Productivity (metrics) and Semantics: A Principal Components Analysis on the Inchoative Construction

Sven Van Hulle, Ghent University, sven.vanhulle@ugent.be
Renata Enghels, Ghent University, renata.enghels@ugent.be

This research attempts to disentangle the theoretical relation between productivity and semantics, based on the results of Principal Components Analysis (PCA) on the inchoative construction in Spanish. The inchoative construction expresses the onset of an event, and incorporates in total over 25 auxiliaries, recruited from various semantic domains, for instance: change of state verbs (*rompió a llorar*, lit. '(s)he broke to cry'), motion verbs (*se echó a reír*, lit. '(s)he threw her/himself to laugh') or put verbs (*se mete a escribir*, lit. '(s)he puts her/himself to write') (Garachana 2017; Enghels & Van Hulle 2018; Fernández Martín 2019).

The study pursues three main goals. First, we dedicate special attention to how the productivity, still a complex concept in linguistics, of a construction can be captured best, including both well-established and more innovative metrics. Productivity in its application to syntactic structure, especially in the framework of usage-based Construction Grammar (Goldberg 2019, Barðdal et al. 2015), refers to the domain of application of a grammatical pattern. More specifically, it concerns "the range of lexical items that may fill the slots of constructions" (Perek 2016: 66), hence their *lexical openness*. The well-known productivity measures that are examined are type/token, hapax/token and hapax/type ratios (Barðdal 2008, Zeldes 2012, Perek 2016). The 'anti-productivity' measures (Van Wettere 2021) include the token frequency of the most token frequent filler(s) and the mean and standard deviation of the frequencies of the three most frequent fillers.

Second, the lexical and semantic openness of the auxiliary slot at macro-level, and of the infinitive slot at micro-level is investigated in detail, applying these well-established and innovative measures. Semantic openness is understood in terms of semantic range and semantic sparsity. Semantic range is defined as the proportion of semantic clusters covered by a *given* micro-construction, within the onomasiological space delineated by the fillers of the *whole set* of micro-constructions (as such it is close to semantic variability, cf. Goldberg 2019). Semantic sparsity, here applied to the whole micro-construction, captures the average semantic diversity of its types, computed on the basis of the average cosine distance between filler pairs (Perek 2016; Lenci 2018). The results reveal that the micro-constructions can be divided in three groups: (1) the highly productive micro-constructions with *comenzar* ('to start'), *empezar* ('to start'), *iniciar* ('to initiate) and *principiar* ('to start'), (2) the less productive variants with *echar* ('to throw') and *romper* ('to break'), and (3) the intermediate types with *arrancar* ('to tear off'), *lanzarse* ('to launch'), *largarse* ('to go away'), *liarse* ('to bind'), *meterse* ('to put'), *ponerse* ('to put') and *saltar* ('to jump').

Third, the interaction between lexical and semantic openness is addressed, as captured by distributional semantic analyses (Perek 2016). A first inspection of the PCA uncovers that semantic range correlates with lexical openness, but that semantic sparsity adds an extra dimension. For example, micro-constructions with high lexical openness (and semantic range) do not necessarily have a high semantic sparsity, since its fillers can form semantically isolated islands, making it semantically rather unproductive. Therefore, the metric of semantic sparsity proves a useful tool to gain more insights in the (semantic) productivity of a construction and enables to leap beyond the scope of the traditional productivity measures.

References

Barðdal, J. 2008. *Productivity: Evidence from case and argument structure in Icelandic* (Vol. 8). Amsterdam: John Benjamins Publishing.

Barðdal, J. et al. 2015. *Diachronic construction grammar*. Amsterdam: John Benjamins Publishing.

Enghels, R., & Van Hulle, S. 2018. El desarrollo de perífrasis incoativas cuasi-sinónimas: entre construccionalización y lexicalización. *Elua 32*. 91–110.

Fernández Martín, P. 2019*. Las perífrasis verbales del español: una perspectiva histórica*. Madrid: Arco/Libros.

Garachana, M. ed. 2017. *La gramática en la diacronía: la evolución de las perífrasis verbales modales en español*. Madrid: Iberoamericana Vervuert.

Goldberg, A. E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton, NJ: Princeton University Press.

Jolliffe, I. T., & Cadima, J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374*(2065).

Lenci, A. 2018. Distributional models of word meaning. *Annual review of Linguistics 4*, 151-171.

Perek, F. 2016. 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, N. 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, A. 2012. *Productivity in argument selection: From morphology to syntax.* Berlin: De Gruyter Mouton.