Annotating complex words to investigate the semantics of derivational processes

Rossella Varvara, Justine Salvadori, Richard Huyghe
University of Fribourg
{firstname.lastname}@unifr.ch

Abstract

In this paper, we present and test an annotation scheme designed to analyse the semantic properties of derived nouns in context. Aiming at a general semantic comparison of morphological processes, we use a descriptive model that seeks to capture semantic regularities among lexemes and affixes, rather than match occurrences to word sense inventories. We annotate two distinct features of target words: the ontological type of the entity they denote and their semantic relationship with the word they derive from. As illustrated through an annotation experiment on French corpus data, this procedure allows us to highlight semantic differences and similarities between affixes by investigating the number and frequency of their semantic functions, as well as the relation between affix polyfunctionality and lexical ambiguity.

Keywords: derivation, affix polyfunctionality, lexical ambiguity, semantic annotation, nominalization

1. Introduction

Derived words (e.g. *arrival*, *impossible*, *exemplify*) constitute a large part of our mental lexicon. Their morphosemantic properties have attracted a growing attention in the last decades, with the development of studies investigating the relationship between form and meaning in derivation (Zwanenburg, 2000; Lieber, 2004; Bauer et al., 2015; Schulte, 2015; Lieber, 2019, a.o.). It has been observed that word-formation processes are often polyfunctional, i.e. each of them may serve a variety of semantic functions by producing derivatives with different types of meaning (Moortgat and van der Hulst, 1981; Lehrer, 2003; Plag et al., 2018; Prćić, 2019, a.o.). For example, the suffix *-er* in English can be used to derive nouns that denote agents (*writer*), experiencers (*hearer*), stimuli (*pleaser*), instruments (*amplifier*), patients (*scratcher*), locations (*smoker*), measures (*fiver*), and inhabitants (*New Yorker*) (Lieber, 2016). In addition, affix polyfunctionality can sometimes lead to the formation of ambiguous words, as in the case of nouns suffixed with *-er* that are ambiguous between an agent and an instrument reading, such as *player* in (1) and (2).

(1) One of the players is arguing with the referee. [AGENT]

(2) The player is damaged and needs to be repaired. [INSTRUMENT]

Previous works have focused on the identification of semantic functions of suffixes, without taking into account their realisation frequency among derived lemmas and tokens. Similarly, as far as lexical ambiguity is concerned, the focus has been put on the number of senses words have, rather than on the frequency of these senses.

In this study, we investigate the polyfunctionality of derivational processes, taking into account not only the number of functions they present, but also the distribution of these functions among tokens, as well as the ambiguity of the words they form. To do so, we annotate 4,500 corpus occurrences of 90 deverbal nouns ending with 6 different suffixes in French. Each occurrence is assigned two labels to account for the meaning of the noun: an ontological type corresponding to the description of the referent, and a relational type corresponding to the semantic relationship with the morphological base. While this annotation scheme allows us to investigate meaning regularities among words and suffixes, the corpus annotation brings three levels of observation for both affix polyfunctionality and derivative ambiguity: (i) the number of meanings observed; (ii) their realisation frequency; (iii) their distribution among items. The reliability of the method is evaluated by considering agreement scores between annotators.

The issue of the disambiguation of derived nouns and the modelling of their semantics is relevant in various areas of computational linguistics and natural language processing. Even though deverbal nouns constitute difficult cases for automatic word sense disambiguation, their processing is necessary for the identification of non-verbal events (Kolya et al., 2013; Zavarella and Tanev, 2013). The modelling of the semantics of derivational processes may also help in copying with the issue of data sparsity. Being able to model derivational semantics can allow to automatically infer the semantics of derived nouns not attested in corpora, for which representations in lexical resources or from word embeddings are not available.

In what follows, after a brief summary of related works (section 2), we introduce the annotation scheme used in the study (section 3), together with a description of the annotated sample, an evaluation of the reliability of the annotation, and a discussion of cases of inter-

annotator disagreement (section 4). In section 5, we present some observations about the polyfunctionality of French suffixes and the ambiguity of deverbal nouns based on the sample we annotated.

2. Related work

Theoretical works on the semantics of derived nouns have frequently aimed at an exhaustive listing of their different possible meanings. Some studies made use of the analysis of corpus data, disambiguating specific instances of derived nouns. However, the annotation of senses of derived nouns in context has proved to be a difficult task even for human annotators. One of the main issue of disagreement is the annotation of ambiguous and semantically complex nouns (see Section 3), which are particularly frequent among deverbal nominalizations.

Martínez Alonso et al. (2013) conducted an annotation task on nouns presenting regular polysemy (not restricted to derived nouns) in three different languages. They found that the agreement score varied across the different polysemous types. While only one obtained a reliable value ($\alpha > 0.6$), i.e. the Animal-Meat type, other polysemous types had low scores of agreement, showing that not all cases of ambiguity can be equally identified.

Valdivia et al. (2013), in an annotation task on a sample of 323 Russian deverbal nouns extracted from the Russian National Corpus (RNC), obtained an average agreement of 0.296 (Fleiss' kappa). Four labels were assigned: 'event', 'state', 'result', and 'underspecified'.

Peris et al. (2010) report the annotation agreement obtained in a task where 5 non-expert annotators had to classify 300 Spanish deverbal nominalizations. Three categories were used: 'event', 'result', and 'underspecified'. After a training on one third of the sample, the annotators reached an agreement of 0.6 (Cohen's kappa).

Barque et al. (2020) annotated 5554 corpus tokens of French simple nouns using 23 WordNet Unique Beginners as semantic tags, reporting an agreement coefficient of 0.649 (Cohen's kappa). In their scheme, the annotation of complex types was allowed and proved to influence inter-annotator agreement. Indeed, considering partial agreement among complex types, the kappa score increases to 0.734.

Our annotation experiment inherits the difficulties of this kind of task, for which an agreement between 0.6-0.7 is already a good result.

3. Annotation scheme

In order to analyse French deverbal nouns in the context of real corpus sentences, we designed a scheme of semantic annotation that can provide reliable information about the semantics of these words. We used a classification of deverbal nouns that fundamentally distinguishes between ontological and relational types, and combines them into a unique class that is identified as their 'complete' semantic type.

The ontological type expresses the nature of the referent denoted by the noun and the category of entities that it represents linguistically. The relational type refers to the semantic relation between the derived noun and its morphological base. While many classifications of deverbal nouns proposed in the literature assimilate ontological and relational types in a single taxonomy, possibly creating conflicting identification issues, it appears that the two dimensions are not strictly interdependent. As can be seen in (3), the same ontological type (ARTEFACT) combines with different relational types, while in (4), the same relational type (RESULT) combines with different ontological types.

(3) a. bâtir 'build' → bâtiment 'building' [ARTEFACT-RESULT]
    b. raser 'shave' → rasoir 'razor' [ARTEFACT-INSTRUMENT]
    c. garer 'park' → garage 'garage' [ARTEFACT-LOCATION]

(4) a. bâtir 'build' → bâtiment 'building' [ARTEFACT-RESULT]
    b. énerver 'irritate' → énervement 'irritation' [STATE-RESULT]
    c. créer 'create' → créature 'creature' [ANIMATE-RESULT]

The ontological classification we used includes 14 simple types and 7 complex types, listed in Table 1. It is based on distributional tests taken from the literature on nominal semantics (Godard and Jayez, 1993; Flaux and Van de Velde, 2000; Huyghe, 2015, a.o.) and described in more detail in Salvadori et al. (2021). For example, a noun that can be used as the subject of the verb *avoir lieu* 'take place', such as *licenciement* 'dismissal' in (5), is classified as EVENT.

(5) Le licenciement de l'employé a eu lieu ce matin.
'The dismissal of the employee took place this morning'

Complex types (also known as *dot types*) characterise nouns that are semantically hybrid. They can refer simultaneously to different ontological aspects of entities and are contextually compatible with predicates that are distinctive of different ontological types (Copestake and Briscoe, 1995; Cruse, 1995; Pustejovsky, 1995; Kleiber, 1999; Asher, 2011; Dölling, 2020; Murphy, 2021). Such is the case of the noun *déclaration* 'statement' in (6), which denotes at the same time the act of making a statement and the content of this statement (i.e. an event and a cognitive object in our ontology, respectively).

As can be seen from the example, the co-existence of different facets of meaning is made explicit by the co-predication: the eventive facet is selected by *effectuer*

Simple types	
ANIMATE	INSTITUTION
ARTEFACT	NATURAL
COGNITIVE	PHENOMENON
DISEASE	PROPERTY
DOMAIN	QUANTITY
EVENT	STATE
FINANCIAL	TIME
Complex types	
ARTEFACT*COGNITIVE	EVENT*NATURAL
ARTEFACT*INSTITUTION	EVENT*PHENOMENON
COGNITIVE*EVENT	EVENT*STATE
EVENT*FINANCIAL	

Table 1: Ontological types

'perform' and the cognitive facet by *selon lequel P* 'according to which P'.

(6) L'hôpital Legouest de Metz a effectué une déclaration selon laquelle il venait d'accueillir deux victimes blessées par balles. (web)
'The Legouest Hospital in Metz made a statement according to which they had just received two victims with gunshot wounds'

In order to account for collective nouns, i.e. nouns that have a plural reference when used in the singular (Flaux, 1999; Lammert, 2006; De Vries, 2019, a.o.), a COLLECTIVE label can be added to both simple and complex types. For instance, *assistance* 'audience' is assigned the label ANIMATE-COLLECTIVE, as it denotes a group of several people. The noun *déménageur* 'mover', by contrast, is annotated only as ANIMATE because it denotes a single animate entity.

Relational types are expressed by labels that correspond to semantic roles, considering that they are semantically equivalent to the roles that derived nouns fulfil with respect to base verb predicates. For example, the relational type of the noun *arrosoir* 'watering can' is INSTRUMENT because *arrosoir* denotes an entity that is used to perform the action denoted by the base verb *arroser* 'water'. We defined a set of 17 semantic roles adapted from VerbNet (Kipper-Schuler, 2005) and LIRICS (Petukhova and Bunt, 2008) (see Table 2 for the list). Since distributional tests cannot be used to differentiate semantic roles, we relied on explicit definitions to identify relational types, as described in Salvadori et al. (2021).

We supplemented the list of traditional semantic roles with a type called TRANSPOSITION for nouns whose role is actually that of transposing the verbal meaning into a noun. The noun *atterrissage* 'landing', for instance, can be considered a TRANSPOSITION, as it roughly denotes the same eventuality as its base verb *atterrir* 'land'. A FIGURATIVE label was also added in order to take into account metaphorical or metonymic meanings that are derived from a given role, but that are not semantically related to the base verb. It is attached directly to the role in question (e.g. *lacer* 'lace up' $\rightarrow$ *lacet*$_1$ 'shoelace' [INSTRUMENT] $\rightarrow$ *lacet*$_2$ 'zigzag' [INSTRUMENT-FIGURATIVE]).

AGENT	INSTRUMENT	RESULT
BENEFICIARY	LOCATION	SOURCE
CAUSE	MANNER	STIMULUS
DESTINATION	PATH	THEME
EXPERIENCER	PATIENT	TOPIC
EXTENT	PIVOT	

Table 2: Relational types

It is important to note that, even if the labels used are the same as those of semantic roles, the annotation of relational types does not concern the semantic role of the derived noun in the sentence of occurrence. Relational types aim at capturing lexical semantic information that is not context-dependent. While a noun like *inspecteur* 'inspector' can be defined lexically as AGENT with respect to the base verb *inspecter* 'inspect', it can be assigned a variety of semantic roles that are not necessarily those of agent in discourse. For instance, *inspecteur* 'inspector' features as the agent in (7), but as the patient in (8) and as the beneficiary in (9). Whatever the semantic role is in context, the derivational relation with the base verb is stable and encoded as such by labels of relational types.

(7) L'inspecteur a interrogé les témoins.
'The inspector interviewed the witnesses'

(8) Le détenu a frappé l'inspecteur.
'The prisoner hit the inspector'

(9) Le commissaire a offert des fleurs à l'inspecteur.
'The superintendent offered flowers to the inspector'

It remains true that polysemous nouns can be lexically associated with different relational types, and therefore be assigned different relational types across sentences of occurrence.

4. Annotation experiment

In this study, we investigate the polyfunctionality of 6 French deverbal suffixes (*-oir*, *-ure*, *-is*, *-ment*, *-aire*, *-ade*) and the ambiguity of their derivatives. We analysed 15 nouns per suffix, based on the semantic annotation of 50 of their occurrences in the French web corpus FRCOW16A (Schäfer and Bildhauer, 2012; Schäfer, 2015). The nouns were randomly selected across different token frequency ranges, based on the frequency distribution of all the nouns derived with the 6 suffixes. We excluded lemmas with a frequency lower than 50, since we needed a sufficient amount of occurrences to investigate lexical ambiguity. Then, for each suffix, we selected 15 lemmas from three frequency ranges: 5

nouns with a token frequency up to the general median value (50-223), 5 nouns with a token frequency from the median value to the third quartile (224-3,799), and 5 nouns with a token frequency higher than the third quartile (3,800-3,966,941). For each lemma, we randomly selected 50 occurrences from the corpus and annotated the sense of the targeted derived noun in context for both ontological type and relational type. We did not consider tokens in which the meaning of the noun was not related to the meaning of the base verb, neither directly nor indirectly (i.e. by means of lexical figure). In total, our sample comprehends 4,500 tokens and 90 different lemmas.

In order to validate our annotation procedure and to assess the reliability of the annotation performed, two samples of the dataset were simultanously annotated by two authors of this paper and adjudicated with the help of the third author. Each sample contained 300 randomly selected tokens for a total of 18 lemmas. Annotators followed the guidelines previously defined to annotate the semantic types of the target words in the context of each sentence. The agreement among the two annotators was computed using Cohen's kappa coefficient[1].

In the first sample, the agreement for the relational type was already almost perfect (κ = 0.848), whereas it was only moderate for the ontological types (κ = 0.539). An analysis of cases of disagreement showed that the main difficulty encountered was the contextual analysis of nouns with complex types. The annotators partially disagreed on the annotation of 67 tokens involving a complex type, i.e. one annotator identified a complex type whereas the other identified a simple or another complex type partially overlapping the first one. For instance, the word *miaulement* ('meowing') in example (10) has been annotated by one annotator as PHENOMENON and by the other as the complex type EVENT*PHENOMENON, since it refers at the same time at something that can be heard and that takes place. In this case, we consider that the annotators only partially disagreed because at least the PHENOMENON facet was identified by both of them.

(10) Les petites chevêches d'Athéna font de jolis miaulements dans la journée mais c'est rare de les entendre.
'Athena's little owls make lovely meowing sounds during the day but it is rare to hear them'

When considering these cases of partial agreement as mere agreement, the kappa coefficient for ontological types increases from 0.539 to 0.786. More generally, it appears that many nouns with complex types are characterised by one dominant facet of meaning (e.g. COGNITIVE as opposed to ARTEFACT for the noun *argumentaire* 'argument document'). In the absence of a distinctive predicate specifically associated with a given facet of a complex type, the two annotators used different strategies, by considering the dominant meaning or the complex meaning as the default type.

We therefore refined the instructions for the contextual annotation of complex types. It was decided that complex types would be the default type when no clear contextual elements drive the interpretation in favour of a specific facet. In other words, in the presence of predicates that are underspecified with respect to facet selection, complex semantic types are annotated as such. This indication proved to be efficient since in the second sample, the agreement score for ontological types increased significantly (κ = 0.815). Indeed, the difference with the agreement score that also included partial agreement was almost null (κ = 0.877). The agreement score for relational types was assessed as substantial (κ = 0.717), albeit lower than that of the first annotation, which is probably due to intrinsic sample differences. After that, the rest of the sample was single-annotated by the three authors of the paper, but problematic cases were discussed in joint annotation sessions.

5. Annotation results

Twenty-three ontological types, 21 relational types, and 62 complete types are included in the final annotated sample. The distinction between ontological and relational categories proves to capture different aspects of the semantics of derived nouns. As can be seen in Fig. 1, the degree of polyfunctionality of the 6 suffixes investigated (i.e. the number of types realised) varies between ontological and relational types. A suffix like *-ment*, which realises a high number of ontological types, can present a low number of relational types. Conversely, a suffix like *-oir*, which realises a low number of ontological types, can present a high number of relational types. Considering the opposite behaviours of *-oir* and *-ment*, it seems that the former has a more cohesive referential meaning, whereas the latter has a more stable derivational relation with its base verbs. The number of complete types can leverage this difference, possibly showing a higher number of functions for suffixes with more ontological or more relational types. It can be noted that the relation between ontological and relational types is not uniform. In particular, there is no general inverse correlation between the number of ontological types and the number of relational types. This relation seems to be specific to each suffix.

The number of types realised per suffix can be regarded as a first measure of polyfunctionality: the higher the number of semantic functions, the more polyfunctional a suffix is. However, this measure does not provide a complete picture of the versatility of suffixes, because the realisation frequency of functions needs to be taken into account as well. It is theoretically possible that two suffixes, although they are associated with

[1]The computation was performed using the kappa2 function of the 'irr' package in R (Gamer et al., 2007; R Core Team, 2013).

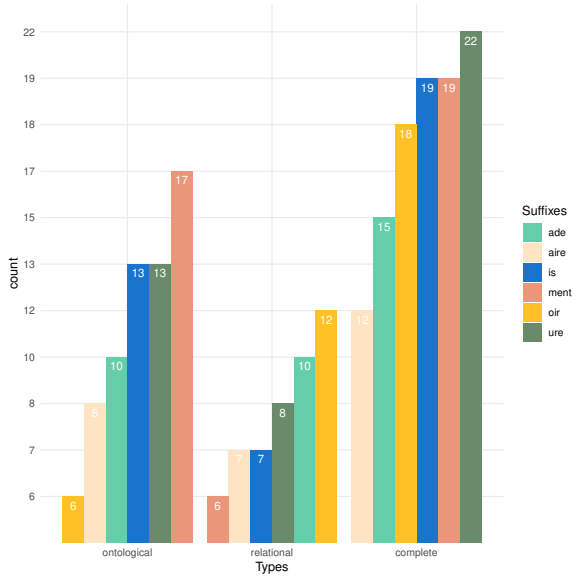

Figure 1: Number of function types realised by each suffix

a different number of semantic functions, can both have only one function that is represented in mostly all of the occurrences of the suffix. In addition to Fig. 1, Fig. 2 shows the proportion of ontological types realised among the corpus tokens of nouns ending with each suffix.[2] The plot illustrates what can be observed through token annotation.

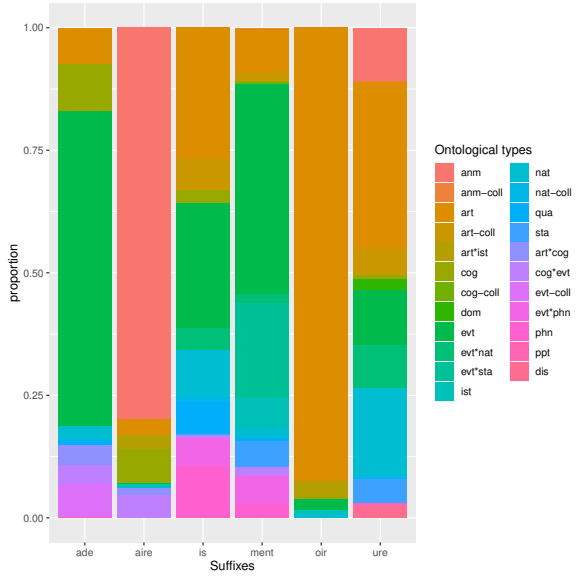

Figure 2: Token frequency of ontological types per suffix

[2]We do not include the plots for relational and complete types due to space limitation and to the large number of functions involved.

The suffix *-oir* realises the same ontological type (ARTEFACT) in most of the occurrences of its derivatives, with the other types covering less than 8% of its total occurrences. Similarly, *-aire* mainly realises animate nouns (e.g. *consignataire* 'consignee'), which constitutes 80% of its tokens. The suffix *-ade* also shows a clear preference for one given ontological type, but in a smaller proportion than *-oir* and *-aire*, with 64% of the occurrences of its derivatives referring to an event. For the last three suffixes, the distribution of ontological types is more diverse, with different types covering larger portions of occurrences. These observations about the distribution of functions go in the same direction as those about the number of ontological types attested: *-oir*, *-aire*, and *-ade* have a lower number of types and a less diverse distribution among tokens.

In a nutshell, the observations for suffix polyfunctionality based on our annotated sample of deverbal nouns relate to two different aspects: the number of types attested per suffix, and the number of tokens each type covers among the occurrences of each suffix. Borrowing terms from the ecological domain, we can think of these two observations as type abundance and relative abundance, i.e. richness and evenness of distribution of functions across suffix tokens, respectively. These properties are two sides of the same coin and should be considered together to fully assess affix polyfunctionality. A diversity index[3] such as the Hill-Simpson index (Hill, 1973; Jost, 2006; Roswell et al., 2021) can be used to consider these two properties together and determine which suffixes are more or less semantically diverse, and by how much. This diversity index not only takes into account the richness of the semantic functions of a suffix, but also the evenness of the token realisation of functions. More precisely, the Hill-Simpson index is equivalent to the inverse of the traditional Simpson index (Simpson, 1949) and is calculated from equation (1), where p_i corresponds to the number of tokens realizing a function *i* divided by the total number of tokens per suffix.

$$\frac{1}{\sum_{n=1}^{S}(p_i)^2} \quad (1)$$

In Table 3, we report the Hill-Simpson values[4] for the 6 suffixes we investigated, computed for ontological types, relational types, and the combination of the two (i.e. complete types).

Considering only ontological types, the Hill-Simpson

[3]Diversity indexes are widely used in ecology to measure the diversity of species. A diversity index like Shannon's entropy has been used in computational linguistics to measure semantic content. See for example Santus et al. (2014) and Padó et al. (2015).

[4]Hill-Simpson indexes are computed using the MeanRarity package (Roswell and Dushoff, 2020) in R. We experimented also with Shannon's entropy and Hill-Shannon index and we found the same ranking of suffixes.

	Ontological	Relational	Complete
-ade	2.29	1.70	2.32
-aire	1.54	3.19	3.54
-is	5.85	2.25	7.08
-ment	4.16	1.80	4.21
-oir	1.16	3.90	3.96
-ure	5.37	2.71	6.64

Table 3: Hill-Simpson diversity index per suffix

index confirms what we already observed from previous analyses (Fig. 1 and Fig. 2): *-oir* and *-aire* have the lowest coefficients of diversity, due to their low number of ontological types and the low evenness of the distribution of their types among tokens. The suffix *-ade* has higher diversity than the previous two, but still lower diversity than *-ment*, *-is* and *-ure*. As for *-ment*, it presents the highest number of different ontological types, but the Hill-Simpson index only ranks it as the third most diverse suffix, given that the distribution of ontological types among its tokens is less even than those of *-is* and *-ure*.

We have discussed the results for ontological types as an illustration of our method, since it would have been more complex to examine a plot with the 60 complete types. However, we should consider complete types to fully represent the polyfunctionality of the suffixes. Looking at the Hill-Simpson index for complete types in Table 3, we can observe that *-ade* is the least diverse of the 6 suffixes. In 64% of the observed tokens, nouns ending in *-ade* refer to an event, whose relational type is always TRANSPOSITION. Despite the high number of complete types of the latter, *-aire* and *-oir* appear to be less diverse than *-is*, *-ment* and *-ure*, even if the difference with *-ment* is weaker. The most diverse suffix appears to be *-is*. Although *-ure* presents the highest number of complete types (fig. 1), it is distributed in less even ways among tokens, and finally evaluated as less diverse, than *-is*.

So far, we have focused on suffix semantics, investigating the semantic functions realised among corpus tokens of nouns ending with different suffixes. We have not considered yet the distribution of semantic functions among lemmas, nor the relationship between the ambiguity of derivatives and the polyfunctionality of suffixes. It can be asked how the different functions of suffixes are distributed among the nouns they derive, and whether more polyfunctional suffixes would derive more ambiguous words. Similarly to our observations for suffixes, we can infer from our annotated data different information for lemmas: the number of functions per lemma, the distribution (and degree of evenness) of these functions among tokens of lemmas, and a Hill-Simpson diversity measure combining these two for each lemma. We report these values averaged across lemmas for each suffix, in order to compare them with the values obtained above.

We first consider the number of semantic types realised by each lemma. In Table 4, we report the average number of different semantic types (considering as semantic types ontological, relational, and complete types) realised by the lemmas derived with each suffix, together with the standard deviation (SD) of these values. It appears that the suffix with the lowest number of complete functions (*-aire*) is also the suffix that produces the less ambiguous nouns (on average 1.53 complete type per lemma). However, this is the only suffix for which there is a direct correspondence between the two rankings. For the other suffixes, the degree of averaged lemma ambiguity is not parallel to the number of types. For example, *-ment* derives more ambiguous nouns than *-is* and *-ure*, although *-is* and *-ure* are associated with an identical and a higher number of complete types than *-ment*, respectively.

	Ontological		Relational		Complete	
	Avg	SD	Avg	SD	Avg	SD
-ade	2.07	1.28	1.80	1.01	2.27	1.49
-aire	1.53	1.19	1.20	0.41	1.53	1.19
-is	2.60	1.64	1.80	0.78	2.73	1.83
-ment	3.40	1.80	1.93	0.46	3.40	1.80
-oir	1.80	1.01	1.67	0.72	2.13	1.13
-ure	3.00	1.60	1.87	0.92	3.13	1.88

Table 4: Ambiguity of lemmas averaged per suffix

In addition to information about the ambiguity of nouns ending with the different suffixes, we should take into consideration the distribution of the different senses observed among these nouns. It can be argued that two words with two senses are not equally ambiguous if one of them realises only one sense among the majority of its tokens, whereas the other shows an even distribution of the two senses among its tokens. As before, we make use of the Hill-Simpson diversity index to consider together type frequency and token frequency of functions among derived words. More precisely, we computed for each lemma the Hill-Simpson diversity index, which tells us how ambiguous a lemma is and how evenly its senses are distributed among tokens. Then we averaged these values across the lemmas ending with each suffix. Results for ontological, relational and complete semantic types are reported in Table 5. We can observe that the ranking obtained is identical to that of the simple average ambiguity (column 6 in Tab. 4): higher ambiguity rates for lemmas per suffix correspond to higher diversity indexes.

We can conclude that the definition and the measurement of lexical ambiguity seem rather straightforward: the degree of ambiguity of derivatives remains the same if we consider the simple count of functions or if we combine it with the evenness of their distribution in the Hill-Simpson measure. The same is not true if we consider suffix polyfunctionality, for which we observed slightly different rankings for the different measures.

	Ontological		Relational		Complete	
	Avg	SD	Avg	SD	Avg	SD
-ade	1.48	0.88	1.31	0.57	1.50	0.91
-aire	1.27	0.61	1.17	0.36	1.27	0.61
-is	1.57	0.83	1.25	0.38	1.60	0.85
-ment	2.07	1.02	1.35	0.37	2.07	1.02
-oir	1.17	0.38	1.31	0.37	1.35	0.41
-ure	1.70	0.69	1.41	0.43	1.71	0.70

Table 5: Hill-Simpson diversity index for lemmas averaged per suffix

Therefore, it appears that suffix polyfunctionality and derivative ambiguity are not clearly correlated, since according to our analyses, each suffix presents its own characteristics for these two features.

6. Conclusion

In this paper, we have presented an annotation experiment on corpus data to investigate the semantics of nouns derived from verbs in French. We focused on 6 nominalizing suffixes, instantiated by 90 nouns and 4,500 corpus occurrences. The annotation represents word senses in a way that allows generalisations on meanings over suffixes and derivatives. It distinguishes semantically the ontological description of the referent and the relation between bases and derivatives. The results from our annotation experiment show that the relation between ontological and relational types is not uniform, but specific to each suffix.

The corpus annotation provides data for considering not only the number of semantic functions associated with suffixes, but also the distribution of these functions among tokens. We combined together these two levels of analysis and computed an index of diversity that can highlight various degrees of polyfunctionality and diversity across nominalizing suffixes. The ranking of suffixes obtained through this index is different from that obtained by simply counting the number of functions, confirming that it is important to take into account the realisation frequency of semantic types in the study of polyfunctionality. A similar evaluation can be conducted at the word level, assessing lexical ambiguity not just as the number of senses displayed, but also considering the evenness of their distribution among tokens.

The sample of nouns annotated in this study can be used as a gold standard for the evaluation of automatic word-sense disambiguation systems. Specifically, given the annotation at the token level, it can be used in the evaluation of contextualized embedding models (Peters et al., 2018; Devlin et al., 2018, a.o.), which seem promising in capturing word ambiguity. It can be hypothesised that (i) the similarity between pairs of token embeddings will be higher if the tokens share the same semantic functions, (ii) less polyfunctional suffixes will have a higher pairwise similarity among their tokens. These hypotheses will be the object of a further study.

7. Acknowledgements

This research was supported by the Swiss National Science Foundation under grant 10001F_188782 ('The semantics of deverbal nouns in French'). We thank the three anonymous reviewers for their helpful comments.

8. Bibliographical References

Asher, N. (2011). *Lexical meaning in context: A web of words*. Cambridge University Press, Cambridge.

Barque, L., Haas, P., Huyghe, R., Tribout, D., Candito, M., Crabbé, B., and Segonne, V. (2020). Frsemcor: Annotating a French corpus with supersenses. In *LREC-2020*.

Bauer, L., Körtvélyessy, L., and Štekauer, P. (2015). *Semantics of complex words*, volume 3. Springer.

Copestake, A. and Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, 12(1):15–67.

Cruse, D. A. (1995). Polysemy and related phenomena from a cognitive linguistic viewpoint. In P. St Dizier et al., editors, *Computational lexical semantics*, pages 33–49. Cambridge University Press.

De Vries, H. (2019). Collective nouns. *Oxford Handbook of Grammatical Number*.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dölling, J. (2020). Systematic polysemy. In Daniel Gutzmann, et al., editors, *The Wiley Blackwell companion to semantics*. John Wiley & Sons, Inc., Hoboken NJ.

Flaux, N. and Van de Velde, D. (2000). *Les noms en français : esquisse de classement*. Editions Ophrys.

Flaux, N. (1999). A propos des noms collectifs. *Revue de linguistique romane*, 63(251):471–502.

Gamer, M., Lemon, J., Fellows, I., and Gamer, M. M. (2007). The irr package.

Godard, D. and Jayez, J. (1993). Types nominaux et anaphores : le cas des objets et des événements. In Walter De Mulder, et al., editors, *Anaphores temporelles et (in-)coherence, Cahiers Chronos*, volume 1, pages 41–58. Rodopi.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2):427–432.

Huyghe, R. (2015). Les typologies nominales: présentation. *Langue française*, (185):5–27.

Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2):363–375.

Kipper-Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

Kleiber, G. (1999). *Problèmes de sémantique: La polysémie en question(s)*. Presses Universitaires du Septentrion, Villeneuve d'Ascq.

Kolya, A. K., Kundu, A., Gupta, R., Ekbal, A., and Bandyopadhyay, S. (2013). Ju_cse: A crf based approach to annotation of temporal expression, event and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 64–72.

Lammert, M. (2006). *Sémantique et cognition: les noms collectifs*. Ph.D. thesis, Strasbourg 2.

Lehrer, A. (2003). Polysemy in derivational affixes. In Brigitte Nerlich, et al., editors, *Polysemy: Flexible Patterns of Meaning in Mind and Language*, pages 217–232. De Gruyter Mouton.

Lieber, R. (2004). *Morphology and lexical semantics*, volume 104. Cambridge University Press.

Lieber, R. (2016). *English nouns: The ecology of nominalization*, volume 150. Cambridge University Press.

Lieber, R. (2019). 3. Semantics of derivational morphology. In Claudia Maienborn, et al., editors, *Semantics - Interfaces*, pages 75–102. De Gruyter Mouton.

Martínez Alonso, H., Sandford Pedersen, B., and Bel, N. (2013). Annotation of regular polysemy and underspecification. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 725–730, Sofia, Bulgaria, August. Association for Computational Linguistics.

Moortgat, M. and van der Hulst, H. (1981). Geïnterpreteerde morfologie. In Rob Knopper, editor, *Woordstructuur*, pages 17–53. Foris.

Murphy, E. (2021). *Linguistic representation and processing of copredication*. Ph.D. thesis, University College London.

Padó, S., Palmer, A., Kisselew, M., and Šnajder, J. (2015). Measuring semantic content to assess asymmetry in derivation. In *Proceedings of the Workshop on Advances in Distributional Semantics, London, UK*, volume 14. Citeseer.

Peris, A., Taulé, M., and Rodríguez Hontoria, H. (2010). Semantic annotation of deverbal nominalizations in the Spanish corpus ancora. In *Proceedings of the ninth International workshop on treebanks and linguistic theories: December 3-4, 2010, University of Tartu, Estonia*, pages 187–198. North European Association for Language Technology.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations.

Petukhova, V. and Bunt, H. (2008). LIRICS semantic role annotation: Design and evaluation of a set of data categories. In Nicoletta Calzolari, et al., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 39–45, Marrakech, Morocco. European Language Resources Association (ELRA).

Plag, I., Andreou, M., and Kawaletz, L. (2018). A frame-semantic approach to polysemy in affixation. *The lexeme in descriptive and theoretical morphology*, 4:467.

Prćić, T. (2019). Exploring the properties of English lexical affixes by exploiting the resources of English general-purpose dictionaries. *Lexikos*, 29:151–179.

Pustejovsky, J. (1995). *The generative lexicon*. MIT Press, Cambridge, MA.

R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roswell, M. and Dushoff, J., (2020). *MeanRarity: Hill diversity estimation and visualisation*.

Roswell, M., Dushoff, J., and Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, 130(3):321–338, March.

Salvadori, J., Barque, L., Pauline, H., Huyghe, R., Lombard, A., Monney, M., Schwab, S., Tribout, D., Varvara, R., and Wauquier, M., (2021). *The semantics of deverbal nouns in French: Annotation guide*. University of Fribourg, Switzerland.

Santus, E., Lenci, A., Lu, Q., and Im Walde, S. S. (2014). Chasing hypernyms in vector spaces with entropy. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 38–42.

Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*, Mannheim. IDS.

Schulte, M. (2015). *The semantics of derivational morphology: A synchronic and diachronic investigation of the suffixes* -age *and* -ery *in English*. Narr, Tübingen.

Schäfer, R. and Bildhauer, F. (2012). Building large corpora from the Web using a new efficient tool chain. In Nicoletta Calzolari, et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 486–493. European Language Resources Association.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, 163(4148):688–688.

Valdivia, G. d., de Valdivia, G., Castellví, J., and Taulé, M. (2013). Morphological and lexical aspect in Russian deverbal nominalizations. In Irina Kor Chahine, editor, *Current studies in Slavic linguistics*, pages 267–280.

Zavarella, V. and Tanev, H. (2013). Fss-timex for tempeval-3: Extracting temporal information from

text. In *Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 58–63.

Zwanenburg, W. (2000). Correspondence between formal and semantic relations. In Geert E. Booij, et al., editors, *Morphologie/Morphology: Ein internationales Handbuch zur Flexion und Wortbildung/An international handbook on inflection and word-formation*, pages 840–850. Walter De Gruyter.