Productivité et sémantique : à la rencontre de la famille de minimiseurs

Peter Lauwers & Margot Van den Heede
(Université de Gand)

La *productivité* peut être définie comme le domaine d'application lexical (potentiel) d'une règle ou d'une construction (Baayen 1992, 2009 ; Barðdal 2008 ; Zeldes 2012). Pendant très longtemps, les études sur la productivité se réduisaient à des comptages de mots, mettant en œuvre une panoplie de mesures quantitatives, partiellement corrélées (Van Wetteere 2021). Ces dernières années, et notamment depuis l'avènement de la sémantique distributionnelle, ces approches axées sur l'ouverture lexicale ont été complétées par des considérations de nature sémantique, dont l'impact ne peut pas être sous-estimé (Barðdal 2008, Goldberg 2019, Perek 2016, Shen & Baayen 2023, Van den Heede & Lauwers 2023).

Nous rappellerons d'abord comment la notion de productivité, qui à l'origine s'appliquait essentiellement aux règles d'affixation (Baayen 1992 ; Dal 2003), a pu être étendue aux structures syntaxiques dans le cadre (« usage-based ») de la Grammaire des Constructions, pour ensuite procéder à un zoom sur les constructions (désormais Cxs) qui font l'objet de la présente communication : les « minimiseurs » (Hoeksema 2002) ou renforçateurs de la négation phrastique en français. En voici un exemple :

Les banques ne prêtent qu'aux riches, pas à ceux qui **n'ont pas un radis** (French Ten Ten corpus, 2017).

Pour cette étude, nous avons réuni 124 micro-cxs dans lesquelles un minimiseur, c'est-à-dire un nom désignant une petite quantité (*gramme*), un petit objet (*miette*), une valeur insignifiante (*rond*), etc., se combine avec la négation phrastique (*ne*) *pas* et une liste de prédicats:

$[(ne) \{Prédicat_x\} pas un(e) N_{min}]_{Cx}$
Appliqué à *radis* : X *ne* {a, vaut, coûte, possède, reste, ...} *pas un radis*

Il s'avère que liste des verbes susceptibles de remplir le « slot » prédicatif varie d'une Cx à l'autre, ce qui illustre bien le continuum de productivité qui ressort de l'application des principales mesures de productivité (*type/token ratio*, *hapax/token ratio*, *hapax/type ratio*, etc.). Notons que pour une comparaison à armes égales, c'est-à-dire à échantillons de taille égale (Zeldes 2012), nous nous limiterons ici l'examen quantitatif aux 19 minimiseurs qui totalisent au moins 100 occurrences dans le *French Ten Ten Corpus* (2017 ; Jakubiček et al. 2013).

Puis, après avoir rappelé les fondements de la sémantique distributionnelle (e.a. Erk 2012) et la façon dont nous l'avons mise en œuvre à partir du *French Cow corpus* (Schäfer 2012 et al., 2015), nous développerons le concept de *densité/sparsité sémantique* pour comparer l'ouverture sémantique des constructions et examiner ses rapports avec la productivité. La densité sémantique, affinée par le biais d'une méthode Monte-Carlo, sera en outre visualisée par des plots MDS (= *MultiDimensional Scaling*, positionnement multidimensionnel) qui projettent les différentes constructions sur un fond de carte onomasiologique.

Munis de cet outillage, nous répondrons aux trois questions suivantes :

1. Quelle est la sparsité sémantique globale des 19 micro-cxs?
2. Comment la sparsité sémantique (= diversité sémantique) se rapporte-t-elle à la productivité *réalisée* (= diversité lexicale)? Les Cxs les moins productives sont-elles aussi les plus denses ?
3. Les Cxs les moins productives s'étendent-elles nécessairement de manière plus cohérente (Barðdal 2008) ? Plus concrètement, leurs hapax sont-ils mieux entourés?

Pour terminer, nous avancerons quelques réflexions sur les relations paradigmatiques au sein de la famille des micro-cxs qui influencent le comportement de celles-ci.

Références bibliographiques

- Baayen, R. H. 1992. Quantitative aspects of morphological productivity. In *Yearbook of Morphology 1991*, ed. Booij, G. E., & J. Marle. Dordrecht: Kluwer Academic Publishers. 109–149.
- Baayen, R. H. 2009. Corpus linguistics in morphology: morphological productivity. In *Corpus Linguistics. An international handbook*, ed. Lüdeling, A. & M. Kyto. Berlin: De Gruyter. 900–919.
- Barðdal, J. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam: Benjamins.
- Dal, G. 2003. Productivité morphologique: définitions et notions connexes. *Langue française* 140: 3-23.
- Erk, K. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass* 6(10). 635–653.
- Goldberg, A. 2019. *Explain me this. Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press
- Hoeksema, J. 2002. Minimaliseerders in het Standaardnederlands. *Tabu* 32. 105–174
- Perek, F. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1): 149–188.
- Shen, T. & R. H. Baayen. 2023. Productivity and semantic transparency: An exploration of word formation in Mandarin Chinese. *The Mental Lexicon*, 1-22.
- Van den Heede, M. & Lauwers, P. 2023. Syntactic productivity under the microscope: the lexical and semantic openness of Dutch minimizing constructions. *Folia Linguistica* 57(3): 723-761.
- 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: 396-428.
- Zeldes, A. 2012. *Productivity in Argument Selection: From Morphology to Syntax*. Berlin: De Gruyter.

Corpus

- Jakubíček, M., A. Kilgarriff, V. Kovář, P. Rychlý & V. Suchomel. 2013. The TenTen corpus family. In *7th International Corpus Linguistics Conference CL*, 125–127. Lancaster
- Schäfer, R. 2015. Processing and querying large web corpora with the COW14 architecture. In *Proceedings of challenges in the management of large corpora (CMLC-3)* (IDS publication server), 28–34.
- Schäfer, R. & F. Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, 486–493.