A distributional assessment of rivalry in word formation

Matías Guzmán Naranjo
Albert-Ludwigs-Universität Freiburg
mguzmann89@gmail.com

Olivier Bonami
Université Paris Cité, LLF, CNRS
olivier.bonami@u-paris.fr

Accepted in *Word Structure*

Abstract

We contrast two views of rivalry in word formation. Under the classical, categorical view, two processes are rivals if they are semantically equivalent. Under the more nuanced, gradient view, two processes can be rivals at different degrees, depending on how frequently they are amenable to be deployed as alternatives to one another.

We propose to use methods from distributional semantics to explore the usefulness of both views. Building on data from French, we first show that distributional differences between average difference vectors capture semantic similarity across derivational processes in a manner comparable to the expectations of expert morphologists. We then propose an operational implementation of the classical view of rivalry based on computational classifiers: processes are rivals if and only if a classifier is unable to discriminate between them. Experimentation with French data shows that this operationalization correctly captures the broad brushes of rivalry, but also reveals finer gradient aspects of competition in the spirit of gradient rivalry.

Keywords: word formation, rivalry, distributional semantics, French

1 Introduction

Much recent literature on word formation (including research reported on in the present issue) documents the division of labor between *rival* morphological processes in competition, showing how phonological, morphological, lexical, semantic, pragmatic, and/or sociolinguistic factors influence the choice of one of the rivals (see among many others Aronoff, 1976; Plag, 1999; Lindsay and Aronoff, 2013; Arndt-Lappe, 2014; Bonami and Thuilier, 2019). These studies generally take for granted which processes should count as rival, with little discussion (see Gardani, Rainer, and Luschützky 2019 and Huyghe and Wauquier 2021 for notable exceptions). Yet the definition of rivalry is notably elusive. We discuss briefly three possible and contrasting views.

When coining the use of the term 'rival' to name competition between morphological processes, Aronoff (1976, p. 37) had a very restrictive view, where he intended rivals to be pairs of processes which differ only in their productivity, and hence differ neither in the semantics they convey nor in the syntactic properties of their inputs and outputs. This position is nicely summarized by Plag (1999, p. 227) as follows:

(1) In general, morphological processes are regarded as rival if they are phonologically distinct but semantically identical.

Notice that on this view, there is a categorical distinction between rival and non-rival processes, entailed by the categorical distinction between semantic identity and semantic difference. While this has the advantage of conceptual clarity, practical concerns immediately arise. It is hard and contentious enough to decide whether two words are true synonyms, it seems hopeless to provide a clean empirical argument to the effect that processes have exactly the same semantics. In addition, given that word formation processes are often polysemous, Plag's early definition sets aside situations of what we call *partial rivalry*, where the set of meanings that two processes can convey overlap without being identical. As a case in point, Plag (1999) documents at length the nested semantics of English denominal verb formation processes, as indicated in Table 1: according to his classification, *-ize* and *-ify* can convey the exact same set of 7 meaning types, and hence count as rivals; they contrast with *-ate*, which is more restricted, and conversion, which is less restricted. Cases can also be documented where there is an *overlap* rather than a nesting between the possible

Morphosemantic Type	Conversion	*-ize*	*-ify*	*-ate*
Locative	✓	✓	✓	
Ornative	✓	✓	✓	✓
Causative	✓	✓	✓	
Resultative	✓	✓	✓	✓
Inchoative	✓	✓	✓	
Performative	✓	✓	✓	
Similative	✓	✓	✓	
Instrumental	✓			
Privative	✓			
Stative	✓			

Table 1: Nested rivalry in English derived verbs according to Plag (1999).

semantics of processes. Consider *-eur* and *-oir* deverbal nouns in French, as exemplified in Table 2: *-eur* readily derives agents or instruments (Huyghe and Tribout, 2015), while *-oir* derives instruments or locations (Luschützky and Rainer, 2013). Given that partial rivalry raises exactly the same

Process	Agent	Instrument	Location
-eur	*nager>nageur* 'swim' 'swimmer'	*tracter>tracteur* 'tow' 'tractor'	—
-oir	—	*hacher>hachoir* 'chop' 'chopper'	*laver>lavoir* 'wash' 'washhouse'

Table 2: Overlapping rivalry in French deverbal nouns in *-eur* and *-oir*.

family of problems as 'full' rivalry, Plag's early definition turns out not to delimit a very useful empirical domain.

Such observations have led some to move from a maximally restrictive to a maximally inclusive definition of rivalry, where it is enough for polysemous processes to share *some* of their meanings to count as rivals. A cogent formulation is provided by Bauer, Lieber, and Plag (2013, p. 33):

(2) Two processes compete [i.e., are rivals] when they both have the potential to be used in the coining of new synonymous forms from the same base.

The characterization in (2) clearly improves on (1) by not requiring general synonymy of the processes, but focussing on particular outcomes of their application. We still have a categorical distinction, but it is based on an existential requirement (there is some context where both processes can apply to express the same meaning) rather than a universal one (every meaning that can be expressed by one process can be expressed by the other and vice versa). However it is hard to see how it can be made operational. (2) crucially depends on *potential* rather than *actual* coining. Hence, where actual attested synonymous doublets (Fradin, 2019) are found, this is strong evidence for rivalry, but the absence or rarity of such doublets is not sufficient to conclude that there is no rivalry: we could have a full complementary distribution between two processes. Where such a complementary distribution is principled (say, because the two processes put contradictory phonological requirements on the base), one cannot sustain the claim that the processes have the potential to apply to the same base; they would hence not count as rivals according to (2).

This observation suggests a variant of the definition in (2) which embraces the gradient nature of similarity between processes rather than aiming for a categorical distinction. Such a view comes out naturally from an onomasiological perspective on word formation. As e.g. Štekauer (2005) highlights, coining happens when a language user builds on the lexical and morphological resources of the language to name a concept with a new word. The concept to be named is determinate for the coiner, but the resources may be more or less fit to the task: different bases paired with different processes might be appropriate, and no one solution need be optimal in all dimensions. Rivalry then happens whenever two processes could apply to some base (not necessarily the same one) with nonzero probability to name the target concept. We can then define the degree of rivalry between two processes as follows:

(3) The degree of rivalry between two processes is the proportion of coining events where both processes could apply among coining events where either process can apply.

Under such a gradient view, maximally competing processes such as *-ize* and *-ify* will have a degree of rivalry close to 1, while completely nonrival processes, e.g. *-ize* and *-able*, which never output lexemes in the same part of speech, will have a degree of rivalry of zero. In between, the definition accomodates partial rivaly as intermediate values, with appropriate granularity. Consider for instance a telling example due to Roché (1997): the established French term for a truck driver, *camionneur*, relies on applying the normally deverbal suffix *-eur* to the noun *camion* 'truck'. The expected suffix here would have been the denominal agent suffix *-ier*. Roché argues convincingly that this would have led to a form *camionnier* that is disfavored by the dissimilative tendencies of French morphophonology, because it has a /j/ in two successive syllables; hence language users arbitrated in favor the 'wrong' suffix in terms of semantics in the interest of having a phonologically better form. Now, clearly, both *-eur* and *-ier* have the potential to be used here: both are attested, and the more unlikely one, namely *-eur*, is the lexicalized outcome. Under the definition in(3), *-eur* and *-ier* would thus have a low but nonzero degree of rivalry.

While all three alternative definitions of rivalry above are coherent, the question is which is most opportune. For instance, maybe the empirical fact of the matter is that processes overlap either fully or very little in their potential scope, in which case the maximalist definition would capture all there is to capture. On the other hand, maybe there is evidence for fine degrees of rivalry, which would justify using the more elaborate gradient definition.

In this paper we propose to approach the identification of rival processes using computational tools from distributional semantics, and assess whether a focus on categorical rivalry as defined in (1) is warranted. In section 2, we justify operationalizing the semantic difference between a derivative and its base as the difference between their vector representations in a distributional vector space – in other words, the path in semantic space to go from the base meaning to the derived meaning. Comparing such difference vectors across word formation processes is a way of assessing their similarity that is inherently gradual, takes into account the relative type frequency of meanings associated with a process, and, being fully automated, can readily be applied on a large scale in a consistent fashion. After presenting the French materials we will use in Section 3, we move on in Section 4 to justifying the claim that difference vectors capture semantic similarity among word formation processes, by comparing similarity across vectors with the opinion of expert morphologists.

Section 5 puts the different views of rivalry to the test. We operationalize a maximalist, categorical view of rivalry in the spirit of (1) by asking whether a computational classifier is capable of telling processes apart by just looking at difference vectors. We show that this method gives results that are broadly compatible with received wisdom on rivalry in French, but also documents gradient structure of theoretical interest.

2 Distributional semantics

We approach the question of rivalry and semantic similarity with the help of distributional semantics. Although it is by no means a new idea (Harris, 1954; Firth, 1957), progress in the development of large lexica and efficient algorithms for inferring word vectors (e.g. Mikolov et al. 2013; Pennington, Socher, and Manning 2014; Bojanowski et al. 2016; or Camacho-Collados and Pilehvar 2020 for an overview) have renewed the interest in distributional semantics within linguistics (Evert, 2014; Turney and Pantel, 2010; Turney, 2012; Miller and Charles, 1991; Erk, 2012; Boleda, 2020) and in particular to explore morphological issues (Padó et al., 2016; Varvara, Lapesa, and Padó, 2021; Guzmán Naranjo and Bonami, 2021; Wauquier, Hathout, and Fabre, 2020; Huyghe and Wauquier, 2020; Amenta, Günther, and Marelli, 2020). The approach crucially relies on the distributional hypothesis, which Lenci (2008, p. 3) states as follows:

(4) The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear.

In practice, what this means is that the meaning of a word can be approximated by information about the words it tends to cooccur with in a corpus. A simple example of this idea is presented in Table 3. In this example we are capturing the meanings of the words *student*, *computer* and *car* as vectors of coocurrences with the verbs *crashes* and *reads* in an tiny imaginary corpus. The rows in the table are in effect two-dimensional vectors. If we represent these graphically as in Figure 1, we clearly see that the distributions of *computer* and *car* are more similar to one another than either is to the distribution of *student*; this captures the intuition that the words *computer* and *car* are more semantically similar to each other than either of them is to *student*. We also see that the distribution of *computer* is somehow intermediate between those of *student* and *car*, highlighting the fact that computers share properties with humans that cars don't. These visual impressions can be made explicit by computing the cosine similarity S, which is just the cosine of the angle between the vectors. In our toy example, $S(\vec{student}, \vec{computer}) = 0.61$, $S(\vec{computer}, \vec{car}) = 0.89$, and $S(\vec{student}, \vec{car}) = 0.20$.

	crashes	reads
student	1	5
computer	4	2
car	4	0

Table 3: Cooccurrence counts in a small toy corpus.

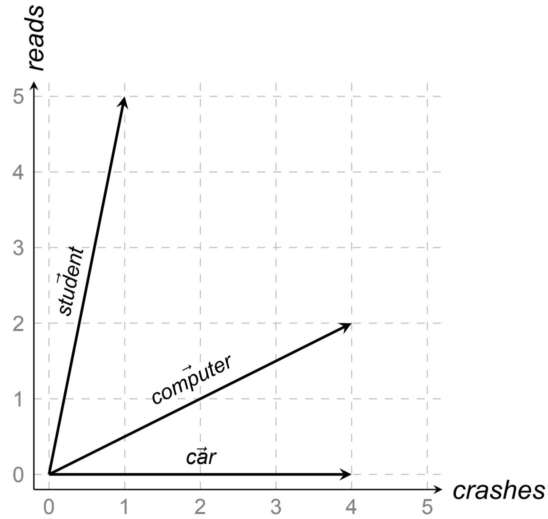

Figure 1: Graphical representation of the cooccurrence vectors from Table 3.

Early approaches to distributional vectors essentially amounted to doing what we just showed with our toy example on a much larger scale, tracking the co-occurrence of many many words with many many other words, leading to vectors of very high dimensionality. A crucial progress has been the design of algorithms that use neural networks to induce vectors as a byproduct of a prediction tasks, rather than directly by counting co-occurrences (see e.g. Mikolov et al., 2013). Such predictive techniques are computationally more efficient, and also turn out to model semantics more accurately (Baroni, Dinu, and Kruszewski, 2014). The downside is that vector dimensions are not directly interpretable in linguistic terms anymore (Boleda, 2020). Since we are not interested in interpreting the vectors directly, this is not a big issue for the present study. What matters for us is that distributional vectors provide for a systematic assessment of semantic similarity that can be derived automatically from a corpus at low computational cost.

We now turn to the use of distributional semantics in the study of word formation. In a seminal study, Marelli and Baroni (2015) propose that the semantic import of a derivational process be modelled as a function mapping vectors for base lexemes to vectors for the corresponding derived lexemes. Their specific proposal is that this function is a linear transformation: each dimension of the derived vector is deduced from a linear combination of the values of the base vector in all dimensions. That function can then be approximated from the data by a series of multivariate linear regressions.

A slightly different view of derivational semantics builds on pairwise comparisons of vectors for pairs of derivationally related words. This is illustrated in Figure 2 with a toy example. In this example, we calculate the difference vectors between *washable* and *wash*, and *drinkable* and *drink*. Because these pairs of words are related by the same process, we expect the difference vectors to be similar to one another (Bonami and Paperno, 2018). Notice that this is true despite the fact that the vectors for *washable* and *drinkable* are quite different from one another, as are those for *wash* and *drink*: the point is not that derivatives using the same process resemble each other, but that the path in semantic space between base and derivative is similar across pairs of words. Also notice that while the two difference vectors are very similar, they are not identical: the semantic effects of derivation varies to some extent across instances of a process, if only because lexicalization may result in the meaning of the derivative to shift in unpredictable ways.[1] However, if the process is not polysemous, we expect difference vectors to be broadly similar.

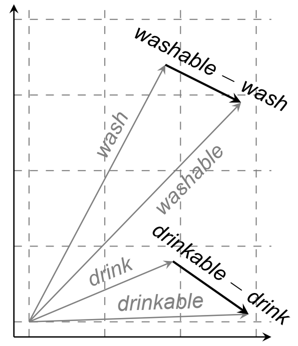

Figure 2: Exemplification of difference vectors relating distributional representations of verbs and derived *-able* adjective.

To get a representation of the semantic effect of a process, we may then average across difference vectors. Averaging across all derivation instances of a derivation process allows us to wash out the quirks of individual derivation instances and arrive at a more accurate representation of the semantics of the process.[2]

In a second step, we can compare average difference vectors for various processes, as a way of assessing how similar these processes are. Figure 3 illustrates this with a toy example, with three average difference vectors for *-ize*, *-ify* and *-able*. Since difference vectors are still semantic vectors, we can apply the same idea of measuring the semantic similarity using cosine. In this example, we expect that the vectors for *-ify* and *-ize* should be more similar to each other (and thus have a higher cosine similarity value), than either of them to the difference vector for *-able*. If this is correct, then we should be able to explore the similarity between derivational processes using distributional semantics.

[1] In fact, Bonami and Paperno (2018) show that the amount of variability across pairs of different vectors is measurably higher in derivation than in inflection.

[2] A possible downside of this method compared to Marelli and Baroni's is that it does not take into account process polysemy. If a process has two main meanings and all derivation instances pick one or the other meaning, then the corresponding difference vectors should cluster in two discrete groups, and their average will point to a direction in vector space which is not a good approximation of any individual vector. The extent to which that is a problem is hard to evaluate though, given that polysemous processes typically give rise to output words that are themselves polysemous; e.g. English *-er* outputs agents or instruments, and many *-er* derivatives are actually ambiguous between an agent and an instrument reading. In such circumstances, linear transformations are just as inadequate for capturing polysemy as average difference vectors, since the word vectors themselves average across meanings. Be that as it may, we rely on averaging difference vectors only for the purposes of our first experiment. The classifiers in experiment 2 build on individual difference vectors rather than averagesx, and hence are not biased by a washing out effect of averaging.

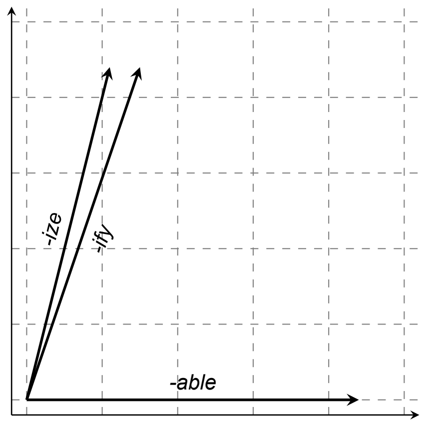

Figure 3: Representing the semantics of derivational processes as average difference vectors.

3 Materials

For this work two resources were needed: a collection of distributional vectors induced from a corpus, and a collection of pairs of derivatives and their bases.[3]

3.1 Distributional data

We used a modified version of the FRCOW16 corpus, an 8.8 billion word web corpus of French compiled in 2017 (Schäfer, 2015; Schäfer and Bildhauer, 2012). Our goal was to derive from this vectors for *lexemes* rather than *words*: as we are focusing on word formation, we want to ignore distributional differences between inflected forms of lexemes, and focus on the distribution of lexemes in the context of other lexemes. To that effect, we set out to use the lemmatization and part of speech (POS) tagging provided with the corpus so as to merge as instances of the same lexeme all tokens with the same lemma and POS tag.

A pilot study showed that grammatical gender was strongly represented in vectors derived from FRCOW, leading to artificially increasing the similarity of nouns of the same grammatical gender, irrespective of semantics. Through experimentation, we found that the lemmatization provided with the corpus was responsible for this bias. We hence corrected the lemmatization as follows.

First, French has a number of portmanteau words realizing as a single item a preposition and some element carrying masculine gender (e.g. *du*, substituting for the combination of preposition *de* 'of' and the masculine singular definite article *le*) where the feminine equivalent consists of two words (e.g. *de la*). As a result, the word *du* is lemmatized as a single lemma *du*, which carries gender information, leading to grammatical gender being a very salient distributional property of nouns. Since our focus is semantics, this is undesirable. We solved the issue by creating a doctored version of the corpus where masculine portmanteau words were systematically analyzed into their component parts (e.g. every *du* is rewritten as *de le*) and given gender-neutral tags.

Second, for out of vocabulary items, the original lemmatization falls back to assuming that the lemma is identical to the inflected wordform. This causes again gender-related problems: rare feminine adjectives and past participles will often be assigned a faulty lemma distinct from that of their masculine counterpart. We solved this by automatically correcting lemmas for feminine adjectives, based on known regularities on the relation between the masculine and feminine forms (Bonami and Boyé, 2005).

[3]The vectors and lexical data are available from the following repository: `https://zenodo.org/record/5799577`.

Third and finally, the original lemmatization is inconsistent in its treatment of paired masculine and feminine nouns such as *directeur* 'male director' and *directrice* 'female director'. Sometimes they share the same lemma, corresponding to the masculine form; sometimes they have distinct lemmas, with no consistency. Whether such pairs of nouns should be considered forms of the same lexeme or distinct lexemes is a contentious issue when dealing with human-denoting nouns (Bonami and Boyé, 2019), but there is no such hesitation when at least one of the two nouns is inanimate: e.g. *batteur* 'whisk' vs. *batteuse* 'thresher'. Hence we opted for the only option that was applicable at scale, namely systematically treating masculine and feminine nouns with different forms as instances of different lexemes; accordingly, all feminine nouns were re-lemmatized to a lemma corresponding to their singular form.

We then set out to derive vectors from the modified corpus, replacing each word by the concatenation of its lemma and POS tag. To this effect we used the Gensim (Řehůřek, 2010) implementation of word2vec (Mikolov et al., 2013) to build the vector space. Hyperparameters were not optimized for lack of time; verification of the stability of results across parameter configurations will have to wait for a future study.[4]

3.2 Morphological data

In addition to the vectors, we need information about base-derivative relations within the French lexicon. An appropriate large-coverage database is currently being compiled within the Démonext project (Namer et al., 2019), but was not available yet at the time our experiments were conducted. Hence we compiled an ad-hoc dataset using a number of existing sources: Hathout and Namer (2014) for deverbal nouns; Tribout (2010) for nouns and verbs related by conversion; Koehl (2012) for deadjectival nouns; Strnadová (2014) for derived adjectives; and Bonami and Thuilier (2019) for derived verbs in *-iser* and *-ifier*.

The final dataset contained 21990 (base, derivative) pairs. We focused on the processes for which at least 50 pairs were present in the resources such that both words are attested at least 50 times in the FRCOW16 corpus: this ensures that we have enough tokens of each word for the vectors not to be too noisy, and enough instances of each type for erratic individual differences between pairs to balance each other out. Finally, we removed pairs for which we did not have semantic vectors.[5]

This resulted in a dataset of 20527 pairs exemplifying 35 processes, as shown in Table 4. The layout of the table groups processes with the same output part of speech (adjectives on the left, nouns in the middle, verbs on the right), and, within each subtable, processes with the same input part of speech. Among verb to noun processes, one can draw a broad semantic distinction between two classes: the proceses at the top of the subtable (*age*, *ance*, etc.) mostly output nouns denoting eventualities (e.g. *repassage* 'ironing'), although they sometime denote other entity types (e.g. locations *garage* 'garage', artefacts *cirage* 'shoe polish'); those in the middle (*-eur*, *-euse*, *-rice*) mostly denote either agents (e.g. *chauffeur* 'driver') or instruments (e.g. *compteur* 'meter, counter'), although other types are attested (e.g. possessors: *détenteur* 'owner'). Such lexical semantic distinctions are not implemented in the resources we used, and hence all derivatives of the same morphological types were lumped together for analysis, irrespective of their exact semantics.

Note that we treat separately situations where the same affix is used with bases and/or derivatives of different parts of speech: hence we distinguish for instance *-eur* forming deverbal nouns or

[4]Hyperparameters: 5 training epochs, 5 negative samples, window size 5, vector size 100, and skipgram representation.
[5]The slight discrepancy in type frequencies is because of how we built the semantic vectors.

Process	Pairs
-aire:N>A	410
-al:N>A	445
CONVERSION:N>A	99
-el:N>A	276
-eux:N>A	399
-ien:N>A	81
-ier:N>A	162
-if:N>A	373
-ique:N>A	1719
-ant:V>A	832
-eur:V>A	297
-é:V>A	651
PST.PART:V>A	190
-if:V>A	134
-Vble:V>A	323
-ième:NUM>A	43

Process	Pairs
-age:V>N	1625
-ance:V>N	94
CONVERSION:V>N	2329
-ée:V>N	60
-erie:V>N	87
-ion:V>N	1951
-ment:V>N	1289
-ure:V>N	74
-eur:V>N	1563
-euse:V>N	505
-rice:V>N	167
-erie:A>N	86
-ité:A>N	1084
-itude:A>N	62
-té:A>N	79

Process	Pairs
-iser:A>V	374
-ifier:A>V	50
CONVERSION:N>V	2337
-iser:N>V	284

Table 4: Derivation processes forming adjectives (left), nouns (middle) and verbs (right).

adjectives, as well as *-iser* forming denominal or deadjectival verbs. Also note that when in doubt, we erred in the direction of making more, rather than fewer, distinctions. For instance, we treat *-té* and *-ité* as distinct, despite the fact that they are usually seen as allomorphs (Koehl, 2012); and we separate eventive deverbal nouns in *-ée*, such as *montée* 'rise', from conversions from past participles of irregular verbs, such as *conduite* 'conduct', despite the fact that Tribout (2012) argues convincingly that both are instances of conversion from the past participles. As we will see, these deliberate decisions will allow us to stress-test our models in an interesting way.

4 Experiment 1: Assessing semantic similarity

With this data in hand, we first explore whether difference vectors for derivational process actually contain semantic information about those derivational processes, and whether that information matches up with what human experts would consider to be semantically similar.

To answer this question, we first averaged across the difference vectors of each process. We do this by taking the mean value for each dimension across all individual vectors of each derivational process. This results in an average difference vector for each process under examination. These average vectors represent the average distributional shift between lexemes related by a process, and are hypothesized to reflect the semantics of that process.

Having calculated the average difference vectors, we then computed pairwise cosine similarities of these vectors and deduce a cosine distance matrix.[6] We then proceeded to perform agglomerative clustering based on that distance matrix, to assess the similarity structure across processes, under the assumption that this structure has the shape of a tree. The result can be seen in Figure 4.[7]

The resulting clustering seems qualitatively reasonable. Processes with the same input and output part of speech cluster together, with two exceptions. First, deverbal adjectives in *-eur* are

[6]Cosine distance is the complement of cosine similarity: $D(\vec{v}, \vec{w}) = 1 - S(\vec{v}, \vec{w})$.

[7]We report here the results of clustering with complete linkage. Experiments with the six other linkage criteria implemented in the `hclust` function in R led to nonidentical but very similar dendrograms. See footnote 11 for details.

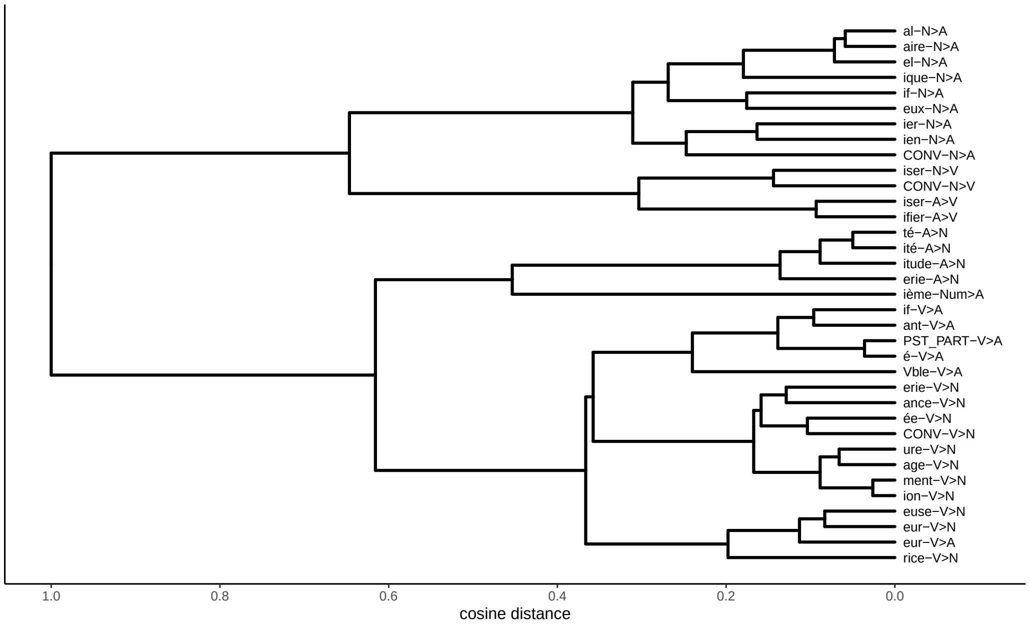

Figure 4: Cosine distance clustering across average difference vectors (0=identical, 1=maximally dissimilar).

grouped with deverbal nouns. This is probably due to noise in the data: the vast majority of adjectives in *-eur* are homopgraphic and morphologically related to a noun in *-eur* (e.g. *moteur* 'engine' (noun) vs. 'driving' (adjective)), leading to widespread tagging errors. Second and more interestingly, we see separate clusters, among deverbal nouns, for eventive vs. agent/instrument nouns: these two subclasses are not more similar to one another than they are to deverbal adjectives. This nicely matches the intuition that the two kinds of deverbal nouns strongly differ.

Within each (input POS, output POS) cluster, groupings also make sense. : notice in particular that *-té* and *-ité* on the one hand, and *-é* and verb to adjective conversion on the other hand, are tighly similar. This is satisfactory, as these are precisely cases where received wisdom would have taken the exponents to be allomorphs of one another and hence assumed a single process rather than two. The suffix *-ée* and verb to noun conversion also form a cluster, although a slighly less tight one: this is due to the fact that we lumped together all conversions, irrespective of whether they were based on the past participle or another verbal stem.

While such qualitative observations are reassuring, it is difficult to be more precise as to how accurate the similarity clustering is, in the absence of a well established gold standard to compare it to: there is no preexisting consensus on which processes is more similar to which. In order to better assess the quality of our results, we set out to ask expert morphologists their opinion. To this effect, we wrote to 21 professional French morphologists, asking them to draw a tree describing semantic similarities and differences between the 35 processes, on the basis of a random sample of 10 pairs of lexemes for each process.[8] Out of the 21 experts we wrote to, 7 provided full trees.[9]

We now need a way to compare the trees produced by experts among themselves and to the tree derived from the vectors. To this end we use the formula given in (5).[10] To find the similarity

[8]Our inclusion criteria were that the experts should hold or be preparing a PhD in linguistics, be native or near native speakers of French, and have published on French derivational morphology.

[9]Instructions and expert trees are available from the following repository: `https://zenodo.org/record/6452950`. The relatively low number of answers is certainly due in part to the fact that we contacted people in the middle of the COVID-19 pandemic. We are all the more grateful to the colleagues who took the time to answer our request.

[10]For this we used dendextend by Galili (2015). Other dendrogram comparison measures were considered but discarded

between two trees T and T', we first determine the set of clusters of that tree; that is, the set of collections of leaf nodes that are dominated by a nonleaf node other than the root. The similarity between the two trees is then given as twice the number of clusters that the two trees share, divided by the summed number of clusters in each tree. This will give us a number between 0 (if the trees share no cluster, i.e. do not agree on any grouping), and 1 (if the trees are identical).

(5)
$$\mathrm{sim}(T,T') = \frac{2 \times |\mathrm{clusters}(T) \cap \mathrm{clusters}(T')|}{|\mathrm{clusters}(T)| + |\mathrm{clusters}(T')|}$$

For a concrete example, consider the two trees in Figure 5. T contains the two clusters $\{a, b\}$ and $\{c, d\}$, while T' contains the two clusters $\{b, c, d\}$ and $\{c, d\}$. As the two trees have exactly one cluster in common out of two, $\mathrm{sim}(T, T') = \frac{2\times1}{2+2} = 0.5$.

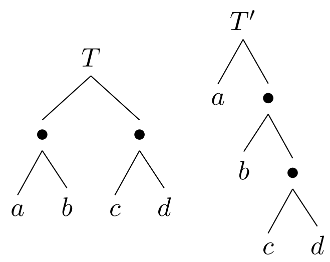

Figure 5: Example of tree similarity measure.

Applying this metric to the seven trees from the experts, and our tree induced from vectors produces the similarity matrix in Figure 6, where darker color indicates stronger similarity. It is immediately striking that there is high variability among experts: while some, e.g. experts 4 and 5, produced very similar trees, others, e.g. experts 6 and 7, diverged strongly. This suggests that the task we are asking experts to carry out is far from obvious, and that experts are not entirely basing themselves on received wisdom to answer.

From examination of the trees and knowledge of the French system, a number of likely causes for these divergences can be established. First, experts vary in the granularity of their classification: some produced binary branching trees, others provided coarser classifications in large groupings. Clearly the broader classes are more consensual than finer distinctions, although there is no obvious way to operationalize that impressionistic judgement. Second, the academic background of the experts seems to have some influence: unsurprisingly, the narrow expertise of an expert colors their choices. Third, the similarity structure of the system is arguably more intricate than a tree can capture, with cases where processes A and B are most similar in one descriptive dimension but B and C are in another. A trivial example of this is taking into account base and derivative part of speech: are deverbal nouns more similar to deverbal adjectives or to deadjectival nouns? In such a situation, and in the absence of explicit instructions, different experts may have chosen to privilege one or the other aspect, leading to different trees.

Against this background, it is clear that we should not place too much value on minute differences between trees. On the other hand, it is striking that the tree induced from vectors does not stand out, either qualitatively or quantitatively. In particular, the similarity values to that tree are

after examination of their results when applied to the expert trees. The problem is that most tree similarity measures are very sensitive to the distance in the tree between pairs of leaf nodes, which in turn is sensitive to the tree's average degree, i.e. the average number of siblings for a parent. As it happens, a major difference among expert trees is that some of them produced a rigidly binary branching tree, while others didn't. Among those that were considered, the reported similarity measure minimizes the influence of such differences.

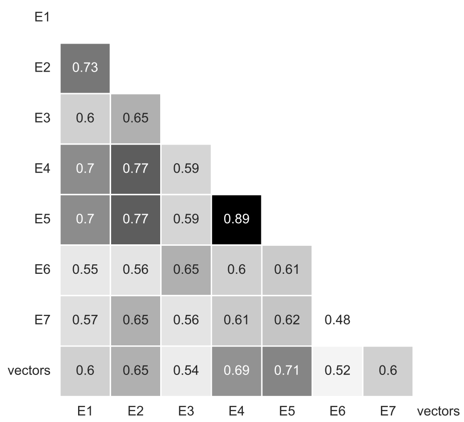

Figure 6: Similarity between trees produced by experts vs. vectors. Darker color indicates stronger similarity.

Tree	E6	E7	E3	**vectors**	E1	E2	E4	E5
Avg. sim.	0.567	0.584	0.597	**0.616**	0.636	0.683	0.693	0.699

Table 5: Average similarity of one tree to all other trees.

within the middle of the distribution of the values found with human experts. This becomes even clearer when we average the similarity of one tree to each of the other trees, as shown in Table 5. The vector-based tree stands in the middle, and thus is completely unremarkable.[11] Hence we may conclude that our classification of processes automatically inferred from distributions compares favorably to those of human experts.

To sum up, our first experiment shows that the difference vectors we calculated do capture information about derivational processes, and that this information is close to what experts in French morphology consider relevant for assessing the similarity of derivational processes. This also means that we have an automated method for systematically measuring semantic similarity in derivational morphology.

5 Experiment 2: Finding rivals

5.1 Methodology

Having established that difference vectors reasonably captures the similarity structure of word formation, we turn back to the issue of identifying rival processes. Remember from the introduction that we want to assess whether a categorical definition of rivalry based on semantic identity can be operationalized, and the extent to which it captures relevant aspects of competition among processes.

We propose to operationalize a categorical definition of rivalry in terms of semantic discriminability. If two processes are such that, from looking at the semantics of a base and that of its derivative, it is (on average) not possible to guess which of the two processes was used to coin the derivative, then clearly we are dealing with rivals.

[11] Using other clustering methods made very little difference in the distance between the vector-derived tree and the expert trees, which ranged from 0.61 to 0.63. This is unsurprising because all vector-derived trees are quite similar to each other, with the distance between them ranging from 0.81 to 0.88.

One advantage of that means of identifying rival processes is that it can easily be assessed in distributional semantic terms. Consider the concrete situation of deadjectival verbs in *-ify* and deverbal adjectives in *-able* in English. The semantic difference between these two processes is considerable: bases are very different, derivatives are very different. As a result, we expect difference vectors between *-able* derivatives and their bases to generally point in a direction in vector space quite different from the direction in which difference vectors between *-ify* derivatives and their bases point. As a result, by just looking at the vector relating two words, say *read* and *readable*, one should be able to guess that it represents an instance of *-able* derivation rather than *-ify* derivation; and correspondingly one should be able from the vector relating *solid* to *solidify* to guess that it represents an instance of *-ify* derivation rather than *-able* derivation. Now let us consider in turn deadjectival verbs in *-ify* vs. *-ize*: this time, our expectation is that the semantic difference between the two processes is at best very small. As a result, we do not expect to be able, from just looking at the difference vector between *solid* and *solidify*, to guess that this is an instance of *-ify* rather than *-ize* derivation.

We operationalize this idea using computational classifiers. A classifier is a program that learns from exposure to data to predict the value of a categorical variable from a series of predictor variables. In our case, the predictor variables are the dimensions of the difference vector between a base and a derivative; the predicted variable is the identity of the process that relates the two words whose vectors we compared. Assuming that the classification method is efficient, the accuracy of a classifier, i.e. the proportion of cases where the classifier predicts the correct answer, is indicative of how hard the classification task is. In our case, if a classifier performs clearly better than chance, we will conclude that the processes are not rivals, as it is possible to discriminate their semantic effects; if a classifier does not perform clearly better than chance, then we are dealing with rivals.

There are dozens of existing classification methods that might have been applied to the problem at hand. In the present study we decided to use Boosting Trees (Chen and Guestrin, 2016), which have proven to have good performance when applied to morphological tasks (Guzmán Naranjo and Bonami, 2021; Bonami and Pellegrini, 2022). Boosting Trees are similar to random forests (Breiman, 2001) in that they fit many simple classification trees to different portions of the data and then combine them to create a much stronger classifier. Unlike random forests, however, Boosting Tree classifiers proceed sequentially, fitting each tree to the residuals of the previous one.

Concretely, we built classifiers for each pair of base part of speech and derivative part of speech in our dataset (V>N, V>A, A>V, A>N, N>V, N>A). For each derivation instance of a base lexeme and derived lexeme, we predict which process relates them from the 100 dimensions of the corresponding difference vector. To avoid overfitting, we perform 10-fold cross-validation. This means that we split our data into 10 groups, and fit 10 different models, each time leaving out one of the groups and then predicting the left-out group. We report the aggregated results of the 10 models on all the left-out data: hence we have a prediction for each of our data points, but that prediction stems from a model that was not trained on that data point.

To evaluate the classifiers we report the overall accuracy and the predictive baseline (i.e. No Information Rate), that is, the result we would get if we always predicted the largest class. We will assume the whole set of processes to be rivals if the accuracy of the classifier does not improve significantly on the baseline. We also present the confusion matrices, which show the errors made by the classifier. The presence of relevant structure in the confusion matrices not captured by the overall accuracy will prove important to our assessment of rivalry.

5.2 Results

5.2.1 Derived nouns

Let us start with deadjectival nouns, which we will comment on in more detail to illustrate how we interpret the results of the experiment. Our dataset includes 4 A>N processes, exemplified below:

(6) a. *-té*: *bon* ‘good’ > *bonté* ‘goodness’

b. *-ité*: *sévère* ‘severe’ > *sévérité* ‘severeness’

c. *-itude*: *apte* ‘apt’ > *aptitude* ‘aptitude’

d. *-erie*: *sauvage* ‘savage’ > *sauvagerie* ‘savagery’

These four processes are nearly synonymous (Koehl, 2012), hence we expect very low discriminability. In particular, we expect discriminability to be lowest for *-té* and *-ité*, which are usually thought of as simple allomorphs. The classification results for these four cases are shown in Figure 7. First note that the overall accuracy is very poor: 86.6% might look good, but this needs to be relativized to the fact that the data is very imbalanced, with 83.1% of the examples being cases of *-ité*. Hence the classifier does not really do a better job that it would have if it had paid no attention to the vector and just predicted that all cases were instances of *-ité*. The confusion matrix confirms

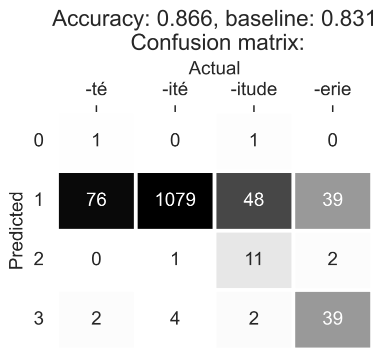

Figure 7: Classifier for deadjectival nouns.

this picture but gives us a little bit more information. Each cell indicates how many of the pairs of lexemes that are actually related by the process given in column were predicted to be related by the process given in row. So for instance, among the 79 actual cases of *-té* derivation, only 1 was correctly predicted to be an instance of *-té*, while 76 were predicted to be an instance of *-ité* and 2 an instance of *-erie*. Background darkness is indicative of proportions of predicted values in the same column: the darker the cell, the more that process was predicted relative to the other possible predictions within the same column.

There are two clear observations stemming from this confusion matrix: first, the largest class, that of *-ité* derivations, plays the role of an attractor. This is to be expected when discriminability is poor: given that more than 80% of its training data are instances of *-ité*, when in doubt, the classifier reasonably defaults to that class. Second, the level of confusion varies from process to process. The classifier is maximally confused between *-té* and *-ité*, doing a catastrophic job of identifying cases of *-té*. It is relatively good at distinguishing *-ité* from *-erie*, with about half the examples correctly classified. *-itude* stands in the middle, with a low but nontrivial number of cases correctly classified.

Overall then we reach a mixed conclusion. On the one hand, overall classifier accuracy is at chance level, wich suggests that these four processes should be considered rivals under a categorical definition of rivalry. On the other hand, detailed examination of the confusion matrix shows evidence for a gradient of discriminability, that matches the gradient of similarity that we documented in the first experiment (see Figure 4).

We now turn to deverbal nouns, which can be related by 11 processes in our dataset. These can be classified in the two broad categories of agent and instrument nouns, as outlined in Section 3.

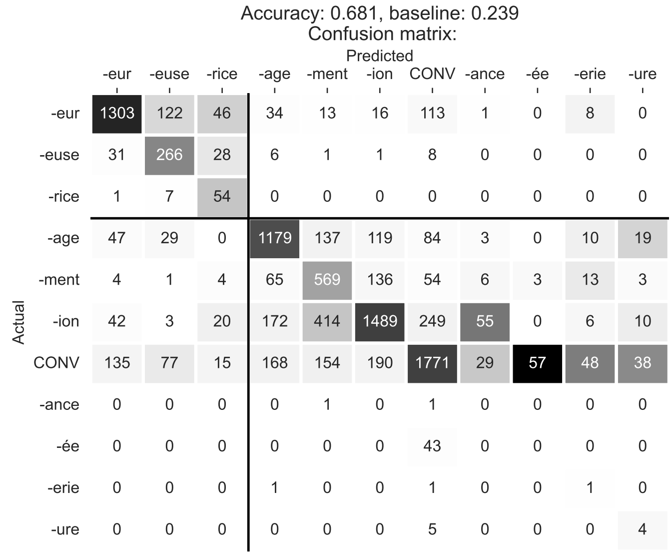

Figure 8: Classifier for deverbal nouns. Dark lines separate agent and event nouns.

Figure 8 shows the results. The overall accuracy is very good: while in absolute terms 68.1% might not look that impressive, it is quite good given that there are 11 classes (hence lots of incorrect options to avoid) and the largest class (conversion) makes up only 23.9% of the data. Moreover, the confusion matrix clearly shows that there is very good discrimination between the agent and event noun forming processes, the separation between which is materialized by the horizontal and vertical thick dark lines. The fact that all cells in the lower left and the upper right quadrants have a very light color indicates that there is little confusion between the two broad families of processes.

If we now turn to discriminability within each of the two broad classes, a much more nuanced picture emerges. Among agent nouns, while there is a significant amount of confusion, with the larger masculine *-eur* class attracting items from the smaller feminine *-euse* and *-rice* classes, the correct class is still the most predicted one within each column. In particular, the classifier did much better than chance in separating out feminine nouns in *-euse* and *-rice*, which we would have expected to be semantically interchangeable. There are many possible causes of this situation, which would require more detailed inverstigation. First, although we made sure that the vectors were not influenced by grammatical gender, for animate nouns gender has obvious semantic consequences, so that denotations of animate nouns in *-eur* are semantically distinguishable from animate *-euse* and *-rice* nouns. Second, remember that nouns in this class do not all denote animate agents; in particular they can also denote instruments. It is quite possible that the proportion of agents vs. instruments in the three classes are different.[12] Third, building on previous literature in sociolinguistics, Wauquier (2020) documents distributionally measurable differences in the valence of feminine agent nouns in *-euse* and *-rice*, *-euse* nouns being more likely to be used to denote low prestige activities. Overall

[12]We are indebted to Timothee Mickus for this observation. See Mickus, Bonami, and Paperno (2019) for discussion of the distributional impact of gender for animate vs. inanimate nouns in *-eur*, *-euse* and *-rice*.

then, there seems to be fine predictive structure among agent noun forming processes that crucially depends on the animate/inanimate distinction, and that would warrant a more detailed study.

Moving on to event nouns, we find an even more nuanced picture. First, prediction accuracy is very low for the 4 low type frequency processes *-ance*, *-ée*, *-erie* and *-ure*. Second, discriminability among the 4 high frequency processes *-age*, *-ment*, *-ion* is far from perfect but not that bad: in each case the correct class is the most frequently predicted; only *-ment* is predicted correctly less than half of the time, the three other ones leading to a correct result three quarters of the time. It is interesting to note that there is more confusion between derivation in *-ment* and *-ion* that between any of these two and *-age*, a result that is at odds with previous observations by Wauquier, Hathout, and Fabre (2020, p. 114). It is also interesting to note that conversion does not stand out as easier to predict, despite the fact that it can convey meanings unavailable to the other processes.(Tribout, 2010)[13]

Overall then, the examination of deverbal nouns provides clear evidence for different degrees of semantic discriminability among derivational processes, but no clear evidence for classes of categorical rivals.

5.2.2 Derived adjectives

We now turn to derived adjectives, starting with denominal formations. As shown in Figure 9, discrimination is reasonably good without being excellent. Examination of the confusion matrix reveals high confusability between formations in *-ien*, *-el*, *-aire*, *-al*, *-if* and *-ique*, which are all mostly predicted to belong to the largest class, that of *-ique*. The situation here is reminiscent of what we observed for deadjectival nouns: these suffixes can easily be argued to be categorical rivals, although there is still a gradient of discriminability, with e.g. *-if* much more frequently correctly predicted than *-ien*. The final three processes (conversion, *-ier*, and *-eux*) stand out in that, despite the attraction power of the largest class, the correct prediction is the most frequent one; this suggests that there is little overlap between the meaning of these processes and those of the first six.

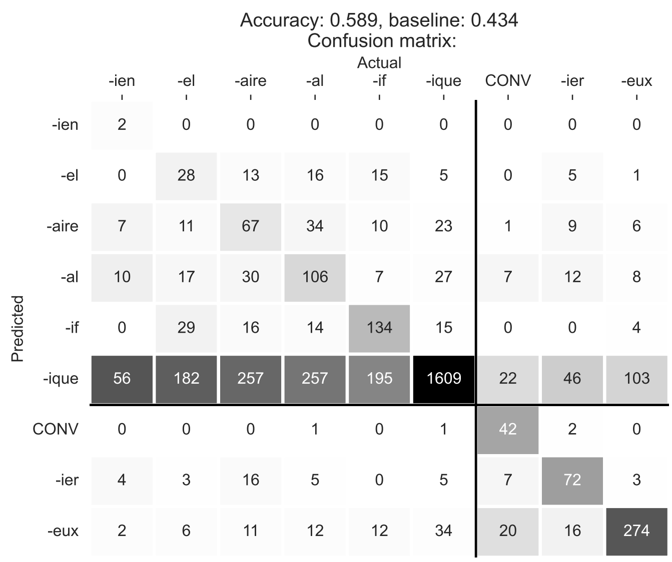

Figure 9: N>A derivations. Dark lines isolate candidate rivals (upper left quadrant) from the rest.

[13] Tribout's (2010, p. 309) careful study of a sample of uncontrovertible V>N conversions concludes that about 20% of them denote an eventuality participant; something that is not impossible (witness *créature* 'creature') but vanishingly rare for the other processes under examination here.

Deverbal adjectives present a strikingly different picture, as shown in Figure 10: the accuracy is higher by 10 points, despite the fact that the baseline is smaller than 10 points. This indicates that deverbal adjectivisation processes are fairly distinct, and hence not prone to rivalry. In fact, the accuracy is artificially dragged down by the decision of treating formations in *-é*, which arguably are converted first conjugation past participles, and "other past participles" as belonging to separate classes: confusion between these two is total, as one would expect given that there is little in the way of correlation of semantics with conjugation class in French. Other than that, we note significant amounts of attraction of *-eur* and *-if* formations to the largest class *-ant*, suggesting that these overlap in meaning.

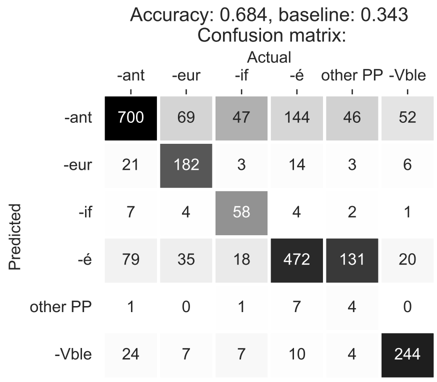

Figure 10: Classifier for deverbal adjectives. Dark square isolates clear rivals.

5.2.3 Derived verbs

Datasets for derived verbs are much poorer, with only two denominal and two deadjectival processes for which we have enough documented data for inclusion in our study.[14] We find however an interesting contrast. For deadjectival verbs, confusion between the two processes is total; hence we have the strongest possible evidence for rivalry. For denominal verbs, the classifier manages to discriminate some cases of *-iser* suffixation from conversion, despite the fact that conversion accounts for nearly 90% of the data. This is reminiscent of Plag's 1999 picture of the relationship bextween the three processes and outlined in Figure 1, and of parallel observations for French by Namer (2009) and Tribout (2010).

5.3 Discussion

The results above suggest that computational classification of processes on the basis of distributional difference vectors is a viable method for capturing basic observations on rivalry via systematic, automated means.

On the one hand, successful classification provides us with a categorical criterion for considering that processes are rivals. This leads to uncontrovertible results conforming to our expectations in three cases, as summarized in (7).

(7) Confirmed collections of rivals

a. Deadjectival nouns: *-té*, *-ité*, *-itude*, *-erie*.

[14]Note that Bonami and Thuilier (2019) found that, for a vast number of verbs in *-iser* and *-ifier*, it was impossible to decide whether their base was a noun or adjective. This helps explain the relatively low type frequencies in Figure 11, and in particular the absence of *-ifier* among denominal processes.

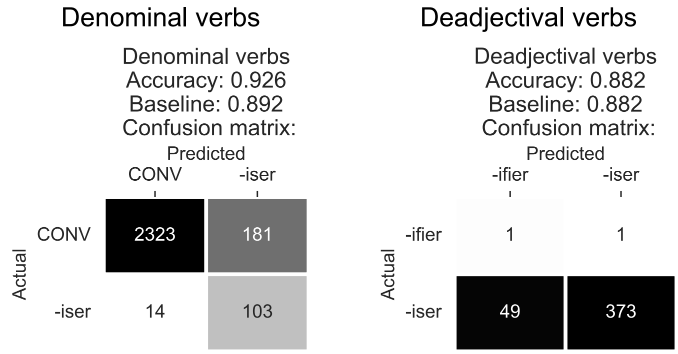

Figure 11: Classifiers for derived verbs

b. Denominal verbs: conversion, *-iser*.

c. Deadjectival verbs: *-iser*, *-ifier*.

Likewise, we met our expectations when concluding that not all deverbal noun forming processes are rivals, and neither are all deverbal or denominal adjective forming processes. Within both sets of derived adjectives, we identified subsets of poorly discriminated processes, which are also candidates for categorical rivalry. Here we cannot conclude as firmly since we did not run classifiers specifically trained on these collections of processes, but the confusion matrices are still strongly suggestive.

(8) Likely rivals

a. Denominal adjectives: *-ien*, *-el*, *-aire*, *-al*, *-if*, *-ique*.

b. Deverbal adjectives: all conversions from past participle, whether regular (*-é*) or not.

On the other hand, we have two types of results that do not conform to our expectations given the simple received view of rivalry as semantic identity. First, we have identified collections of processes which, while far from being perfectly discriminated, are still different enough that classifiers the classifiers separate them much better than chance would predict. These are summarized in (9).

(9) Unexpected discriminability:

a. Agent nouns: *-eur*, *-euse*, *-rice*.

b. Event nouns: *-age*, *-ment*, *-ion*, conversion.

We conjecture that three types of factor may explain these unexpected results. First, polysemy may be involved: two processes may have the same range of possible readings without having the same preferences among these readings, leading to statistical regularities that make prediction possible but were not identified in the literature. Second, distributional properties not reflective of semantics may be at play; see for instance (Wauquier and Bonami, 2021) for suggestive observations on distributional differences between native (e.g. *-age*, *-euse*) and learned (e.g. *-ion*, *-rice*) derivational processes. Third and finally, true semantic differences between processes might be at play. Finer-grained studies will be needed to determine which of these factors are actually relevant, but it is entirely plausible that all three be.

The second result not conformant to expectations is more fundamental. Although successful classification gives us a natural threshold allowing us to take a binary decision, our classifiers are reflective of the inherently gradient nature of competition among word formation processes: classification can be successful or unsuccessful to different degrees, in a way that reflects the gradience of semantic similarity across processes already documented in our first experiment. As such then, the results suggest that, while an operational categorical definition of rivalry is workable, it does not do justice to the subtleties of competition in word formation; and hence the gradient view outlined in (3) is fully supported by the data.

Having recapitulated what we have done, it is worth noting that there are two more things we have left aside. First, we have not attempted to distinguish different types of partial rivalry. As a case in point, in the introduction we made a distinction between nested rivalry, where one process has more potential meanings than another, and overlapping rivalry, where two processes have some meaning types in common but neither can be said to be more general than the other. While this would be a topic for a different paper, we may speculate that our method could be extended to explore that distinction: all other things being equal, if process A is semantically more general than process B, we expect instances of A to be easier to classify than instances of B on average, as there will be cases of instances of A incompatible with the meaning of B, but not the other way around.[15]

Second, we have not questioned the conventional position that rivalry, seen as competition among word formation processes for the derivation of the same type of meanings from the same class of bases, is a natural focus of attention in the study of competition in morphology. Yet it is not self-evident. As we already hinted at in the introduction, our interest in rivalry stems from the fact that it corresponds to one of the situations language users face when they need to coin a new lexeme. But in that situation, they are definitely not limited to considering a set of processes applied to a single base: they might consider multiple relevant bases (think e.g. of *incarceration* vs. *imprisonment* vs. *jailing*), or combine the choice of a morphological strategy with that of a syntactic construction (think e.g. of the choice of compounding in *inflection class* vs. coining of an attributive denominal adjective in *inflectional class*; see Strnadová (2018) for a discussion of related situations in French). In addition, recent research suggests that coining decisions are influenced not just by a single base but by multiple members of the morphological family that is to be extended: hence Bonami and Thuilier (2019) shows that rivalry between *-iser* and *-ifier* is influenced by whether the family contains both a noun and a related adjective, while Bonami and Strnadová (2019) shows that rivalry between *-euse* and *-rice* in the formation of feminine agent/instrument nouns can be predicted by the identity of the process forming the corresponding event noun. Summing up then, rivalry is one case in a rich landscape of alternatives a speaker chooses from probabilistically when coining a new lexeme. While we focussed here on the more manageable question of rivalry rather than the broader question of probabilistic coining, the basic methodology we proposed should readily extend to the exploration of the larger question.

6 Conclusion

In this paper we have argued that distributional semantics offers a solid basis for exploring the semantics of derivational processes at scale, by showing that it can capture the conventional notion of morphological rivalry. First, we showed that difference vectors between bases and derivatives,

[15] Note however that such comparisons should be made on datasets where the size of classes are balanced: if one class is larger than the other, classifiers will tend to favour that class when in doubt, making comparisons across classes difficult. Having large, balanced datasets for a variety of processes is an empirical challenge that goes beyond the scope of the present research project.

once averaged across instances of the same process, capture relevant aspects of the gradient semantic similarity between processes: in particular, hierarchical clustering applied to these average difference vectors yields results comparable to those produced by expert morphologists. Second, we proposed to diagnose whether processes are rivals by examining whether a computational classifier can discriminate them on the basis of their distributional import. We showed that this led to results mostly congruent with received wisdom about rivalry in French, but also revealed a richer gradient of differentiation between processes. We argue that focusing on this richer, noncategorical landscape of coining opportunities is a promising avenue for future research on competition in word formation.

Acknowledgements

We are deeply grateful to Lucie Barque, Bruno Cartoni, Bernard Fradin, Pauline Haas, Nabil Hathout, Franz Rainer, and Marine Wauquier for providing the expert trees reported on in section 3. Versions of the work were presented at the 2020 meeting of the *Societas Linguistica Europaea*, the 2021 Fribourg workshop on affix rivalry, the Paradigmo 2 workshop, and colloquia at Université Paris Cité and Ohio State University. We thank organizers and audience members for insightful feedback, and in particular Maria Copot, Richard Huyghe, Timothee Mickus, Andrea Sims, Rossella Varvara, and Marine Wauquier. We also thank two anonymous reviewers and the guest editors for useful comments and suggestions.

This research was partially conducted while Matías Guzman Naranjo was a postdoc at Université Paris Cité. This work was partially supported by the Démonext project (ANR 17-CE23-0005) as well as a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083); and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement 834050).

References

Amenta, Simona, Fritz Günther, and Marco Marelli (2020). "A (distributional) semantic perspective on the processing of morphologically complex words". In: *The Mental Lexicon* 15.1, pp. 62–78.

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

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.

Baroni, Marco, Georgiana Dinu, and Germán Kruszewski (2014). "Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Baltimore, Maryland: Association for Computational Linguistics, pp. 238–247.

Bauer, Laurie, Rochelle Lieber, and Ingo Plag (2013). *Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov (2016). "Enriching Word Vectors with Subword Information". In: *arXiv preprint arXiv:1607.04606*.

Boleda, Gemma (2020). "Distributional Semantics and Linguistic Theory". In: *Annual Review of Linguistics* 6.1, pp. 213–234.

Bonami, Olivier and Gilles Boyé (2005). "Construire le paradigme d'un adjectif". In: *Recherches Linguistiques de Vincennes* 34, pp. 77–98.

— (2019). "Paradigm uniformity and the French gender system". In: *Perspectives on morphology: Papers in honour of Greville G. Corbett*. Ed. by Matthew Baerman, Oliver Bond, and Andrew Hippisley. Edinburgh: Edinburgh University Press, pp. 171–192.

Bonami, Olivier and Denis Paperno (2018). "Inflection vs. derivation in a distributional vector space". In: *Lingue e Linguaggio* 17.2, pp. 173–195.

Bonami, Olivier and Matteo Pellegrini (2022). "Derivation predicting inflection: A quantitative study of the influence of derivational history on inflectional behavior in Latin". In: *Studies in Language* 46.

Bonami, Olivier and Jana Strnadová (2019). "Paradigm structure and predictability in derivational morphology". In: *Morphology* 29.2, pp. 167–197.

Bonami, Olivier and Juliette Thuilier (2019). "A statistical approach to affix rivalry: French *-iser* and *-ifier*". In: *Word Structure* 12.1, pp. 4–41.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32.

Camacho-Collados, Jose and Mohammad Taher Pilehvar (2020). "Embeddings in Natural Language Processing". In: *Proceedings of the 28th International Conference on Computational Linguistics: Tutorial Abstracts*. Barcelona, Spain (Online): International Committee for Computational Linguistics, pp. 10–15.

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA). KDD '16. New York, NY, USA: ACM, pp. 785–794.

Erk, Katrin (2012). "Vector space models of word meaning and phrase meaning: A survey". In: *Language and Linguistics Compass* 6.10, pp. 635–653.

Evert, Stefan (2014). "Distributional semantics in R with the wordspace package". In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: System demonstrations*, pp. 110–114.

Firth, John R. (1957). “Modes of meaning”. In: *Papers in Linguistics, 1934-1951.* Oxford: Oxford University Press, pp. 190–215.

Fradin, Bernard (2019). “Competition in derivation: what can we learn from duplicates?” In: *Competition in Morphology.* Ed. by Francesco Gardani, Hans-Christian Luschützky, and Franz Rainer. U. L’Acquila. Berlin: Springer, pp. 67–93.

Galili, Tal (2015). “dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering”. In: *Bioinformatics.*

Gardani, Francesco, Franz Rainer, and Hans C. Luschützky (2019). “Competition in morphology: a historical outline”. In: *Competition in inflection and word-formation.* Ed. by Franz Rainer, Francesco Gardani, Wolfgang U. Dressler, and Hans C. Luschützky. Dordrecht: Springer, pp. 3–36.

Guzmán Naranjo, Matías and Olivier Bonami (2021). “Overabundance and Inflectional Classification: Quantitative Evidence from Czech”. In: *Glossa: a journal of general linguistics* 6.1.

Harris, Zelig S. (1954). “Distributional Structure”. In: *Word* 10, pp. 146–162.

Hathout, Nabil and Fiammetta Namer (2014). “Démonette, a French derivational morpho-semantic network”. In: *Linguistic Issues in Language Technology* 11.5, pp. 125–168.

Huyghe, Richard and Delphine Tribout (2015). “Noms d’agent et noms d’instrument: le cas des déverbaux en *-eur*”. In: *Langue française* 185, pp. 99–112.

Huyghe, Richard and Marine Wauquier (2020). “What’s in an agent?” In: *Morphology* 30.3, pp. 185–218.

— (2021). “Distributional semantic insights on agentive suffix rivalry in French”. In: *Word Structure* 14.3, pp. 354–391.

Koehl, Aurore (2012). “La construction morphologique des noms désadjectivaux suffixés en français”. PhD thesis. Université de Lorraine.

Lenci, Alessandro (2008). “Distributional semantics in linguistic and cognitive research”. In: *Rivista di Linguistica* 20, pp. 1–31.

Lindsay, Mark and Mark Aronoff (2013). “Natural Selection in Self-Organizing Morphological Systems”. In: *Morphology in Toulouse: Selected Proceedings of Décembrettes 7.* Ed. by Nabil Hathout, Fabio Montermini, and Jesse Tseng. Munich: Lincom Europa, pp. 133–153.

Luschützky, Hans C. and Franz Rainer (2013). “Instrument and place nouns: a typological and diachronic perspective”. In: *Linguistics* 51, pp. 1301–1359.

Marelli, Marco and Marco Baroni (2015). “Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics”. In: *Psychological Review* 122, pp. 485–515.

Mickus, Timothee, Olivier Bonami, and Denis Paperno (2019). “Distributional Effects of Gender Contrasts Across Categories”. In: *Proceedings of the 2nd Meeting of the Society for Computation in Linguistics.*

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). “Efficient Estimation of Word Representations in Vector Space”. In: *CoRR* abs/1301.3781.

Miller, George A and Walter G Charles (1991). “Contextual correlates of semantic similarity”. In: *Language and cognitive processes* 6.1, pp. 1–28.

Namer, Fiammetta (2009). *Morphologie, Lexique et Traitement Automatique des Langues.* London: Hermès Science Publishing.

Namer, Fiammetta, Lucie Barque, Olivier Bonami, Pauline Haas, Nabil Hathout, and Delphine Tribout (2019). “Demonette2 — Une base de données dérivationnelles du français à grande échelle : premiers résultats”. In: *Actes de TALN.* Toulouse, France.

Padó, Sebastian, Aurélie Herbelot, Max Kisselew, and Jan Šnajder (2016). “Predictability of distributional semantics in derivational word formation”. In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 1285–1296.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning (2014). “GloVe: Global Vectors for Word Representation”. In: *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543.

Plag, Ingo (1999). *Morphological productivity*. Berlin: Mouton de Gruyter.

Řehůřek, Radim (2010). “Software Framework for Topic Modelling with Large Corpora”. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50.

Roché, Michel (1997). “*Briard*, *bougeoir* et *camionneur* : dérivés aberrants, dérivés possibles”. In: *Silexicales 1: Mots possibles et mots existants*. Ed. by Danielle Corbin, Bernard Fradin, Benoît Habert, Françoise Kerleroux, and Marc Plénat, pp. 241–250.

Schäfer, Roland (2015). “Processing and Querying Large Web Corpora with the COW14 Architecture”. In: *Proceedings of Challenges in the Management of Large Corpora*, pp. 28–34.

Schäfer, Roland and Felix 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*, pp. 486–493.

Štekauer, Pavol (2005). *Meaning Predictability in Word Formation*. Amsterdam: John Benjamins.

Strnadová, Jana (2014). “Les réseaux adjectivaux: Sur la grammaire des adjectifs dénominaux en français”. PhD thesis. Université Paris Diderot et Univerzita Karlova V Praze.

— (2018). “Lexeme equivalence or rivalry of lexemes?” In: *The lexeme in descriptive and theoretical morphology*. Ed. by Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo, and Fiammetta Namer. Berlin: Language Science Press, pp. 509–525.

Tribout, Delphine (2010). “Les conversions de nom à verbe et de verbe à nom en français”. PhD thesis. Université Paris Diderot.

— (2012). “Verbal stem space and verb to noun conversion in French”. In: *Word Structure* 5, pp. 109–128.

Turney, Peter D (2012). “Domain and function: A dual-space model of semantic relations and compositions”. In: *Journal of artificial intelligence research* 44, pp. 533–585.

Turney, Peter D and Patrick Pantel (2010). “From Frequency to Meaning: Vector Space Models of Semantics”. In: *Journal of artificial intelligence research* 37, pp. 141–188.

Varvara, Rossella, Gabriella Lapesa, and Sebastian Padó (2021). “Grounding Semantic Transparency in Context”. In: *Morphology*, pp. 1–38.

Wauquier, Marine (2020). “Confrontation Des Procédés Dérivationnels et Des Catégories Sémantiques Dans Les Modèles Distributionnels”. Toulouse: Toulouse 2.

Wauquier, Marine and Olivier Bonami (2021). “Social gender and derivational morphology: a distributional study of the gendered import of learned morphology in French”. Presented at the 4th International Symposium on Morphology.

Wauquier, Marine, Nabil Hathout, and Cécile Fabre (2020). “Semantic discrimination of technicality in French nominalizations”. In: *Zeitschrift für Wortbildung/Journal of Word Formation* 4.2, pp. 100–119.