



HAL
open science

Nominalisations sans base verbale suffixées en -ion , -age et -ment du français : conditions morphologiques

Alice Missud, Florence Villoing

► To cite this version:

Alice Missud, Florence Villoing. Nominalisations sans base verbale suffixées en -ion , -age et -ment du français : conditions morphologiques. 8e Congrès Mondial de Linguistique Française (CMLF 2022), Jul 2022, Orléans, France. pp.07002, 10.1051/shsconf/202213807002 . hal-04106392

HAL Id: hal-04106392

<https://hal.science/hal-04106392v1>

Submitted on 25 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Nominalisations sans base verbale suffixées en *-ion*, *-age* et *-ment* du français : conditions morphologiques

Alice Missud¹, Florence Villoing^{2*}

¹ MoDyCo, UMR 7114 (CNRS & Université Paris Nanterre) & LaTTiCe, UMR 8094 (CNRS, Paris Sorbonne Nouvelle & ENS), France.

² MoDyCo, UMR 7114 (CNRS & Université Paris Nanterre), France.

Résumé. Cet article se propose d'étudier les conditions de formation des nominalisations en *-age*, *-ment* et *-ion* du français qui semblent se former directement sur des bases nominales ou adjectivales, le verbe intermédiaire n'étant pas attesté ou attesté plus tard. L'étude se donne pour double objectif de montrer (i) que les nominalisations du français sans base verbale ne concernent pas seulement la suffixation en *-ion* déjà identifiée dans la littérature, mais aussi les suffixations en *-age* et en *-ment* ; (ii) que ces nominalisations répondent à des conditions particulières : elles ne relèvent que de certaines familles morphologiques de triplets $X_{N/Adj} - X_V - X_V\text{-ion/-age/-ment}_N$, celles qui se distinguent par une haute fréquence d'apparition en corpus. La recherche est menée sur un corpus de très grande taille (frCOW), permettant de constituer une base de données suffisamment importante pour utiliser des mesures quantitatives significatives.

Abstract. Nominalisations without verbal base suffixed with *-ion*, *-age* and *-ment* in French: morphological conditions. This article aims to study the conditions that allow the formation of French nominalizations suffixed with *-age*, *-ment* and *-ion* which seem to be formed directly on nominal bases. The study has two objectives: (i) to show that French nominalizations without a verbal base do not only concern the *-ion* suffixations already identified in the literature, but also the *-age* and *-ment* suffixations; (ii) to show that these nominalizations meet specific conditions: they must be part of morphological families of $X_{N/Adj} - X_V - X_V\text{-ion/-age/-ment}_N$ triplets, that are distinguished by a high frequency of occurrence in the corpus. The research is conducted on a very large corpus (frCOW), allowing for a database large enough to use meaningful quantitative measures.

1 Introduction

Cet article se propose d'étudier les conditions de formation des nominalisations en *-ion*, *-age*, *-ment* et du français qui semblent se former directement sur des bases nominales comme dans (1).

- (1) a. *-ion*
*macbeth*_N → *macbethisation*_N (*macbethiser*_V)¹
*translucide*_{Adj} → *translucidation*_N (*translucider*_V)

* Corresponding author : villoing@parisnanterre.fr

*disneyland*_N → *disneylandification*_N (*disneylandifier*_V)

b. -age

*bandelette*_N → *bandelettage*_N (*bandeletter*_V)

*bestof*_N → *bestofage*_N (*bestofer*_V)

*dofus*_N → *dofusage*_N (*dofuser*_V)

c. -ment

*stupide*_{Adj} → *enstupiderment*_N (*enstupider*_V)

*bruyère*_N → *embruyèrment*_N (*embruyérer*_V)

En effet, la base verbale en lien avec ces nominalisations n'apparaît pas dans le corpus étudié ou y figure avec une très faible fréquence (laissant présager, dans ce cas, que le verbe s'est formé après la nominalisation). Par exemple, la nominalisation *disneylandification*_N en *-ion* en (1a) est morphologiquement reliée au nom *disneyland*_N, alors que la base verbale attendue *disneylandifier*_V est absente du corpus de base de l'étude. Il en va de même pour les nominalisations en *-age* en (1b) ou en *-ment* en (1c). Les verbes de base, même s'ils ne sont pas attestés, sont néanmoins parfaitement reconnaissables, parce que la nominalisation révèle

- (i) la forme phonologique de l'affixe verbalisateur sous-jacent (sauf s'il s'agit d'un verbe dénominal ou déadjectival convers), et/ou
- (ii) les relations sémantiques du schéma morphologique dérivant le verbe non attesté (ou faiblement attesté).

Pour reprendre les exemples de (1), les nominalisations laissent apparaître le suffixe ou le préfixe verbalisateur attendu et leur interprétation inclut celle associée aux dérivations dénominales ou déadjectivales formant les verbes non réalisés.

Certaines de ces formations ont déjà été repérées dans la littérature, en particulier les nominalisations en *-isation* et en *-ification* (2).

- (2) a. *Burundi*_N → *Burundisation*_N (*se burundiser*_V) ;
*Belmondo*_N → *belmondisation*_N (*se belmondiser*_V) (Dal 2004 : 84) ;
 b. *contraventionnel*_N → *contraventionnalisation*_N (*contraventionnaliser*_V) ;
*brèche*_N → *bréchification*_N (*bréchifier*_V) (Roché 2009 : 154)

Ces nominalisations dérivées directement de noms ou d'adjectifs, sans verbe intermédiaire, ont principalement servi d'argument contre une morphologie à base morphème et orientée, alimenté les démonstrations en faveur d'une morphologie paradigmatique (suivant notamment les travaux de Plag (2003) et Bauer, Lieber et Plag (2013) à partir de faits similaires en anglais), et permis de développer l'hypothèse de la création d'un nouveau suffixe (*-isation*) résultant de la succession fréquente des suffixes *-iser* et *-(at)ion* (Dal 2004, Roché 2009, Namer 2009, Dal & Namer 2015, Lignon et al. 2014). Cependant, elles n'ont jamais fait l'objet d'un traitement systématique et n'ont porté que sur certaines formations réduites du français (les nominalisations en *-isation*). La disponibilité de corpus de grande ampleur permet aujourd'hui d'accéder à un grand nombre de données originales qui rend possible une étude plus vaste, systématique et quantifiée. C'est dans cette perspective que s'inscrit notre recherche, exploitant le corpus frCOW (Schäfer & Bildhauer 2012 ; Schäfer 2015) tiré de la Toile (voir § 3.1.) et la base de données lexicale VerNom (Missud et al. 2020), construite automatiquement à partir de frCOW et contenant plus de vingt-cinq mille paires verbe-nom (avec leurs fréquences) couvrant les principales suffixations nominalisatrices du français.

L'étude que nous présentons se donne un double objectif :

- (i) montrer que les nominalisations du français sans base verbale ne concernent pas seulement la suffixation en *-ion* mais aussi les suffixations en *-age* et en *-ment*,
- (ii) identifier, en s'appuyant sur des données authentiques quantifiées, les conditions morphologiques qui permettent ces nominalisations sans base verbale, afin de pouvoir prédire lesquelles sont possibles.

La section suivante (§ 2.) s'attache à expliciter la méthodologie précise qui a permis de récolter les données. La section 3. est consacrée à la mise au jour des conditions morphologiques qui permettent la formation des nominalisations sans base verbale. La dernière section (§4.) propose un bilan et dessine quelques perspectives de recherches possibles.

2. Collecte des données

Afin d'identifier les nominalisations du français en *-ion*, *-age* et *-ment* construites sur base nominale ou adjectivale (sans que le verbe intermédiaire attendu ne soit attesté, ou soit attesté plus tard que la nominalisation correspondante), et ce, dans leur plus grande diversité, nous avons dû avoir recours à une grande quantité de données (les cas n'étant pas aussi nombreux que les nominalisations déverbiales standard). Nous avons choisi de collecter les données en utilisant frCOW, actuellement le plus large et le plus récent corpus web disponible pour le français, comptabilisant 9 milliards de mots extraits du web (Schäfer & Bildhauer 2012 ; Schäfer 2015). Les données de frCOW sont annotées selon leur catégorie syntaxique (*Part Of Speech-tagged*) et lemmatisées à l'aide de l'outil d'annotation TreeTagger. Elles sont également renseignées du point de vue de leur fréquence. Nous avons ainsi collecté tous les noms, verbes et adjectifs qui y figurent avec leur fréquence. Si aucun lemme n'était attribué à une forme, nous l'avons lemmatisée nous-mêmes en additionnant la fréquence de chaque forme après avoir supprimé les marques de flexion, puis décidé d'une forme lemmatique (en *-er -ir, -re* pour les verbes et singulier pour les noms et adjectifs).

A partir de ce corpus, nous avons constitué deux bases de données inédites : la première recense les nominalisations construites sur la base de verbes eux-mêmes dérivés (formant ainsi des triplets $X^2_{N/Adj} - X_V - X_V -ion/-age/-ment_N$ dont nous montrerons l'influence dans la formation des nominalisations sans base verbale) et la deuxième liste l'ensemble des nominalisations en *-ion, -age* et *-ment* sans base verbale du français.

2.1 Constitution des familles de triplets standard $X_{N/Adj} - X_V - X_V - ion/-age/-ment_N$

La première base de données constituée à partir de l'extraction des lemmes de frCOW est constituée de familles de triplets $X_{N/Adj} - X_V - X_V -ion/-age/-ment_N$ dont les lexèmes étaient tous attestés dans le corpus. La méthode utilisée correspond à celle décrite dans Missud et al. (2020) pour la constitution de la base de données lexicale VerNom.

Pour les paires $X_V - X_V -ion/-age/-ment_N$, nous avons recherché les bases verbales potentielles de tous les noms se terminant par *-ion, -age* et *-ment*. Notre méthodologie a consisté à i) tronquer les suffixes des noms à finale en *-ion, -age* et *-ment*, ii) tronquer les marques de flexion infinitives des verbes (*-er, -ir* ou *-re*), iii) appairer les lemmes des radicaux tronqués des noms et des verbes qui correspondaient parfaitement, ou qui correspondaient jusqu'aux deux derniers graphèmes (afin de prendre en compte

l'allomorphie). Ensuite, pour les paires $X_{N/Adj} - X_V$ associées aux nominalisations, nous avons recherché les bases nominales et adjectivales des verbes correspondant à un nom suffixé (N-ion, N-age ou N-ment). Pour les besoins de cette étude, nous nous sommes concentrés sur les principaux schémas morphologiques qui permettent de former des verbes dénominaux ou désadjectivaux : les suffixations en *-iser*, *-ifier*, les préfixations en *é-*, *a-*, *en-* et la conversion du nom en verbe. La préfixation en *dé-*, qui peut être dérivée soit d'un nom, soit d'un adjectif, soit d'un verbe, n'a pas été prise en compte car elle nécessitait d'identifier manuellement la base exacte du dérivé puisque plusieurs chemins dérivationnels sont possibles (par exemple, la base du verbe *déboutonner* peut-être soit le nom *bouton* soit le verbe *boutonner* ; la base du verbe *désaler* peut correspondre soit au nom *sel* ou au verbe *saler*). Par ailleurs, la suffixation en *oyer-* (*guerre* → *guerroyer*) avait trop peu d'exemples pour être pris en compte. Afin d'identifier les bases nominales et adjectivales des verbes, nous avons tronqué les affixes déverbaux *-iser*, *-ifier*, *é-*, *a-*, *en-* (et leurs allomorphes potentiels comme *-yser*, *-izer*, *-yzer*, *em-*) des verbes, ainsi que le marquage infinitif en *-er*, *-ir* et *-re*. Les radicaux ainsi obtenus ont ensuite été appariés avec tous les adjectifs et noms collectés à partir de frCOW. Nous avons autorisé une correspondance parfaite entre les formes tronquées jusqu'aux deux derniers graphèmes afin de prendre en compte certains cas d'allomorphie, notamment lors de la correspondance des verbes convertis avec leurs bases. Certaines des bases des verbes convers ont également été récupérées en utilisant la base de données de Tribout (2010), qui contient des cas de conversion verbale de bases nominales supplétives. Les familles de triplets $X_{N/Adj} - X_V - X_V - ion/-age/-ment_N$ ont ensuite été filtrés manuellement pour s'assurer que les noms, adjectifs et verbes étaient sémantiquement liés. Ces premières étapes nous ont permis de constituer une base de données composée de 2465 triplets $X_{N/Adj} - X_V - X_V - ion/-age/-ment_N$. Le nombre de triplets par schéma est détaillé dans le tableau 1 et quelques exemples de triplets sont présentés de (3) à (8).

Tableau 1. Nombre de triplets $X_{N/Adj} - X_V - X_V - ion/-age/-ment_N$ selon leur construction pour chaque schéma de nominalisation

Nominalisations	Schéma morphologique de formation des verbes base des nominalisations						
	<i>-iser</i>	<i>-ifier</i>	conversion	<i>a-</i>	<i>en-</i>	<i>é-</i>	Total
<i>-ion</i>	1218	170	116	7	16	14	1541
<i>-age</i>	40	7	306	14	108	54	529
<i>-ment</i>	16	1	147	45	155	30	395
Total	1274	178	569	66	279	98	2465

(3) ***-iser***

a. ***-ion*** : *signal*_N – *signaliser*_V – *signalisation*_N / *sensible*_A – *sensibiliser*_V – *sensibilisation*_N

b. ***-age*** : *carbone*_N – *carboniser*_V – *carbonisation*_N / ∅³

c. ***-ment*** : *cicatrice*_N – *cicatriser*_V – *cicatrisement*_N / *aigu*_A – *aiguiser*_V – *aiguisement*_N

(4) **-ifier**

- a. **-ion** : capsule_N – encapsuler_V – encapsulation_N / pur_A – purifier_V – purification_N
- b. **-age** : code_N – codifier_V – codifiage_N / ∅
- c. **-ment** : mort_A – mortifier_V – mortifiement_N / ∅

(5) **conversion N/A → V**

- a. **-ion** : instrument_N – instrumenter_V – instrumentation_N / actif_A – activer_V – activation_N
- b. **-age** : frein_N – freiner_V – freinage_N / compact_A – compacter_V – compactage_N
- c. **-ment** : fleur_N – fleurir_V – fleurissement_N / raide_A – raidier_V – raidissement_N

(6) **en-**

- a. **-ion** : capsule_N – encapsuler_V – encapsulation_N / ∅
- b. **-age** : bouteille_N – embouteiller_V – embouteillage_N / ∅
- c. **-ment** : cadre_N – encadrer_V – encadrement_N / laid_A – enlaidir_V – enlaidissement_N

(7) **é-**

- a. **-ion** : ventre_N – éventrer_V – éventration_N / long_A – élonger_V – élongation_N
- b. **-age** : tripe_N – étriper_V – étripage_N / chaud_A – échauder_V – échaudage_N
- c. **-ment** : merveille_N – émerveiller_V – émerveillement_N / court_A – écourter_V – écourtement_N

(8) **a-**

- a. **-ion** : crédit_N – accréditer_V – accreditation_N / bête_A – abêtir_V – abêtisation_N
- b. **-age** : jour_N – ajourer_V – ajourage_N / plat_A – aplatir_V – aplatissage_N
- c. **-ment** : ligne_N – aligner_V – alignement_N / faible_A – affaiblir_V – affaiblissement_N

2.2 Constitution des paires X_{N/Adj} - X-ion/-age/-ment_N sans verbe intermédiaire

Les nominalisations dépourvues de base verbale, quant à elles, ont été collectées en identifiant tous les dérivés en *-ion*, *-age* et *-ment* qui ne pouvaient pas être appariés avec un verbe existant dans frCOW mais avec un nom ou un adjectif du corpus. Dans ces cas, l'appariement morphologique entre X_{N/Adj} et X-ion/-age/-ment_N a consisté à supprimer « ion », « isation », « ification », « isage », « ifiage », « isement », « ifiement », « ement », « ment » à la fin, ou « a », « en » et « é » au début du lemme du dérivé présumé et à trouver, dans le corpus, un nom ou un adjectif qui correspondait parfaitement. Dans certains cas, et de la même manière que pour les nominalisations, nous avons autorisé une correspondance parfaite entre les formes tronquées jusqu'au dernier graphème afin de tenir compte de l'allomorphie. Par exemple, les finales liées à la suffixation de *tunisification* (c'est-à-dire *-cation*) ont été supprimées, laissant « tunisi » s'accorder avec le nom *tunisie*, puisque le dernier graphème ne doit pas nécessairement être identique. Le nombre de N/Adj et N-ion/N-age/N-ment qui correspondaient et pour lesquels aucun verbe associé n'a pu être trouvé dans frCOW est indiqué dans le tableau 1, ci-dessus.

Compte tenu du nombre important de paires Nom-Nominalisation ou Adjectif-Nominalisation collectées par cette méthode et de la difficulté à pouvoir toutes les analyser manuellement, nous en avons sélectionné un échantillon aléatoire correspondant à 40% de chaque type de paires ; nous avons ainsi conservé 40% de paires en *-ion* (918 paires), 40% de paires en *-age* (593 paires) et 40% de paires en *-ment* (667 paires), que nous avons ensuite filtrées manuellement. La première ligne du tableau 2 montre le nombre de paires

que nous avons échantillonnées pour chaque suffixation nominalisatrice, la seconde montre le nombre de paires qui ont été nettoyées manuellement et finalement conservées.

Tableau 2. Nombre de paires $X_{N/Adj}$ - X -ion/-age/-ment_N collectées et sélectionnées

	-ion	-age	-ment	Total
Nombre total de paires N/Adj - N	918	593	667	2178
Nombre total de paires sélectionnées	168	85	12	265

La sélection manuelle a consisté à ne garder les paires que si le verbe était totalement absent ou attesté plus tard que la nominalisation sur la Toile. Le fait que le verbe supposé base de la nominalisation soit absent de la Toile ne signifie pas nécessairement qu'il n'a jamais été prononcé ou écrit nulle part auparavant. Nous avons cependant considéré que son absence (ou sa très faible fréquence) suggérait très probablement que le locuteur qui avait construit une telle nominalisation avait une intuition de la forme et du sens de ce que pourrait être le verbe de base mais que cette intuition ne dépendait pas de la réalisation effective de ce verbe.

Afin de déterminer si un verbe est manquant (ou attesté plus tard sur la Toile que la nominalisation correspondante recueillie dans frCOW), nous avons recherché les dates de première attestation des dérivés et des verbes présumés (inférés à partir de la forme de la nominalisation) en utilisant le moteur de recherche Google. Le choix d'une recherche sur Google plutôt que frCOW est motivé par le fait que ce moteur de recherche fournit des dates (ce que ne propose pas frCOW) et des résultats plus étendus (dans la mesure où frCOW date de 2016). Lorsque les lexèmes n'apparaissaient pas sur Google, nous les avons également recherchés sur Twitter, car les tweets, parfois plus susceptibles de contenir des néologismes, ne sont pas toujours référencés sur Google. La méthodologie que nous avons choisie repose sur les résultats que Google (et Twitter, dans une certaine mesure) a mis à la disposition du public au cours du mois de novembre 2021. Les lexèmes ont été recherchés en utilisant des guillemets stricts (par exemple, "macbethiser"), ce qui évite d'obtenir des réponses très bruitées. Nous avons également recherché les variantes orthographiques ("macbethiser", "macbetiser"...) et les formes fléchies ("macbethisé", "macbethisera"... / "macbethisations"...) de chacun des lexèmes.

Le moteur de recherche Google fournit les dates de publication des pages web, où un terme apparaît, directement sur la page du résultat de la recherche, ainsi que le nombre de "hits" (résultats) pour une recherche. Notre méthodologie repose principalement sur la récolte des dates indiquées sur la page des résultats de la recherche, mais dans certains cas, nous avons cliqué sur les liens des résultats et cherché les dates directement sur la page pour nous assurer que la date de publication de la page web correspondait à la date de première attestation du terme - ce qui n'est pas toujours le cas avec les livres, synopsis, journaux, dictionnaires, encyclopédies ou citations rapportées qui peuvent être publiés bien avant la page web dans laquelle ils apparaissent.

Pour éviter d'obtenir trop de résultats (dont la lourdeur aurait compromis la vérification manuelle), et même dans les cas où la nominalisation semblait plus fréquemment utilisée que le verbe, nous avons ignoré les paires pour lesquelles (i) le verbe présumé apparaissait dans plus de 10 pages de résultats (chacune comprenant de nombreux liens, ce qui alourdissait les recherches de première attestation) (ii) et dont les dates d'attestation étaient antérieures à 1950 (seuil en-deça duquel nous considérons les dérivés comme non néologiques). Si, en suivant ces conditions, le verbe était absent de Google, (ou

apparaissait plus tard que le dérivé compte tenu de leurs dates d'attestation les plus anciennes sur Google), les paires $X_{N/Adj}$ -*X-ion/-age/-ment*_N ont été collectées. Si le verbe était complètement absent, nous avons inféré sa forme à partir de la structure morphologique de la nominalisation. Le nombre total de paires sélectionnées est indiqué dans la deuxième ligne du tableau 2. Une très faible proportion de paires en *-ment* a été conservée (seulement 12) par rapport au nombre initial de paires collectées (667). Ceci est dû au fait que la majorité des paires contenaient des noms en *-ment* mal orthographiés ou provenant de dictionnaires en ligne de l'ancien ou du moyen français, ce qui remet donc en question la productivité réelle du suffixe déverbal *-ment* lorsque celle-ci est évaluée à l'aide d'un corpus massif tiré de la Toile (comprenant nécessairement des ressources telles que lexiques et dictionnaires ne recensant pas toujours de lexèmes du français contemporains), constitué automatiquement et difficile à nettoyer manuellement. Quelques exemples de paires $X_{N/Adj}$ -*X-ion/-age/-ment*_N où le verbe n'est pas attesté (ou plus tard que la nominalisation) sont présentés de (9) à (12) (les verbes non attestés ou de faible fréquence sont indiqués entre parenthèses).

- (9) **-iser**
 a. **-ion** : *Macbeth*_N – (*macbethiser*_V) – *macbethisation*_N / *violent*_A – (*violentiser*_V) – *violentisation*_N
- (10) **-ifier**
 a. **-ion** : *personnage*_N – (*personnagifier*_V) – *personnagification*_N / *brun*_A – (*brunifier*_V) – *brunification*_N
- (11) **conversion N/A → V**
 a. **-ion** : *mammifère*_N – (*mammiférer*_V) – *mammifération*_N / *translucide*_A – (*translucider*_V) – *translucidation*_N
 b. **-age** : *bandelette*_N – (*bandeletter*_V) – *bandelettage*_N / *déficient*_A – (*déficienter*_V) – *déficientage*_N
- (12) **en-**
 a. **age** : *paume*_N – (*empaumer*_V) – *empaumage*_N / ∅
 b. **-ment** : *bruyère*_N – (*embruyérer*_V) – *embruyèment*_N / *stupide*_A – (*enstupider*_V) – *enstupidement*_N

Notre étude des nominalisations dépourvues de base verbale présente donc l'originalité de s'appuyer sur des données authentiques, extraites systématiquement d'un important corpus issu de la Toile, et permettant, de ce fait, une approche quantitative.

3. Conditions morphologiques à la formation des nominalisations sans base verbale.

Comme précisé dans l'introduction, nous cherchons à mettre au jour les conditions morphologiques qui permettent l'apparition de nominalisations sans verbe de base en *-ion*, *-age* et *-ment* du français.

3.1 Les familles morphologiques de triplets standard $X_{N/Adj} \rightarrow X_V \rightarrow X_V$ *-ion/-age/-ment*_N du français

Bien que les nominalisations étudiées ne soient pas construites sur une base verbale qui leur préexiste, elles intègrent phonologiquement et/ou sémantiquement un verbe intermédiaire. En cela, elles appartiennent à des familles morphologiques au sens de Hathout 2009, Bonami & Strnadová 2019, Namer & Hathout 2020, c'est-à-dire qu'elles appartiennent à un

ensemble de mots reliés morphologiquement entre eux de sorte que n'importe lequel des membres de cette famille est relié à n'importe quel autre membre de cette famille. Plus précisément, les nominalisations en question relèvent de familles morphologiques partielles en forme de triplets (cf. 13) dont les membres sont liés entre eux par une succession de deux dérivations :

$$(13) X_{N/Adj} \rightarrow X_V \rightarrow X_V\text{-ion/-age/-ment}_N$$

- une première dérivation formant des verbes sur base nominale ou adjectivale, par suffixation en *-iser* (14a.) ou en *-ifier* (14.b.), ou par préfixation en *a-* (14.c.), *en-* (14.d.), *é-* (14.e.), et *dé-* (14.f.) ou par conversion (14g.)

- et une seconde dérivation formant les nominalisations déverbiales standard en *-ion*, *-age* et *-ment*.

- (14) a. *-iser* : $signal_N \rightarrow signaliser_V \rightarrow signalisation_N$
 b. *-ifier* : $pur_A \rightarrow purifier_V \rightarrow purification_N$
 c. *a-* : $crédit_N \rightarrow accréditer_V \rightarrow accréditation_N$
 d. *en-* : $laid_A \rightarrow enlaidir_V \rightarrow enlaidissement_N$
 e. *é-* : $tripe_N \rightarrow étripier_V \rightarrow étripage_N$
 f. *dé-* : $amiante_N \rightarrow désamianter_V \rightarrow désamiantation_N$
 g. conversion : $frein_N \rightarrow freiner_V \rightarrow freinage_N$

On comptabilise ainsi sept familles morphologiques partielles comprenant des triplets qui impliquent des noms ou des adjectifs, servant de base à la dérivation de verbes eux-mêmes base de noms d'action (cf. 15). Ces familles représentent 26.4% des nominalisations en *-ion*, *-age* et *-ment* dans la base de données VerNom (les 74.6% restants correspondant aux nominalisations dérivées d'un autre type de verbe que ceux étudiés ici, qu'il soit simple, construit au moyen d'un autre schéma morphologique ou mal orthographié).

- (15) Familles morphologiques partielles de triplets $X_{N/Adj} - X_V - X_V\text{-ion/-age/-ment}_N$
- | | | |
|----|---|------------------------------------|
| a. | $X_{N/Adj}$ - Xiser _V - Xisation _N | <i>signalisation</i> _N |
| b. | $X_{N/Adj}$ - Xifier _V - Xification _N | <i>purification</i> _N |
| c. | $X_{N/Adj}$ - aX _V - aXation _N | <i>accréditation</i> _N |
| d. | $X_{N/Adj}$ - enX _V - enXment _N | <i>enlaidissement</i> _N |
| e. | $X_{N/Adj}$ - éX _V - éXage _N | <i>étripage</i> _N |
| f. | $X_{N/Adj}$ - déX _V - déXation _N | <i>désamiantation</i> _N |
| g. | $X_{N/Adj}$ - X _V - Xage _N | <i>freinage</i> _N |

Relever d'une de ces familles morphologiques de triplets constitue la première condition à la formation des nominalisations sans base verbale que nous étudions. Cependant, chacune de ces sept familles ne donnent pas nécessairement lieu à ces nominalisations particulières : nos données révèlent que seules quatre d'entre elles le permettent. Une autre condition préside donc à leur formation. Les sections qui suivent visent à montrer qu'il faut également que ces familles morphologiques de triplets, dans le cas des nominalisations standard, se distinguent par une haute fréquence d'apparition en corpus. La prise en compte de cet autre facteur, décisif à leur formation, permet de prédire quelles nominalisations dépourvues de bases verbales sont possibles en français (et quelle est leur structure morphologique).

3.2 Les familles morphologiques de triplets permettant des nominalisations sans base verbale.

Afin de déceler lesquelles de ces sept familles morphologiques partielles de triplet permettent la formation de nominalisations sur base nominale ou adjectivale, nous avons examiné la structure morphologique des dérivés dénominatifs et désadjectivatifs en *-ion*, *-age* et *-ment* pour déceler quel type de verbes (non attestés) elles présupposent. Comme déjà précisé ci-dessus, nous nous sommes appuyées sur les formes des affixes verbalisateurs présents sur ces nominalisations (formes affixales en *-iser*, *-ifier*, *a-*, *en-* et *é-* ou aucune trace pour la conversion), ainsi que sur la valeur sémantique qu'ils véhiculent (seul critère disponible, en l'occurrence, pour la conversion). Les proportions de chaque type de verbe présumé sont calculées sur le total des nominalisations sans base verbale attestée et figurent dans le tableau 3.

Tableau 3. Proportions de nominalisations sans base verbale selon la forme de la nominalisation

	<i>-iser</i>	<i>-ifier</i>	conversion	<i>a-</i>	<i>en-</i>	<i>é-</i>	Total
<i>-ion</i>	0.51	0.08	0.08	-	-	-	0.67
<i>-age</i>	-	-	0.27	-	0.01	-	0.29
<i>-ment</i>	-	-	< 0.01	-	0.02	-	0.02
Total	0.51	0.08	0.36	-	0.04	-	1.0

L'étude de notre base de données contenant les paires $X_{N/Adj} - X_V$ *-ion/-age/-ment* _N révèle que plus de la moitié (51%) du total des nominalisations sans base verbale sont suffixées en *-ion* et comprennent une trace formelle du suffixe verbalisateur *-iser* (bien que le verbe soit attesté plus tard ou jamais sur internet). Ce résultat peut expliquer que seules les nominalisations en *-isation* aient fait l'objet d'études jusqu'à présent. En revanche, les nominalisations sans base verbale suffixées en *-age* et *-ment* ne présentent jamais de trace d'un verbe en *-iser*. Il en va de même pour *-ifier* dont on ne rencontre de trace formelle que parmi les nominalisations en *-ion* (8%). Les signes de conversion verbalisatrice, eux, sont surtout observés avec les nominalisations en *-age* (27%, par exemple *bandelette* – *bandeletage* (*bandeletter*)), mais pas exclusivement puisque 8% de ces supposés verbes convers sont observés avec les nominalisations en *-ion*, et moins de 1% avec les nominalisations en *-ment*. Pour ce qui concerne les nominalisations présupposant une base verbale préfixée non (ou plus tardivement) attestée, nos données n'ont révélé aucun signe de préfixation en *a-* et *é-*, au contraire de la préfixation en *en-* dont on observe des traces principalement avec les dérivés en *-ment* (2%), et, en plus petite proportion, avec les dérivés en *-age* (1%) (voir les exemples de (9) à (12) ci-dessus).

Les quatre familles morphologiques de triplets préférées des trois suffixations nominalisatrices sur base nominale ou adjectivale apparaissent donc clairement : la nominalisation en *-ion* sans base verbale relève préférentiellement des familles incluant des verbes en *-iser* (7a.) et en *-ifier* (7b.), la nominalisation sans base verbale en *-age* relève de la famille dans laquelle des verbes convertis sont impliqués (7c.), et la nominalisation en *-ment* relève de la famille supposant des verbes préfixés par *en-* (7d.).

- (7) Familles morphologiques de triplets permettant des nominalisations sans base verbale.
- a. $X_{N/Adj} - X_{iser}_V - X_{isation}_N$
 $\Rightarrow X - (Xiser) - Xisation$ *macbethisation*_N

- b. $X_{N/Adj}$ - **Xifier**_V-**Xification**_N
 ↳ X-(Xifier)-Xification *disneylandification_N*
- c. $X_{N/Adj}$ -**enX**_V-**enXment**_N
 ↳ X-(enX)-enXment *enstupidement_N*
- d. $X_{N/Adj}$ -**X**_V-**Xage**_N
 ↳ X-(X)-Xage *bandelettage_N*

Une fois établies les préférences des nominalisations sans base verbale pour ces quatre familles partielles en forme de triplets, nous avons cherché à déterminer ce qui les distinguaient des autres. Nous montrons ci-dessous que les familles qui peuvent donner lieu à des nominalisations sans base verbale sont celles qui, dans le cas standard, se distinguent nettement quantitativement.

3.3 Les familles partielles de triplets standard quantitativement saillantes

En effet, les préférences observées en faveur des quatre familles morphologiques donnant lieu à des nominalisations sans verbe ne sont que le reflet des préférences pour ces mêmes familles dans le cas des nominalisations standard (où le verbe est attesté avant la nominalisation). Comme le montre le tableau 4 (renseigné à partir de notre base de données contenant les familles de triplets $X_{N/Adj} \rightarrow X_V \text{ dérivé} \rightarrow X_V \text{-ion/-age/-ment}_N$ et où la proportion la plus élevée pour chaque colonne est en gras), les nominalisations standard en *-ion*, *-age* et *-ment* dérivées de bases verbales construites ont des attirances morphologiques marquées pour certaines bases : la suffixation en *-ion* sélectionne principalement les verbes dérivés en *-iser* (49%) ou *-ifier* (5%), la suffixation en *-age* sélectionne principalement les verbes convertis (11%) et une très petite proportion de verbes préfixés en *dé-* (4%)⁴ et en *é-* (1%). La suffixation en *-ment*, quant à elle, sélectionne essentiellement des verbes préfixés en *en-* (5%) et une petite proportion de verbes préfixés en *a-* (1%).

Tableau 4. Proportions de chaque type de triplet $X_{N/Adj} \rightarrow X_V \text{ dérivé} \rightarrow X_V \text{-ion/-age/-ment}_N$

	<i>-iser</i>	<i>-ifier</i>	<i>conversion</i>	<i>a-</i>	<i>en-</i>	<i>é-</i>	<i>dé-</i>	Total
<i>-ion</i>	0.49	0.05	0.04	< 0.01	< 0.01	< 0.01	0.01	0.62
<i>-age</i>	0.01	< 0.01	0.11	< 0.01	0.03	0.01	0.04	0.22
<i>-ment</i>	< 0.01	< 0.01	0.05	0.01	0.05	< 0.01	0.01	0.15
Total	0.51	0.07	0.2	0.03	0.09	0.03	0.06	1.0

Ces préférences valent également pour les nominalisations les moins fréquentes (qui apparaissent moins de 10 fois dans notre corpus) que nous avons également examinées dans l'espoir d'y relever des néologismes, moins exposés à la variation lexicale que les nominalisations de haute fréquence, et permettant, à ce titre, un accès plus direct aux propriétés des schémas morphologiques. Le constat est clair : les préférences sont constantes car elles apparaissent dans les deux cas, lorsque le verbe est attesté et lorsqu'il ne l'est pas (voir tableau 5, également renseigné à partir notre base de données contenant les familles de triplets $X_{N/Adj} \rightarrow X_V \text{ dérivé} \rightarrow X_V \text{-ion/-age/-ment}_N$). Ainsi, dans le cas des dérivations en chaînes $X_{N/Adj} \rightarrow X_V \text{ dérivé} \rightarrow X_V \text{-ion/-age/-ment}_N$, les nominalisations néologiques suffixées en *-ion* sélectionnent elles aussi préférentiellement des bases verbales en *-iser* (50%) ou *-ifier* (5%) ; les néologismes en *-age* continuent de préférer les bases

verbales converties (9%), et dans une moindre mesure, les verbes préfixés par *dé-* (5%) ou *é-* (1%). Quant aux néologismes en *-ment*, ils confirment leur prédilection pour les bases verbales préfixées en *en-* (4%) et *a-* (1%). Il y a donc une corrélation directe entre la structure morphologique des nominalisations dépourvues de base verbale et celle privilégiée par les nominalisations standard.

Tableau 5. Proportions de triplets comprenant des nominalisations dont le verbe est attesté et apparaissant moins de 10 fois.

	<i>-iser</i>	<i>-ifier</i>	<i>conversion</i>	<i>a-</i>	<i>en-</i>	<i>é-</i>	<i>dé-</i>	Total
<i>-ion</i>	0.5	0.05	0.05	< 0.01	< 0.01	< 0.01	0.1	0.63
<i>-age</i>	0.01	< 0.01	0.09	< 0.01	0.03	0.01	0.05	0.22
<i>-ment</i>	< 0.01	0.05	< 0.01	0.01	0.04	< 0.01	0.01	0.14
Total	0.53	0.05	0.2	0.01	0.07	0.02	0.08	1.0

Le degré de préférence des suffixes nominalisateurs pour certaines bases verbales dérivées, ou ce que l'on pourrait appeler la « spécialisation » des suffixes nominalisateurs, varie selon chaque suffixe, et cette variation est plus marquée encore pour les nominalisations sans base verbale. Nous évaluons la spécialisation d'un suffixe nominalisateur sur la base de deux critères : (i) d'une part le degré de préférence qu'il manifeste pour un type de base verbale construite, caractérisée par sa structure morphologique et (ii) d'autre part, le degré de préférence qu'une base verbale dérivée caractérisée par sa structure morphologique manifeste pour un suffixe nominalisateur. Ainsi, la nominalisation en *-ion* semble plus spécialisée que celle en *-age* parce qu'elle concentre très fortement ses préférences sur les bases verbales en *-iser* et *-ifier* qui ne sont jamais sélectionnées par les nominalisations en *-age* ou *-ment* (et inversement). La nominalisation en *-age*, elle, semble plus spécialisée que celle en *-ment* parce qu'elle est une des rares à sélectionner préférentiellement les bases verbales converties, mais semble néanmoins moins spécialisée que la nominalisation en *-ion*, dans la mesure où ce type de base verbale sert aussi aux dérivés en *-ion* et en *-ment*. Et enfin, la nominalisation en *-ment* sélectionne des bases verbales qui peuvent également l'être par d'autres nominalisations, et ce pour une faible quantité de données relativement aux autres, ce qui conduit à l'envisager comme peu spécialisée.

Par ailleurs, en examinant la fréquence de type de chaque suffixe déverbal, il est frappant de remarquer qu'une corrélation entre la hiérarchie de spécialisation et le classement du nombre de type se dégage nettement : la hiérarchie de spécialisation (*-ion* > *-age* > *-ment*) suit le classement des trois suffixes, comme le montre la figure 1.

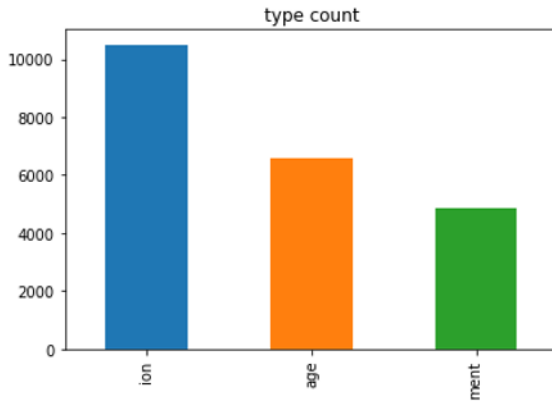


Figure 1. Nombre de paires V-Nominalisation en fonction du schéma de nominalisation du français (VerNom)

En effet, *-ion* étant également le plus fréquent de tous les suffixes nominalisateurs, suivi de *-age* et de *-ment*, la saillance des préférences des schémas de nominalisation semble bien corrélée à leur fréquence globale. Cette dernière observation pourrait conduire à penser que la spécialisation d'un schéma morphologique influe sur sa fréquence, voire sa productivité. En effet, il se pourrait que la saillance des préférences décrites assure aussi partiellement la productivité des suffixes nominalisateurs *-ion*, *-age* et *-ment* : puisque la famille morphologique de triplet est fréquente, des raccourcis sont permis et les nominalisations n'ont plus besoin d'attendre la disponibilité d'un verbe correspondant à leurs préférences pour être construites ; elles peuvent directement sélectionner un nom ou un adjectif morphologiquement et sémantiquement pertinent.

4. Bilan et perspectives

Notre étude part du constat que les nominalisations sur base nominale ou adjectivale dont le verbe intermédiaire n'est pas attesté, ou peu attesté, présupposent néanmoins morphologiquement et sémantiquement ce verbe, et, de ce fait, relèvent de familles morphologiques partielles en forme de triplets $X_{N/Adj} - X_V - X_V\text{-ion/-age/-ment}_N$. Nous avons montré, dans cet article, que ces nominalisations particulières ne sont pourtant pas possibles dans toutes les familles existantes mais sont restreintes à quatre d'elles. En effet, comme l'a révélé notre étude quantitative des préférences affichées par chaque suffixe nominalisateur pour une base verbale supposée, seules les familles comprenant la nominalisation en *-ion* sur base verbale dérivée en *-iser* et *-ifier*, la nominalisation en *-age* sur base verbale convertie et la nominalisation en *-ment* sur base verbale préfixée par *en-* peuvent donner lieu à des nominalisations sans base verbale attestée.

En outre, nous avons montré que les préférences observées dans le cas où le verbe n'est pas attesté ou est attesté après la nominalisation ne sont que le reflet de la forte fréquence que ces familles morphologiques de triplets affichent dans le cas des nominalisations standard. Ainsi, la très haute saillance des familles morphologiques concernées expliquerait la possibilité que le verbe intermédiaire des triplets ne soit pas nécessairement réalisé ou qu'il le soit avec une très faible fréquence. A contrario, si la famille constituée de triplet n'est pas quantitativement assez saillante (c'est-à-dire assez fréquente dans le cas standard), alors la formation d'une nominalisation en *-ion*, *-age* ou *-ment*, directement sur une base nominale ou adjectivale, ne serait pas possible et l'étape de verbalisation serait nécessaire.

De plus, notre étude a permis de préciser que ces préférences se font à des degrés divers et sont hiérarchisées : ainsi le suffixe *-ion* affiche les préférences les plus fortes pour les verbes suffixés en *-iser* et en *-ifier* et inversement (*-iser* et *-ifier* n'apparaissent que dans les triplets comprenant une nominalisation en *-ion*), par rapport au suffixe *-age*, qui lui préfère les bases verbales converties (mais accepte aussi les bases préfixées par *en-*). Le suffixe *-age* est également plus spécialisé que le suffixe *-ment* qui, certes, préfère les verbes préfixés par *en-* mais dans des proportions assez proches de celles de *-age*, et accepte, par ailleurs, tout type de base verbale construite, aucune ne lui étant plus spécifique. La saillance des préférences des schémas de nominalisation est corrélée à leur fréquence globale, puisque *-ion* est également le plus fréquent de tous les suffixes, suivi de *-age* et de *-ment*, ce qui pourrait suggérer que plus un affixe est spécialisé, plus il est fréquent et inversement.

Notre recherche mériterait d'être approfondie, notamment par la caractérisation sémantique des familles morphologiques partielles en forme de triplets, en fournissant des annotations plus fines des bases verbales potentielles et des bases nominales ou adjectivales. Comme nous l'avons constaté, certaines familles morphologiques sont nettement plus représentées que d'autres : c'est le cas des familles qui contiennent un toponyme ou un anthroponyme, un potentiel verbe de changement d'état en *-iser* et un nom dérivé en *-ion* comme dans *frankenstein*, *frankensteiniser* et *frankensteinisation*. En effet, les noms en *-isation* sans base verbale attestée semblent principalement liés à des noms antroponymiques ou toponymiques (voir aussi Dal 2004). Il en va de même pour les familles de triplets qui contiennent un nom d'instrument ou d'objet, un verbe converti instrumental et un dérivé en *-age* : les nominalisations en *-age* sur base nominales présupposent des verbes convertis et sont principalement liées aux artefacts concrets et aux verbes instrumentaux. Cette analyse sémantique pourra constituer une prochaine étape de notre recherche.

Par ailleurs, notre étude conduit à d'autres sujets de recherche, tels que (i) la formation de nouveaux affixes par agglutination d'affixes résultant d'une succession initiale de dérivations (cf. les notions d'« affix telescoping » de Haspelmath 1995 ou de « rule conflation » de Stump 2017) et d'autre part (ii) la formation de verbes par rétroformation. Pour ce qui concerne (i), les travaux cités en introduction ont déjà évoqué la possible création d'un nouveau suffixe *-isation*. Nos recherches permettent d'interroger également l'apparition d'un nouveau suffixe *-ification*, d'un nouvel affixe parasynthétique en *en-X-ment* (bien que nos données, trop peu nombreuses, manquent de représentativité pour permettre, pour l'heure, une généralisation), et éventuellement, d'un nouveau suffixe nominalisateur *-age* sur base nominale ou adjectivale (le verbe absent, relevant d'une conversion, n'a pas de réalisation phonologique). Pour ce qui concerne (ii), se pose toujours la question des critères permettant d'affirmer que le verbe est construit par rétroformation sur la nominalisation ou non. Comme l'évoque un de nos relecteurs anonyme (que nous remercions au passage), la sémantique très restreinte des verbes associés à ces nominalisations pourrait argumenter en faveur d'une formation par rétroformation. En effet, parmi toutes les valeurs sémantiques possibles associées à la suffixation en *-iser* (cf. Namer 2009, Bonami & Thuilier 2019) seule celle directement liée à l'interprétation de la nominalisation correspondante est activée. L'étude précise, que nous projetons de mener, des propriétés sémantiques associées aux nominalisations dépourvues de base verbale attestée, permettra certainement d'éclairer cette question.

Références bibliographiques

- Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English Morphology*. Oxford: Oxford University Press.
- Bonami, O., & Thuilier, J. (2019). A statistical approach to rivalry in lexeme formation: French-iser and -ifier. *Word structure*, 12(1), 4-41.
- Bonami, O. & Strnadova, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology* 28, no 2 (2019), 167-197.
- Dal, G. (2004). *Vers une morphologie de l'évidence : d'une morphologie de l'input à une morphologie de l'output* (Doctoral dissertation, Université Lille 3).
- Dal, G., & Namer, F. (2015). La fréquence en morphologie : pour quels usages?. *Langages*, (1), 47-68.
- Haspelmath, M. (1995). The growth of affixes in morphological reanalysis. *Yearbook of Morphology 1994*. Springer, Dordrecht, 1995. 1-29.
- Hathout, N. (2009). *Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie*. Habilitation à diriger des recherches. Université de Toulouse II-Le Mirail. Toulouse.
- Lignon, S., Namer, F., & Villoing, F. (2014). De l'agglutination à la triangulation ou comment expliquer certaines séries morphologiques. In *SHS Web of Conferences* (Vol. 8, pp. 1813-1835). EDP Sciences.
- Missud, A., Amsili, P. & Villoing, F. (2020). VerNom : une base de paires morphologiques acquise sur très gros corpus. *Actes de la 27e conférence sur le Traitement Automatique des Langues Naturelles* (TALN 2020) : 305-313.
- Missud, A. & Villoing, F. (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.
- Namer, F. (2009). *Morphologie, lexique et traitement automatique des langues*, pp. 444. Hermès-Lavoisier.
- Namer, F., & Hathout, N. (2020). ParaDis and Démonette—From Theory to Resources for Derivational Paradigms. *The Prague Bulletin of Mathematical Linguistics*, 114(1), 5-34.
- Plag, I. (2003). *Word-formation in English*. Cambridge University Press.
- Roché, M. (2009). Un ou deux suffixes ? Une ou deux suffixations ?. B. Fradin, F. Kerleroux et M. Plénat. *Aperçus de morphologie du français* (pp.143-173), Presses Universitaires de Vincennes.
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In P. Baaski, H. Biber, E. Breitenender, M. Kupietz, H. Langen & A. Witt (eds.), *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)*, Lancaster: UCREL IDS.
- Schäfer, R. & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In N. C. C. CHAIR, K. CHOUKRI, T. DECLERCK, M. U. DOÄYAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS (eds.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (pp. 486-493), Istanbul, Turkey: European Language Resources Association (ELRA).
- Stump, G. (2017). Rule conflation in an inferential-realizational theory of morphotactics. *Acta Linguistica Academica*, 64(1), 79-124.
- Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français* (Doctoral dissertation, Université Paris Diderot (Paris 7)).

¹ Les verbes non ou faiblement attestés apparaissent entre parenthèses.

² X est une variable qui représente le radical du lexème de base, qu'il s'agisse d'un nom ou d'un adjectif.

³ Le signe \emptyset indique ici que le corpus ne compte aucun exemple de nominalisation qui relève d'un triplet dont la base puisse être adjectivale.

⁴ Les verbes dénominaux et déadjectivaux préfixés en *dé-* n'ont pas été pris en compte lors de la collecte des verbes manquants car ils peuvent être facilement confondus avec les verbes déverbaux en *dé-* (comme dans *faire* → *défaire*), notamment lors de la collecte automatique des données. En effet, aucune information formelle permet de décider si le verbe *déboutonner* est construit sur la base nominale *bouton* (au moyen du préfixe verbalisateur dénominal à interprétation privative *dé-* ou s'il est construit au moyen du préfixe *dé-* sélectionnant des bases verbales (ici *boutonner*) et exprimant l'action contraire à celle du verbe de base (*déboutonner* 'action inverse de *boutonner*').