

Quantitative measures of affix rivalry

Justine Salvadori Rossella Varvara Richard Huyghe

September 13, 2023

4th International Symposium of Morphology

Affix rivalry

- **Affix rivalry** occurs between affixes that have equivalent semantic functions and compete in the formation of derivatives¹
- E.g., several suffixes can be used to form **agent nouns** in French

(1)	a.	<i>signataire</i> 'signatory'	/agent/
	b.	<i>combattant</i> 'fighter'	/agent/
	c.	<i>déménageur</i> 'mover'	/agent/
	d.	<i>magicien</i> 'magician'	/agent/
	e.	<i>bijoutier</i> 'jeweler'	/agent/
	f.	<i>exorciste</i> 'exorcist'	/agent/

¹ Aronoff (1976); Plag (1999); Fábregas (2010); Arndt-Lappe (2014); Schulte (2015); Fernández-Domínguez (2017); Bonami and Thuilier (2019); Fradin (2019); Gardani et al. (2019); Naccarato (2019); Radimský and Stichauer (2021); Denistia et al. (2022); Huyghe and Varvara (2023); a.o.

Polyfunctionality

- Affixes are rarely strictly equivalent due to their **polyfunctionality**²

	agent	instrument	beneficiary	inhabitant	container	partisan
<i>-aire</i>	✓	–	✓	–	✓	–
<i>-ant</i>	✓	✓	✓	–	–	✓
<i>-eur</i>	✓	✓	–	–	–	–
<i>-ien</i>	✓	–	–	✓	–	–
<i>-ier</i>	✓	✓	–	–	✓	–
<i>-iste</i>	✓	–	–	✓	–	✓

Table 1: Subset of semantic types realized by 6 polyfunctional suffixes in French

² Zwanenburg (2000); Prčić (2019); Salvadori and Huyghe (2023)

Degrees of rivalry

- Different **degrees of rivalry** can be postulated depending on how (dis)similar affixes are³

	agent	instrument	beneficiary	inhabitant	container	partisan
<i>-aire</i>	✓	–	✓	–	✓	–
<i>-ant</i>	✓	✓	✓	–	–	✓
<i>-eur</i>	✓	✓	–	–	–	–
<i>-ien</i>	✓	–	–	✓	–	–
<i>-ier</i>	✓	✓	–	–	✓	–
<i>-iste</i>	✓	–	–	✓	–	✓

Table 1: Subset of semantic types realized by 6 polyfunctional suffixes in French

³ Huyghe and Wauquier (2021); Guzmán Naranjo and Bonami (2023)

Quantifying affix rivalry

- A **coefficient** of competition may be useful to compare situations of rivalry both within languages and cross-linguistically
- **Objective**: Explore measures of **semantic similarity** between polyfunctional affixes that can be used to approach their partial rivalry

Similarity measures

- We consider **two measures** drawn from studies in **ecology**
 - the **Sørensen** index⁴
 - the **Percentage similarity** coefficient⁵
- They both range from **0** (full dissimilarity) to **1** (identity)

⁴ Sørensen (1948)

⁵ Odum (1950)

Similarity measures

- The measures highlight different aspects of **functional similarity**

	Sørensen index	Percentage similarity coefficient
Uses...	Presence/absence data	Abundance data
Considers...	# of distinct functions realized by rival affixes	# of derivatives realizing each function of rival affixes
Quantifies similarity based on...	Proportion of shared functions	Type frequencies

Table 2: Overview of the two similarity measures

Similarity measures

Sørensen index

$$S = \frac{2|A \cap B|}{|A| + |B|}$$

A = set of functions of Affix α

B = set of functions of Affix β

$A \cap B$ = set of functions common to α and β

Percentage similarity coefficient

$$PS = \frac{2 \sum_{i=1}^p \min(N_{i\alpha}, N_{i\beta})}{\sum_{i=1}^p (N_{i\alpha} + N_{i\beta})}$$

$N_{i\alpha}$ = number (i.e., the abundance) of derivatives with Affix α that realize Function i

$N_{i\beta}$ = number of derivatives with Affix β that realize Function i

p = total number of functions observed for α and β

- The Sørensen index (*S*) returns the **same value** when the proportion of shared functions is the same...

...regardless of the **frequency** of realization of functions

	F1	F2	F3	F4	<i>S</i>	<i>PS</i>
Affix A	30	30	30	30	.86	.75
Affix B	40	40	40	0		
Affix A	30	30	30	30	.86	.33
Affix C	110	5	5	0		

Table 3: Number of derivatives per semantic function (F1-F4) and similarity scores (*S*, *PS*) obtained for pairs of rival affixes

Fake data

- The Percentage similarity coefficient (*PS*) returns the **same value** when the ratio between the minimal number of derivatives instantiating shared functions and the total number of derivatives formed with rival affixes is the same...

...regardless of the **number** of shared functions

	F1	F2	F3	F4	S	PS
Affix A	20	20	20	0	.67	.67
Affix B	20	20	0	20		
Affix C	10	10	20	20	1	.67
Affix D	20	20	10	10		

Table 4: Number of derivatives per semantic function (F1-F4) and similarity scores (S, PS) obtained for pairs of rival affixes

Forms of rivalry

- Rivalry comes in different flavors⁶

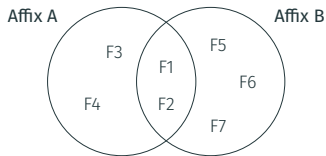


Figure 1: Overlapping rivalry

A and B share only part of their respective functions

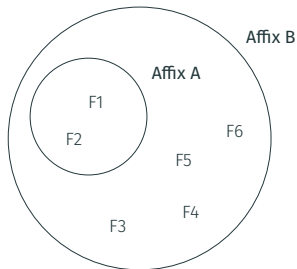


Figure 2: Nested rivalry

All functions of A are also realized by B

⁶ Plag (1999); Guzmán Naranjo and Bonami (2023)

- The Sørensen index and the Percentage similarity coefficient can be complemented with additional information about **dissimilarity structure**⁷
- We consider two **complementary measures**
 - **Balanced richness** (for the Sørensen index)
 - **Balanced abundance** (for the Percentage similarity coefficient)

⁷ Baselga (2013); Legendre (2014)

Complementary measures

	Balanced richness	Balanced abundance
Uses...	Presence/absence data	Abundance data
Considers...	# of distinct functions realized by rival affixes	# of derivatives realizing each function of rival affixes
A score of 0 indicates...	nestedness	nestedness
A score of 1 indicates...	symmetric functional overlap	even distribution of derivatives across unshared functions

Table 5: Overview of the two complementary measures

Complementary measures

Balanced richness

$$BR = \frac{\min(|A \setminus B|, |B \setminus A|)}{\max(|A \setminus B|, |B \setminus A|)}$$

A = set of functions of Affix α

B = set of functions of Affix β

$X \setminus Y$ = the relative complement of Set Y in Set X

$\min(a, b)$ = the smaller of two number a and b

Balanced abundance

$$BA = \frac{\min \left(\sum_{j=1}^q N_{j\alpha}, \sum_{k=1}^r N_{k\beta} \right)}{\max \left(\sum_{j=1}^q N_{j\alpha}, \sum_{k=1}^r N_{k\beta} \right)}$$

$N_{j\alpha}$ = number of derivatives with Affix α that realize Function j

$N_{k\beta}$ = number of derivatives with Affix β that realize Function k

q = total number of functions of α but not of β

r = total number of functions of β not of α

- Balanced richness (**BR**) returns the **same value** when the ratio of unshared functions is the same...

...regardless of the **frequency** of realization of functions

	F1	F2	F3	F4	S	BR	PS	BA
Affix A	20	20	20	0	.67	1	.67	1
Affix B	20	20	0	20				
Affix A	20	20	20	0	.67	1	.67	.10
Affix C	29	29	0	2				

Table 6: Number of derivatives per semantic function (F1-F4), incidence- (S and BR) and abundance-based (PS and BA) scores obtained for pairs of rival affixes

Fake data

- Balanced abundance (**BA**) returns the **same value** when the total number of derivatives instantiating unshared functions is the same...

...regardless of the **distribution** of unshared functions between the two affixes

	F1	F2	F3	F4	F5	F6	S	BR	PS	BA
Affix A	10	30	30	0	0	0	.29	.67	.14	1
Affix B	10	0	0	20	20	20				
Affix C	10	15	15	15	15	0	.29	.25	.14	1
Affix D	10	0	0	0	0	60				

Table 7: Number of derivatives per semantic function (F1-F6), incidence- (S and BR) and abundance-based (PS and BA) scores obtained for pairs of rival affixes

- The **potential** of the 4 measures was explored using real linguistic material, viz. rival suffixes used to form deverbal nouns in French
- 3 **eventive** (*-ade, -ment, -ure*) and 3 **agentive** (*-aire, -ant, -eur*) suffixes were selected
- A random sample of **600 French deverbal nouns** (100 per suffix) was retrieved from the French web corpus FRCOW16A⁸

⁸ Schäfer and Bildhauer (2012); Schäfer (2015)

Semantic analysis

- Each derived noun was analyzed using a **double classification**⁹ that distinguishes between
 - the **ontological** description of the referent
 - the **relation** with the eventuality denoted by the base verb

(2)	a.	<i>bâtir</i> 'build'	→	<i>bâtiment</i> 'building'	/artefact-result/
	b.	<i>raser</i> 'shave'	→	<i>rasoir</i> 'razor'	/artefact-instrument/
	c.	<i>garer</i> 'park'	→	<i>garage</i> 'garage'	/artefact-location/
(3)	a.	<i>bâtir</i> 'build'	→	<i>bâtiment</i> 'building'	/artefact-result/
	b.	<i>énervé</i> 'annoy'	→	<i>énervement</i> 'annoyance'	/state-result/
	c.	<i>créer</i> 'create'	→	<i>créature</i> 'creature'	/animate-result/

⁹ Haas et al. (2022); Salvadori and Huyghe (2023)

Computation of scores

- 61 combined semantic types (i.e., that include an **ontological type** and a **relational type**) were identified
- Semantic types that were observed only once per suffix were removed from the sample to maximize the chances that they correspond to semantic functions
- The 4 measures were applied to the 6 suffixes based on the **782 word meanings/37 functions** identified in the dataset

Results

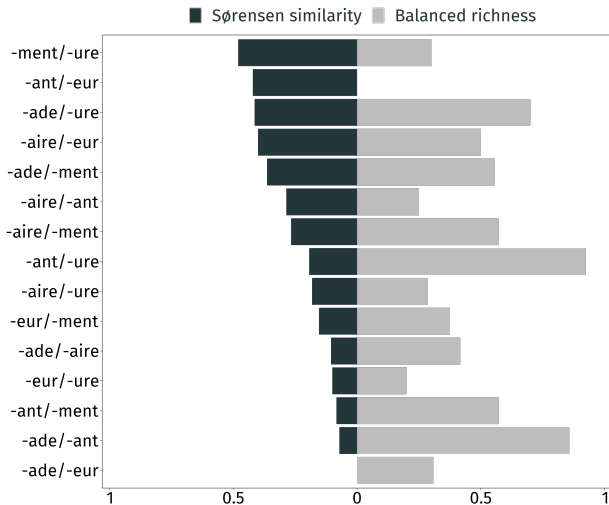


Figure 3: Scores for incidence-based measures (pairs of suffixes are ordered from top to bottom by decreasing similarity)

Results

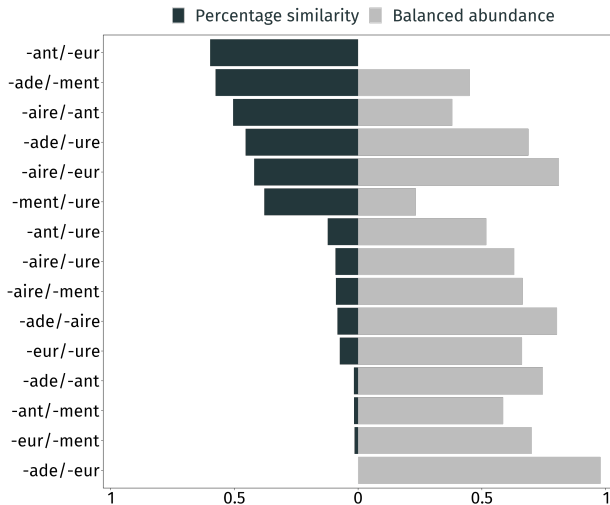


Figure 4: Scores for **abundance-based measures** (pairs of suffixes are ordered from top to bottom by decreasing similarity)

Comparison of similarity scores

- The *S* and *PS* scores are **correlated** (Mantel test: $r = .875$, $p < .01$)
- Suffixes that have many functions in common also tend to present a relatively similar distribution of derivatives across shared functions
- Some qualitative differences can be highlighted

Comparison of similarity scores

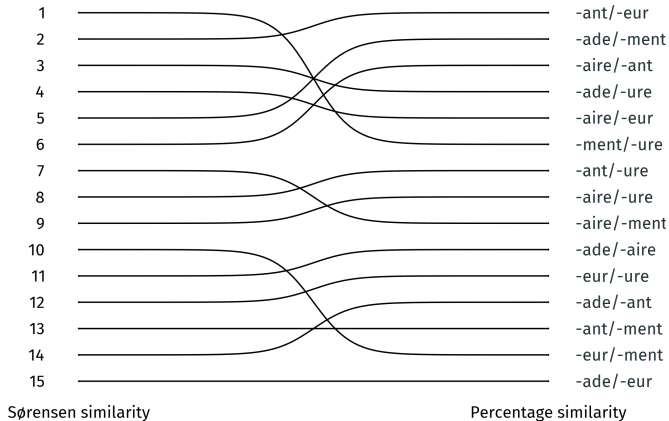


Figure 5: Ranking of the suffix pairs according to *S* vs. *PS*

- The study introduced
 - coefficients of **similarity** between rival affixes (*S*, *PS*)
 - complementary indices to analyze **dissimilarity structures** (*BR*, *BA*)
- The potential of the measures was explored through the analysis of a sample of 600 nouns formed with 6 nominalizing suffixes in French

- The metrics should be considered a **first step** towards a comprehensive measurement of morphological competition
- Further refinement is needed to take into consideration additional factors such as
 - the productivity of word-formation processes¹⁰
 - the co-realization of functions in ambiguous derivatives

¹⁰ Corbin (1987); Plag (1999); Bauer (2001); Fernández-Domínguez (2013)

Thank you!



References i

- Arndt-Lappe, S. (2014). Analogy in suffix rivalry: The case of English *-ity* and *-ness*. *English Language & Linguistics*, 18(3):497–548.
- Aronoff, M. (1976). *Word formation in Generative Grammar*. MIT press, Linguistic Inquiry Monographs, Cambridge, MA.
- Baselga, A. (2013). Separating the two components of abundance-based dissimilarity: Balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, 4(6):552–557.
- Bauer, L. (2001). *Morphological productivity*. CUP.
- Bonami, O. and Thuilier, J. (2019). A statistical approach to rivalry in lexeme formation: French *-iser* and *-ifier*. *Word structure*, 12(1):4–41.
- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Max Niemeyer Verlag.
- Denistia, K., Shafaei-Bajestan, E., and Baayen, R. H. (2022). Exploring semantic differences between the Indonesian prefixes PE- and PEN- using a vector space model. *Corpus Linguistics and Linguistic Theory*, 18(3):573–598.
- Fábregas, A. (2010). A syntactic account of affix rivalry in Spanish nominalizations. *The syntax of nominalizations across languages and frameworks*, pages 67–92.
- Fernández-Domínguez, J. (2013). Morphological productivity measurement: Exploring qualitative versus quantitative approaches. *English Studies*, 94(4):422–447.
- Fernández-Domínguez, J. (2017). Methodological and procedural issues in the quantification of morphological competition. *Competing patterns in English affixation*, pages 67–117.
- Fradin, B. (2019). Competition in derivation: What can we learn from French doublets in *-age* and *-ment*? In *Competition in inflection and word-formation*, pages 67–93. Springer.
- Gardani, F., Rainer, F., and Luschützky, H. C. (2019). Competition in morphology: A historical outline. In *Competition in inflection and word-formation*, pages 3–36. Springer.
- Guzmán Naranjo, M. and Bonami, O. (2023). A distributional assessment of rivalry in word formation. *Word Structure*, 16(1):87–114.
- Haas, P., Barque, L., Huyghe, R., and Tribout, D. (2022). Pour une classification sémantique des noms en français appuyée sur des tests linguistiques. *Journal of French Language Studies*, pages 1–30.
- Huyghe, R. and Varvara, R. (2023). Affix rivalry: Theoretical and methodological challenges. *Word Structure*, 16(1):1–23.

References ii

- Huyghe, R. and Wauquier, M. (2021). Distributional semantics insights on agentive suffix rivalry in French. *Word Structure*, 14(3):354–391.
- Legendre, P. (2014). Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, 23(11):1324–1334.
- Naccarato, C. (2019). Agentive (para)synthetic compounds in Russian: A quantitative study of rival constructions. *Morphology*, 29(1):1–30.
- Odum, E. P. (1950). Bird populations of the Highlands (North Carolina) Plateau in relation to plant succession and avian invasion. *Ecology*, 31(4):587–605.
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. De Gruyter Mouton.
- Prčić, T. (2019). Exploring the properties of English lexical affixes by exploiting the resources of English general-purpose dictionaries. *Lexikos*, 29:151–179.
- Radimský, J. and Stichauer, P. (2021). Nomina actionis in the diachrony of Italian: A paradigm-based model of competition. *Lingue e linguaggio*, 20(1):33–55.
- Salvadori, J. and Huyghe, R. (2023). Affix polyfunctionality in French deverbal nominalizations. *Morphology*, 33.
- Schulte, M. (2015). *The semantics of derivational morphology: A synchronic and diachronic investigation of the suffixes -age and -ery in English*. Narr Dr. Gunter.
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In Bański, P., Biber, H., Breiteneder, E., Kupietz, M., Lungen, H., and Witt, A., editors, *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*, pages 28–34. Institut für Deutsche Sprache.
- Schäfer, R. and Bildhauer, F. (2012). Building large corpora from the Web using a new efficient tool chain. In Calzolari, N., Choukri, K., Declerck, T., Dogan, M. U., Maegaard, B., Mariani, J., Odiijk, J., and Piperidis, S., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 486–493. European Language Resources Association.
- Sørensen, T. 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.
- Zwanenburg, W. (2000). Correspondence between formal and semantic relations. In Booij, G. E., Lehmann, C., and Mugdan, J., editors, *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An international handbook on inflection and word-formation*, pages 840–850. Walter De Gruyter.