

Signes liés non-affixaux dans le lexique : de l'inventaire des bases liées en français à leur description lexicographique

Non-affixal bound signs in lexicons: census and
lexicographical description of bound bases in French

François Nemo

Université d'Orléans. Laboratoire ligérien de linguistique, France
Francois.nemo@univ-orleans.fr
<https://orcid.org/0000-0003-1216-7564>

Binène Horchani

Université d'Orléans. Laboratoire ligérien de linguistique, France
binenehorhani@gmail.com
<https://orcid.org/0000-0002-3381-6132>

Abstract

Historically, lexicography has given a relatively secondary place to infralexical signs, and among those bound signs, an even more marginal attention has been paid to non-affixal infralexical signs, such as the *trag-* of *tragique*, which nevertheless constitute a significant portion of the minimal constituents of a language. After clarifying the criteria that define them, in relation to the diversity of affixal constructions or in terms of form variation, the procedure leading to the exhaustive inventory of such bases in French is described step by step, using the treatment of adjectives in French as an example. Are further discussed i) the difficulties encountered at each step, especially regarding the polymorphy of the bases, and how to face them; ii) the question of the integration of a dictionary of bound bases in French into existing electronic dictionaries; iii) the question of the nature of the semantic description such bases may receive, given the impossibility to provide definition of what they would refer to; iv) the question of what information each entry must provide to be complete. It is shown, in particular, that the core element in the lexicographical description of these bases is the description of their *conditions of use*, and therefore of the presuppositions inherited by the words in which they are used. Several illustrations of this lexicographical treatment of bound bases are provided.

Keywords: bound bases, affixal constructions, census, polymorphy, lexicographical glosses

1. LES BASES LIÉS : UN ANGLE MORT DE LA LEXICOGRAPHIE

La lexicographie n'a dans son histoire accordé qu'une place secondaire aux signes infralexicaux pour des raisons principalement sémantiques. Parmi ces "building blocks" (Quinion, 2005), une attention encore plus marginale a été accordée aux signes infralexicaux non affixaux, bases liées non autonomes présentes à l'intérieur de mots, à l'instar de *com-* ou *trag-* dans *comique* et *tragique*, ou de *-spir-* dans *inspirer*, *expirer*, *respirer*, *aspirer*.

Loin d'être traitées comme ayant « le même statut lexical que les mots » (Corbin, 2022 [1991]), le rôle de ces composants lexicaux dans la structuration du lexique, au travers des liens interlexicaux associés au partage de telles bases, n'a que rarement donné lieu à des traitements systématiques¹, une exception notable étant le traitement par le *Robert méthodique* (1982) et sa discussion détaillée dans Corbin (2022 [1991], pp. 737-779) des morphèmes non-autonomes du français tels que la base *authent-* de *authentique*. L'objet du présent article va être dans ce contexte d'aborder les nombreuses questions que pose le recensement de ces bases ainsi que leur description lexicographique, interrogrant ainsi les deux types d'hypothèses de toute démarche lexicographique que sont les hypothèses relatives à la délimitation de ce qui doit être décrit, les entrées, et celles qui concernent la façon de les décrire.

Or, comme souvent, ces deux types d'hypothèses, externes et internes au sens de Ducrot (1980), s'avèrent ne pas être indépendantes les unes des autres, les secondes contraignant souvent les premières, ce qui conduit à ne pas traiter comme des entrées les constituants lexicaux qui ne se prêtent pas à une description lexicale en termes de définition parce qu'ils ne sont pas le nom de quelque chose, alors même qu'ils constituent une large part des signes minimaux d'une langue comme le français.

Éliminées de la liste des entrées lexicographiques, ces constituants fondamentaux représentent de ce fait le principal angle mort de la lexicographie classique, ainsi que des morphologies « à base de mot », qui ont pu présenter ces unités comme dénuées de statut d'unité sémantique (Aronoff, 1976) en tant que morphèmes vides de sens (*meaningless*).

Or, une telle affirmation est triplement problématique, en cela qu'elle :

- repose sur l'idée que les morphèmes n'ont pas de sens parce qu'ils ont trop de sens distincts, ce qui est une contradiction logique ;
- impliquerait pour être cohérente que les mots ayant plusieurs sens ne soient pas non plus des unités sémantiques, ce qui viderait le lexique de toute unité sémantique et se situe à l'opposé de la démarche lexicographique, dont le rôle principal est de décrire cette diversité de sens ;

¹ Voir néanmoins Corbin (1985), et, pour le seul vocabulaire savant, Cottez (1980).

– implique de ne pas envisager non plus l’homonymie comme alternative à la polysémie, et donc d’exclure tout regroupement homonymique des entrées.

Or, si la base liée « *-ceive* » (Aronoff, 1976, p. 14) dans les verbes *perceive*, *deceive*, *conceive*, et par conséquence la base « *-cept* » des noms associés (*perception*, etc.) se voient ainsi contester tout statut sémantique, il est tout à fait possible en réalité de grouper les emplois de ces bases par clusters sémantiques, par exemple en rapprochant *contraception* et *conception* (d’un enfant), ou encore *interception* et *réception* (d’un message), puis de créer des clusters de clusters avant de trancher in fine entre polysémie et homonymie, par exemple à propos du « *-cept* » de *déception*, en interrogeant sa relation de présupposition avec la base de *espoir*, *espérer* ou *espérance* à laquelle renvoie la glose lexicographique « Action d’être déçu quand une espérance ne se réalise pas » (CNRTL). Ce qui, loin d’impliquer que l’on doive ignorer ce type d’unité, implique au contraire qu’une démarche lexicographique authentique doit leur être appliquée.

2. RECENSER ET DECRIRE LES BASES LIEES

Toute démarche de recensement des unités en question, permettant d’établir la liste des entrées à traiter et d’envisager un dictionnaire des bases liées, et toute réflexion sur ce que pourrait être leur traitement lexicographique, commence par une définition des critères qui permettent d’identifier ces bases liées individuellement, et une clarification des questions qui se posent lorsque ces critères sont appliqués à l’ensemble des mots d’une langue, concernant la variabilité des formes identifiées et l’existence de familles de bases apparentées.

2.1. CRITERES PRELIMINAIRES D’IDENTIFICATION DES BASES LIEES

Quantitativement, les bases à recenser sont essentiellement des *composants sémantiques non affixaux de constructions affixales*, composants qui doivent être récurrents, autrement dit observables dans d’autres mots.

Sachant qu’une construction affixale est une construction incluant au moins un affixe, la plupart des bases liées seront donc des *bases liées affixées*, les autres étant des *bases liées affixales* partagées, autrement dit des composants partagés par plusieurs affixes, à l’instar de la base *hyp-* dans les préfixes *hyper* et *hypo*, qui apporte à ces affixes un thème sémantique partagé en tant que présupposition alors qu’ils sont en relation sémantique d’antonymie.

2.1.1. FLEXIBILITE FORMELLE

Ces bases s'avèrent² avoir en français des formes flexibles, formes qui sont et doivent être leurs formes orales, les formes écrites masquant très souvent le partage de base, en particulier quand cette variation est due à des graphies issues de plusieurs langues pour les composants dits néoclassiques. Il en est ainsi de relations de forme telles que la relation permutative entre *form* et *morph*-, qui associe la base orale liée \mɔʁf\ à la forme orale \fɔʁm\ du mot *forme* alors que les formes écrites sont respectivement issues du grec et du latin. Ou encore de la forme orale de la base affixale *end*- de *endo*- (e.g. *endogamie*), à savoir \ɑ̃d\, elle aussi en relation permutative avec la forme orale de *dans*, à savoir \dɑ̃\, de même signification.

La variation de forme est aussi parfois archiphonémique, à l'instar de celle que l'on observe pour les préfixes *sub*- et *sup*-, qui constituent dans leurs emplois des paires antonymiques (e.g. *sup-érieur* vs *sub-alterne*) avec partage de présupposé sémantique³ et d'une forme incluant l'archiphonème partagé par *b* et *p*, noté *b^p* ou *p^b*.

La variation de forme est aussi phonémique, à l'instar de la variation des bases de la paire *ignore* / *ignare*.

2.1.2. CONTINUITE SEMANTIQUE

Parler de composants sémantiques implique que le partage d'une forme soit associée à une proximité de sens démontrable. Ainsi, le constituant *-spir-* de *respirer* sera une base liée dès lors que sa récurrence dans les mots *inspirer*, *expirer*, *aspirer* est aussi la récurrence d'un même scénario présupposé de circulation de l'air dans et hors d'un corps.

Les bases ne sont en revanche pas identifiées sur la base d'intuitions de décomposition, pas plus que leur identification n'impose de justifier de leurs relations avec le reste du mot ou du statut des autres composants. Ainsi la base dans *instituer* est-elle obtenue par observation de sa récurrence pertinente dans *constituer*. Il en est de même pour la paire *occurrence* / *récurrence*.

² Si ce type de rapport de forme est souvent considéré axiomatiquement comme accidentel, il s'avère concerner, quand il est effectivement testé sur l'ensemble d'un lexique, la quasi-totalité des lexèmes, ce qui interdit radicalement de le considérer comme statistiquement non significatif et d'exclure sans test ce type de relations.

³ Décrire les conditions d'emploi d'un signe, sa signification, suppose de décrire *tous* les référents présupposés par le signe et la relation entre ceux-ci et non les seuls référents nommés. Le préfixe *sup*- désigne ainsi l'opposé de ce que désigne *sub*-, alors même que chacun d'entre eux présuppose l'existence de ce que désigne l'autre et ne peut être défini sans lui (Nemo, 2020).

2.1.3. DIVERSITE DES CONSTRUCTIONS AFFIXALES ET SELECTION DES MOTS AFFIXES

Les constructions affixales étant très diverses en français – les affixes pouvant être associés aussi bien à des mots (e.g. *héroïque*), des bases liées (e.g. *comique*), des bases cranberry (e.g. *ineffable*), des expansions cranberry⁴ (e.g. *microbe*) ou d'autres affixes (e.g. *multiple*) – il est crucial pour extraire les bases liées d'établir une liste de tous les mots comportant des affixes, ce qui impose souvent de démontrer la présence d'au moins un affixe par un double test de récurrence et de continuité sémantique, le principe méthodologique mis en œuvre étant qu'identifier la présence d'un affixe est possible sans avoir à analyser par avance le reste du mot. Ainsi, *ineffable* est reconnu comme adjectif à affixe, avec une base *-eff-* à tester ensuite, du fait de la double pertinence sémantique directe du préfixe *in-* et du suffixe *-able*.

2.1.4. LA QUESTION DE LA DERIVE SEMANTIQUE

Bien qu'il s'agisse en réalité d'une variante du même problème, le recensement n'a pas vocation à identifier les cas très nombreux, dits le plus souvent de *dérive sémantique*, dans lesquels la base d'un mot affixé existe comme mot autonome mais avec une pertinence partielle (ou nulle⁵) comme c'est le cas pour *décoller* quand il s'agit d'un avion, dont la pertinence partielle est strictement thématique, la présence de *colle* étant analogique et non référentielle, contrairement au *papier peint décollé* qui lui était bien collé. Ce problème concernant la quasi-totalité des constructions affixales à base de mots, il n'a pas vocation à donner lieu à un recensement, son traitement lexicographique prenant la forme d'une section dédiée aux emplois dits par extension.

2.1.5. RECURRENCE DE BASE ET INSERTION CONSTRUCTIONNELLE

Tester la récurrence d'une base n'implique pas que la base en question soit observée dans la même construction. Ainsi, peut-on identifier une base *-vor-* dans *dévorer* par observation de sa récurrence dans le mot *vorace* mais aussi dans le mot *carnivore*.

⁴ Si un segment cranberry est un segment non affixal vide de sens associé à un mot (*cran-* dans *cranberry*, Bloomfield, 1933), une expansion cranberry inclut aussi des formes uniphonémiques comme le *n* de *tourne* (vs *tour*) ou de *hiverne* (vs *hiver*) tandis qu'une base cranberry est un segment non affixal vide de sens associé à des affixes (ici *-eff-*). Les segments cranberry n'étant pas des signes ne peuvent être nommés *morphème*, et ne l'étaient pas chez Bloomfield. Les composants liés non affixaux discutés plus loin sont quant à eux des composants qui ne sont ni des bases ni des affixes – à l'instar du *m-* de *me*, *moi*, *mon*, *ma*, *mes* ou du *p* ou *m* de *père* / *mère*, *paternel* / *maternel*, *paterner* / *materner* – et qui sont associés à des récurrences sémantiques.

⁵ Comme c'est le cas pour *dérober*, sans rapport avec le sens du nom *robe*. La base française *rob-*, si elle a donné le verbe anglais *to rob*, n'existe plus comme base libre (*rober*) et n'est pas récurrente.

2.2. ILLUSTRATION : RECENSEMENT DES BASES ASSOCÉES A DES ADJECTIFS

La démarche de recensement mise en œuvre va être illustrée ici par l'exemple des bases liées associées en français à des adjectifs. Elle consiste en l'application successive d'un ensemble de tests à un ensemble d'adjectifs sélectionnés.

2.2.1. SELECTION DES ADJECTIFS A AFFIXES NON-FLEXIONNELS

La mise en œuvre de l'arbre de test est précédée des deux étapes consistant à ;

- isoler d'abord, dans une liste formée de 14320 adjectifs, les adjectifs à affixe des adjectifs simples ;
- isoler ensuite dans la liste des adjectifs à affixes, les adjectifs à affixes non flexionnels (participé passé ou présent) qui seront soumis à l'arbre de test.

S'agissant d'isoler les adjectifs sans affixes, le principe évoqué plus tôt consiste à ne retirer que les adjectifs dont il est certain qu'ils ne contiennent pas d'affixe(s), comme *bleu* ou *triste*, sans éliminer les cas dans lesquels la présence d'affixe n'est initialement que possible, l'intuition sémantique n'ayant donc aucun rôle à jouer dans la démarche.

Ainsi la présence potentielle du préfixe *re-* dans les adjectifs *rétif*, *réticent*, *revêche*, *réfractaire*, *récalcitrant*, *rébarbatif*, *répréhensible* ou *reciproque* (ou les verbes *répugner*, *regimber*, *renâcler*) ne suppose pas l'identification initiale de bases comme *-tif* ou *-vêche*. Sans pouvoir rentrer ici dans le détail de la procédure d'identification de la présence potentielle d'un affixe (Nemo, 2014), il faut noter que dans les versions automatisées de la création de la liste des mots à affixe, est testée la récurrence potentielle de l'affixe envisagé dans la liste de synonymes et antonymes, procédure très fiable puisque la récurrence d'un *ré-* (ou d'un *re-*) est observée par exemple dans 6 des 18 premiers synonymes⁶ de *rétif*, avec à chaque fois des bases opaques.

2.2.2. ARBRE DE TESTS

La mise en œuvre de l'arbre de test consiste ensuite à :

- tester la présence d'un mot dans l'adjectif, avec son sens exact⁷, test qui est nommé *test d'héritage lexical*. Il donne lieu à des sous-tests pour les cas d'alomorphie⁸ (e.g. *vér-idique* et *vrai*) ;

⁶ Par défaut, toutes les gloses lexicographiques et références dans cet article à des données lexicographiques sont issues de la version en ligne du CNRTL.

⁷ Ainsi, *chanter* est-il présent avec un sens exact dans *chanteur* alors que *viser* n'est pas présent avec son sens exact dans *vision* ou *visuel*, même si *viser* présuppose *voir*.

⁸ Présence d'un mot sous une autre forme que sa forme libre, avec identité de sens et identité grammaticale.

- relancer un test identique, si le premier test est positif, à propos du mot détecté, et ce récursivement ;
- tester, si le premier test est négatif ou si les tests suivants le sont après relance, une récurrence de la base dans d'autres mots, avec des sens proches et pertinents. Il donne lieu à des sous-tests concernant la variabilité des formes associées à la récurrence (polymorphie).

Dans le cadre d'un recensement exhaustif des bases du français, l'extraction de base ne se limite pas aux cas où la base liée est directement affixée, mais se fait récursivement pour les mots dérivés (étape 2), ce qui conduit pour *fragmentaire* à tester la présence d'une base dans *fragment*.

Le test d'héritage lexical (étape 1) vise à isoler les unités lexicales qui incluent au moins une unité lexicale autonome, et donc à séparer, les adjectifs à base libre (e.g. *héroïque*) des adjectifs à base non-autonome (e.g. *mirifique* ou *magnifique*).

La troisième étape vise à séparer les bases récurrentes des bases non-récurrentes⁹, en prenant en compte les questions de variations de forme ;

2.2.3. RECURRENCE S ÉMANTIQUE ET ALLOCONSTRUCTIONALITÉ

Une récurrence de base n'ayant pas comme nous l'avons vu à être isoconstructionnelle, elle peut être trouvée dans toute position grammaticale.

Si l'on considère par exemple l'adjectif *immense*, la base *-mens-* s'avère récurrente dans les noms comme *dimension* ou *mesuration*, ceci autour d'un thème sémantique très homogène, tournant autour d'un scénario de *mesure* et donc de la base du verbe *mesurer*. Cette mise en relation revisite en réalité en l'inversant ce que les gloses lexicographiques des mots qui les partagent ont toujours fait sans le rendre explicite, à savoir d'expliquer les mots opaques comme *immense*, par leurs polymorphes¹⁰, ici *mesur-er*, comme dans la glose « *étendue mesurable d'un corps ou d'un objet* » pour *dimension*, ou encore « *qui est illimité, qui n'a ni bornes ni mesures* » pour *immense*.

De même, la base suffixée de *mental* se retrouve-t-elle dans la base préfixée de l'adjectif *démence* / *dément* et du nom *démence*. Lien qui là aussi transforme *mental* en descripteur privilégié pour la description des sens de *démence* : *détérioration mentale*, *trouble mental*, *aliénation mentale*.

⁹ Les bases non récurrentes, à l'instar de la base *-eff-* dans *ineffable*, dont l'existence n'est reconnue qu'au travers de la certitude de la présence des affixes dans le mot, sont loin d'être rares en français.

¹⁰ Les polymorphes sont des réalisations différentes d'un même morphème (e.g. *roul(-er)* vs *roue*) dans des lexèmes distincts, alors qu'un allomorphe est une réalisation différente d'un même lexème, sans variation sémantique ni grammaticale et avec contrainte de distribution complémentaire. Un lexème est un signe grammaticalement défini, à l'instar des préfixes *sup-* et *sub-*, polymorphes antonymes comme vu précédemment du morphème *su^{p/b}* dont la signification correspond aux présupposés sémantiques qu'ils partagent (e.g. *supérieur* vs *subalterne*)

2.2.4. RECURRENCE SEMANTIQUE ET POLYMORPHIE

La récurrence de bases passe souvent par un type spécifique de variation de forme, à savoir la polymorphie. Contrairement à l'allomorphie qui ne concerne en fait qu'un lexème¹¹, n'étant associée à aucune variation sémantique ou grammaticale (catégorielle) et imposant une distribution complémentaire, la polymorphie est une relation entre lexèmes, associant différenciation sémantique et parfois grammaticale avec une différenciation partielle de forme.

La question de la polymorphie se pose dès que des différences de forme limitées coexistent avec des différences de sens limitées et s'accroissent lorsque le profilage¹² des éléments sémantiques partagés varie significativement.

Pour la paire à base isomorphe *implicite* / *explicite*, l'évidence d'un lien sémantique entre les deux adjectifs est immédiate, leurs sens étant antonymiques, ce qui implique une unité thématique et le partage d'un thème sémantique. Mais si l'on considère ensuite les verbes *impliquer* et *expliquer* et les noms *explication* et *implication* se pose une double question, celle du lien entre *implicite* et *implication*, et ensuite celle du lien entre *impliquer* et *expliquer*, l'unité thématique étant en partie masquée. S'il suffira pour la première question de constater qu'*implication* peut être un hyponyme d'*implicite*¹³, il faudra pour la seconde décrire en détail les conditions d'emplois et mettre à jour les présupposés partagés dans tous les emplois ou au moins une partie de ceux-ci (profils), à savoir l'existence d'une relation entre deux réalités A et B telle que l'existence de A rend logique l'existence de B, pour décrire la signification de la base.

Toute lexicographie des bases liées étant exposée à devoir justifier des rapprochements que la lexicographie des mots fait constamment sans avoir à les justifier¹⁴, le principe méthodologique à mettre en œuvre est à l'inverse de s'appuyer sur les rapprochements lexicographiques effectués dans les gloses de mots pour établir une récurrence polymorphique de façon incontestable. Démarche qui pourra aller du

¹¹ Un lexème est une unité sémantique grammaticalement définie observable dans une séquence linguistique, alors qu'un morphème est une unité sémantique non-définie grammaticalement qui est partagée par des lexèmes (Nemo, 2020). Ainsi, le morphème /ni/ est-il présent aussi bien dans le verbe *ni(er)* que dans la conjonction *ni* et le préfixe *in-* qui sont trois lexèmes distincts. De même, les formes *no*, *not*, *non-* en anglais sont-ils des lexèmes distincts et des polymorphes du même morphème. De même enfin, le morphème /able/ en anglais existe aussi bien comme suffixe que comme composant de *to be able to*, qui sont des lexèmes distincts. La signification d'un morphème est constituée des présupposés sémantiques partagés par tous les lexèmes dans lesquels il est employé.

¹² Au sens de Cadiot & Visetti (2001), les profils sont des strates intermédiaires entre signification et sens, partagés par un ensemble d'emplois d'un signe mais pas tous. Ils sont aussi parfois nommés interprétation-type.

¹³ L'inverse étant faux, un implicite n'étant pas forcément une implication.

¹⁴ Une glose lexicographique utilisant *poésie* (et non *poète*) pour expliquer *poétique* n'a pas à justifier de ce décalage, alors que la description de *-plic-* et *-pliqu-* comme variantes de la même base pourra être mise en question.

simple repérage de relation de coréférence ou de relation d'hyponymie / hyponymie, comme dans la glose suivante :

*L'implication est l'acte de suggérer quelque chose sans le dire directement (l'aspect **implicite**), ou la conséquence logique d'une chose.*

Mais il peut aussi passer par la recherche récursive automatisée d'éléments de forme dans les paraphrases proposées par des gloses lexicographiques. Comme l'illustrent les gloses du mot *arduous* en anglais dans le *American Heritage Dictionary* (AHD).

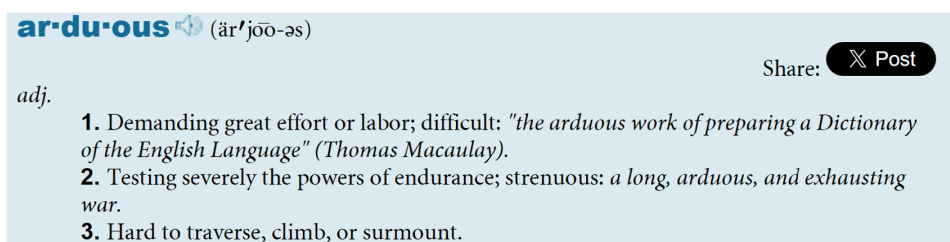


Figure 1. *American Heritage Dictionary*

Source : site web.

à propos desquelles on observe la présence directe de *hard* comme descripteur direct pour le troisième sens ainsi que, après test récursif du descripteur *endurance* dans le second sens, la présence de *hard* à la fois comme descripteur du présupposé principal de ce mot et dans un exemple.

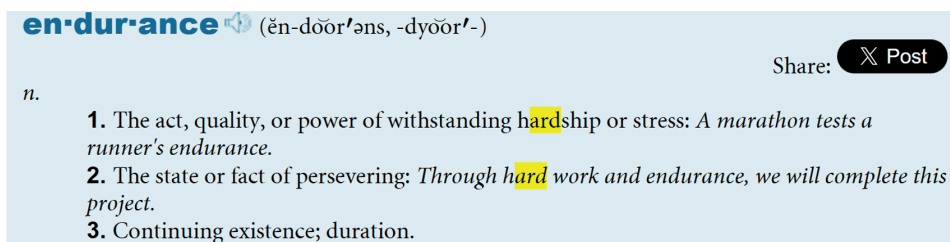


Figure 2. *American Heritage Dictionary*

Source : site web.

Le tout faisant qu'associer la base d'*arduous* avec son polymorphe *hard* n'est pas une spéculation hasardeuse mais la simple explicitation des relations tacitement introduites dans la lexicographie préexistante de *arduous*.

D'autres techniques, mesurant la substituabilité des paraphrases ou l'interopérabilité des descripteurs, peuvent aussi être mises en œuvre. Elles consistent à se demander d'abord si la glose produite pour une base de forme FA serait pertinente pour décrire un mot de forme FB, par exemple si dans la glose « *Ce qui est impliqué par quelque chose, ce qui est contenu sans être exprimé* » associée à *implication*, la partie « *ce qui est contenu sans être exprimé* » serait acceptable comme glose pour le mot *implicite*. Et à chercher à repérer ensuite la récurrence de mêmes descripteurs D1, D2 *etc.* dans les gloses lexicographiques des mots FA et FB dont la proximité des bases est testée. Largement automatisable dans le cadre d'une démarche de recensement, cette technique permet d'identifier les présupposés sémantiques partagés, y compris pour tester des relations d'antonymie.

Cette approche peut, notamment pour le vocabulaire technique, être étendue à des définitions non lexicographiques¹⁵ telles que « *le **béton matricé** consiste à appliquer une matrice ou **trame** (sorte de pochoir géant de 1m de large) sur un support encore frais* » qui mettent directement en relation de substituabilité référentielle la base liée de *matr-ice* et son polymorphe *trame*. Ce qui permet de multiplier les traces de la réalité cognitive de la relation identifiée dans des domaines d'expériences variés, tels que :

La paroi renferme une trame polysaccharidique comportant deux phases : la matrice, amorphe et continue, essentiellement de nature pectique et hémicellulosique, et la charpente fibrillaire de cellulose. (Unisciel, Université des Sciences en Ligne)

Bi-trame Une image bi-trame se compose de trois matrices rectangulaires d'entiers: une matrice de luminance (Y) et deux matrices de chrominance (Cb et Cr). (Union Internationale des Télécommunications IUT H.262, 02/2000)

Une fois envoyée sur le site Des_frgs, cette image matrice va servir de trame sur laquelle viendront s'afficher d'autres images fixes ou animées récupérées sur le Web pour composer une image mosaïque finale. (Fourmentaux & Drouhin, 2003)

La polymorphie en français prend des formes variées et parfois plus surprenantes, notamment lorsqu'elle concerne les emplois libres de la base liée. Souvent en effet, il y aura bien récurrence polymorphique de la base, mais comme mot, et plus souvent encore, à la fois récurrence de la base sous forme de base et sous forme de mot. A l'instar de la base liée de *rot-ation* et *rot-or* qui a *tor(dre)* comme polymorphe¹⁶, ou en anglais de la base de *astr-onomy* et *astr-ology*, qui a *star* comme polymorphe¹⁷.

¹⁵ Telles que nomenclatures, ontologies ou entrées Wikipédia.

¹⁶ La non-linéarité des morphèmes est massive en français, alors que les lexèmes sont presque toujours linéarisés.

¹⁷ Directement ou indirectement, cette relation est utilisée dans les gloses lexicographiques, p. ex. celles du *American Heritage Dictionary*, pour décrire la forme libre en faisant appel à la forme liée.

Et enfin de la base de *fict-if* et *fiction*, dont la base antonymique et polymorphe est la base de *factuel*.

2.2.5. RECURRENCE DE BASES ET ETYMOLOGIE

L'identification de récurrences de base est une approche *sens / forme et non forme / sens*, même si toute recherche de récurrences commencera par celle de formes isomorphes, la contrainte de partage de présupposés sémantiques partagés constituant toujours la contrainte dominante.

Cela conduit à ce que soit identifiés des liens d'une autre nature que les liens étymologiques. Associer la base de *grégaire* à celle d'*agréger* n'est en rien une démarche étymologique, visant à identifier la seconde comme étant à l'origine de la première ou l'inverse, mais une démarche consistant à identifier un scénario sémantique partagé en synchronie entre formes proches, et donc une paire sens / formes, exactement comme l'identification d'un scénario partagé entre *com-ique* et *com-édie* permet de reconnaître *com-* comme une base.

Il s'en suit que si les récurrences de base peuvent être héritées du passé, et correspondre à ce que dira une explication étymologique, à l'instar de ce qui serait le cas pour la base *-mens-* de *immense* citée précédemment, l'inverse sera donc faux et ce pour plusieurs raisons.

La première raison est que la proximité de forme conduit fréquemment à ce qu'un emploi nouveau d'un mot soit rattaché à une autre base que sa base d'origine¹⁸, rendant non pertinente du fait de la rupture sémantique la référence à l'origine en question.

La seconde est que les bases sont constamment remobilisables par des emplois qui ne sont pas en continuité avec les emplois précédents. Ainsi la création du couple *implicature / explicature* par transformation d'*implication*, a donné à *explicature* un sens antonymique en forte continuité avec celui d'*explicite* et en rupture forte avec celui d'*expliquer*.

La troisième, liée à la précédente, est que les transferts d'un sens stable entre formes proches partageant une base, à l'instar de la relation entre *explicite* et *explicature*, sont omniprésents, toute approche faisant l'hypothèse inverse d'une modification du sens d'une forme stable¹⁹ ne pouvant être ni adéquate ni adaptée à l'identification des bases liées.

¹⁸ Ainsi l'emploi linguistique de *macron* quand il est défini par Wikipédia comme « *diacritical mark* », passe-t-il du sens de *macro* (vs *micro*) au sens de *marque*. Seule une partie des rapprochements de ce type est associée dans les explications étymologiques à des formules de type « *sous l'influence de* », les autres comme ici *mark* pour *macron* sont seulement introduits.

¹⁹ Cette approche conduirait pour notre exemple à vouloir décrire le sens d'*explicature* comme obtenu par transformation du sens d'*explication* ou d'*expliquer* sans passer par une relation de polymorphie avec *explicite*.

On observe en effet que la diversité des profils sémantiques associés à un morphème – que celui-ci ait des emplois libres ou seulement des emplois liés – le sont non pas à tel ou tel lexème dans lequel celui-ci est employé mais à l'ensemble de ceux-ci, avec une constante transférabilité de la polysémie d'une forme lexémique à l'autre (Nemo & Horchani, 2018). Il s'en suit que toute approche faisant l'histoire des sens associés à des formes lexémiques particulières ne pourra que constater des évolutions sémantiques erratiques, alors même que les sens en question ont très largement été transférés de formes proches sans changement de sens, ce dont il est précisément question pour l'identification de récurrence de bases.

Plus généralement, comme l'illustrent les morphèmes et lexèmes de quantité²⁰, il n'y a en fait pas de relations d'héritage sémantique qui soient dépendantes des formes employées : si le rapport entre *quatre* comme lexème et la forme orale *quar(t)* est une relation de polymorphie, celle entre *quatre* et la base de *quarante* est techniquement analysable comme une relation d'allomorphie dérivationnelle, avec comme résultat que *quarante* n'est pas en relation d'héritage lexical avec *quar(t)*. Dès lors qu'un présupposé sémantique unique comme 4 sera partagé par un grand nombre de formes, dont un nombre important de formes liées, il sera nécessaire pour leur description lexicographique de traiter ensemble toutes ces formes et toutes leurs relations sans faire une histoire sémantique des formes.

3. DESCRIPTION LEXICOGRAPHIQUE DES BASES LIÉES : PRINCIPES ET ILLUSTRATIONS

Aborder la question de la description lexicographique des bases liées ouvre trois séries de question. La première, liée à la polymorphie des bases et à la question de la consultation, est celle de l'adressage lexical et des relations entre bases. La seconde, liée au fait que les bases liées ne nomment jamais quelque chose – ce qui interdit au lexicographe de chercher à définir ce qu'elles nomment – est celle du type de description qui peut leur être associée. La dernière est celle des rubriques à renseigner pour leur description.

3.1. INTEGRATION AUX DICTIONNAIRES EXISTANTS OU DICTIONNAIRE AUTONOME

Tout recensement des bases liées rend possible la création d'un dictionnaire spécialisé qui leur soit dédié, pouvant aussi inclure les bases allomorphiques.

²⁰ La base liée de *quant*-ité est par ailleurs un polymorphe de la forme *compt*- dans *compter* ou *compte*.

Si globalement le principe d'un adressage alphabétique des entrées peut être conservé, il faut noter qu'il poserait néanmoins un problème de consultation de celui-ci. Comme les premiers dictionnaires du français lorsqu'ils réunissaient des familles de mots associées à des familles de formes plutôt que des listes de mots définis alphabétiquement, le problème est que l'utilisateur du dictionnaire ne connaîtra souvent pas la forme des bases qu'il pourrait consulter.

Ce problème ne se pose néanmoins pas pour les versions électroniques des dictionnaires, qui peuvent être associées à des interfaces et ne sont pas soumises à des contraintes de taille.

La démarche la plus adaptée est donc d'intégrer pour consultation l'information relative aux bases liées à la description des mots dans lesquelles elles sont employées. Autrement dit d'en faire une rubrique dans le traitement de ces mots, avec lien unique sur une entrée autonome dédiée pour l'ensemble de ceux-ci.

Ainsi est-il possible pour *intérieur*, *extérieur*, *postérieur* et *antérieur* d'indiquer qu'ils sont formés sur une base formelle commune²¹, à la description lexicographique de laquelle un lien permet d'accéder. Une telle approche pour être pertinente doit être affinée par un ancrage sur des sens spécifiques des lexèmes concernés. Ainsi, les quatre adjectifs ne sont-ils à rapprocher que pour les emplois spatiaux des deux derniers, *intérieur* et *extérieur* n'ayant pas d'emplois temporels. Seuls ces derniers en revanche sont à mettre en relation avec la base présente dans *externe* et *interne*, *posterne* et *anterne* n'existant pas. La nécessité d'atteindre ce degré de précision est cruciale comme l'illustre la base liée *-morphe* dans *amorphe* en français, emploi qui n'a de lien pour son sens usuel avec *forme* que pour l'interprétation que l'on retrouve dans « *être en forme* » et non pour ses emplois « géométriques » (e.g. tracé).

Cette approche, qui permet aussi de surmonter le problème d'incompatibilité entre un adressage alphabétique et la non-linéarité des bases, n'interdit pas une publication autonome parallèle d'un dictionnaire de bases, mais est la seule approche compatible avec une logique de consultation.

3.2. GLOSE SEMANTIQUE

Les bases liées n'ayant de référent dénominatif que dans le cadre du mot dans lequel elles sont employées, leur glose est exposée à un double risque de :

- circularité, consistant à expliquer *morn-* par *morning* (AHD), *dynam-* par *dynamique* ;
- d'inconsistance (e.g. expliquer *loqu-et* par *loque*).

²¹ Techniquement, une base est ce qui reste d'un mot à affixe(s) une fois le ou les affixes retirés, donc ici *-terne* ou *-térieur*. Analytiquement, ces bases peuvent donner lieu à décomposition (partage de *-ter-* entre *-terne* et *-térieur*), mais lexicographiquement, ce sont dans ce cas les récurrences qui doivent être introduites et décrites.

Si expliquer une base suffixée par la même base avec un autre suffixe est souvent circulaire, cela peut être parfois, par exemple en rapprochant *prolifique* par *proliférer* une manière commode de forcer le lecteur à fusionner en un unique scénario ce que les deux mots ont en commun.

Echapper à ces formes de circularité passe en réalité par la description de la contribution sémantique de la base au sens des mots dont elle est un constituant et non par celle du sens en question. Ce qui revient à mettre l'accent sur la description des présupposés partagés par tous les mots dans lesquels elle est employée et en premier lieu sur les référents présupposés et non nommés. Dans la paire *hermano* / *hermana* (frère, sœur) en espagnol, la base *herman-* introduit en effet comme présupposé quatre personnes (2 parents et au moins 2 enfants) et un ensemble de relations entre elles, indépendamment de ce qui est ensuite nommé par l'adjonction des marques -o et -a, à savoir une seule de ces quatre personnes. Décrire ce qui doit se passer depuis le début pour que la base puisse être employée à propos de quelque chose est donc l'objectif premier de toute glose d'une base. A partir de quoi :

- la base *-tens-* des antonymes *extensif* / *intensif* et de *extension*, indique ainsi qu'il est question de la quantité de quelque chose qui peut être mise dans un espace donné, ou de la réduction de l'espace que quelque chose peut occuper. Cette base est le polymorphe de *tasser* et *sat-urer*, qui partagent cette thématique (Nemo, 2025), la base *sat-* évoquant plus précisément l'impossibilité ou non d'ajouter quelque chose à un contenant ;
- la base *com-* de *comique* / *comédie* indique que quelque chose fait rire quelqu'un, ce qui est aussi présupposé par la base de son polymorphe *moqu(er)* ;
- la base *mar-* de *marraine* et *marâtre* indique que quelqu'un remplace ou est à même de remplacer une mère après décès de celle-ci pour s'occuper de ses enfants. Cette base est un polymorphe de *mère*.
- la base *-nonc-* des verbes *prononcer*, *énoncer*, *annoncer* indique que quelque chose a été dit, en particulier publiquement. Cette base est un polymorphe de *son*, *sonner*, *résonner*, etc.
- la base *méd-* de *médical*, *médicament* et *médecin* indique qu'une personne malade ou blessée est prise en charge.

3.3. INFORMATIONS LEXICOGRAPHIQUES

Outre la description des présupposés partagés par l'ensemble des emplois, le format de description d'une base liée doit :

- rappeler la distribution de celle-ci, autrement dit les mots dans lesquels elle est employée ;
- décrire la contribution de la base à l'interprétation et au sens de ces mots ;

- introduire aux polymorphes de la base décrite ;
- décrire la ou les familles de bases auxquelles elle est associée, et les clusters dénominatifs les associant²².

Ce qui conduira pour la distribution à une définition extensionnelle :

pit- : base²³ de *piteux* et *pitié* et de leurs dérivés.

splend- : base de *splendide*, *splendeur*, *resplendir* et de leurs dérivés.

ép- : base d'*épique* et *épopée*.

Définition qui pourra être justifiée :

La base d'*octogénnaire* est associée d'une part à *quinquagénnaire*, *septuagénnaire*, etc. ainsi qu'à *génération*, au sens de « *Ensemble de ceux qui vivent à une même époque et qui ont sensiblement le même âge* ».

Ce qui conduira pour la polymorphie à l'introduction des polymorphes :

Thé- : base des mots *théologie*, *théocratie*, *athée*, *polythéisme*. Polymorphe de la base *dé-* de *déesse*, *déité*, *déiste*.

Lum- : base des mot *lumière*, *lumineux* et *illuminer*. Polymorphe des bases de *luciole* et *luire*.

ru+ : base des mots *rur-al*, *rust-ique*, *rustre*. Polymorphe de la base *urb-* des mots évoquant ce qui n'est pas rural. Indiquent respectivement que quelque chose a un lien avec la campagne ou la ville.

Dynam- : base de *dynamique*, *dynamie* ou *dynamo*. Indique que quelque chose qui est en mouvement n'est pas ralenti ou accélère. Polymorphe d'*anim-er*.

Hydr- base d'*hydrater* et des composés incluant *hydro-*. Polymorphe de la base *-ride* associée au préfixe privatif *-a* dans *aride*, au sens de « *sans eau* ».

Autor- : base d'*autoriser*, *autorité* et *autoritaire*, Polymorphe de la base d'*ordonner* et *ordre*.

²² Une famille de bases réunit toutes les bases, liées ou non, qui sont en relation de polymorphie. Ainsi, *occl-* dans *occlusion*, *loqu-* dans *loquet* et *clo(s)* forment une famille de bases. La notion de cluster dénominatif renvoie quant à lui au fait que les polymorphes sont souvent utilisés pour nommer des réalités ontologiquement proches, par exemple dans les paires *reine* / *règne*, *hiverner* / *hiberner* *aorte* / *artère*, *beugler* / *meugler* ou *mound* / *mount* en anglais.

²³ Ce type de base, à l'instar de la base *lex-* de *lexique*, est un constituant du lexique à part entière, qui est notamment mobilisable pour la création de mots nouveaux (e.g. les mots *lexèmes* ou *lexie* en linguistique).

Anx- : base d'*anxieux* et *anxiété*. Polymorphe de *ang-* dans *angoisse*. Indique que quelqu'un souffre par anticipation ou réactivation d'un événement traumatisant.

Pig- : base de *pigment* et polymorphe²⁴ des bases *pic-* de *pictural* et *pix-* de *pixel*.

Frag- : base de *fragment* et polymorphe de la base *frac-* de *fracture* ou *fracasser*

Seg- : base de *segment* et polymorphe de la base / *sek* / de *section* ou de *séquence*

Grég- ; base de *grégaire*. Polymorphe de la base d'*agréger* dans son sens de « *se rassembler en un tout compact* ».

Suppl- : base interne de *supplémentaire*. Polymorphe de *plus*, de même sens.

Autant que possible, la glose devra décrire la relation sémantique entre indications morphémiques et emplois de la base (i.e. présupposés partagés et sens dénominatif ou autre) :

Sen- : base des adjectifs ou noms *sénile*, *senior*, *sénescence*. Indique que ce dont il s'agit

concerne l'âge avancé du référent. Associée soit à des mots évoquant directement cet âge, ou des réalités associées à cet âge, telles que la réduction des facultés cognitives.

Obst- : base d'*obstruction*, *obstacle* mais aussi d'*obstiné*, *s'obstiner* ou *nonobstant*. Indique que le mouvement de quelque chose ou l'action de quelqu'un est empêché. Associée au morphème négatif *-in-*, indique que quelqu'un ne s'arrête pas (alors qu'on s'attendrait au contraire). Associée dans *nonobstant* aux affixes *-ant* et *non-*, ou elle a le sens de « Sans être empêché par qqch., sans s'y arrêter ».

-morphe : base d'*amorphe*, qui indique quand il réfère au « caractère peu énergétique », qu'il est question de quelqu'un qui n'a pas une activité ou forme physique importante ou même normale.

Comme évoqué à propos de la relation *matrice* / *trame*, il est souvent possible de décrire de façon précise la polysémie dénominate des emplois de la base liée, la relation *trame* / *matrice* étant ainsi soit une relation de synonymie, soit une relation de type hyperonyme vs hyponyme.

L'entrée peut aussi comporter une description de la construction dans laquelle la base est employée :

²⁴ On notera que la neutralisation du voisement est régulière en français, créant une polymorphie systématique.

-Atern- : base de *maternel*, *paternel* et *fraternel* ainsi que de *materner*, *maternité*, *fraternité* ou *paternaliste*. Se combine avec les composants liés non-affixaux *p-*, *m-* et *fr-*, eux aussi récurrents dans d'autres mots, Allomorphe de la base *-ère* de *père*, *mère* et *frère*, Indique l'existence de relations de parenté, dont la nature est précisée par les composants non affixaux *p-*, *m-* et *fr-*.

Enfin, comme cela a déjà été souligné, toute lexicographie des bases liées a vocation à associer comme exemples et preuves des mises en relation qu'elle opère des exemples issus de la lexicographie des mots ou encore de définitions encyclopédiques au sens large.

4. CONCLUSION

La question de savoir si un lexique doit être considéré comme formé de familles de mots ou comme une liste de mots a été au cœur de la naissance de la lexicographie.

La prise en compte des bases liées redécouvre cette question sous un angle inédit, à savoir l'existence de relations interlexicales qui, autant que les lexèmes, méritent d'être décrites.

Or, si l'intégration des bases liées à la lexicographie d'une langue a des aspects complexes, elle donne paradoxalement l'occasion de rendre explicites et de mettre au premier plan des relations constamment mobilisées en arrière-plan par la lexicographie des mots, tout en systématisant la démarche en prenant en compte l'omniprésence de la polymorphie en français.

BIBLIOGRAPHIE

- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge (Mass.) : The MIT Press.
- Cadiot, P. & Visetti, Y.-M. (2001). *Pour une théorie des formes sémantiques*. Paris : P.U.F.
- Corbin, D. (1985). Les bases non autonomes en français ou comment intégrer l'exception dans un modèle lexical. *Langue Française* 66, 54-76. <https://doi.org/10.3406/lfr.1985.6345>
- Corbin, D. (2022 [1991]). *Morphologie dérivationnelle et structuration du lexique*. Lille : Presses universitaires du Septentrion.
- Cottez, H. (1980). *Dictionnaire des structures du vocabulaire savant : éléments et modèles de formation*. Paris : Le Robert.
- de Fourmentraux, J.P. & Drouhin, R. (2003). Les coulisses de l'œuvre net art. *Ligeia CNRS*, 45-48, 116-122.
- Le Robert Méthodique. Dictionnaire méthodique du français actuel* (1982). Paris : Le Robert.
- Nemo, F. (2014). Interprétabilité ou grammaticalité ? Les listèmes comme interface entre sémantique et morphologie. *Revue de Sémantique et Pragmatique*, 35-36, 105-144. <https://doi.org/10.4000/rsp.1582>
- Nemo, F. (2020). Le lexique entre conditions d'emploi (des morphèmes) et référence lexicale. In G. Kleiber et al. (éds.), *Lexique et référence. Res Per Nomen VII* (pp. 409-426). Reims : Editions et Presses Universitaires de Reims (EPUR).
- Nemo, F. (2025). Différenciation lexicale et construction linguistique de l'espace. In M. Krazem, D. Saulan & S. Tchougounnikov (éds.), *Espaces du sujet parlant Intime et public, lecture et écriture* (pp. 155-178). Dijon : EUD.
- Nemo, F. & Horchani, B. (2018). Accounting for transcategorial morphemes: Theoretical and methodological issues. *Cognitive Linguistic Studies*, 5.1, 77-105. <https://doi.org/10.1075/cogls.00014.nem>
- Quinion, M. (2005). *Ologies and isms: A dictionary of word beginnings and endings*. Oxford : Oxford University Press.