What French eventive nominalizations without verbal bases tell us about the salience of paradigmatic networks

Alice Missud & Florence Villoing
Université Paris Nanterre - MoDyCo (CNRS)

1 Introduction

French nominalizations in *-ion*, *-age* and *-ment* can derive from constructed verbs(1).

(1) a. *créole*$_N$ 'Creole' → *créoliser*$_V$ 'to creolize' → *créolisation*$_N$ 'creolization'
b. *plan*$_N$ 'plan' → *planifier*$_V$ 'to plan' → *planification*$_N$ 'planning'
c. *jour*$_N$ 'gap/chink' → *ajourer*$_V$ 'to perforate' → *ajouration*$_N$ 'perforation'
d. *cadre*$_N$ 'frame' → *encadrer*$_V$ 'to frame' → *encadrement*$_N$ 'framing'
e. *goutte*$_N$ 'drop' → *égoutter*$_V$ 'to drain' → *égouttage*$_N$ 'draining'
f. *frein*$_N$ 'brake' → *freiner*$_V$ 'to brake' → *freinage*$_N$ 'braking'

In such cases, the successive derivation schemes constitute a derivational family of triplets (thus constituting a derivational paradigm in the sense of Bauer 1997 and Štekauer 2014) (2) that consists of a verb constructed on a nominal or adjectival base either by affixation (suffixation in *-iser* (1a), in *-ifier* (1b), prefixation in *a-* (1c), *en-* (1d), *é-* (1e), or by conversion (1f)), and of a resulting nominalization in *-ion*, *-age* or *-ment*. These successive formations do not constitute a case of double suffixation as previously discussed in cases of interfixation or affixal offset (Plénat 2005, Roché 2009) since each derivative carries a specific semantic value. They are not relevant to the bracketing paradox (Pesetsky 1985, Sproat 1992, Spencer 1988, Harley 2010) either, because the interpretations of the derivatives are in line with the morphological schemas they derive from. They also do not constitute a case of parasynthetic formation because missing verbs often eventually appear after the derived nominalization with the expected semantic interpretation.

(2) N / Adj → V → N-*ion* / N-*age* / N-*ment*

Although these formations are not a majority, they still represent a substantial proportion of derivatives: 26.4% of the nouns in *-ion*, *-age* and *-ment* in VerNom (Missud et al. 2020) are derived from a constructed verbal base (as assessed in Missud & Villoing 2020). Analyzing the data revealed a certain proportion of coinages in *-ion*, *-age* and *-ment* for which the constructed verbal base (denominal or deadjectival) does not appear in corpora or occurs at a very low frequency (3).

(3) a. *-ion*: *macbeth*$_N$ 'Macbeth' → *macbethisation*$_N$ 'macbethization' (°macbethiser$_V$ 'to macbethize'), *translucide*$_A$ 'translucent' → *translucidation*$_N$ 'translucentation' (*translucider*$_V$ 'to make translucent'), *disneyland*$_N$ 'Disneyland' → *disneylandification*$_N$ 'disneylandification' (*disneylandifier*$_V$ 'to disneylandify')
b. *-age*: *bandelette*$_N$ 'strip' → *bandelettage*$_N$ 'strip application' (*bandeletter*$_V$ 'to apply strips'), *bestof*$_N$ 'best-of' → *bestofage*$_N$ 'making a best-of' (°*bestofer*$_V$ 'to make a best-of'), *dofus*$_N$ 'Dofus (video game)' → *dofusage*$_N$ 'playing Dofus' (°*dofuser*$_V$ 'to play Dofus')
c. *-ment*: *stupide*$_A$ 'stupid' → *enstupidement*$_N$ 'making stupid' (°*enstupider*$_V$ 'to make stupid'), *bruyère*$_N$ 'heather' → *embruyèrement*$_N$ 'proliferation of heather' (*embruyérer*$_V$ 'to proliferate heather')

These rare or nonexistent verbs constitute the focus of our study; we hypothesize that such nominalizations are formed directly on the base noun or adjective as the intermediate derivation

in the family of triplets (verb formation) is not lexicalized. The successive derivations of triplets in (2) becomes (4).

(4) N / Adj (→ °V) → N-*ion* / N-*age* / N-*ment*

The degree symbol (°) indicates that the verb is possible but not attested in corpora, as previously used by French morphologists following D. Corbin. We analyze them as potential words: as widely discussed in morphology, they are the conceptual (and not actual) result of a productive rule (Booij 1977, Rainer 2012 for an extensive state-of-the-art). They are words that have not been lexicalized, or even attested, but nonetheless seem perfectly acceptable as they meet all the criteria that would make the rule derive them, while not interfering with an already existing form with the same meaning. Although such words have been identified (but not explained) by Roché (2007) and Lignon et al. (2014) in the case of *-ion* nominalizations, our data show that it also concerns *-age* and *-ment* suffixations.

Thus, despite its absence or low frequency, the verb is perfectly identifiable and can be easily interpreted. When looking at the corresponding *-ion*, *-age* or *-ment* nominalization, it is identifiable in i) its form as the nominalization reveals the phonological form of the verbalizing schema although the verb did not appear (3a), as well as in ii) its semantics - specifically in cases of conversion which do not show any affix (3b).

The purpose of our research is to identify the conditions that allow such coinages. We will show that a fundamental condition for these types of formations is that they correspond to a derivational network – a stable and identifiable formal and semantic relation between two or multiple lexemes, in the sense of Hathout (2011), Roché (2017), Bonami & Strnadová (2019), Fradin (2020), that is very salient in the French lexicon (i.e. occurring more frequently that expected in our data). The very high salience of such derivational network implies that some member of a triplet does not necessarily have to be coined, or can be coined with a very low frequency as previously shown (regarding other data) in Villoing & Namer (2012), Roché (2017). However, if the triangular relation in the paradigmatic network is not salient or frequent enough, then the direct derivation from noun/adjective to N-*ion*, N-*age* or N-*ment* is not possible and the verbalization step is necessary.

2 Collecting the data

The data we used were first extracted from frCOW (Schäfer & Bildhauer 2012, Schäfer 2015), a massive French web corpus consisting of 1.9 billion words. We collected every word tagged as a noun ending in either *-ion*, *-age* or *-ment*, as well as all nouns and adjectives. As we were looking for N-*ion*, N-*age* and N-*ment* for which no corresponding verb could be found, we first deleted all nouns that were already included in VerNom (Missud et al. 2020), a lexical database automatically constructed from frCOW consisting of verb-noun pairs and covering *-ion*, *-age* and *-ment* suffixations. We kept N-*ion*, N-*age*, N-*ment*, other nouns and adjectives that were not lemmatized in the distributed version of frCOW, hoping that we would find neologisms as we hypothesized that newly-coined derivatives would be more likely to lack a corresponding verb than lexicalized ones. The remaining nouns in *-ion*, *-age* and *-ment* were then matched automatically with the other nouns and adjectives using regular expressions. As for morphological matching, if the potential suffixed nouns differed from the base nouns and adjectives on the formal level, they were expected to exclusively show signs of an *-iser* or *-ifier* suffixation or an *en-*, *é-* or *a-* prefixation[1]. Otherwise, in the case of verb to noun conversion, only the last syllable of the base noun could differ from the stem of the suffixed noun. In view of the large number of

[1] *dé-* prefixation was not taken into account as the confusion between deverbal *dé-* (as in *boutonner* 'to button' → *déboutonner* 'to unbutton') and denominal *dé-* (as in *os* 'bone' -> *désosser* 'to bone') required a time-consuming step of manual annotation.

	Description of the N-N pairs	Number of items	Number of selected triplets
Sample 1	30% of the pairs containing a noun in *-ion* that shows signs of an *-iser* or *-ifier* suffixation, a conversion, or an *en-*, *é-* or *a-* prefixation	918	186
Sample 2	30% of the pairs containing a noun in *-age* that shows signs of an *-iser* or *-ifier* suffixation, a conversion, or an *en-*, *é-* or *a-* prefixation	593	80
Sample 3	30% of the pairs containing a noun in *-ment* that shows signs of an *-iser* or *-ifier* suffixation, a conversion, or an *en-*, *é-* or *a-* prefixation	667	8

Table 1: Table 1: Extracted samples and selected triplets

pairs that were collected, we created 3 random samples that consisted of 30% of the pairs we found for each of the nominalization schemas. The details are shown in Table 1.

For all the pairs in the samples above, as frCOW does not provide dates, we looked for the suffixed noun using the Google search engine (strictly, using quotation marks) in order to find their earliest date of attestation if the base noun or adjective was semantically related. Although the method relies heavily on the web pages that Google allows users to access, this appeared to be the most convenient way to look for attestation dates since it is still able to capture many book release dates, news articles, forums and dictionaries. Some nouns that did not exist in Google's database were also searched on Twitter.

As we were looking for coinages, suffixed nouns that got more than 10 pages of results and that appeared before 1950 were ignored. Otherwise, we looked for a corresponding verb (inflected or not). In some cases, we also looked for spelling variants (for example: *mic**k**aeljacksoniser* instead of *mic**h**aeljacksoniser*). If the verb's earliest attestation date was later than the noun's according to the search results, or if no verb could be found at all, we considered it as an unattested verb and collected the base noun, the unattested verb and the suffixed noun as triplets[2].

3 Results

The concrete evidence that these nominalizations lacking a corresponding attested verb belong to salient paradigmatic networks in the French lexicon is reflected by the fact that the unattested verbs that the nominalizations infer correspond to the schemas that *-ion*, *-age* and *-ment* prefer. As previously shown in Missud & Villoing (2020), *-ion*, *-age* and *-ment* all display salient distinct preferences when it comes to constructed base selection. *-ion* is by far the most specialized as it strongly favors *-iser* verbs and is the one that selects *-ifier* verbs the most. It can also select a great proportion of converted verbs. *-age* is less categorical although it strongly favors converted verbs, as well as a significant proportion of the *é-* verbs of the data. While *-ment* is the one that has the least salient preferences, it still shows a preference for *en-* verbs and converted verbs (although converted verbs are mostly selected by *-age*). The proportions in Table 2 show the preferences of *-ion*, *-age* and *-ment* when a verb is lacking (or attested later than the nominalization) in the paradigm by dividing the number of occurrences of each cell by the total number of occurrences of the data.

As shown in Table 2, nominalizations without verbs in *-ion* suppose the existence of an *-iser* derived verb, as in the general case. The same applies to nouns in *-age* with no verb that suppose a converted verb, and nouns in *-ment* that suppose a denominal verb in *en-*. What is striking is that the more salient the preference is in the general case, the more it shows in cases where

[2]Note that the number of selected triplets containing a noun in *-ment* is extremely low compared to the number of items initially collected: this is due to the fact that *-ment* coinages are rare, and most *-ment* nouns in the sample were misspelled or taken from various Ancient and Middle French dictionaries on the web, which calls into question the actual productivity of the suffix (as assessed in Missud et al. 2020).

	-iser	*-ifier*	*conversion*	*a-*	*en-*	*é-*
-ion	**0.51**	0.08	0.08	-	-	-
-age	-	-	**0.27**	-	0.01	-
-ment	-	-	<0.01	-	**0.02**	-

Table 2: Proportions of infrequent verbs according to their construction for each verbalization schema (division of each cell by the total)

no verb is attested. *-ion* is the only suffixation that can derive nouns in *-is(ation)* (51% of all data) and *-ifi(cation)* (8%), *-ment* suffixation constructs *en*-X-*ment* nouns in most cases; only one nominalization in *-ment* implying a converted verb was identified (*zizi*, *ziziement* → °zizier'), and the semantic relation between the members of this triplet is unclear (*zizi* can refer to an onomatopoeia or a willy). *-ion* derives a similar proportion of *-ification* nouns and nouns linked to an unattested converted verb (8% in both cases), which reflects a tendency for converted verbs that was already identified in the general case and that might have been extrapolated because of the higher proportion of unattested converted verbs that have been collected. Although *é-* prefixed verbs were mostly represented amongst *-age* triplets in the general case, *-age* preference for *en-* is reflected here as *en*-X-*age* represents 2% of the data. The strongest preference for converted verbs is apparent as 27% of the data consist of *-age* nouns implying a converted verb.

The proportion of constructed verbs that the nominalizations select is a good indicator of the salience of the derivational network, and subsequently of the potential nominalization that lacks a corresponding verbal base. Inversely, nominalizations without a verb correctly predict the preferred verbal bases that the nominalizations select in the general case. As a result, it could be that the salience of such preferences also partially ensures the productivity of *-ion*, *-age* and *-ment*: since the paradigmatic network is productive (e.g. N/A → V-*iser* → N-*ion*, or N/A → converted verb → N-*age*), shortcuts are allowed and nominalizations no longer need to wait for the availability of a verb that corresponds to their preferences to be constructed; they can directly select a noun or an adjective that fits morphologically and semantically.

References

Bauer, Laurie. 1997. Derivational paradigms. In *Yearbook of morphology 1996*, 243–256. Springer.

Bonami, Olivier & Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2). 167–197.

Booij, Geert Evert. 1977. *Dutch morphology: A study of word formation in generative grammar*. Dordrecht: Foris.

Fradin, Bernard. 2020. Characterizing derivational paradigms. In J. Fernández-Domínguez, A. Bagasheva & C. Lara-Clares (eds.), *Empirical approaches to linguistic theory*, vol. 16, 49–84. Leiden: Koninklijke Brill.

Harley, Heidi. 2010. Affixation and the mirror principle. *Interfaces in linguistics* 166. 186.

Hathout, Nabil. 2011. Une approche topologique de la construction des mots: propositions théoriques et application à la préfixation en anti. *Des unités morphologiques au lexique* 251–318.

Lignon, Stéphanie, Fiammetta Namer & Florence Villoing. 2014. De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In *Shs web of conferences*, vol. 8, 1813–1835. EDP Sciences.

Missud, Alice, Pascal Amsili & Florence Villoing. 2020. Vernom: une base de paires morphologiques acquise sur très gros corpus (vernom: a french derivational database acquired on a massive corpus). In *Actes de la 6e conférence conjointe journées d'études sur la parole (jep, 33e édition), traitement automatique des langues naturelles (taln, 27e édition), rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (récital, 22e édition). volume 2: Traitement automatique des langues naturelles*, 305–313.

Missud, Alice & Florence Villoing. 2020. The morphology of rival –ion, –age and –ment selected verbal bases. In Dany Amiot & Delphine Tribout (eds.), *Lexique*, vol. 26, 29–52. Presses Universitaires de Lille.

Pesetsky, David. 1985. Morphology and logical form. *Linguistic inquiry* 16(2). 193–246.

Plénat, Marc. 2005. Rosinette, cousinette, putinette, starlinette, chipinette: décalage, infixation et épenthèse devant–ette. In Choï Jonin I., Bras M., Dagnac A. & Rouquier M. (eds.), *Questions de classification en linguistique : méthodes et descriptions. mélanges offerts au professeur christian molinier*, 275–298. Berne, Peter Lang.

Rainer, Franz. 2012. Morphological metaphysics: virtual, potential, and actual words. *Word Structure* 5(2). 165–182.

Roché, Michel. 2007. Logique lexicale et morphologie: la dérivation en-isme. *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse* 45–58.

Roché, Michel. 2009. Un ou deux suffixes? une ou deux suffixations? In B. Fradin, F. Kerleroux & M. Plénat. (eds.), *Aperçus de morphologie du français*, 143–173. Presses Universitaires de Vincennes.

Roché, Michel. 2017. Les familles dérivationnelles : comment ça marche ?, Toulouse : Université Toulouse 2 Jean Jaurès.

Schäfer, Roland. 2015. Processing and querying large web corpora with the COW14 architecture. In Piotr Banski, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Längen & Andreas Witt (eds.), *Proceedings of challenges in the management of large corpora 3 (cmlc-3)*, UCREL Lancaster: IDS. `http://rolandschaefer.net/?p=749`.

Schäfer, Roland & Felix Bildhauer. 2012. Building large corpora from the web using a new efficient tool chain. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet UÄŸur DoÄŸan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, 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). `http://rolandschaefer.net/?p=70`.

Spencer, Andrew. 1988. Bracketing paradoxes and the english lexicon. *Language* 663–682.

Sproat, Richard. 1992. Unhappier is not a" bracketing paradox". *Linguistic Inquiry* 23(2). 347–352.

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

Villoing, Florence & Fiammetta Namer. 2012. Composition néoclassique en-logue et en-logiste: Les noms en-logue sont-ils encore des noms de spécialistes? *La composition néoclassique, Verbum, Nancy: Presses Universitaires de Nancy. DOI* 10. 213–231.