

# Syntactic productivity under the microscope: the lexical and semantic openness of Dutch minimizing constructions

Margot Van den Heede, Ghent University, mcvdnhee.vandenheede@ugent.be

Peter Lauwers, Ghent University, peter.lauwers@ugent.be

**Keywords:** syntactic productivity, distributional semantics, minimizing constructions

This study investigates different aspects of syntactic productivity (Barðdal 2008) and its relation to semantics, based on a comparison of 43 (Netherlandic) Dutch minimizing constructions. Syntactic productivity has a fundamental role in language and refers to “the range of lexical items that may fill the slots of constructions” (Perek 2018: 66). (Lexical) minimizing (micro-)constructions consist of the negator *geen* ‘no’, a predicate and a minimizer such as *fuck* that is used to reinforce sentential negation (Hoeksema 2002). Productivity for these constructions is measured at the level of the predicate slot:

(1) Het interesseert me eigenlijk geen ene fuck ‘I don’t really give a single fuck’

Our first goal is to examine the correlations between a set of 7 productivity measures (*TypeTokenRatio*, *HapaxTokenRatio*, *HapaxTypeRatio*) (Baayen 2009), some of which relate to conventionalization (*FrTop1*, *MeanFrTop3*, *SDTop3*, the *Alpha* parameter of the Zipf curve) and which are thus measures of anti-productivity (Van Wetteere 2021). These metrics capture distinct but complementary aspects of productivity and constitute a “multidimensional productivity complex” (Zeldes 2012). In order to chart their relations, a Principal Components Analysis (PCA) is performed (Kassambara 2017).

Second, we aim to disentangle the relation between lexical and semantic openness, as captured by distributional semantic analyses (Perek 2016). The interrelations between the frequency measures presented above and semantic variables remain far from clear. To this end, two semantic variables based on distributional semantics, *SemanticRange* and *SemanticSparsity*, are included in the PCA. Semantic range is defined as the proportion of semantic clusters covered by a given construction within the semantic space delineated by the fillers of the whole set of minimizing constructions (as such it is close to semantic variability, cf. Goldberg 2019). Semantic sparsity (aka density) captures the overall semantic diversity of the types of one specific construction, computed on the basis of the average pairwise cosine distances (Perek 2016).

Minimizing constructions constitute a particularly interesting testing ground, since no less than 43 minimizing constructions could be compared at equal sample size (i.e. 100 tokens per construction). The data come from the Netherlandic Dutch section of the Dutch Web 2014 corpus (nlTenTen14, Jakubiček et al. 2013), available on Sketch Engine.

The analysis reveals two underlying dimensions of openness. The first dimension corresponds, as expected, to the opposition between closed constructions and open constructions, both lexically (high *TypeTokenRatio* and *HapaxTokenRatio*) and semantically (high *SemanticRange*). Interestingly, the PCA also identifies a second dimension of productivity, which is correlated with *HapaxTypeRatio* and *SemanticSparsity*, showing that there is still room for innovation ‘behind’ entrenched, high token-frequent predicates. In addition, even lower type-frequent constructions may exhibit considerable semantic diversity (on average) compared to their type-frequent sister constructions. This sparsity is at odds with the principle of semantic coherence (Barðdal 2008). Finally, our study also highlights the amplifying role of analogy among minimizers.

## References

- 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* (Constructional approaches to language 8). Amsterdam: John Benjamins.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton University Press.
- Hoeksema, Jack. 2002. Minimaliseerders in het Standaardnederlands. *Tabu* 32(3-4). 105-174.
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, 125-127. Lancaster.
- Kassambara, Alboukadel. 2017. *Practical guide to principal component methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra*. STHDA.
- Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: a case study. *Linguistics* 54(2). 149-188.
- Perek, Florent. 2018. Recent change in the productivity and schematicity of the *way*-construction: a distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65-97.
- Van Wettere, Niek. 2021. Productivity of French and Dutch (semi-)copular constructions and the adverse impact of high token frequency. *International Journal of Corpus Linguistics* 26(3). 396-428.
- Zeldes, Amir. 2012. *Productivity in argument selection: From morphology to syntax*. Berlin: De Gruyter Mouton.