Olivier Bonami & Matías Guzmán Naranjo

Distributional evidence for derivational paradigms

Abstract: This paper provides evidence from distributional semantics on the importance of paradigmatic relations in word formation systems. We argue that, if paradigmatic relations play a systemic role, that role should be measurable by examining whether pairs of lexemes derived from a common base (e.g. *baker* and *bakery* derived from *bake*) have semantic properties that are more readily predictable from one another than they are from the semantics of the base. We implement that idea using statistical models that predict the distributional vector of a lexeme based on that of another lexeme in the same derivational family. Applying the method on French data, we show that some (but by no means all) pairs of derivational processes give rise to undisputably paradigmatic properties and hence provide evidence of a systemic role for paradigmatic relations.

Keywords: derivational paradigms, semantics, analogical modeling

1 Introduction

There is a variety of ways lexemes may be morphologically related. Within a morphological family, some lexemes stand in a clear base-derivative relation, e.g. *lover* derives from *love*; others stand in a determinate but indirect relation, e.g. *lover* and *lovable* are indirectly related through their shared base *love*. In many other cases, while the two lexemes are clearly related, the exact nature of that relation is hard if not impossible to determine; think e.g. of accidental cases such as the relation between *social* and *society* or more systematic cases such as the relation between *pessimism* and *pessimist*. Clearly then, lexemes in a family entertain a variety of PARADIGMATIC RELATIONS, only some of which are base-derivative relations.

The idea that paradigmatic relations in general play an important role in word formation has received renewed interest by some, and skepticism by others, in the last few years; witness individual papers such as Štekauer (2014); Bauer (2019) as well as collections such as Hathout & Namer (2018, 2019); Fernández-Domínguez et al. (2020). The debate centers on two issues: the extent to which

Olivier Bonami & Matías Guzmán Naranjo, Université Paris Cité, CNRS, Laboratoire de linguistique formelle, e-mail: olivier.bonami@linguist.univ-paris-diderot.fr, Eberhard Karls Universität Tübingen, e-mail: mguzmann89@gmail.com

derivational families can be said to share properties of inflectional paradigms, and the existence and importance of predictability relations among members of a derivational family that cannot be reduced to their shared relationship to a common base.

In this paper we focus on the latter issue, and argue that hard quantitative evidence is needed to settle it and establish whether nontrivial paradigmatic relations play a systematic role in word formation. To this end, we deploy on French data methods from distributional semantics to assess the semantic interpredictability between pairs of lexemes derived from a common base, and compare it to semantic predictability from the base. We conclude that definitive evidence for nontrivial paradigmatic relations can be found in some but not all corners of the word formation system.

The structure of this paper is as follows. Section 2 presents an overview of the literature on derivational paradigms, and contextualizes our study. Section 3 gives a brief introduction to key ideas from distributional semantics, and formalizes our hypothesis. In Section 4 we present our dataset and the methods for this study. Section 5 discusses the results and Section 6 concludes.

2 Derivational paradigms

2.1 Motivations for derivational paradigms

Conventional theories of word formation typically assume that derivational morphology is organized in terms of oriented (base, derivative) relationships, where each morphologically complex unit is motivated by a prior and morphologically simpler unit. This is formulated differently in different traditions: in a morpheme-based approach, the complex unit is formed by adding a morpheme to the simpler unit; in lexeme-based approaches of the tradition initiated by Aronoff (1976),[1] the meaning and shape of a derived lexeme is motivated by the application of a Lexeme Formation Rule to another base lexeme. Both approaches however share the view that derivational families are normally structured as rooted trees (Stump, 2019), with one unit serving as the ultimate ancestor for the whole family. Figure 1 illustrates this for the family of the English noun *center*.

A consequence of this view of derivational families is that not all situations of predictable morphological relatedness have the same value. As a case in point,

1 As Aronoff (1994, 7) clarifies, the term 'word-based' was an unfortunate choice for the approach developed there. We follow the more recent and accurate usage here.

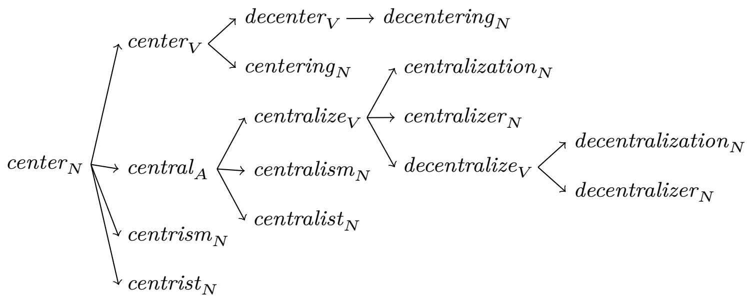

Fig. 1: Example of a derivational family structured as a rooted tree

consider the pair of lexemes (*centralizer*, *decentralizer*) from Figure 1: while there is a large series of pairs of English nouns that differ formally by the presence of the prefix *de-* and implement parallel semantic contrasts, this perceptible morphological relatedness is assumed not to be a derivational relation, but a consequence of the existence of a common base *centralize* from which both derive directly (*centralize*→*centralizer*) or indirectly (*centralize*→*decentralize*→*decentralizer*).

Yet there is ample and well-known evidence for the existence of situations where the rooted-tree model of derivational families fails to capture the full extent of morphological relatedness involving derivational processes. We document some of these in the remainder of this subsection (see also Bauer et al. 2013, chap. 23 for a compendium of the evidence from English). Our list is by no means meant to be exhaustive, but should give the reader a feel for the nature and importance of the phenomena under consideration; see among many others van Marle (1984); Becker (1993); Bauer (1997); Booij (1997); Roché (2011b); Hathout & Namer (2014b); Strnadová (2014b) for more detailed discussion.

2.1.1 Back-formation

Probably the most well-known relevant situation is BACK-FORMATION. The term describes cases where there is clear historical or morphological evidence that the existence of a morphologically more complex unit motivated the coining of the simpler one rather than the other way around. A well-known example is English *bartend* from *bartender*: because English has NN compounds but no NV compounds, it is clear that the verb was modeled on the agent noun and not the other way around. While back-formation is usually considered to be an interesting but abnormal use of morphological resources, Namer (2012) provides compelling evidence for the highly productive use of backformation to coin verbs from neoclassical compounds in French.

2.1.2 Multimotivation

A different but important family of situations that challenge the received view are cases of MULTIMOTIVATION, where multiple members of a derivational family function as motivators for a derivative. The simplest subtype is the situation where it is undecidable which of two or more derivation trees should be postulated for a given derivative. Corbin (1976) famously discusses the case of French *assymmétrique*, whose English cognate *asymmetrical* shares the same properties: it is undecidable whether it should be seen as deriving from *asymmetry* by suffixation of *-ical*, or from *symmetrical* by prefixation of *a-*. As Figure 2 illustrates, one would be tempted to model this by having two converging derivation relations leading to the same form, if that did not go against the rooted tree view of derivational families. Importantly however, while it is unpleasant to have to choose arbitrarily one of the two rooted trees (i.e. including one of the two dashed arrows in the figure but not the other), there is no morphological generalization that is not captured by both; hence arguably nothing valuable is lost by making that choice.

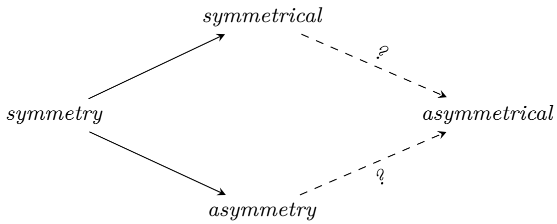

Fig. 2: The position of *asymmetrical* in its derivational family

Multimotivation is more problematic in cases where it combines with a FORM-CONTENT MISMATCH (Hathout & Namer, 2014b). In such situations, the lexeme that is the manifest formal base for a derivative is distinct from the (related) lexeme from which the meaning of the derivative seems to follow. Hence *frequentist* is based formally on *frequent*, but its meaning is based on that of *frequency*: a frequentist values frequency, not things that are frequent. This is represented graphically in Figure 3. Likewise, *pessimistic* is based formally on *pessimist* but relates semantically more readily to *pessimism*. While such situations have a family resemblance with the case of *asymmetrical* discussed above, they crucially differ in that no choice of a rooted tree really captures the situation: since two lexemes in the family make complementary contributions to motivating the derivative, neither can be ignored.

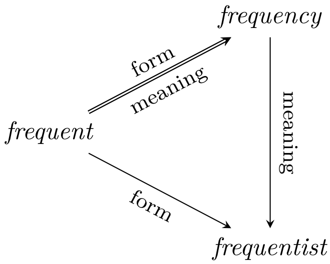

Fig. 3: The position of *frequentist* in its derivational family. Double arrows indicate convergent motivation by form and meaning, while single arrows indicate motivation in one dimension only.

More situations of multimotivation can and should be documented. In the preceding example, the two motivators make orthogonal contributions, since one is crucial for form and the other for meaning. Sometimes two motivators are crucial for meaning, with the derivative being ambiguous between a reading relating to one or the other (Strnadová, 2014b). For instance, *senatorial* has both a reading derived from *senator* (as in *senatorial family*, a family consisting of senators) and a reading derived from *senate* (as in *senatorial palace*, the palace hosting the senate). Because the two readings are closely related, it would be odd to treat this as a situation of homonymy, with two distinct lexemes derived independently, rather than polysemy, with a single derived lexeme. But if polysemy is posited, then we have a single lexeme whose semantics is jointly motivated by two separate other members of the derivational family. This is depicted in Figure 4.

2.1.3 Cross-formation

Becker (1993) calls CROSS-FORMATION situations where pairs of lexemes stand in a relation of mutual motivation with no reason to assume that one direction of derivation is preferrable to the other.

As Becker notes, conversion is a prime source of situations of cross-formation. Since conversion pairs do not contrast in their morphological complexity, form

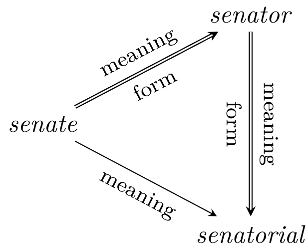

Fig. 4: The position of *senatorial* in its derivational family

provides no strong argument for directionality. As discussed by Marchand (1963), sometimes semantics does not help either: for instance, it makes just as much sense to see the nouns *judge* as deriving from the verb *judge*, by analogy to the derivation of *leader* from *lead*, or the verb *judge* as deriving from the noun *judge*, by analogy to the derivation of *deputize* from *deputy*. We may depict such situation of cross-formation as in Figure 5.

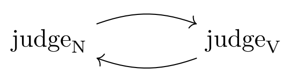

Fig. 5: Conversion as cross-formation

Reporting on a detailed investigation of verb-noun conversion pairs in French, Tribout (2020) shows that the situation exemplified with *judge* is general. She argues convincingly that the only unambiguous evidence for directionality in conversion comes from situations where the prior derivational history makes it obvious which member of the pair came first: for instance, *parlement* 'parliament' is the source for *parlementer* 'negotiate' rather than the other way around, because it obviously derives from *parler* 'speak' through the *-ment* noun-forming deverbal suffix; conversely, *décharge* 'dump, discharge' must derive from *décharger* 'unload, discharge', because of the presence of the verb-forming deverbal suffix *dé-*.[2] Analyzing 626 such cases with the help of a set of semantic relations derived from those of Plag (1999), Tribout observes that 85% of the pairs are related by a reversible semantic function with both directions of derivation attested. As a consequence, semantics does not motivate a direction of derivation in the vast

2 Compare Barbu Mititelu et al. (2023), which relies on etymological information from the *Oxford English Dictionary* to establish directionality, a method Tribout explicitly argues against.

majority of cases. Zooming out to the much larger set of conversion pairs where derivational history provides no motivation either, she concludes that absence of clear directionality is the rule rather than the exception in French verb-noun conversion pairs.

A different type of cross-formation involves substitutive morphology, and is nicely illustrated by *-ist*/*-ism* pairs and their cognates in various European languages. While there is no doubt that *-ist* and *-ism* derivatives can productively be formed by simple suffixation (witness *Trump*, *Trumpism*, *Trumpist*), it has long been observed[3] that the lexicon contains numerous pairs that are readily relatable to one another without being relatable to a common base (e.g. *fascism*, *fascist*; *optimism*, *optimist*; *masochism*, *masochist*). In addition, even where a common base clearly exists, the semantic relationship between the two derived terms is often more crisp and predictable than their relationship to their common base. For a telling example, consider the triple (*social*, *socialism*, *socialist*). While there is a nonarbitrary relation between *social* and its two derivatives, the meaning of the latter can hardly be predicted from that of the former. It is worth citing in full the Oxford English Dictionary's definition of the main reading of *socialism*,[4] to highlight the convoluted relation between its meaning and that of *social*.

> A theory or system of social organization based on state or collective ownership and regulation of the means of production, distribution, and exchange for the common benefit of all members of society; advocacy or practice of such a system, esp. as a political movement. Now also: any of various systems of liberal social democracy which retain a commitment to social justice and social reform, or feature some degree of state intervention in the running of the economy.

By contrast, *socialist* is simply defined as 'an advocate or supporter of socialism'; arguably, there is a transparent, bidirectional relationship between *socialism* (the doctrine of socialists) and *socialist* (an adherent to socialism), which is lacking between the two derivatives and their formal base. This is depicted graphically in Figure 6, with thicker lines indicating stronger motivation. Parallel observations hold for dozen if not hundreds of comparable triples of words.[5]

3 van Marle (1984) traces back observations to that effect in the Dutch tradition to a 1944 paper by Gerlach Royen. See also Roché (2011a) for extensive discussion of the French facts.

4 Consulted online on December 21, 2021.

5 See Rainer (2018) for a fascinating tour of the evolving meanings of *capitalist* and *capitalism* in European languages from the 17th century on, suggesting that the tight parallemism between *-ism* and *-ist* formations in the context of political doctrines is a relatively recent development.

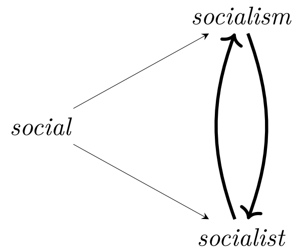

Fig. 6: The position of *socialism* and *socialist* in their derivational family

These observations suggest that speakers are attuned to a close relationship between *-ism* and *-ist* lexemes, allowing them to productively derive one from the other, independently of whether or to what extent a relationship to a common base can be inferred. That is, there is evidence for the existence of strong morphological regularities linking the two series of suffixed nouns in both directions, that are more reliable than those linking them to their formal base.

2.2 Derivational paradigms

The existence of situations of back-formation, multimotivation, and cross-formation, has been documented for decades. Yet they have had limited impact of mainstream morphological theorizing. The most typical position, still held by many, is dismissal: while their existence is acknowledged, they are deemed not to fall within the scope of morphology, either as a matter of principle or because they are assumed to be rare exceptions.

The seminal study of van Marle (1984) made a strong case that such phenomena should not be ignored by morphologists. Van Marle argues that PARADIGMATIC RELATIONS grounding what he calls SECONDARY or ANALOGICAL coinings need to be recognized, but that these are qualitatively different from derivational relations and should be treated outside the derivational system. This view was then popularized in the context of Construction Morphology (Booij, 2010) through the distinction between first-order (classical derivation) and second-order (paradigmatic relations) schemas, and adopted under a different name (daughter vs. sister schemas) by Relational Morphology (Jackendoff & Audring, 2020); see Audring (2019) for a recent and cogent discussion.

A more radical type of reaction is to conclude that unexpected word formation strategies should lead us to rethink entirely the architecture of the derivational system. Under this view, paradigmatic relations structure derivational families across the board, irrespective of whether it is possible to make sense of some of these relations in terms of oriented (base, derivative) relationships. Variants of this view are defended, among others, by Robins (1959); Becker (1993); Bochner (1993); Bauer (1997); Štekauer (2014); Bonami & Strnadová (2019); Namer & Hathout (2020). Crucial to this type of approach is the observation of parallels between inflectional paradigms and derivational families, leading to some notion of a DERIVATIONAL PARADIGM seen as a system of morphological relations that is parallel across derivational families. In particular, just as in inflection, the organization of derivationally-related words in terms of syntactic and semantic contrasts should be seen as conceptually orthogonal to the nature of the formal relationship they entertain.

There are important conceptual differences both between these two families of views and across studies within each of the two groups.[6] These are of little consequence for the present work, which focuses mostly on empirical evidence for a systemic role of paradigmatic relations in derivational morphology. For concreteness though, we will rely on Bonami & Strnadová's (2019) theoretical elaboration, according to which a paradigmatic system is a set of partial families of morphologically related words aligned using parallel contrasts of content. Figure 7 exhibits a simple example of a derivational paradigmatic system in this sense: the semantic contrasts between verb and agent noun, verb and action noun, agent noun and action noun, are the same (at some appropriate level of granularity) across all three derivational families, despite the use of different formal strategies (three different affixes for the action nouns, use of different types of stem allomorphs for the agent nouns). Once such broad semantic contrasts have been identified as motivating alignment across partial families, each collection of aligned lexemes can be seen as filling the same cell in a derivational paradigm. Bonami & Strnadová argue that this allows one to capture similarities between paradigmatic organization in inflection and derivation while doing justice to important differences such as the open-endedness of derivational families and the stronger semantic predictability of inflectional relations (Bonami & Paperno, 2018). Note that, in the spirit of abstractive word-based morphology (Blevins, 2006), there is no notion that some morphological relations between pairs of cells are prior to others: the fully connected graph of all pairs of cells is worthy of investigation, and any member of a paradigm is partially predictive all other

6 There are interesting parallels to be drawn here with the distinction, within approaches to inflection, between realizational approaches such as Paradigm Function Morphology (Stump, 2001, 2016), which use separate mechanisms to derive wordforms from stems and to capture direct implicative relations between wordforms, and abstractive word-based approaches (Blevins, 2006, 2016), which argue that all derivations from subword units are superfluous.

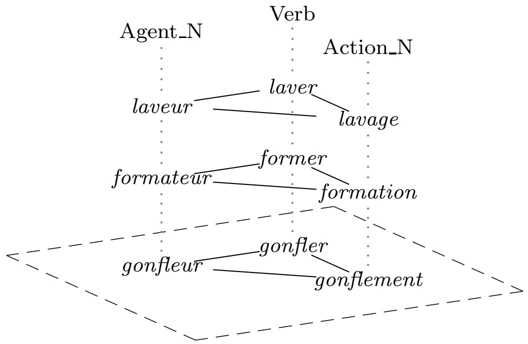

Fig. 7: Example of a derivational paradigmatic system. Horizontal planes represent partial derivational families; horizontal edges represent aligned semantic contrasts; and vertical dotted lines materialize aligned lexemes, filling the same cell in their respective derivational paradigms.

members; the focus of inquiry is in assessing quantitatively the strength of these predictability relations, without any preconception about some lexemes being more 'truly' connected than others.[7]

2.3 Towards systematic evidence for derivational paradigms

In the previous paragraphs we outlined some the main empirical arguments for the view that paradigmatic relations play an important role in the organization of morphological families. It is striking that, while some of these arguments have been on the table for decades, many morphologists remain unconvinced. We submit that one of the causes of that situation is the lack of quantitative evidence bearing on the issue: the debate is not whether some cases exist where morphological structure is used to coin new words that are not derivatives of a simpler base; rather, the debate is whether these are common enough to be considered part of the core of morphology, warranting radical changes to one's understanding of its architecture.

Hence we are on the lookout not just for paradigmatic effects in derivation, but for paradigmatic effects that are systematic. Bonami & Strnadová (2019) provide relevant evidence based on predictibility of *form* rather than *meaning*. Building on previous quantitative work on the implicative structure of inflectional

7 In this context, the representations in Figures 2–6 can all be seen as partial impressionistic representations of predictability relations within derivational paradigms.

paradigms (Ackerman et al., 2009; Ackerman & Malouf, 2013), they examine how predictive the shape of words in each cell in a derivational paradigm is of the shape of words in other cells in that derivational paradigm, using conditional entropy of phonological shape classes as a measure of average predictability.[8] Focusing on French verbs and related agent and action nouns, they show that on average, the shape of an agent noun is more readily predictable from that of the corresponding action noun than from that on the base verb; while the shape of an action noun is equally hard to predict from the base verb as from the agent noun. This is represented graphically in Figure 8, where a higher entropy value indicates more unpredictability. Note also the interesting observation that the derivatives provide better evidence for the shape of their base than the other way around.

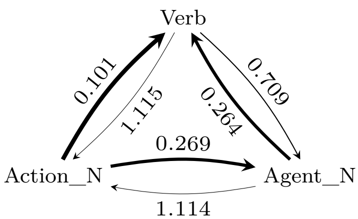

Fig. 8: Form predictability in French derivational paradigms, as measured by conditional entropy. Thicker lines correspond to higher predictability (and hence lower entropy).

Our goal in this paper is to explore whether similar evidence can be found looking at meaning rather than form. We focus on situations where two derivational processes[9] readily apply to the same base, making it possible to compare semantically triples of a base and two derivatives that are maximally similar from a formal point of view. Note that this is one of the configurations that led to the observation of cross-formations. In this situation, the mainstream view

8 More precisely, they examine the relationship between the phonological shape of citation forms, and rely on the modeling strategy detailed in (Bonami & Beniamine, 2016): reported numbers are the conditional entropy of the shape alternation linking two forms given relevant information on the shape of the predictor. All numbers were computed using Sacha Beniamine's Qumín package (Beniamine, 2018) applied to data from the Démonette database (Hathout & Namer, 2014a).

9 For the purposes of this study, we individualize derivational processes by form alternation and part of speech of the input and output; hence neither affix polysemy nor affix rivalry play a role in individualizing processes.

predicts that on average, the semantics of the base should be a better predictor of the semantics of each derivative than the derivatives are of each other.

Anecdotal evidence certainly points in the direction of that situation being common. Consider the two French triples in Figure 9. In the one on the left, the verb is polysemous with two main readings, and each of its two derivatives only has one reading, relating to one of the two readings of the verb. As a result, there is only a very loose semantic connection between the two derived nouns, while both relate very clearly to the base. In the one on the right, both derivatives have a predictable meaning 'user of pumps', but both have been lexicalized with a much more specific reading. While predictability of the derived meaning from the base meaning is already lower in this case, the connection between the meaning of the two derivatives is even thinner.

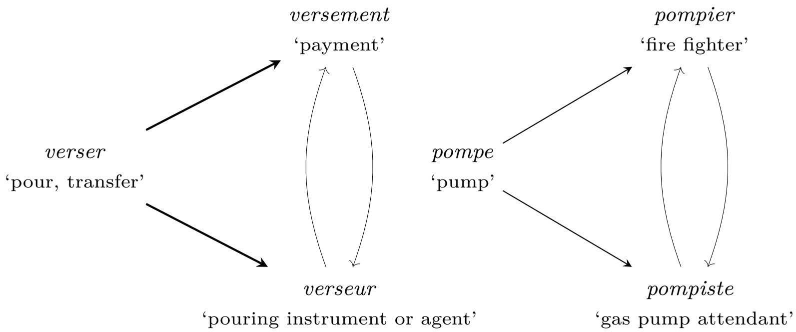

Fig. 9: Two French triples illustrating base predictiveness

These two examples contrast directly with that of *social*, *socialism* and *socialist* pictured in Figure 6, where the semantic relationship between the two derivatives is both more transparent and more predictable than their relations to the adjectival base. What we do not know at this point is how prevalent each of these situations is: if we compare *-ment* and *-eur* (or *-ier* and *-iste*) suffixation in French, is it generally true that the base is a better predictor of derivative semantics, or did we just find a cute but isolated example? And what about *-isme* and *-iste*? Stated more broadly, our research question is then: are there pairs of processes applicable to the same bases where on average, the derivatives are more interpredictable than they are from the base? If the answer is yes, then this is strong evidence that paradigmatic relations between derivatives from the same base must be taken at face value.

Our assessment of the semantic interpredictability of morphologically related lexemes will rely on distributional semantics, to which we now turn.

3 Distributional semantics for morphology

In this paper we rely on a distributional representation of meaning (Harris, 1954; Firth, 1957). Under the distributional hypothesis, the meaning of a word is reflected in its distribution, so that words with similar meanings will appear in similar contexts. In practice, this view of semantics allows us to operationalize the meaning of a word as a numeric vector which we can induce automatically from large corpora. While current implementations of distributional models are not transparent in how they represent the meanings of words, the insight that words with similar meanings have similar distributions is directly captured in terms of the cosine similarity between two vectors (Lenci, 2008). Early work on distributional semantics employed count models, that is, the vector of a word was estimated directly by counting its coocurrences with other words in a corpus (Evert, 2014; Turney & Pantel, 2010; Turney, 2012; Miller & Charles, 1991; Erk, 2012; Baroni et al., 2014). More recent implementations of distributional semantics (e.g. Mikolov et al. 2013b) work by training a neural network to predict each word from its context, and then using the representation used by the network as the distributional vector of the word.[10] Since we are not interested in interpreting the individual vectors themselves, we will work with vectors induced using neural networks.

Distributional vectors have been put to use in all kinds of contexts in computational semantics and natural language processing. Interestingly for our purposes, one observation is that they can be used to draw semantic analogies to a surprising level of accuracy (Mikolov et al., 2013b; Pennington et al., 2014), to the point that accuracy of semantic analogies has quickly become a standard way of assessing the quality of a vector space. To take a concrete example, consider the situation depicted if Figure 10. We know the vectors for the words *Paris*, *France*, and *Colombia*, depicted in black, and we want to find a candidate vector for the word that is semantically to *Colombia* what *Paris* is to *France*. One simple way of doing this is to substract from the vector for *Paris* the vector for *France* and then add back the vector from *Colombia* (both operations depicted in blue). The empirical observation is then that this predicted vector is quite close in vector space (as measured by cosine similarity) to the actual vector for *Bogotá*, depicted in red. And there is a good reason for this: because $(France, Paris)$ and $(Colombia, Bogotá)$ stand in the same semantic relation, the difference vectors $\vec{Paris} - \vec{France}$ and $\vec{Bogotá} - \vec{Colombia}$ are expected to be nearly identical.

10 Note that count models are still used in cases in which interpretability is more important that predictive accuracy (Boleda, 2020; Varvara et al., 2021)

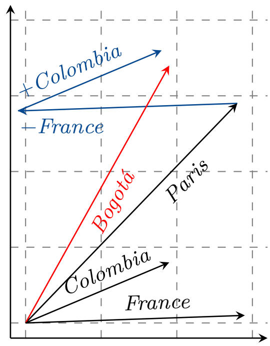

Fig. 10: Semantic analogies using distributional vectors: Bogotá is to Colombia as Paris is to France

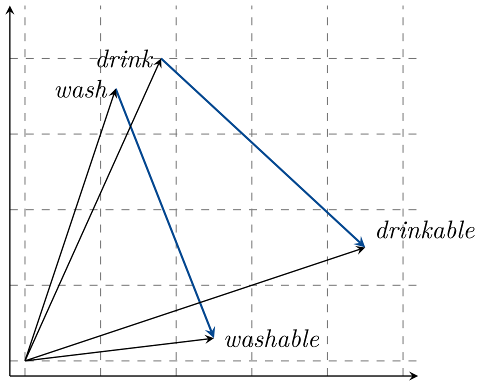

Fig. 11: Example of the distributional representation of *-able*

To illustrate semantic analogies we purposefully used an example where there is no morphological relation between the words under consideration. However if we move back to morphology, at least in simple cases, we expect pairs of words related by the same formal process to stand in the same relation, and hence for the difference vectors between bases and derivatives to be similar to one another (Marelli & Baroni, 2015). Figure 11 illustrates this idea with the example of *able* adjectives in English. While *wash* and *drink* on the one hand and *washable* and *drinkable* on the other hand might be quite dissimilar, we expect the vector from *wash* to *washable* and the vector from *drink* to *drinkable* to be very similar to each other.

This fact effectively allows us to understand the semantics of a morphological process as a function linking the meanings of two words, as first advocated in detail

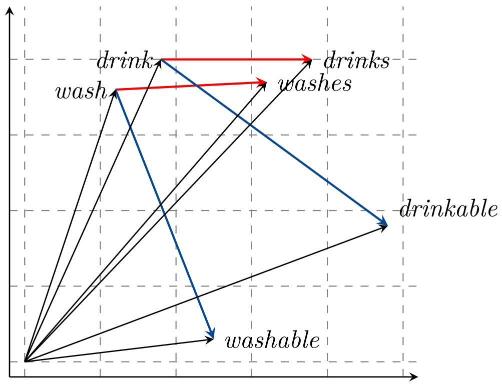

Fig. 12: Difference between derivation and inflection in distributional semantics

by Marelli & Baroni (2015). For example, Bonami & Paperno (2018) compared the degree of predictability in derivational and inflection morphology. Their study showed that inflectional processes have more predictable semantics than derivational processes. Figure 12 illustrates this idea. In this example, the vectors from the infinitive to third singular are more consistent (more similar in length and direction) than the vectors going from the infinitive to the corresponding *-able* adjective.

Other examples of relevant studies leveraging distributional semantics to study semantic aspects of derivational processes are Padó et al. (2016); Lapesa et al. (2018); Amenta et al. (2020); Wauquier et al. (2020); Huyghe & Wauquier (2020, 2021); Varvara et al. (2021), whereas (Mickus et al., 2019; Guzmán Naranjo, 2020; Guzmán Naranjo & Bonami, 2021) use it to study inflection. We cannot do a complete review of literature on the matter in this paper, for a more exhaustive discussion of how distributional semantics has been used in linguistic theory see Boleda (2020).

We are now in a position to state our research question in terms of distributional semantics. Remember that we want, for a pair of processes that apply to the same bases, to compare the semantic interpredictability of two derivatives among themselves to how predictable each derivative is of its base. To this end we can set up partial paradigmatic systems in the sense of Bonami & Strnadová (2019) consisting of triples of words (x_i, y_i, z_i), such that all (x_i, y_i) pairs instantiate the same derivational process, and likewise all (x_i, z_i) pairs instantiate a second process, as shown in Table 1.

We now can tabulate the corresponding triples of distributional vectors $(\vec{x}_i, \vec{y}_i, \vec{z}_i)$ and compare the average predictability of each of the morphosemantic relations represented in the table. We are on the lookout for situations such as

Tab. 1: Abstract representation of a set of triples of words exemplifying two processes applying to the same base.

Base	**Process**$_1$	**Process**$_2$
$\mathbf{x}_1$	$\mathbf{y}_1$	$\mathbf{z}_1$
$\mathbf{x}_2$	$\mathbf{y}_2$	$\mathbf{z}_2$
$\mathbf{x}_3$	$\mathbf{y}_3$	$\mathbf{z}_3$
...	...	...

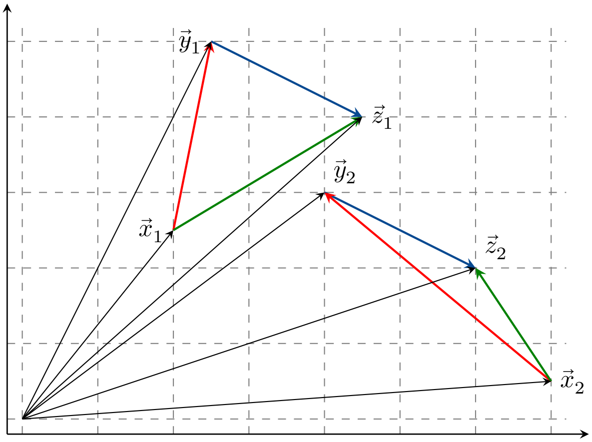

Fig. 13: Situations of interest, where derivatives are more interpredictable (blue vectors) than either is from their common base (red and green vectors).

that illustrated in Figure 13: here, the $\vec{y}_i$s and the $\vec{z}_i$s relate to one another in a (more or less) uniform way, while the relations between $\vec{x}_i$s and the $\vec{y}_i$s vary widely; likewise for the relations between $\vec{x}_i$s and the $\vec{z}_i$s. If such situations can be documented, then they constitute strong evidence for models that take into account paradigmatic relations.[11] Note that what matters here is whether the difference vectors $\vec{z}_1 - \vec{y}_1$ and $\vec{z}_2 - \vec{y}_2$ are similar to one another, not whether $\vec{y}_1$ and $\vec{z}_1$ (or $\vec{y}_2$ and $\vec{z}_2$) are similar: we aim to assess the systematicity of contrasts across morphological families, not the proximity between members of a family.[12]

Before proceeding to explain how we implemented this idea, it is worth emphasizing what the structure of our argument is. Our aim is to compare the prediction of a classical view of derivation, where only (base, derivative) relations constitute linguistic knowledge, to that of a paradigmatic model, where all relations between pairs of lexemes in a family may constitute such knowledge. The prediction of a classical model is that a base should always be the best predictor of the properties of its derivatives; the prediction of a paradigmatic model is that this need not be the case, not that it can't be the case. Hence any situation of the type represented in Figure 13 is evidence for paradigmatic models. Situations where the base is the best predictor are irrelevant to the comparison, as the two types of models do not make contrasting predictions on these. Despite this, we will report on all the comparisons we have conducted, not

11 Note that such situations are the analogue for predictability of meaning of the situations documented by Bonami & Strnadová (2019) for predictability of form: the base is a poorer predictor of properties of a derivative than another member of the derivational family.

12 In a study of eventive nominalizations in German, Varvara et al. (2021) rely on comparisons of base and derivative vectors to evaluate the average transparency of different rival processes. It is worth clarifying why we do not use the same methodology and follow instead the lead of Marelli & Baroni (2015); Bonami & Paperno (2018); Mickus et al. (2019). First, Varvara et al.'s methodology makes sense when comparing *rival* processes that have the same types of inputs and outputs. This is not the case for us, leading to problems: the similarity between pairs of vectors is bound to be heavily influenced by the corresponding words sharing a part of speech or ontological type. Hence we expect e.g. deverbal agent and event nouns to be more similar to one another than they are to their base, just by virtue of both being nouns. Likewise we expect verbs and event nouns to be more similar than verbs and agent nouns, because they both denote eventualities. Clearly these facts are orthogonal to our research question, which pertains to the diversity of semantic relations between pairs of words that stand in the same formal relation. Second, Varvara et al.'s (2021) careful study led to mixed results, and in particular did not confirm the hypothesis that average cosine similarity between base and derivative reflects intuitions on the relative transparency of different processes. This contrasts with the robust results of Marelli & Baroni (2015) and Bonami & Paperno (2018) relying on basically the same methodology we are adopting here.

just those that provide evidence on our main research question, as these turn out to raise interesting questions for future research.

4 Materials and methods

All materials used in this study (datasets, vector spaces, and R scripts) are available at https://zenodo.org/record/5799577.

4.1 Dataset

For this project we compiled a dataset of derivational processes in French by combining information from various sources: Hathout & Namer (2014a) for relations between verbs and nouns; Tribout (2010b) for nouns and verbs related by conversion; Koehl (2012) for deadjectival nouns; Strnadová (2014a) for derived adjectives; and Bonami & Thuilier (2019) for derived verbs in *-iser* and *-ifier*. To these we added new datasets on derived nouns in *-isme*, *-iste*, *-ier* and *-erie* automatically extracted from the GLÀFF lexicon (Hathout et al., 2014) and curated by hand.

The analysis focuses on the processes for which at least 50 (base,derivative) pairs were present in the resources, such that both lexemes are attested at least 5 times in the FRCOW corpus (Schäfer & Bildhauer, 2012). For those processes admitting bases in more than one part of speech, (e.g. *-age*: *laver* 'wash' > *lavage* 'washing', *feuille* 'leaf' > *feuillage* 'foilage'), if one part of speech accounts for 90% or more of the types, only types where the base has that part of speech were kept; this led to dismissing denominal *-age* and *-eur* derivatives but keeping *-isme* and *-iste* derivatives with both nominal and adjectival bases. Overall, this initial selection step resulted in a cumulative dataset of $21,990$ (base, derivative) pairs each exemplifying one of 35 derivational processes.

From this dataset we extracted paradigmatic systems corresponding to two processes sharing a base, and selected for analysis those for which more than 150 triples (base, process 1 derivative, process 2 derivative) were documented. Table 2 shows the selected pairs of processes and the number of triples documenting each. As can be seen, almost all processes derive nouns, with the exception of *-ant* which builds deverbal adjectives. Table 3 shows a sample triple for each pair of processes. Note that action noun forming processes are overrepresented in the sample, and that some pairs of processes are rivals (e.g. age:V>N, ment:V>N) for which there happens to be a large enough number of doublets in our data sources for comparison to be possible.

Tab. 2: Selected pairs of processes with the number of corresponding lexeme triples in the final datasets

Process$_1$	Process$_2$	Sample size
age:V>N	conversion:V>N	833
age:V>N	eur:V>N	584
age:V>N	ment:V>N	354
ant:V>A	ment:V>N	302
conversion:V>N	eur:V>N	679
conversion:V>N	ment:V>N	377
ier:N>N	erie:N>N	151
eur:V>N	ion:V>N	514
eur:V>N	ment:V>N	342
isme:A/N>N	iste:A/N>N	277

Tab. 3: Sample triple for each selected pair of processes

Processes	Base	Derivative$_1$	Derivative$_2$
(age:V>N, conversion:V>N)	givrer	givrage	givre
(age:V>N, eur:V>N)	racler	racleur	raclage
(age:V>N, ment:V>N)	encaisser	encaissage	encaissement
(ant:V>A, ment:V>N)	percer	perçant	percement
(conversion:V>N, eur:V>N)	découvrir	découverte	découvreur
(conversion:V>N, ment:V>N)	défausser	défausse	défaussement
(ier:N>N, erie:N>N)	verre	verrier	verrerie
(eur:V>N, ion:V>N)	dévorer	dévoreur	dévoration
(eur:V>N, ment:V>N)	porter	porteur	portement
(isme:A/N>N, iste:A/N>N)	Europe	européisme	européiste

We build a 100-dimensional vector space using a modified version of FrCow (Schäfer, 2016) and the skipgram variant of the Word2Vec algorithm (Mikolov et al., 2013a).[13] These vectors are lexeme-based rather than word-based: we

13 We used the gensim implementation of Word2Vec (Řehůřek, 2010). Hyperparameters were as follows: 100 dimensions, 10 iterations, negative sampling, a window of 5 tokens and a minimum occurrence of 5 tokens. We chose these parameters after several tests, and based on previous experience. In particular, we experimented with both the cbow and the did not apply Word2vec, to the the wordforms, but to the concatenation of the word's lemma and its part of speech tag, as provided by the corpus. This is justified by the goal of studying lexeme-level derivational properties, and hence abstracting away from the special distributional properties of citation forms. It also maximizes the number of tokens representing each of the lexemes of interest.[14]

4.2 Assessing semantic predictability

To be able to compare semantic predictability across morphological relations, we first need to lay out a method for assessing precitability in each individual case. That is, given a set of pairs of derivationally related predictor and target lexemes $\{(p_1, t_1), \dots, (p_n, t_n)\}$, we need to define a prediction algorithm mapping vectors for p_is to vectors for t_is, and assess the quality of the result for each pair.

The simplest way of doing this relies directly on Mikolov et al.'s ideas on semantic analogies already discussed in Section 3: To predict $\vec{t}_i$ from $\vec{p}_i$, pick another pair of words linked by the same morphological relation (p_j, t_j), and compute the vector $\vec{p}_i + \vec{t}_j - \vec{p}_i$. This simple method is quite noisy, as idiosyncratic

skipgram variant of the algorithm for this and other studies. Experience shows that vectors obtained with skipgram give rise to better performance for vector-to-vector prediction tasks. Both the skipgram and the cbow vector space are available for examination on the Zenodo repository.

14 Initial experimentation showed that vectors tend to overemphasize the role of grammatical gender: grammatical gender is by far the most prominent distributional distinction among nouns. Since our vectors are lexeme-based, this should not be the case (inflectional variation in determiners, adjectives and verbs agreeing in gender with nouns should be neutralized), and points to poor lemmatization. Although this is not crucial for the present study, it is for a separate study for which we used the same vectors. Accordingly, we slightly modified the corpus to compensate for that effect. First, portmanteau forms like *du* 'of_the.MAS', which had been tokenized as a single form and thus lemmatized as *du*, were re-lemmatized as *de* + *le*. This avoids an asymmetry between the lemmatization of masculine *du* and feminine *de la* 'of the.FEM'. Second, lemmatization of feminine nouns and adjectives was not coherent: sometimes feminine nouns (e.g. *institutrice* 'female teacher' were lemmatized to the corresponding masculine noun (e.g. *instituteur* 'male teacher' rather than seen as their own lemma; while sometimes feminine adjectives were lemmatized to the feminine (e.g. *gibbeuse* 'gibbous.FEM' rather than the conventional masculine form (e.g. *gibbeux* 'gibbous.FEM'). We corrected such cases to the extent possible by checking the lemmatization against the large lexicon in development in the Démonext project (Namer et al., 2019) and correcting automatically the lemmatization through string searches where that was possible. Examination of a random sample suggests that the number of remaining gender-related lemmatization errors is below 1%.

properties of the pair of lexemes used for analogy (p_j, t_j) will inevitably have an influence on quality of prediction. This problem can be mitigated by using a set of pairs of lexemes rather than a single pair as our analogical base, and derive from this the average difference vector between predictors and targets. By adding this average difference vector to $\vec{p}_i$, we get a predicted value for $\vec{t}_i$ which evens out idiosyncratic differences between pairs of lexemes standing in the same morphological relation. This is the method used e.g. in Mickus et al. (2019).

Marelli & Baroni (2015) identify a potential problem with this and related methods and propose an elegant solution.[15] They note that derivational morphology often leads to predictably different outcomes depending on semantic properties of the base it applies to; for instance, the English prefix *re-*, giving rise to iterative readings when attached to an activity verb (*resing* means 'sing one more time') but to a restitutive reading when attached to an accomplishment verb (*reopen* means 'open what was previously closed'). Compare such a case with that of the polysemy of the suffix *-er*, which readily forms either agent or instrument nouns from verbs. In both cases we have a form of affix polysemy, but in the former, the output meaning is predictable from semantic properties of the base, whereas in the latter, it is not (or at least not to the same degree). This is precisely a difference in semantic predictability that the average difference vector method is unlikely to be able to capture.

To avoid that problem, Marelli & Baroni (2015) relate predictor and target vectors using a linear transformation. That is, instead of predicting each dimension in the target vector from just the same dimension in the predictor, each dimension in the target is predicted by a linear model taking all dimensions of the predictor as input. Thus there are as many linear models as there are vector dimensions; in our case this would lead to 100 models with the following structure in R formula notation:

$$
\begin{aligned}
\texttt{target\_val\_1} &\sim \texttt{pred\_val\_1} + \texttt{pred\_val\_2} + \cdots + \texttt{pred\_val\_100} \\
\texttt{target\_val\_2} &\sim \texttt{pred\_val\_1} + \texttt{pred\_val\_2} + \cdots + \texttt{pred\_val\_100} \\
&\vdots \\
\texttt{target\_val\_100} &\sim \texttt{pred\_val\_1} + \texttt{pred\_val\_2} + \cdots + \texttt{pred\_val\_100}
\end{aligned}
$$

With our data this tended to overfit the models, which resulted in relatively poor predictions of unseen test items. To compensate for this, we used a model

15 Marelli & Baroni (2015) actually discuss this in the context of a critique of yet another way of addressing the semantics derivational morphology, by building a vector for an affix from the distribution of all words containing that affix, and then combining base and affix vectors using a linear model (Lazaridou et al., 2013). However their argument applies *mutatis mutandis* to the average difference vector approach.

structure that is conceptually similar to Marelli and Baroni's, but less rich. We first computed the 10 main principal components of the 100 dimensions for the vectors. Then we trained a single model predicting, for each dimension, the target value from the predictor value and the 10 principal components of the full predictor vector. This is shown in the formula below, where `dimension` is a nominal variable indicating the dimension of interest, `target_val` and `pred_val` are the respective values of the target and predictor vector for that dimension, and `PC1`, …, `PC10` are the 10 main principal components of the input vector.

$$\texttt{target\_val} \sim \texttt{pred\_val} * \texttt{dimension} + \texttt{PC1} + \texttt{PC2} + \cdots + \texttt{PC10}$$

Note that their are two differences between our modelling strategy and Marelli & Baroni's: the use of principal components rather than full vectors, and the use of a single model to predict all dimensions rather than one model per dimension.

We then used 10-fold cross-validation to assess how well the model performed on unseen data. We split the data into 10 groups, and fitted the model using 9 of the groups. We then try to predict the left out group, and repeat for all 10 groups, and for all (predictor,target) pairs. To evaluate how well the model performed, we calculated the cosine similarity between the cross-validated predicted vector and the actual vector for each target.

4.3 Comparing predictability across morphological relations

Remember that we want to test the hypothesis that for some pairs of processes applicable to the same base, the semantic relationship between the derivatives is more predictable than the relationship of either derivative to their base. To that effect we need to assess semantic predictability between pairs of cells within a small paradigmatic system consisting of aligned sets of three forms, a base and two derivatives. This entails that, for each pair of processes under consideration, we have six sets of (predictor, target) pairs to consider; these are depicted graphically in Figure 14. Our main goal is to compare the accuracy of the $D_1 \rightarrow D_2$ and $D_2 \rightarrow D_1$ prediction relations to that of predictions from the base ($B \rightarrow D_1$, $B \rightarrow D_2$). However we will also comment on prediction of the base from the derivatives ($D_1 \rightarrow B$, $D_2 \rightarrow B$).

As a concrete example, for the process pair (*-eur*, *-ment*), we build six models (all with crossvalidation): `base>eur`, `eur>base`, `base>ment`, `ment>base`, `ment>eur`, and `eur>ment`. The model `eur>ment` learns to predict *-ment* from *-eur* forms, and we evaluate its performance as the cosine between the predicted vector and the actual observed vector. Table 4 shows the result on a sample triple of

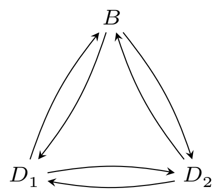

Fig. 14: The six prediction relations within a system of two derivational processes applying to the same bases

Tab. 4: Illustration of model performance evaluated as cosine similarity between predicted actual target vectors

Prediction relation	Sample predictor	Sample target	Sample performance	Average performance
base>eur	*accorder*	*accordeur*	**0.640**	**0.676**
eur>base	*accordeur*	*accorder*	**0.753**	**0.689**
base>ment	*accorder*	*accordement*	**0.849**	**0.633**
ment>base	*accordement*	*accorder*	**0.869**	**0.637**
eur>ment	*accordeur*	*accordement*	**0.712**	**0.615**
ment>eur	*accordement*	*accordeur*	**0.493**	**0.600**

lexemes. Here we see that, for this triple, we get best performance for prediction of the *-ment* derivative from the base verb, and worse performance for prediction of the *-eur* derivative from the *-ment* derivative. Note also the asymmetry of performance scores, which varies from minimal when comparing `base>ment` to `ment>base`, to major when comparing `eur>ment` to `ment>eur`. This might seem counterintuitive but is to be expected: the first row in the table reports the cosine similarity between a predicted vector and the actual vector for *accordeur*, while the second reports the cosine between between a predicted vector and the actual vector for *accorder* (we are not using cosine to compare the vectors for two words, but two vectors for the same word). Likewise, the first and last rows report different sample performances, despite the fact that the target and hence the actual vector in the comparison is the same: on the first row the predicted vector stems from a model trained on `base>eur` pairs and applied to the actual vector for *accorder*, while on the last row the predicted vector stems from a model trained on `ment>eur` pairs and applied to the actual vector for *accordement*

We thus have an estimated cosine similarity between a predicted vector and an actual vector for each model and each (predictor,target) pair. At this point we could try to address our research question directly by comparing the average cosine similarity of each of the six models documenting a pair of process. By way of an example, The last column of Table 4 shows such averages for the models under consideration, suggesting a better accuracy of prediction of derivatives from their bases than among themselves.

We want to be more cautious, however, and to evaluate how certain we are about these averages cosine similarities. We expect that there will be some degree of randomness in the prediction results stemming from at least two sources. First, the accuracy of the semantic distributional vectors themselves is not the same for all words, because Word2Vec models build better representations for more frequent words. Second, there is some random variation in the predictions stemming from the cross-validation. For this reason, comparing mean cosine distances directly would be careless. For example, considering again the numbers Table 4, we would want to know how confident we are that the predictive performance of the `base>ment` model, estimated at 0.633, is indeed higher than that of the `eur>ment` model, estimated at 0.615. Similarly, we want to take into account the variance around the mean: we can be more certain about a mean value if there is little variance across the individual estimates.

To quantify our uncertainty about the individual mean cosine estimates, we fit a separate Bayesian Beta regression for each of the 10 datasets under consideration. Remember that for each dataset we have 6 linear models producing predicted vectors for each of the 6 pairwise prediction relations between a base and two derivatives. The Bayesian model estimates the mean cosine similarity between predicted vector and actual vector on the basis of just a nominal variable indicating which of the 6 prediction relation this measurement exemplifies. That is, it predicts the Cosine column from the Relation column in the sample dataset in Table 5.[16]

Because we are working with cosine similarities, all values are bounded between 0 and 1, which means that we can use a Beta distribution as our likelihood.[17] We thus used brms (Bürkner, 2017, 2018) and Stan (Carpenter et al., 2017) to fit a Beta regression model to each of the 10 datasets. In addition to an estimation of mean cosine similarity for each of the 6 prediction relations, the model assesses the uncertainty of that estimation in the form of a posterior distribution, which allows us to assess how strong the evidence for a difference in mean is.[18]

16 We tried adding the log frequency of the base and derived forms as predictors to see whether this had any effect the accuracy of the predictions but we were not able to find any noticeable effects.

17 Technically cosine similarities can include 0 and 1, which is not covered in Beta regression, but in practice our data contained neither zeroes nor ones.

18 See Gelman et al. (2013) and Gelman et al. (2020) for justification of the use of regression models rather than classical hypothesis testing to assess uncertainty.

Tab. 5: Sample input to the Bayesian model for the pair of processes (*-eur*, *-ment*). The predictor and target for which the performance observation was computed are shown for illustration, they are not part of the input to the model.

Relation	Predictor	Target	Cosine
base>eur	*accorder*	*accordeur*	**0.640**
base>eur	*verser*	*verseur*	**0.580**
...	...	...	...
base>ment	*accorder*	*accordement*	**0.849**
base>ment	*verser*	*versement*	**0.737**
...	...	...	...
texttttment>eur	*accordement*	*accordeur*	**0.493**
ment>eur	*versement*	*verseur*	**0.629**
...	...	...	...

5 Results

In this section we discuss the results of the models. Instead of presenting coefficient tables we built conditional effects plots. A conditional effect plot shows the expect mean value for each predictor level, as well as the 95% uncertainty interval (the interval where 95% of the posterior probability density lies). We conclude that we have clear enough of a difference between two estimated means if the uncertainty intervals do not overlap. Figures 15-22 show the conditional effects plots for four pairs of processes which illustrate our findings. Plots for the remaining pairs of processes can be found in the appendix.

First, Figure 15 shows the conditional effects plot for (*-eur*, *-ment*). We observe that predictions between the base an *-eur* or *-ment* derivative are clearly better than predictions between the two derivatives. Hence this pair of processes do not provide evidence for the importance of a paradigmatic relation between derivatives: formal bases are the best predictors of their derivatives.

In contrast, Figure 16 shows the best example of paradigmatic effects in our sample. In this case, predicting the *-isme* lexeme from *-iste* or the other way around, was considerably easier than predicting either from the base. Interestingly, in this case we also see that going from the semantics of the *-isme* and *-iste* derived forms to the semantics of the base is considerably harder than going from the base to the derived form.

A similar, although less extreme situation, visible with *-ier* and *-erie*, as shown in Figure 17. For this pair of processes, predicting between derived forms was easier than predicting from the base, but the evidence is far from being overwhelming: there is considerable overlap in the posterior distribution. Hence,

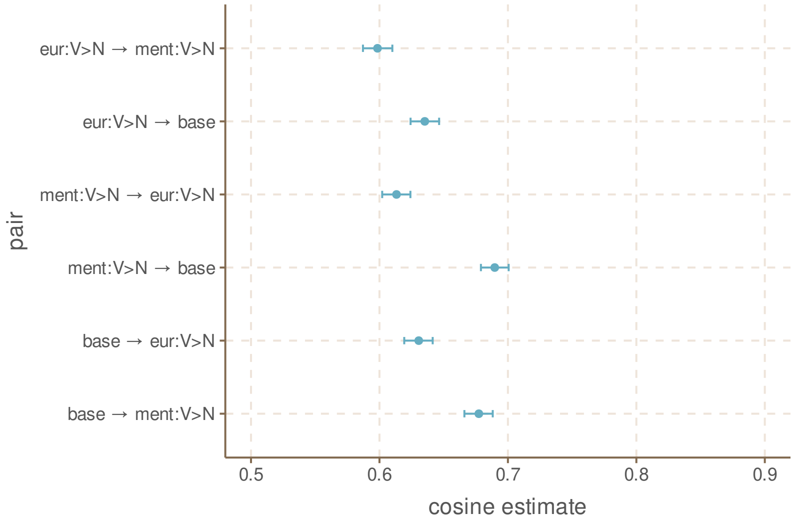

Fig. 15: Comparing -*eur* and -*ment* suffixation

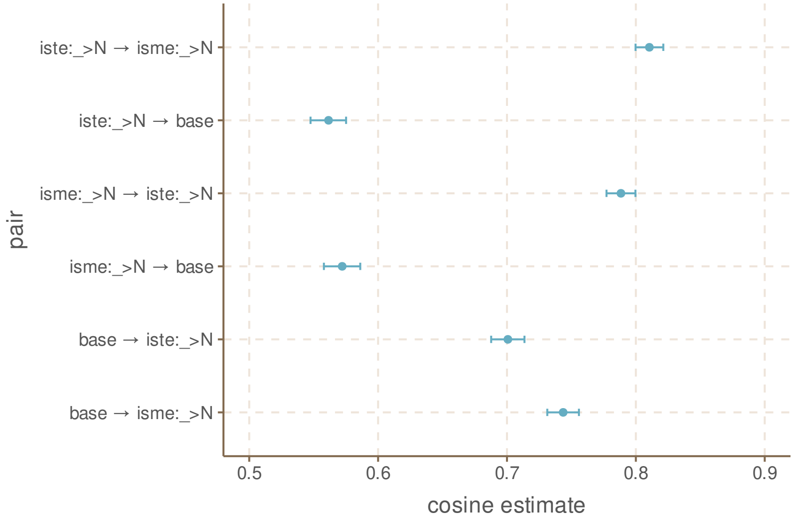

Fig. 16: Comparing -*isme* and -*iste* suffixation

even though we do not see a big asymmetry as with *-isme* and *-iste*, the evidence ponts in the direction of a paradigmatic effect. On the other hand, and predicting the base semantics from the semantics of the derived forms was somewhat harder.

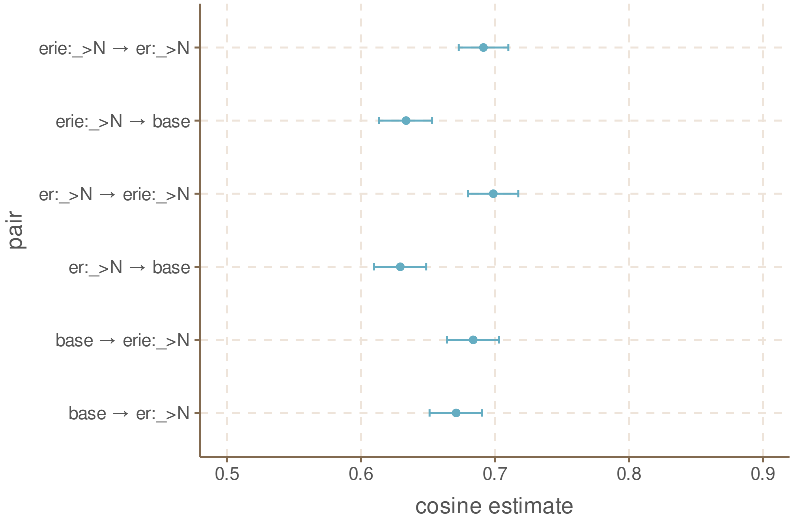

Fig. 17: Comparing *-ier* and *-erie* suffixation

The cases discussed so far have one property in common, namely that predicting derivative from base is easier than or equally difficult as predicting base from derivative. Interestingly, this does not hold for all pairs of processes. Figure 18 shows the effects for conversion and *-ment*. In this case we see that predicting between the derivatives is harder than predicting from the base, but also that predicting the base semantics from derivatives is equally hard as the other way around. This suggests that there is not a clear directionality of prediction between base and derivative. While this is an interesting observation worthy of further investigation, it has no bearing on the theoretical distinction between classical and paradigmatic approaches.

The preceding examples illustrate all the situations found in the data. Remaining plots can be found in the appendix, but are summarized in Table 6. For each pair of processes, the table indicates in the first two columns whether prediction of each derivative is easier from the other derivative or from the base; a superior sign (‘>’) indicates that there is clear evidence that it is, that is, the measured estimate is higher and the uncertainty intervals do not overlap. An

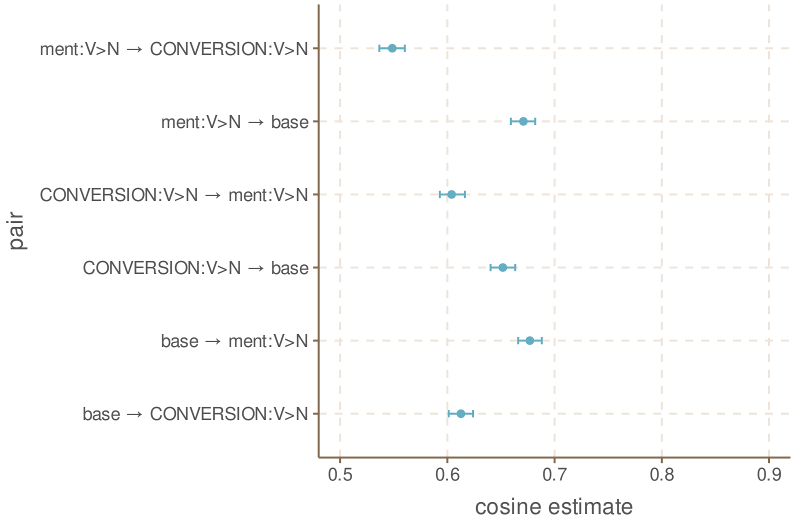

Fig. 18: Comparing conversion and *-ment* suffixation

equal sign indicates that there is no clear evidence either way, as the intervals overlap; an inferior sign indicates that the evidence goes in the opposite direction. Using the same conventions, the last two columns indicate whether prediction of a derivative from its base is easier than prediction of the base from the derivative.

Tab. 6: Summary of the evidence for paradigmatic relations

	$D_2 \to D_1$ vs. $B \to D_1$	$D_1 \to D_2$ vs. $B \to D_2$	$B \to D_1$ vs. $D_1 \to B$	$B \to D_2$ vs. $D_2 \to B$
(age:V>N, conversion:V>N)	$<$	$<$	$=$	$<$
(age:V>N, eur:V>N)	$=$	$>$	$<$	$=$
(age:V>N, ment:V>N)	$<$	$<$	$<$	$=$
(ant:V>A, ment:V>N)	$<$	$<$	$=$	$=$
(conversion:V>N, eur:V>N)	$<$	$=$	$>$	$<$
(conversion:V>N, ment:V>N)	$<$	$<$	$>$	$=$
(er:N>N, erie:N>N)	$=$	$=$	$<$	$<$
(eur:V>N, ion:V>N)	$=$	$<$	$=$	$=$
(eur:V>N, ment:V>N)	$=$	$<$	$=$	$=$
(isme:N>N, iste:N>N)	$>$	$>$	$<$	$<$

As the reader can check verify, only 1 out of the 10 pairs of processes provides clear evidence for a paradigmatic effect (two '>' signs in the first two columns); however, only 4 are such that the base is clearly a better predictor of a class of derivatives than another member of the morphological family (two '<' signs in the first two columns). In the remaining 5 cases, the evidence is somewhat mixed, and the base has no clear privileged predictive status.

6 Discussion

In Section 2 we contrasted the predictions of a theory of morphology purely based on oriented base-derivative relations to one which recognizes other kinds of paradigmatic relations. The former entails that, although there can be accidental exceptions, at the level of the system the base should always be the best predictor of a derivative within its family. The latter view on the other hand suggests that for some processes this does not hold. Note the important lack of symmetry: the paradigmatic view does not entail that formal bases are *never* good predictors, but only that they need not be.

The results presented in section 4 hence provide exactly the kind of evidence we were looking for: paired nouns in *-isme* and *iste* are undoubtedly better predicted by each other than by their common formal base. The data for *-ier* and *-erie* nouns points in the same direction, although the evidence is less clear, probably because of a smaller dataset. Note that it is unsurprising that *-isme* and *-iste* exhibit paradigmatic effects, as they are the poster child for paradigmatic relations in derivation (see e.g. Becker 1993; Booij 2010; Roché 2011a). On the other hand, we are basing our conclusions here only on those cases where a formal base is present, whereas much of the usual argumentation on these nouns focuses on the large number of cases where no formal base is to be found. The novelty of this paper is to provide an operational method to investigate contrasts of semantic predictability in derivational families, which is only part of the relevant evidence. The fact that none of the remaining 8 pairs of processes led to similar results is not particularly concerning for the paradigmatic view, which predicts the *existence* of strong predictability among derivatives, not their high prevalence; after all, if they were highly prevalent, there would be no debate. However, the gradient of predictability we have observed suggests a direction for future work: once we have clearly established the reality of systematic paradigmatic relations among derivatives, it is worth exploring why these are on average less reliable than those between bases and derivatives.

While this does not directly address our main research question, another interesting observation concerns the interpredictability of bases and derivatives. Our expectation is that bases be better predictors of their derivatives than the other way around. Derivatives tend to be less polysemous than their bases, either because the derived meaning elaborates on a single sense of the base (Fradin & Kerleroux, 2003); or, when that is not the case, because the distribution of senses is less uniform in derivatives (Anselme et al., 2021). In addition, derivatives tend to be less frequent than their bases (Harwood & Wright, 1956; Hay, 2001), giving them less room for a variety of uses (and in particular for polysemy). Finally, Kotowski & Schäfer (2023) provide direct distributional evidence that derivatives resulting from the same process tend to be semantically less diverse than the bases they derive from. As a result of these observations, we expect the larger diversity of base semantics to impair the predictability of their meaning from the derivative's meaning.

This prediction is largely borne out in our data: we find clear evidence for higher predictability in the base-derivative direction in 8 cases out of 20, and clear evidence to the contrary in only 2 cases, which both correspond to different samples of verb to noun conversion. This is probably not an accident. First, conversion is generally assumed to be more polysemous than affixal processes (Plag, 1999; Tribout, 2010a), to the point that some have argued it to be essentially semantically unconstrained (e.g. Clark & Clark 1979 and Aronoff 1980 on noun to verb conversion). Second, deciding on the orientation of conversion is notoriously difficult: for the relationship between nouns and verbs in French, after careful empirical consideration, (Tribout, 2020) concludes that orientation is undecidable in a majority of cases. These two observations lead us to expect conversion to behave differently from other processes: while it would be very interesting to explore the situation in more detail, and understand why and how exactly conversion is different, the empirical results on directionality are overall compatible with our expectations.

One limitation of the present study is that we have not provided evidence that the cosine similarity between predicted and actual vectors does indeed reflect semantic unpredictability. Establishing this is not trivial: the quality of our vectors is clearly variable, all the more so because we used a very low frequency threshold for inclusion in the dataset, in the interest of getting enough types. As a result, there are many individual cases where it is unclear why the words are judged to be predictable or unpredictable. Given this situation, only a quantitative evaluation of average correlation between human unpredictability judgements and model predictions would truly allow us to assess quality. This is clearly beyond the scope of the present paper.

In the absence of such hard evidence, the best we can do here is to provide some circumstantial evidence that the models are capturing relevant distinctions. Focusing on *-isme* and *-iste* derivatives, we looked manually at extreme cases of quality of prediction of *-iste* derivatives from the base or from the *-iste* derivative. Among the 10 *-iste* derivatives that are best predicted from the base, we find cases like *biologiste* 'biologist' and *dynamiste* 'dynamist'. In the first instance, the corresponding *biologisme* names a specific philosophical doctrine. The noun *biologiste* can name a follower of that doctrine, but it also has a much more general and frequent meaning transparently related to that of the base *biologie* 'biology'. Conversely, *dynamiste* names a follower of the doctrine of *dynamisme*, but *dynamisme* 'dynamism' also has a more general and more frequent meaning that relates readily to the base adjective *dynamique* 'dynamic'. At the other end of the spectrum, cases where predictability of the *-iste* derivative from the *isme* derivative is maximal while predictability from the base is minimal typically involve an adjectival base with a fairly broad meaning (*social* 'social', *individuel* 'invidual') or that is clearly polysemous (e.g. *gauche* 'left side', 'political left') and the *-isme* and *-iste* derivatives name two closely related concepts of a doctrine and a follower of that doctrine, that is more specialized than the meaning of the base.

To sum up then, although a more thorough evaluation would be in order, a superficial examination of model predictions suggests that these do capture the relative predictability of different pairs of lexemes, and hence that these can be trusted as informing us on the existence of semantically stable paradigmatic relations.

A second limitation of this study is that we ignored two important phenomena at the core of contemporary research on derivational semantics, as is evident from other chapters in this volume. First, affix polysemy (see e.g. Lieber 2023; Plag et al. 2023; Schäfer 2023) is not taken into account directlty: our vectors lump together tokens of a lexeme corresponding to different senses, and no effort has been made to select monosemous lexemes for analysis. It is hence possible that our results are influenced by different affixes exhibiting different degrees of polysemy, leading to different levels of predictability. Second, we also abstracted away from affix rivalry (see e.g.Plag et al. 2023; Huyghe et al. 2023), which could have an incidence on predictability among derivatives: if there is truly rivalry between two processes, then we expect the semantic contrasts between doublets to be random, leading to poor interpredictability. In both cases, we were limited by the lack of good operationalizations of the relevant variables (sense distributions in corpus data, degree of polysemy of affixes, degree of rivalry of affixes). We are hopeful that future research will overcome these limitations.

7 Conclusion

In this paper we first reviewed the literarure on derivational paradigms. After exhibiting various types of morphological facts that are taken as evidence for paradigmatic relations in derivation, we highlighted the existence of two different tradition for integrating these relations in morphological theory, typified by seminal work by van Marle (1984) and Robins (1959) respectively. We then argued that while the literature provides many examples of anecdotal evidence for the importance of paradigmatic relations, systematic evidence was required for the relevant phenomena not to be dismissed as epiphenomenal and hence outside the focus of morphological theory.

We proposed to use distributional semantics to provide exactly that type of evidence: systematic paradigmatic relations should give rise to strong semantic predictability among derivatives, which can be operationalized using distributional methods. We went on to examine 10 pairs of processes in French derivational morphology, and showed that for at least one of these, we have clear evidence that on average, pairs derivatives are more interpredictable than either is predictable from their common base. This is precisely the kind of systematic paradigmatic relation we were looking for. In that sense, we have thus provided strong evidence that paradigmatic relations among derivatives from a common base cannot be dismissed as epiphenomenal.

It is worth noting that we have only examined one type of evidence bearing on the reality of derivational paradigms. Although various configurations challenging the conventional view of derivation as grounded in oriented base-derivative relations were identified in section 2, we focused entirely on the particular situation of paired derivatives with a common and well-attested base. As such our conclusions are inherently limited: in particular they have little bearing on the distinction between approaches that take paradigmatic relations to be secondary to ordinary morphology, in the tradition of van Marle (1984), and those that conceptualize all derivational relations as paradigmatic, in the tradition of Robins (1959). However, we have exemplified how distributional methods can be used to bear on investigating structured morphological relatedness quantitatively. We hope this study to lay the groundwork for a larger research program that would address empirically the challenges raised by paradigmatic structure in word formation.

Acknowledgment: We thank the audience from the workshop at the workshop on Derivational Semantics, 2021 DGfS meeting, as well as two anonymous reviewers and the editors for the present volume for useful discussion and comments.

Funding: This research was partially conducted while Matías Guzman Naranjo was a postdoc at Université de Paris. 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).

Appendix: remaining conditional effect plots

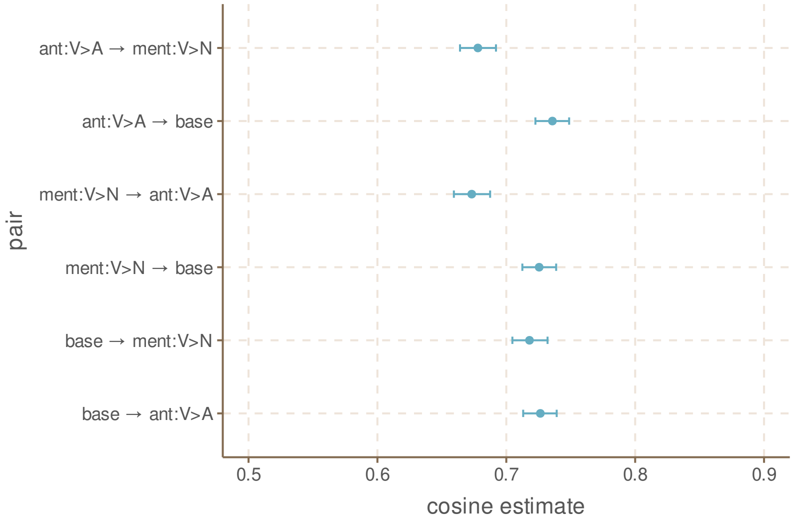

Fig. 19: Comparing *-ant* and *-ment* suffixation

Bibliography

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar*, 54–82. Oxford: Oxford University Press.

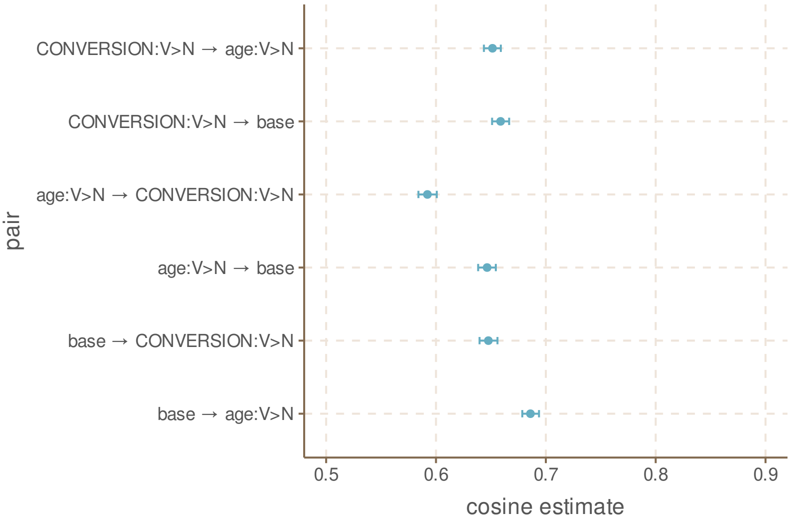

Fig. 20: Comparing conversion and *-age* suffixation

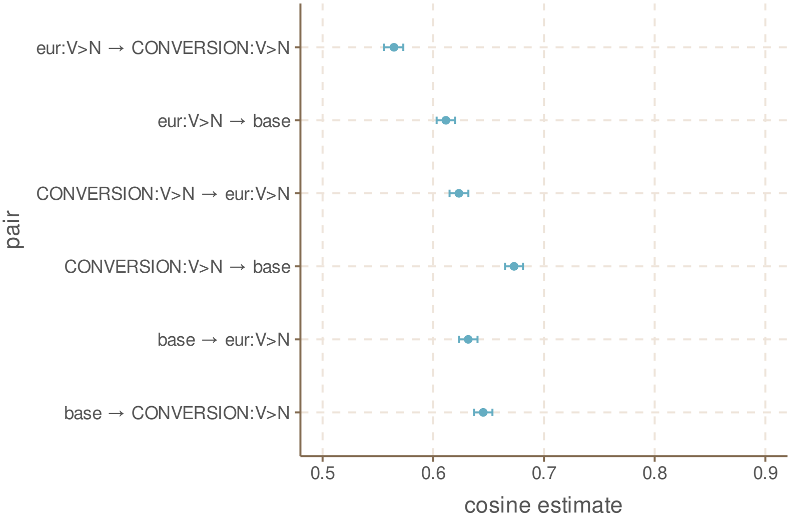

Fig. 21: Comparing conversion and *-eur* suffixation

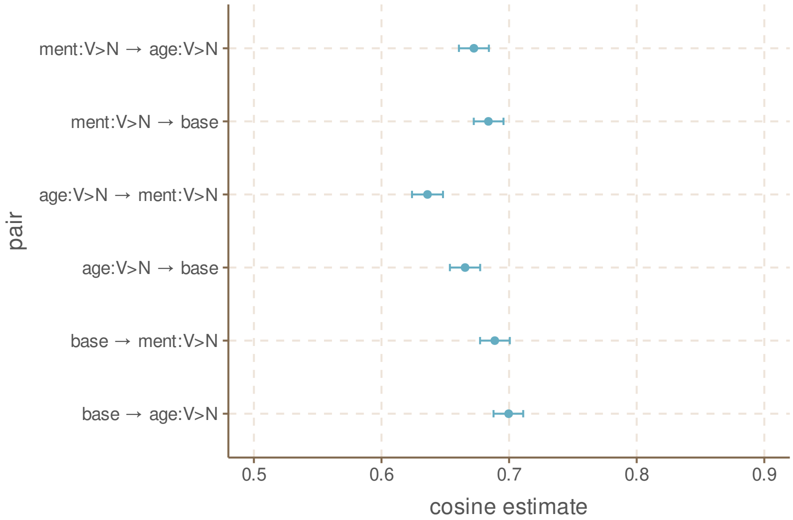

Fig. 22: Comparing *-age* and *-ment* suffixation

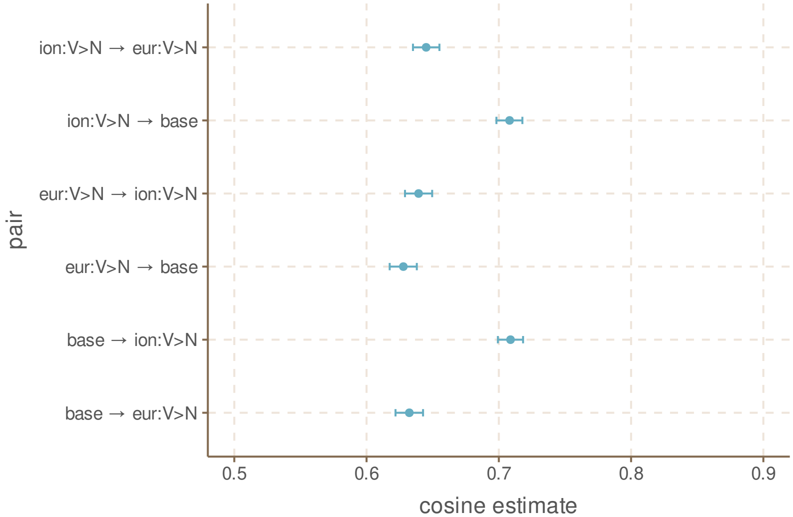

Fig. 23: Comparing *-eur* and *-ion* suffixation

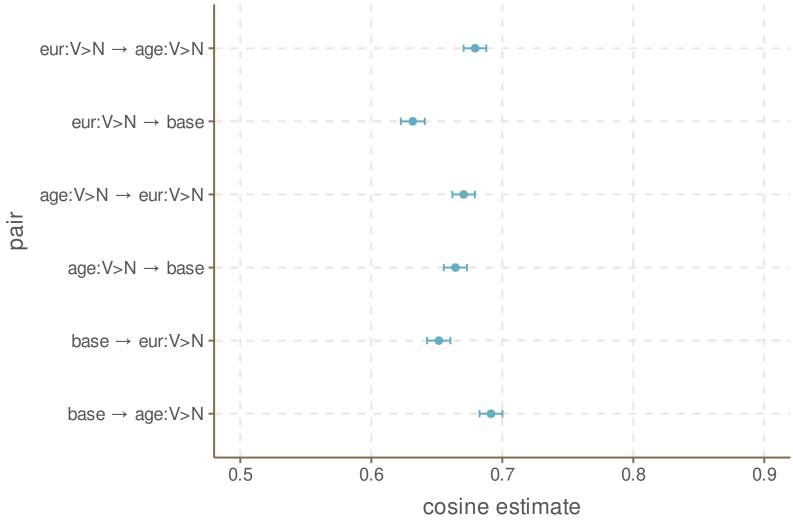

Fig. 24: Comparing *-age* and *-eur* suffixation

Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: the low conditional entropy conjecture. *Language* 89. 429–464.

Amenta, Simona, Fritz Günther & Marco Marelli. 2020. A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon* 15(1). 62–78.

Anselme, Rémi, Olivier Bonami & Heather Burnett. 2021. Polysémie et troncation des noms en -ion en français. *Verbum* XLIII(1). 97–118.

Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge: MIT Press.

Aronoff, Mark. 1980. Contextuals. *Language* 744–758.

Aronoff, Mark. 1994. *Morphology by itself*. Cambridge: MIT Press.

Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word Structure* 12(3). 274–296.

Barbu Mititelu, Verginica, Gianina Iordăchioaia, Svetlozara Leseva & Ivelina Stoyanova. 2023. The meaning of zero nouns and zero verbs. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Baroni, Marco, Georgiana Dinu & 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)*, 238–247. Baltimore, Maryland: Association for Computational Linguistics. 10.3115/v1/P14-1023. https://aclanthology.org/P14-1023.

Bauer, Laurie. 1997. Derivational paradigms. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1996*, 243–256. Dordrecht: Kluwer.

Bauer, Laurie. 2019. Notions of paradigm and their value in word-formation. *Word Structure* 12(2). 153–175.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Becker, Thomas. 1993. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1993*, 1–25. Dordrecht: Kluwer.

Beniamine, Sacha. 2018. *Typologie quantitative des systèmes de classes flexionnelles*: Université Paris Diderot dissertation.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42. 531–573.

Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.

Bochner, Harry. 1993. *Simplicity in generative morphology*. Berlin: Mouton de Gruyter.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6(1). 213–234. 10.1146/annurev-linguistics-011619-030303. https://doi.org/10.1146/annurev-linguistics-011619-030303.

Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182.

Bonami, Olivier & Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e Linguaggio* 17(2). 173–195.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197. 10.1007/s11525-018-9322-6.

Bonami, Olivier & Juliette Thuilier. 2019. A statistical approach to affix rivalry: French *-iser* and *-ifier*. *Word Structure* 12(1). 4–41.

Booij, Geert. 1997. Autonomous morphology and paradigmatic relations. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1996*, 35–53. Dordrecht: Kluwer.

Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.

Bürkner, Paul-Christian. 2017. Brms: An R package for bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28. 10.18637/jss.v080.i01.

Bürkner, Paul-Christian. 2018. Advanced bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411. 10.32614/RJ-2018-017.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li & Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles* 76(1). 1–32. 10.18637/jss.v076.i01.

Clark, Eve V & Herbert H Clark. 1979. When nouns surface as verbs. *Language* 767–811.

Corbin, Danielle. 1976. Peut-on faire l'hypothèse d'une dérivation en morphologie? In Jean-Claude Chevalier (ed.), *Grammaire transformationnelle : syntaxe et lexique*, 47–91. Lille: Presses Universitaires de Lille.

Erk, Katrin. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass* 6(10). 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*, 110–114.

Fernández-Domínguez, Jesús, Alexandra Bagasheva & Cristina Lara Clares. 2020. *Paradigmatic relations in word formation*. Brill. https://brill.com/view/title/57860.

Firth, John R. 1957. Modes of meaning. In *Papers in linguistics, 1934-1951*, 190–215. Oxford: Oxford University Press.

Fradin, Bernard & Françoise Kerleroux. 2003. Troubles with lexemes. In S. Scalise & A. Ralli B. Geert, J. De Cesaris (ed.), *Selected papers from the third mediterranean morphology* *meeting*, 177–196. IULA-Universitat Pompeu Fabra (Barcelona). https://halshs.archives-ouvertes.fr/halshs-00122209.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari & Donald B Rubin. 2013. *Bayesian data analysis* Texts in Statistical Science. Boca Raton, Florida: Chapman & Hall/CRC third edition edn.

Gelman, Andrew, Jennifer Hill & Aki Vehtari. 2020. *Regression and other stories* Analytical Methods for Social Research. Cambridge: Cambridge University Press.

Guzmán Naranjo, Matías. 2020. Analogy, complexity and predictability in the Russian nominal inflection system. *Morphology* 30(3). 219–262.

Guzmán Naranjo, Matías & Olivier Bonami. 2021. Overabundance and inflectional classification: Quantitative evidence from czech. *Glossa: a journal of general linguistics* 6(1).

Harris, Zelig S. 1954. Distributional structure. *Word* 10. 146–162.

Harwood, F. W. & Alison M. Wright. 1956. Statistical study of English word formation. *Language* 32(2). 260–273.

Hathout, Nabil & Fiammetta Namer. 2014a. Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology* 11(5). 125–168.

Hathout, Nabil & Fiammetta Namer. 2014b. Discrepancy between form and meaning in word formation: the case of over- and under-marking in french. In Franz Rainer, Wolfgang U. Dressler, Francesco Gardani & Hans Christian Luschützky (eds.), *Morphology and meaning (selected papers from the 15th international morphology meeting, vienna, february 2012)* Current Issues in Linguistic Theory, 177–190. Amsterdam: John Benjamins.

Hathout, Nabil & Fiammetta Namer (eds.). 2018. *Lingue e linguaggio xvii:2: Defining paradigms in word formation*.

Hathout, Nabil & Fiammetta Namer (eds.). 2019. *Morphology 29:2: Paradigms in word formation: what are we up to?*

Hathout, Nabil, Franck Sajous & Basilio Calderone. 2014. GLÀFF, a large versatile French lexicon. In *Proceedings of lrec 2014*, .

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39(6). 1041–1070.

Huyghe, Richard, Alizée Lombard, Justine Salvadori & Sandra Schwab. 2023. Semantic rivalry between French deverbal neologisms in *-age*, *-ion* and *-ment*. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Huyghe, Richard & Marine Wauquier. 2020. What's in an agent? *Morphology* 30(3). 185–218. 10.1007/s11525-020-09366-2.

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

Jackendoff, Ray & Jenny Audring. 2020. *The texture of the lexicon*. Oxford: Oxford University Press.

Koehl, Aurore. 2012. *La construction morphologique des noms désadjectivaux suffixés en français*: Université de Lorraine dissertation.

Kotowski, Sven & Martin Schäfer. 2023. Semantic relatedness across base verbs and derivatives: Quantitative and distributional analyses of English *out*-prefixation. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Lapesa, Gabriella, Lea Kawaletz, Ingo Plag, Marios Andreou, Max Kisselew & Sebastian Padó. 2018. Disambiguation of newly derived nominalizations in context: A distributional semantics approach. *Word Structure* 11(3). 277–312.

Lazaridou, Angeliki, Marco Marelli, Roberto Zamparelli & Marco Baroni. 2013. Compositional-ly derived representations of morphologically complex words in distributional semantics. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)*, 1517–1526. Sofia, Bulgaria: Association for Computational Linguistics. https://aclanthology.org/P13-1149.

Lenci, Alessandro. 2008. Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica* 20. 1–31.

Lieber, Rochelle. 2023. Ghost aspect and double plurality. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Marchand, Hans. 1963. On a question of contrary analysis with derivationally connected but morphologically uncharacterized words. *English Studies* 44. 176–187.

Marelli, Marco & Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122. 485–515.

van Marle, Jaap. 1984. *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Mickus, Timothee, Olivier Bonami & 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 & Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *CoRR* abs/1301.3781. http://arxiv.org/abs/1301.3781.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado & Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Miller, George A & Walter G Charles. 1991. Contextual correlates of semantic similarity. *Language and cognitive processes* 6(1). 1–28.

Namer, Fiammetta. 2012. Nominalisation et composition en français: d'où viennent les verbes composés? *Lexique* 20. 173–205.

Namer, Fiammetta, Lucie Barque, Olivier Bonami, Pauline Haas, Nabil Hathout & Delphine Tribout. 2019. Demonette2 — Une base de données dérivationnelles du français à grande échelle : premiers résultats. In *Actes de taln*, Toulouse, France. https://halshs.archives-ouvertes.fr/halshs-02275652/document.

Namer, Fiammetta & Nabil Hathout. 2020. ParaDis and Démonette – from theory to resources for derivational paradigms. *The Prague Bulletin of Mathematical Linguistics* 114. 5–33.

Padó, Sebastian, Aurélie Herbelot, Max Kisselew & 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*, 1285–1296.

Pennington, Jeffrey, Richard Socher & Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)*, 1532–1543.

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

Plag, Ingo, Lea Kawaletz, Sabine Arndt-Lappe & Rochelle Lieber. 2023. Analogical modeling of derivational semantics. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Rainer, Franz. 2018. Word formation and word history: The case of CAPITALIST and CAPITALISM. In Olivier Bonami, Gilles Boyé, Georgette Dal, Hélène Giraudo & Fiammetta Namer (eds.), *The lexeme in descriptive and theoretical morphology*, 43–65. Berlin: Language Science Press.

Řehůřek, Radim. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the lrec 2010 workshop on new challenges for nlp frameworks*, 45–50.

Robins, R. H. 1959. In defense of WP. *Transactions of the Philological Society* 58. 116–144.

Roché, Michel. 2011a. Quel traitement unifié pour les dérivations en *-isme* et en *-iste*? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (eds.), *Des unités morphologiques au lexique*, 69–143. Hermès Lavoisier.

Roché, Michel. 2011b. Quelle morphologie? In Michel Roché, Gilles Boyé, Nabil Hathout, Stéphanie Lignon & Marc Plénat (eds.), *Des unités morphologiques au lexique*, 14–39. Hermès Lavoisier.

Schäfer, Martin. 2023. Splitting *-ly*'s: using word embeddings to distinguish derivation and inflection. In Sven Kotowski & Ingo Plag (eds.), *The semantics of derivational morphology: Theory, methods, evidence*, Berlin: De Gruyter.

Schäfer, Roland. 2016. Commoncow: Massively huge web corpora from commoncrawl data and a method to distribute them freely under restrictive eu copyright laws. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on language resources and evaluation (lrec 2016)*, 4500–4504. Portorož, Slovenia: European Language Resources Association (ELRA).

Schäfer, Roland & 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*, 486–493.

Strnadová, Jana. 2014a. *Les réseaux adjectivaux: Sur la grammaire des adjectifs dénominaux en français*: Université Paris Diderot et Univerzita Karlova V Praze dissertation.

Strnadová, Jana. 2014b. Multiple derivation in French denominal adjectives. In Sandra Augendre, Graziella Couasnon-Torlois, Déborah Lebon, Clément Michard, Gilles Boyé & Fabio Montermini (eds.), *Proceedings of the 8th décembrettes* (Carnets de grammaire 22), 327–346. Toulouse: CLLE-ERSS.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure* (Cambridge Studies in Linguistics 93). Cambridge: Cambridge University Press.

Stump, Gregory T. 2016. *Inflectional paradigms*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2019. Some sources of apparent gaps in derivational paradigms. *Morphology* 29. 271–292.

Tribout, Delphine. 2010a. How many conversions from verb to noun are there in French? In Stefan Müller (ed.), *Proceedings of the 17th international conference on head-driven phrase structure grammar*, 341–357. Stanford, CA: CSLI Publications.

Tribout, Delphine. 2010b. *Les conversions de nom à verbe et de verbe à nom en français*: Université Paris Diderot dissertation.

Tribout, Delphine. 2020. Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French. *Zeitschrift für Wortbildung/Journal of Word Formation* 4(2). 187–207.

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

Turney, Peter D & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37. 141–188.

Varvara, Rossella, Gabriella Lapesa & Sebastian Padó. 2021. Grounding semantic transparency in context. *Morphology* 1–38.

Štekauer, Pavol. 2014. Derivational paradigms. In Rochelle Lieber & Pavol Štekauer (eds.), *The oxford handbook of derivational morphology*, 354–369. Oxford: Oxford University Press.

Wauquier, Marine, Nabil Hathout & Cécile Fabre. 2020. Semantic discrimination of technicality in french nominalizations. *Zeitschrift für Wortbildung/Journal of Word Formation* 4(2). 100–119.