Richard Huyghe, Alizée Lombard, Justine Salvadori, and Sandra Schwab

Semantic rivalry between French deverbal neologisms in *-age*, *-ion* and *-ment*

Abstract: This study investigates the semantic aspects of the rivalry between French nominalizing suffixes *-age*, *-ion* and *-ment*. To control for lexicalization effects on derivational semantics, a sample of neologisms ending with the three suffixes is examined. A detailed systematic description of base verbs and derived nouns is provided, taking into account their lexical ambiguity. A total of 501 verb-noun pairs are described with respect to lexical aspect, semantic role assignment properties, and nominal semantic type. Different statistical methods are used to evaluate the relative influence of these properties on suffix selection, the predictability of suffix distribution, and the gradient nature of the rivalry between the three suffixes. Results reveal the importance of discriminative properties such as the referential type of the noun and the ability for event-denoting verbs and nouns to have a result state interpretation. Different degrees of distinctiveness and rivalry can be identified between the three suffixes. It appears that *-age* and *-ment* compete more with each other than they do with *-ion*.

Keywords: derivation, semantics, affix rivalry, neologism, nominalization

1 Introduction

The suffixes *-age*, *-ion* and *-ment* are arguably the most productive suffixes used to form eventuality-denoting deverbal nouns in French. Their semantic rivalry[1] has attracted a lot of attention in the last decades, but no consensus has yet emerged

1 Affix rivalry is understood here in a broad sense, based on the similarity of derivational patterns (Aronoff 1976; Baayen et al. 2013; Arndt-Lappe 2014; Schulte 2015; Bonami & Thuilier 2019; Dressler et al. 2019; Naccarato 2019; Varvara 2020; Huyghe & Wauquier 2021; a.o.). Two or more affixes are regarded as rivals if, in at least one of their patterns, they apply to base words from the same lexical class, and derive words from the same lexical class and with the same coarse-grained meaning. For a theoretical and historical overview of morphological competition, see Gardani et al. (2019).

Richard Huyghe, Alizée Lombard, Justine Salvadori, Sandra Schwab, Université de Fribourg, email of corresponding author: richard.huyghe@unifr.ch

https://doi.org/10.1515/9783111074917-006

as to whether these suffixes are equivalent forms or not, and what their distinctive properties could be. Some authors have claimed that there is no systematic difference between *-age*, *-ion* and/or *-ment* and that these suffixes are related to the same derivation type, as morphological exponents of the same word-formation rule or as possible allomorphs of a single affix expressing 'action' (Zwanenburg, 1984; Debaty-Luca, 1986; Dal et al., 2018). Others argue that different constraints apply to the three suffixes, but diverge on the constraints or focus on different aspects of the morphosemantic patterns associated with the suffixes. For instance, Dubois (1962) and Lüdtke (1978) suggest that *-age* contrasts with *-ment* in that it preferentially selects transitive verbal bases, whereas Kelling (2001) considers that the essential difference between *-age* and *-ment* is not the transitivity of the base verb, but the degree of agentivity of its subjects (*-age* selecting more agentive base verbs than *-ment*). Martin (2010) advocates a multifactorial analysis, involving agentivity, but also the length of the eventive chain (including more subevents for *-age* than for *-ment*), the incremental relation between events and themes (observable in the case of *-age* but not of *-ion* and *-ment*), and processive ontology (nouns in *-age* denoting physical processes more frequently than nouns in *-ment* and *-ion*). Fradin (2016, 2019) defends the idea that *-age* selects base verbs which denote more controlled or more concrete events than *-ment*, and that nominalizations in *-ment* denote states more frequently than nominalizations in *-age*. Finally, Wauquier (2020) argues that *-age* is preferentially used in technical domains and is therefore more specialized than *-ion* and *-ment*.

A limitation of existing studies is that they rarely provide quantitative information about the differences observed between the nominalizing suffixes, which hinders the possibility of generalization. Moreover, studies usually focus on lexicalized words, especially on morphological doublets (i.e. lexemes based on the same verbal input but ending with different suffixes). Although not infrequent, doublets remain the exception.[2] They may not reveal all differences between competing affixes and when lexicalized, they could exhibit idiosyncratic properties. More generally, lexicalized derivatives integrate all sorts of variations due to diachronic evolution, onomasiological needs, lexical competition, etc. (Corbin, 1987; Plag, 1999; Kawaletz & Plag, 2015). They do not directly allow for an exploration of derivational semantics (i.e. semantic correlates of morphological processes) as opposed to lexical semantics (i.e. semantic properties of estab-

2 As an indication, the Démonette database, which is a large morphological resource for French (Hathout & Namer, 2014), contains 404 cases of doublets in *-age* and *-ment* (13.9% and 16.5% of the nouns in the database ending in *-age* and *-ment*, respectively), 35 cases of doublets in *-ion* and *-ment* (1.2% and 1.4% of the nouns ending in *-ion* and *-ment*, respectively), and 26 cases of doublets in *-age* and *-ion* (0.9% of the nouns in both cases).

lished words in the lexicon), if one assumes that lexical semantics is based on derivational semantics but can further specify word meaning.

Our goal in this paper is to contribute to the study of the semantic rivalry between *-age*, *-ion* and *-ment* by investigating non-lexicalized derivatives in contemporary French. We will examine a sample of neologisms ending in *-age*, *-ion*, *-ment* in a quantitative approach and provide a detailed systematic description of both base verbs and derived nouns, taking into account their lexical ambiguity. It will be asked whether significant semantic differences can be observed between the three suffixes, and to what extent they could explain suffix distribution. In addition, considering affix rivalry as a gradient phenomenon, we will discuss whether different degrees of rivalry can be identified between the competing suffixes.

The article is structured as follows. In Section 2, we present the method used to sample and describe the semantic properties of French neologisms ending in *-age*, *-ion* and *-ment*. In Section 3, we report the results of the semantic description and present some differences observed between the three suffixes. In Section 4, we investigate the relative importance of the semantic properties and examine how they combine to predict suffix distribution. Finally, in Section 5, we discuss the gradient nature of the rivalry between the three suffixes. It will be concluded that *-age*, *-ion* and *-ment* are tendentially associated with different semantic operations, especially with respect to the type of entities or eventualities denoted by suffixed nouns. However, differences between the three suffixes are neutralized in many cases. The semantic overlap is particularly important in the case of *-age* and *-ment*, which compete with each other more than they do with *-ion*.

2 Method

Our study is based on the analysis of a sample of 300 French deverbal neologisms ending in *-age*, *-ion* and *-ment* (100 nouns per suffix). In this section, we present the method used to collect the neologisms, the semantic properties described for each verb-noun pair, and the annotation protocol we followed.

2.1 Data sampling

Neologisms in *-age*, *-ion* and *-ment* were extracted from the FRCOW16A corpus, which is a large French web corpus containing 10.8 billion tokens (Schäfer & Bildhauer, 2012; Schäfer, 2015). Words from the corpus ending in *-age(s)*, *-ion(s)*, and *-ment(s)* were filtered automatically using large lists of existing word forms taken

from the *Lexique* (New et al., 2004) and *Lefff* (Sagot, 2010) resources. Filtered data were abundant and noisy, due among other things to many misspellings and irrelevant forms. As a consequence, we randomly ordered candidate words and for each suffix selected the first 100 nouns that would satisfy the conditions C1-C3 below.

C1: The noun is morphosemantically related to a verb present either in lexicographic resources (*Le Petit Robert*, *Wiktionnaire*) or in the reference corpus (FRCOW16A). At least one meaning of the noun can be analyzed in relation to one meaning of the verb, whether the noun denotes the same eventuality as the verb or a participant in that eventuality. A noun such as *barillage* 'putting into barrels' was thus excluded because we could not find any attestation of the verb *bariller* with the meaning of 'put into barrels'.

C2: The noun is not analyzable as derived both from a verb and from an existing noun in *-age*, *-ion* or *-ment*. Possible nominal compounds or prefixed denominal nouns are discarded, on the assumption that existing nouns in *-age*, *-ion* and *-ment* may influence by analogy the form of neologisms. Accordingly, nouns such as *coenseignement* 'co-teaching' and *photocoagulation* 'photocoagulation' were excluded from the sample, as they can be analyzed as derived from *coenseigner* 'co-teach' and *photocoaguler* 'photocoagulate' or from *enseignement* 'teaching' and *coagulation* 'coagulation'.[3]

C3: The noun is not strictly a technical term. Although unknown to most speakers, technical terms may be lexicalized in specialized languages and there-

3 We excluded candidate nouns with the prefixes *dé-* and *re-* that are formally a prefixed version of an existing noun ending in *-age*, *-ion* or *-ment*. *Dé-* and *re-* are mostly deverbal prefixes, and it is uncertain whether they can select nominal bases or not. However, even deverbal *déXsuff* and *reXsuff* nouns might be formed by analogy with an existing *Xsuff* noun. To test this possibility, we selected 100 prefixed verbs in FRCOW16A for which there was an attested nominalization in *-age*, *-ion* or *-ment*: half of the verbs prefixed with *re-*, the rest prefixed with *dé-*. For each nominalization, we then checked in FRCOW16A whether there was a competing noun with the same prefix and one of the two other suffixes. We also searched for non-prefixed nominalizations corresponding to the three prefixed nominalizations. For example, for *débureaucratiser* 'debureaucratize' and *débureaucratisation* 'debureaucratization', we found that *bureaucratisation* 'bureaucratization' was also attested in the corpus, but that *débureaucratisage*, *débureaucratisement*, *bureaucratisage* and *bureaucratisement* were not. Finally, we performed a logistic regression analysis on the collected data. The fitted regression model was: $Prob\_Exist\_prefXsuff = -3.61 + 4.19 \times Exist\_Xsuff$. Likelihood ratio-tests showed a significant effect ($p < 2.2e\text{-}16$). The probability that a nominalization in *-age*, *-ion* or *-ment* prefixed with *re-* or *dé-* exists is strongly influenced by the existence of a corresponding non-prefixed nominalization. In other words, there is an important formal attractiveness between prefixed nominalizations and their non-prefixed equivalents. Such formal analogies may bias the analysis of the relationship between suffix selection and semantic properties of verb-noun pairs, hence the exclusion of the words concerned.

fore not be neologisms. Technical terms absent from standard lexicographic resources can be identified through their corpus occurrences, when these clearly relate to specialized discourse. For example, listed candidates such as *abergeage* 'feodal contract', *carassonage* 'vineyard trellis repair', *chanfreinage* 'bevelling', *enzymage* 'enzyming', *trancanage* 'crosswinding', and *panotage* 'panning' were excluded from the sample. During the selection, it appeared that technical terms were much more frequently found among *-age* candidates than among *-ion* and *-ment* candidates. Although we did not precisely quantify the difference, this seems to confirm previous claims made by Dubois (1962), Fleischman (1980), Uth (2010) and Wauquier (2020) about the technicality of *-age*.

Examples of neologisms in *-age*, *-ion* and *-ment* satisfying C1-C3 and included in the sample are given in (1).

(1) a. affolage 'panic', brancardage 'stretcher bearing', corbeillage 'trashing', dandinage 'swaying', implorage 'imploring', militage 'campaining', oubliage 'forgetting', pixelisage 'pixelization', suspectage 'suspecting', visitage 'visit'

b. alternation 'alternating', christification 'christification', colmatation 'filling-in', confortation 'comfort', expulsation 'expulsion', foiration 'screwing up', poutinisation 'putinization', romanticisation 'romanticization', rutilation 'shine', semestrialisation 'semesterization'

c. absorbement 'absorption', atténuement 'softening', bedonnement 'paunch', cernement 'encirclement', ficellement 'tying up', ignorement 'ignoring', malaxement 'kneading', résiliement 'termination', subissement 'putting up with', trompement 'cheating'

2.2 Annotated properties

A number of syntactic and semantic properties are mentioned in the literature as possible discriminating factors between *-age*, *-ion* and *-ment*, including:

- the transitivity of the base verb;
- the semantic type of the derived noun;
- the lexical aspect of both verbs and nouns;
- the semantic roles assigned by verbs and nouns to their arguments.

In this study, we propose a systematic analysis of these properties. For each verb-noun pair included in the sample, we annotated a series of features related to the above-mentioned properties, while also taking into account the lexical ambiguity of both the verb and the noun. The description was based on controlled manual

annotation and precise definitions of the annotated features. The general principles and linguistic tests used to analyze the properties of verbs and nouns are detailed in an annotation guide available in the supplementary material of the paper.

We analyzed the semantic type of the nouns by distinguishing between their ontological and relational descriptive properties, each noun being doubly classified. Ontological types relate to the nature of the referents, whereas relational types depend on the semantic relation with the base. Existing classifications of nominalizations often assimilate the two kinds of properties, possibly leading to confusion in semantic descriptions. Ontological and relational types belong to different taxonomies since, as illustrated in (2) and (3), an ontological type can be related to different relational types and conversely.

(2) a. bâtir 'build' → bâtiment 'building' [ARTEFACT-RESULT]
b. fixer 'fasten' → fixation 'fastener' [ARTEFACT-INSTRUMENT]
c. garer 'park' → garage 'garage' [ARTEFACT-LOCATION]

(3) a. bâtir 'build' → bâtiment 'building' [ARTEFACT-RESULT]
b. énerver 'irritate' → énervement 'irritation' [STATE-RESULT]
c. traduire 'translate' → traduction 'translation' [COGNITIVE-RESULT]

Thirteen ontological simple types are distinguished based on distributional tests taken from the literature on French nominal semantics (Godard & Jayez, 1996; Flaux & Van de Velde, 2000; Huyghe, 2015; Haas et al., 2022). Some of them combine to form complex types, in which case characteristic predicates of different simple types are contextually compatible (Copestake & Briscoe, 1995; Cruse, 1995; Pustejovsky, 1995; Kleiber, 1999; Asher, 2011; Dölling, 2020; a.o.). For example, the noun *déclaration* 'statement' in (4) instantiates a complex type of event and cognitive object. The eventive facet is selected by *effectuer* 'perform' and the cognitive facet by *selon lequel P* 'according to which P'.

(4) 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'

Relational types are based on the semantic roles used to analyze arguments, complemented with a transpositional type for nouns that denote roughly the same kind of eventualities as their base verb (i.e. with respect to the event/state distinction). We defined a set of 17 semantic roles adapted from *Verbnet* (Kipper-Schuler, 2005) and *Lirics* (Petukhova & Bunt, 2008). Since distributional tests cannot be

used to differentiate semantic roles, we relied on explicit definitions to identify the roles assigned to the arguments of verbs and nouns.

Lexical aspect was decomposed into four basic features (dynamicity, durativity, telicity, and post-phase) and analyzed using linguistic tests proposed in the literature (Vendler, 1967; Dowty, 1979; Rothstein, 2004; Haas et al., 2008; Filip, 2012; a.o.). telicity was encoded by default with a delimited internal argument, and annotated as variable for degree achievements (Abusch, 1986; Bertinetto & Squartini, 1995; Hay et al., 1999; Rothstein, 2008; a.o.). Other aspectual features are binary. We labelled as 'post-phase' the possibility for a dynamic eventuality to include a durative result state, as illustrated in (5) with *partir* 'leave' vs. *arriver* 'arrive' (Piñón, 1997, 1999; Apothéloz, 2008; Fradin, 2011; Haas & Jugnet, 2013).

(5) Julie {est partie/?est arrivée} pendant deux jours.
'Julie {left/arrived} for two days'

The possibility of being interpreted as a result state when combined with temporal complements or with the verb *durer* 'last', as in (6), was used as a criterion for the identification of nominal post-phase.

(6) L'exclusion a duré six jours, avant que les services éducatifs régionaux n'ordonnent sa réintégration. (web)
'The exclusion lasted six days before regional educational services ordered his reinstatement'

To account for the polysemy of nominalizations, the different meanings of each verb and noun were carefully distinguished and systematically paired. Ambiguous nouns were assigned one entry per meaning in the database. The lexical ambiguity of a given form was identified through the variation of at least one annotated property. Verbal and nominal lexemes were paired based on the principle of closest semantic correspondence: If a verb or a noun is ambiguous, the verbal and nominal lexemes that share the most aspectual and role-assigning properties are paired together. For instance, two eventive meanings were found for the noun *croquement* 'crunching/sketching', illustrated by occurrences such as *le croquement des glaçons* 'the crunching of ice cubes' and *le croquement des tatouages* 'the sketching of tattoos'. These two meanings are, respectively, punctual and durative, and associated with the subcategorization of a patient and a result argument. They were paired with two different verbs (*croquer* 'crunch' and *croquer* 'sketch') exhibiting the same distinctive features as the two nominal lexemes, and were annotated in both cases as instantiating a transpositional relational type. Two meanings were also found for the noun *retouchement* 'retouching/modification': an event and an artefact meaning (*opérer un retouchement des paupières* 'perform

an eyelid retouching' vs. *des retouchements blancs sur graphite* 'white modifications on graphite'). But in this case both meanings are related to the same meaning of the verb *retoucher* 'retouch', i.e. a transitive accomplishment verb subcategorizing an agent and a patient. *Retouchement* was then assigned two entries linked to the same base verb, analyzed respectively as a transpositional and a resultative relational type.

2.3 Annotation protocol

Verb-noun pairs included in the sample were annotated in a double-blind process and adjudicated with the help of a third annotator. The different meanings of each noun were identified through the occurrences in FRCOW16A, complemented with examples taken from the web. Ten nouns per suffix were used for joint training and refinement of the annotation guidelines; then, two annotation sessions of 45 nouns per suffix were conducted. Inter-annotator agreement scores over the two sessions were calculated using Cohen's kappa, as well as prevalence-adjusted and bias-adjusted kappa (PABAK). The latter compensates for the overvaluation of disagreement cases with Cohen's kappa when categories are highly unbalanced in value distribution (Byrt et al., 1993). Agreement scores for each annotated verbal and nominal property are presented in Table 1. They indicate an overall substantial inter-annotator agreement[4] and can be regarded as evidence of the operationality of the linguistic categories we used to describe the properties of verbs and nouns.

The 300 nouns in the sample were finally associated with 501 meanings (162 for *-age*, 168 for *-ion*, 171 for *-ment*), ranging from 1 to 4 meanings per noun, with an average of 1.67. These 501 meanings are related to 418 distinct verbal meanings (123 for *-age*, 153 for *-ion*, 142 for *-ment*). In 107 cases of ambiguity (59.1%), multiple nominal meanings associated with the same form are derived from multiple verbal

4 Various scales have been proposed to interpret kappa values. According to the reference scale defined by Landis & Koch (1977), the agreement is "substantial" when kappa scores range from 0.61 to 0.80 and "almost perfect" when they range from 0.81 to 1.00. In the alternate scale introduced by McHugh (2012), the agreement is said to be "moderate" between 0.60 and 0.79, "strong" between 0.80 and 0.90, and "almost perfect" between 0.90 and 1.00. Beyond the inevitable arbitrariness of those evaluation scales, inter-annotator agreement should be interpreted with regard to the specific nature of each annotation task and the inherent fuzziness of the phenomena described.

Tab. 1: Inter-annotator agreement per property.

	Observed agreement	Kappa	PABAK
V transitivity	0.96	0.92	0.93
V Dynamicity	0.98	0.56	0.96
V durativity	0.86	0.59	0.72
V telicity	0.82	0.72	0.73
V Post-phase	0.78	0.65	0.67
V Role of subject	0.79	0.71	0.76
V Role of object	0.83	0.78	0.82
V Role of oblique	0.90	0.61	0.90
N Ontological type	0.83	0.77	0.82
N Relational type	0.93	0.78	0.92
N Dynamicity	0.95	0.85	0.92
N durativity	0.87	0.71	0.80
N telicity	0.85	0.79	0.77
N Post-phase	0.83	0.73	0.74
N Role of 1st arg.	0.78	0.72	0.77
N Role of 2nd arg.	0.80	0.71	0.78
N Role of 3rd arg.	0.95	0.60	0.93
Average	0.87	0.72	0.82

meanings, following the pattern in (7). In 74 cases of ambiguity (40.9%), they are derived from the same verbal meaning, following the pattern in (8).[5]

(7) Multibase ambiguity
V1 → N1
V2 → N2

(8) Single-base ambiguity
V1 → N1
V1 → N2

5 It is uncertain whether N2s in single-base ambiguities result from a derivational operation or from a lexical figure such as metonymy (see Ferret & Villoing 2015 for a related discussion on the formation of instrument-denoting nouns in *-age*). Evidence in favor of the morphological or figurative construction is difficult to provide. For a given N1/N2 type, the existence of derived N2s without a corresponding N1 in the lexicon shows the existence of a morphological pattern, but does not imply that any N2 with the semantic type considered results from derivation. Conversely, the existence in the lexicon of metonymically-related underived N1s and N2s shows the existence of a figurative pattern, but does not imply that any N2 results from metonymy. It could also be hypothesized that the two types of patterns are not necessarily exclusive of each other, and that morphological and figurative derivations actually combine to favor the formation of ambiguous words (Huyghe, 2021).

The 501 verb-noun pairs vary with respect to annotated properties, and it can be asked whether these variations depend on the suffix used to form deverbal nouns.

3 Observed differences

In this section, we describe the distribution between *-age*, *-ion*, *-ment* according to the different properties we analyzed, as a first approach to suffix similarity and distinctiveness. Due to space limitations, we present only a subset of the results.[6] Some properties of our sampled verb-noun pairs deserve special attention, either because they have been much debated in the literature, or because they show important variation between neologisms formed with the three suffixes. We focus on six properties: (i) three verbal properties frequently discussed in studies on *-age*, *-ion* and *-ment*, and (ii) three nominal properties distributed across the suffixes with particularly salient differences—and that will prove in further analyses to have a major influence on suffix selection (see Section 4).

3.1 Verb transitivity

As shown in Figure 1, the three suffixes mostly select transitive base verbs, but the preference is more pronounced for *-age* (76.4% of transitive base verbs) than for *-ion* and *-ment* (58.8% and 59.9% of transitive base verbs, respectively).[7] To some extent, this result supports the observations of Dubois (1962) and Lüdtke (1978) who state that *-age* tends to select transitive bases and *-ment* intransitive bases. Counterexamples mentioned by Kelling (2001) and Martin (2010) do not seem to affect the general tendency towards a more frequent selection of transitive bases with *-age*. Nevertheless, the tendency only holds true in the perspective of suffix comparison. The suffix *-age* does not privilege transitive over intransitive base verbs with regard to the distribution of transitive and intransitive verbs in

6 The complete description of the 17 verbal and nominal properties is available in the supplementary material of the paper.

7 The quantitative results in Sections 3.1, 3.2 and 3.3 are based on the number of distinct verbal lexemes in the dataset, as opposed to their frequency as a base in the dataset. Verbal lexemes present in multiple entries, i.e. associated with different nominal lexemes in single-base ambiguity configuration, were counted only once in the statistics. This counting is appropriate for a discussion of previous claims about the rivalry between *-age*, *-ion* and *-ment*, given that existing studies do not take into account nominal ambiguity in the comparison of the base verbs selected by the different suffixes.

the lexicon. As a comparison, the lexical resource *Les Verbes Français* (Dubois & Dubois-Charlier, 1997) contains 19580 transitive and 6029 intransitive verbal lexemes (in the sense intended here). That distribution is not significantly different from that of verbs deriving *-age* nouns in our sample (χ^2 (1, N = 25732) = 0.0001, p = .99). By contrast, the distribution between transitive and intransitive is clearly unbalanced in favor of intransitive verbs with *-ment*, even if *-ment* selects mostly transitive bases (χ^2 (1, N = 25751) = 21.56, p = 3.4e-06).

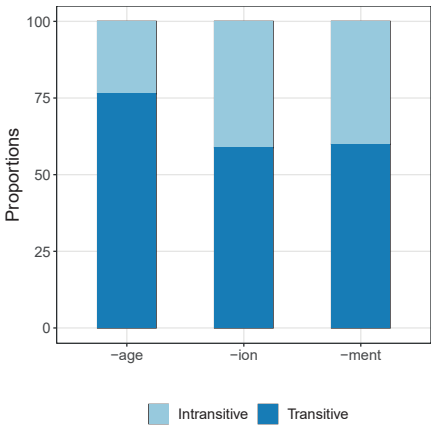

Fig. 1: Transitive and intransitive base verbs per suffix (%).

3.2 Semantic role of verb subjects

Important differences can be observed between *-age*, *-ion* and *-ment* with respect to semantic roles assigned to their base verb subjects. As illustrated in Figure 2, *-age* has a strong predilection for base verbs subcategorizing agent subjects, whereas *-ion* prefers base verbs with cause or patient subjects (especially verbs allowing for the causative-anticausative alternation). The suffix *-ment* is more similar to *-age* than to *-ion*, but it selects less agentive and more patientive verbs than *-age*, as well as verbs assigning more diverse roles to their subjects (with higher proportions of theme, stimulus, pivot subjects than the two other suffixes). This specificity echoes Kelling's statement that "the French suffix *-age* combines with verb stems whose first argument is proto-agentive, whereas the French suffix *-(e)ment* combines with verb stems whose first argument is less proto-agentive" (Kelling, 2001, 155). However, this contrast should only be regarded as a tendency, since

prototypical agentive base verbs regularly combine with *-ment* to form neologisms (e.g. *déblatèrement* 'badmouthing', *mitraillement* 'machine-gunning', *retapement* 'refurbishment').

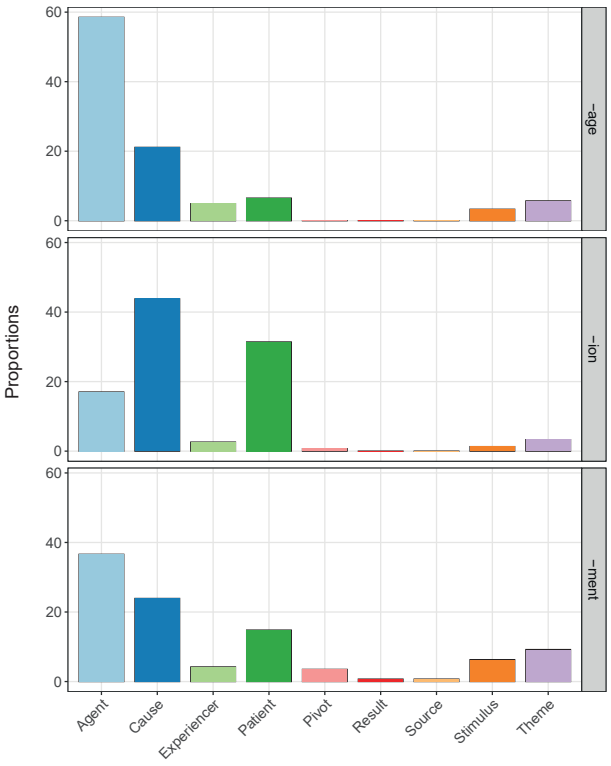

Fig. 2: Semantic roles assigned to base verb subjects per suffix (%).

3.3 Verb telicity

The data we collected do not show any significant difference between *-age* and *-ment* with respect to base verb telicity (χ^2 (1, N = 265) = 3.43, p = .18). In both cases, about half of the base verbs are telic, less than half atelic, and a minority of verbs with variable telicity can be observed (see Figure 3). There is a clear contrast with *-ion*, which shows a strong preference for verbs with variable telicity

(66.0% of the base verbs). These results contradict the findings of Martin (2010), who argues that the selection between *-age*, *-ment* and *-ion* is not correlated with the aspectual properties of base verbs. The fact that Martin focuses on psychological verbs and their nominalizations, and does not encode variable telicity as a distinct feature could explain this difference. It can also be noted that in our data, most verbs with variable telicity and deriving *-ion* nouns are suffixed with *-iser* (84.2%)—and reciprocally, that most verbs ending in *-iser* are nominalized with *-ion* and have variable telicity (81.7%). Not only does *-ion* select more *-iser* verbs than the two other suffixes (see Table 2), but most *-iser* verbs are also verbs with variable telicity in our sample (see Table 3).[8] The fact that verbs in *-iser* preferentially form neologisms ending in *-ion* rather than *-age* and *-ment* confirms the morphological tendency reported by Missud & Villoing (2020). The semantic counterpart of that morphological tendency appears to be the predilection for verbs with variable telicity.

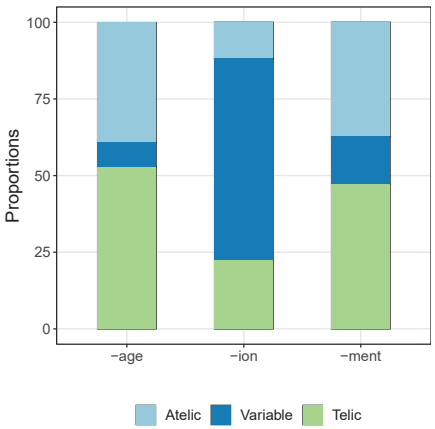

Fig. 3: Telicity of base verbs per suffix (%).

8 It can be asked whether *-isation* should be analyzed as a suffix in contemporary French, given its productivity and the fact that some nouns in *-isation* do not have a corresponding verb in *-iser* (Lignon et al., 2014; Dal & Namer, 2015; Cartier, 2018; Missud & Villoing, 2020). Our sample only includes nouns in *-isation* for which an existing verb in *-iser* is attested, to ensure at least the possibility of analyzing neologisms in *-isation* as derived from verbs.

Tab. 2: Number of base verbs ending or not in *-iser* per suffix (percentages by row).

	V ending in *-iser*	V not ending in *-iser*
-age	1 (0.8%)	122 (99.2%)
-ion	103 (67.3%)	50 (32.7%)
-ment	0 (0%)	142 (100%)

Tab. 3: Number of base verbs with or without variable telicity ending or not in *-iser* (percentages by row).

	V with variable telicity	V without variable telicity
V ending in *-iser*	86 (82.7%)	18 (17.3%)
V not ending in *-iser*	47 (15%)	267 (85%)

3.4 Nominal ontological type

The semantic type of derived nouns is one of the most discriminative properties between *-age*, *-ion* and *-ment*. The most contrasted properties pertain to ontological types, as opposed to relational types. Since *-age*, *-ion* and *-ment* are mostly transpositional suffixes, differences between the three suffixes with respect to relational types are only marginal—the main difference observed is the propensity of *-ment* to denote more results than *-age* and *-ion*.[9] Ontological types are more diverse, as can be seen in Figure 4, and some suffix peculiarities can be observed. A specificity of *-ion* is that it forms mostly nouns with a complex type combining eventive and stative descriptions (66.7% of the nouns). Most neologisms in *-ion* denote dynamic eventualities that involve a salient state, which is also strongly correlated with the variable telicity of the nouns inherited from the base verbs: 90.2% of the *-ion* nouns with a complex event-state type have variable telicity (e.g. *compaction* 'compacting', *turquification* 'turkification', *verdurisation* 'greenification'). In other words, *-ion* clearly privileges the derivation of nouns that denote a progressive change of state. Neologisms in *-ion* also denote events (in 20.8% of the cases), but other ontological types are poorly represented. Unlike *-ion*, the suffixes *-age* and *-ment* form mostly event nouns (52.5% and 38.0% of the nouns respectively) and are more likely to derive nouns that denote artefacts (8.0% and 7.6% of the nouns) and complex entities combining events and cogni-

9 Chi-squared statistics for the distribution between resultative and non-resultative nominal types indicate a significant distinction in the case of *-age*/*-ment* (χ^2 (1, N = 333) = 9.600, p = .0019) and *-ion*/*-ment* (χ^2 (1, N = 339) = 18.086, p = 2.1e-05), but not in the case of *-age*/*-ion* (χ^2 (1, N = 330) = 1.4367, p = .2306).

tive objects (8.0% and 5.3% of the nouns). The suffix *-ment* forms less strictly eventive nouns than *-age*, but more nouns with a complex event-state type (e.g. *crispement* 'clenching') and more nouns denoting states (e.g. *déconcertement* 'confusion'). The specificity of *-age*, besides the fact that it is the suffix that forms the most event nouns, is that it is the only one in our data to derive domain-denoting nouns—domains being defined as non-occurrential activities, i.e. dynamic eventualities that do not 'take place' as spatio-temporally individuated events (e.g. *marketage* 'marketing', *aquarellage* 'watercoloring', *militage* 'campaining'). We identified 20 domain-denoting neologisms, all of which are suffixed with *-age*, which reflects the affinity between *-age* and domain description.

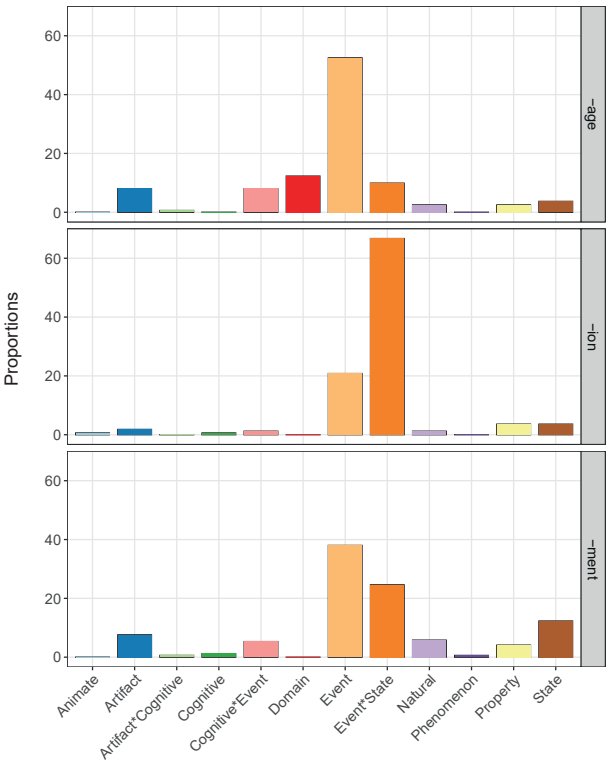

Fig. 4: Semantic type of derived nouns per suffix (%).

3.5 Nominal post-phase

Another property that varies considerably across nominalizations in *-age*, *-ion* and *-ment* is the ability to denote eventualities with a post-phase and to be interpreted as the result state of an event (see Figure 5). The suffix *-ion* contrasts with *-age* and *-ment* in that it forms mostly nouns that can have a post-phase interpretation, which is directly related to the predilection of *-ion* for complex event-state eventualities. Many nouns denoting a gradable change of state can be contextually interpreted as denoting the state that results from the changing process.

The suffixes *-age* and *-ment* derive mostly nouns that denote eventualities without post-phase, but this tendency is more prominent in the case of *-age* (77.2% of the derived nouns) than in the case of *-ment* (59.1% of the derived nouns). Transpositional nouns in *-age* can frequently change the post-phase feature of the base verb: 19.1% of those nouns do not inherit the post-phase specification of their base verb, while that proportion is only of 3.3% for *-ion* and 4.1% for *-ment*.[10] Changes observed between verbs and nouns always consist in the loss of post-phase interpretation. For example, *sortage* 'taking out', unlike *sortir* 'take out', cannot have a post-phase interpretation. Whereas the duration complement in (9-a) can be related to a result locative state, it necessarily characterizes a motion process in (9-b).

(9) a. Il a sorti son appareil photo pendant dix minutes.
'He took out his camera for ten minutes'
b. Le sortage de l'appareil photo a duré dix minutes.
'Taking the camera out took ten minutes'

3.6 Nominal durativity

durativity as an aspectual property applies only to nouns that denote eventualities and is comparable to post-phase in this respect. The specificity of nominal durativity is that it distinguishes *-ment* from both *-age* and *-ion*. The suffix *-ment* forms more nouns that denote punctual eventualities, i.e. achievements, than its rivals (e.g. *démissionnement* 'resignation', *trinquement* 'clinking of glasses', *heurtement* 'knock'). durative eventualities are denoted by 74.7%, 81.0%, and 55.6% of the neologisms suffixed with *-age*, *-ion*, and *-ment*, respectively. By con-

10 The difference observed with respect to preservation of verbal post-phase is significant between *-age* and *-ion* (χ^2 (1, N = 317) = 14.7637, p = 1.22e-04), between *-age* and *-ment* (χ^2 (1, N = 289) = 10.9067, p = 9.58e-04), but not between *-ion* and *-ment* (χ^2 (1, N = 282) = 0.1032, p = .7480).

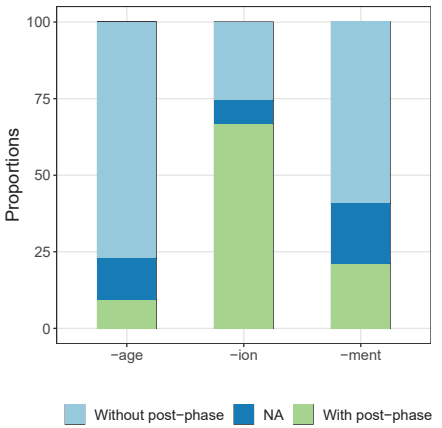

Fig. 5: Post-phase of derived nouns per suffix (%).

trast, non-durative eventualities are denoted by 11.7%, 11.3%, and 24.6% of the neologisms suffixed with *-age*, *-ion*, and *-ment*, respectively (see Figure 6). Overall, 52.5% of the nouns that denote achievements in our sample are suffixed with *-ment*.

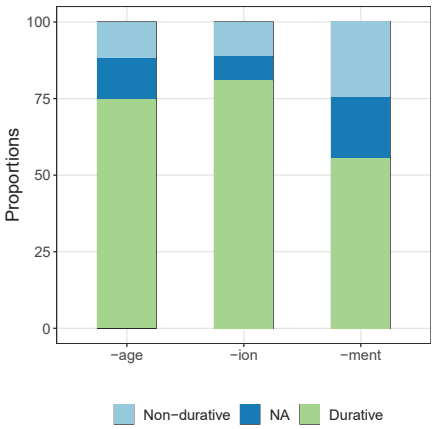

Fig. 6: Durativity of derived nouns per suffix (%)

Verbal durativity is more frequently preserved through nominalization than verbal post-phase, and the punctual aspect of derivatives in *-ment* is inherited from their base verb. Only 1.3% and 2.2% of the eventuality-denoting nouns in *-ion* and

-ment differ from their base with respect to durativity, whereas the proportion is 9.3% in the case of *-age*.[11] The change of aspect with *-age* can be observed for nouns derived from achievement verbs and denoting durative events (e.g. *déposer* 'drop off' vs. *déposage* 'dropping off'). As in the case of post-phase, *-age* appears to be the suffix that least preserves the lexical aspect of the base.

4 Combined influence of semantic properties

Semantic properties of verbs and nouns jointly influence the selection of *-age*, *-ion* and *-ment*. It can be asked how these properties combine to determine the distribution of the three suffixes, which properties are the most influential, and how much of the suffix distribution they can explain. In this section, we first examine correlations between the semantic properties we encoded, then use a random forest analysis to assess their relative importance and the predictability of suffix selection.

4.1 Correlations between properties

As a preliminary to statistical analysis of joint influence, we estimated pairwise correlations between the 17 annotated properties. The correlations were first evaluated using Pearson's chi-squared tests, with simulated *p*-value based on 2000 replicates for categories with expected counts less than 5. Out of 136 pairs, 98 show a significant relationship at $p < .05$. The most uncorrelated properties are the semantic role of the third nominal argument, nominal durativity, and the semantic role of the verb oblique, which have a non-significant relationship with 10, 9 and 8 other properties, respectively.

We used Cramér's V tests to evaluate the strength of the correlation for significant associations. Results are reported in Figure 7. The properties with the greatest number of strong correlations are the semantic role of the first nominal argument, verbal telicity, and nominal telicity, which are correlated respectively with 12, 9, and 9 other properties at Cramér's $V > 0.5$. Special attention can be paid to semantic properties present both in the verbal and nominal domains, such as aspectual properties and argument roles. These appear to be strongly but not fully

11 The difference observed with respect to preservation of verbal durativity is significant between *-age* and *-ion* (χ^2 (1, N = 295) = 9.7434, p = .0114), between *-age* and *-ment* (χ^2 (1, N = 277) = 6.406, p = .0018), but not between *-ion* and *-ment* (χ^2 (1, N = 292) = 0.3496, p = .5544).

correlated, which suggests that the preservation of semantic features across categories is not necessarily observed (even in the case of eventuality-denoting nominalizations), as already noted in Sections 3.5 and 3.6. Among aspectual features, dynamicity is subject to important variation (Cramér's V = 0.40), with possible aspectual shifts from dynamic verbs to stative nouns, as in the case of nominal result state reading. The correlation is stronger for telicity (Cramér's V = 0.94), durativity (Cramér's V = 0.88), and post-phase (Cramér's V = 0.83), but with differences that indicate the possibility of cross-categorial variation, especially in the case of post-phase.

The preservation of argument structure cannot be assessed as directly as that of lexical aspect, because of syntactic differences between verbal and nominal arguments, and of the possible variation in nominal argument position. However, high correlations can be observed between semantic roles of verbal objects and first nominal arguments (Cramér's V = 0.93), and semantic roles of verbal obliques and second nominal arguments (Cramér's V = 0.91), presumably because of semantic preservation. The correlation between verb subjects and nominal arguments is somewhat weaker, since the highest correlation coefficient observed between their semantic roles is 0.84, which indicates a possible alteration of the subject argument in nominalization.

The considerable amount of correlations observed in our data requires the use of an adapted statistical method to analyze the combined and relative influence of semantic properties on suffix selection.

4.2 Relative importance of properties

We used a random forest algorithm to determine how verbal and nominal properties can jointly predict the distribution of *-age*, *-ion* and *-ment*, and individually contribute to the prediction. random forests are an ensemble method designed to predict a response variable with respect to a set of possible explanatory variables (Breiman, 2001; Tagliamonte & Baayen, 2012; Levshina, 2020). They operate by averaging predictions from a large number of conditional inference trees, themselves resulting from binary recursive partitioning of data according to predictors. random forests are based on the randomization of both the data subsamples used as training sets in decision trees and the subset of predictors tested at each node of each tree. They provide reliable information about predictive accuracy and can be used to assess the relative importance of predictor variables. Being non-parametric and able to handle high dimensional data with correlated and interacting variables, random forests are well adapted to the analysis of our data,

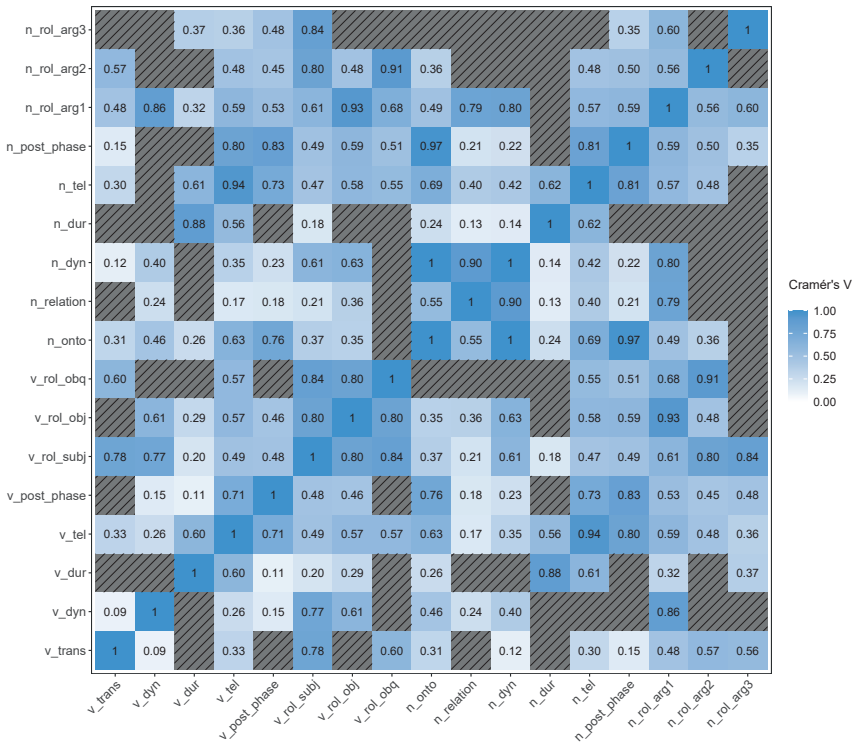

Fig. 7: Correlations between nominal and verbal properties based on Cramér's V tests. Coefficients are not computed for non-significant associations.

given its sparseness (i.e. the small number of observations with respect to the number of predictors) and the correlations observed between semantic properties.

We conducted a random forest analysis based on 3000 conditional inference trees, with random samples of 5 predictor variables at each tree node.[12] The developed model shows an accuracy of 62.3%. The important proportion of wrongly predicted cases can be interpreted as the effect of semantic indistinguishability between the three suffixes. A closer look at the discrepancies between observed and predicted data reveals important differences between the three suffixes (see Table 4). The most predictable suffix is *-ion*, followed by *-age* and then *-ment*. The discrepancy between observed and predicted suffixes varies according to suffix

12 The random forest analysis was performed using the party (Hothorn et al., 2006; Strobl et al., 2007, 2008) and permimp (Debeer & Strobl, 2020) packages in R.

pairs. The suffixes *-age* and *-ment* are more frequently confused with each other than they are both with *-ion*: 68.0% of the wrongly predicted cases for *-age* are mistaken for *-ment*, whereas 60.3% of the wrongly predicted cases for *-ment* are mistaken for *-age*. As for *-ion*, it is more frequently confused with *-ment* than with *-age*, since 66.7% of the wrongly predicted cases for *-ion* are mistaken for *-ment*. This result suggests the existence of degrees of rivalry between the three suffixes, *-age* and *-ment* competing more with each other (as being more confusable and therefore more similar) than they both compete with *-ion*, and *-ion* competing more with *-ment* than with *-age*. The gradient aspects of rivalry will be further explored in the next section.

Tab. 4: Classification of deverbal neologisms in random forest model according to their suffix. Percentages indicate the proportion of reference items predicted as *-age*, *-ion* or *-ment* (in columns). The most abundant items in each predicted group are indicated in bold.

		Reference		
		-age	***-ion***	***-ment***
Prediction	*-age*	**103** (63.6%)	24 (14.3%)	51 (29.8%)
	-ion	12 (7.4%)	**113** (67.3%)	24 (14.1%)
	-ment	47 (29.0%)	31 (18.4%)	**96** (56.1%)

We analyzed the contribution of each semantic property to the random forest prediction by using a conditional computation of variable importance that takes into account predictor correlations (Debeer & Strobl, 2020).[13] The outcome of the analysis is presented in Figure 8. Results indicate an alternation of nominal and verbal properties in variable ranking, with some predominance of the former, three nominal properties being found among the four most important variables. Almost all properties contribute to the prediction of suffix distribution (with the exception of verb dynamicity and role of third nominal argument), which confirms both the rel-

13 As noted by Debeer & Strobl (2020), there is no consensus on the exact nature of variable importance. It can be defined in a more or less marginal or partial perspective, depending on whether the impact of a predictor is evaluated independently or conditionally upon other predictors. When predicting the distribution of *-age*, *-ion* and *-ment*, the main difference between conditional and unconditional computations of variable importance concerns nominal and verbal telicity (ranked 14th and 15th in conditional estimation, but 3rd and 4th in unconditional estimation). This difference can be explained by the fact that (i) telicity in itself is an important discriminative factor between the three suffixes, and is therefore important in a marginal perspective, but (ii) telicity is one of the most correlated predictors, presumably highly redundant, and therefore minor in a partial perspective.

evance of the selected predictors and the semantic complexity of the distinction between *-age*, *-ion* and *-ment*. The most influential property is the semantic ontological type of the noun, whose impact on suffix selection is described in Section 3.4. The discriminative capacity of ontological type as opposed to relational type is not unexpected since the three suffixes compete in having mostly the same relational output. Nevertheless, the major importance of ontological type compared to all encoded properties shows that derivational semantics determines not only relations, but also referential descriptions.

Among the aspectual properties of verb-noun pairs, post-phase appears to be the most important variable. Post-phase properties of related nouns and verbs may diverge in some cases, with significant differences between the three suffixes, which explains why both are highly ranked. Verbal post-phase notably influences the distribution between *-ion* on the one hand (which selects mostly verbs with post-phase) and *-age* and *-ment* on the other (which select mostly verbs without post-phase). Nominal post-phase further distinguishes *-age* from *-ment*, as described in Section 3.5. Nominal durativity is another important aspectual variable in that it contributes to the distinction of *-ment*, which tends to form more achievement nouns than *-age* and *-ion*, as indicated in Section 3.6.

The semantic role of verbal oblique arguments is ranked as the most influential role assignment property, partly because it is one of the most uncorrelated features in the dataset. Its contribution concerns mainly locative roles, indicating predilections for base verbs that assign the roles of location (*-age*), theme (*-ment*), source (*-ion*), and destination (*-age* and *-ment*) to their oblique arguments. The semantic role of the verb subject plays an important role as well, following the preferences presented in Section 3.2 and related to the salient agentivity of *-age* base verbs, and the important causativity and patientivity of *-ion* base verbs. The role of the second argument of the noun (ranked as 3rd role assignment property) varies between *-age* and *-ion*, which select agentive and causative arguments, respectively. However, the distinctive contribution of that property concerns *-ment*. Nouns suffixed with *-ment* exhibit a relative preference for monovalent structure (in which case the second argument is absent) or for theme external arguments (especially in co-theme configuration, as in *le convergement de X avec Y* 'the converging of X with Y').

5 Gradient rivalry

As argued by Huyghe & Wauquier (2021), affix rivalry can be considered a gradient phenomenon. Affixes can be seen as more or less rivaling depending on the se-

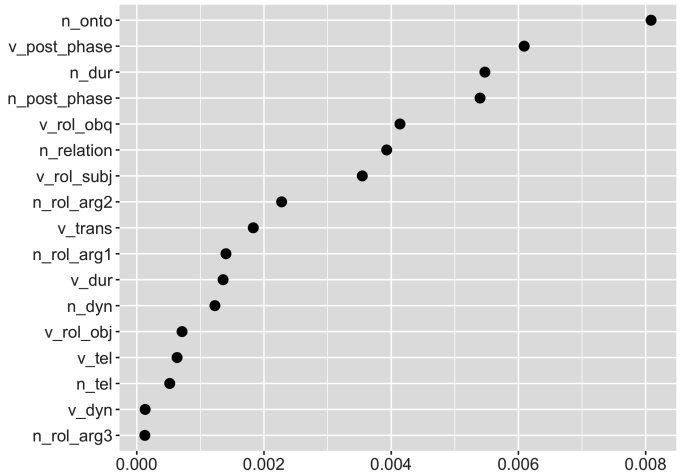

Fig. 8: Conditional importance of predictor variables for suffix selection.

mantic granularity with which morphological competition is evaluated. Furthermore, given that many affixes are polyfunctional, affix rivalry can vary according to (i) the proportion of semantic functions shared between affixes, and (ii) the frequency of lexical realization of these functions. Rivalry between *-age*, *-ion* and *-ment* is usually inferred from the fact that these suffixes can form event nouns. However, finer semantic distinctions can be made, and both the proportion of specific semantic functions the three suffixes have in common and the frequency with which they instantiate these functions may be variable. Results from Sections 3 and 4 suggest the existence of different degrees of rivalry between *-age*, *-ion* and *-ment*. In this section, we further investigate suffix similarity and gradient rivalry by using unsupervised methods, based on cluster analysis and dimensional reduction of our semantic dataset.

5.1 Cluster analysis

A way to approach gradient rivalry is to perform a cluster analysis of words with competing affixes, based on their distinctive properties. Clustering consists in grouping together similar objects of a dataset, so that objects in the same cluster are more similar than objects in different clusters. Dividing the sample of neologisms into clusters based on semantic properties and examining the distribution

of the three suffixes in these clusters can provide information about their degree of similarity or dissimilarity, and therefore of rivalry.

To examine the distribution of neologisms in *-age*, *-ion* and *-ment* in semantically similar groups, we first built a distance matrix based on our dataset, using a dissimilarity measure that can handle categorical data (Gower distance).[14] Then we applied a hierarchical clustering to the matrix using a linkage method that creates clusters in such a way that the variance of the merged clusters is minimized (Ward linkage). Finally, we split the dataset into three clusters to be compared with the distribution between the three suffixes. The result is presented in Table 5. No one-to-one correspondence between suffixes and clusters can be observed, but clustering and suffix distribution are not independent (χ^2 (4, N = 501) = 145.83, $p <$ 2.2e-16). Each cluster is dominated by a different suffix, although with important differences in cluster size: 59.7%, 13.8% and 26.5% of the items are grouped in the first, second and third clusters, respectively. Almost four-fifths of the *-age* items fall into the same cluster, whereas *-ion* is characterized by the existence of a cohesive distinctive subgroup containing three-fifths of its items. The specificity of *-ment* is to be mostly represented in a group in which it is not dominant (Cluster 1).

The clustering reveals similarity differences between the three suffixes. It shows the distinctiveness of *-ion* and the close rivalry of *-age* and *-ment*. Cluster 3 contains 76.1% of *-ion* items, whereas Clusters 1 and 2 include, respectively, 44.1% of *-age* and 38.1% of *-ment*, and 32.9% of *-age* and 48.1% of *-ment* —all rates being weighted by the proportion of *-age*, *-ion* and *-ment* verb-noun pairs in the sample. This result confirms that *-age* and *-ment* compete more with each other than they both do with *-ion*. More marginally, *-ion* seems to be more similar to *-ment* than to *-age*, since Cluster 3 contains more of the former than of the latter, which is in line with the prediction analysis reported in Table 4. These observations support the existence of degrees of rivalry between pairs of suffixes, *-age* and *-ment* being by far the closest rivals, followed by *-ion* and *-ment*, and then by *-age* and *-ion*.

5.2 Dimensional reduction

We used a dimensionality reduction method to visualize neighborhood relationships between items in our dataset and evaluate the degree of similarity between the three suffixes based on the encoded semantic properties. The method we used is the t-distributed stochastic neighbor embedding (t-SNE), which is based on a

14 The cluster analysis was performed using the cluster package in R (Maechler et al., 2021).

Tab. 5: Distribution of deverbal neologisms in clusters based on 17 properties. Percentages indicate the proportion of items per cluster for each suffix (in columns). The most abundant items in each cluster are indicated in bold.

Cluster	*-age*	*-ion*	*-ment*
1	**127** (78.4%)	54 (32.2%)	116 (67.8%)
2	22 (13.6%)	13 (7.7%)	**34** (19.9%)
3	13 (8.0%)	**101** (60.1%)	21 (12.3%)

probabilistic interpretation of similarities between objects and aims to preserve neighborhood relationships between data points in a high-dimensional space, when reducing it to a two- or three-dimensional space (Maaten & Hinton, 2008). Two objects with a high probability of being neighbors in the high-dimensional space are expected to also have a high probability of being neighbors in the reduced dimensional space. We applied the t-SNE algorithm to our dataset,[15] and then mapped the three suffixes onto the resulting two-dimensional t-SNE plot, in order to examine the correspondence between semantically consistent groups of items and suffix distribution. The result of this operation is presented in Figure 9.

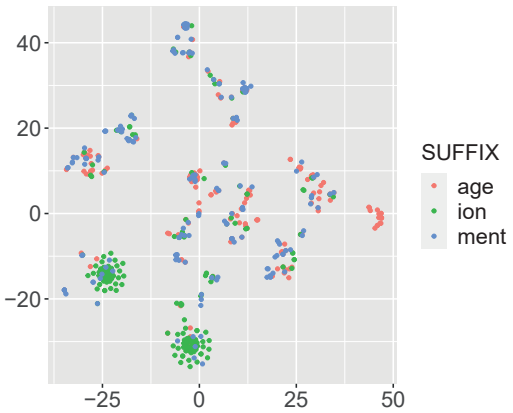

Fig. 9: t-SNE reduction of the dataset (501 deverbal neologisms analyzed for 17 properties) with suffix mapping. Each dot represents a neologism.

Two large and dense clusters including a vast majority of *-ion* items can be observed at the bottom left of the figure, showing the distinctiveness of *-ion*. A

15 The t-SNE analysis was performed using the Rtsne package in R (Krijthe, 2015).

detailed analysis of these clusters reveals that they comprise nouns with variable telicity, denoting events with a stative facet, and derived from verbs that allow post-phase reading. The base verbs are transitive verbs with causative subjects in one cluster, and intransitive verbs with patient subjects in the other. The two clusters thus distinguish nominalizations in *-ion* derived from causative and anticausative constructions. The rest of the data is more scattered in the two-dimensional space and less homogeneous with respect to suffix distribution. Although *-ion* is not absent, *-age* and *-ment* are the most represented suffixes in these areas. The most indistinguishable items are nouns with non-variable telicity that denote events, possibly with a cognitive facet, and to a lesser extent artefacts, animates, cognitive objects, natural objects, phenomena, and properties. Small isolated groups can be identified, such as the ones in ([25,35],[0,10])[16] and in ([-35,-25],[10,15]), which are equally populated with *-age* and *-ment* items, corroborating the semantic proximity between the two suffixes. Two groups exclusively or almost exclusively include one suffix. A cluster of *-age* can be observed in ([40,45],[-5,5]), and a cluster of *-ment* in ([-10,0],[35,45]). When mapped onto nominal ontological types, these two groups appear to be constituted of domain-denoting and state-denoting nouns, respectively. They reveal marginal distinctive uses of each suffix.

Overall, the analysis of the dimensionally reduced dataset confirms that *-age*, *-ion* and *-ment* cannot be fully distinguished based on semantic properties of bases and derivatives. It also suggests that distinctive semantic functions can be identified for rival suffixes, but with highly variable frequencies in terms of lexical realization, which results in different degrees of dissimilarity, and conversely of rivalry, between suffixes.

6 Conclusion

In this paper, we have investigated the rivalry between the suffixes *-age*, *-ion* and *-ment* in the formation of deverbal nouns in French. We focused on the semantic aspects of the rivalry in a quantitative approach and, to control for lexicalization effects on derivational semantics, examined a sample of neologisms ending with the three suffixes.

Some suffix preferences can be observed with respect to semantic properties of base verbs and derived nouns. Different discriminative properties combine with

16 The numbers between commas in square brackets denote intervals, whereas the numbers between commas in parentheses denote x and y coordinates.

a variable influence on suffix selection. The most important are the semantic ontological type of the noun, the ability to have a result state interpretation for both verbs and nouns, the nominal features of durativity and telicity, and the semantic role assigned to base verb subjects and obliques. Our results support some existing claims in the literature, but also demonstrate the importance of previously unconsidered properties. In addition, the quantitative approach allowed us to evaluate discriminative properties as tendencies rather than as categorical distinctions and to estimate their respective effect on suffix selection.

The main differences observed between the three suffixes are the following. The suffix *-age* is the most oriented towards the expression of dynamicity. Nouns in *-age* mostly refer to strictly dynamic eventualities, with a distinctive ability to denote domains of activities. They tend to block result state interpretation, and *-age* is able to modify the lexical aspect of base verbs with respect to post-phase and durativity. It also selects more transitive and agentive verbs than the two other suffixes. Neologisms ending in *-ion* have a preference for the denotation of gradable changes of state. They combine eventive and stative facets, have variable telicity and possible post-phase reading. They correspond to causative and patientive verbal constructions, selecting verbs that allow for the causative-anticausative alternation. The predilection of *-ion* for change-of-state denotation is related morphologically to its affinity for verbs ending in *-iser* and to the productivity of *-isation* in contemporary French. The suffix *-ment* is the one with the least salient distinctive properties and consequently appears as the least predictable of the three suffixes. It is nevertheless characterized by the formation of more state-denoting nouns than *-age* and *-ion* and by the selection of less dynamic verbs. In addition, *-ment* selects verbs with more stimulus and theme subjects than the two other suffixes and derives more achievement nouns.

Despite these preferences, the three suffixes are not always distinguishable semantically. Considering semantic properties of bases and derivatives, it is often difficult to predict which suffix will be selected. In a number of cases, preferences are neutralized, i.e. the same kinds of verbs and nouns are involved in the derivational process, and the suffixes seem to be interchangeable. However, the neutralization capacity is not the same for all pairs of suffixes and various degrees of rivalry can be identified. In particular, our observations indicate that *-age* and *-ment* compete more with each other than they both do with *-ion*.

The fact that the properties we examined do not allow a complete differentiation of *-age*, *-ion* and *-ment* leads to various hypotheses. Other factors than those considered here could play a role in the resolution of the rivalry. Additional semantic properties could be investigated, such as the type of polysemy associated with derivational patterns, with respect to ambiguity configuration and meaning specifications. Non-semantic (e.g. phonological, morphological, stylistic, diachronic)

characteristics can also be considered (Uth, 2010; Missud & Villoing, 2020). An investigation of the combined effects of these factors is necessary to provide a full account of the rivalry between the three suffixes. Nevertheless, it can also be hypothesized that affix rivalry is not ruled by fine-grained discrete discrimination, and that competing affixes have an inherent element of indistinction. Productive rival suffixes such as *-age*, *-ion* and *-ment* in French may overlap in distribution. Their coexistence in the morphological system is accompanied by semantic preferences and could ultimately favor the creation of niches (Lindsay & Aronoff, 2013; Aronoff, 2016). Due to the limited size of the sample of derivatives studied here, the preferences we observed for *-age*, *-ion*, *-ment* may not be exhaustive. Other niches than those addressed in the study could appear by taking into account larger sets of nouns (e.g. sound denotation in the case of *-ment* derivatives, such as *bruissement* 'rustling', *chuintement* 'hissing', *couinement* 'squealing', *craquement* 'crack', *crépitement* 'crackling', *crissement* 'grating', *grincement* 'grinding').

Acknowledgments

This research was supported by the Swiss National Science Foundation under grant 10001F_188782 ('The semantics of deverbal nouns in French'). We are grateful to the two anonymous reviewers and to the editors for their helpful comments on an earlier version of this paper. All remaining errors are our own.

Supplementary material

The annotation guidelines, dataset, statistical scripts and descriptive results used in this study are available online at the following URL: https://github.com/neologisms-annotation/deverbal-nouns-rivalry

Bibliography

Abusch, Dorit. 1986. *Verbs of change, causation and time*. Stanford: CSLI report.

Apothéloz, Denis. 2008. *Entrer quelques instants* vs *arriver quelques instants*: Le problème de la spécification de la durée de l'état résultant. *Verbum* 30. 199–219.

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

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

Aronoff, Mark. 2016. Competition and the lexicon. In Annibale Elia, Claudio Iacobini & Miriam Voghera (eds.), *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII Congresso Internazionale di Studi della Società di Linguistica Italiana (Fasciano, Salerno 26-28 settembre 2013)*, 39–51. Roma: Bulzoni Editore.

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

Baayen, R Harald, Anna Endresen, Laura A Janda, Anastasia Makarova & Tore Nesset. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37(3). 253–291.

Bertinetto, Pier Marco & Mario Squartini. 1995. An attempt at defining the class of gradual completion verbs. In Pier Marco Bertinetto, Valentina Bianchi, James Higginbotham & Mario Squartini (eds.), *Temporal reference aspect and actionality 1: Semantics and syntactic perspectives*, 11–26. Torino: Rosenberg and Sellier.

Bonami, Olivier & Juliette Thuilier. 2019. A statistical approach to rivalry in lexeme formation. French *-iser* and *-ifier*. *Word structure* 12(1). 4–41.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45(1). 5–32.

Byrt, Ted, Janet Bishop & John B Carlin. 1993. Bias, prevalence and kappa. *Journal of Clinical Epidemiology* 46(5). 423–429.

Cartier, Emmanuel. 2018. *De l'extraction d'information à la représentation sémantique*. Paris: Université Paris 13 Habilitation dissertation.

Copestake, Ann & Ted Briscoe. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics* 12(1). 15–67.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer Verlag.

Cruse, D. Alan. 1995. Polysemy and related phenomena from a cognitive linguistic viewpoint. In Patrick St Dizier & Evelyne Viegas (eds.), *Computational lexical semantics*, 33–49. Cambridge: Cambridge University Press.

Dal, Georgette, Nabil Hathout, Stéphanie Lignon, Fiammetta Namer & Ludovic Tanguy. 2018. Toile *versus* dictionnaires: Les nominalisations du français en *-age* et en *-ment*. In Franck Neveu, Bernard Harmegnies, Linda Hriba & Sophie Prévost (eds.), *Congrès Mondial de Linguistique Française - CMLF 2018*, Paris: Institut de Linguistique Française. https://doi.org/10.1051/shsconf/20184608003.

Dal, Georgette & Fiammetta Namer. 2015. La fréquence en morphologie: Pour quels usages? *Langages* 197(1). 47–68.

Debaty-Luca, Thierry. 1986. *Théorie fonctionnelle de la suffixation (appliquée principalement au français et au wallon du centre)*. Genève: Librairie Droz.

Debeer, Dries & Carolin Strobl. 2020. Conditional permutation importance revisited. *BMC Bioinformatics* 21(1). 1–30.

Dowty, David R. 1979. *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht: D. Reidel Publishing Company.

Dressler, Wolfgang U, Lavinia Merlini Barbaresi, Sonja Schwaiger, Jutta Ransmayr, Sabine Sommer-Lolei & Katharina Korecky-Kröll. 2019. Rivalry and lack of blocking among italian and german diminutives in adult and child language. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 123–143. Cham: Springer.

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

Dubois, Jean & Françoise Dubois-Charlier. 1997. *Les verbes français*. Paris: Larousse.

Dölling, Johannes. 2020. Systematic polysemy. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann & Thomas Ede Zimmermann (eds.), *The Wiley Blackwell companion to semantics*, 1–27. Hoboken, NJ: John Wiley & Sons.

Ferret, Karen & Florence Villoing. 2015. French N-*age* instrumentals: Semantic properties of the base verb. *Morphology* 25. 473–496.

Filip, Hana. 2012. Lexical aspect. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 721–751. Oxford: Oxford University Press.

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

Fleischman, Suzanne. 1980. *The French suffix -age: Its genesis, internal growth, and diffusion*. Ann Arbor, MI: University Microfilms International.

Fradin, Bernard. 2011. Remarks on state-denoting nominalizations. *Recherches Linguistiques de Vincennes* 40. 73–99.

Fradin, Bernard. 2016. L'interprétation des nominalisations en N-*age* et N-*ment* en français. In Franz Rainer, Michela Russo & Fernando Sánchez Miret (eds.), *Actes du XXVIIe Congrès International de Linguistique et Philologie Romanes*, 53–66. Nancy: ATILF.

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

Gardani, Francesco, Franz Rainer & Hans Christian Luschützky. 2019. Competition in morphology: A historical outline. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 3–36. Cham: Springer.

Godard, Danièle & Jacques Jayez. 1996. Types nominaux et anaphores: Le cas des objets et des événements. *Cahiers Chronos* 1. 41–58.

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

Haas, Pauline, Richard Huyghe & Rafael Marín. 2008. Du verbe au nom: Calques et décalages aspectuels. In Jacques Durand, Benoît Habert & Bernard Laks (eds.), *Congrès Mondial de Linguistique Française - CMLF 2008*, 2051–2065. Paris: Institut de Linguistique Française.

Haas, Pauline & Anne Jugnet. 2013. De l'existence des prédicats d'achèvements. *Lingvisticæ Investigationes* 36(1). 56–89.

Hathout, Nabil & Fiammetta Namer. 2014. Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology* 11(5). 208–219.

Hay, Jennifer, Christopher Kennedy & Beth Levin. 1999. Scalar structure underlies telicity in 'degree achievements'. In Tanya Matthews & Devon Strolovitch (eds.), *Proceedings of SALT 9*, 127–144. Ithaca, NY: CLC Publications.

Hothorn, Torsten, Peter Buehlmann, Sandrine Dudoit, Annette Molinaro & Mark Van Der Laan. 2006. Survival ensembles. *Biostatistics* 7(3). 355–373.

Huyghe, Richard. 2015. Les typologies nominales: présentation. *Langue Française* (1). 5–27.

Huyghe, Richard. 2021. Building a lexical database to investigate the semantics of French deverbal nouns. *Resources and Tools for Derivational Morphology (DeriMo 2021)* 21.

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

Kawaletz, Lea & Ingo Plag. 2015. Predicting the semantics of English nominalizations: A frame-based analysis of *-ment* suffixation. In Laurie Bauer, Lívia Körtvélyessy & Pavol Štekauer (eds.), *Semantics of complex words*, 289–319. Berlin: Springer.

Kelling, Carmen. 2001. Agentivity and suffix selection. In Miriam Butt & Tracy H. King (eds.), *Proceedings LFG 2001*, 147–162. Stanford, CA: CSLI Publications.

Kipper-Schuler, Karin. 2005. *Verbnet: A broad-coverage, comprehensive verb lexicon*. Philadelphia, PA: University of Pennsylvania dissertation.

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

Krijthe, Jesse H. 2015. *Rtsne: T-distributed stochastic neighbor embedding using barnes-hut implementation*. https://github.com/jkrijthe/Rtsne. R package version 0.15.

Landis, J Richard & Gary G Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174.

Levshina, Natalia. 2020. Conditional inference trees and random forests. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 611–643. Berlin: Springer.

Lignon, Stéphanie, Fiammetta Namer & Florence Villoing. 2014. De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer & Sophie Prévost (eds.), *Congrès Mondial de Linguistique Française - CMLF 2014*, 1813–1835. Paris: Institut de Linguistique Française.

Lindsay, Mark & Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Nabil Hathout, Fabio Montermini & Jesse Tseng (eds.), *Morphology in Toulouse: Selected proceedings of Décembrettes 7*, 133–153. München: LINCOM Europe.

Lüdtke, Jens. 1978. *Prädikative Nominalisierungen mit Suffixen im Katalanischen, Spanischen und Französischen*. Tübingen: Niemeyer.

Maaten, Laurens van der & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9(11). 2579–2605.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert & Kurt Hornik. 2021. *cluster: Cluster analysis basics and extensions*. https://CRAN.R-project.org/package=cluster. R package version 2.1.2.

Martin, Fabienne. 2010. The semantics of eventive suffixes in French. In Monika Rathert & Artemis Alexiadou (eds.), *The semantics of nominalizations across languages and frameworks*, 109–141. Berlin: De Gruyter.

McHugh, Mary L. 2012. Interrater reliability: The kappa statistic. *Biochemia Medica* 22(3). 276–282.

Missud, Alice & Florence Villoing. 2020. The morphology of rival *-ion*, *-age* and *-ment* selected verbal bases. *Lexique* 26. 29–52.

Naccarato, Chiara. 2019. Agentive (para)synthetic compounds in Russian: A quantitative study of rival constructions. *Morphology* 29(1). 1–30.

New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. *Lexique 2*: A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36(3). 516–524.

Petukhova, Volhaand & Harry Bunt. 2008. LIRICS semantic role annotation: Design and evaluation of a set of data categories. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard,

Joseph Mariani, Jan Odijk, Stelios Piperidis & Daniel Tapias (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, 39–45. Marrakech, Morocco: European Language Resources Association (ELRA).

Piñón, Christopher. 1997. Achievements in an event semantics. In Aaron Lawson & Eun Cho (eds.), *Proceedings of SALT 7*, 273–296. Ithaca, NY: CLC Publications.

Piñón, Christopher. 1999. Durative adverbials for result states. In Sonya Bird, Andrew Carnie, Jason D. Haugen & Peter Norquest (eds.), *WCCFL18: Proceedings of the 18th West Coast Conference on Formal Linguistics*, 420–433. Somerville, MA: Cascadilla Press.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin: De Gruyter.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.

Rothstein, Susan. 2004. *Structuring events: A study in the semantics of lexical aspect*. Oxford: Blackwell Publishing.

Rothstein, Susan. 2008. Two puzzles for a theory of lexical aspect: Semelfactives and degree achievements. In Johannes Dölling, Tatjana Heyde-Zybatow & Martin Shäfer (eds.), *Event structures in linguistic form and interpretation*, 175–198. Berlin: De Gruyter.

Sagot, Benoît. 2010. The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*, Valletta, Malta: European Language Resources Association (ELRA).

Schulte, Marion. 2015. Polysemy and synonymy in derivational affixation: A case study of the English suffixes *-age* and *-ery*. *Morphology* 25(4). 371–390.

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

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

Strobl, Carolin, Anne-Laure Boulesteix, Thomas Kneib, Thomas Augustin & Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9(1). 1–11.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8(25). 1–21.

Tagliamonte, Sali A & R Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–78.

Uth, Melanie. 2010. The rivalry of French *-ment* and *-age* from a diachronic perspective. In Monika Rathert & Artemis Alexiadou (eds.), *The semantics of nominalizations across languages and frameworks*, 215–244. Berlin: De Gruyter.

Varvara, Rossella. 2020. Constraints on nominalizations: Investigating the productivity domain of Italian *-mento* and *-zione*. *Zeitschrift für Wortbildung/Journal of Word Formation* 4(2). 78–99.

Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Wauquier, Marine. 2020. *Confrontation des procédés dérivationnels et des catégories sémantiques dans les modèles distributionnels*. Toulouse: Université Toulouse Jean Jaurès dissertation.

Zwanenburg, Wiecher. 1984. Word formation and meaning. *Quaderni di Semantica* 5(1). 130–142.