doi:10.1515/ling-2015-0043.
- [7] A. Goldberg, *Constructions at Work: The Nature of Generalization in Language*, Oxford University Press, 2005. URL: <https://doi.org/10.1093/acprof:oso/9780199268511.001.0001>. doi:10.1093/acprof:oso/9780199268511.001.0001.
- [8] H. Baayen, On frequency, transparency and productivity, *Yearbook of morphology* 1992 (1993) 181–208.
- [9] R. H. Baayen, 41. Corpus linguistics in morphology: Morphological productivity, in: A. Lüdeling, M. Kytö (Eds.), *Corpus Linguistics*, Mouton de Gruyter, 2008, pp. 899–919. URL: <https://www.degruyter.com/document/doi/10.1515/9783110213881.2.899/html>. doi:10.1515/9783110213881.2.899
- [10] J. Barðdal, *Productivity: Evidence from Case and Argument Structure in Icelandic*, 2008. doi:10.1075/cal.8.
- [11] Y. Song, S. Shi, J. Li, H. Zhang, Directional skip-gram: Explicitly distinguishing left and right context for word embeddings, in: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 175–180. URL: <https://aclanthology.org/N18-2028>. doi:10.18653/v1/N18-2028.

## Partial productivity is the rule rather than the exception

---

***Ewa Dąbrowska***

***Friedrich-Alexander-Universität Erlangen-Nürnberg***

Most linguists take it as self-evident that linguistic generalizations are, or should be, formulated in terms of abstract variables and that they apply “across the board”, that is to say, to all items belonging to the class defined by the variable (unless they are pre-empted by more specific knowledge). Of course, there are exceptions, especially in morphology – but, it is often assumed, behind every exception there is a default rule that applies “elsewhere”.

From its very beginning, construction grammar emphasized the importance of low-level patterns which are subject to various (sometimes highly idiosyncratic) restrictions; however, it also tacitly assumed that the highly abstract patterns that generative linguists tend to focus on are a key design feature of language. In this paper, I argue that while such “big mean rules” exist, they may be a special case: an exception rather than the norm. I summarize a number of studies which suggest that even when a highly general pattern arguably exists at the community level, it is not necessarily represented in individual speakers’ minds. Furthermore, in some cases, languages lack a default pattern even at the community level: in other words, a language might not offer a standard solution for some types of communicative problems (not just in word formation but in all areas of grammar).

I also reflect on why modern linguists have tended to overemphasize general rules and principles. This is partly due to a strong focus on languages spoken by very large numbers of speakers, and, within those languages, on norms that apply to written rather than spoken texts; partly to the fact that language users often rely strongly on preconstructed units when these fit their communicative needs and on a variety of avoidance strategies when no suitable prefab is available; partly to theoretical commitments which affect judgments; and in some cases, to more or less deliberate attempts at mystification.



## Measuring the correlation of productivity rates for English libfixes

---

**Romain Delhem & Caroline Marty**

*Université Paris Nanterre & Sorbonne Université*

In linguistics, productivity is defined and measured in very diverse ways. It is traditionally assessed with type frequency (Bybee 1985), type–token ratio (Norde & Van Goethem 2017), or potential productivity (Baayen & Lieber 1991). Additional measures have been proposed, involving a comparison with the number of lexical bases (relative frequency, Hay 2021); the difference with the number of words listed in a dictionary (Bauer 2001: 156–157); low- frequency types and not just hapaxes (weighted potential productivity, Pustyl'nikov & Schneider-Wiejowski 2009); or the number of neologisms for a given decade (renewal rate, Berg 2020).

We wanted to assess whether those measures were correlated and whether they reflected the same aspects of productivity. To do so, we selected a closed set of data made up of libfixed words. Libfixes are a subtype of suffixes that are created through the ‘liberation’ of a part of word (e.g. *-holic* from *alcoholic*) (Zwicky 2010). These new suffixes only keep some of the semantic features associated to the original lexemes, and discard others. They can be considered as productive, as each of them yields several types (e.g. *workaholic*, *shopaholic* or *chocoholic* for *-holic*).

We identified eight central libfixes of English (*-athon*, *-holic*, *-licious*, *-nomics*, *-rama*, *-scape*, *-tastic*, *-topia*, *-zilla*). Based on results from the Corpus of Historical American English, we calculated the diachronic productivity of each of them between 1820 and 2020. We did the same for four established derivational suffixes of English which are unambiguously productive (*-age*, *-dom*, *-hood*, *-some*) and were therefore used as a baseline.

The first interesting result is that the correlation between the number of hapaxes and the number of neologisms is extremely high (between 0,818 and 0,995 depending on the suffix, with  $\rho = 0,936$  on average). This indicates that the number of hapaxes for a given morphological pattern on a given decade is a very good indication of the number of words created with this pattern. It can therefore be used with confidence when there is no access to diachronic data.

We then calculated the correlation coefficient for each of the measures described earlier and each of the suffixes (both libfixes and established derivational suffixes). Among all those measures, only two pairs exhibit strong correlation, though this could be expected since those pairs of measures use the same variable:

- type–token ratio and weighted potential productivity ( $\rho = 0,935$  on average), which use the number of tokens as the denominator;
- type frequency and renewal rate ( $\rho = 0,686$  on average), which use the number of types.

Other measures are correlated with one another only very weakly, or not at all. This tends to show that those measures do not evaluate the same thing: for instance, type frequency reflects diversity of a pattern while renewal rate reflects its capacity to regenerate itself over time. This means either that productivity is a multi-faceted notion with various components or that it has to be defined more rigorously and associated with only one measure.

### **References**

Baayen, Harald & Lieber, Rochelle. 1991. Productivity and English derivation : a corpus-based study. *Linguistics* 29. 801–843.

- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Berg, Kristian. 2020. Changes in the productivity of word-formation patterns: Some methodological remarks. *Linguistics* 58 (4). 1117–1150.
- Bybee, Joan. 1985. *Morphology: A study of the relation between meaning and form*. Philadelphia: Benjamins.
- Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative?. *Linguistics* 39 (6). 1041–1070.
- Norde, Muriel & Van Goethem, Kristel. 2018. Debonding and clipping of prefixoids in Germanic: Constructionalization or constructional change?. In Booij, Geert (ed.), *The construction of words. Advances in Construction Morphology*, 475–518. Dordrecht: Springer.
- Pustynnikov, Olga & Schneider-Wiejowski, Karina. 2009. Measuring morphological productivity. In Köhler, Reinhard (ed.), *Issues in quantitative linguistics*, 106–125. Lüdenscheid: RAM-Verlag.
- Zwicky, Arnold. Libfixes. *Arnold Zwicky's Blog*, 2010 (visited November 30, 2023) <<https://arnoldzwicky.org/2010/01/23/libfixes/>>

## Through the learning glass: what kind of structure emerges from usage?

---

***Dagmar Divjak***

*University of Birmingham*

“Too many notes, my dear Mozart!”, is the famous objection that Emperor Joseph II voiced against Mozart’s *The Marriage of Figaro*. “Just cut a few, and it will be perfect!”

In this talk I will argue that linguistics is infected by the same malady, albeit in the incarnation of too many meaningful abstractions. The expectation for such abstractions to exist and to be meaningful provides a cognitively unrealistic starting position for usage-based linguists.

Drawing on work done with the Out of our Minds team [outofourminds.bham.ac.uk], I will argue that cutting down on abstractions, and removing the requirement for these abstractions themselves to be meaningful, is needed to meet the requirements of cognitive commitment. I will illustrate that an approach grounded in what can be detected in and learned from input provides valuable insights into the kinds of descriptions that might lay claim to cognitive plausibility.

## ***These words had my teeth on edge: Examining the role of morphological productivity in acceptability judgments of hapax legomena***

***Kim-Kristin Droste***

***Universität Osnabrück***

This paper examines the roles of type frequency, quantitative productivity measures, and semantic variability in acceptability judgments of novel coinages. This case study is concerned with morphological constructions with locative prefixoids, in particular the *up-x* schema.

Usage-based approaches assume that constructional experience is the basis for selectional restrictions and the productivity of open slots in constructions (see, e.g., Bybee & Eddington 2006 for an account of syntactic constructions). Barðdal (2008: 28) argues that the productivity of a construction is determined by the inverse correlation between type frequency and semantic coherence: Schemas with a higher type frequency and lower coherence are relatively open, whereas schemas with a relatively low type frequency and high semantic coherence are extensible through analogy. A similar approach is described by Goldberg (2019: 65–66) under the name of coverage: The higher the variability of attested types and the higher the type frequency, the higher the coverage in hyper-dimensional conceptual space, and the more productive a construction. Consequently, novel coinages that are similar to attested types are judged as more acceptable.

First, these criteria are examined using corpus data. All types of the *up-x* construction were extracted from the BNC (2001). Following Marchand's (1969: 109) classification, these types were assigned to three categories: Prepositional combinations (*upstairs*), prefix verbs (*uphold*), and nouns resembling AN compounds (*uplight*). Collostructional analysis (Stefanowitsch & Gries 2003; Flach 2021) was conducted to identify selectional restrictions of the subschemas. Besides estimating the semantic coherence of these schemas, their productivity was calculated by various quantitative measures (Baayen 2009; Goldberg 2019).

Second, the ELEXIS English Web 2020 corpus (Jakubíček et al. 2022) was searched for hapax legomena of the *up-x* construction. These hapaxes, which are taken as operationalizing novel coinages, were assigned to the three subschemas mentioned above. A selection of these items was added to an acceptability judgment task, which is currently underway as an online survey. Items were embedded in their original context to facilitate the interpretation of the word. On a 5-point Likert scale for sentences like the following, participants were asked to rate how natural or unnatural the word seems to them:

- (1) *Daily tours of the park are operated by the hotel. Those seeking transportation upbay may apply for a boater permit or contact authorized charter boat operators.* (ELEXIS, modified)
- (2) *This sort of crazy behavior of reality persisted far uplife, even unto the time when I left school.* (ELEXIS, modified)

Hapaxes that are part of subschemas with higher measures of productivity are predicted to be judged as more acceptable. In addition, hapaxes that are more similar to previous constructional experience, i.e., attested types in the BNC dataset, should be judged as more acceptable than dissimilar hapaxes. To illustrate, *upbay* should be judged more acceptable due to its similarity to the frequent types *upriver* and *upstream*, whereas *uplife* should be judged less acceptable as it is not similar to the types attested in the BNC. This paper thus aims to examine how speakers' intuitions about novel coinages relate to corpus-based measures of the productivity of the constructions that instantiate them.

### **References**

- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 899–919. Berlin; New York: De Gruyter Mouton.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: John Benjamins.
- Bybee, Joan & David Eddington. 2006. A usage-based approach to Spanish Verbs of “becoming.” *Language* 82(2). 323–355.
- Flach, Susanne. 2021. Collostructions: An R implementation for the family of collostructional methods. <https://sfla.ch/collostructions/>.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Jakubíček, Miloš, Vít Suchomel, Frederico Martelli & Roberto Navigli. 2022. Semantically annotated corpora. ELEXIS - European Lexicographic Infrastructure.
- Marchand, Hans. 1969. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. 2nd edn. München: C. H. Beck'sche Verlagsbuchhandlung.
- Stefanowitsch, Anatol & Stefan Th Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- The British National Corpus*, version 2 (BNC World). 2001. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. <http://www.natcorp.ox.ac.uk/>.

## Diachronic changes and different patterns of (morphological) productivity: the case of causative construction in Persian

**Pegah Faghiri**

*University of Amsterdam*

In causative constructions a new A-argument (the causer) is added to the predicate (Song 2001), and as valency-changing constructions, morphological causatives are frequent cross-linguistically (Song 2013). Two closely related languages, Modern Persian and Modern Tajik, have both inherited a morphological causative construction, with different diachronic paths, resulting in two different patterns of productivity. Middle Persian, spoken in the Iranian territory between 3<sup>rd</sup> – 7<sup>th</sup> centuries, is the ancestor of New Persian (cf. e.g. Lazard 1975). Modern Persian (18<sup>th</sup> – present day) – spoken in Iran known as *Farsi* and in Afghanistan as *Dari* – and Modern Tajik – spoken in Tajikistan – both evolved from Early New Persian and in spite of some significant difference share much of their underlying features (Windfuhr & Perry 2009: 416).

In Middle Persian the suffix *-ēn* added to the (present) stem of intransitive verbs yield transitive predicates: *šaw* ‘go’/*šaw-ēn* ‘cause to go’; *est* ‘stand’/*est-ēn* ‘make stand’. A few double transitive verbs, e.g., *ozan* ‘kill’/*ozan-ēn* ‘cause to kill’ are also found in existing corpora (Skjærvø 2009: 213). It is safe to claim that this was a productive construction, as it was also used as a verbalization strategy, e.g., *āgāh*<sub>Adj</sub> ‘aware’ → *āgāh-ēn*<sub>Verb</sub> ‘make aware’. The construction survived into New Persian and exists in both Modern Persian, as *-ān*, and in Modern Tajik, as *-on*: e.g. *res* ‘arrive’ alternates with ‘make arrive’ – *res-ān* in Persian and *res-on* in Tajik. But while it is fully productive in Tajik, it is restricted in Persian, where it only applies to a small subset of (intransitive and transitive) verbs: *duz-on* (\**duz-ān* in Persian) ‘have something sewn’ is only possible in Tajik (Windfuhr & Perry 2009: 448). Persian uses periphrastic causativation constructions, such as serial verbs, e.g. ‘give sew’.

The Tajik suffix *-on* is, in effect, as productive as *-ēn* was in Middle Persian and similarly used as a verbalization strategy: *xušk*<sub>Adj</sub> ‘dry’/*xušk-on*<sub>V</sub> ‘make dry’. This strategy is not as productive in Persian, where such predicates are preferably expressed as light verb constructions: *xošk*<sub>Adj</sub> ‘dry’ → *xošk kardan* ‘make dry (lit. dry do)’. Also, the (light) verb *kardan* ‘do’ can be causativized in Tajik (*kun* ‘do’/*kun-on* ‘make do’), e.g., *remont kun-on* ‘have (something) repaired’, but not Persian: \**kun-ān* ‘(intended) make do’. Hence, there is a series of (transitive) light verb constructions for which morphological causativation is only available in Tajik.

Whereas lexical restrictions are common in valency-changing construction cross-linguistically, causative constructions seem to be less lexically restricted than their mirror-image counterpart anticausative. Restricted causatives usually apply to patientive monovalent verbs, such as ‘be pregnant’, which may receive a semantic explanation (Van Lier & Messerschmidt 2022). Yet, in the case of Persian it is hard to define the subset based on semantic features: ‘fear’, ‘sleep’, ‘sit’, ‘arrive’, ‘dry’, ‘run’, ‘burn’, ‘dance’, ‘eat’, ‘drink’, ‘wear’, ‘understand’. The restriction hence seems to be purely lexical and may have a diachronic explanation related to usage frequency and/or the prevalence of periphrastic, e.g., light verb and serial verb, constructions in this language.

NB.: Persian verbs have two distinct stems, one used in past tenses and one in present tenses:

*kard*<sub>past</sub>-*an*<sub>infinitive</sub> ‘do’; *kard*<sub>past</sub>-*am*<sub>1SG</sub> ‘I did’ vs. *kon*<sub>present</sub>-*am*<sub>1SG</sub> ‘I do’

### References

Lazard, Gilbert. 1975. The Rise of the New Persian Language. In R.N. Frye (ed.) *Cambridge History of Iran*, 595–632. Cambridge: Cambridge University Press.

Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax*. Harlow: Pearson Education.



Song, Jae Jung. 2013. Nonperiphrastic causative constructions. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Skjærvø, Prods Oktor. 2009. Middle West Iranian. In G. Windfuhr (ed.), *The Iranian Languages*, Routledge Family Series, 196-278. London and New York: Routledge.

van Lier, Eva and Messerschmidt, Maria. "Lexical restrictions on grammatical relations in voice and valency constructions" *STUF - Language Typology and Universals*, vol. 75, no. 1, 2022, pp. 1-20. <https://doi.org/10.1515/stuf-2022-1047>

Windfuhr & Perry 2009. Persian and Tajik. In G. Windfuhr (ed.), *The Iranian Languages*, Routledge Family Series, 416-544. London and New York: Routledge.

## Frequency effects and the representation of light verb construction: a priming study of alternating Persian LVCs

**Pegah Faghiri, Eva van Lier & Monique Flecken**

*University of Amsterdam*

This talk studies the effect of productivity, as part of speakers' knowledge, in mental representation of light verb constructions (LVCs). Noun-verb LVCs, e.g., *bāzi kardan* 'play' (lit. 'game do'), are prevalent in Persian's verbal lexicon. We present findings from a priming experiment focusing on the alternation between two light verbs (LVs): *kardan* 'do' and *zadan* 'hit', both among productive LVs, knowing that 'do' is the LV par excellence. Whereas many LVCs obligatorily occur with one of these LVs, even when sharing the same meaning, e.g., *harf zadan* vs. *sohbat kardan* 'talk', some alternate, e.g. *šāmpu<sub>N</sub> kardan<sub>V</sub>/zadan<sub>V</sub>* 'shampoo<sub>V</sub>'. Alternating LVCs show (lexical) preferences: both *atse kardan* and *atse zadan* ('sneeze do/hit') mean 'sneeze', but there is a clear preference in usage for *kardan*.

We conducted an online priming experiment using 34 alternating LVCs. 126 participants read sentences (primes) then saw pictures (targets) they had to describe with LVCs (1). Following previous research on Germanic dative alternation (e.g. Bernolet & Hartsuiker 2010, Jaeger & Snider 2013, Segaert et al. 2014), we expected to find an effect of lexical preferences of LVCs manipulated in prime sentences (preferred vs. dispreferred LVs), as well as an effect of the overall high frequency of *kardan*. Following our Construction-based view of Persian LVCs, we expected participants to recognize noun-predicate combinations in target items as alternating LVCs and predicted that both LVs would be activated.



- (1) The woman is combing (i.e. [**comb<sub>N</sub>**]-hitting/doing) her hair.

In a Construction-based view, LVCs belong to constructions of different levels, organized hierarchically from lexically specific to general (cf. Samvelian & Faghiri 2014). Considering the event 'sweeping using a broom' (2), while the noun gives the core meaning, the choice of the verb depends on the argument structure, i.e., active (3a) vs. passive (3b) voice. Three options are available for the active voice 'do', 'hit', and 'pull'. On the general level, Sweeping Construction belongs to the active noun-LV Construction (4). Both the frequency of specific LVs occurring with specific Ns in alternating LVCs and the overall frequency of LVs varies. This is schematically illustrated in Figure 1; ('do'/'hit') alternating LVCs are marked as the green intersection. At the lexical level, some nouns are strongly associated with 'do' and with 'hit'. At the general level, they belong to "active LVCs", which is overall more strongly associated with 'do' than with 'hit' (or any other LV).

- (2) Sweeping Construction

(NO)        N1                                *jāru*    V<sub>[+Agent: kardan 'do', zadan 'hit', kešidan 'pull']; -Agent: šodan; 'become', xordan 'collide']</sub>  
 (Agent)    Ground/Location broom  
 'N0 sweeps N1' or 'N1 is swept'

- (3) a.        *Ali*    *in*    *otāq=rā*        *xub*    *jāru*    *kard/zad/kešid*  
               Ali    this    room=DOM    well    broom do/hit/pull.PST.3SG  
               'Ali swept this room well.'
- b.        *in*    *otāq*    *xub*    *jāru*    *na-šod*  
               this    room    well    broom    NEG-become.PST.3SG  
               'This room is not well swept.'

(4) Active Construction

NO (N1) (N2) N V<sub>[+Agent: kardan, zadan, kešidan, etc.]</sub>  
 Agent (Arg1) (Arg2)

We found no priming effects in the LV choice, suggesting that lexical preferences of prime LVCs were overridden by the prevalence of *kardan*. For completion times, we found a speed-up with dispreferred prime LVs, when not reusing the prime LV (see Figure 2). We conclude that when a participant is primed with an alternating LVC, the lexically specific level and general constructional levels are simultaneously accessed. At the constructional level, *kardan* ‘do’ is most strongly activated due to its frequency; at the lexical level, both LVs associated with the target noun are activated, regardless of the prime. Nevertheless, dispreferred prime LVs lead to faster co-activations, without affecting the LV choice, which mainly depends on the lexical preference of the target LVC.

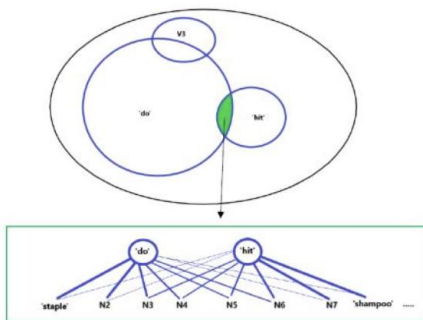


Figure 1. Schematic representation of ‘do’/‘hit’ alternating LVCs.

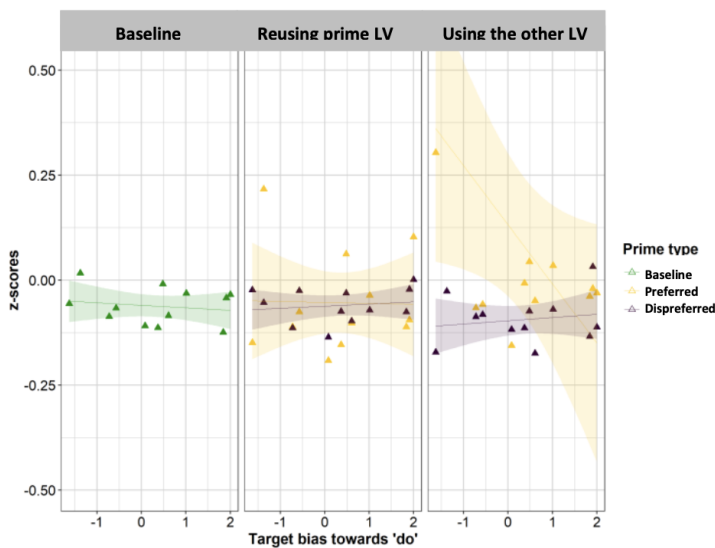


Figure 2. Completion times by prime preference and target bias for different response types.

**References**

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, 114(3), 455–461. <https://doi.org/10.1016/j.cognition.2009.11.005>

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83. <https://doi.org/10.1016/j.cognition.2012.10.013>

Samvelian, P., & Faghiri, P. (2014). Persian complex predicates: How compositional are they? *Semantics-Syntax Interface*, 1(1), 43–74.

Segaert, K., Weber, K., Cladder-Micus, M., & Hagoort, P. (2014). The influence of verb-bound syntactic preferences on the processing of syntactic structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1448–1460. <https://doi.org/10.1037/a0036796>

## Can we reliably measure productivity independently of token frequency? A statistical test based on Zipf-Mandelbrot generated data

**Quentin Feltgen**

**Ghent University**

The empirical study of productivity has benefitted early on from quantitative measures (Baayen & Lieber 1991), such as the type frequency (realized productivity) and the hapax-token ratio (potential productivity). However, one common drawback of these measures is that they heavily depend on sample size, in a way that is not easily sorted out (Baayen 2001 : 29, Gaeta & Ricca 2006, Pankratz et al. 2019). This becomes a real hurdle for the study of diachronic changes of productivity, since token frequency is a major confounding factor to most quantitative measures (Barðdal et al. 2024). Several methods have been offered to alleviate this bias : relying on quantities with a smaller dependency on sample size, such as the hapax-type ratio (Van Wettere 2022), or the entropy, either standard or normalized (Sundquist 2020, Pankratz et al. 2019); taking random samples of a fixed size (Flach 2021, Valdeson 2022); comparing the number of types in a decade with an equally-sized sample drawn from the common pool of tokens (Säily 2011); extrapolating the vocabulary growth curve based on a mathematical modeling (Lüdeling & Evert 2005, Hartmann 2018).

The contribution of this presentation is threefold. First, we argue that the difficulties to derive a proper productivity measure stem from a) the Zipf-Mandelbrot organization over the types (Baayen 2001, Evert 2004, Zeldes 2012), which implements a non-linear relationship between types and tokens; b) the scaling properties of the fitted parameters of the Zipf-Mandelbrot distribution itself, meaning that the parameters of the distribution and of the associated growth curve are themselves sample-size dependent (Baayen 2001, Feltgen 2020); and c) the non-trivial deviations from this distribution that natural language productive exhibit.

Second, we use the Zipf-Mandelbrot distribution to generate synthetic data mimicking diachronic data, with decade-varying sample size. This synthetic data enables us to provide a large-scale statistical test of the main quantitative measures of productivity, in order to assess their sample size dependency. All methods mentioned *infra* have been submitted to that test. We show that several of them achieve sample size independence, and that the obtained productivity profiles are consistent across these methods, hinting at their capability to measure an actually relevant empirical property of constructional patterns.

Third, we apply these well-behaving productivity measures over ten schematic constructions in English, based on data for both the COCA corpus (Davies 2008-) to assess their Zipf-Mandelbrot profile, and from the COHA corpus (Davies 2010) to track diachronically their productivity profile : *a hint of N, extremely ADJ, get used to V, in the midst of N* (Désagulier 2022), *keep Ving, let alone V* (Cappelle et al. 2015), *never to V, on the edge of N, so ADJ a* (Rudnicka 2021), *way too ADJ* (Feltgen 2020). We then provide an overview of their associated diachronic trajectories of productivity, in relationship with their token frequency trajectories.

### References

- Baayen, R. Harald. 2001. *Word frequency distributions*. Dordrecht: Kluwer Academic Publishers.
- Baayen, R. Harald. & Rochelle Lieber. 1991. Productivity and English Derivation: A Corpus-Based Study. *Linguistics* 29. 801–43.
- Barðdal, Jóhanna, Renata Enghels, Quentin Feltgen, Sven Van Hulle & Peter Lauwers. 2024. Productivity in Diachrony. In Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin É Kiss, Joseph Salmons, & Alexandra Simonenko (eds.), *The Wiley Blackwell Companion to Diachronic Linguistics*. Hoboken: John Wiley & Sons, Inc.

- Cappelle, Bert, Edwige Dugas & Vega Tobin. 2015. An Afterthought on Let Alone. *Journal of Pragmatics* 80. 70–85.
- Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Davies, Mark. .2010. *The Corpus of Historical American English (COHA)*. Available online at <https://www.english-corpora.org/coha/>.
- Desagulier, Guillaume. 2022. Changes in the Midst of a Construction Network: A Diachronic Construction Grammar Approach to Complex Prepositions Denoting Internal Location. *Cognitive Linguistics* 33. 339–386.
- Evert, Stefan. 2004. A Simple LNRE Model for Random Character Sequences. In Gérald Purnelle, Cédric Fairon & Anne Dister (eds.), *Proceedings of JADT 2004*, 411–422. Louvain: Presses universitaires de Louvain.
- Feltgen, Quentin. 2020. Diachronic Emergence of Zipf-like Patterns in Construction-Specific Frequency Distributions: A Quantitative Study of the Way Too Construction. *Lexis. Journal in English Lexicology*, 16.
- Flach, Susan. 2021. From Movement into Action to Manner of Causation : Changes in Argument Mapping in the into -Causative. *Linguistics* 59. 247–83.
- Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian Word Formation: A Variable-Corpus Approach. *Linguistics* 44. 57–89.
- Hartmann, Stefan. 2018. Derivational Morphology in Flux: A Case Study of Word-Formation Change in German. *Cognitive Linguistics* 29. 77–119.
- Lüdeling, Anke & Stefan Evert. 2005. The Emergence of Productive Non-Medical-Itis. Corpus Evidence and Qualitative Analysis. In Stephan Kepser & Marga Reis (eds.), *Linguistic Evidence. Empirical, Theoretical, and Computational Perspectives*, 351–70. Berlin; New York: Mouton de Gruyter.
- Pankratz, Elizabeth, Titus von der Malsburg & Shravan Vasishth. 2022. Shannon Entropy Is a More Comprehensive and Principled Morphological Productivity Measure than the Standard Alternatives. <https://doi.org/10.31234/osf.io/vkyug>
- Rudnicka, Karolina. 2021. So-Adj-a Construction as a Case of Obsolescence in Progress. In Svenja Kranich & Tine Breban (eds.), *Lost in Change: Causes and Processes in the Loss of Grammatical Elements and Constructions*, 51–73. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Säily, Tanja. 2011. Variation in Morphological Productivity in the BNC: Sociolinguistic and Methodological Considerations. *Corpus Linguistics and Linguistic Theory*, 7(1). 119–141.
- Säily, Tanja. 2016. Sociolinguistic Variation in Morphological Productivity in Eighteenth-Century English. *Corpus Linguistics and Linguistic Theory* 12. 129–51.
- Sundquist, John D. 2020. Productivity, Richness, and Diversity of Light Verb Constructions in the History of American English. *Journal of Historical Linguistics* 10. 349–88.
- Valdeson, Fredrik. 2022. Lexical Variation in the Double Object Construction in 19th and 20th Century Swedish. In Ida Larsson & Erik M. Petzell (eds.), *Morphosyntactic Change in Late Modern Swedish*, 99-144. Berlin: Language Science Press.
- Van Wettene, Niek. 2022. The hapax/type ratio: An indicator of minimally required sample size in productivity studies?. *International Journal of Corpus Linguistics*, 27(2). 166-190.
- Zeldes, Amir. 2012. *Productivity in Argument Selection*. Berlin; Boston: Walter de Gruyter.



## Readable or legible? The bound form of the postclassical Greek doing affix

**Victoria Fendel**

*University of Oxford*

Support-verb constructions are combinations of a verb and a noun that fill the predicate slot of a sentence, such as *to make a suggestion* in *I made the suggestion that she join us*. The paper is interested in the (support) verb. The support is semi-lexical and semi-grammatical by Boye's (2023) criteria in that it cannot be focused or anaphorically resumed but it can be modified (e.g. *I quickly made the suggestion*). The support verb selects by collocational range for the (predicative) nouns permissible with it such that support-verb-construction families, i.e. the combination of a predicative noun with various support verbs, are not predictable (Kamber 2008). Support verbs thus differ from auxiliary verbs.

The paper is interested in the support verb *ποιέομαι poieomai* 'to do' in post-classical (3<sup>rd</sup> c. BC to 7<sup>th</sup>/8<sup>th</sup> c. AD) literary and documentary Greek and specifically in its bound (i.e. as an affix) rather than its unbound (i.e. as a support verb) form (e.g. *λογοποιέομαι logopoieomai* 'to say' vs *λόγον ποιέομαι logon poieomai* 'to make a comment'). By means of a corpus-based approach (*Thesaurus Linguae Graecae, Duke Database of Documentary Papyri*), the paper addresses three questions: (i) What measures of productivity are available in a corpus language such as post-classical Greek with the additional difficulty of a noisy dataset? (ii) How does the productivity of the semi-lexical and semi-grammatical affix differ in literary as opposed to documentary corpora of post-classical Greek? (iii) Does the synthetic form constitute a competitor to the analytic form such that reallocation takes place (e.g. English *readable* vs *legible*)?

The paper finds that type/token counts even when controlling for hapaxes are not sufficient but that especially the parameters of idiolect and so-called leader words (Burdy 2019; Brucale & Mocciaro 2016; Marini 2010) play a role, that the bound form of the support verb behaves very differently in the literary and documentary sources as regards productivity, and that reallocation is support-verb-construction specific and does not affect all members of the family of constructions.

### References

Boye, Kasper. 2023. Grammaticalization as Conventionalization of Discursively Secondary Status: Deconstructing the Lexical–Grammatical Continuum. *Transactions of the Philological Society* 121(2). 270–292. <https://doi.org/10.1111/1467-968X.12265>.

Brucale, Luisa & Egle Mocciaro. 2016. Composizione verbale in latino: il caso dei verbi in - facio, -fico. In Paolo Poccetti (ed.), *Latinitatis rationes: Descriptive and historical accounts for the Latin language*, 279–297. Berlin ; Boston: Mouton De Gruyter.

Burdy, Philipp. 2019. On the importance of leader words in word formation: The popular transmission of the Latin abstract-forming suffix -io in French. *Word Structure* 12(1). 42–59. <https://doi.org/10.3366/word.2019.0138>.

Kamber, Alain. 2008. *Funktionsverbgefüge-- empirisch: eine korpusbasierte Untersuchung zu den nominalen Prädikaten des Deutschen* (Reihe Germanistische Linguistik). Vol. 281. Tübingen: Max Niemeyer.

Marini, Emanuela. 2010. L'antipassivo in greco antico: ποιείσθαι come verbo supporto in Aristotele. *Journal of Latin Linguistics* 11(1). 147–180. <https://doi.org/10.1515/joll.2010.11.1.147>.

# Productivity of verb-medium colligations expressing change-of-state events - Corpus Linguistics Research

*Martin Godts & Miriam Taverniers*

*Ghent University*

Life involves change, evolution, modification, growth, and decline. Most objects, entities, and concepts, and we and our bodies, are subject to change. Events involving change(s) are referred to as ‘change-of-state’ events. Change-of-state events involve a verb that denotes a change in state as referred to above, and a related noun phrase that can either initiate or undergo the change. For instance, *She opens the door.* vs. *The door opens.* and *The cat broke the vase.* vs. *The vase broke.* The functional, syntactic, lexical, and semantic features of clauses encoding such change-of-state events have received ample attention in earlier research.

In English, especially ambi-transitive verbs (such as *open* and *break* above), named ‘labile’ verbs in Haspelmath (1993), are fit to express changes of state. Levin (1993) classifies ±400 of these change-of-state verbs in 35 lexical groups and shows how they typically function in an (anti)causative alternation. Francis et al. (1996) offer a syntactic classification and propose 70 semantic groups. In Halliday’s (1985, 1994) Systemic Functional Linguistics, a complying noun phrase is called a ‘medium’, as it can function as the subject or the object of the event in an implementation of a lexically construed ergative clausal alignment.

A change-of-state may thus be either expressed in a causative way, i.e., someone or something causes (or more precisely, instigates or triggers) the medium to change, or in a factual way, i.e., the medium inherently, seemingly independently, changes, in a way made possible by the known context or the medium’s properties – see Davidse (1992). McMillion (2006) attests 800+ English labile verbs capable of expressing a change-of-state event, and the corpus linguistics research of Godts & Taverniers (2022) extends this list of labile, ergative verbs with corpus data, including 142 particle verbs. Our present collection includes corpus data for 1088 labile verbs.

In this paper, we first investigate different perspectives to determine the extent, the scope, and the evolution of clausal expressions of change-of-state events. By clustering and mapping the verbs and mediums of (anti)causative alternants, we measure productivity between two alternating constructions – see Perek (2015), Clausner & Croft (1997). Our WordNet-linked research adds both verbal and nominal lexical information to the web corpus user data. We show how verbs and mediums build a verb-medium colligational core, with verbs restricting medium use and mediums restricting verb use, thus licensing the use of the verb-medium core on the basis of functional, syntactic, lexical, and semantic factors. This includes the presence of prepositional objects or adverbials which add to the acceptance and comprehension, and also licensing through specifying and clarifying discourse contexts, as shown in research by (inter alia) Davidse (2011), Rappaport Hovav (2014), Artemis, Anagnostopoulou & Schäfer (2015).

Then, with all related corpus data available, we visualize how productive change-of-state events involving labile verbs are, in terms of type frequency (synchronic corpus linguistics research), and in terms of type evolution (diachronic corpus linguistics research). We follow up on the growing semantic diversity of the semantic classes involved while mapping the syntactic features of the (anti)causative alternants in the user data. A multi-dimensional dynamic network appears, similar to what Diessel (2023) describes. We observe that verb-medium colligations expressing change-of-state events evolve, parallel with the lexical expansion of a language, applying available morphological means to form appropriate change-of-state construals, extending the verb-medium core with as much detail as needed to comprehend, and to license the alternant(s) used.

Whether we express the change-of-state event in a causal or factual way seems not to be dependent on diachronic evolution, but appears to be rather linked to personal style and literary genre.

## **References**

Alexiadou, Artemis, Anagnostopoulou, Elena and Florian Schäfer. 2015. *External Arguments in Transitivity Alternations. A layering Approach*. Oxford: Oxford University Press.

Clausner, Timothy & Croft, William. 1997. Productivity and Schematicity in Metaphors. *Cognitive Sciences* 1 July 1997.

Davidse, Kristin. 1992. Transitivity/ergativity: the Janus-headed grammar of actions and events. In Martin Davies & Louise Ravelli (eds.) *Advances in Systemic Linguistics: Recent Theory and Practice*, 105-135. London/New York: Printer.

\_\_\_\_\_. 2011. Alternations as a heuristic to verb meaning and the semantics of constructions. *Morphosyntactic Alternations in English - Functional and Cognitive Perspectives*. Equinox eBooks Publishing, United Kingdom. p. 11 - 37 Sep 2011 – Version used as retrieved from Researchgate.net on June 30, 2020.

Diessel, Holger. 2023. *The Constructicon: Taxonomies and Networks*. Published online by Cambridge University Press: 20 May 2023. DOI: <https://doi.org/10.1017/9781009327848>

Francis, Gill; Susan Hunston & Elizabeth Manning et al. 1996-2020. *Collins COBUILD Grammar Patterns – Ergative verbs*, chapter 7.

Godts, Martin & Taverniers, Miriam. 2022. Researching verb-medium colligations, poster & abstract, Grammar and Corpora Conference, Ghent University 30-6 till 7-2 2022.

Halliday, M.A.K. 1985,1994. 4th Ed, revised by Christian M.I.M. Matthiessen. *Halliday's introduction to Functional Grammar*. London, New York: Routledge.

Haspelmath, Martin. 1993. More on the typology of inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (eds.) *Causatives and Transitivity*, 88-120. Amsterdam: John Benjamins Publishing Company.

Levin, Beth. 1993. *Verb classes in English and alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

McMillion, Allan. 2006. *Labile Verbs in English. Their Meaning, Behavior and Structure*. PhD dissertation, Stockholm University.

Perek, Florent (2015). *Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives*. Amsterdam: John Benjamins.

Rappaport Hovav, Malka. 2014. Lexical content and context: The causative alternation in English revisited. *Lingua* 141. 8-29.

## **Corpus resources**

<http://www.sketchengine.eu> (enTenTen21 web corpus)

<https://wordnet.princeton.edu> (lexical database)

## Productivity: using a system of learned constructions to express new messages

---

**Adele Goldberg**

*Princeton University*

In order to communicate, we each learn a complex, dynamic system of constructions, a CONSTRUCTIONNET. Mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via competition-driven learning (STATISTICAL PREEMPTION). To express novel messages, we must combine familiar constructions in new ways; such productive combinations have given us *wugs*; *tweeted*; *humble brag*; *Ok, Boomer*, and *is (not) a thing*. Productive combinations of constructions also allow us to talk about a time *three hairstyles ago* or explain that *we napped our way across the Atlantic*. Granted certain caveats, evidence is reviewed that novel combinations are generally judged less acceptable to the extent that there exists a “better” (conventional) way of expressing the same intended message-in-context (e.g., *say to me* > ?*say me*; *succeeded in doing* > ?*succeeded to do*).

## Productivity as a mosaic: Disentangling dimensions of productivity in theory and practice

---

**Stefan Hartmann**

*Heinrich Heine Universität Düsseldorf*

The concept of productivity has been discussed controversially for many decades now (see e.g. Plag 2006, Baayen 2009 for overviews). There is a growing consensus now that productivity has to be regarded as a multidimensional phenomenon, which in turn calls for a combination of multiple measures in operationalizing it. The aim of this talk is to disentangle different dimensions of productivity on a conceptual level from a usage-based and constructionist perspective, and discussing consequences for operationalizing productivity. To do so, I will draw on a case study of the comparative correlative construction (*the x-er the y-er*) in German (*je x-er umso/desto y-er*). In particular, I will argue that we have to distinguish, among other aspects, between (i) the productivity of a linguistic pattern in the language use of an individual vs. a specific population, (ii) different degrees of productivity as suggested by Barðdal (2008), and (iii) semantic variability in addition to purely lexical and morphosyntactic aspects. In my case study, I will discuss how frequency analyses can be combined with distributional-semantic and sociolinguistic approaches to get a fuller picture of the productivity of a constructional pattern.

### References

- Baayen, R. Harald. 2009. Corpus Linguistics in Morphology: Morphological Productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics* (HSK 29.2), 899–919. Berlin, New York: De Gruyter.
- Barðdal, Jóhanna. 2008. Productivity: Evidence from Case and Argument Structure in Icelandic (Constructional Approaches to Language 8). Amsterdam/Philadelphia: John Benjamins.
- Plag, Ingo. 2006. Productivity. In Bas Aarts & April M. S. McMahon (eds.), *The handbook of English linguistics* (Blackwell Handbooks in Linguistics), 537–555. Malden, MA ; Oxford: Blackwell Pub.

## Productivity and creativity: Insights from the 5C model of constructional creativity

---

**Thomas Hoffmann**

*KU Eichstätt-Ingolstadt | Hunan Normal University*

Humans are an incredibly creative species – our minds have evolved to a degree that has enabled us to think original thoughts and come up with novel solutions to a great number of problems. One domain of human cognition that has recently received considerable attention in cognitive linguistics is linguistic creativity. Over the past couple of years, several publications have contributed new and interesting cognitive linguistic findings on the topic. This paper gives an overview of creativity research from the fields of linguistics and psychology and shows how these findings are relevant for cognitive approaches to linguistic creativity. A particular focus will be on Glăveanu's (2013) 5A framework of creativity, which offers the most comprehensive model of creativity and takes into account the various aspects that interact in any creative act (actors, audience, artefacts, actions and affordances). Next, the cognitive framework in which most linguistic creativity research is currently being carried out in, Construction Grammar (Goldberg 2019; Hilpert 2019; Hoffmann 2022) will be discussed. After summarizing the results from constructionist research on verbal creativity, a constructionist model of Glăveanu's framework will be presented – the '5C model' of constructional creativity (constructors, co-constructors, constructs, constructional blending and the constructional network). This model will detail the role of constructional networks (Diessel 2019) in creative acts and it will be argued that Conceptual Blending (Turner 2018, 2020) is the domain-general cognitive process that creates creative (as well as non-creative) constructs.

### References

- Diessel, Holger. 2019. *The Grammar Network: How Linguistic Structure is Shaped by Language Use*. Cambridge: Cambridge University Press.
- Glăveanu, Vlad P. 2013. Rewriting the language of creativity: The five A's framework. *Review of General Psychology* 17(1), 69–81.
- Goldberg, Adele E. 2019. *Explain Me This: Creativity, Competition and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Hilpert, Martin. 2019. *Construction Grammar and its Application to English*. 2nd ed. Edinburgh: Edinburgh University Press.
- Hoffmann, Thomas. 2022. *Construction Grammar: The Structure of English*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Turner, Mark. 2018. The role of creativity in multimodal construction grammar. *Zeitschrift für Anglistik und Amerikanistik* 66(3), 357–370.
- Turner, Mark. 2020. Constructions and creativity. *Cognitive Semiotics* 13(1).



## ***Token by token: A bottom-up approach to NPN productivity***

---

***Lara Höttecke***

***Universität Osnabrück***

Symmetrical noun-preposition-noun (NPN) patterns like *face to face*, *wave after wave* and *bit by bit* have been shown to be syntactically versatile and cross-linguistically pervasive (Jackendoff 2008; Poß 2010; Sommerer & Baumann 2021; Roch et al. 2010). Besides this, NPNs are remarkably diverse and high in overall token frequency, which raises the question as to how we can grasp their apparent productivity. A common way of measuring productivity is calculating a pattern's type-token ratio (Baayen 2009: 901–902). However, it is far from trivial to decide what counts as a type. Previous research on NPN patterns has mainly proposed preposition-specific types such as *N after N* (Jackendoff 2008; Sommerer & Baumann 2021; Sommerer 2022). In these accounts, one type can be assigned multiple meanings, and various types can overlap in their meaning. The types are therefore merely form-based rather than derived from form and function; the calculation of productivity and the modelling of constructional networks based on these types would then somewhat conceal their fuzzy functional boundaries.

In order to address this problem, the talk will present a bottom-up clustering approach to modelling constructional networks which operationalizes cognitive mechanisms such as analogy and categorization via similarity (Langacker 2013: 16–17; Bybee 2010: 7). Instead of exploring individual tokens as instantiations of formally defined types, the procedure takes the entire productive range as a starting point and seeks to motivate types from there. The study is based on a random sample of about 3,000 tokens of NPN patterns retrieved from the British National Corpus (The British National Corpus 2001) that are annotated for syntactic and semantic variables as well as cooccurrence patterns. The model of the constructional network is created in a four-step clustering and visualization procedure that includes 1. the calculation of a distance matrix based on similarities between individual data points (Gower 1971), 2. visualization and dimension reduction via UMAP (McInnes et al. 2018), 3. clustering (Hahsler & Piekenbrock 2022) and 4. Visual reconstruction of the global structure. Results show that NPN patterns, despite their formal commonalities, are considerably more complex than previously assumed. Furthermore, it becomes evident that research on linguistic productivity can profit from the insights yielded by a token-based approach.

### **References**

- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: an international handbook* (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)v.29/2), vol. 2, 899–919. Berlin: Walter de Gruyter.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic* (Constructional Approaches to Language 8). Amsterdam: John Benjamins Publishing Company.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Goldberg, Adele. 2019. *Explain Me This*. Princeton: Princeton University Press.
- Gower, John C. 1971. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* 27(4). 857–871.
- Hahsler, Michael & Matthew Piekenbrock. 2022. *dbscan: Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Related Algorithms*.
- Jackendoff, Ray. 2008. Construction After Construction and its Theoretical Challenges. *Language* 84(1). 8–28.

- Langacker, Ronald W. 2013. *Essentials of Cognitive Grammar*. Oxford, New York: Oxford University Press.
- McInnes, Leland, John Healy & James Melville. 2018. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. <https://arxiv.org/pdf/1802.03426>.
- Poß, Michaela. 2010. *Under Construction: Cognitive and Computational Aspects of Extended Lexical Units*. Utrecht: LOT.
- Roch, Claudia, Katja Keßelmeier & Antje Müller. 2010. Productivity of NPN sequences in German, English, French, and Spanish. 157–162.
- Sommerer, Lotte. 2022. Day to day and night after night: Temporal NPN constructions in Present Day English. In Lotte Sommerer & Evelien Keizer (eds.), *English Noun Phrases from a Functional-Cognitive Perspective: Current Issues* (Studies in language companion series volume 221), 363–394. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Sommerer, Lotte & Andreas Baumann. 2021. Of absent mothers, strong sisters and peculiar daughters: The constructional network of English NPN constructions. *Cognitive Linguistics* 32(1). 97–131.
- The British National Corpus*. 2001. Oxford University Computing Services on behalf of the BNC Consortium.

## Cross-linguistic structural priming and its role in contact-induced language change: Ungrammatical priming in Turkish-German bilinguals

**Hanife Ilen, Gunnar Jacob & Helen Engemann**

*University of Mannheim*

Research investigating the psycholinguistic foundations of contact-induced language change suggests that new structures may enter a target language through cross-linguistic structural priming in bilinguals. Specifically, a structure which initially only exists in the L2 may get primed during bilingual conversations. Priming of this kind may eventually lead to the production of L1 sentences with the primed structure, causing the new structure to gradually become fully productive in the L2 through structural priming (e.g. Kootstra & Şahin, 2018; Kootstra & Muysken, 2019).

While cross-linguistic structural priming effects as such have been demonstrated for a wide variety of different language pairs and structural alternations (van Gompel & Arai, 2018), the vast majority of studies have investigated structures which are grammatical in both languages. However, the claim that cross-linguistic priming is also involved in contact-induced language change implicitly assumes that such effects can emerge even for structures which are ungrammatical in the target language. The present study investigates this assumption by looking at cross-linguistic ungrammatical priming for comparatives between Turkish and German. While Turkish comparatives are formed analytically, by inserting the comparative marker *daha* in front of an adjective, German comparatives are instead synthetic in nature, i.e. formed by attaching the comparative suffix *-er*.

In a self-paced reading experiment with a sample of 36 Turkish-German bilinguals (mean age 22.39, 28 female, 8 male), participants read German target sentences such as (2), which contained grammatically incorrect German analytic comparatives (*mehr interessant\**). These target sentences were preceded by either a Turkish prime sentence such as (1a), which included a Turkish analytic comparative (*daha konforlu*), or an otherwise identical Turkish control prime sentence such as (1b), which instead contained an indicative form (*konforlu*):

- (1a) Jülide büyük bir evde oturmayı **daha konforlu** buluyor. (comparative prime)  
 (1b) Jülide büyük bir evde oturmayı **konforlu** buluyor. (indicative prime)  
 'Jülide finds it (more) comfortable to live in a big house.'

- (2) Hannah findet naturwissenschaftliche Fächer **mehr interessant** als sprachliche Fächer.\*  
 'Hannah finds scientific subjects more interesting than linguistic subjects.'\*

As shown in Table 1, the results for the critical segment containing the ungrammatical German comparative (*mehr interessant*) showed significantly faster reading times for targets following Turkish comparative primes than following otherwise identical indicative primes. Priming aside, the results also revealed a significant adaptation effect for the ungrammatical German comparative, with reading times for the critical segment gradually getting faster during the experiment.

In an additional grammaticality-judgment study in which the German target sentences from the self-paced reading experiment were rated on a 7-point scale, a new sample of 35 Turkish-German bilinguals rated the sentences as significantly more grammatical (mean rating 2.45) than a sample of 30 monolingually raised German speakers (mean rating 1.77).

In sum, these results suggest that, at least in comprehension, cross-linguistic structural priming can even occur when the respective structure is ungrammatical in the target language. Our findings are consistent with

the idea that new grammatical structures may enter a language through cross-linguistic ungrammatical priming in bilingual speakers.

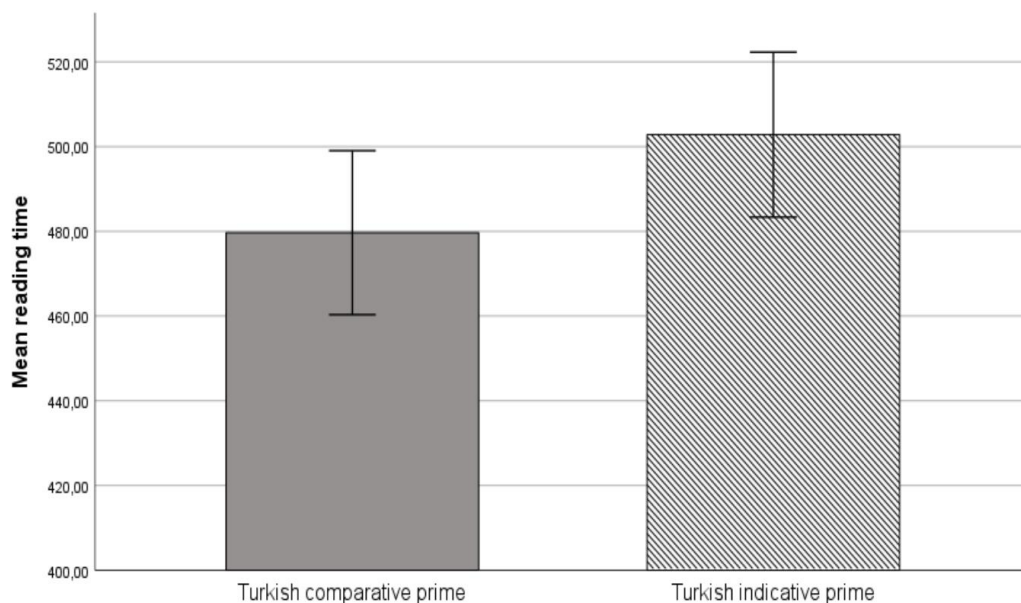


Figure 1. Mean reading times for the critical comparative segment (e.g. ‘mehr interessant’) in German target sentences by condition.

	Estimate	Std. Error	df	t value	Pr(> t )	
(Intercept)	6.123	0.041	44.496	150.351	< .001	***
prime type	0.055	0.022	738.554	2.474	.014	*
item position	-0.015	0.004	34.410	-4.219	< .001	***
prime type x item position	-0.001	0.003	736.981	-0.310	.757	

Formula:  $\log RT(\text{comparative}) \sim \text{prime type} * \text{item position} + (1 + \text{item position} | \text{participantID}) + (1 | \text{itemID})$

Table 1. Linear mixed-effects model predicting log-transformed reading times for the critical segment.

## References

- Kootstra, G. J., & Muysken, P. (2019). Structural priming, levels of awareness, and agency in contact-induced language change. *Languages*, 4(3), 65.
- Kootstra, G. J., & Şahin, H. (2018). Crosslinguistic structural priming as a mechanism of contact-induced language change: Evidence from Papiamentto-Dutch bilinguals in Aruba and the Netherlands. *Language*, 94(4), 902-930.
- Van Gompel, R. P., & Arai, M. (2018). Structural priming in bilinguals. *Bilingualism: Language and Cognition*, 21(3), 448-455.

## Alternating predicates in Romanian: priming at thematic role level

**Mihaela Iliaoaia**

**Ghent University**

This presentation deals with the Romanian construction in (1a–b) and (2a–b), in which certain predicates select for two alternating argument structures Dat-Nom (1a) vs. Nom-Dat (1b), or Acc-Nom (2a) vs. Nom-Acc (2b):

(1a) Fetei	îi	place	istoria	(1b) Istoria	îi	place	fetei
girl.DAT	her.DAT	pleases	history.the.(NOM)	history.the.(NOM)	her.DAT	pleases	girl.DAT
'The girl likes history'				'History pleases the girl'			
(2a) Pe Maria	o	enervează	indiferența	(2b) Indiferența	o	enervează	pe Maria
PE Maria.ACC	her.ACC	irritates	indifference.the.(NOM)	indifference.the.(NOM)	her.ACC	irritates	PE Maria.ACC
'Maria gets irritated by indifference'				'Indifference irritates Maria'			

In the literature, this type of construction is called an *alternating predicate construction*. Recent studies show that alternating predicates are found in several old and modern Indo-European (IE) languages (Barðdal et al. 2019; Barðdal 2023: Ch. 3; Somers & Barðdal 2022). The alternating predicate constructions are a type of oblique subject construction that selects for two distinct and opposed argument structures: DAT-NOM or ACC-NOM vs. NOM-DAT or NOM-ACC (Barðdal et al. 2019), of which DAT-NOM and ACC-NOM are often analyzed as cases of topicalization (Dobrovie-Sorin 1987). Remarkably, when the order of the arguments is DAT-NOM or ACC-NOM, the dative or the accusative takes on the subject role, while when the order is NOM-DAT or NOM-ACC, the nominative behaves as a subject, hence refuting the topicalization analysis. Note that each element preserves its thematic role when the word order is inverted. Among IE languages, Romanian shows a high number of accusative subject constructions (Van Peteghem 2016) and a fair number of dative subject constructions, although not all of them may be of the alternating type.

By means of a series of psycholinguistic experiments, I intend to unveil what triggers the choice for one argument structure or the other in the speaker's mind. From the inventory of possible factors influencing the speaker's choice, priming and verb bias are under scrutiny here. The concept of *verb bias* refers to the fact that some verbs may have a strong preference for one of the alternating argument structures they select for (Bernolet et al. 2014). As for *structural priming*, this refers to the tendency of speakers to reuse structures from the immediately preceding (unrelated) discourse (Bock 1986; Scheepers et al. 2017).

The study discussed in this presentation consists of a psycholinguistic experiment meant to provide an answer to the following research questions: (i) is priming driven by thematic roles as argued by Pappert & Pechmann (2014) or by event structure as suggested by Ziegler et al. (2018)? (ii) is there any correlation between cross structural priming and verb bias?

In order to answer these research questions, I designed a production psycholinguistic experiment by using a sentence generation task (programmed in PCIBex Farm) and presented it to a sample of 48 Romanian native speakers recruited via Prolific. While a previous experiment has shown that structural priming influences the speaker's choice of a specific argument structure over the other when the priming and the target constructions are of the same type, the present experiment aims at verifying whether priming works also across different argument structure constructions (e.g. can a DAT-NOM argument structure be obtained when an ACC-NOM argument structure has been provided as priming?). The outcomes of this experiment would bring more clarity to the discussion on whether priming is triggered by thematic roles – case in which a DAT-NOM argument structure is expected to be obtained when the participant is primed with an ACC-NOM structure (Pappert & Pechmann 2014), or by event structure – case in which no effect of priming is expected (Ziegler et al. 2018). The preliminary results confirm the importance of verb bias, since speakers tend to reuse the primed argument structure from the preceding context to a lesser extent in contexts where verb bias is present.

### References

- Barðdal, Jóhanna. 2023. *Oblique Subjects in Germanic: Their Status, History and Reconstruction*. To appear with Mouton de Gruyter.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The Alternating Predicate Puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1), 107–170.
- Bernolet, Sarah, Timothy Colleman, & Robert. J. Hartsuiker. 2014. The “sense boost” to dative priming: Evidence for sense-specific verb-structure links. *Journal of Memory and Language* 76, 113–126.
- Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18, 355–387.
- Dobrovie-Sorin, Carmen. 1987. A propos de la structure du groupe nominal en roumain. *Rivista di Grammatica Generativa*. Vol. 11. Unipress. 123–152.
- Pappert, Sandra, and Thomas Pechmann. 2014. Priming word order by thematic roles: No evidence for an additional involvement of phrase structure. *Quarterly Journal of Experimental Psychology* 67.11: 2260-2278.
- Scheepers, Christoph, Raffray, Claudine N., & Andriy Myachykov. 2017. The lexical boost effect is not diagnostic of lexically-specific syntactic representations, in *Journal of Memory and Language* 95, 102–115.
- Somers, Joren & Jóhanna Barðdal. 2022. Alternating Dat-Nom/Nom-Dat Verbs in Icelandic: An Exploratory Corpus-Based Analysis. *Working Papers in Scandinavian Syntax* 107: 83–110.
- Van Peteghem, Marleen. 2016. Verbs of pain and accusative subjects in Romanian. *Atypical predicate-argument relations* 33. 3.
- Ziegler, Jayden, Jesse Snedeker, and Eva Wittenberg. 2018. Event structures drive semantic structural priming, not thematic roles: Evidence from idioms and light verbs. *Cognitive Science* 42.8: 2918-2949.

## **Structural priming and the emergence of the *to*-dative in Middle English - The psycholinguistics of historical language change**

***Gunnar Jacob, Lena Kaltenbach & Mirja Maier***

*University of Mannheim*

It has been suggested that structural priming, i.e. a speaker's tendency to re-use syntactic structures encountered or produced shortly before, may constitute a key psycholinguistic mechanism behind grammatical language change (e.g. Kootstra & Muysken, 2019). Specifically, new grammatical structures may consolidate themselves and spread further within a language through structural priming. However, experimental studies on this issue are faced with a principled methodological problem: For cases of language change in previous centuries, it is obviously impossible to conduct experimental psycholinguistic research, while for phenomena which are currently in the middle of changing, it is not yet clear how this process will develop in the future.

The present study attempts to avoid this problem by simulating a structural priming experiment on the basis of historical corpus data. We investigate the possible role of structural self-priming in a well-documented instance of historical language change, i.e. the emergence of prepositional-object (PO) ditransitives in Middle English (e.g. McFadden, 2002; Trips & Stein, 2019). While only sentences with the double-object (DO) structure, such as (1), existed in Old English, sentences with the alternative prepositional-object (PO) structure, such as (2), started to emerge in Middle English from around 1250 AD, and consistently increased in frequency between 1250-1350 AD.

We relied on Gries' (2005) approach for investigating structural (self-) priming in contemporary corpora to simulate a structural priming experiment with corpus data from the M2 (1250-1350 AD) and M3 (1350-1420 AD) sections of the Penn-Helsinki Parsed Corpus of Middle-English (PPCME2). A corpus search for ditransitive sentences with a PO or DO structure yielded a total of 2927 ditransitive prime sentences. These were categorized as either 'PO', 'DO', or 'PO-similar'. The 'PO-similar' category consisted of sentences such as (3), which superficially resembled a PO in purely structural terms, but did not constitute an alternative for a DO. For each prime sentence, we determined the grammatical structure of the next ditransitive sentence following the prime within the same historical text (i.e. the equivalent of a target sentence in an experimental study).

Figure 1 shows the proportions of PO, DO, and PO-similar target sentences, separately for each of the three prime types. The results revealed a self-priming effect, with a significantly higher proportion of PO targets following PO primes than following DO primes (and vice versa). This suggests that authors were indeed primed by the structure of the preceding prime sentence while choosing a structure for the target sentence. Interestingly, the results showed an increased proportion of PO targets not only after PO primes, but also after PO-similar primes. This indicates that the PO structure can also be primed by structures which only superficially resemble a PO structure. Thus, it is possible that such structures may have contributed to the consolidation of the PO structure in the English language, through structural priming.

Priming effects were strongest when prime and target were separated by only a small number of intervening non-ditransitive sentences, and gradually decreased with more sentences between prime and target. This result is consistent with the findings from experimental studies on structural priming, in which the magnitude of the priming effect was also affected by the number of intervening sentences (e.g. Branigan, Pickering & Cleland, 1999).

With respect to the role of structural priming in language change, our results are consistent with the claim that the new PO structure consolidated itself within the English language through structural self-priming in medieval authors. The results also suggest that it is possible to detect traces of the psycholinguistic mechanisms driving language change in historical corpora. Practical problems with investigating structural

priming in historical corpus data, such as the repetition of structures for stylistic reasons or individual structural preferences, and possibilities to control for them in the analyses are discussed.

(1) (...) I wyll +geue you tresur wythout nombyr (CMMIRK,113.3103)

*(I will give you treasure without number)*

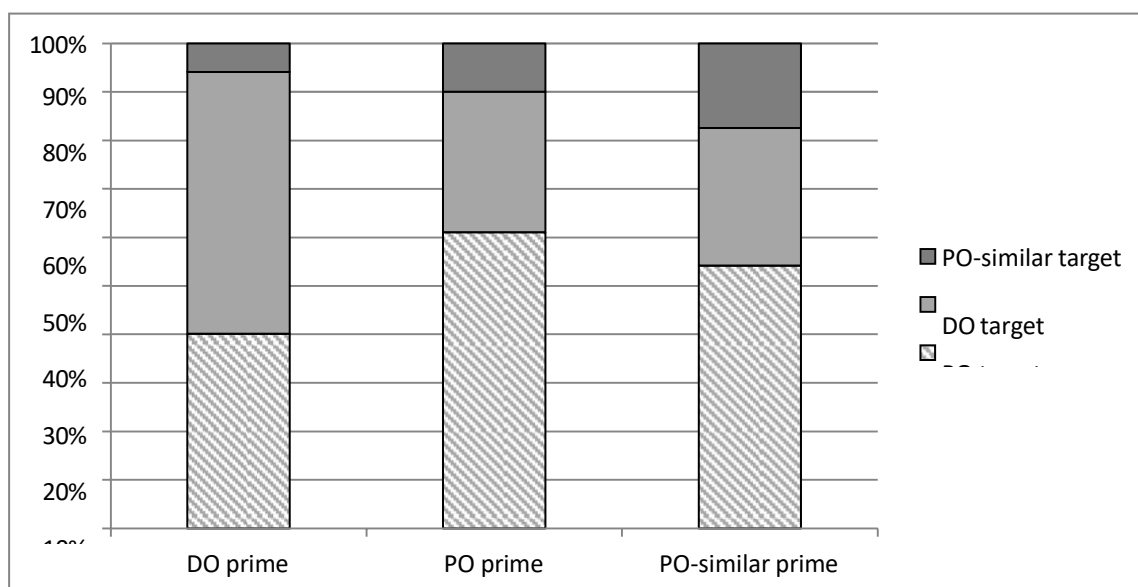
(2) (...) and gave the scabbard Excaliber to her lover (CMMALORY, 655.4476)

*(and gave the scabbard Excalibur to her lover)*

(3) (...) he +trewe doun +te emperour to +te er+te (CMPOLYCH, VI,435.3188)

*(he threw down the emperor to the earth)*

**Figure 1.** Proportions of ‘PO’, ‘DO’, and ‘PO-similar’ target sentences by prime type.



## References

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, 6, 635-640.

Gries, S. Th. (2005). Syntactic Priming: A Corpus-based Approach. *Journal of Psycholinguistic Research*, 34 (4), 365–399.

Kootstra, G. J. & Muysken, P. (2019). Structural Priming, Levels of Awareness, and Agency in Contact-Induced Language Change. *Languages*; 4(3). 65.

McFadden, T. (2002). The rise of the to-dative in Middle English. In D. W. Lightfoot (Ed.), *Syntactic effects of morphological change* (pp. 107–123). Oxford University Press.

Trips, C., & Stein, A. (2019). Contact-Induced Changes in the Argument Structure of Middle English Verbs on the Model of Old French. *Journal of Language Contact*, 12 (1), 232–267.



## Could Zipf's word-frequency law originate from sample-space reduction driven by systematic underspecification of intended meanings during word selection?

**Gerard Kempen & Karin Harbusch**

*Max Planck Institute for Psycholinguistics & Computer Science Department, University of Koblenz*

In 1936/1949, George Zipf identified a remarkably simple type of frequency distribution of words in English novels. Having ranked words in descending order of relative frequency (proportions interpretable as probabilities), he found that the product of probability  $p(w_i)$  of word  $w_i$  and its rank  $r_i$  has a more or less constant value ( $C$ ), valid within a substantial region of the frequency spectrum:  $C \approx p(w_i)r_i$ , equivalent to a power law  $p(w_i) = C/r_i^\alpha$  for  $\alpha \approx 1$ . Subsequent corpus studies in other languages and other text genres revealed that power laws tend to be the best fitting class of word frequency distributions, although  $\alpha$  may deviate somewhat from 1.

Ten years ago, in an influential review of studies on the origin of Zipf's Law, Steven Piantadosi (2014) concluded that none of the mathematical-statistical proposals meet an essential psycholinguistic criterion: being rooted in plausible assumptions about lexical selection.

Shortly thereafter, a new theory was put forward in several papers coauthored by Bernat Corominas-Murtra, Stefan Thurner, Rudolf Hanel et al. (2015/2016). They situate the origin of power-law phenomena in *Sample-Space Reduction* (SSR) processes, explaining this notion in terms of an appealing metaphor. Imagine standing atop a staircase throwing balls down the steps. Bouncing down, the balls will hit ("visit", "sample") all or some of the steps in top-down order, hitting the same step at most once, always halting at the lowest step (top frequency). Crucially, the balls never rebound to higher steps (unidirectionality). With each hit/sample, the number of possible next steps (the remaining "sample space") decreases. The authors present computer simulations and analytical proofs showing how during long sequences of random throws the frequencies of visits to individual steps converge to power laws with  $\alpha=1$ . This occurs under a wide range of unidirectional sampling regimens, and with nearly all "prior distributions" (initial patterns of wider/narrower steps yielding more/less likely hits).

We propose here to link this statistical model to the plausible psycholinguistic assumption that speakers and writers strongly tend to *underspecify* the concepts (meanings) they wish to convey. Rather than expressing specific, detailed concepts by selecting the words that precisely match those concepts, they often use words carrying less specific (less accurate, vaguer, more indeterminate) meanings — as evidenced, e.g., by studies on "basic-level categories," "good enough" language processing, and heavy use of pronouns and other pro-forms. This tendency creates 'upward pressure' in the resulting frequency distributions, for three reasons: (1) Unspecific words fit wider ranges of communicative situations than their specific counterparts, moving up their ranks; (2) this indirectly enhances the accessibility of unspecific words in the lexicon (e.g., due to higher activation levels); and (3) directionality reversals through *overspecification* are rare, easily entailing truth-value violations. The emerging distributions tend to be "Zipfian" even if the prior distribution of concepts to be expressed is non-Zipfian.

In our talk we present computer simulations supporting the hypothesis that large-scale concept underspecification could be crucial as driving force behind SSR and the ensuing Zipfian shape of word-frequency distributions. (Neuro)cognitive and corpus-linguistic techniques that allow identification of (specific) intended/activated concepts/meanings independently of the (subspecific) words actually used (cf. Fernandino et al., 2022; Kempen & Harbusch, 2019; Koranda et al., 2022) are needed to put the hypothesis on a solid empirical footing.

## References

- Corominas-Murtra, Bernat, Rudolf Hanel & Stefan Thurner (2015). Understanding scaling through history-dependent processes with collapsing sample space. *PNAS*, *112*, 5348-5354.
- Corominas-Murtra, Bernat, Rudolf Hanel & Stefan Thurner (2016). Extreme robustness of scaling in sample space reducing processes explains Zipf's law in diffusion on directed networks. *New J. Phys.*, *18*, 093010.
- Fernandino Leonardo, Jia-Qing Tong, Lisa Conant, Colin Humphries & Jeffrey Binder (2022). Decoding the information structure underlying the neural representation of concepts. *PNAS*, *119*, e2108091119.
- Kempen, Gerard & Karin Harbusch (2019). Mutual attraction between high-frequency verbs and clause types with finite verbs in early positions: corpus evidence from spoken English, Dutch, and German. *Lang. Cog. & Neurosci.*, *34*, 1140-1151.
- Koranda, Mark, Martin Zetterstein & Maryellen MacDonald (2022). Good-enough production: Selecting easier words instead of more accurate ones. *Psychol. Science*, *33*, 1440-1451
- Piantadosi, Steven (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.*, *21*, 1112-1130.
- Thurner Stefan, Rudolf Hanel, Bo Liu & Bernat Corominas-Murtra (2015). Understanding Zipf's law of word frequencies through sample-space collapse in sentence formation. *J. R. Soc. Interface*, *12*, 20150330.
- Zipf, George (1936). *The psycho-biology of language*. Abingdon, UK: Routledge.
- Zipf, George (1949). *Human behavior and the principle of least effort*. Reading, MA: AddisonWesley.

## Individual Variation in German Noun Phrase Agreement

---

**Hendrik Kligge, Ewa Dąbrowska & Thorsten Piske**

*Friedrich-Alexander-Universität Erlangen-Nürnberg*

German NP agreement is a complex phenomenon that seems to be learned quite early but mastered comparably late (cf. Szagun, 2011), arguably around the age of 8 (cf. Ulrich et al., 2016). Reasons for the complexity of NP agreement are, amongst others, many-to-many mappings of form and meaning as well as co-dependent elements within the NP: If the first element of a NP is a determiner or an adjective, it inflects for number, case and gender (a), (b). If the determiner is followed by one or more adjectives, the adjectives inflect according to their ‘weak forms’ -e or -en (c). If there is no determiner, but two adjectives, both generally show number, case and gender agreement (d). If it is a masculine or neuter dative NP, there are two accepted alternatives of either parallel or alternating inflection (d), which are commonly studied as a ‘case of doubt’ (cf. Moulin, 2002; Münzberg & Hansen, 2020; Nübling, 2011).

- (a) der Mann – die Frau – das Kind
- (b) großer Mann – große Frau – großes Kind
- (c) der große, (schöne) Mann – die große, (schöne) Frau – das große, (schöne) Kind
- (d) großer, schöner Mann – große, schöne Frau – großes, schönes Kind
- (e) großem, schönem Mann – großem, schönem Mann

Our hypothesis explores the impact of literacy and language aptitude as well as phonological salience of the -m/-n variation on mental representations and linguistic productivity of German NP agreement. To investigate this, we examined a total of 55 German native teenagers attending the 8th grade of a ‘Realschule’.

The study engaged participants in a written gap-filling task with nonce and real words, a verbal repetition task as well as a number of tasks for cognitive predictors. Results indicate a high level of overall command of the morphological system, with generally low error rates for the control condition. However, individual variations emerge, especially in the dative condition. Higher literacy levels correlate with lower error rates, yet challenges with dative noun phrase agreement persist. While overall error rates for most test conditions recede with increasing levels of literacy, the variability and the error rates for dative NPs remain comparably high across the board and the emerging patterns used by some individual speakers do not seem to be systematic. Virtually all possible patterns that would be categorized as ungrammatical by standard reference grammars can be found (e.g., (f), (g), (h), (i), (j)).

- (f) \*dem großem, schönem Mann
- (g) \*dem großem, schönen Mann
- (h) \*dem großen, schönem Mann
- (i) \*den großem, schönem Mann
- (j) \*den großen, schönem Mann

The findings suggest more potential gray areas in individual mental representations of morphosyntactic phenomena than usually assumed, evoking the conclusion that some individual speakers may not be fully productive with them.

### References

Moulin, C. (2002). Varianz innerhalb der Nominalgruppenflexion. Ausnahmen zur sogenannten Parallelflexion der Adjektive im Neuhochdeutschen. *Germanistische Mitteilungen*, 52, 73–97.

Münzberg, F., & Hansen, S. (2020). Starke vs. schwache Flexion aufeinanderfolgender attributiver Adjektive: Mit hohem technischen/technischem Aufwand. *Bausteine einer Korpusgrammatik des Deutschen*, 99-130 Seiten. <https://doi.org/10.17885/HEIUP.BKGD.2020.0.24237>

Nübling, D. (2011). Unter großem persönlichem oder persönlichen Einsatz? Der sprachliche Zweifelsfall adjektivischer Parallel- vs. Wechselflexion als Beispiel für aktuellen grammatischen Wandel. In K.-M. Köpcke & A. Ziegler (Eds.), *Grammatik—Lehren, Lernen, Verstehen: Zugänge zur Grammatik des Gegenwartsdeutschen*. De Gruyter.

Schmitz, K. (2006). Indirect objects and Dative case in monolingual German and bilingual German/Romance language acquisition. In D. Hole, A. Meinunger, & W. Abraham (Eds.), *Studies in Language Companion Series* (Vol. 75, pp. 239–268). John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.75.11sch>

Szagan, G. (2019). *Sprachentwicklung beim Kind* (7. überarbeitete Auflage). Beltz.

Ulrich, T., Penke, M., Berg, M., Lüdtke, U., & Motsch, H.-J. (2016). Der Dativerwerb—Forschungsergebnisse und ihre therapeutischen Konsequenzen. *Logos*, 24(3), 176–190.

## Development of productive preterit rules in the grammar of Swedish-speaking children

**Therese Landh**

*Uppsala University*

Swedish is commonly described as having four verbal conjugations. The preterit forms of the first, second and third conjugations take a dental suffix (*-de*, *-dde* or *-te*, predictable on a phonological basis). The fourth conjugation consists of the Swedish strong verbs, which typically have vowel shift and no preterit suffix (see table 1).

*Table 1: Swedish stems and preterit forms by conjugation.*

Conjugation	Stem	Preterit
1st	<i>hoppa</i> 'jump'	<i>hoppade</i>
2nd	<i>hyr</i> 'rent' <i>lås</i> 'lock'	<i>hyrde</i> <i>låste</i>
3rd	<i>tro</i> 'believe'	<i>trodde</i>
4th (examples of some strong patterns)	<i>gråt</i> 'cry' <i>rid</i> 'ride' <i>skär</i> 'cut'	<i>grät</i> <i>red</i> <i>skar</i>

The first conjugation, comprising verbs with stem final *a*, is often the only class ascribed productive status. Novel verbs adhere to the first conjugation pattern (because novel verbs are formed by suffixing *-a*): *Google* becomes *googla* 'google', preterit *googlade*. However, the other conjugations, including the fourth, have been productive historically when non-novel verbs have changed class (Lundberg, 1921; Strik, 2015; Wessén, 1965).

Research on verb inflection in child language (see Veres, 2004 for Swedish; Bleses et al. 2011 for a comparison to other Scandinavian languages) suggests that while Swedish-speaking children master the first conjugation first, overgeneralization errors almost exclusively result in second or third conjugation forms. Additionally, while overgeneralizations of vowel shift forms are rare, an increase is observed from age four to eight (Veres, 2004). No comprehensive explanation has been proposed for the Swedish data, but it is possible that they may be elucidated in light of the *Tolerance Principle* (Yang, 2016), a formula describing when it is more economic for children to form productive rules rather than list lexical forms individually.

The primary aim of the current study is to determine how verb inflection patterns are affected by age: what differences can be observed in experimental data from four, six, eight and ten-year-old Swedish-speaking children? I extend on previous research through an elicitation experiment where 128 children inflect verbs and nonsense verbs for past tense. In this talk, I present preliminary results and discuss what they reveal about the development of productive rules for verb inflection in Swedish. The *Tolerance Principle* (Yang, 2016) is discussed as a possible candidate for explaining the results.

The test items were selected to better understand productivity from a synchronic and diachronic perspective: many of the verbs included belong to ablaut classes that have proven sensitive to change historically (gaining new members or losing members to the dental suffix) – nonsense verbs were also formed keeping this in mind. The inclusion of nonsense verbs also enables assessment of how children handle novel verbs that do not end with *a*, and thus do not adhere to the productive first conjugation pattern. Since Veres (2004) observes increased overgeneralization of strong forms from four to eight years, the current study also includes ten-year-olds to determine if there is a further increase.

Preliminary data indicates that the dental suffix is the main source for overgeneralization errors and for production of nonsense forms for all age groups. However, children also produce strong ablaut forms for weak verbs and nonsense verbs, increasing with age. The strong overgeneralizations and nonsense forms produced seem to align with patterns of high type frequency as well as patterns that have been sensitive to change historically.

### **References**

Bleses, D., Basbøll, H., Vach, W., 2011. Is Danish difficult to acquire? Evidence from Nordic past- tense studies. *Language and Cognitive Processes* 26:8, 1193–1231.

Lundberg, J. 1921. *Studier över andra starka verbalklassen i nysvenskan med hänsyn tagen även till övriga starka verbalklasser*. Diss. Lunds universitet.

Strik, O. 2015. *Modelling Analogical Change: A History of Swedish and Frisian Verb Inflection*. Groningen Dissertations in Linguistics 137.

Veres, U. 2004. *Input and production in Swedish children's acquisition of past tense*. Diss. Göteborgs universitet.

Wessén, E. 1992 [1965]. *Svensk språkhistoria: I. Ljud och ordböjningslära*. Nytryck i nordiska språk 4.

Yang, C. 2016. *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. MIT Press.

## Productivity and semantic sparsity within a family of minimizing constructions

*Peter Lauwers & Margot Van den Heede*

*Ghent University*

This talk deals with the complex interplay between productivity and semantics. We compare a family of minimizing (sister) constructions in French that contain a lexically filled nominal slot (= *minimizer*, i.a. Hoeksema 2002) and a verb slot that hosts a variable set of verbs, such as [\_\_\_<sub>v</sub> *pas un iota*]:

*La droite la plus bête du monde ne change pas d'un iota* (FrenchTenTen2017)  
'The world's most stupid right-wing doesn't change a jota'

The relationship between productivity and semantics is still far from clear. First, although high(er) type frequency (*realized productivity*, Baayen 2009) often entails high(er) semantic diversity (~ schematicity), both in diachrony (Trousdale 2008: 170-171) and in synchronic comparison (Barðdal 2008; Van Goethem & Norde 2020), this correlation has been partly challenged (Perek 2020). Second, as to *extensibility* (Barðdal 2008), a concept that is operationalized here on the basis of hapaxes (cf. Baayen's *potential productivity*, i.a. Baayen 2009), part of the research has focused on the *local* semantic conditions under which extensibility is most probable to occur, both in subsequent diachronic stages (Perek 2016) and in acceptability judgments (Suttle & Goldberg 2011; Goldberg 2019). These studies highlight the importance of semantic similarity and, more specifically, the semantic density of the neighborhood of the new coinage (Perek 2016), its "coverage" (Goldberg 2019). The *global* semantic properties of extending constructions have been examined as well, showing that semantic *variability* (Suttle & Goldberg 2011) and, for low type-frequent constructions, its opposite, viz. semantic *coherence* (Barðdal 2008), are vectors of extensibility.

Against this background, we zoom in on three central questions:

- (i) Does high lexical diversity (TypeTokenRatio) necessarily entails wide semantic diversity?
- (ii) Is semantic similarity more important for hapaxes than non-hapaxes, suggesting that new types are more likely to appear in dense neighborhoods?
- (iii) Do constructions with lower lexical diversity (realized productivity) need to extend more coherently than constructions with higher lexical diversity?

For the sake of comparability (Gaeta & Ricca 2006; Zeldes 2012), we focus on the 19 most entrenched minimizers in the *French TenTen2017* corpus and keep the sample size constant ( $n = 100$  tokens). We systematically distinguish between *productivity* as *lexical diversity*, *realized* (TypeTokenRatio) and *potential* (TypeTokenRatio), and *semantic diversity*. For the latter, we use home-made type-based word-embeddings extracted from the *French Cow Corpus* (Schäfer 2015; Schäfer & Bildhauer 2012). More specifically, we computed for each minimizing construction its *semantic sparsity* (or *density*, its complement), that is the mean pairwise cosine dissimilarity between i) the types occurring in the verb slot (*global\_sparsity*) and ii) the hapaxes and their nearest neighbors (*hapax\_nearest-neighbor\_sparsity*). To control for the effect of type frequency on the mean semantic distance, raw figures have been replaced by *z\_scores* obtained via Monte Carlo random sampling.

Our main findings: i) lexical diversity does not significantly correlate with semantic diversity; ii) hapaxes are not better surrounded than non-hapaxes, and, iii) an inverse correlation ( $r = -0.59$ ;  $p = 0.01$ ) between realized productivity (TypeTokenRatio) and *hapax\_nearest-neighbor\_sparsity*, the most lexically diverse constructions extending more coherently. This surprising result can mainly be explained by the specifics of the minimizing construction: minimizers exhibit incipient grammaticalization, with different degrees of bleaching and opacification, which expose them to the vagaries of collocational conventionalization of the two slots.

## References

- Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 109-149. Dordrecht: Kluwer Academic Publishers.
- Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 181-208. Dordrecht: Kluwer Academic Publishers.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An international handbook*, 900-919. Berlin: De Gruyter.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: Benjamins.
- Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: a variable-corpus approach. *Linguistics* 44(1). 57-89.
- Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2019. Explain me this. *Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Hoeksema, Jack. 2002. Minimaliseerders in het Standaardnederlands. *Tabu* 32. 105-174.
- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(2). 149-188.
- Perek Florent. 2020. Productivity and Schematicity in constructional change. In Lotte Sommerer & Elena Smirnova (eds.), *Nodes and Networks in Diachronic Construction Grammar*, 141-166. Amsterdam/Philadelphia: Benjamins.
- Suttle, Laura & Adele Goldberg. 2011. The partial productivity of constructions as induction. *Linguistics* 49(6). 1237-1269.
- Trousdale, Graeme. 2008. A constructional approach to lexicalization processes in the history of English: evidence from possessive constructions. *Word Structure* 1. 156-177.
- Van Goethem, Kristel & Muriel Norde. 2020. Extravagant “fake” morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425-458.
- Zeldes, Amir. 2012. *Productivity in Argument Selection: From Morphology to Syntax*. Berlin: De Gruyter.

## Corpora

COW corpus:

- Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. *Proceedings of challenges in the management of large corpora (CMLC-3)*, IDS publication server. 28-34.
- Schäfer, Roland. & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*. 486-493.

TenTen corpus:

- Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, 125–127. Lancaster.



## Transitivization as productivity in transitivity: an experimental investigation in French

*Suzanne Lesage & Richard Huyghe*

*University of Fribourg*

Transitivity is widespread across languages and can encode a multitude of semantic relationships between verbs and direct objects, although with some cross-linguistic variation (Haspelmath 2015). The transitive construction is usually considered to be highly productive (Barðdal 2006), which manifests in two closely related phenomena:

- (i) a large proportion of verbs are transitive;
- (ii) novel verbs tend to be transitive.

Productivity in transitivity can also be observed in the extension of the transitive construction to existing intransitive verbs. This phenomenon is documented for English, in which some intransitive verbs can be used transitively (Callies 2018):

- (1) *He graduated (from) Harvard summa cum laude.* (COCA, 1993, NPR Weekend)

Cases of transitivization have also been reported in French by Serbat (1994) and Condamines (2013, 2017), and one may wonder under which conditions they occur. Corpus research highlights three forms of transitivization:

- **Objectivization** consists in making an oblique argument or an adjunct of an intransitive verb into a direct object (2a);
- **Causativization** consists in applying a causative-anticausative alternation pattern to a verb that is standardly used in the anticausative construction, thus adding a causer participant realized as the subject of the verb (2b);
- **Ex nihilo transitivization** adds a direct object out of the blue, i.e. without inferring it from any argument or adjunct in the verb's intransitive use (2c).

- (2) a. *cliquer l'icône* ("click the icon")  
 b. *mousser le lait* ("froth the milk")  
 c. *dormir une nuit paisible* ("sleep a peaceful night")

In addition, transitivization can be realized in different syntactic structures, including a direct object construction with lexicalized arguments (3a), a cliticized construction with the object in a pronominal form (3b), and a passive construction in which the novel object is promoted as the subject (3c).

- (3) a. *J'ai cliqué l'icône.* ("I clicked the icon.")  
 b. *Je l'ai cliquée.* ("I clicked it.")  
 c. *L'icône a été cliquée.* ("The icon has been clicked.")

The objective of our study is to determine (i) whether the different types of transitivization (as illustrated in (2)) influence the possibility of transitivization, and (ii) whether transitivization depends on differences in syntactic realizations (as illustrated in (3)), possibly in interaction with transitivization types.

We ran a 4\*3 acceptability judgement experiment on transitive sentences in French. The manipulated factors in the experiment are:

- transitivization type (between items: baseline (standard transitive verbs) vs. objectivization vs. ex nihilo vs. causativization);
- syntactic realization (within items: NP object vs. object clitic vs. subject of a passive).

24 experimental items and 24 fillers were created for the experiment. The (3\*24=72) experimental items were split into 3 lists following the Latin square. Participants had to rate the acceptability of sentences on a scale from 0 to 100. Preliminary results (based on responses from 150 participants, see Figure 1) suggest that ex nihilo constructions are less acceptable than other types of transitivization. Furthermore, ex nihilo transitivization and causativization contrast with objectivization in being sensitive to syntactic realization, with higher acceptability rates for NP objects than for passives and clitic objects.

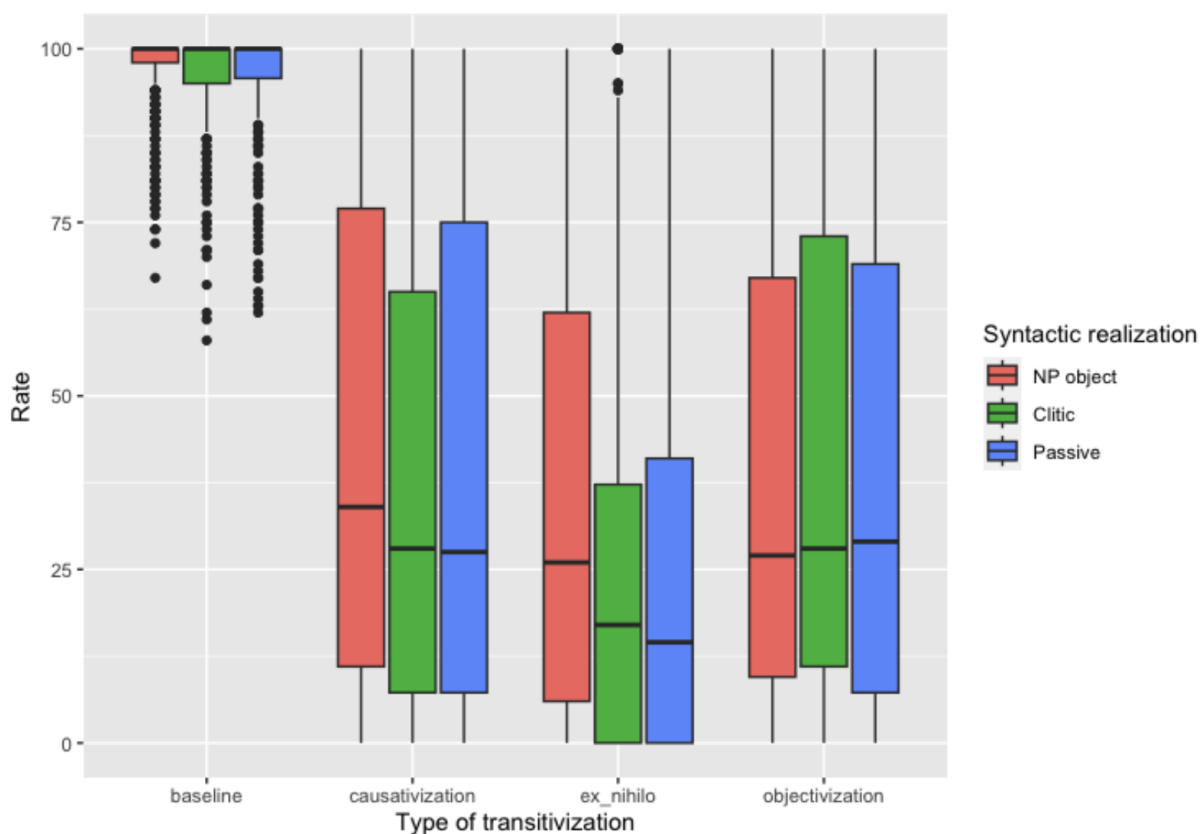


Figure 1. Acceptability judgements on transitive sentences depending on transitivization types and syntactic relations.

## References

- Barðdal, J. (2006). Predicting the productivity of argument structure constructions. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 32, No. 1, pp. 467-478).
- Callies, M. (2018). Patterns of direct transitivization and differences between British and American English. *Changing structures: Studies in constructions and complementation*, 151-167.
- Condamines, A. (2013). Quand la passion autorise la transitivisation d'un circonstanciel de lieu. *Journal of French Language Studies*, 23(3), 335-355.
- Condamines, A. (2017). The emotional dimension in terminological variation. *Multiple perspectives on terminological variation*, 18, 11.
- Haspelmath, M. (2015). Transitivity prominence. *Valency classes in the world's languages*, 1, 131-147.
- Serbat, G. (1994). Le golf: une démocratisation et une transitivisation galopantes. *L'information grammaticale*, 62(1), 57-58.

## The productivity of morphological and syntactic evaluative constructions in Italian: exploring networks and daughters

**Francesca Masini & Flavio Pisciotta**

*University of Bologna & University of Salerno*

Constraining the generative power of constructions and defining their actual degree of productivity represents a major challenge in Construction Grammar [1] and Construction Morphology [2] given their “constructions-all-the-way-down” design.

Traditionally, the productivity of constructions has been evaluated:

- by quantitative productivity measures (type/token ratio, hapax/token ratio, etc.), initially applied to morphology [3] and later adapted to syntactic constructions [4] along with the notions of schematicity and entrenchment [5][6];
- by the qualitative analysis of construction-internal (e.g., phonological, semantic, morphological) constraints on open slots (e.g., affixes selecting bases belonging to a specific word (sub)class).

However, productivity measures sometimes fail to represent actual facts. For instance, in the domain of evaluative morphology, it was observed that low-frequency evaluative affixes are recognized by quantitative measures to be more productive than more frequent (and, intuitively, more regularly exploited) evaluative and derivational affixes [7][8]. While evaluative affixes have been recognized as highly productive (e.g., in languages like Italian) [9], these results pose questions on how much productivity measures can account for their actual availability and entrenchment.

Our hypothesis is that the peculiar behavior of evaluatives can be better understood by looking not only at construction-internal constraints, but also at the portion of constructional network [1][10] that immediately surrounds the construction. Namely, even though evaluative complex words may be stored if they are frequent enough (Italian *cucchiaino* spoon.DIM ‘teaspoon’) or if they acquire an idiosyncratic meaning (Italian *bocconcino* bite.DIM ‘delicacy’), normally they are formed on the spur of the moment. In this sense, schemas in evaluative morphology should tend to behave more similarly to syntactic constructions [7] than to other word-formation schemas, as they tend to be instantiated more by constructs (e.g., concrete examples used in actual discourse and not retained in memory) than by daughter constructions that are lexically specified and stored (the predominant case in canonical word-formation).

Since generally studies on productivity have not compared syntactic and morphological constructions, we will explore this direction by looking at the productivity and entrenchment of a sample of evaluative constructions expressing approximation in Italian [11] including both established morphological (*pseudo-X*, *semi-X*) and syntactic (e.g., *specie/sorta di X* ‘kind/sort of X’) constructions, along with emerging ones that fall in-between morphology and syntax (e.g., *non-X* ‘not-X’, *mezzo-X* ‘half-X’). As a benchmark, we also included the productive derivational prefix in- ‘un-/in-’ [7].

First, we extract 500 occurrences for each of the above-mentioned constructions from the CORIS corpus [12]. We compare their productivity by computing various measures, including those traditionally applied to assess productivity (e.g., type/token ratio, hapax/token ratio), but also more recently exploited ones [13] to evaluate different aspects of the type and token distributions. Then, we assess the entrenchment of the studied constructions by means of a familiarity judgement test [14]. Participants are asked to rate on two six-point Likert scales the acceptability and familiarity of a sample of 3 (real) examples for each construction.

Finally, we compare the familiarity/acceptability scores of each construction with their productivity scores in order to:

- understand the relationship between familiarity and productivity in different types of constructions: we expect evaluative affixes to behave partly similarly to syntactic constructions in terms of productivity (due to the ‘construct’ nature of their daughters in the network), with emerging affixes lying in between and displaying a higher degree of interspeaker variability in their entrenchment;
- assess which of the productivity measures can account better for the actual availability of constructions at different levels of entrenchment.

## References

- 1 Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- 2 Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- 3 Baayen, Harald R. 1991. Quantitative aspects of morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 109-149. Dordrecht: Kluwer Academic Publishers.
- 4 Zeldes, Amir. 2012. *Productivity in Argument Selection. From Morphology to Syntax*. Berlin: De Gruyter.
- 5 Barðdal, Johanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.
- 6 Perek, Florent. 2020. Productivity and schematicity in constructional change. In Lotte Sommerer & Elena Smirnova (eds.), *Node and networks in Diachronic Construction Grammar*, 141-166. Amsterdam: John Benjamins.
- 7 Gaeta, Livio & Davide Ricca. 2003. Italian prefixes and productivity: a quantitative approach. *Acta Linguistica Hungarica* 50(1-2), 93-112.
- 8 Albair, Joshua. 2010. What is the state of evaluative affixes in contemporary English? *Début: the undergraduate journal of languages, linguistics and area studies* 1(1), 1-16.
- 9 Bauer, Laurie. 1997. Evaluative morphology: in search of universals. *Studies in Language* 21(3), 533-575.
- 10 Diessel, Holger. 2019. *The grammar network: how linguistic structure is shaped by language use*. New York, NY: Cambridge University Press.
- 11 Masini, Francesca, Muriel Norde & Kristel Van Goethem (eds.) 2023. Approximation in Morphology. *Zeitschrift für Wortbildung / Journal of Word Formation*, Special Issue 7(1).
- 12 Rossini Favretti, Rema, Fabio Tamburini & Cristiana De Santis. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, 27-38. Munich: LINCOM-Europa.
- 13 Van den Heede, Margot & Lauwers, Peter. 2023. Syntactic productivity under the microscope: the lexical and semantic openness of Dutch minimizing constructions. *Folia Linguistica* 57(3). 723-761.
- 14 Caldwell-Harris, Catherine, Jonathan Berant & Shimon Edelman. 2012. Measuring mental entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics. In Dagmar Divjak & Stefan Th. Gries (eds.), *Frequency effects in language representation*, 165-194. Berlin: De Gruyter.

## Productivity, Creativity, and Genre: The Case of Hebrew Possessive Constructions

**Nurit Melnik**

*The Open University of Israel*

### INTRODUCTION

In this paper, we explore the productivity of two constructions across two genres by using corpus-based measures. We show that productivity is a complex theoretical construct that cannot be reduced to one measure. Nevertheless, we argue that it is a part of speakers' linguistic knowledge, and that this knowledge is not only construction-specific but also includes sensitivity to genre. Furthermore, we investigate how this knowledge is exploited for the purpose of linguistic creativity.

The two construction in the focus of this study are two competing pronominal possessive constructions in Hebrew: an analytic construction, marked with a inflected possessive preposition *Sel* 'of' (1), and a synthetic construction in which a pronominal suffix is attached to the possessed nominal (2).

- |   |  |
|---|--|
| <p>(1) <i>ha-fem fel-i</i><br/> <i>the-name of-POSS.1S</i><br/>         'my name'</p> | <p>(2) <i>fm-i</i><br/> <i>name-POSS.1S</i><br/>         'my name'</p> |
|---|--|

The two variants are perceived to share the core meaning of possession, yet they are not completely synonymous; They have different histories and each has a characteristic discourse profile (Dubnov, 2000; Erb, 2022).

### PRODUCTIVITY AND GENRE

We used corpora of two distinct genres. For spoken language we consulted two small-scale corpora: CoSIH (Izre'el et al., 2002) and The Haifa Corpus (Maschler et al., 2021). The formal written genre was represented by IAHLTwiki, a UD-treebank of Wikipedia entries (Zeldes et al., 2022). Our findings regarding various productivity measures that have been proposed in the literature (Baayen, 1991, 2001, 2009) are presented below.

	Spoken		Written	
	Preposition	Suffix	Preposition	Suffix
tokens	653	172	118	2096
realized productivity (types)	286	55	105	668
type-token ratio	0.438	0.320	0.890	0.319
# hapax legomena	222	34	92	351
potential productivity	0.340	0.198	0.780	0.170

Table 1: Productivity measures

The *realized productivity* measures reveal that in the spoken language corpora the prepositional construction is considerably more productive than the suffixed one. Conversely, in the written corpus a much higher measure is found for the suffixed construction, suggesting that it is the preferred structure in written language. As for *potential productivity*, in both genres the prepositional construction receives a higher productivity score. We interpret the consistency of this measure across corpora as an indication that the prepositional construction maintains its expressive power and is more likely to be used for neologisms.

Our findings thus show that productivity is a complex construct, which is nevertheless a fundamental component of speakers' knowledge of language. Speakers (subconsciously) utilize grammatical resources, i.e. their productive constructions, as a function of the genre in which they communicate.

#### PRODUCTIVITY AND CREATIVITY

One aspect in which speakers' knowledge regarding the productivity of their resources is expressed is their linguistic creativity. Acknowledging that the notion of creativity is notoriously difficult to define (Runco & Jaeger, 2012), we adopt Simonton's (2012) definition which lists *novelty*, *utility* and *surprise* as necessary conditions. For language, we take utility to mean comprehensibility. Yet, while comprehensibility considerations drive language use towards conventionality, Haspelmath (1999) proposes the Maxim of Extravagance: Talk in such a way that you are noticed. This additional (competing) motivation, he claims, promotes linguistic innovations.

We suggest that the distinction between productivity and creativity is not categorical but rather a matter of degree, and is based on the element of *surprise*. We illustrate this with innovative instantiations, found in informal writing, of the less productive schema, which is also the dis-preferred one in the spoken genre.

One type of innovation is illustrated in (3), where an innovative variant of the masculine suffixed *iSt-o* 'his wife' appears with a feminine possessive suffix. We suggest that this is an instance of *productivity*. The innovative form fills a lexical gap brought by social change. Although the form is innovative, its construction choice is not surprising; it is coined by analogy to an existing form.

- (3) *paḡafti et iŋt-o / iŋt-a*  
*met.1S ACC woman-POSS.3SM / woman-POSS.3SF*  
 'I met his wife / her wife.'

A different type of innovation is illustrated in (4), where a borrowed English word (*ex*) used informally in Hebrew appears with a possessive suffix.

- (4) *eks-it-i ḡara be-arad*  
*ex-SF-POSS.1S lives.SF in-Arad*  
 'My ex lives in Arad.'

Unlike (3), the use of suffixation here is intended to create an aesthetic effect. The association of a suffix, which is perceived as belonging to a formal register, with a colloquial borrowings is creative and "eye-catching". An alternative prepositional expression would go unnoticed.

#### References

- Baayen, Harald. 1991. Quantitative aspects of morphological productivity. In *Yearbook of morphology 1991*, 109–149. Springer.
- Baayen, R Harald. 2001. *Word frequency distributions* Text, Speech and Language Technologies 18. Springer Science & Business Media.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdelling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 900–919. Berlin/New York: Mouton de Gruyter.
- Dubnov, Keren. 2000. Kinuyei hakinyan haxavurim vehaprudim shel shem ha-etsem ba-ivrit ha-meduberet; synthetic and analytic possessive pronouns related to nouns in spoken Hebrew. *Balshanut Ivrit; Hebrew Linguistics* 47. 21.

Erb, Ittamar. 2022. From synchrony and diachrony and back: the case of Hebrew pronominal possessives. MA thesis, Tel-Aviv University.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043—1068.

Izre'el, Shlomo, Benjamin Hary & Giora Rahav. 2002. Designing CoSIH: The corpus of Spoken Israeli Hebrew. *International Journal of Corpus Linguistics* 6(2). 171–197.

Maschler, Yael, Hilla Polak-Yitzhaki, Stav Fishman, Carmit Miller Shapiro, Netanel Goretsky, Gallith Aghion & Ophir Fofliger. 2021. The Haifa Corpus of Spoken Hebrew. <https://sites.google.com/humanities.haifa.ac.il/corpus>.

Runco, Mark A & Garrett J Jaeger. 2012. The standard definition of creativity. *Creativity Research Journal* 24(1). 92–96.

Simonton, Dean Keith. 2012. Taking the US patent office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal* 24(2-3). 97–106.

Zeldes, Amir, Nick Howell, Noam Ordan & Yifat Ben Moshe. 2022. A secondwave of UD Hebrew treebanking and cross-domain parsing. In *Proceedings of EMNLP 2022*, Abu Dhabi, UAE.

## Competing productive rules and ongoing change: evidence from Frisian verbal morphology

**Anne Merkuur**

*Fryske Akademy / Rijksuniversiteit Groningen*

This paper discusses recent data on changes in Frisian verbal inflection that shed an interesting light on morphological productivity in relation to language change. Frisian is a minority language in the Netherlands, spoken by nearly half a million bilingual (Frisian- Dutch) speakers (Provinsje Fryslân, 2020).

Unlike neighboring languages such as English, German and Dutch, Frisian has two classes of regular verbs: class I with an infinitive ending in *-a*, and class II ending in *-ja*. Both classes have their own inflectional paradigms. The past tense of class I verbs is similar to Dutch verbs as it exhibits a dental suffix *-te/-de* in the simple past:

### (1) Paradigms of class I and II

	Person/Nr	Class I	Class II
PRS	1SG	bak $\emptyset$	wurk j e
	2SG	bak st	wurk e st
	3SG	bak t	wurk e t
	PL	bak e	wurk j e
PST	1SG	bak te $\emptyset$	wurk e $\emptyset$
	2SG	bak te st	wurk e st
	3SG	bak te $\emptyset$	wurk e $\emptyset$
	PL	bak te n	wurk e n

Both classes are open and able to attract new members (Dyk, 2015), and class membership is nearly arbitrary (Merkuur, 2010). In terms of productivity, class II has traditionally been presented as the default, or only productive class (Popkema, 2006).

It is this seemingly clear image of the Frisian verbal system that we will overturn by confronting it with actual empirical data, thereby raising several questions for any theory on productivity in verbal morphology. These empirical data stem from a filling-in-the-gaps questionnaire (n=321) reported in Merkuur (2021) in which inflections of class II verbs were elicited. The results provide us with two facts that complicate the picture sketched above. The first of which is that speakers assign fairly novel verbs to both classes and productively inflect these verbs along the lines of both paradigms in (1). In other words, both classes seem to provide productive ways of inflecting Frisian verbs. This raises the question how and why the verbal system accommodates two sets of productive inflectional rules, which seem to be in competition since in theory they can both apply to any novel item?

Secondly, a recent and ongoing change seems to be breaking down the neat distinction between both paradigms. The results of the questionnaire show that although participants are fairly consistent in inflecting class II verbs along the lines of their paradigm, younger participants tend to inflect one form of class II verbs more and more with the dental suffix *-te/-de*, i.e. as if they were class I verbs: the second person singular in the past tense. *Do wurkest* thus becomes *do wurktest*:

### (2) Development of 2SG past of regular class II verbs

	norm	alternative
<i>wurkje</i>	<i>do wurk-e-st</i>	<i>do wurk-te-st</i>



work.INF      you work-PSTclassII-2SG      you work-PSTclassI-2SG

Why does this only happen to the second person singular past and how can we account for the actuation of this change? Zooming in on differences between participants who use these new forms structurally and participants who show a lot of intra-individual variation, we specifically find that this change seems to be ongoing and on the rise. Whereas L2 learners and older L1 speakers hardly use them or only use them on occasion, 1 in 5 younger L1 learners (born after 1980) uses them structurally and thus seems to apply a productive rule.

Looking at these developments through the lens of productivity may help us understand the mechanisms behind this change and provide more insight into whether and when this change will continue and what consequences such a shift in the verbal system might have for the coexistence of these two classes with competing productive inflectional rules.

### **References**

Dyk, S. (2015). *Verbal inflection*. Retrieved from: <http://www.taalportaal.org/>

Merkuur, A.A. (2010). *Conversie in het Fries*. Paper, University of Amsterdam.

Merkuur, A.A. (2021). *Changes in Modern Frisian verbal inflection*. Amsterdam: LOT publications.

Popkema, J. (2006). *Grammatica Fries*. Utrecht: Prisma Woordenboeken en Taaluitgaven.

Provinsje Fryslân. (2020). *De Fryske Taalatlas 2020: Fryske taal yn byld*  
<https://friesland.databank.nl/report/Taalatlas%202020%20%28Nederlands%29.pdf>

## Constructionalization and (limits on) productivity in emerging varieties: the case of Mozambican Portuguese

*Alice Mevis*

*Universidade Católica Portuguesa*

Mozambican Portuguese (henceforth MP) is an emerging variety of Portuguese undergoing fast-paced nativization since the country's independence from Portugal in 1975. Since MP is developing in a highly multilingual context, language contact thereby plays an important role. Recently, issues of productivity have been addressed in the context of emerging varieties of English, within the Dynamic Model framework (Schneider 2007). According to this model – initially designed for World Englishes but easily adapted to other pluricentric languages – MP situates itself between Phase 3, i.e., *Nativization*, a stage marked by large-scale linguistic effects (including syntactic innovations), and Phase 4, *Endonormative Stabilization*, in which the novel linguistic norms are accepted as adequate also in formal usage (Hoffman & Brunner 2022). An illustrative case of this dynamic in MP is the constructionalization of Recipient passives (Traugott & Trousdale 2013). Mozambican speakers have the possibility, with ditransitive verbs, to select either the Recipient or the Theme as subject of the passive clause (see (1) and (2) respectively), whereas in standard Portuguese only the Theme stands available for passivization (2).

- (i) *Estou muito feliz por essa oportunidade que fui dado para falar-vos acerca de uma das lições essenciais na vida humana.* (CP)  
'I'm very happy for this opportunity I was given to speak to you about one of the most essential lessons in life.'
- (ii) *Fiquei impressionado quando me foi dada a oportunidade de ver o que as nossas crianças são capazes de fazer.* (CP)  
'I was impressed when the opportunity was given to me to see what our children are able to achieve.'

Within the Dynamic Model, there is a claim that varieties in later stages should exhibit an increase in construction productivity, as the number of speakers grows (Hoffman & Brunner 2022). While this hypothesis aims at assessing productivity levels across varieties, our objective is to focus on a case of variety-specific productivity. The following research questions will guide our investigation: (i) How productive is this new MP construction and what are the constraints bearing on it? (ii) How can we characterize this construction in terms of the two important dimensions of productivity and schematicity? (iii) Is it possible to spot constructional changes in the making and make predictions as to future developments? (iv) What is the role of related schemas in the development of this construction? To address these questions, we carried out a corpus study aiming at an in-depth description of the constructional schema underpinning the formation of Recipient passives. To do this we relied on the Mozambican subsection of the *corpus do português* (CP) – *web/dialects* (Davies 2016) made up of over 27 million tokens retrieved from the internet.

Our research shows that even though the Recipient passive construction displays typical patterns of productivity, such as increase in type frequency and schematicity, productivity is far from an all-or-nothing phenomenon. Although the construction was attested with a total of 56 transfer verbs, this set is more semantically constrained for Recipient than for Theme passives, and also narrower in MP than in English or Bantu languages. Furthermore, while the Recipient passive construction commits to the requirement of semantic coherence, which helps increase its productivity (Barðdal 2008), there are other mechanisms at play (e.g., interactions with related constructional patterns or verb-specific preferences) that might influence the overall productivity of the construction. Finally, while there is some evidence towards more productivity and

extensibility, the current sociolinguistic situation in Mozambique and its inherent variability make it hard to have a clear view on the future trajectory of this constructional pattern. In particular, it is difficult to tease apart the effects of language contact from MP language-internal developments such as schematization.

## **References**

Barðdal, J. (2008). Productivity: Evidence from Case and Argument Structure in Icelandic. *Constructional Approaches to Language* (8). Amsterdam: Benjamins.

Davies, M. (2016-). *Corpus do Português: Web/Dialects*. One billion words; 4 countries. <http://www.corpusdoportugues.org/web-dial/>.

Gonçalves, P. (2010). *A génese do português de Moçambique*. Lisboa: Imprensa Nacional – Casa da Moeda.

Hoffmann, T. & Brunner, T. (2022). Construction Grammar meets the Dynamic Model. In: A. Ngefac, H.-G. Wolf & T. Hoffmann (eds.) *World Englishes and Creole Languages Today. Vol. 1: The Schneiderian Thinking and Beyond*. München: Lincom, 25–38.

Schneider, E. (2007). *Postcolonial English: Varieties around the world*. Cambridge University Press.

Traugott, E. & Trousdale, G. (2013). *Constructionalization and constructional changes*. Oxford University Press.

## **A quantitative study of avertivity. How Romanian *a fi pe cale* (lit. ‘to be on the way’) became a productive avertive construction**

***Pahonțu Serafino Beatrice***

***Université Paris Cité, LLF & University of Bucharest***

Avertivity ‘was on the verge of V-ing, but did not’ (Kuteva 1998) is a grammatical category characterized by aspectuality - expressing proximative meaning (i.e. the pre-phase of the event (Heine 1994)), temporality - pastness, and modality - counterfactuality. The present study investigates the avertive Romanian construction *a fi pe cale* (lit. ‘to be on the way’) by assessing its productivity through post-constructionalization constructional change (Traugott & Trousdale 2013) and also by examining its manifestation in the mind of the language user. We adopted a constructional approach (Goldberg 2013), exploring how co-textual factors (frequency type) influence the use of the avertive construction (token type) over time.

The diachronic perspective relies on a corpus of 1758 texts (233,126,081 tokens) provided by the Romanian Institute of Linguistics, ranging from the 16th to the 20th century. We annotated 976 aspectual occurrences according to the following formal schema: *(auxiliary) + pe cale + verb/deverbal noun*. Figure 1 depicts the polyfunctional nature of this construction, with proximative usage tracing to the early 19th century and avertive usage to the late 19th century. This supports the grammaticalization path from proximative to avertive (see Schwellenbach 2019 for Ibero-Romance languages). This avertive construction defines a schema that includes two free slots to be filled, respectively {auxiliary} and {lexical verb}. Firstly, the combinatorics with verb tenses show that past tenses account for only 16.2% of the total sample, correlating with a reduced frequency of proximative in the past and avertive uses. The construction always expresses avertivity with perfectives, while the imperfect tense determines the overlap of proximative and avertive. Secondly, the lexical verb strongly prefers telic predicates (achievement verbs, Vendler 1957) in both avertive and proximative uses. Moreover, at this stage, the avertive is context-dependent, appearing less frequently in monoclausal sentences (4.4%) than in subordinate clauses (12.9%). In the temporal subordinate clause cf. (1), an event is conveyed in its preparatory phase, with the main clause explicitly indicating the non-realization of the event in question. This suggests that the avertive meaning is a contextual implicature (pragmatic avertive function).

The synchronic analysis includes 635 annotated occurrences extracted from the *roTenTen16* corpus (3,142,636,172 tokens). The following results reveal an evolution in aspectual uses in the 21st century: there is an increase in the frequency of perfective tenses (32.6%) and the construction more frequently expresses the counterfactual meaning (62.2%) than the proximative meaning (37.8%) in the imperfect tense. Consequently, proximative has yielded ground to avertive. Regarding the lexical verb slot, the preference for achievement verbs remains consistent for both avertive and proximative. Concerning syntactic context, there is a rise in avertive use in monoclausal sentences cf. (2), and this occurs with both temporal markings. These results support the conventionalization of the avertive meaning (i.e. semantic avertive), indicating that the avertive use of *a fi pe cale* is the outcome of a semantic enrichment process, involving the conventionalization of the counterfactual implicature (see Traugott & Köning 1991). Additionally, we experimentally tested this hypothesis in two forced-choice tasks, under the following conditions: in monoclausal sentences, with both perfective and imperfective marking, with achievement cf. (3) and accomplishment verbs. The high percentages of avertive responses (over 90% with achievement verbs, irrespective of verbal tense, and over 84% with accomplishment verbs) validate our corpus results. Therefore, the avertive usage of *a fi pe cale* exhibits a high degree of *entrenchment* (Langacker 1987).

The observed form-meaning pairings provide empirical evidence that the productivity of the avertive *a fi pe cale* is connected to schematicity and entrenchment. At the conference, we will present the detailed results of our statistical analyses.

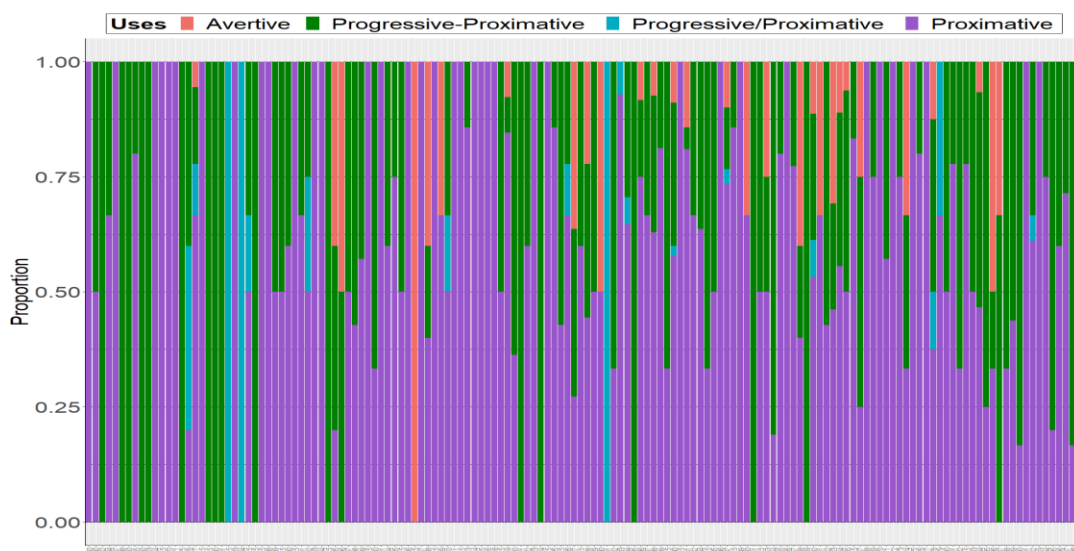


Figure 1: Aspectual uses of *a fi pe cale* in the diachronic corpus (19th – 20th century)

- (1) Tocmai în clipa când erai pe cale de a birui pe un  
 just in moment.DEF when be.IPFV.2SG on way of INF defeat.INF DOM an  
 dușman [...], te-a doborât. (1938\_DIAMANDI, D., 380). enemy  
 [...], ACC.2SG-PERF.3SG defeated  
 'Just when you were about to defeat an enemy, [...] he defeated you.'
- (2) Prin anii 1830, opt state americane au fost pe cale de a  
 by years.DEF 1830 eight states Americans PERF.3SG been on way of INF  
 intra în faliment. (roTenTen16, 257)  
 enter.INF in failure  
 'In the 1830s, eight American states nearly failed.'
- (3) Un avion {era<sub>IPFV</sub> / a fost<sub>PERF</sub> pe cale} să se prăbușească din cauza condițiilor meteo.  
 'A plane almost crashed due to the weather conditions'.

## References

- Goldberg, A. 2013. Constructionist approaches. Dans Thomas Hoffmann et Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 15-31. Oxford: Oxford University Press.
- Heine, B. 1994. On the Genesis of Aspect in African Languages. The Proximative. In K. E. Moore et. al. (eds.): *Special Session on Historical Issues in African Linguistics*. Berkeley: Berkeley Linguistics Society, 35-46.
- Kuteva, T. 1998. On identifying an evasive gram: action narrowly averted. In *Studies in Language*, 22 (1), 113-160.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar*. Vol.1: *Theoretical prerequisites*. Vol.2: *Descriptive applications*. Stanford: Stanford University Press.
- Schwellenbach, S. 2019. *Avertiv und Proximativ: Eine korpusbasierte synchrone und diachrone Untersuchung der romanischen Sprachen*. Berlin, Boston: De Gruyter.

Traugott, E. C. & E. König. 1991. The semantics- pragmatics of grammaticalization revisited. In: Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization. Vol. 1. Focus on Theoretical and Methodological Issues*. Amsterdam & Philadelphia: John Benjamins, 189-218.

Traugott, E. Closs & G. Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Vendler, Z. 1957. *Linguistics in Philosophy*. Ithaca: Cornell University Press.

## What lexical biases tell us about schematicity and productivity

**Dirk Pijpops**

*Universiteit Antwerpen*

Productivity is directly related to schematicity. That is, schematic constructions are productive in and of themselves, as new lexical material can readily be placed in their open slots, while concrete, ready-made exemplars do not feature any open slots that can accommodate for new lexical material. Cognitively stored exemplars can still be productive, but only through analogy. We know that language users can process language both through schematic constructions (Perek and Goldberg 2017) and ready-made exemplars (Tremblay and Baayen 2010; Dąbrowska 2014), but when do they use either? In this talk, I will propose a corpus-based method to determine for a case of morphosyntactic variation whether exemplar processing takes place. The key is to look for specific lexical biases that are created through contamination phenomena, such as constructional or lectal contamination, that crucially rely on exemplar processing (Pijpops and Van de Velde 2016, Pijpops 2022). I will present two lines of evidence for the link between these lexical biases and exemplar processing.

The first line of evidence comes from agent-based simulation. These are computer simulations wherein a number of agents, representing language users, interact with each other according to simple rules set by the researcher, with the goal of studying what is required for certain emergent behavior to take place (cf. Steels 2011). In this way, an agent-based simulation of lectal contamination was built. The results show that when the agents employ cognitive exemplar processing, the expected lexical biases develop (cf. Figure 1a), whereas when they use schematic constructions, no such biases emerge (cf. Figure 1b).

The second line of evidence is based on corpus research. Two corpus studies are presented, one on the Dutch partitive genitive, as in (1), and another on the variation between the determiners *zulke* ‘such’ and *zo’n* ‘such’, as in (2) (Van Olmen and van der Auwera 2014). On the one hand, both constitute similar cases of language variation: they both concern variation in the noun phrase, and both exhibit an outspoken difference between Belgian and Netherlandic Dutch, with the partitive genitive variant without *-s* and the determiner *zo’n* being more popular in Belgian Dutch. On the other hand, they are crucially different in that exemplar processing is much more likely for the partitive genitive. First, the variation in question concerns a bound morpheme, the *-s* ending, rather than a free morpheme. Second, the internal structure of the partitive genitive is highly atypical for Dutch grammar, while the phrases with *zulke* and *zo’n* follow the canonical structure of the Dutch noun phrase. Third, while both the partitive genitive and *zulke/zo’n*-phrases are productive, the partitive genitive appears to occur with a more limited set of lexical material. The results of the corpus studies that show that the expected lexical biases appear to be present for the partitive genitive, while they were not observed for the *zulke-zo’n* alternation.

- (i) a. *iets leuks*  
‘something fun’
- b. *iets leuk*  
‘something fun’
- (ii) a. *zulke praatjes*  
‘such talks’
- b. *zo’n praatjes*  
‘such talks’

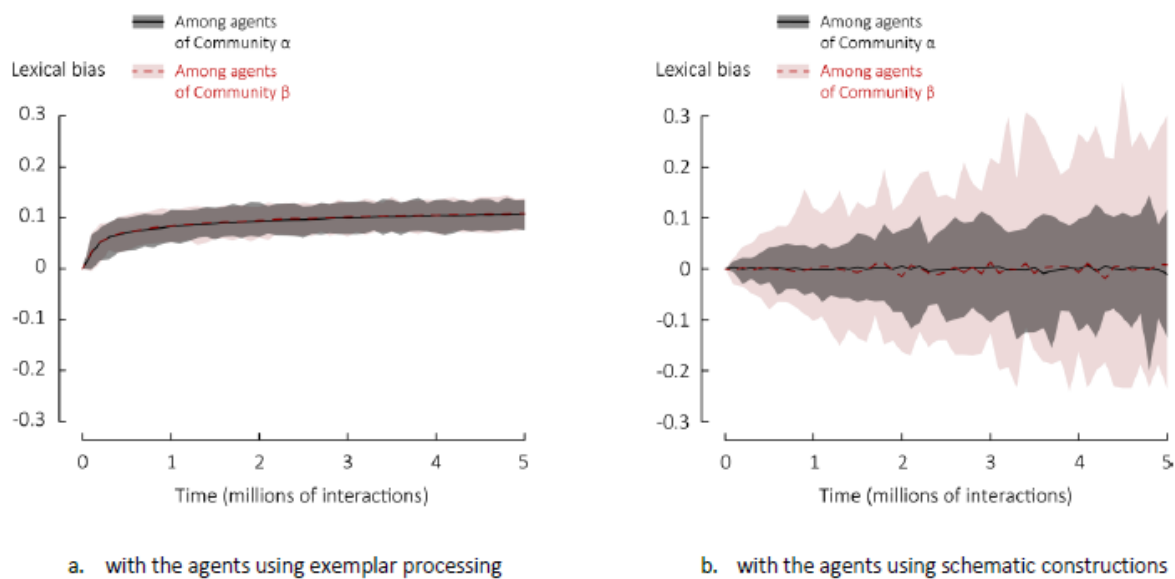


Figure 1: The development of a lexical bias in an agent-based simulation of lectal contamination. Lexical bias is measured as the preference of a word for one morphosyntactic variant over the other. The graphs show average, minimum and maximum results of 100 executions of the simulation with the same parameter settings. When exemplar processing is used, a lexical bias consistently develops, which is not the case when schematic constructions are used.

## References

- Dąbrowska, Ewa. 2014. Recycling utterances: A speaker's guide to sentence processing. *Cognitive Linguistics* 25(4). 617–653.
- Perek, Florent and Adele Eva Goldberg. 2017. Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition* 168. 276–293.
- Pijpops, Dirk and Freek Van de Velde. 2016. Constructional contamination: How does it work and how do we measure it? *Folia Linguistica* 50(2), 543–581.
- Pijpops, Dirk. 2022. Lectal contamination: Evidence from corpora and from agent-based simulation. *International Journal of Corpus Linguistics* 27(3), 259–290.
- Steels, Luc. 2011. Modeling the cultural evolution of language. *Physics of Life Reviews* 8(4). 339–356.
- Tremblay, Antoine and Rolf Harald Baayen. 2010. Holistic Processing of Regular Four-word Sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In David Wood (ed.), *Perspectives on formulaic language: acquisition and communication*, 151–173. London: Continuum.
- Van Olmen, Daniël and Johan van der Auwera. 2014. Over zo'n en zo meer. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument. Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 215–228. Leuven: Leuven University Press.



## From English borrowings to parts of the French system: the *-ing* suffix and velar nasal phoneme

**Arielle Roadman & Corentin Mazet**

*Indiana University*

English borrowings into French are increasingly studied (e.g. Poplack et al., 1988; Chelsey, 2010; Ten Hacken & Panocová, 2020), with *-ing* forms becoming numerous (e.g. *parking, bowling...*). While they were borrowed without awareness of morphological structure (Gottlieb & Furiassi, 2020), this awareness may be growing which begs the question of the status of the associated velar nasal into the French phonemic system. Thus, we ask: what are the morphological status and phonological repercussions of the introduction of *-ing* in Hexagonal French?

Morphological reanalysis may lead *-ing* to becoming productive in Hexagonal French. Bases borrowed with *-ing* can be derived: *stretching* → *stretcher* ‘to stretch’ (Lewis, 2007); suggesting the morphology of *-ing* to be understood. Using a matched guise technique (Casesnoves & Sankoff, 2004), we ran a study to probe native speakers’ judgments of nonce derivations containing *-age* (French counterpart to *-ing*), *-ness*, and *-ing* attached to 10 English and 10 French stems. The three suffixes are not treated equally, with French speakers accepting *-age* and *-ing* and rejecting *-ness* on French stems. On English stems, English suffixes are preferred, but these words are generally disfavored. These results support the productive nature of the suffix in French, as it patterns with the attested French suffix *-age*.

Historically, the velar nasal phoneme is absent from the French phonemic inventory (Walter, 1983; Walker, 2001), but the decreasing phonetic adaptation of English *-ing* borrowings has left its status unclear. Older studies reject it (Greenleaf, 1921), but more recently, studies attest the [ŋ] production without testing its phonemic status (Picone, 1996; Walker, 2001; Lewis, 2007). A marginal phoneme, restricted to the *-ing* suffix, is another possibility (see Bazell, 1952).

Participants recorded themselves reading the sentences in the aforementioned study, which included attested and novel borrowings and other cases of English /ŋ/. Stimuli were further balanced in phonetic context. Productions were analyzed by trained phoneticians to identify the phoneme, and results confirmed the changing status of the velar nasal. Indeed, the velar nasal was overwhelmingly produced by participants, across conditions, with no effect of sociolinguistic variables and not restricted to the *-ing* morpheme. These results suggest a now phonemic status to the phoneme in French, which contrasts with previous studies, as the production of the velar nasal was generally viewed to be either allophonic or morpheme restricted.

Borrowings in Hexagonal French are understudied compared to those in other varieties and the phonemic status of the velar nasal and the productivity of *-ing* had not been studied empirically before. This morpho-phonological study helps us look at the evolution of the status of the morpheme and the phoneme in a holistic way. Results support the integration of both the morpheme *-ing*, which seems to be productive in Hexagonal French, and of the velar nasal phoneme, whose status seems to be trending towards phonemicization.

### References

- Bazell, C.E. (1952). Phonemic and morphemic analysis. *WORD*, 8(1), 33-38.
- Casesnoves Ferrer, R., & D. Sankoff. (2004). “The Valencian Revival: Why Usage Lags behind Competence.” *Language in Society* 33: 1–31. Print.
- Chelsey, P. (2010). Lexical borrowings in French: anglicisms as a separate phenomenon. *French Language Studies*. 20, 231-251.

- Coveney, A. (2001). *The sounds of contemporary French: articulation and diversity*. Exeter: Elm Bank Publications.
- Dominicy, M. (2000). La dynamique du système phonologique en français. *Français moderne*, 68(1), 17-30.
- Furiassi, C. & Gottlieb, H. (2015). Pseudo-English, studies on false anglicisms in Europe. *Language Contact and Bilingualism*, 9, 1-35.
- Goldsmith, J.A. (1995). *The Handbook of Phonological theory*. Cambridge, Massachusetts: Blackwell.
- Greenleaf, J.H. (1921). Le [ɲ] ou "n mouillé". *The Modern Language Journal*, 5(7), 365-369. Lewis, J. C. (2007). *The -ing suffix in French*. [Master's thesis, University of North Dakota]. Morin, Y. C. (2009). Histoire des systèmes phonique et graphique du français. *Romanische Sprachgeschichte/Histoire linguistique de la Romania*, 3, 2907-2926.
- Mott, B. (2015). The rise of the English -ing form in Modern Spanish: A source of pseudo Anglicisms. *Language Contact and Bilingualism*, 9, 175-195.
- Pergnier, M. (1989). *Les anglicismes*. Paris: Presses Universitaires de France.
- Picone, M. D. (1996). *Anglicisms, Neologisms and Dynamic French*, Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26(1), 47-104. 10.1515/ling.1988.26.1.47
- Redman, R. (1978). L'adaptation phonétique des emprunts à l'anglais en français. *La linguistique*, 14(1), 111-124.
- Rifelj, C. (1996). False friends or true? Semantic anglicisms in France today. *The French Review*. 69(3), 409-416.
- Spence, N.C.W. (1987). *Faux amis and faux anglicismes : problems of classification and definition*. *Forum for Modern Language Studies*. 23(2), 169-183.
- Ten Hacken, P., & Panocová, R. (2020). *Word Formation, Borrowing and their Interaction*. Edinburgh : Edinburgh University Press.
- Walter, H. (1983). La nasale vélaire /ŋ/ un phonème du français?. *Langue française*, 60, 14-29. Walker, D.C. (2001). *French sound structure*. Calgary: University of Calgary Press.
- Walker, J. (2015). False Anglicisms in French: A measure of their acceptability for English speakers. *Language Contact and Bilingualism*, 9, 159-175.

## Productivity at the margins of polysemy: the case of Persian *faqare* ‘vertebra’

**Reza Soltani & Laura Janda**

*UiT, The Arctic University of Norway*

In this study we examine how the term *faqare* ‘vertebra’ has become grammaticalized (Hopper & Traugott, 2003, p. 2) from an anatomical term for spinal bones into a numeral classifier in Persian. Numeral classifiers are a subtype of noun classifier systems and are found in quantifying expressions that consist of a classifier and a head noun modified by a numeral (Example 1). They usually categorize nouns based on semantic properties such as animacy, humanness, sex, shape, and size (Aikhenvald, 2000, p. 21). We use corpus data from as early as the 7th century (early Modern Persian) to explore the grammaticalization of *faqare*. In addition, we draw on contemporary corpus data to investigate the productivity of this numeral classifier in Persian and argue that *faqare* behaves like a general numeral classifier.

Grammaticalization of *faqare* originates from a visual metaphor based on the physical similarity of a line of bones with verses and paragraphs which motivates the use of *faqare* to refer to texts and units thereof. Subsequent semantic extensions result in polysemy and in a shift in the central meaning of *faqare* from the domain of texts and documents to the domain of legal proceedings and crime including both objects (such as permits and bonds) and events (such as robberies and murders).

Contemporary corpus data reveal a radial category of meanings, where the most prototypical meanings refer to legal concepts and related semantic fields (Figure 1). Analysis of data shows that in addition to the core usage, *faqare* is used with nouns that lie outside of this radial category but are associated with the core usage by the context of use. The large number of hapaxes in the data as well as the schematicity of the numeral construction also indicates that *faqare* can be used productively (Bybee, 2010, p.67) in numeral constructions (Table 1). This means that the range of use of *faqare* has extended to the point that it can function as a general classifier, albeit with associations to certain contexts and styles. *Faqare* is generally considered a specific numeral classifier for legal documents, crimes, and car accidents (Mache, 2012, p. 76), but corpus analysis indicates that the scope of this numeral classifier goes beyond such nouns. Moreover, despite other numeral classifiers, *faqare* can be used with nouns that can be coupled with both objects and events.

1.
  - (a) 287 *faqare*            *ček*  
     287 vertebra.CLF    check  
     ‘287 checks.’
  - (b) 19 *faqare*            *serqat*  
     19 vertebra.CLF    robbery  
     ‘19 robberies’
  - (c) 180 *faqare*            *lāstik*  
     180 vertebra.CLF    tire  
     ‘180 tires’
  - (d) *yek* *faqare*            *pedar*  
     One vertebra.CLF    father  
     ‘one father’

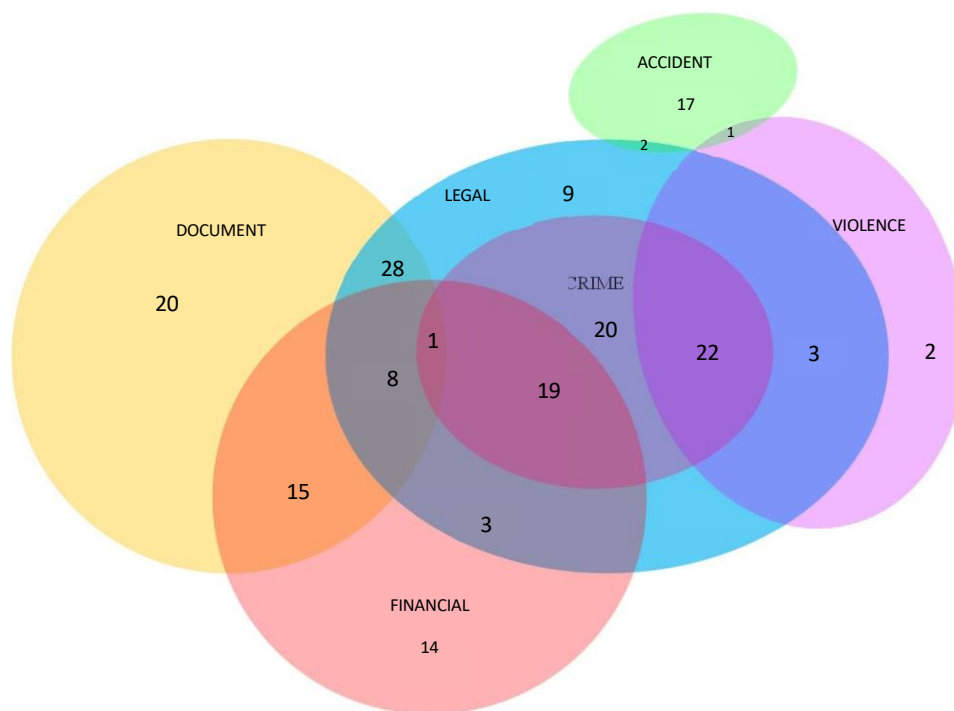


Figure 1. Venn diagram of radial category of *faqare*

Corpora token size	<i>faqare</i> token frequency		Type and (token) frequency		
	Total	As a classifier	Core	Periphery	Hapax
749,194,690	13,677	9013	184 (8383)	72 (355)	275

Table 1. Corpus statistics and *faqare* frequency

### References

- Aikhenvald, A. Y. (2000). *Classifiers: A typology of noun categorization devices*. New York: Oxford University Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2 ed.). New York: Cambridge University Press.
- Mache, A. (2012). *Numeral Classifiers in Persian*. Muenchen: LINCOM.

## The schematicity and productivity of alternating Dat-Nom/Nom-Dat verbs in German

**Joren Somers, Anna Jessen & Jóhanna Barðdal**

*Ghent University*

It has been shown that Modern German possesses a class of verbs which we have chosen to refer to as Dat-Nom/Nom-Dat verbs (Barðdal, Eythórsson & Dewey 2019; Somers, Jensen & Barðdal 2024; Somers et al. 2024, *inter alia*). The prime syntactic feature of these verbs is that they possess two diametrically-opposed argument structure constructions, Dat-Nom and Nom-Dat, instead of only one, i.e. either Dat-Nom or Nom-Dat. This means that both nominal arguments may take initial position in declarative clauses without there being a change in meaning or focus, as is illustrated by examples (1a–b) below:

- (1) a. *Dem Mann gefällt das Buch.*  
 the.DAT man likes the.NOM book  
 'The man likes the book.'
- b. *Das Buch gefällt dem Mann.*  
 the.NOM book be.to.liking the.DAT man  
 'The book is to the man's liking.'

However, a large-scale corpus study by Somers et al. (2024) has shown that there is wide within-class variation, in that some verbs tend more towards the nominative-before-dative order, whereas others are more strongly attracted towards the dative-before-nominative order.

The current paper investigates the schematicity and concomitant productivity of these alternating Dat-Nom/Nom-Dat constructions. Here we focus on the network link between the two constructions and argue that it is productive, in the sense that the relation between the two constructions is active and psychologically real in the minds of speakers. Thus, we argue that the alternation exists at a high level of schematicity (cf. Barðdal's 2008: Ch. 2 discussion of the relation between productivity and schematicity) even for verbs that are skewed more towards either the nominative-before-dative order or the reverse dative-before-nominative order. This conclusion is based on three psycholinguistic experiments in which alternating verbs demonstrate a robust priming effect of 8%, thus demonstrating a certain degree of extensibility.

### References

- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: John Benjamins.
- Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1): 107–170.
- Somers, Joren, Gard B. Jensen & Jóhanna Barðdal. 2024. Subjecthood and argument structure of cognate/synonymous Dat-Nom/Nom-Dat verbs across German and Icelandic. Submitted.
- Somers, Joren, Torsten Leuschner, Ludovic De Cuypere & Jóhanna Barðdal. 2024. A corpus-based analysis of the Dat-Nom/Nom-Dat alternation in German: A corpus study. Under revision for *Zeitschrift für Sprachwissenschaft*.

## The productivity of the Evaluative Attribution Construction in Modern English

---

**Veronika Stampfer**

*Friedrich-Alexander-Universität Erlangen-Nürnberg*

This study is concerned with constructs such as the following:

1. **He** still regards **himself** as a **working farmer**.ARCHER-1979-NEW
2. **We** consider **it** almost **one of the necessities of life**.ARCHER-1875-ADV
3. **He** finds **the bromide** to be **most suitable**.ARCHER-1864-MED
4. **We** no longer think of **your form of worship** as **idolatrous**.ARCHER-1966-SER

All of them share the same underlying Argument-Structure-Construction which expresses that an **Attributer** attributes an **Attribute** to an **Attributee**. In contrast to other linguistic structures that include what Quirk et al. (1985) call “object complement”, this attribution does not change the state of the Attributee like the “resultative construction” but is a mere subjective evaluation which does not have to be true. (Halliday 1967, 63. Compare: The sentence *He considers the couple husband and wife* doesn’t mean the couple is actually married, whereas *He pronounced the couple husband and wife* does make them a married couple.) This is why the construction, which is also known as “Qualifying Construction” (Schneider 1997) or “AGENT–AFFECTED–JUDGEMENT”-construction (Herbst, Uhrig 2009), will be called Evaluative Attribution Construction in this study.

Despite the semantic similarity, there is a lot of variation regarding the form of this construction. Aside from the verb, the construction also differs in the Attribute slot which can, e.g., be a noun phrase as in (1) and (2) or an adjective phrase as in (3) and (4). Moreover, the construction can connect the Attributee with Attribute slot with either the preposition *as* as in (1) and (4), without any filler as in (2) or with *to be* as in (3).

The goal of this study is, first, to determine how to structure this construction in the most appropriate way, e.g., if it is one construction or several allostructions of a more abstract constructeme. (cf. Cappelle 2006, Coleman 2011, Perek 2015) Second, an overview of its variation and the development should be delineated, also with reference to how each potential allostruction develops in terms of productivity. The method that will be applied is to closely analyze the fillers of those constructions and create collo-profiles for certain periods of time. Those collo-profiles will be compared to see in which contexts and combination of variables the type frequencies of the slot fillers as well as the token frequencies for each type increased or decreased most.

For this, the ARCHER corpus will be used to see how the productivity of the construction(s) develops over the period from 1600 to 1990 and find out how it differs between the verbs *regard*, *consider*, *find* and *think* as well as between the variants of the fillers of the other slots. The results can, on the one hand, be taken as an indication for the factors that play a role in the frequency decline of *think* in this construction after being one of the most prototypical verbs to express this meaning in Old English. On the other hand, this may shed light on the rise of *consider* and *regard* in this construction in frequency and productivity after they were borrowed from French during the Middle English period.

### References

ARCHER Corpus. <https://www.projects.alc.manchester.ac.uk/archer/>.

- Cappelle, Bert. 2006. "Particle Placement and the Case for 'Allostructions'." *Constructions* 1: 1–28.
- Colleman, Timothy. 2011. "Ditransitive Verbs and the Ditransitive Construction: A Diachronic Perspective." *Zeitschrift für Anglistik und Amerikanistik* 59 (4): 387–410.
- Halliday, Michael A. K. 1967. "Notes on transitivity and theme in English, Part 1". *Journal of Linguistics* 3: 37-81.
- Herbst, Thomas and Peter Uhrig. 2009. Erlangen Valency Patternbank. A Corpus-Based Research Tool for Work on Valency and Argument Structure Constructions. Also available at [www.patternbank.fau.de](http://www.patternbank.fau.de).
- Perek, F. 2015. *Argument Structure in Usage-Based Construction Grammar*. Amsterdam, Philadelphia: Benjamins.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1986. *A comprehensive grammar of the English language*, 24th edn. London, New York: Longman.
- Schneider, Edgar W. 1997. "As as 'is'. Is as 'is'?". In Udo Fries, Viviane Müller and Peter Schneider, eds. *From Ælfric to the New York Times: Studies in English Corpus Linguistics*. (Language and Computers: Studies in Practical Linguistics 19.) Amsterdam and Atlanta, GA: Rodopi, 33-50.

## Alternating predicates in Romanian and Bulgarian: A comparative study on the Harry Potter corpus

***Bilyana Todorova & Mihaela Iliaia***

*South-West University, Bulgaria; Ghent University*

This talk presents a comparative Romanian – Bulgarian study of a special type of predicates illustrated in (1a–b) for Romanian and in (2a–b) for Bulgarian, which occur with an accusative experiencer and a nominative stimulus.

<p>(1a) <i>Pe Maria o enervează indiferența</i>          PE Maria.ACC her.ACC irritates indifference.the.(NOM)          'Maria gets irritated by indifference'</p>	<p>(1b) <i>Indiferența o enervează pe Maria</i>          indifference.the.(NOM) her.ACC irritates PE Maria.ACC          'Indifference irritates Maria'</p>
<p>(2a) <i>Мене ме интересува историята</i>          me.ACC me.ACC interests history.the.(NOM)          'I am interested in history'</p>	<p>(2b) <i>Историята ме интересува</i>          history.the.(NOM) me.ACC interests          'History interests me'</p>

This kind of structure is known as *alternating predicate construction* and it is special in that the involved predicates select for two alternating and perfectly opposed argument structures: DAT-NOM or ACC-NOM vs. NOM-DAT or NOM-ACC (Barðdal et al. 2019), of which DAT-NOM and ACC-NOM are traditionally analyzed as cases of topicalization (Dobrovie-Sorin 1987), but more recently are argued to instantiate cases of non-canonical subject marking (Barðdal et al. 2019; Barðdal 2023: Ch. 3; Somers & Barðdal 2022). Among Indo-European languages, Romanian shows a high number of such accusative experiencer predicates (Van Peteghem 2016), though not all of them may be of the alternating type. In contrast, in Bulgarian, the accusative experiencer construction is argued to be restricted to a limited set of lexical psychological verbs (Tisheva & Dzhonova 2022), and to come in competition with a generalized dative experiencer construction (cf. Friedman & Joseph 2018).

The present study focuses on the accusative predicates and it aims at identifying the peculiarities of the construction in a comparative perspective in Romanian and in Bulgarian. Two hypotheses are at the origin of this research. The first hypothesis refers to the tendency in South-Slavic languages, to replace the accusative experiencer construction with a dative construction (Fedriani & Manzelli 2014, 87; Friedman & Joseph 2018). The second hypothesis puts forward a tendency observed among the Standard Average European (SAE) languages, including Romanian, toward canonical marking of core arguments (Haspelmath 2001; Bossong 1998).

By means of a synchronic corpus study, we investigate how frequent and how semantically spread the accusative experiencer construction is in the two languages. To do that, we created a complete Harry Potter corpus for each of the two languages. Given the nature of the corpus, – a parallel translation, we will observe the choices that were made with respect to the use of the studied construction, especially whether the use of the accusative experiencer construction in a language matches the use of that construction in the same context in the other language.

The main aim of this research is the study of the construction productivity, with special focus on its extensibility (Barðdal 2008). Therefore, a series of exploratory measurements will be applied such as token, type, and hapax frequency (Baayen 2009), as well as semantic density (Perek 2016), measurement which equates a high semantic spread of a construction with a high productivity. This combination of measurements will allow us to understand the synchronic tendencies of the accusative experiencer construction in the two languages investigated.

At a first glance at the data, the accusative experiencer construction is expected to be both less frequent and less expanded semantically in Bulgarian compared to Romanian. Will the data on the accusative experiencer



construction confirm the claim that Romanian goes against the SAE tendency toward canonical marking of core arguments? This is very less probable given the rivalry the investigated construction faces from the middle nominative experiencer construction in this language.

## **References**

Baayen, R. Harald (2009), Corpus linguistics in morphology. Morphological productivity, in: Lüdeling, Anke/Kytö, Merja (edd.), *Corpus linguistics. An international handbook*, Berlin/New York, De Gruyter, 900–919.

Barðdal, Jóhanna (2008), *Productivity. Evidence from case and argument structure in Icelandic*, Amsterdam/Philadelphia, John Benjamins Publishing.

Barðdal, Jóhanna (2023), *Oblique Subjects in Germanic: Their Status, History and Reconstruction*. To appear with Mouton de Gruyter.

Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey (2019), The Alternating Predicate Puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames* 11(1), 107–170.

Bossong, Georg (1998), Le marquage de l'expérient dans les langues de l'Europe, in *Actance et valence dans les langues de l'Europe*, J. Feuillet (ed.), Berlin/New York: Mouton/de Gruyter, 259–294.

Fedriani, Chiara & Gianguido Manzelli (2014), Costruzioni esperienziali con esperiente in accusative nelle lingue slave meridionali, in: Bonola, Anna Paola/Cotta Ramusino, Paola/Goletiani, Liana (edd.), *Studi italiani di linguistica slava*, Firenze, Firenze University Press, 75–94.

Friedman, Victor A. & Brian D. Joseph (2018), Non-nominative and depersonalized subjects in the Balkans, in: Barðdal, Jóhanna/Pat-El, Na'ama/Carey, Stephen Mark (edd.), *Non-canonically case-marked subjects*. The Reykjavík-Eyjafjallajökull papers, Amsterdam/Philadelphia, John Benjamins Publishing, 23–54.

Haspelmath, Martin (2001), Non-canonical marking of core arguments in European languages, in A. Aikhenvald, R.M.W. Dixon & M. Onishi (eds.), *Non-canonical Marking of Subjects and Objects*, Amsterdam: Benjamins, 53.

Dobrovie-Sorin, Carmen (1987), A propos de la structure du groupe nominal en roumain. *Rivista di Grammatica Generativa*. Vol. 11. Unipress. 123–152.

Somers, Joren & Jóhanna Barðdal (2022), Alternating Dat-Nom/Nom-Dat Verbs in Icelandic: An Exploratory Corpus-Based Analysis. *Working Papers in Scandinavian Syntax* 107: 83–110.

Tisheva, Yovka & Marina Dzhonova (2022), Syntactic characteristics of emotive predicates in Bulgarian: A corpus-based study. In *Proceedings of the 5th International Conference on Computational Linguistics in Bulgaria (CLIB 2022)*, Sofia, Bulgaria. Department of Computational Linguistics, IBL – BAS, 75–80.

Van Peteghem, Marleen (2016), Verbs of pain and accusative subjects in Romanian. *Atypical predicate-argument relations* 33. 3.

## Individual differences in productivity: Intra- and extralinguistic determinants in evaluations of “creative” uses of grammatical patterns

**Anouk Van den Stock, Anne-Sophie Ghyselen & Timothy Coleman**

**Universiteit Gent**

Constructions are rarely fully productive. In that sense, our linguistic creativity is “constrained in ways that can be hard to articulate” (Goldberg 2019: 1). Several scholars in CxG have described productivity in relation to the concept of creativity, and to Sampson’s (2016) F-creativity in particular (Hoffmann 2018, 2019; Bergs 2019). While the effects of several intralinguistic factors on the productivity of constructions have received ample attention, the roles of social and cognitive factors have been largely neglected thus far (cf. Hoffmann 2018). Yet, in view of the pervasive individual differences attested in speakers’ grammatical knowledge and grammatical representations in work by, among others, Dąbrowska (2012, 2018 etc.), we may expect a degree of between-speaker variation in the extent to which language users (i) extend constructions creatively, and (ii) evaluate creative extensions produced by other speakers as acceptable. Moreover, research in psychology demonstrates that general creativity is influenced by individual variables, such as gender, general intelligence, and personality traits, which suggests that such factors may be relevant to linguistic creativity, too. The present study therefore investigates inter-individual variation in evaluations of productive/creative instantiations of grammatical constructions by combining currently known intralinguistic and currently unexplored extralinguistic variables in a single design.

More than 700 native speakers of Dutch participated in an online acceptability rating experiment in which they evaluated both conventional and “unconventional/productive /creative” instantiations of two selected Dutch argument structure patterns, namely the *weg*-pattern in (1) (see e.g., Verhagen 2003) and the *krijgen*-passive in (2) (see e.g., Coleman 2015), on a 7-point Likert scale.

- (i) Hij baande/zocht/toeterde/elleboogde zich een weg door de menigte.  
‘He made/searched/honked/elbowed his way through the crowd.’
- (ii) Els kreeg een kaartje aangeboden/opgeplakt/geleverd/toevertrouwd.  
‘Els was presented/stuck on/delivered/entrusted a card.’ (lit. ‘E. got the card presented/...’)

The findings indicate considerable inter-individual variation in the extent to which speakers evaluate productive/creative instantiations of the patterns at stake positively or negatively. The results of ordinal regression analyses reveal (i) that participants’ ratings are influenced by their social backgrounds, linguistic experiences, and personality traits, and (ii) that intralinguistic and extralinguistic variables are inextricably linked to each other.

### References

- Bergs, Alexander. 2019. What, if anything, is linguistic creativity?. *Gestalt Theory* 41(2). 173-183.
- Colleman, Timothy. 2015. Constructionalization and post-constructionalization: The constructional semantics of the Dutch *krijgen*-passive in a diachronic perspective. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (ed.), *Diachronic Construction Grammar*, 213-255. Amsterdam/Philadelphia: John Benjamins.
- Dąbrowska, Ewa. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2(3). 219-253.

Dąbrowska, Ewa. 2018. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* 178. 222-235

Goldberg, Adele. 2019. *Explain me this. Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.

Hoffmann, Thomas. 2018. Creativity and construction grammar: Cognitive and psychological issues. *Zeitschrift für Anglistik und Amerikanistik* 66(3). 259-276.

Hoffmann, Thomas. 2019. Language and creativity: a Construction Grammar approach to linguistic creativity. *Linguistics Vanguard* 5(1). 20190019.

Sampson, Geoffrey. 2016. Two ideas of creativity. In Martin Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and Philosophy of Language*, 15-26. Bern: Peter Lang.

Verhagen, Arie. 2003. Hoe het Nederlands zich een eigen weg baant. Vergelijkende en historische observaties vanuit een constructie-perspectief. *Nederlandse taalkunde* 8. 328-346.