Quantitative measures of affix rivalry

Justine Salvadori Université de Fribourg
Rossella Varvara Université de Fribourg
Richard Huyghe Université de Fribourg

1 Background

Affix rivalry occurs between affixes that have equivalent semantic functions and can therefore compete in the formation of derivatives (Lindsay & Aronoff, 2013; Arndt-Lappe, 2014; Fradin, 2019; Huyghe & Varvara, 2023; a.o.). However, equivalence may be established only between some of the functions of polyfunctional derivational processes. According to Lieber (2016), for example, the English suffixes *-ation* and *-al* can both derive event (*conversation*, *portrayal*) and result (*coloration*, *acquittal*) nouns, but only the former can be used to derive instrument (*decoration*) and agent (*administration*) nouns.

The fact that rival affixes are not always strictly equivalent entails that morphological competition should be considered a gradient relationship. Semantic differences observed between rival affixes can be more or less important, and affixes can be seen as more or less rivaling depending on how close they are semantically. This gradient nature of affix rivalry calls for an appropriate, i.e. quantified, assessment. Ideally, a coefficient of competition should be provided so that different situations of rivalry can be compared both within languages and cross-linguistically.

This work introduces two similarity measures drawn from studies in ecology that can be used to assess degrees of rivalry between polyfunctional affixes: the Sørensen index (Sørensen, 1948), which quantifies how similar two affixes are according to the proportion of functions they share; and the Percentage similarity coefficient (as a complement to the Percentage difference index proposed by Odum, 1950), which quantifies how similar two affixes are considering type frequencies. Two complementary measures — Balanced richness (for the Sørensen index) and Balanced abundance (for the Percentage similarity coefficient) — are also provided to further analyze the semantic dissimilarity between rival affixes. For instance, they can help identify nestedness, i.e. when the functions of an affix A are a subset of the functions of an affix B, and overlap, i.e. when two affixes A and B have functions in common but also specific functions that are not covered by B and A, respectively (Plag, 1999; Guzmán Naranjo & Bonami, 2023; a.o.).

2 Case study

French deverbal suffixes often compete for morphosemantic functions (Dubois, 1962; Thiele, 1987; Huyghe & Wauquier, 2021; a.o.). In order to explore the potential of the proposed measures, we selected six of them for a case study: 3 eventive suffixes (*-ade*, *-ment*, *-ure*) and 3 agentive suffixes (*-aire*, *-ant*, *-eur*). Given that morphological competition can only be investigated through the lexicon, a random sample of 100 French deverbal nouns formed with each suffix was retrieved from the French web corpus FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015). To identify functions, each collected noun was then semantically analyzed using a double classification (Salvadori & Huyghe, 2023) that distinguishes between the ontological description of the referent (e.g. animate entity, artifact, event) and the relation with the eventuality denoted by the base verb (e.g. agent, instrument, result). In total, 21 ontological and 18 relational classes were considered and assigned to nouns using linguistic tests and

definitions taken from the literature (Flaux & Van de Velde, 2000; Petukhova & Bunt, 2008; Haas et al., 2022; a.o.). The different measures were finally applied to the 6 suffixes based on the 782 word meanings and 37 functions identified in the dataset.

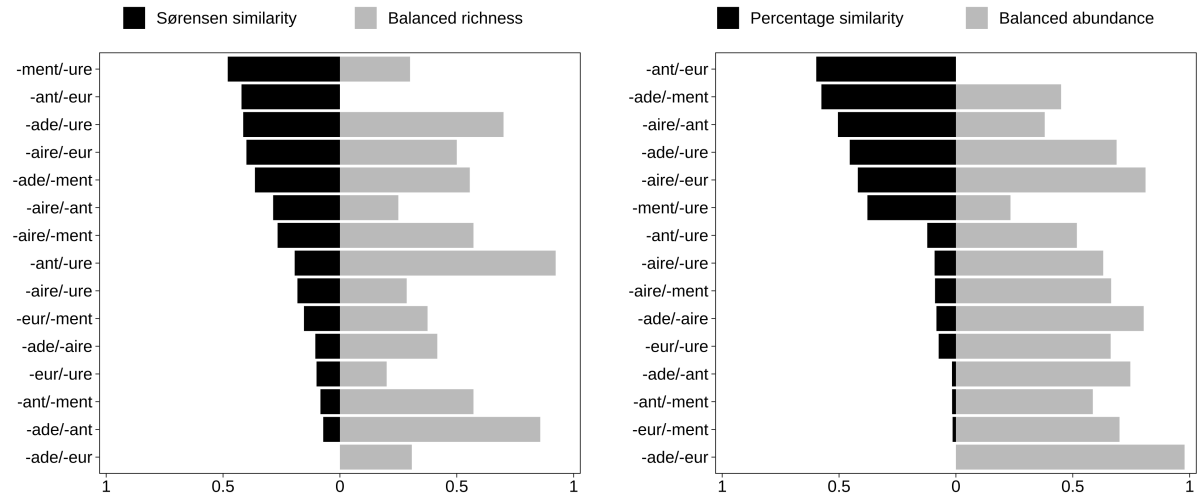

Figure 1: Scores for the incidence- (Sørensen similarity and Balanced richness) and abundance-based (Percentage similarity and Balanced abundance) measures. Pairs of suffixes are ordered from top to bottom by decreasing similarity.

Overall, the results of the case study support the need to approach affix rivalry as a gradient phenomenon. As shown in Figure 1, there are no perfect rivals in the sample and almost all suffixes compete — even in very small proportions. It remains that the pairs composed of suffixes belonging to the same semantic group (i.e. agentive or eventive) obtain higher scores than those contrasting two types of suffixes.

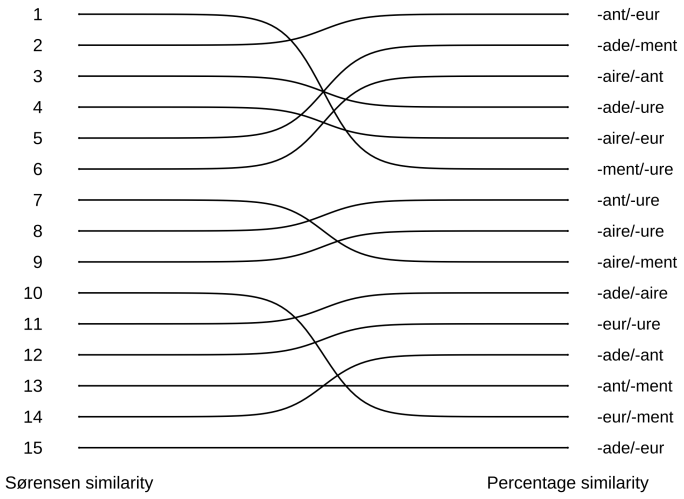

Figure 2: Ranking of the suffix pairs according to the Sørensen vs. Percentage similarity measures.

The proposed measures highlight different facets of similarity relationships and complement each other accordingly. As incidence-based measures, the Sørensen and Balanced richness indices allow in-depth investigation of functionality structures. As abundance-based measures, the Percentage similarity and Balanced abundance indices can weight functional rivalry by realization frequency and shed a different light on the sharing of functions. The comparison between the two types of measures informs on the architecture of rivalries and on the (in)congruence between the number of shared functions and the number of derivatives that instantiate these functions. In this sample, there is a strong and significant correlation between the Sørensen and the Percentage similarity scores (Mantel test: $r = .868$, $p < .01$). This suggests that suffixes that have many functions in common also tend to present a relatively even distribution of derivatives across shared functions, although some exceptions can be noted. For example, while the suf-

fixes *-ment* and *-ure* are the most similar according to the Sørensen index, they lose 5 places in the ranking based on the Percentage similarity measure (see Figure 2), meaning that, although they share a high number of functions, they realize them at different frequencies.

3 Conclusion

This work introduces different measures of affix rivalry and explores their potential through the analysis of a sample of 600 nouns formed with 6 nominalizing suffixes in French. The metrics presented in the study should be considered a first step toward a comprehensive measurement of morphological competition. They do not account for the diachronic evolution and change in productivity that can affect rivalry in the long run, nor do they inform about the availability of an affix when coining new words at a given point in time. In the future, these similarity indices could be examined diachronically and could also be combined with productivity measures to improve the assessment of rivalry.

References

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: the case of English *-ity* and *-ness*. *English Language & Linguistics* 18(3). 497–548.

Dubois, Jean. 1962. *Étude sur la dérivation suffixale en français moderne et contemporain : essais d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Larousse.

Flaux, Nelly & Danièle Van de Velde. 2000. *Les noms en français : esquisse de classement*. Editions Ophrys.

Fradin, Bernard. 2019. Competition in derivation: what can we learn from French doublets in *-age* and *-ment*? In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 67–93. Berlin: Springer.

Guzmán Naranjo, Matías & Olivier Bonami. 2023. A distributional assessment of rivalry in word formation. *Word Structure* 16(1). 86–113.

Haas, Pauline, Lucie Barque, Richard Huyghe & Delphine Tribout. 2022. Pour une classification sémantique des noms en français appuyée sur des tests linguistiques. *Journal of French Language Studies* 1–30. doi:http://doi.org/10.1017/S0959269522000187.

Huyghe, Richard & Rossella Varvara. 2023. Affix rivalry: theoretical and methodological challenges. *Word Structure* 16(1). 1–23. doi:https://doi.org/10.3366/word.2023.0218.

Huyghe, Richard & Marine Wauquier. 2021. Distributional semantics insights on agentive suffix rivalry in French. *Word Structure* 14(3). 354–391.

Lieber, Rochelle. 2016. *English nouns: the ecology of nominalization*, vol. 150. Cambridge University Press.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. *Morphology in Toulouse: Selected Proceedings of Décembrettes* 7. 133–153.

Odum, Eugene P. 1950. Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology* 31(4). 587–605.

Petukhova, Volha & Harry Bunt. 2008. LIRICS semantic role annotation: design and evaluation of a set of data categories. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis & Daniel Tapias (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 39–45. European Language Resources Association.

Plag, Ingo. 1999. *Morphological productivity: structural constraints in English derivation*. Berlin: Walter de Gruyter.

Salvadori, Justine & Richard Huyghe. 2023. Affix polyfunctionality in french deverbal nominalizations. *Morphology* 33(1). 1–39.

Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lüngen & Andreas Witt (eds.), *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*, 28–34. Institut für Deutsche Sprache.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the Web using a new efficient tool chain. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 486–493. European Language Resources Association.

Sørensen, Thorvald A. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskabs Biologiske Skrifter* 5. 1–34.

Thiele, Johannes. 1987. *La formation des mots en français moderne*. Presses de l'Université de Montréal.