Université Paris Nanterre

200 av. de la République
92001 Nanterre Cedex
www.parisnanterre.fr

Membre de l'université Paris

École doctorale 139 :
Connaissance, langage, modélisation
MoDyCo (UMR, 7114)

Yifeng Lin

La formation des noms d'instruments par suffixation en français

une perspective cognitive

Thèse présentée et soutenue publiquement le *12/09/2025*
en vue de l'obtention du doctorat de Sciences du langage: linguistique générale et française de l'Université Paris Nanterre
sous la direction de Mme Florence Villoing (Université Paris Nanterre)
et de Mme Juliette Thuilier (co-encadrante, Université Toulouse II Jean Jaurès)

Rapporteur :	M. Guillaume Desagulier	Professeur, Université Bordeaux Montaigne
Rapporteur :	M. Fabio Montermini	Directeur de recherche, Université Toulouse II Jean Jaurès
Membre du jury :	M. Richard Huyghe	Professeur, Université de Fribourg (Suisse)
Membre du jury :	Mme Delphine Tribout	Maîtresse de Conférences, Université de Lille

Remerciements

Je tiens tout d'abord à exprimer ma plus sincère gratitude à mes deux directrices de thèse. Je remercie chaleureusement Professeure Florence Villoing. C'est elle qui m'a offert cette précieuse opportunité de réaliser un doctorat sous sa direction, me permettant ainsi de concrétiser un rêve nourri depuis ma licence : devenir chercheur en morphologie. Tout au long de ces trois années, elle m'a guidé avec rigueur, patience et un grand professionnalisme. De la définition du sujet jusqu'à la rédaction finale, en passant par l'exploration des théories linguistiques, elle m'a constamment apporté un soutien précieux et des éclairages stimulants. Son accompagnement a été particulièrement décisif dans la conception et le développement du cadre théorique central de cette thèse, la *Néo-Morphologie Cognitive* : c'est grâce à ses encouragements, son soutien et ses conseils attentifs que ce cadre théorique a pu voir le jour et s'affiner. J'ai toujours beaucoup appris de ses commentaires, à la fois profonds, pertinents et constructifs. J'apprécie particulièrement la clarté et la franchise avec lesquelles elle formule ses remarques. Durant ces trois années, sa compréhension et sa bienveillance ont également été d'un grand réconfort. Avoir été son doctorant est un véritable honneur.

Je remercie également très sincèrement Mme Juliette Thuilier d'avoir accepté d'être ma co-encadrante. Sa présence dans l'encadrement de cette thèse a été pour moi d'un grand enrichissement, tant sur le plan méthodologique que rédactionnel. Elle m'a apporté une aide essentielle à toutes les étapes importantes du travail doctoral : méthodologie, traitement des données, analyses statistiques et rédaction académique. Ses conseils ont toujours été précis et pertinents, notamment en ce qui concerne l'organisation de la structure du texte et la formulation des idées, ce qui m'a permis de progresser de manière significative. Son souci du détail et sa grande sensibilité aux enjeux de clarté argumentative m'ont non seulement aidé à améliorer mes compétences rédactionnelles, mais aussi à développer un style scientifique plus mûr. Tout au long du processus d'écriture, elle m'a accompagné avec patience, exigence et générosité, et je lui en suis profondément reconnaissant.

Merci également aux rapporteurs de cette thèse, M. Guillaume Desagulier et M. Fabio Montermini, pour le temps considérable qu'ils ont consacré à la lecture attentive de mon manuscrit, ainsi que pour la rigueur et la précision de leurs rapports. Leurs remarques, à la fois exigeantes, stimulantes et constructives, ont grandement contribué à l'enrichissement et à la clarification de ce travail. Leurs retours m'ont permis d'approfondir certaines dimensions théoriques et de mieux mettre en valeur les apports de ma démarche. Leur lecture fine et leur bienveillance ont représenté un appui précieux dans l'aboutissement de ce parcours.

Merci aussi aux examinateurs du jury, M. Richard Huyghe et Mme Delphine Tribout. Les échanges avec M. Huyghe ont été particulièrement enrichissants, et ses travaux ont profondément nourri ma réflexion. Je suis très reconnaissant d'avoir pu

bénéficier de son regard critique et bienveillant lors de la soutenance. Je remercie également Mme Tribout pour sa participation au jury et pour la qualité de ses observations lors de la soutenance. J'ai toujours suivi ses travaux avec beaucoup d'intérêt, et c'était un privilège de pouvoir bénéficier de son regard sur ce travail.

Je remercie les enseignants et collègues du Modyco pour leur soutien et leur aide dans l'utilisation des équipements, le traitement des données, les échanges lors des séminaires ainsi que dans les tâches du quotidien. Je tiens à exprimer une reconnaissance particulière à Mme Aude Da Cruz Lima pour son aide précieuse dans la mise en place de la base de données, qui a permis à mon travail de recherche de progresser de manière indispensable. Je tiens également à remercier les représentants des doctorants du laboratoire, dont l'engagement a grandement facilité mon intégration au sein de l'équipe et des activités collectives. Je remercie aussi chaleureusement le laboratoire pour le soutien financier qui m'a permis de participer aux Journées de Linguistique Cognitive. Je remercie également l'École doctorale 139 pour le soutien administratif apporté au cours de ces années de doctorat.

Je tiens également à remercier tout particulièrement ma directrice de mémoire de master, Mme Chunhong Yu. Durant mon parcours de master, elle m'a accordé une grande confiance et un soutien constant, me permettant de choisir librement mon orientation de recherche et de la suivre avec conviction. Lorsque j'ai décidé de me consacrer à la morphologie, elle m'a témoigné une réelle compréhension, notamment en rédigeant une lettre de recommandation déterminante, grâce à laquelle j'ai pu accéder au doctorat. Sa pensée académique et sa vision de l'enseignement ont eu une influence profonde sur moi, et continuent encore aujourd'hui à nourrir ma réflexion de manière subtile. Je lui suis profondément reconnaissant pour le soutien et les encouragements qu'elle m'a apportés à cette époque.

Je remercie tout particulièrement ma mère, qui m'a toujours soutenu, encouragé et accompagné dans tous les aspects de ma vie depuis ma naissance. Sa présence a été une force essentielle tout au long de ce parcours. Je remercie également mon père pour l'influence profonde et le soutien qu'il m'a apportés dans les premières années de ma vie. Il reste pour moi une source de force intérieure et de réconfort. Je suis également reconnaissant envers les autres membres de ma famille, notamment mes frères, mes sœurs, mes oncles et mes tantes, pour leur compréhension, leur attention et leur soutien constants au fil des années.

Je remercie aussi mes amis, qui m'ont encouragé dans les moments les plus difficiles et partagé ma joie dans chaque petit progrès. Leur présence m'a apporté un soutien humain irremplaçable.

Enfin, je remercie le China Scholarship Council (CSC) pour le soutien financier complet qu'il m'a accordé tout au long de mon doctorat, sans lequel ce travail n'aurait pas pu être mené à bien.

Résumé

Cette thèse a pour objectif d'étudier la formation des noms d'instruments par suffixation en français et la compétition entre quatorze suffixes spécialisés dans cette formation (*-eur*, *-euse*, *-rice*, *-oir*, *-oire*, *-ier*, *-ière*, *-et*, *-ette*, *-on*, *-ant*, *-ante*, *-ail*, *-aille*), dans la perspective de la linguistique cognitive. Pour ce faire, nous avons constitué une base de données à partir du *Grand Robert* (2005), comprenant 1338 noms d'instruments construits par ces suffixes et annotés selon neuf variables linguistiques réparties sur cinq niveaux : phonologique, morphologique, syntaxique, sémantique et diachronique. L'analyse a été menée selon une approche mixte, combinant méthodes qualitative et quantitative, afin d'articuler les interprétations linguistiques et les résultats empiriques.

Tout d'abord, en nous appuyant sur la théorie des prototypes en linguistique cognitive ainsi que sur des travaux en philosophie et en linguistique, nous avons proposé une nouvelle définition des noms d'instruments prototypiques. Cette définition nous a permis d'identifier six sous-classes ontologiques prototypiques d'instruments, que nous avons ensuite intégrées dans une proposition de classification structurée. En parallèle, nous avons mené une étude descriptive afin de mettre en évidence les propriétés préférentielles associées à chaque suffixe, et avons mis en œuvre une modélisation par forêt aléatoire pour déterminer les variables et modalités les plus discriminantes susceptibles de prédire le choix des suffixes instrumentaux. Les résultats montrent notamment l'importance des sous-classes ontologiques. En nous appuyant sur la notion de niche écologique, nous avons renouvelé la définition de la niche morphologique et proposé une formule originale pour mesurer le chevauchement entre niches morphologiques. Dans ce cadre, une analyse diachronique a permis d'identifier trois modes de compétition. Nous avons également mis en évidence, à partir de la productivité diachronique, trois phases distinctes (XIIe–XVIe siècle, XVIIe–XVIIIe siècle et XIXe–XXe siècle), et nous avons proposé cinq catégories en fonction de leurs caractéristiques. Enfin, à partir des travaux en *Construction Morphology* (Booij 2010), nous avons proposé un nouveau cadre théorique *Néo-Morphologie Cognitive* (NMC), qui apporte une contribution originale à la linguistique cognitive, en particulier à l'étude de la dérivation morphologique et qui ouvre des pistes de recherche pour de futurs travaux.

Table de matières

Remerciements i

Résumé iii

Liste des Figures vii

Liste des Tableaux ix

Introduction 1

Chapitre 1 Etat de l'art sur les noms d'instrument 7

1.1 La classification des noms 7

1.2 Définitions et approches selon les perspectives sémantiques et morphologiques 11

1.2.1 Définition des noms d'instrument selon les sémanticiens 11

1.2.2 Les noms d'instrument construits selon les morphologues 18

1.3 Etudes des suffixes instrumentaux en français 22

1.3.1 Suffixes instrumentaux 23

1.3.2 Propriétés linguistiques étudiées 25

1.3.3 Relation entre les suffixes et les sous-classes ontologiques 33

1.4 Critiques 37

Chapitre 2 La compétition morphologique 41

2.1 La compétition écologique 41

2.2 Niche morphologique 43

2.2.1 Lindsay & Aronoff (2013) 43

2.2.2 Missud (2023) 46

2.3 Deux approches influentes pour la compétition morphologique 48

2.3.1 Approche monofactorielle 49

2.3.2 Approche multifactorielle 55

2.4 Critiques 59

Chapitre 3 Cadres théoriques 64

3.1 Fondements théoriques et nouveau modèle proposé 65

3.1.1 Cadre théorique de départ : *Morphologie de Construction* 65

3.1.2 Théories d'inspiration 74

3.1.3 Propositions pour une Néo-Morphologie Cognitive 85

3.2 Définition et classification proposées des noms d'instruments 89
3.2.1 Dans la perspective ontologique 89
3.2.2 Dans la perspective linguistique 96
3.2.3 Classification des sous-classes ontologiques 113
Chapitre 4 Méthodologies 117
4.1 Méthodologie de constitution de la base des données 118
4.1.1 Constitution des listes des entrées lexicales retenues 119
4.1.2 Constitution de la base de données 121
4.1.3 Organisation de la base de données 122
4.2 Méthodologie d'annotation de la base de données 124
4.2.1 Au niveau phonologique 124
4.2.2 Au niveau morphologique 125
4.2.3 Au niveau syntaxique : catégorie syntaxique de la base du dérivé 128
4.2.4 Au niveau sémantique 129
4.2.5 Au niveau diachronique 133
4.3 Analyses et Modélisations statistiques 134
4.3.1 Tests de Chi carré et Test de Fisher 134
4.3.2 UpsetR 136
4.3.3 Forêt aléatoire 137
4.3.4 Formule pour mesurer le chevauchement des niches 140
Chapitre 5 Construction des noms d'instruments 142
5.1 Au niveau syntaxique 143
5.2 Au niveau phonologique 144
5.2.1 Taille du radical de la base du dérivé 144
5.2.2 Segment phonologique final du radical 147
5.3 Au niveau morphologique 152
5.3.1 Classe flexionnelle du verbe de la base du dérivé 153
5.3.2 Structure morphologique de la base du dérivé 154
5.3.3 Espaces thématiques 165
5.4 Au niveau sémantique 170
5.4.1 Profil sémantique assumé par la base du dérivé 170
5.4.2 Sous-classes ontologiques du dérivé 176
5.5 Au niveau diachronique 180
5.6 Discussion 183

Chapitre 6 Compétition morphologique entre schémas suffixaux instrumentaux 189

6.1 Prétraitement et regroupement des modalités 190

6.2 Panorama des niches morphologiques 193

6.2.1 Définitions de niche et de sous-niche morphologiques proposées ... 194

6.2.2 Chevauchement de niches morphologiques 199

6.2.3 Visualisation de niche et de sous-niche morphologiques 202

6.3 Mode de compétition morphologique 210

6.3.1 Compétition absolue 212

6.3.2 Compétition alternative 214

6.3.3 Compétition non-absolue 216

6.4 Modélisations statistiques 221

6.4.1 Modèle Général 221

6.4.2 Sous-modèles 227

6.5 Discussion 237

Chapitre 7 Mécanismes et Représentations lexicales de la Néo-Morphologie Cognitive 242

7.1 Principes architecturaux du modèle 242

7.1.1 Module de profilage conceptuel 243

7.1.2 Module de restructuration métonymique 245

7.1.3 Trajectoire cognitive 246

7.1.4 Espaces d'intégration morphosémantique 247

7.1.5 Système morphologique 248

7.2 Représentation lexicale de NMC 250

7.2.1 Architecture de la représentation lexicale en NMC 250

7.2.2 Stratégies fondées sur les bases verbales 251

7.2.3 Stratégies fondées sur les bases non-verbales 254

Conclusion générale et perspectives 259

Annexes 266

Bibliographie 295

Liste des Figures

3.1 Organisation hiérarchique des constructions dérivationnelles, articulée en plusieurs niveaux d'abstraction....68
3.2 Représentation lexicale du lexème PRESENTATEUR....70
3.3 Structure d'argument du lexème α....79
3.4 Réseau de l'intégration conceptuelle....84
3.5 Schéma du modèle cognitif du processus dérivationnel de NMC....87
3.6 Arbre de classification ontologique des noms d'instruments....116
4.1 Recherche avancée dans *Le Grand Robert* 2023....120
4.2 Exemple de la visualisation effectuée par UpsetR....137
4.3 Exemple simple de l'arbre de décision....138
6.1 Schéma représentant toutes les variables linguistiques avec leurs modalités possibles....195
6.2 Niche morphologique d'un hypothétique schéma morphologique n°1....196
6.3 Niche morphologique du schéma morphologique n°2....197
6.4 Chevauchement des niches morphologique n°1 et n°2....198
6.5 La niche morphologique du suffixe -*eur* et ses sous-niches....205
6.6 La niche morphologique du suffixe -*euse* et ses sous-niches....207
6.7 La niche morphologique du suffixe -*oir* et ses sous-niches....208
6.8 La niche morphologique du suffixe -*ier* et ses sous-niches....209
6.9 Compétition absolue dans la sous-classe ontologique Appareil....213
6.10 Compétition alternative dans la sous-classe ontologique Outil....214
6.11 Compétition alternative dans la sous-classe ontologique Machine....215
6.12 Compétition non-absolue dans la sous-classe ontologique Instrument de captation....217
6.13 Compétition non-absolue dans la sous-classe ontologique Récipient....218
6.14 Compétition absolue dans la sous-classe ontologique Substance chimique....219
6.15 Importance des variables dans le modèle de forêt aléatoire (MODÈLE GÉNÉRAL), selon la mesure de *mean decrease in accuracy*....224
6.16 Importance des variables dans le modèle de forêt aléatoire (Sous-modèle 1),

selon la mesure de *mean decrease in accuracy* 229
6.17 Importance des variables linguistiques dans la prédiction des quatre suffixes instrumentaux (Sous-modèle 1) 232
6.18 Importance des variables dans le modèle de forêt aléatoire (Sous-modèle 2) 234
6.19 Importance des variables linguistiques dans la prédiction des quatre schémas instrumentaux, selon le Sous-modèle 2 236
7.1 Espaces d'intégration morphosémantique 247
7.2 La représentation du lexème EPLUCHEUR 253
7.3 La représentation du lexème CAFETIERE 255
7.4 La représentation lexicale du lexème CHIMIQUIER.....257

Liste des Tableaux

1.1 Les suffixes instrumentaux en français 23
1.2 Sous-classes des noms d'instruments 34
1.3 Sous-classification ontologique des instruments selon les suffixes 35
3.1 Tests d'identification Instrument 100
3.2 Résultat des tests pour les noms des Outils 103
3.3 Résultat des tests pour les noms des Instruments de captation 105
3.4 Résultat des tests pour les noms des Appareils 107
3.5 Résultat des tests pour les noms des Machines 109
3.6 Résultat des tests pour les noms des Récipients 111
3.7 Résultat des tests pour les noms des Substances chimiques 113
4.1 Nombre d'entrée totales et sélectionnées pour les suffixes étudiés 121
4.2 Exemple de la base de données annotées 123
4.3 Distribution des radicaux dans les paradigmes pour 8 verbes français 127
4.4 Thèmes nominaux et leur utilisation dans la dérivation 128
4.5 Inventaire de rôles sémantiques et leur définition 131
5.1 Proportion des catégories syntaxiques par suffixes instrumentaux 144
5.2 Distribution de la taille syllabique du radical de la base du dérivé selon les suffixes instrumentaux (en pourcentage) 145
5.3 Catégorisation des schémas en fonction des types de syllabes 147
5.4 Mode d'articulation des segments phonologiques finaux du radical 148
5.5 Segment phonologique final du radical 149
5.6 Les phonèmes préférentiels terminant les radicaux selon les suffixes 149
5.7 Les phonèmes finaux des radicaux qui ne sont pas choisis par les suffixes 151
5.8 Classification des schémas selon les types de conditions linguistiques 152
5.9 Distribution de la classe flexionnelle préférentielle du verbe de la base du dérivé selon les suffixes 153
5.10 Distribution de la structure morphologique des bases selon les suffixes formant des noms d'instrument 155
5.11 Proportion des différentes catégories de lexèmes simples selon les schémas 157

5.12 Répartition des suffixes instrumentaux selon les catégories syntaxiques et les structures morphologiques des lexèmes.....157
5.13 Proportion des procédés morphologiques de dérivation.....159
5.14 Proportion des préfixations par suffixes instrumentaux 159
5.15 Proportion des suffixations par suffixes instrumentaux.....161
5.16 Proportion des conversions par suffixes instrumentaux 164
5.17 Proportion des compositions par suffixes instrumentaux 165
5.18 Proportion des thèmes verbaux par suffixes instrumentaux 167
5.19 Proportion des thèmes nominaux et adjectivaux par suffixes instrumentaux 169
5.20 Proportion des verbes dans la Structure de Qualia par suffixes instrumentaux 171
5.21 Proportion des rôles sémantiques et des projections cognitives par suffixes instrumentaux.....173
5.22 Proportion des rôles sémantiques et des projections cognitives par suffixes instrumentaux.....177
5.23 Sous-classes ontologiques préférentielles par suffixes instrumentaux 177
5.24 Proportion des noms d'Instrument de captation par suffixes.....179
5.25 Proportion des noms de récipient par suffixes.....179
5.26 Répartition diachronique de la productivité des suffixes instrumentaux (XIIe–XXe siècle) 180
5.27 Catégorisation diachronique des suffixes instrumentaux selon leur productivité et leur évolution diachronique 182
6.1 Classement de nombre de modalités activées selon les schémas morphologiques.....200
6.2 Chevauchement directionnel des niches morphologiques entre les suffixes morphologiques instrumentaux du français 201
6.3 Le nombre de sous-niches et le pourcentage selon les schémas instrumentaux 203
6.4 Le nombre de lexèmes et leur pourcentage selon les schémas 222
6.5 Matrice de confusion du modèle final de la forêt aléatoire Modèle Général basée sur les propriétés linguistiques. Les vrais positifs sont surlignés en vert 227
6.6 Matrice de confusion du modèle final **(Sous-modèle 1)**, fondé sur les

propriétés linguistiques 231
6.7 Matrice de confusion du modèle final de la forêt aléatoire Sous-modèle 2 basée sur les propriétés linguistiques 236

Introduction

Ce travail doctoral a pour objectif d'analyser, de manière à la fois quantitative et qualitative, la suffixation des noms d'instruments ainsi que leur compétition, dans une perspective de linguistique cognitive. Plus précisément, l'analyse quantitative sert principalement à décrire la répartition et les préférences des suffixes instrumentaux en fonction de leurs propriétés linguistiques, tout en permettant, lorsque cela s'avère pertinent, de tester statistiquement l'existence de relations significatives entre les différentes variables. L'approche qualitative complète cette démarche en interprétant les résultats quantitatifs à la lumière des contraintes linguistiques, lesquelles influencent notamment le choix des suffixes. Enfin, le choix d'une perspective cognitive se justifie principalement par deux raisons. Premièrement, la notion de compétition morphologique repose sur un modèle *usage-based* (Bybee 2006), en accord avec les principes de la linguistique cognitive. En effet, comme le souligne Goldberg (2019 : 74 ; 77), les formes les plus fréquentes sont également les plus accessibles et privilégiées, un principe valide aussi pour la compétition morphologique dans le domaine de morphologie dérivationnelle. Deuxièmement, cette perspective offre une voie explicative permettant de saisir les motivations cognitives qui sous-tendent la sélection de la base et du dérivé (Schmid 2016 : 103). Ainsi, la combinaison des approches quantitatives et qualitatives, dans la perspective cognitive, se révèle particulièrement adaptée pour étudier la formation des noms d'instruments par suffixation et leur compétition morphologique.

La formation des noms d'instruments par suffixation concerne la formation des noms instrumentaux par l'ajout de suffixes à valeur instrumentale à une base morphologique. D'après la littérature existante, nous avons recensé quatorze suffixes instrumentaux en français (1) : *-ail, -aille, -ant, -ante, -ier, -ière, -et, -ette, -eur, -euse, -oir, -oire, -on, -rice.* Ces suffixes ont fait l'objet de nombreux travaux (Sliwa 1990 ; Roché 1998 ; Dal 1999 ; Namer & Villoing 2008 ; Ferret & Villoing 2015 ; Huyghe & Tribout 2015).

(1)

a. RAPPORTER $_V$ > RAPPORTEUR $_N$

b. TONDRE $_V$ > TONDEUSE $_N$

c. CALCULER $_V$ > CALCULATRICE $_N$

d. RASER $_V$ > RASOIR $_N$

e. PASSER $_V$ > PASSOIRE $_N$

f. SOUFFLER $_V$ > SOUFFLET $_N$

g. SONNER $_V$ > SONNETTE $_N$

h. LORGNER $_V$ > LORGNON $_N$

i. DESHERBER $_V$ > DESHERBANT $_N$

j. IMPRIMER $_V$ > IMPRIMANTE $_N$

k. EVENTER $_V$ > EVENTAIL $_N$

l. SONNER $_V$ > SONNAILLE $_N$

m. PLAFOND $_N$ > PLAFONNIER $_N$

n. CAFE $_N$ > CAFETIERE $_N$

Les travaux existants se concentrent principalement sur les propriétés sémantiques de la base du dérivé et du dérivé (Dubois 1962 ; Sliwa 1990 ; Spence 1990 ; Dal 1999 ; Rainer 2005 ; Namer & Villoing 2008 ; Rosenberg 2008 ; Ferret & Villoing 2015 ; Huyghe & Tribout 2015 ; Lin 2021 ; Rainer 2021). Plus précisément, la plupart de ces études portent sur la polysémie des suffixes et sur les sous-classes ontologiques qu'ils tendent à privilégier. Par exemple, Namer & Villoing (2008) ont élaboré des tests linguistiques permettant de distinguer les noms d'instruments et les noms de lieux construits avec les suffixes *-oir*/*-oire*. En ce qui concerne les affinités entre suffixes et sous-classes ontologiques, plusieurs auteurs ont montré que certains suffixes présentent des préférences marquées. A titre d'exemple, selon Rainer (2005), le suffixe *-eur* dérive le plus souvent des noms d'appareils, tandis que le suffixe *-euse* est plutôt associé à des noms de machines.

Bien que les travaux mentionnés se révèlent fructueux, plusieurs problématiques demeurent. Tout d'abord, comme le soulignent Bauer et al. (2013) et Plag (2018), la suffixation est sensible à plusieurs niveaux linguistiques : phonologique, morphologique, syntaxique et sémantique. Or, les études existantes se concentrent principalement sur la dimension sémantique. Il semble donc que les autres dimensions linguistiques soient encore peu explorées. De plus, s'agissant de la relation entre les suffixes instrumentaux et les sous-classes ontologiques qu'ils dénotent, les résultats des études antérieures apparaissent souvent divergents. Par exemple, Dubois (1962), Rainer (2005, 2021) et moi-même (Lin 2021) soutenons que les noms de machine sont généralement construits avec le suffixe *-euse*. En revanche, Spence (1990) conteste cette idée, estimant que les noms de machine ne se limitent pas à la suffixation en *-euse*. Par exemple, les suffixes *-eur* et *-ier* permettent aussi de dériver les noms de machine (APLATIR > APLATISSEUR ; TUNNEL > TUNNELIER). Ces divergences montrent qu'il

manque à ce jour une classification claire des sous-classes ontologiques associées aux noms d'instruments, ainsi qu'une définition solide et partagée de ce que sont les noms d'instruments prototypiques.

En outre, dans les études précédemment évoquées, les cadres théoriques adoptés — tels que la morphologie lexématique (Aronoff 1994) ou la Morphologie de la Construction (Booij 2010) — se focalisent le plus souvent sur la relation entre une base existante et le dérivé, sans prendre en compte l'origine ou la dynamique de la sélection de la base. Par conséquent, celle-ci est envisagée comme donnée et statique. Or, cette vision s'avère réductrice, car les bases des dérivés instrumentaux sont en réalité plus hétérogènes et dynamiques que ne le supposent ces théories évoquées. Par exemple, pour les noms dénominaux, leur représentation sémantique inclut des verbes sous-jacents, ce qui permet d'établir une structure argumentale. Dans cette structure, la base nominale du dérivé peut occuper un rôle sémantique spécifique. Or, selon les cas, différentes bases peuvent être associées à des rôles sémantiques distincts. Dans les cas de PLAFONNIER et CAFETIERE, leurs verbes sous-jacents sont respectivement FIXER et INFUSER. Ainsi, pour les bases PLAFOND et CAFE, en relation avec le verbe sous-jacent du dérivé évoqué, n'occupent pas le même rôle sémantique : PLAFOND occupe le rôle de Lieu (*Le plafonnier est fixé au plafond pour éclairer la chambre*), tandis que CAFE correspond à un Patient (*On infuse du café avec la cafetière*). Cela reflète une différence cognitive importante dans la sélection des bases, qui ne peut être pleinement expliquée par les théories existantes reconnues.

Enfin, comme l'illustre l'exemple (1), les quatorze suffixes recensés peuvent tous dériver des noms d'instruments. Cette situation constitue donc une véritable compétition morphologique entre les suffixes instrumentaux (au sens de Lieber 2021). En effet, pour la construction d'un nouveau lexème, un seul suffixe est sélectionné parmi les possibles. Par exemple, pour désigner le récipient destiné à ARROSER, pourquoi le suffixe *-oir* a-t-il été sélectionné pour former ARROSOIR, plutôt que *-eur*, *-euse*, *-ier* ou *-ière* ? Il convient donc d'analyser plus en détail cette compétition afin d'identifier le(s) variable(s) les plus discriminante(s) dans le choix des suffixes instrumentaux en français.

À partir de ces problématiques, il émerge une problématique générale : dans une perspective cognitive, quels sont les noms d'instruments prototypiques et quelles propriétés linguistiques président à leur formation par suffixation en français ? Pour y répondre, les questions de recherche suivantes sont posées :

1. Quelles propriétés linguistiques sont privilégiées par chacun des suffixes instrumentaux spécifiques ?
2. Quels sont les noms d'instruments prototypiques ? Et quelles sous-classes ontologiques peut-on considérer comme prototypiques pour ces noms ?
3. Dans une perspective cognitive, quelle est la relation cognitive entre la base et le dérivé, et comment cette relation peut-elle être interprétée dans un cadre théorique cohérent ?
4. Dans le cadre de la compétition morphologique, quelles variables linguistiques s'avèrent les plus discriminantes pour différencier les suffixes concurrents ?

Dans cette thèse, nous proposons une nouvelle définition des noms d'instruments, dans une perspective intégrative combinant les approches ontologique et linguistique. Les noms d'instruments désignent un artefact qu'utilise un agent (directement ou indirectement via une énergie extérieure - naturelle, électrique, chimique) pour mener une action relativement à une entité. Cette entité peut être considérée comme un patient si elle est affectée, comme un thème ou un stimulus si elle ne l'est pas (permettant ainsi d'engendrer une nouvelle entité relative), ou comme une cible si elle est déplacée. Nous nous inscrivons dans l'approche de la morphologie lexématique (Aronoff 1994) ainsi que dans le cadre de la linguistique cognitive pour étudier la suffixation des noms d'instruments. Nous avons choisi le dictionnaire *Le Grand Robert* (2005) comme source de données, et nous analysons 1338 noms d'instruments construits par 14 suffixes instrumentaux évoqués. Ces noms sont annotés selon 8 variables linguistiques réparties sur 4 niveaux linguistiques (phonologie, morphologie, syntaxe et sémantique). Ces quatre niveaux reflètent les dimensions que Bauer et al. (2013) et Plag (2018) considèrent comme pertinentes pour la suffixation.

Par ailleurs, pour aborder la compétition morphologique entre les suffixes instrumentaux en jeu, nous avons d'abord adopté la perspective proposée par Lindsay et Aronoff (2013). Ces auteurs introduisent la notion de « niche » en morphologie afin d'expliquer les phénomènes de compétition entre affixes. Sur cette base, Missud (2023) a développé cette définition dans le cadre d'une approche multifactorielle. Selon cette approche, diverses contraintes linguistiques interagissent pour conditionner le choix des suffixes. En nous inspirant également des travaux de Hutchinson (1957) et de Hurlbert (1978), nous proposons de nouvelles définitions de la niche morphologique et de la sous-niche, dont l'étude permet de mieux comprendre comment la niche

morphologique d'un schéma se construit à partir de modalités activées dans des lexèmes construits. Nous proposons également une formule permettant de mesurer le chevauchement entre niches et de les visualiser. Enfin, pour identifier les variables discriminantes pour prédire le choix des suffixes, depuis une décennie récente, plusieurs auteurs (Arndt-Lappe 2014 ; Bonami & Thuilier 2019 ; Missud 2023 ; Bobkova 2023 ; Thuilier et al. 2023) ont utilisé les outils statistiques, tels que la régression logistique, la forêt aléatoire en modélisant les données pour prédire le choix. Cette approche a abouti à une réussite remarquable. Nous nous inscrivons dans la lignée de ces auteurs et nous adoptons cette approche en utilisant la forêt aléatoire pour prédire le choix des suffixes instrumentaux.

Cette thèse est structurée en sept chapitres principaux. Le chapitre 1 est consacré à l'état de l'art sur les noms d'instruments. Nous y présentons les principales études portant sur leur définition ainsi que sur les suffixes instrumentaux évoqués, ce qui nous conduit à formuler les problématiques et hypothèses de la recherche. Le chapitre 2 porte sur la littérature relative à la compétition morphologique. Nous y introduisons la notion de niche morphologique, puis passons en revue les différentes approches mobilisées pour analyser la compétition entre suffixes. Le chapitre 3 est consacré aux cadres théoriques. Dans un premier temps, nous présentons la Morphologie de Construction (CM). En analysant de manière critique ses limites, nous intégrons ensuite des apports théoriques complémentaires afin d'y remédier. Sur cette base, nous proposons la Néo-Morphologie Cognitive (NMC), un nouveau cadre théorique de la dérivation fondé sur les principes de la linguistique cognitive. Par ailleurs, ce chapitre comprend une redéfinition des noms d'instruments prototypiques, ainsi qu'une identification et classification des sous-classes ontologiques associées.

Le chapitre 4 est consacré à la méthodologie. Il présente les méthodologies de constitution et d'annotation de la base de données, ainsi que les outils statistiques et de visualisation mobilisés. Le chapitre 5 propose une analyse qualitative approfondie, complétée par des analyses descriptives quantifiées. Nous y analysons d'abord les propriétés linguistiques préférentielles des suffixes étudiés en synchronie, puis nous examinons leur évolution diachronique. Le chapitre 6 est centré sur la compétition morphologique. Nous y formulons les définitions de niche et de sous-niche morphologiques, et utilisons une formule que nous proposons pour mesurer le chevauchement entre les suffixes. Une analyse diachronique est également menée pour explorer les modes de compétition. Le chapitre se conclut par une modélisation par

forêt aléatoire, destinée à identifier les variables et modalités linguistiques les plus discriminantes entre les suffixes concurrents. Enfin, le chapitre 7 porte sur les mécanismes et les représentations lexicales de la Néo-Morphologie Cognitive en vue d'illustrer comment fonctionne la NMC afin d'interpréter la formation des noms d'instruments par suffixation dans la perspective cognitive.

Chapitre 1 Etat de l'art sur les noms d'instrument

Les noms d'instruments occupent une place cruciale dans les recherches menées en sémantique et en morphologie. En sémantique, ils peuvent être envisagés soit comme un type nominal, à l'instar d'une sous-classe d'instrument des noms d'artefact, soit comme un argument syntaxique du prédicat, c'est-à-dire comme un rôle sémantique d'instrument. En fonction de la structure morphologique du dérivé, les noms d'instruments peuvent être construits par divers procédés morphologiques tels que la suffixation, la conversion, la composition, etc. Dans ce chapitre, nous présenterons les définitions des noms d'instruments et les travaux effectués en sémantique et en morphologie. Avant d'aborder les définitions des noms d'instruments, nous commencerons par une classification des noms afin de mieux comprendre les notions relatives aux noms d'instruments (§1.1). Ensuite, nous montrerons les définitions des noms d'instruments selon les sémanticiens et les recherches menées (§1.2). De surcroît, nous présenterons les noms d'instruments construits morphologiquement selon les morphologues et les approches employées destinées à analyser les noms construits en morphologie (§1.3). Pour terminer, nous examinerons les études portant sur les suffixes formateurs de noms d'instruments en français (§1.4).

1.1 La classification des noms

En tant que catégorie syntaxique (ou lexicale), le nom occupe une place essentielle dans la langue. Du point de vue de la typologie linguistique, il constitue une catégorie syntaxique répandue dans la majorité des langues parmi lesquelles le système lexical ne se limite pas à une simple dichotomie (Jin 2011 : 111). Morphosyntaxiquement, en français, « les noms ont, un genre, masculin ou féminin ; dans l'énoncé, ils sont en outre spécifiés pour le nombre, singulier ou pluriel » (Abeillé et al. 2021 : 367). Sémantiquement, les noms se divisent en deux groupes : les noms communs et les noms propres. Précisément, « les noms communs permettent de décrire des entités, c'est-à-dire tout ce dont on parle : êtres vivants ou non, objets concrets ou

abstraits, réels ou imaginaires, etc » (Abeillé et al. 2021 : 398). En revanche, les noms propres désignent « une entité unique identifiable par les interlocuteurs dans le contexte » (Abeillé et al. 2021 : 428). Dans cette thèse, l'accent étant particulièrement mis sur la dimension sémantique, en particulier sur les noms d'instruments, il convient donc d'approfondir les noms communs sur ce plan.

Van de Velde (1995), influencé par Husserl, introduit les distinctions entre noms de matière et noms abstraits à travers deux hypothèses majeures. Premièrement, elle postule une corrélation entre l'indépendance ou la dépendance des objets et leur caractère homogène ou hétérogène. L'indépendance signifie que l'objet existe comme un individu autonome, tandis que la dépendance implique que la matière n'existe qu'en tant que partie d'un ensemble ; l'homogénéité signifie que toute partie de l'objet est identique à l'ensemble, tandis que l'hétérogénéité implique que l'objet comporte des compositions distinctes. Par exemple, une bague est identifiée comme objet indépendant, alors que l'or reste une matière dépendante. Deuxièmement, elle avance l'idée que les objets concrets ne présentent pas le même degré d'indépendance : une table est totalement indépendante, alors que l'huile devient indépendante seulement lorsqu'elle est conditionnée (par exemple, dans une bouteille de 500 ml). En s'appuyant sur Van de Velde (1995), Flaux et Van de Velde (2000) explorent les propriétés linguistiques (syntaxiques, sémantiques, morphologiques) pour établir une classification purement linguistique des noms. À partir de six paires de critères (concret/abstrait, extensif/intensif, dénombrable/indénombrable, animé/inanimé, humain/non humain, naturel/fabriqué), elles parviennent à identifier différentes classes de noms (*par ex.* : noms d'objets naturels, fabriqués, collectifs, de qualité, de sentiments ou d'état).

Dans cette classification, les noms d'objets fabriqués, particulièrement pertinents pour notre thèse, sont subdivisés en objets physiques (outils, récipients, bâtiments) et non physiques (objets langagiers et musicaux). Pour les objets fabriqués physiques, les auteurs dégagent trois structures syntactico-sémantiques spécifiques pour identifier les noms d'artefacts : l'indication de la finalité par un complément en « à » (*fil à couper le beurre*), la présence potentielle d'un complément thématique agentif (*Cette table faite par Luc*) et le complément de matière introduit par « en » (*Table en bois fabriquée par Jacques*). À l'inverse, les noms d'objets fabriqués non physiques, dénombrables, ne sont pas liés à l'espace mais au temps, permettant l'usage d'un complément temporel (*dix minutes de sonate*) tout en excluant une quantification précise (Flaux & Van de

Velde 2000 : 51).

Dans la même ligne d'analyse syntaxique, Cadiot (1991) examine la distinction prépositionnelle entre « à » et « avec » dans les constructions instrumentales. Il montre que les contraintes morphosyntaxiques liées aux déterminants varient selon ces prépositions : « à » n'autorise que l'article défini singulier, tandis que « avec » permet une plus grande liberté déterminative (*abattre à la hache* vs *avec une hache*).

Bien que les tests proposés par Cadiot (1991), Van de Velde (1995) et Flaux & Van de Velde (2000) s'appuient davantage sur l'intuition des locuteurs que sur des corpus à grande échelle, ils ont fourni une base précieuse pour des travaux ultérieurs. Ainsi, de nouveaux tests linguistiques ont émergé pour mieux identifier certaines classes spécifiques, notamment les noms d'instruments et les noms d'agent (Namer & Villoing 2008 ; Huyghe & Tribout 2015) (voir le chapitre 3 pour plus de détails). Dans cette lignée, Haas et al. (2022) proposent une série de tests distributionnels destinés à classer sémantiquement les noms en trois grands pôles (noms d'entités, noms de situations, et classes intermédiaires). Les auteurs se concentrent en particulier sur l'identification des noms d'artefacts en distinguant artefacts simples et complexes, à travers trois tests spécifiques permettant d'isoler les purs artefacts (non animés, validant uniquement les tests 1) des artefacts complexes (passant plusieurs tests simultanément) (Haas et al. 2022 : 9). Précisément, concernant les artefacts complexes, les noms hybrides de la classe ARTEFACT*INSTITUTION passent simultanément les tests 1 et les tests 2. En revanche, les noms hybrides de la classe ARTEFACT*OBJET-COGNITIF passent les tests 1 et les tests 3. Ces travaux montrent l'importance cruciale des tests distributionnels et des critères syntaxico-sémantiques pour classer les noms de manière fine et cohérente, notamment dans l'objectif précis de cette recherche centrée sur les noms d'instruments.

Tests 1 servant à identifier les noms d'artefact :

(a) N[humain] a {fabriqué / déchiré} Dét N {à tel endroit / à tel moment}

(b) N[humain] a construit Dét N + Compl de matière + Adj de couleur

(c) N[humain] a confectionné Dét N + Compl d'arôme

Les exemples (1), (2) et (3) illustrent respectivement les tests (a), (b) et (c).

(1) Luc a fabriqué cette chaise en 2022. / Luc a déchiré la peinture de Marie hier soir. [Cf. Test (a)]

(2) Les ouvriers ont construit cet incinérateur en acier bleu. [Cf. Test (b)]

(3) Louis a confectionné un gâteau au lait / une glace au miel. [Cf. Test (c)]

Tests 2 servant à identifier les noms d'institution :

(d) Dét N a été fondé en telle année

(e) N[humain] a été nommé à la tête de (Dét) N

Illustrons ces deux tests avec les exemples (4) et (5).

(4) L'UNESCO a été fondée en 1945. [Cf. Test (d)]

(5) Gabriel Attal a été nommé à la tête de l'Hôtel de Matignon le 9 janvier 2024. [1][Cf. Test (e)]

Les noms d'objet cognitif sont identifiés à l'aide des tests 3 :

(f) Dét N selon lequel P est difficile à admettre

(g) N[humain] écoute Dét[plur] N décousus de Vincent

(h) N[humain] a écrit Dét N {intéressant / pertinent}

Les exemples (6) à (8) permettent de mieux comprendre ces tests.

(6) {L'évidence / L'hypothèse / L'idée} selon laquelle l'eau manquera demain sur Terre est difficile à admettre. (Haas et al. 2022 : 66) [Cf. Test (f)]

(7) Marie écoute les {élucubrations / mensonges / propos} décousu(e)s de Vincent. (Haas et al. 2022 : 66) [Cf. Test (g)]

(8) Albert Camus a écrit un roman intéressant, L'étranger. [Cf. Test (h)]

Ces tests indiquent que les linguistes cités utilisent souvent leur intuition combinée à des critères purement linguistiques pour classer les noms d'instrument. Bien que cette méthode ait fait ses preuves, elle présente une certaine limite : tandis que Namer & Villoing (2008) et Huyghe & Tribout (2015) visent spécifiquement des sous-classes telles que les noms d'instrument et d'agent, d'autres travaux se cantonnent généralement à des catégories sémantiques supérieures. En effet, plusieurs sémanticiens se sont attachés à examiner spécifiquement des sous-classes de noms, dont les noms d'instruments, objet principal de cette thèse, et sur lesquels nous nous concentrerons dans la section suivante.

[1] https://fr.wikipedia.org/wiki/Gabriel_Attal consulté le premier décembre 2024.

1.2 Définitions et approches selon les perspectives sémantiques et morphologiques

Dans cette section, nous allons aborder les définitions et les recherches sur les noms d'instruments en sémantique et en morphologie. L'instrument était originalement un concept purement sémantique. De nombreux sémanticiens le définissent sous différents angles. Explorés ultérieurement ou parallèlement par les morphologues, en tant qu'une des nominalisations, les noms d'instruments ont été définis dans les recherches correspondantes en morphologie. A partir de ce fil conducteur, nous commencerons par les définitions et les recherches en sémantique (§1.2.1). Ensuite, nous présenterons les définitions et les recherches proposées par les morphologues (§1.2.2).

1.2.1 Définition des noms d'instrument selon les sémanticiens

Les noms d'instruments jouent un rôle important en sémantique. Afin d'examiner les noms d'instruments, les sémanticiens ont tenté de proposer plusieurs approches, parmi lesquelles l'approche ontologique, l'approche relationnelle syntactico-sémantique et l'approche cognitive sont remarquables. La notion d'instrument relève naturellement de définitions différentes selon les différentes approches. Parmi elles, l'approche ontologique se caractérise par la valorisation de facteurs extralinguistiques. En revanche, l'approche relationnelle syntactico-sémantique est plutôt ancrée sur les relations linguistiques entre le prédicat et ses participants dans une phrase. L'approche cognitive critique, hérite et développe ces deux approches. Concrètement, elle adapte la plupart de la terminologie de l'approche relationnelle syntactico-sémantique et les concepts fondamentaux de l'approche ontologique. Nous explorerons chacune de ces trois approches en détail dans les parties suivantes.

1.2.1.1 L'approche ontologique

Le concept d'ontologie a été proposé pour la première fois par Aristote. En tant qu'une partie de la métaphysique, il s'applique à « l'être en tant qu'être » (Aristote, 1840 : 213). « Les sujets de préoccupation comprennent : (1) quelles entités existent ou sont censées exister ; (2) comment ces entités sont classées dans une hiérarchie, comment elles sont liées les unes aux autres, comment elles sont classées en fonction

de similitudes ou de différences, etc. » (Mei 2017). Ce concept philosophique a exercé une profonde influence sur les disciplines ultérieures (*par. ex.* l'informatique, la psychologie cognitive, la linguistique, etc.). Ainsi, les sémanticiens ont introduit cette perspective et proposé l'approche ontologique pour analyser le sens des mots et le lien entre les mots et les référents qu'ils dénomment (Carter 1998 : 15). Concrètement, les sémanticiens accordent de l'importance aux caractéristiques et à la taxonomie des référents. Pour l'instrument, les sémanticiens le considèrent généralement comme une sous-classe des artefacts et répertorient les sous-classes de l'instrument, telles que l'outil (Riemer 2010 : 274 ; Polguère 2016 : 229).

Avant tout, il est nécessaire d'introduire les définitions de l'artefact. En tant qu'hyponyme de l'artefact, l'instrument hérite des propriétés générales de l'artefact ; or, dans une taxonomie, l'hyperonyme exprime précisément les caractéristiques partagées par ses hyponymes (Geeraerts 2010 : 83 ; Pustejovsky & Batiukova 2019 : 184). Ainsi, comprendre ce qu'est un artefact permet de mieux cerner ce qu'est un instrument. Selon Monod (1970 : 25), un artefact est « un produit de l'activité d'un être vivant ». Van de Velde (2000 : 51) complète cette définition en soulignant que les artefacts présupposent l'existence d'un modèle idéal préalable et qu'ils intègrent dans leur définition les propriétés commandées par leur finalité. Ces deux traits — production intentionnelle et finalité — constituent donc les critères de base auxquels un instrument doit également satisfaire.

Sur cette base, la question se pose : quelles caractéristiques spécifiques permettent de distinguer les instruments des autres types d'artefacts (*par. ex.* aliments, œuvres d'art, décorations) ? Pour y répondre, plusieurs définitions de l'instrument ont été proposées. Nous nous concentrons sur les travaux menées dans le cadre du *Lexique Génératif* (Pustejovsky 1996) et de la morphologie lexicaliste, car ils sont conformes à notre positionnement de cette thèse (voir le Chapitre 3).

Dans le cadre du *Lexique Génératif* (Pustejovsky 1996), Faber (2012 : 125, 162) définit l'instrument comme un objet utilisé pour réaliser une action, délimité par des propriétés liées à la manipulation, à la fonction et au résultat. Wu et Yuan (2019 : 63) précisent que l'instrument est un objet dont dépend une action ou un comportement. D'ailleurs, du point de vue du *Lexique Génératif,* l'idée d'« action à effectuer » se manifeste généralement par un verbe. Pustejovsky (1996), en s'inspirant de la théorie aristotélicienne des quatre causes, a proposé la Structure de Qualia, qui distingue quatre rôles : formel, constitutif, agentif et télique (Geeraerts 2010 : 150) (voir chapitre 3 pour

les détails). En ce qui concerne les noms d'instruments, Faber (2012 : 155) soutient que la définition de l'instrument repose principalement sur les rôles formel et télique, dans la mesure où l'instrument est mieux défini en termes de fonction ou de finalité. En effet, le rôle formel correspond à ce qui distingue l'objet dans un domaine plus large, par exemple l'orientation, la taille, la forme, la dimension, la couleur ou la position ; le plus souvent, il s'agit de son type. Le rôle télique, quant à lui, se rapporte à la fonction ou au but de l'objet. Le rôle constitutif peut également être pris en compte, notamment lorsque l'instrument est composé de plusieurs parties fonctionnelles. Il s'agit alors de la relation entre l'objet et ses composantes (matière, poids, éléments constitutifs). Par ailleurs, du point de vue de la sémantique distributionnelle et de la morphologie lexicaliste, Huyghe et Tribout (2015 : 106), quant à eux, le définissent comme « une sous-catégorie d'artefact caractérisée par la description d'une certaine forme de fonctionnalité ».

En résumé, selon l'approche ontologique, un instrument est un objet manipulable, destiné à permettre l'exécution d'une action, et doté de fonctionnalités préalables. La présence de cette fonctionnalité opératoire et d'un lien direct avec une action sont précisément les critères qui le distinguent des autres artefacts. Par exemple, la distinction entre les noms d'objets décoratifs et ceux d'instruments réside dans leur caractère opératoire : les noms d'objets décoratifs renvoient à des artefacts dépourvus de fonctionnalité opératoire, c'est-à-dire qu'ils ne sont pas conçus pour être manipulés afin d'accomplir une action.

Sur le plan de la classification, la classe ontologique "instrument" regroupe elle-même plusieurs sous-classes. D'après Faber (2012 : 125), les outils, machines et dispositifs relèvent tous de cette classe. Wu & Yuan (2019 : 172) précisent que l'instrument comprend des sous-classes telles que : outil, véhicule, arme, meuble, instrument de musique, appareil électrique, article de papeterie, équipement sportif, etc. Chacune de ces sous-classes ontologiques peut à son tour être subdivisée en sous-classes plus spécifiques.

Ainsi, les définitions des noms d'instruments montrent l'importance des facteurs extralinguistiques (finalité, manipulation, domaine d'usage). Dans la partie suivante, nous présenterons une approche complémentaire, à savoir l'approche relationnelle syntactico-sémantique, qui s'appuie à l'inverse sur des critères strictement linguistiques pour définir cette classe de noms.

1.2.1.2 L'approche relationnelle syntactico-sémantique

Contrairement à la grammaire générative transformationnelle de Chomsky (1957), qui relègue la composante sémantique au second plan au profit de la syntaxe, Fillmore rompt avec ce cadre et, dans sa *Grammaire des cas* (*Case Grammar*, 1968), inaugure une approche relationnelle syntaxico-sémantique explorant en profondeur l'interdépendance entre structures syntaxiques et rôles sémantiques. Il a ainsi fourni une liste de cas portant sur les relations sémantiques des prédicats, en particulier verbaux, tels que l'agent, le patient, le thème, l'instrument, le lieu, etc.

Selon cette approche, l'instrument est défini comme « la force ou l'objet inanimé qui intervient causalement dans l'action ou l'état désigné par le verbe »[2] (Fillmore 1968 : 46), et il est censé être inanimé (Cruse 2000 : 283). Certains linguistes (Carnie 2007 ; Pustejovsky & Batiukova 2019) partagent cette vision, en y ajoutant des considérations ontologiques : « l'instrument est un objet avec lequel l'action est effectuée »[3] (Carnie 2007 : 231) ; « l'instrument est une entité utilisée pour effectuer une action » [4] (Pustejovsky & Batiukova 2019 : 29).

D'autres chercheurs ont modifié la définition de Fillmore (1968) en précisant que l'instrument est toujours utilisé par un agent pour accomplir une action. Ainsi, Ježek (2016 : 118) affirme que « l'instrument est requis par l'agent pour effectuer l'action »[5], et L'Homme (2020 : 133) le définit comme « l'argument qui fait référence à l'objet utilisé par un agent pour effectuer une action »[6]. Comme mentionné ci-dessus, Fillmore (1968) a étudié la relation syntaxico-sémantique entre le prédicat et les autres constituants dans une phrase. Les cas sont projetés sous forme de syntagmes au niveau syntaxique, se réalisant comme fonctions syntaxiques telles que sujet, complément ou attribut. D'après Fillmore (1968 : 43), lorsque l'agent ou l'instrument apparaît seul dans la phrase, il peut occuper la fonction de sujet (exemples 9 -10). Toutefois, lorsque l'agent et l'instrument coexistent, le sujet est généralement occupé par l'agent, et non l'instrument.

---

[2] Le texte original : "the case of the inanimate force or object causally involved in the action or state identified by the verb".

[3] Le texte original : "The object with which an action is performed is called the instrument."

[4] Le texte original : "The entity used to perform an action."

[5] Le texte original : "Instrument is required by the agent to perform the action."

[6] Le texte original : "The argument that refers to the object used by an agent to perform an action."

(9) John broke the window. (l'agent *John* occupe la fonction sujet)

(10) A hammer broke the window. (l'instrument *a hammer* occupe la fonction sujet)

(11) John broke the window with a hammer. (*John* joue le cas agent en sujet, *a hammer* occupe la fonction oblique)

Cette position est reprise par Schlesinger (1981 : 241), qui reconnaît une productivité limitée pour la construction transitive avec un instrument en sujet. Il souligne néanmoins que l'instrument est analogue à l'agent, car tous deux sont impliqués dans l'action : « l'instrument "exécute" également une action, de manière analogue à l'agent »[7] (Schlesinger 1981 : 241, repris par Taylor 2003 : 273). Taylor (2003 : 273) précise que « Schlesinger fait l'observation très pertinente que les catégories de rôles sémantiques d'agent, d'instrument, d'expérienceur, etc. ne sont en aucun cas aussi claires que de nombreux exposés de la grammaire des cas (*par ex.* Fillmore 1968) voudraient nous le faire croire. »[8] En somme, selon lui, ces rôles thématiques ne permettent pas toujours d'expliquer certains phénomènes syntaxico-sémantiques. Depuis les années 80, les linguistes cognitivistes ont tenté d'assimiler les deux approches évoquées, à savoir l'approche ontologique et l'approche relationnelle syntactico-sémantique, et se sont efforcés de réinterpréter les phénomènes linguistiques au niveau syntaxico-sémantique dans une perspective cognitive. Nous allons dans la partie suivante (§ 1.2.1.3) présenter cette approche.

1.2.1.3 L'approche cognitive

En linguistique, la distinction entre les connaissances encyclopédiques et les connaissances linguistiques occupe depuis longtemps une place primordiale en sémantique et constitue un sujet incontournable pour les sémanticiens. Comme mentionné dans la partie précédente, les linguistes génératives tentent de marginaliser les connaissances encyclopédiques en supposant que la grammaire est autonome et que la sémantique est secondaire (Geeraerts 2010 : 182). Par contre, selon les linguistes cognitivistes, il n'est pas nécessaire d'établir une frontière stricte entre ces deux types de connaissances, car « la connaissance encyclopédique est étroitement

[7] Le texte original : "the instrument also 'performs' an action, analogous to the agent."

[8] Le texte original : "Schlesinger makes the highly pertinent observation that the semantic role categories of agent, instrument, experiencer, etc. are by no means as clear-cut as many expositions of case grammar (e.g. Fillmore 1968) would have us believe."

liée à la manière dont les mots sont utilisés » (Taylor 2003 : 87). La conceptualisation facilite ainsi la construction du sens en donnant accès aux connaissances encyclopédiques et en favorisant la catégorisation des concepts et des référents (Mei 2017 ; Pustejovsky & Batiukova 2019 : 102).

Par exemple, dans *Le Grand Robert*, le mot « chat » est défini comme un « petit mammifère familier à poil doux, aux yeux oblongs et brillants, à oreilles triangulaires, se nourrissant de petits animaux (traditionnellement, de souris qu'il aime à chasser) et de la nourriture que ses maîtres lui servent ». Dans cette définition, il va de soi que les informations sur le poil, les yeux et les oreilles proviennent directement des référents du monde réel. Ces informations favorisent une définition précise et contribuent à la conceptualisation mentale de « chat ». En conséquence, pour les linguistes cognitivistes, il est essentiel d'intégrer l'approche ontologique à l'approche relationnelle syntactico-sémantique pour étudier le sens des unités linguistiques.

Dans cette perspective, Fillmore (1977,1982) a proposé un nouveau cadre théorique, la *Frame Semantics*. Dans cette théorie, la notion de « frame » joue un rôle fondamental, car elle repose essentiellement sur la description des participants impliqués en termes de rôles sémantiques, précisant le rôle de chacun dans l'événement global et comment cela affecte les autres participants (Pustejovsky & Batiukova 2019 : 166-167). Le « frame » se définit comme « une représentation mentale d'une situation concrète ou d'un état de fait abstrait »[9] (Schmid 2016 : 105). En d'autres termes, « [l]a théorie des frames s'intéresse précisément à la manière dont le langage peut servir à mettre en perspective une conceptualisation sous-jacente du monde : il ne s'agit pas seulement de percevoir le monde à travers des modèles conceptuels, mais aussi de reconnaître que ces modèles peuvent être verbalisés de diverses façons. »[10] (Geeraerts 2010 : 225).

Fillmore conserve la plupart des rôles sémantiques issus de la *Case Grammar* (Fillmore 1968). Toutefois, contrairement à cette dernière, Fillmore (1982) insiste sur le fait que la langue ne peut pas être séparée du « frame », et que toute expression linguistique est toujours liée au « frame » pertinent, activé dans la mémoire humaine (Wang & Wu 2020 : 191). En effet, il existe de nombreux scénarios dans une langue.

[9] Le texte original en anglais : "the scene as a mental representation of a concrete situation or an abstract state of affairs".

[10] Le texte original en anglais : "Frame theory is specifically interested in the way in which language may be used to perspectivize an underlying conceptualization of the world—it is not just that we see the world in terms of conceptual models, but those models may be verbalized in different ways".

Nous prenons un « frame » précis pour illustrer cela. Dans l'exemple (12), le verbe « voler » constitue le verbe principal du scénario de VOL (Pustejovsky & Batiukova 2019 : 167). Dans ce scénario, la définition de vol correspond à des « situations dans lesquelles une Personne prend des Biens d'une Victime ou d'une Source. Les Moyens par lesquels cela est accompli peuvent également être précisés » (Pustejovsky & Batiukova 2019 : 167).

(12) Leslie a VOLÉ une montre avec une longue perche de bambou.

Dans l'événement exprimé par le verbe VOLER, plusieurs éléments lexicaux sont associés : *Leslie* est une personne et remplit le rôle sémantique d'Agent, la *montre* représente les biens et joue le rôle de Thème, tandis que la *longue perche en bambou* constitue le moyen par lequel est accompli le vol et joue le rôle de l'Instrument. Dans cet exemple, l'instrument est un objet utilisé par une personne (l'agent) pour effectuer une action sur une autre entité (Pustejovsky & Batiukova 2019 : 167).

Outre la *Frame Semantics* (Fillmore 1977 ; 1982), une approche cognitive complémentaire a été proposée par Langacker (1987) dans le cadre de la *Grammaire Cognitive*, également développée à partir des deux approches évoquées précédemment. D'un point de vue cognitif, Langacker (1987) a tenté de mettre en évidence des isomorphismes entre les référents du monde réel et les systèmes linguistiques. Cette approche conserve une grande partie de la terminologie proposée par Fillmore (1968) tout en y ajoutant des contraintes ontologiques sur les rôles sémantiques prototypiques. S'enracinant dans la théorie de Fillmore (1968), Langacker (1987) a réinterprété la notion de rôles sémantiques comme un outil de description des structures conceptuelles : plutôt qu'une liste rigide, il s'agit de repères cognitifs souples, activés chaque fois qu'un locuteur se représente ou décrit un événement. Langacker (1987 ; 2008) a également proposé une série de rôles archétypiques qui organisent probablement notre conception des événements. Ces rôles incluent l'agent, le patient, l'instrument, la cible, l'expérienceur et le zéro (Langacker 2008 : 356). Selon Langacker (1987 ; 2008 ; repris 2014 : 216), « le rôle d'instrument correspond à un objet inanimé manipulé par un agent pour affecter un patient ; c'est par l'intermédiaire de cet instrument que l'énergie est transmise de l'agent au patient. La présence d'un instrument médiant ce transfert d'énergie entre l'agent et le patient constitue un trait optionnel des actions

prototypiques. » [11] De cette définition émergent deux rôles sémantiques centraux (agent et patient) ainsi qu'une contrainte ontologique (objet inanimé) essentiels à la définition de l'instrument. Toutefois, Ungerer & Schmid (2006) ont développé cette définition et ajouté la cible comme contrainte ontologique à la définition des noms d'instruments. Ils soulignent que l'instrument peut intervenir dans la chaîne d'agent-cible en tant qu'intermédiaire (2006 : 184), dans l'exemple (13), *à la main* est un instrument intermédiaire. Ils constatent ainsi que la chaîne d'agent-cible implique divers rôles archétypiques tels que l'agent, le thème, la cible et l'instrument.

(13) Luc donne ce cadeau à Louis à la main.

En conséquence, Ungerer & Schmid (2006) estiment qu'il faudrait intégrer le thème et la cible pour préciser la définition de l'instrument. Selon eux, un instrument est donc « une entité utilisée par un agent pour mener une action à une autre entité, servant d'intermédiaire dans le transfert de force ».

Jusqu'à présent, nous avons examiné trois approches influentes selon les perspectives sémantiques : ontologique, relationnelle syntactico-sémantique et cognitive. Dans la section suivante, nous étudierons les noms d'instruments selon les morphologues.

1.2.2 Les noms d'instrument construits selon les morphologues

Du point de vue de leur structure morphologique, les noms d'instruments peuvent être soit simplex, soit construits, par la suffixation, la composition ou la conversion. Précisément, pour ce qui concerne ces schémas morphologiques, les morphologues ont appliqué les approches sémantiques aux analyses des noms d'instruments construits morphologiquement et proposent une approche spécifique, à savoir l'approche sémantique relationnelle (§1.2.2.1) et les approches mixtes (§1.2.2.2).

1.2.2.1 L'approche sémantique relationnelle en morphologie

Certains morphologues (Booij 1986 ; Corbin 1987 ; Sliwa 1990 ; Fradin 2003) ont appliqué *Case Grammar* (Fillmore 1968) à la description de la structure morphologique du dérivé. Dans cette perspective, ils font l'hypothèse selon laquelle les

[11] Le texte original : "the instrument role is that of an inanimate object manipulated by an agent to affect a patient; it is through the instrument that energy is transmitted from the agent to the patient. The presence of an instrument mediating the transfer of energy from agent to patient is an optional feature of prototypical actions.".

rôles sémantiques (ou les fonctions thématiques) informent du type de relation d'un argument par rapport au prédicat (Sliwa 1990 : 58). En d'autres termes, un schéma morphologique a pour fonction de nominaliser l'un des participants sémantiques de la base verbale. Le dérivé hérite du rôle sémantique que ce participant occupe dans l'événement décrit par le verbe de base. (Schlesinger 1979 : 322 ; Booij 1986 : 503 ; Sliwa 1990 : 58 ; Fradin & Winterstein 2013 : 11 ; Bauer et al. 2013 : 38). Ainsi, les morphologues s'intéressent surtout aux noms déverbaux, par exemple la suffixation nominale déverbale, la conversion de verbe à nom et la composition Verbe-Nom. Nous prenons quelques exemples de la suffixation (14)-(17) pour bien illustrer le cas où un argument est nominalisé.

(14) Nom d'agent : ACHETEUR (Huyghe et Tribout 2015 : 104)

(15) Nom de patient : TIROIR (Namer et Villoing 2008 : 1545)

(16) Nom d'instrument : CLIMATISEUR (Huyghe et Tribout 2015 : 104)

(17) Nom de lieu : PROMENOIR (Namer et Villoing 2008 : 1552)

Concernant l'exemple (14), ACHETEUR remplit le rôle sémantique d'agent par rapport à sa base verbale : *l'acheteur achète quelque chose*. Dans l'exemple (15), concernant TIROIR, le suffixe *-oir* nominalise l'argument patient du prédicat verbal de base en remplissant le rôle sémantique de patient : *tirer un tiroir* (Namer et Villoing 2008 : 1545). Pour l'exemple (16), l'argument de la base verbale (CLIMATISER) est nominalisé par le suffixe *-eur* comme le nom d'instrument : *climatiser avec le climatiseur*. Enfin, dans l'exemple (17), PROMENOIR remplit le rôle sémantique de lieu par rapport à sa base verbale : *se promener sur le promenoir*.

Venons-en à la définition des noms d'instruments dans cette approche. D'après Lieber (2021 : 257), « les noms d'instrument sont les arguments du verbe servant à effectuer une action. ». Ainsi, selon cette définition, les noms d'instruments sont définis dans la perspective purement syntaxico-sémantique. A partir de cette définition, en morphologie, les noms d'instruments sont prototypiquement des noms déverbaux, car seuls les noms déverbaux possèdent la base verbale (*par ex.* : RASOIR, EPLUCHEUR, TONDEUSE).

Cette approche tend à traiter comme des noms d'instrument certains noms non-artefacts tels que : BERCEUSE (objet informationnel), ASSURANCE (objet financier), MACHOIRE (partie du corps) (Sliwa 1990 ; Namer & Villoing 2008 ; Huyghe & Tribout 2015). Cela rejoint cependant la notion d'instrument prototypique, dans la mesure où

la fonctionnalité constitue un critère central pour la définition des noms d'instruments.

Par conséquent, afin de mieux délimiter les noms d'instruments prototypiques, les auteurs ont essayé d'ajouter des contraintes ontologiques à la définition des noms d'instruments. Autrement dit, les noms d'instruments prototypiques devraient avant tout être des artefacts. Selon Sliwa (1990 : 59), l' « instrument » désigne l'ensemble d'objets à l'aide desquels l'agent humain exerce une action exprimée par le verbe de base.

En outre, en plus de l'ajout d'une contrainte ontologique à la définition, Namer & Villoing (2008 : 1542) excluent encore les noms de parties du corps. Elles proposent deux propriétés particulières sur les noms d'instruments, à savoir la création humaine et l'usage particulier. Pour les noms déverbaux, « un nom d'instrument dénote un objet concret artefactuel. [...] L'instrument possède deux propriétés : celle de provenir d'une création humaine, et celle d'être destiné à un usage particulier. » (Namer & Villoing 2008 : 1542). Dans ces travaux ultérieurs, toute définition sémantique d'un nom précisera sa valeur ontologique.

Huyghe & Tribout (2015 : 101) partagent l'opinion de Namer & Villoing (2008 : 1542). Les noms d'instruments se définissent comme « un nom déverbal qui dénote l'artefact prototypiquement utilisé pour réaliser l'action décrite par le verbe de base » (Huyghe & Tribout 2015 : 101). Ils ont remarqué que le facteur ontologique joue un rôle prépondérant pour définir les noms d'instruments : « les noms d'instruments constituent un type nominal à prédominance ontologique. Il s'agit de noms d'artefacts qui, sans la condition morphologique, pourraient être inclus dans une hiérarchie des noms d'objets. » (Huyghe et Tribout 2015 : 111). A l'aide de tests linguistiques, ils précisent aussi : « Ces disparités sont déterminées par (i) la nature du type ontologique impliqué et (ii) la spécificité de la prédication actionnelle sous-jacente. Dans le cas des noms d'instruments, la description d'artefacts relègue au second plan la composante prédicative issue de la base morphologique. Les objets étant fabriqués, leur fonctionnalité est prédéterminée, ce qui sature la description nominale et implique le rôle dénominatif des N d'instruments » (Huyghe et Tribout 2015 : 111).

Par ailleurs, Kerleroux (2007) et Bauer et al. (2013) tentent aussi d'étendre les notions de noms d'agent et de nom d'instrument aux noms non-déverbaux construits avec le suffixe *-eur* à l'aide des noms de base (noms événementiels) pour obtenir une interprétation instrumentale. Concrètement, le dérivé a une interprétation instrumentale, parce qu'à partir des noms de base (noms événementiels), la base est affectée ou impliquée par l'agent avec le dérivé. « Lorsque le suffixe *-er* est attaché à des bases

nominales, adjectivales ou adverbiales, les dérivés ainsi peuvent dénoter des noms d'agents ou des noms d'instruments, tout comme les noms déverbaux en *-er*. Dans de tels cas, la relation verbale est implicite dans le contexte ou peut être inférée à partir de la nature de la base non verbale. Ainsi, par exemple, un CONFECTIONER désigne une personne qui fabrique ou vend des confections, un CRICKETER désigne une personne qui joue au cricket, et un FREIGHTER désigne un objet qui transporte du freight. Ces interprétations résultent des types d'activités auxquelles la base nominale pourrait, en théorie, être associée »[12] (Bauer et al. 2013 : 233).

Ainsi, à partir de ces travaux menés, les noms d'instruments sont un terme plutôt sémantico-ontologique, car dans ce cas le facteur ontologique joue un rôle plus central que la dépendance syntactico-sémantique. La composante prédicative issue de la base morphologique ne constitue plus la condition nécessaire pour identifier les noms d'instruments. D'autres morphologues (Roché 1998 ; Bauer et al. 2013) reconnaissent également l'importance du facteur ontologique pour définir les noms d'instruments. Ils ont combiné l'approche ontologique et l'approche relationnelle pour proposer une approche mixte en vue d'interpréter les noms d'instruments. La section suivante est consacrée à la présentation de ces approches mixtes.

1.2.2.2 Les approches mixtes

Roché (1998) et Roché & Plénat (2010) ont intégré l'approche ontologique et l'approche relationnelle pour considérer les noms non-déverbaux comme des noms instruments. Ils utilisent les verbes sous-jacents dans la représentation sémantique de la base du dérivé pour interpréter la relation entre la base et le dérivé dans un procès événementiel. Ces verbes peuvent souvent être déduits de connaissances encyclopédiques relatives au dérivé, dans la mesure où ces dernières sont liées aux caractéristiques ontologiques des référents et participent à la construction des lexèmes.

Roché (1998) a exploré le rapport sémantique fondamental "Ndér contient Nbase" dans ses recherches sur les suffixes *-ier* et *-ière*. Précisément, il réserve l'emploi du terme « instrument » à l'approche sémantique relationnelle et prend en compte les noms d'artefact représentant le rôle sémantique au procès dans la représentation

[12] Le texte original : "Where we find *-er* on nominal, adjectival, or adverbial bases, such forms can denote agents or instruments, just as deverbal -er forms can. In such cases, the verbal relation is implied by context or can be inferred from the nature of the non-verbal base. So for example, a confectioner is someone who makes or sells confections, a cricketer someone who plays cricket, and a freighter something that carries freight. Such interpretations follow from the sort of activities that the base nouns could conceivably be involved in".

sémantique du dérivé. Nous reprenons un exemple de Roché (1998 : 17), concernant le lexème TISONNIER pour bien illustrer cette idée.

(21) On secoue les tisons avec un tisonnier.

Dans l'exemple (21), syntaxiquement, le TISONNIER est introduit toujours avec la préposition *avec*. Sémantiquement, elle représente plutôt l'instrument, non pas en tant que classe d'objets, mais par son rôle par rapport au procès. Le verbe SECOUER assigne le rôle instrument à TISONNIER. Le sens de TUNNELIER, MOULURIERE, FACTURIERE, DENTELLIERE est construit de la même façon. Ces exemples sont étroitement liés à la situation où, dans une approche ontologique, les noms d'instruments dénominaux et désadjectivaux ne disposent pas de base verbale. Dans ces cas, le verbe sous-jacent peut être déduit à partir des connaissances encyclopédiques. Cette observation fournit ainsi un cadre pertinent pour l'analyse des noms d'instruments dont les bases sont nominales ou adjectivales

Il convient de noter que bien que Kerleroux (2007) et Bauer et al. (2013) tentent de chercher un lien pour interpréter la relation entre la base non-verbale et le dérivé instrumental, Roché (1998) prend le terme de rôles sémantiques dans le sens d'argument d'un procès qui est impliqué d'une façon sous-jacente dans le nom d'instrument. En revanche, dans le cas de Kerleroux (2007) et Bauer et al. (2013), il explicite que la valeur événementielle est portée par le nom de base lui-même. C'est un peu différent de ce que propose Roché (1998), car concernant le lexème comme TISONNIER, TISON ne désignera jamais un événement, à la différence de CONFECTION ou CRICKET.

Il va de soi que dans cette approche, le verbe sous-jacent dans la représentation sémantique dans un procès événementiel est aussi important que la base verbale pour étudier les noms suffixés. Les types de noms dépendent souvent des rôles sémantiques nominalisés et prédits par leur prédicat dans la représentation sémantique de la base du dérivé. Naturellement, pour déduire le verbe sous-jacent, il faudrait avoir recours aux connaissances encyclopédiques sur ce dérivé. Par conséquent, le facteur ontologique joue un rôle crucial pour cette approche.

1.3 Etudes des suffixes instrumentaux en français

Dans cette section, nous allons d'abord présenter les suffixes qui dérivent les

noms d'instruments en français tels que présentés dans les travaux de recherche antérieurs (§1.3.1). Nous essaierons ensuite d'analyser les suffixes étudiés selon les propriétés linguistiques afin de connaître les intérêts des morphologues et d'identifier d'éventuelles lacunes. (§1.3.2). Enfin, nous discuterons des travaux concernant la sous-catégorisation ontologique des instruments selon les suffixes (§1.3.3).

1.3.1 Suffixes instrumentaux

Etant donné que dans cette thèse nous intégrerons les approches ontologiques et relationnelles, nous recensons tous les suffixes instrumentaux selon les recherches menées par les morphologues ou lexicologues dans ces deux approches. Il existe une dizaine de suffixes utilisés pour dériver les noms d'instruments en français (Tableau 1.1).

Tableau 1.1 Les suffixes instrumentaux en français

Type	Suffixes	Références
Déverbaux	*-eur/-euse/-rice*	Darmesteter (1877) ; Meyer-Lübke (1895) ; Dubois (1962) ; Benveniste (1975) ; Spence (1990) ; Rainer (2005) ; Fradin (2005) ; Rosenberg (2008 ; 2011) ; Huyghe & Tribout (2015) ; Grevisse & Goosse (2016) ; Wauquier (2020 : 93) ; Lin (2021) ; Rainer (2021)
	-oir/-oire	Dubois (1962) ; Sliwa (1990) ; Namer & Villoing (2008) ; Rosenberg (2008) ; Ferret (2013) ; Fradin & Winterstein (2013) ; Lin (2021) ; Rainer (2021) ; Salvadori & Huyghe (2021)
	-ail/-aille	Rainer (2021) ; Rey & Rey-Debove (2022)
	-on	Salvadori & Huyghe (2022)
	-ant/-ante	Rosenberg (2008 ; 2011) ; Knittel (2017a ; 2017b)
	-age	Ferret & Villoing (2015)
Dénominaux	*-ier/-ière*	Dubois (1962) ; Corbin & Temple (1994) ; Roché (1998) ; Grevisse & Goosse (2016) ; Lin (2021)
Déverbaux/ Dénominaux	*-et/-ette*	Corbin (1986) ; Dal (1999) ; Fradin et al. (2003) ; Rainer (2021)

Selon le Tableau 1.1, nous avons regroupé les suffixes en trois groupes en fonction des schémas morphologiques. Les noms d'instruments sont majoritairement déverbaux. En outre, en ce qui concerne l'intérêt des morphologues, les suffixes *-eur/euse/-rice* sont plus populaires que les autres suffixes, suivis par les suffixes *-oir/-oire* et *-ier/-ière*. Bien que les suffixes *-et/-ette*, *-ant/-ante*, *-on* et *-ail/-aille* aient moins

été étudiés, les morphologues ou lexicologues ont reconnu qu'ils peuvent dériver directement les noms d'instrument dans la perspective onomasiologique (Dal 1999 ; Rainer 2021) (*par ex.* BLERET, SONNETTE, EMULSIFIANT, IMPRIMANTE, LORGNON, SONNAILLE, EVENTAIL). Ainsi, nous allons conserver ces suffixes dans notre thèse. Par contre, bien que selon Ferret & Villoing (2015), les noms en *-age* puissent présenter une interprétation instrumentale, il nous semble que ce sens n'est pas construit par le suffixe, mais le résultat de l'évolution sémantique. A titre d'exemple, concernant ALLUMAGE, son premier sens enregistré dans *Le Grand Robert* (2005) est « action d'allumer un foyer ». Toujours d'après *Le Grand Robert* (2005), à partir de 1866, dans le contexte d'allumer quelque chose avec un instrument, *via* la métonymie conceptuelle, ALLUMAGE acquiert un nouveau sens « Inflammation du mélange gazeux provenant du carburateur d'un moteur à explosion ». De même, pour ATTELAGE, BANDAGE, EMBRAYAGE, GARNISSAGE, GOMMAGE et MAQUILLAGE ont subi les mêmes changements sémantiques que ALLUMAGE. Ce phénomène relève donc davantage de la sémantique que de la morphologie. Ainsi, on exclut le suffixe *-age* de cette thèse.

En outre, d'après Knittel (2017a ; 2017b), les suffixes *-ant*/*-ante* pourraient dériver des noms de substances chimiques dans la perspective ontologique. Bien que le suffixe *-if* permette aussi de former les noms de substances chimiques (DETERSIF, MINORATIF, VOMITIF), ils sont souvent convertis à partir de la base adjectivale. Par exemple, VOMITIF est construit au XIVème siècle comme un adjectif, dont le sens est « qui provoque le vomissement. ». Ainsi, les noms en *-if* dénotant les noms d'instruments sont censés être construits par conversion plutôt que par suffixation, car aucun nom en *-if* a un premier sens instrumental. Bien que de nombreux noms suffixés en *-ant* aient les mêmes lexicalisations que le suffixe *-if,* le suffixe *-ant* permet de dériver, directement et d'une manière autonome sans être converti des adjectifs, les noms de substances chimiques (*par ex.* EMULSIFIER > EMULSIFIANT, ELUER > ELUANT, FLEGMATISER > FLEGMATISANT) depuis le XIXe siècle, et plus particulièrement au XXe siècle où il est en plein épanouissement. Ainsi, on suppose qu'à travers ces lexèmes, les locuteurs pourraient réanalyser les noms en *-ant* et considérer ce suffixe comme un suffixe nominal servant à dériver les substances chimiques. Il faudrait donc inclure ce suffixe dans notre recherche.

En définitive, nous conservons 14 suffixes à étudier : *-ail*, *-aille, eur/-euse/-rice, -oir/-oire, -ier/-ière, -et/-ette, -on, -ant/-ante*. Dans les parties suivantes, nous allons présenter les recherches menées par les morphologues sur ces suffixes.

1.3.2 Propriétés linguistiques étudiées

Dans cette recherche, nous considérerons quatre niveaux linguistiques : la phonologie, la morphologie, la syntaxe et la sémantique. L'analyse sera organisée selon les propriétés linguistiques, celles-ci étant considérées comme des facteurs conditionnant le choix des suffixes (Plag 2018 : 172).

1.3.2.1 Propriétés phonologiques

Roché (1998, 2003) et Plénat (2005) s'intéressent aux contraintes phonologiques qui exercent une influence sur la formation des noms. Ils ont émis l'hypothèse selon laquelle la sélection des suffixes est déterminée par des propriétés phonologiques, telles que la taille du radical du dérivé, la similarité phonologique entre la base et le suffixe, le segment phonologique final du radical du dérivé, etc.

Tout d'abord, en ce qui concerne la taille du radical du dérivé, Roché (1998) a observé que lorsque la base du dérivé est monosyllabique, les suffixes *-ier* et *-ière* ont tendance à recourir à des bases supplétives. De plus, Roché (2003 : 110) a noté que le suffixe *-on* a une propension à sélectionner des bases dissyllabiques. De manière similaire, Plénat (2005 : 8) a constaté que le suffixe *-ette* tend également à s'attacher à des bases dissyllabiques.

En outre, selon Plénat (2005 : 16), le décalage désigne un moyen d'éviter « la rencontre à courte distance de phonèmes identiques ou similaires ». Par exemple, concernant le lexème RONDELETTE, nous observons que sa base est ROND, mais pour éviter la rencontre à courte distance de phonèmes identiques ou similaires, l'interfixe *-el-* est inséré entre le base ROND et le suffixe *-ette*. Cette notion est, en effet, en lien avec la taille du radical, car le décalage augmente la taille du radical par l'ajout de l'interfixe. Toutefois, « le décalage est banni lorsque la base compte trois (vraies) syllabes ou plus, le suffixe *-ette* est exceptionnel (dans le cas des bases en /–t/ ou en /–d/) quand elle est dissyllabique, et variable lorsqu'elle ne comporte qu'une syllabe » (Plénat 2005 : 5). C'est pourquoi le suffixe *-ette* préfère majoritairement les radicaux dissyllabiques.

Pour terminer, quant au segment phonologique final du radical du dérivé, Concernant les suffixes *-ier* et *-ière*, Roché (1998) a constaté que ces suffixes évitent les radicaux de base se terminant par les phonèmes /i/, /ʎ/, /j/ et /ɲ/, car ces phonèmes

partagent des propriétés phonologiques avec le premier segment phonologique des suffixes. « Sur le plan phonologique, on a pu observer que le /i/ constitue un obstacle plus fort que les autres finales vocaliques ; que /ɲ/ est un obstacle plus fort que /ʎ/ ; que le /j/ est plus gênant à l'attaque de la finale qu'à la rime. » De plus, le phénomène de décalage joue un rôle important dans la sélection du segment phonologique final du radical. Pour le suffixe *-on*, Roché (2003 : 99, 110) a observé que les consonnes palatales et prépalatales sont largement représentées devant ce suffixe, en raison du fait que les interfixes se terminent fréquemment par ces segments. Par ailleurs, le suffixe *-ette* privilégie les radicaux dissyllabiques se terminant par des sonantes telles que /n/, /l/, /r/ ou même /j/, ce qui représente environ 75 % des dérivés dissyllabiques (Plénat 2005 : 8). Ce phénomène s'explique également par le fait que l'interfixe se termine fréquemment par les segments /n/ ou /l/.

1.3.2.2 Propriétés morphologiques

Dans cette section, nous présentons les études traitant des genres grammaticaux et des espaces thématiques dans la dérivation des noms d'instruments.

Concernant les espaces thématiques, notamment pour le verbe, le cadre théorique retenu postule que les verbes comprennent un espace thématique à 14 thèmes : les thèmes 1 à 13 (Bonami & Boyé 2003 ; Bonami, Boyé & Kerleroux 2009) ainsi que le thème 0 (Tribout 2010). Parmi ceux-ci, le thème 1 et le thème 13 sont mobilisés pour la suffixation (Bonami, Boyé & Kerleroux 2009 ; Fradin & Winterstein 2013), le thème 3 pour la composition (Villoing 2009), et les thèmes 0, 12 et 13 pour la conversion (Tribout 2010). Des exemples pour chaque thème évoqué sont présentés en (18 - 20).

(18) suffixation :

a. thème 1 : PRESENTER > *présent-able*
b. thème 13 : PRESENTER > *présentat-eur*

(19) composition :

a. thème 3 : LAVER > *lave-linge*

(20) conversion :

a. thème 0 : FINIR > *fin-ir*
b. thème 12 : MONTER > *mont-ée*
c. thème 13 : CERTIFIER > *certificat*

Pour le nom, il existe 3 thèmes, parmi lesquels le thème 2 (*par ex.* PLACARD > PLACARDER) et le thème L servent à la construction des lexèmes (*par ex.* POINT >

PONCTUER) (Plénat 2008, repris par Tribout 2010). D'ailleurs, les suffixes ont leurs propres caractéristiques pour le choix des thèmes. Concrètement, les suffixes *-eur* et *-rice* préfèrent le thème 13) et ont un lien paradigmatique avec le suffixe *-ion* (Anscombre 2001 : 40 ; Bonami, Boyé & Kerleroux 2009 : 121 ; Grevisse & Goosse 2016 : 180) (*par ex.* FORMER > *format-eur* ~ *format-rice* ~ *format-ion*), parce que certains noms construits par ces suffixes sont empruntés du latin. Par contre, selon Anscombre (2001 : 37) et Fradin & Winterstein (2013 : 1), les suffixes *-oir* et *-ant* ont tendance à sélectionner le thème 1 (*par ex.* RASER > *ras-oir* ; DESHERBER > *désherb-ant*). A la différence des deux cas évoqués, le suffixe *-oire* aurait en même temps deux thèmes verbaux préférentiels : le thème 1 et 13 (Fradin et Winterstein 2013 : 4) et un lien paradigmatique avec le suffixe *-ion* (Grevisse 2016 : 184) (*par ex.* BASSINER > *bassin-oire* ; OBSERVER > *observat-oire* ~ *observat-ion*).

En ce qui concerne le genre, les morphologues ne sont pas unanimes sur sa fonction pour la construction des noms. Dubois (1962) considère que pour un suffixe par rapport à la forme masculine, la forme féminine en est la variante. Ils se distinguent pour dériver différentes classes de noms. En revanche, Spence (1990) et Wauquier (2020) ont trouvé que le genre ne permet pas de distinguer les différentes classes des noms entre eux, car ils permettent de dériver les mêmes classes de noms : les noms d'agent et les noms d'instruments. Ainsi, ils soulignent qu'il est nécessaire de les considérer comme des suffixes autonomes.

Selon Dubois (1962 : 44), concernant le suffixe *-eur*, ses formes féminines *-euse* et *-rice* en sont la variante dans la construction des noms. En effet, avant la révolution industrielle, les suffixes *-eur*, *-euse* et *-rice* désignaient principalement des noms d'agents, ce qui explique leur emploi restreint à cette époque. Les formes féminines *-euse* et *-rice* sont utilisées pour indiquer le sexe des agents. Avec la révolution industrielle, a émergé le sens instrumental pour les suffixes *-eur*, *-euse* et *-rice*. D'après cet auteur, les formes féminines sont utilisées pour construire les noms d'instruments afin de les différencier de la forme masculine utilisée pour dériver les agents manipulant ces instruments (*par ex.* DEPANNEUR/DEPANNEUSE et PERFORATEUR/PERFORATRICE). « Le féminin forme marquée, était tout désigné pour indiquer ce qui a été, pendant quelque temps le substitut fréquent (machine) de l'agent (ouvrier) ; lorsque la machine s'est imposée dans la production, la différenciation entre *-eur/-euse/-rice* n'a plus eu la même importance morphologique, et *-eur/-euse* devinrent de simples variantes libres. » (Dubois 1962 : 44)

Cependant, Spence (1990 : 34) n'est pas d'accord avec le point de vue de Dubois (1962). Selon lui, le suffixe *-euse* peut désigner à la fois les noms d'instrument et les noms d'agent manipulant ces instruments (*par ex.* PEIGNEUSE et TRICOTEUSE). Il a proposé qu'il faudrait considérer les formes féminines comme des suffixes autonomes. Précisément, « les dérivés en *-eur* au sens instrumental sont plus fréquents que les dérivés en *-euse* au sens instrumental et il existe encore des noms en *-euse* qui n'ont pas de correspondance en *-eur*. Concrètement, les machines agricoles, textiles et à bois sont généralement désignées par les noms se terminant par le suffixe *-euse* : ARRACHEUSE, EGRENEUSE, FANEUSE, FAUCHEUSE, LIEUSE, LAINEUSE. La plupart des ouvriers du textile étaient des femmes. » (Spence 1990 : 34).

Le résultat de la thèse de Wauquier (2020 : 93) a étayé l'argument de Spence (1990). Selon elle, l'agentivité se manifeste différemment dans les noms déverbaux masculins en *-eur* par rapport à ceux en *-euse* et *-rice*. De plus, ces deux suffixes féminins forment des noms qui présentent des propriétés sémantiques distinctes. (Wauquier 2020 : 113-114)

1.3.2.3 Propriétés syntaxiques

Au niveau syntaxique, deux propriétés sont généralement étudiées par les morphologues : la catégorie syntaxique de la base du dérivé, ainsi que le type de base verbale du point de vue de l'approche argumentale du verbe (Grimshaw 1990). Pour ce qui est de la catégorie syntaxique de la base, nous avons divisé les suffixes à l'étude en deux groupes : ceux qui sélectionnent une base verbale et ceux qui sélectionnent une base nominale. S'agissant des suffixes qui privilégient une base verbale, il existe *-eur*, *-euse*, *-rice*, *-oir*, *-oire*, *-ant*, *-ante*, *-et*, ainsi que *-ette* (Dal 1999 ; Anscombre 2001 ; Namer & Villoing 2008 ; Roy & Soare 2012 ; Huyghe & Tribout 2015 ; Grevisse & Goosse 2016 ; Knittel 2017a ; Knittel 2017b). Par exemple, on trouve des noms tels que RAPPORTER $_{V}$ > RAPPORTEUR $_{N}$, TONDRE $_{V}$ > TONDEUSE $_{N}$, RASER $_{V}$ > RASOIR $_{N}$. Par contre les suffixes *-ier*/*-ière* privilégient la base nominale (Roché 1998). A titre d'exemple, les deux suffixes dérivent des noms, tels que : PLAFOND $_{N}$ > PLAFONNIER $_{N}$, CAFE $_{N}$ > CAFETIERE $_{N}$.

En outre, du point de vue de la structure argumentale du verbe de base (Grimshaw 1990), certains suffixes se différencient par la structure argumentale de base verbale. Selon Roy et Soare (2012 : 222), à la différence du suffixe *–eur*, le suffixe *-ant* peut admettre des bases inaccusatives, des bases statives ou de perception, ainsi que des

verbes réfléchis. Nous prenons quelques exemples (cités par Anscombre 2001 : 40 ; Roy et Soare 2012 : 222) pour bien illustrer. Dans l'exemple (21), le verbe ARRIVER est un verbe inaccusatif, car son sujet correspond généralement au thème, et il s'agit d'un verbe intransitif. Dans ce cas, le suffixe *-ant* est sélectionné pour former le nom désignant « personne qui arrive », tandis que le suffixe *-eur* est beaucoup moins acceptable pour dériver ce lexème. De même, dans les exemples (22) à (24), seul le suffixe *-ant* s'avère nettement plus approprié que *-eur*. Concrètement, pour le verbe VOIR, qui exprime une perception, le suffixe *-ant* est choisi pour désigner « personne qui voit ». Dans l'exemple (23), APPRENANT dérive du verbe APPRENDRE, qui est faiblement dynamique, c'est-à-dire plutôt statif. Enfin, pour l'exemple (24), la base de PLAIGNANT est effectivement SE PLAINDRE, un verbe réfléchi.

(21) ARRIVER : un arrivant vs. *un arriveur (cf. des bases inaccusatives)

(22) VOIR : un voyant vs. #un voyeur (cf. des bases de perception)

(23) APPRENDRE : un apprenant vs. *un appreneur (cf. des bases stratives)

(24) SE PLAINDRE : un plaignant vs. *un plaigneur (cf. des verbes réfléchis)

1.3.2.4 Propriétés sémantiques

Dans cette partie, nous examinons les propriétés sémantiques des suffixes à étudier, à travers la question centrale de leur polysémie fonctionnelle (agent/instrument, lieu/instrument, instrument/patient, etc.). De nombreux travaux morphologiques (Darmesteter 1877 ; Meyer-Lübke 1895 ; Dubois 1962 ; Sliwa 1990 ; Rainer 2005 ; Namer & Villoing 2008 ; Rosenberg 2008 ; Huyghe & Tribout 2015 ; Salvadori & Huyghe 2022) ont soutenu l'idée que les suffixes déverbaux et non déverbaux concernés ne se limitent pas à la formation des noms d'instrument, mais présentent une polysémie systématique. Pour éclairer cette problématique, nous regroupons les contributions existantes en deux ensembles : les suffixes déverbaux d'une part, les suffixes non déverbaux (noms dénominaux ou désadjectivaux) d'autre part, afin de mettre en évidence les régularités sémantiques et les facteurs déterminants dans la construction du sens.

Nous commençons par les suffixes déverbaux. Deux grandes orientations se dégagent. L'une se caractérise par les études essentiellement empiriques (Darmesteter 1877 ; Meyer-Lübke 1895 ; Dubois 1962 ; Benveniste 1975 ; Sliwa 1990 ; Dal 1999 ; Fradin 2005 ; Namer & Villoing 2008 ; Rosenberg 2008, 2011 ; Huyghe & Tribout 2015 ; Knittel 2017a, 2017b ; Wauquier 2020 ; Salvadori & Huyghe 2022) en explorant

la polysémie des suffixes sur la base d'attestations lexicales ; d'autres travaux adoptent une approche plus explicative en mobilisant les propriétés sémantiques et syntaxiques de la base verbale pour rendre compte des contraintes de dérivation (Ferret 2013 ; Fradin & Winterstein 2013).

Le cas du suffixe *-eur*, *-euse* et *-rice* cristallise particulièrement ce débat parmi les morphologues. Ces suffixes sont considérés comme polysémiques, car ils permettent de dériver des noms d'agent (PRESENTATEUR), d'instrument (LOCALISATEUR) et de lieu (BOUDEUSE). Dans le cadre des recherches axées sur la polysémie, Darmesteter (1877 : 47), pionnier à explorer ce sujet, a trouvé que les suffixes *-eur*/*-euse*/*-rice* sont polysémiques depuis le XIXe siècle. Selon lui, le suffixe *-eur* dérive initialement les noms d'agent. A partir du XIXe siècle, les suffixes nominaux *-eur* et *-euse* dénotant des noms d'instruments étaient le résultat de l'ellipse des expressions « appareil + adjectif en *-eur* » et « machine + adjectif en *-euse* ». Par exemple, ENREGISTREUR est l'ellipse de l'expression « appareil + enregistreur » et SECHEUSE est l'ellipse de l'expression « machine + sécheuse ». Avec cette grammaticalisation, ces suffixes permettent, d'une manière automne, de former les noms d'instrument en français (Rainer 2005 : 28).

Les propos de Darmesteter ont été étayés par les travaux de Dubois (1962), Benveniste (1975), Spence (1990), Fradin (2005), Rosenberg (2008 ; 2011), Huyghe & Tribout (2015) et Wauquier (2020). Dubois (1962 : 40-44) a précisé que les suffixes *-eur*/*-euse*/*-rice* sont polysémiques depuis la révolution industrielle. Or, Spence (1990) partage partiellement l'opinion de Dubois (1962). Selon lui, ces suffixes deviennent polysémiques avant la révolution industrielle. Par exemple, RAPPORTEUR désigne « instrument à l'aide duquel on peut mesurer les angles ou construire (rapporter) un angle d'une mesure donnée. » en 1680, qui est formé avant la révolution industrielle. A ce moment-là, le suffixe *-eur* permet déjà de dériver les noms d'instruments.

De plus, Benveniste (1975 : 61) a mis en évidence la polysémie agent/instrument des N-*eur* dans la perspective ontologique. En d'autres termes, la polysémie du suffixe *-eur* n'est pas déterminée linguistiquement, mais dépend de contraintes extralinguistiques et socio-culturelles. « Les mots en *-(t)eur* désignent des hommes ou des instruments. On ne devinerait pas, si on ne le savait pas, que chauffeur s'applique à un homme, brûleur à un appareil, et il est d'ailleurs inévitable, dans une civilisation de plus en plus mécanisée, que les tâches humaines s'assimilent à des fonctions d'instruments. » Benveniste (1975 : 61).

Dans la perspective sémasiologique, en synchronie, Rosenberg (2008 ; 2011) a identifié la polysémie du suffixe *-eur* pour dénoter les noms d'agents et les noms d'instruments Elle a précisé que le suffixe *-eur* construit plus de noms d'agent que d'instrument. En diachronie, « le sens Agent est de loin le plus rentable. Le sens Instrument ne devient rentable qu'aux 19e et 20e siècles, et il ne devient disponible qu'au 14e siècle. » (Rosenberg 2008 : 159). En outre, en s'appuyant sur les tests linguistiques proposés par Namer & Villoing (2008), Huyghe & Tribout (2015) ont trouvé une différence entre les noms d'instruments en *-eur* et ceux d'agent en *-eur*. Concrètement, en ce qui concerne les noms d'instruments en *-eur*, le facteur ontologique joue un rôle important pour prédire ce type sémantique. Par contre, le verbe de base occupe une place primordiale pour définir les noms d'agent en *-eur*. « Dans le cas des N d'agents, [...] la dénotation d'entités animées rend la prédication agentive contingente, les animés pouvant effectuer une grande variété d'actions. La composante prédicative prend, de ce fait, plus d'importance que dans le cas des N d'instruments, et la dépendance syntactico-sémantique des noms est plus manifeste. » (Huyghe et Tribout 2015 : 110)

A la différence des auteurs ci-dessus, Wauquier (2020) a appliqué la méthode distributionnelle pour explorer les interprétations des suffixes *-eur*/*-euse*/*-rice*. Enfin, en plus des sens instrument/agent, Fradin & Winterstein (2013 : 11) ont trouvé encore que le suffixe *-euse* permet de dériver les noms de lieu, par exemple : BOUDEUSE, CAUSEUSE, COIFFEUSE.

Les suffixes *-oir* et *-oire* ont également fait l'objet d'une importante attention concernant la question de polysémie (Meyer-Lübke 1895 ; Sliwa 1990 ; Namer & Villoing 2008 ; Rosenberg 2008 ; Fradin & Winterstein 2013 ; Ferret 2013 ; Salvadori & Huyghe 2022). Plus précisément, Meyer-Lübke (1895) observe dès le XIXe siècle une ambivalence instrument / lieu dans des noms comme BAIGNOIRE ou ARROSOIR. Cette idée est reprise par Sliwa (1990), qui identifie d'autres rôles sémantiques (Patient, Cause) dérivés par ces suffixes. A partir de tests linguistiques, Namer & Villoing (2008) montrent que les dérivés en *-oir* sont sensibles à la transitivité du verbe de base : les verbes transitifs favorisent des interprétations instrumentales, tandis que les verbes intransitifs donnent lieu à des interprétations locatives (*par ex.* DORMIR > DORTOIR).

Rosenberg (2008), quant à elle, reprend cette hypothèse en la généralisant dans une perspective sémasiologique : en synchronie, le sens Instrument est dominant, suivi du Locatif ; en diachronie, ces deux sens sont attestés très tôt, sans qu'on puisse établir

une dérivation unidirectionnelle entre eux. Cette interprétation est confortée par Salvadori & Huyghe (2022), qui recourent à une analyse probabiliste et concluent que les fonctions artéfact-instrument et artéfact-location couvrent ensemble 76 % des emplois annotés du suffixe *-oir*.

Plusieurs travaux récents ont renouvelé l'analyse de la polysémie des suffixes déverbaux, en explorant les relations entre propriétés sémantiques des bases verbales et types dénominatifs dérivés. Dans cette perspective, Ferret (2013) propose d'examiner si les suffixes *-oir/-oire* sont sensibles à certaines propriétés sémantiques de la base verbale. Elle montre que le trait [+/-manière] joue un rôle déterminant dans la distinction entre noms d'instruments et noms de lieu. Si la base verbale porte le trait [+manière], le dérivé a plus de chances de renvoyer à un instrument. Sinon, le dérivé a plus de chances de renvoyer à un lieu. Par exemple, dans RASOIR, dérivé du verbe RASER, la présence de trait [+manière] motive l'interprétation instrumentale (exemple 25). En revanche, dans des cas comme LAVOIR, la base LAVER manquant du trait [+manière] empêche l'interprétation instrumentale. Le LAVOIR désigne alors un lieu et non un instrument (exemple 26).

(25) Jean a rasé la tête de Paul → Jean a utilisé un objet pour raser la tête de Paul (Ferret & Villoing 2015 : 485).

(26) Jean a lavé la chemise de Paul → Jean a lavé cette chemise dans un lavoir.

Fradin & Winterstein (2013) adoptent une approche différente : ils mettent en avant des contraintes structurelles supplémentaires pour qu'un argument verbal soit interprété comme un instrument, telles que l'individualité et la réutilisabilité. Pour eux, l'interprétation locative repose plutôt sur des propriétés aspectuelles du verbe : *Coinitial* (PLONGEOIR), *Cofinal* (CRACHOIR), *Transitoire* (COULOIR) ou *Statique* (ABATTOIR), lesquelles correspondent à des schémas directionnels liés au mouvement du participant (ci, cf, tr, st). Par exemple, dans COULOIR, l'objet est dit transitoire dans un voisinage pendant un intervalle indéfini, ce qui rend le repérage temporel difficile (Fradin & Winterstein 2013 : 14). Ces deux approches, bien que distinctes, l'une centrée sur les traits sémantiques (Ferret 2013), l'autre sur les contraintes de construction événementielle (Fradin & Winterstein 2013), convergent sur l'idée que les suffixes comme *-oir* mobilisent des régularités sémantiques internes à la base verbale, expliquant leur tendance à produire tant des noms d'instrument que des noms de lieu.

En marge de ces suffixes fréquemment étudiés, d'autres suffixes déverbaux

présentent eux aussi une polysémie fonctionnelle. Ainsi, Dal (1999) attribue au suffixe *-ette* une capacité à dériver divers types de noms : agent (COUSETTE), objet (GRIMPETTE), instrument (SONNETTE), et lieu (OUBLIETTE). Cette polysémie, bien que moins systématisée que celle de *-oir*, témoigne d'une souplesse dérivationnelle comparable.

Concernant le suffixe *-ant*, Knittel (2017a, 2017b) met en évidence sa polysémie, en distinguant notamment les noms d'agent (comme ASSISTANT) et les noms non animés (comme DESHERBANT). Elle observe que ces derniers tendent à se comporter comme des noms d'instruments. Cette interprétation s'appuie sur le fait que, tout comme les instruments, ces noms, fréquemment liés à des substances chimiques, ne requièrent pas de contrôle agentif, car leurs propriétés intrinsèques leur confèrent une capacité d'action autonome. Cette tendance s'inscrit également dans une évolution diachronique. Rosenberg (2008 : 158) souligne qu'à partir du XIXe siècle, le sens instrumental du suffixe *-ant* devient de plus en plus productif, et qu'au XXe siècle, il atteint une rentabilité comparable à celle du sens agentif. Ces résultats mettent en évidence une évolution sémantique du suffixe : celui-ci en vient désormais à former, outre les noms d'agent, d'autres noms.

S'agissant des suffixes non déverbaux, la polysémie semble beaucoup plus rare. En effet, seuls les suffixes *-ier*/*-ière* présentent une productivité sémantique comparable. Corbin et Temple (1994) leur attribuent une capacité à dériver, sur la base d'indices extralinguistiques, aussi bien des noms d'instruments que des noms de personne ou d'arbre (*par ex.* BARBE > BARBIER ; POMME > POMMIER). Roché (1998) reprend cette analyse et l'enrichit en montrant que ces suffixes permettent également la dérivation de noms de récipient (*par ex.* CAFE > CAFETIERE). Ces différentes contributions soulignent que comme les suffixes *-eur*, *-euse*, *-rice*, *-oir* et *-oire*, les suffixes *-ant*, *-ier* et *-ière*, se caractérisent aussi par une extension sémantique structurée, qui reflète une diversification progressive de leurs fonctions dénominatives au sein du lexique.

1.3.3 Relation entre les suffixes et les sous-classes ontologiques

Dans la section 1.2.2.2, nous avons mentionné que certains morphologues considèrent que la forme féminine peut exister en tant que suffixe autonome par rapport à la forme masculine, car ils se distinguent par la sémantique. De plus, certains morphologues et lexicologues (Dubois 1962 ; Spence 1990 ; Rainer 2021) ont trouvé encore que les suffixes différents permettent d'instancier respectivement les sous-

classes ontologiques différentes. Concernant les sous-classes proposées, ils ne sont pas souvent unanimes, ne justifient pas leur classification, et ne définissent pas les frontières entre les sous-classes ontologiques. Leurs sous-classes semblent choisies intuitivement. Nous allons présenter d'abord les sous-classes des noms d'instruments évoquées par les auteurs. Ensuite, nous allons aborder la sous-classe ontologique des instruments selon les suffixes.

Concernant les sous-classes des noms d'instruments, les morphologues (en synchronie) proposent une série de sous-classes des noms d'instruments que nous présentons dans le tableau 1.2.

Tableau 1.2 Sous-classes (pour leur définition, voir la Section 3.2) des noms d'instruments

sous-classes	Références
Instrument de captation	Dubois (1962) ; Nagórko (1981) ; Oczkowska (1983) ; Grevisse & Goosse (2016); Lin (2021)
Machine	Dubois (1962); Nagórko (1981) ; Oczkowska (1983); Grzegorczykowa (1984); Sliwa (1990) ; Grevisse & Goosse (2016) ; Rainer (2021) ; Lin (2021)
Appareil	Dubois (1962); Nagórko (1981) ; Oczkowska (1983) ; Grzegorczykowa (1984) ; Sliwa (1990) ; Grevisse & Goosse (2016) ; Wang (2016) ; Lin (2021)
Outil	Grzegorczykowa (1984) ; Sliwa (1990) ; Rainer (2021) ; Lin (2021)
Substances chimiques	Oczkowska (1983) ; Grzegorczykowa (1984) ; Rainer (2021)
Organe	Nagórko (1981) ; Oczkowska (1983) ; Sliwa (1990)
Moyens de transport	Oczkowska (1983)
Meubles	Oczkowska (1983)
Récipient	Sliwa (1990) ; Grevisse & Goosse (2016) ; Rainer (2021); Lin (2021)
Installations	Oczkowska (1983)
Pièce de vêtement	Rainer (2021)
Ustensile	Grevisse & Goosse (2016) ; Lin (2021)

Parmi les sous-classes proposées, Instrument de captation, Machine, Appareil, Outil et Récipient sont mentionnés plus fréquemment que les autres. De surcroît, en fonction de différentes perspectives, les morphologues ou lexicologues ont regroupé ces sous-classes de noms d'instruments. Selon la manière d'effectuer une action sur le patient avec les instruments, Oczkowska (1983 : 42-43) divise les sous-classes en deux groupes. Concrètement, certaines sous-classes ontologiques permettent de servir directement à effectuer une action (des instruments, des appareils, installations, machines, substances, parties du corps). D'autres sous-classes ontologiques servent à être intermédiaires et à aider à effectuer une action (moyens de transport, meubles, etc) (cité par Sliwa 1990). Prenant pour jalon l'avènement de la Révolution industrielle, Dubois (1962) divise les sous-classes d'instruments en deux groupes : les anciens et les

nouveaux instruments. Selon lui, les anciens instruments se caractérisent par leur manipulation à la main, soit l'outil, l'instrument de captation et le récipient. Les nouveaux instruments, quant à eux, se caractérisent par la mécanisation et la technologie et comprennent la machine et l'appareil.

De plus, Sliwa (1990) a utilisé l'autonomie de l'utilisation des instruments pour distinguer deux groupes de sous-classes de noms d'instruments. Certains sont non autonomes et ne fonctionnent pas par eux-mêmes. Cela concerne les instruments (outils) au sens propre, les organes et les récipients. D'autres sont au contraire autonomes et peuvent fonctionner seuls une fois mis en marche par l'agent humain. Ce sont alors des machines ou des appareils. A la différence des auteurs évoqués, Rainer (2021 : 1) a classé les noms d'instruments, en fonction de leurs caractéristiques prototypiques, comme des instruments prototypiques et moins prototypiques. L'auteur a précisé que parmi eux, outil et machine sont prototypiques A l'opposé, Récipient, Substance chimique, Pièce de vêtement sont moins prototypiques, car ces sous-classes ontologiques se situent à la frontière des noms d'instruments et des autres classes de noms, telles que le lieu, le thème. Par exemple, selon Rainer (2021 : 1), les noms de récipients se situent à la frontière entre les noms d'instruments et ceux de lieu. « Un encrier est un artefact conçu pour contenir de l'encre et, à ce titre, il est proche d'un instrument, mais il peut aussi être considéré comme un lieu où l'encre est stockée. »[13]

Après avoir introduit les sous-classes ontologiques des noms d'instruments et leur regroupement, nous allons examiner la sous-classification ontologique des instruments selon les suffixes. Cette démarche se justifie par le fait que certains auteurs (Dubois 1962 ; Spence 1990 ; Fradin 2003 ; Devos & Taeldeman 2003 ; Rainer 2021 ; Lin 2021) constatent que les suffixes permettent de désigner certaines sous-classes spécifiques.

Tableau 1.3 Sous-classification ontologique des instruments selon les suffixes

Suffixes	sous-classes et références
-eur	- Appareil (Dubois 1962 ; Grevisse & Goosse 2016 ; Lin 2021) ; - Machine (Rainer 2021) ; - Substance chimiques (Devos & Taeldeman 2003) ; - Instrument de captation (Lin 2021)
-euse	- Appareil (Rainer 2021) ; - Machine (Grevisse & Goosse 2016 ; Rainer 2021 ; Lin 2021)
-ier	- Appareil (Dubois 1962) ; - Récipient (Roché 1998 ; Grevisse & Goosse 2016) ; - Ustensile (Roché 1998 ; Grevisse & Goosse 2016)
-ière	- Récipient (Grevisse & Goosse 2016) ;

[13] Le texte originale : "an inkpot is an artifact designed to contain ink and as such close to an instrument, but can also be viewed as a place where ink is stored."

	- Ustensile (Grevisse & Goosse 2016)
-oir	- Instrument de captation (Rainer 2021 ; Lin 2021) ; - Outil (Dubois 1962 ; Rainer 2021 ; Lin 2021)
-oire	- Instrument de captation (Rainer 2021) ; - Outil (Dubois 1962 ; Rainer 2021)
-ant	- Substance chimiques (Devos & Taeldeman 2003 ; Knittel 2017 a)
-ante	- Substance chimiques (Knittel 2017 a)
-et	- Outil (Fradin et al. 2003 ; Rainer 2021) ; - Ustensile (Fradin et al. 2003)
-ette	- Outil (Fradin et al. 2003 ; Rainer 2021) ; - Ustensile (Fradin et al. 2003)

Dans le Tableau 1.3, on observe que certains suffixes (*-eur*, *-ier*/*-ière*, *-oir*) permettent de dériver une plus grande diversité de sous-classes ontologiques dénotées, tandis que d'autres suffixes (*-euse*, *-ant*/*-ante*, *-et*/*-ette*) semblent associés à un nombre plus restreint de sous-classes ontologiques dénotées. D'après Spence (1990 : 30), lorsque les anciens et les nouveaux instruments ont continué à coexister, différents suffixes étaient mobilisés pour désigner différentes sous-classes d'instrument. Par exemple, concernant le cas ARROSOIR/ARROSEUR/ARROSEUSE, ARROSOIR désigne un récipient destiné à l'arrosage, ARROSEUR dénote un « appareil d'arrosage automatique » (Le Grand Robert 2005) et ARROSEUSE est un « véhicule muni d'un réservoir d'eau et destiné à l'arrosage des voies publiques » (Le Grand Robert 2005). Autrement dit, ces suffixes se distinguent par les sous-classes différentes instanciées. En revanche, il existe des exceptions qui permettent de dénoter la même sous-catégorie avec des suffixes différents. A titre d'exemple, BUTTOIR/BUTTEUR dénotent tous deux une « petite charrue utilisée pour butter (des plantes) » (Le Grand Robert 2005) et ils sont synonymes. Autrement dit, il n'y a pas de différence sur la sous-classification ontologique selon les suffixes.

En utilisant les données dictionnairiques du *Grand Robert* (2005), j'ai examiné (Lin 2021) en diachronie la préférence et l'évolution des sous-catégorisations selon les suffixes. J'ai étayé une partie des résultats de recherches évoquées (Dubois 1962 ; Sliwa 1990 ; Rainer 2021) : les suffixes *-eur* et *-oir* préfèrent construire des noms d'Instrument de captation ; la formation des noms d'Appareil a une préférence pour-*eur* ; la formation des noms de Machine a tendance à sélectionner le suffixe *-euse* ; la formation des noms d'Outil a une préférence pour *-oir.* De plus, j'ai également trouvé que pour les noms des instruments de captation, le suffixe *-oir* est le patron avant le XIXe siècle, et après le XIXe siècle le suffixe *-eur* devient le nouveau patron en remplaçant le suffixe *-oir*. Enfin, concernant les noms de récipients et d'ustensile, il ne

semble pas y avoir de suffixe préférentiel.

Jusqu'à présent, nous avons présenté des approches et des travaux de recherche sur les noms d'instruments. Nous avons constaté que les morphologues et les sémanticiens ont exploré ce sujet sous différents angles. Cependant, il existe encore des lacunes et des problèmes sur ces approches et ces recherches. Nous nous concentrerons sur ce point dans la section suivante.

1.4 Critiques

Les approches et les recherches connexes présentent plusieurs problématiques et lacunes manifestes qui méritent d'être analysées. Nous commençons par les approches sur les noms d'instruments. Comme mentionné précédemment, selon l'approche relationnelle syntactico-sémantique, les auteurs privilégient les facteurs syntactico-sémantiques pour étudier les noms d'instruments et marginalisent les facteurs ontologiques, de sorte que certains noms abstraits et corporels sont considérés comme des noms d'instruments (Lieber 2021). Cependant, cela ne correspond pas aux noms d'instruments prototypiques dans notre cognition, car les noms d'instruments sont plutôt des artefacts manipulables (Langacker 2008). Afin de remédier à ce problème, Roché (1998) ajoute une contrainte ontologique à la définition des noms d'instruments, ce qui favorise l'émergence des approches mixtes. Cette approche mixte prend en compte les noms concrets représentant des participants dans un procès. Selon cette approche, concernant les noms dénominaux et les noms désadjectivaux, si le dérivé représente le rôle sémantique instrument du procès dans la représentation sémantique du dérivé, ils sont considérés comme des noms d'instruments. Dans cette thèse, nous allons suivre l'approche mixte proposée par Roché (1998) pour étudier les noms d'instruments construits morphologiquement par suffixation.

Cependant, bien que Roché (1998) valorise le facteur ontologique dans l'approche mixte, les noms d'instruments prototypiques ne sont pas systématiquement exploités. Pour mieux délimiter les noms d'instruments prototypiques, selon Rainer (2021), les sous-classes d'instruments sont essentielles. « Comme c'est le cas pour d'autres catégories de formation de mots, celles qui sont désignées par le nom d'instrument et le nom de lieu ne forment pas des ensembles conceptuellement homogènes, mais plutôt des ensembles de sous-classes conceptuellement

apparentées » (Rainer 2021 : 1).[14] Cependant, ces auteurs évoqués (Roché 1998 ; Namer & Villoing 2008 ; Huyghe & Tribout 2015) ne fournissent pas de sous-classes d'instruments prototypiques permettant une meilleure délimitation des noms d'instruments. Ainsi, un des objectifs de cette thèse vise à identifier des sous-classes représentant les noms d'instruments prototypiques.

Notre état de l'art a permis de pointer que les travaux en morphologie se concentrent principalement sur les suffixes instrumentaux déverbaux, en raison de l'influence de l'approche relationnelle. Les noms non déverbaux sont moins étudiés et interprétés, bien qu'ils constituent une partie essentielle des noms d'instruments, représentant une proportion non négligeable. Sans prise en compte de ces derniers, il serait impossible d'avoir une vision complète du panorama des noms d'instruments morphologiquement construits. Par conséquent, il est nécessaire d'en discuter avec la même rigueur que celle appliquée aux noms déverbaux.

Concernant les propriétés linguistiques étudiées sur ces suffixes, elles ne sont pas abordées avec la même précision. En effet, les études existantes se cantonnent principalement à la sémantique, aux espaces thématiques au niveau morphologique ainsi qu'à la catégorie syntaxique de la base du dérivé. Cependant, certaines propriétés de la base sont encore peu prises en compte. Outre les traits phonologiques (nombre de syllabes, segment final) et morphologiques (structure interne, groupe de conjugaison) déjà mentionnés, d'autres facteurs syntaxiques, sémantiques et pragmatiques peuvent aussi orienter le choix du suffixe (Plag 2018 : 172). Pour comprendre la formation des noms d'instruments, il faut donc intégrer l'ensemble de ces propriétés et les analyser de façon systématique.

En ce qui concerne le genre, les morphologues ne partagent pas les mêmes avis. Concrètement, certains considèrent la forme féminine comme la variante de la forme masculine (Rosenberg 2008 ; Namer & Villoing 2008 ; Fradin & Winterstein 2013 ; Huyghe & Tribout 2015), alors que d'autres proposent que la forme féminine constitue un suffixe autonome pour dériver des noms d'instrument (Spence 1990 ; Wauquier 2020 ; Lin 2021). Selon ces derniers auteurs, les formes féminines se distinguent de la forme masculine par leur sémantique. Wauquier (2020) a notamment validé les différences dans ces deux cas lors de l'étude des noms d'agent. Elle constate que par

[14] Le texte original : "As is the case with other categories of word formation, those referred to with instrument noun and place noun do not constitute conceptually homogeneous sets, but sets of conceptually related subcategories."

rapport au suffixe *-eur*, les suffixes *-euse* et *-rice* sont des suffixes autonomes. Dans ce présent travail, nous allons suivre l'idée de Wauquier (2020) pour vérifier si les suffixes tels que *-euse*, *-rice*, *-oire*, *-ière* sont des suffixes autonomes.

Par ailleurs, la question de la relation entre les suffixes et les sous-classes d'instruments n'a jamais été traitée à partir de définitions précises des sous-classes, ce qui pourrait entraîner une hétérogénéité des critères et des résultats obtenus. À l'exception de mes propres travaux (Lin 2021), les recherches ne reposent pas sur des données extensives. On peut s'interroger sur la validité de ces observations à l'échelle de la langue. Bien que j'aie réalisé une recherche empirique sur ce sujet (Lin 2021), je me suis appuyé sur les définitions du *Grand Robert* (2005), lesquelles ne permettent pas d'extraire clairement les caractéristiques des sous-classes. Par exemple, RASOIR est classé dans le groupe des instruments de captation, mais la notion de RASOIR n'inclut aucune caractéristique liée à la perception ou à la captation. Au contraire, RASOIR correspond mieux aux critères du groupe « outil ». Ainsi, la classification du *Grand Robert* (2005) ne présente pas les garanties de cohérence nécessaires à une étude linguistique approfondie. Il serait donc urgent de traiter ce problème en vérifiant et reclassant les sous-classes selon de nouvelles définitions.

Enfin, dans les travaux précédemment évoqués, la sélection de la base morphologique y est généralement considérée comme un processus statique, c'est-à-dire comme une forme lexicale déterminée au préalable et considérée comme stable, sans questionner le mécanisme cognitif qui conduit à son choix précis. Autrement dit, ces recherches s'arrêtent à l'observation que tel ou tel lexème sert de base à une dérivation, mais elles n'expliquent guère pourquoi et comment ce lexème particulier a été sélectionné dans le processus dérivationnel. L'absence de perspective cognitive empêche ainsi de saisir pleinement la dynamique interne de la dérivation des noms d'instruments en français.

En guise de conclusion, les travaux examinés présentent certaines limites, notamment en ce qui concerne la notion des noms d'instruments prototypiques et les propriétés linguistiques prises en compte pour l'analyse des suffixes instrumentaux. À la lumière de ces analyses critiques, nous proposons les hypothèses suivantes :

- Il est possible d'identifier des sous-classes prototypiques de noms d'instruments dans une perspective mixte (ontologique et linguistique), afin de mieux cerner ce que sont les noms d'instruments prototypiques (voir la Section

3.2).

- Certains suffixes instrumentaux privilégient certaines propriétés linguistiques spécifiques. Leur choix est souvent conditionné par des contraintes linguistiques de divers niveaux linguistiques : phonologiques, morphologiques, syntaxiques, sémantiques (voir le chapitre 5).
- Les formes féminines des suffixes se distinguent de leurs formes masculines non seulement par leur sémantique, mais également par d'autres propriétés linguistiques (voir le chapitre 5).
- Les sous-classes ontologiques, auxquelles appartiennent les noms construits, jouent un rôle dans le choix des suffixes instrumentaux (voir le chapitre 5).
- Enfin, la base morphologique n'est pas préalablement donnée. Son choix résulte d'un processus de sélection dynamique et dépend d'une motivation cognitive. Dans une perspective cognitive, cette sélection repose sur le profil sémantique assumé par la base dans la représentation sémantique du dérivé, activées *via* le profilage conceptuel (voir la Section 3.1.3). Chaque suffixe instrumental tend à privilégier certains profils sémantiques assumés par la base dans la représentation sémantique du dérivé. (voir la section 3.1 et le chapitre 5)

Dans le chapitre 2, nous nous concentrerons sur la compétition morphologique, afin de comprendre comment les propriétés linguistiques activées interagissent dans la construction d'un lexème et quelles sont celles qui jouent un rôle discriminant dans la différenciation des suffixes lors de leur compétition.

Chapitre 2 La compétition morphologique

Ce chapitre est consacré à la compétition morphologique, un phénomène central en morphologie, et qui est au cœur de l'étude des suffixes qui font l'objet de notre recherche. La compétition morphologique « apparaît lorsque des affixes rivaux (c'est-à-dire construisant le même sens) interagissent de telle sorte qu'un mot dérivé peut être construit avec l'un des schémas morphologiques, empêchant les autres de former des mots synonymes » (Lieber 2021). Selon Lindsay & Aronoff (2013), la compétition morphologique dans le lexique est analogue à la compétition en écologie : le lexique repose sur une pluralité de propriétés linguistiques (phonologiques, morphologiques, syntaxiques et sémantiques), comparables aux ressources disponibles pour les organismes vivants. Cette analogie s'étend au principe d'exclusion compétitive formulé par Gause (1934), selon lequel deux espèces en compétition pour une même niche écologique ne peuvent pas coexister durablement. De même, en morphologie, les schémas en compétition ne coexistent généralement pas, à moins que l'un d'eux ne change de niche ou ne se spécialise davantage, afin d'éviter l'extinction (Aronoff 2019). Afin de mieux comprendre les dynamiques propres à la compétition morphologique, nous commencerons par une présentation de la compétition écologique (§2.1). Nous aborderons ensuite les définitions de la niche morphologique, notion centrale en morphologie, et leur apport aux recherches sur la compétition morphologique (§2.2). Nous présenterons ensuite deux approches influentes ayant nourri les travaux récents sur cette question (§2.3). Enfin, nous conclurons ce chapitre par quelques remarques sur les notions et études existantes portant sur la compétition morphologique (§2.4).

2.1 La compétition écologique

Dans cette section, nous nous concentrons sur la compétition écologique. Nous commencerons par présenter les définitions de la niche écologique, avant d'examiner le principe d'exclusion compétitive formulé par Gause (1934), qui a inspiré Aronoff (2019), qui a proposé une approche écologique de la compétition morphologique. La niche écologique joue un rôle fondamental dans la dynamique de la compétition entre espèces. À ce sujet, plusieurs définitions ont été proposées par les écologistes, parmi lesquels Grinnell (1917) et Hutchinson (1957), chacun mettant en lumière des aspects

différents de cette notion. Grinnell (1917) est le premier à avoir défini le concept de niche. Selon lui, la niche correspond à « tout ce qui conditionne l'existence d'une espèce à un endroit donné », incluant à la fois des facteurs abiotiques (température, humidité, précipitations) et biotiques (disponibilité de nourriture, compétiteurs, prédateurs, abris, etc.). Cette conception, enracinée dans la pensée darwinienne, insiste sur l'importance des dimensions spatiales de l'environnement (Ge 2008 : 216). Hutchinson (1957), quant à lui, fait évoluer cette définition en proposant une approche plus abstraite et formalisée. Il définit la niche comme un hypervolume « où chaque dimension de l'espace représente une ressource (alimentaire, en matériaux, spatiale, offre en cachette, substrats ou perchoirs, etc.) ou une condition (température, précipitations, acidité, etc.) de l'environnement. La quantité de ressources varie dans l'espace et dans le temps en fonction de l'activité de l'espèce. Les conditions et les ressources sont des conditions limitantes qu'on peut hiérarchiser pour étudier la vulnérabilité de l'espèce dans l'environnement. » [15]

Par ailleurs, la définition de la niche écologique proposée par Hutchinson (1957) a continué d'inspirer de nombreux écologistes, notamment en ce qui concerne la mesure des niches. Hutchinson soulignait que les niches écologiques des différentes espèces se chevauchent partiellement selon certaines dimensions. En raison de cette distribution inégale des ressources et du chevauchement partiel des niches, l'indicateur de chevauchement des niches constitue un outil conceptuel particulièrement pertinent pour transposer cette perspective au domaine de la morphologie. Il offre ainsi une référence précieuse pour mesurer le chevauchement des modalités entre les niches morphologiques et analyser leurs relations réciproques. Nous procéderons à sa présentation et à son calcul dans le chapitre 6.

En outre, nous allons présenter le principe d'exclusion compétitive, proposé par le biologiste et évolutionniste soviétique Georgy Gause, et largement adopté en écologie. Selon Pocheville (2011 : 899), cette idée est déjà proposée par Darwin (1872 : 85) et a été développée à partir de la définition de la niche écologique proposée par Grinnell (1917). Ce principe concerne la compétition entre deux populations qui ne peuvent coexister dans la même niche écologique (Gause 1934). Concrètement, Gause (1934) stipule que « les populations de deux espèces ayant des exigences écologiques

[15] Traduction en provenance de Wikipédia., « Niche écologique » [en ligne], disponible à l'adresse : https://fr.wikipedia.org/wiki/Niche_%C3%A9cologique (consulté le 10 juin 2024).

identiques, c'est-à-dire exploitant une ressource limitante unique, ne peuvent coexister indéfiniment dans un milieu stable et homogène » (Hardin 1960 : 1292). Dans ce cadre, seule une différenciation des niches permet d'atténuer la compétition et de permettre leur coexistence.

En résumé, nous avons présenté deux définitions influentes de la niche écologique ainsi que le principe d'exclusion compétitive. Ces notions occupent une place centrale en écologie et éclairent notre compréhension des définitions de la compétition morphologique qui seront abordées dans la section suivante. Elles constituent en outre des références théoriques solides pour étayer notre perspective sur la compétition morphologique, et facilitent l'application du chevauchement des niches en vue de mesurer le chevauchement entre différentes niches morphologiques, c'est-à-dire la proportion de modalités linguistiques partagées entre elles.

2.2 Niche morphologique

Inspirés par Gause (1934), Lindsay & Aronoff (2013) ont introduit une perspective écologique en morphologie en établissant une analogie entre le lexique et un écosystème. Depuis cette proposition fondatrice, la recherche morphologique dans cette direction s'est largement développée, donnant lieu à une production scientifique foisonnante (Aronoff 2019 ; Lin 2021 ; Missud 2023 ; Bobkova 2023). Le concept de niche, devenu central, joue désormais un rôle fondamental dans l'analyse de la compétition morphologique. Dans cette lignée, plusieurs morphologues se sont saisis de cette notion pour approfondir les mécanismes de coexistence et de différenciation dans le lexique. En s'inspirant des travaux de Lindsay & Aronoff (2013) et en y intégrant une modélisation statistique, Missud (2023) a proposé une conceptualisation multifonctionnelle de la niche morphologique, offrant ainsi une approche novatrice de cette problématique. Nous présenterons la notion de niche morphologique telle qu'elle a été initialement formulée par Lindsay & Aronoff (2013) (§2.2.1), avant d'examiner le cadre développé par Missud (2023) (§2.2.2).

2.2.1 Lindsay & Aronoff (2013)

Comme mentionné précédemment, le principe d'exclusion compétitive de Gause (1934) est élaboré à partir du concept de niche écologique proposé par Grinnell (1917), qui met l'accent sur la distribution des espèces. Selon ce principe, deux espèces

occupant la même niche écologique ne peuvent pas coexister durablement : la plus efficace survit, l'autre disparaît (Gause 1934, repris par Aronoff 2019 : 41). A la lumière de ce principe, Lindsay & Aronoff (2013) ont observé une analogie entre les écosystèmes et le lexique, et ont proposé une métaphore écologique pour penser la structure lexicale. Dans cette perspective, les schémas morphologiques occupent des niches linguistiques spécifiques, ce qui leur permet de coexister et de survivre au sein du lexique. Concrètement, selon Lindsay & Aronoff (2013), chaque propriété linguistique peut constituer une niche linguistique : phonologique, morphologique, sémantique, ou pragmatique, entre autres. Pour subsister, un suffixe doit occuper au moins une de ces niches, qui constitue alors sa condition de survie. À défaut, il risque la disparition. C'est cette répartition dans des niches distinctes qui permet la coexistence entre affixes, notamment en contexte de compétition morphologique.

Les auteurs identifient ainsi quatre types de niches linguistiques fondées sur des contraintes linguistiques (Lindsay & Aronoff 2013 : 4, repris et traduit pas Missud 2023 : 26) :

- Niche phonologique : « les procédés peuvent être contraints de sélectionner certaines bases en raison de certaines contraintes prosodiques ou phonologiques. » [16] Par exemple, en anglais, le suffixe *-ify* préfère les radicaux monosyllabiques, tandis que son rival *-ize* tend à s'attacher à des radicaux polysyllabiques.
- Niche morphologique : « les procédés peuvent s'attacher à des bases préalablement construites à l'aide d'un procédé particulier. »[17] Par exemple, le suffixe *-ical* privilégie les radicaux se terminant en *-olog-* (*par ex.* MORPHOLOGICAL plutôt que MORPHOLOGIC).
- Niche sémantique : « les procédés peuvent se différencier sur la base des interprétations sémantiques de leurs dérivés. »[18] À ce titre, *-hood* et *-ship*, deux suffixes exprimant historiquement un « état », se sont différenciés : *-ship* étant limité à une lecture « stage-level », tandis que *-hood* peut avoir une lecture « stage-level » ou « individual-level » (Aronoff & Cho 2001).

[16] Le texte original : "The usage of an affix can be restricted by stress, syllables, or prosody (of the stem or the output of the affixed word)."

[17] Le texte original : "An affix can attach strictly to another specific affix, a phenomenon known as potentiation."

[18] Le texte original : "the affixes actually settle into a situation in which they are no longer competing."

- Niche pragmatique : « les procédés rivaux peuvent dénoter les mêmes sens mais être utilisés dans des registres spécifiques. »[19] Par exemple, les suffixes d'origine latine ou française comme *-esque* ou *-ian* sont typiques des registres formels, alors que *-ish* s'emploie plutôt en registre informel.

Par ailleurs, Lindsay & Aronoff (2013) s'appuient sur le principe d'exclusion compétitive pour expliquer la compétition morphologique. Selon eux, ce principe s'applique parfaitement à l'évolution des suffixes dans le lexique. Ils présentent deux cas pour illustrer ce propos : l'un concerne la compétition entre suffixes, l'autre traite de l'allomorphie. Concernant la compétition entre suffixes, *-ment* et *-ion* sont tous deux considérés comme des suffixes événementiels en anglais. Avant le XVIIe siècle, leur productivité était comparable. Cependant, en anglais, le suffixe *-ment*, affaibli, a été progressivement supplanté par *-ion*, dont la productivité s'est accrue grâce à des emprunts continus à partir du XVIIe siècle (Lindsay & Aronoff 2013 : 8 ; Aronoff 2016 : 3). En outre, ils mobilisent ce principe pour expliquer certains phénomènes d'allomorphie. Selon eux : « il n'existe aucun fondement permettant de considérer qu'une variante distributionnelle soit dérivée d'une autre. Dès lors, aucune signification perceptuelle particulière ne saurait être attribuée à la forme dite « elsewhere ». »[20] (Aronoff 2019 : 52). À titre d'exemple, ils citent les suffixes pluriels en anglais <-*s*> et <-*en*>. Selon leur analyse, aucun des deux n'est dérivé de l'autre ; le suffixe<-*en*> est relativement indépendant de <-*s*> et a réussi à trouver une niche morphologique, celle des noms hérités de la déclinaison faible du vieil anglais, comme en témoigne OX > OXEN. Ce cas illustre une distribution complémentaire par rapport à son rival <-*s*>.

À la suite de ces travaux, Aronoff et Lindsay ont continué à souligner l'importance de ce principe en morphologie. Selon eux : « l'axiome de Gause (1934) nous dit que la compétition finira toujours par être résolue. Mais il faut toujours un certain temps pour que la compétition se résolve d'elle-même. Cette période n'a pas de durée déterminée. » (Aronoff & Lindsay 2016 : 46). Notamment, Aronoff (2023 : 57) considère le principe d'exclusion compétitive comme particulièrement efficace pour expliquer la compétition morphologique, notamment en sémantique. Précisément, d'un point de vue sémantique, le suffixe *-free* présente une valeur évaluative (Grandi & Körtvélyessy 2015, repris par Aronoff 2023 : 58). Contrairement au suffixe *-less*, qui

[19] Le texte original : "An affix can be constrained to a specific register."

[20] Le texte original en anglais : "there is no sense in which one distributional variant is derived from another. Hence no special perceptual significance can be imputed to an elsewhere form."

se limite à signaler l'absence de la base du dérivé, le suffixe *-free* exprime en outre que l'absence de la base chez le référent est perçue comme souhaitable ou valorisée. Cependant, bien qu'il mentionne la définition écologique de la « niche » selon Hutchinson (1957) dans Aronoff (2023 : 57), il ne développe pas cette perspective théorique : la référence à la niche reste superficielle et sert principalement à illustrer le principe d'exclusion, sans toutefois développer de manière approfondie la perspective écologique dans l'analyse morphologique.

En définitive, à travers les travaux de Lindsay et Aronoff (2013, 2016), mais aussi d'Aronoff (2019, 2023), on peut résumer leur position comme suit : la compétition entre schémas morphologiques ne peut se résoudre que par l'extinction de l'un des éléments, ou par leur différenciation fonctionnelle, permettant leur survie. Cette compétition peut durer un certain temps avant de se stabiliser. Inspirée par cette approche, Missud (2023) poursuit cette réflexion en proposant une nouvelle définition de la niche morphologique, fondée sur une modélisation statistique. La section suivante sera consacrée à la présentation de cette perspective novatrice.

2.2.2 Missud (2023)

Dans la section précédente, nous avons vu que Lindsay & Aronoff (2013) ont introduit une perspective écologique afin d'expliquer la compétition morphologique, une contribution qui a exercé une influence notable sur ce domaine. S'inscrivant dans leur lignée, Missud (2023) poursuit cette approche en développant le concept de niche linguistique dans une perspective morphologique renouvelée.

Contrairement à la définition originale proposée par Lindsay et Aronoff (2013), qui mettait l'accent sur la distribution des niches linguistiques, Missud (2023) redéfinit la niche morphologique comme un ensemble de propriétés linguistiques distinctives issues de différents niveaux d'analyse linguistique. C'est cet ensemble qui constitue une niche, et non la seule spécialisation fonctionnelle. Elle y intègre également une contrainte paradigmatique, qu'elle considère comme un facteur central dans l'organisation morphologique. Plus précisément, selon Missud (2023 : 27) : « La théorie des niches morphologiques implique que l'organisation générale de la langue, en perpétuelle évolution, justifie qu'il est difficile de résoudre la compétition de façon catégorique par le recensement de propriétés discriminatives discrètes. Les schémas morphologiques s'organisent de sorte à investir progressivement des niches, et les contraintes recensées reflètent donc des tendances. En outre, le lexique semble être

organisé de sorte à autoriser la compétition mais à contraindre progressivement les schémas morphologiques à investir chacun des espaces de manière privilégiée pour survivre. »

Autrement dit, pour Missud (2023), une niche morphologique est définie par un ensemble de propriétés linguistiques (phonologiques, syntaxiques, sémantiques, etc.) distinctives. Certaines de ces propriétés peuvent être partagées par plusieurs schémas morphologiques, tandis que d'autres sont spécifiques, permettant ainsi de différencier un schéma morphologique de ses concurrents. Ces propriétés constituent alors les conditions de survie du suffixe dans le système linguistique.

Nous allons à présent examiner un exemple concret afin d'illustrer de manière plus claire la définition de la niche morphologique proposée par Missud (2023 : 134). Prenons le cas du suffixe *-ion* dans le nom DARWINISATION. Les propriétés linguistiques possibles qui composent sa niche morphologique sont les suivantes :

- Propriétés phonologiques :
  - La longueur de la base (en caractères) : 6
  - La nature du dernier caractère de la base : consonne
  - La nature du dernier phonème de la base : consonne
- Propriétés morphologiques :
  - Le schéma à l'origine de la construction du verbe : la suffixation en *-iser*
- Propriétés syntaxiques :
  - La catégorie grammaticale de la base : nom
  - La catégorie grammaticale du dérivé : verbe
- Propriétés sémantiques :
  - Le type ontologique de la base : animé
  - Le type sémantique du verbe : verbe similatif

La différence entre les conceptions de la niche linguistique chez Lindsay & Aronoff (2013) et Missud (2023) se situe à trois niveaux : l'unité de la niche, ses composantes, ainsi que la notion de niche écologique qui leur sert de cadre théorique. Commençons par l'unité de la niche. Pour Lindsay & Aronoff (2013), chaque propriété linguistique constitue une niche linguistique à part entière pour les schémas morphologiques. À l'inverse, chez Missud (2023), la niche est définie comme un ensemble de propriétés linguistiques distinctives.

Ainsi, les conceptions divergent sur ce qu'est une niche morphologique. Pour Lindsay & Aronoff (2013), une niche morphologique est définie par une propriété

linguistique ; les schémas morphologiques cherchent alors à occuper des niches linguistiques (dont la niche morphologique fait partie) pour assurer leur survie. En revanche, pour Missud (2023), les schémas évoluent au sein des niches morphologiques, lesquelles sont formées d'un ensemble de propriétés linguistiques distinctives, incluant mais dépassant les seules propriétés morphologiques. Dans cette perspective, la propriété morphologique n'est qu'un élément parmi d'autres à l'intérieur d'une niche morphologique.

Enfin, une divergence importante concerne le modèle écologique sous-jacent à ces définitions. Comme nous l'avons vu, Lindsay et Aronoff (2013) fondent leur approche sur le principe d'exclusion compétitive (Gause 1934), lui-même basé sur la définition spatiale de la niche chez Grinnell (1917), centrée sur la distribution et l'habitat. À l'inverse, Missud (2023) redéfinit la niche morphologique comme un ensemble de propriétés linguistiques distinctives issues de différents niveaux d'analyse. Certaines de ces propriétés peuvent être partagées entre plusieurs suffixes, mais seules les propriétés distinctives permettent de définir la niche propre à un suffixe donné.

Cette approche présente une certaine affinité avec la définition multidimensionnelle de la niche chez Hutchinson (1957), qui conçoit la niche comme un espace structuré par des variables multiples, issues de différentes ressources. Toutefois, un point fondamental les distingue : chez Hutchinson, toute variable mobilisée par une espèce, qu'elle soit distinctive ou non, fait partie intégrante de la niche, ce qui autorise le chevauchement entre niches. En revanche, chez Missud (2023), la niche morphologique est limitée aux seules propriétés distinctives, ce qui exclut par principe toute logique de recouvrement. Bien que Missud (2023) développe de manière novatrice le concept de niche morphologique proposé par Lindsay & Aronoff (2013) pour rendre compte de la compétition entre suffixes, certaines limites demeurent dans son approche. Celles-ci seront discutées plus en détail dans la section 2.4. Dans la section 2.3, nous aborderons diverses approches et travaux portant sur la compétition morphologique, en particulier dans le contexte de l'étude du français.

2.3 Deux approches influentes pour la compétition morphologique

Après avoir présenté les notions fondamentales relatives à la compétition

morphologique, cette section se concentrera sur les principales approches théoriques et les travaux empiriques consacrés à ce phénomène. Plus précisément, la littérature distingue deux grandes approches sur lesquelles se fondent la plupart des analyses : une approche monofactorielle et une approche multifactorielle. Nous commencerons par examiner l'approche monofactorielle, avant de nous tourner vers l'approche multifactorielle, plus récente et innovante.

2.3.1 Approche monofactorielle

En lien avec la compétition de la niche écologique de Lindsay & Aronoff (2013), les auteurs qui adoptent cette approche postulent qu'il existe une propriété linguistique (phonologique, morphologique, sémantique, syntaxique, etc.) particulièrement discriminante pour le choix des suffixations. Nous désignons ainsi cette approche comme l'approche monofactorielle. Rappelons qu'à la différence de la manière dont Missud (2023), l'approche monofactorielle consiste à prendre en compte une seule variable. Toutefois, l'approche monofactorielle dans Missud (2023), comprend plusieurs variables au même niveau linguistique. Elle valorise le même niveau linguistique comme le prédicteur des suffixes. Ainsi, dans ce présent travail, selon nous, l'approche de catégorisation proposée par Missud (2023) relève d'un modèle multifactoriel, plusieurs variables étant prises en compte pour prédire le choix des suffixes. Nous présentons d'abord la perspective synchronique, puis la perspective diachronique, en suivant l'ordre des niveaux linguistiques les plus souvent traités dans la littérature : phonologique, morphologique, syntaxique, et sémantique.

Tout d'abord, en ce qui concerne les recherches en synchronie, nous commençons par le niveau phonologique. À ce niveau, seule la taille du radical de la base du dérivé est mentionnée. Par exemple, en anglais, Lindsay (2012) et Lindsay & Aronoff (2013) ont constaté que le suffixe *-ify* tend à sélectionner des bases monosyllabiques (*par ex.* PURE > PURIFY), tandis que le suffixe *-ize* privilégie les bases polysyllabiques (*par ex.* PRIORITY > PRIORITIZE). Sur le plan morphologique, la structure de la base du dérivé, l'étymologie de cette base, ainsi que le groupe de conjugaison du verbe peuvent être étudiés. Parmi ces propriétés, la structure de la base du dérivé est la plus approfondie. Ainsi, en anglais, Anshen & Aronoff (1981) ont constaté que, par rapport au suffixe *-ness* (*par ex.* SAD > SADNESS), le suffixe *-ity* préfère les bases suffixées en *-able* (*par ex.* SUITABLE > SUITABILITY). En français, Koehl (2012), qui a étudié les suffixes nominaux désadjectivaux, a mis en évidence le fait que le suffixe *-ité* tend à sélectionner

des bases en *-able*, *-ique*, *-if*, *-eux*, *-al*, *-aire*. En outre, le suffixe *-ise* favorise les bases en *-ard* et *-ant*, tandis que les suffixes *-esse* et *-eur* ont tendance à éviter les adjectifs déjà suffixés. Des exemples sont présentés en (1) :

(1) a. LISIBLE > LISIBILITE
b. HISTORIQUE > HISTORICITE
c. SPORTIF > SPORTIVITE
d. DANGEREUX > DANGEROSITE
e. NATIONAL > NATIONALITE
f. ALIMENTAIRE > ALIMENTARITE
g. DEBROUILLARD > DEBROUILLARDISE
i. EMOUVANT > EMOUVANTISE
j. FAIBLE > FAIBLESSE
k. DOUX > DOUCEUR

Toujours pour ce qui concerne le français, Strnadová (2014) s'est intéressée à la suffixation des adjectifs dénominaux. Elle identifie trois niches morphologiques : *-aire* a une propension à sélectionner les bases nominales en *-ité* (*par ex.* SANTE > SANITAIRE) ; *-el* préfère les noms déverbaux en *-ion* (*par ex.* DERIVATION > DERIVATIONNEL) ; et *-ique* tend à choisir les noms suffixés en *-iste* (*par ex.* JOURNALISTE > JOURNALISTIQUE). Par ailleurs, les suffixes événementiels ont suscité une attention particulière. Missud & Villoing (2020 : 29) et Villoing & Gueye (2024) examinent les structures morphologiques privilégiées par ces suffixes. Concrètement, le suffixe *-ion* sélectionne les bases verbales dérivées en *-iser* et *-ifier* (*par ex.* MODELISER > MODELISATION ; AMPLIFIER > AMPLIFICATION), *-age* privilégie les bases verbales converties (*par ex.* FREINER > FREINAGE), et *-ment* préfère les préfixations déverbales à interprétation itérative ou inversive (*par ex.* REACCOMPAGNER > REACCONPAGNEMENT ; DECOMMANDER > DECOMMANDEMENT), ainsi que les bases verbales issues de conversions dénominatives (*par ex.* NOYAUTER > NOYAUTEMENT).

En ce qui concerne l'étymologie de la base du dérivé, certains morphologues (Lindsay 2012 ; Lindsay & Aronoff 2013) relèvent son influence sur le choix des suffixes. Lindsay (2012) se concentre sur les suffixes anglais *-ity* et *-ness*, et soutient que *-ity* a pu se maintenir grâce à des niches morphologiques correspondant à des bases dérivées à l'aide de suffixes latins (*par ex.* PRODUCTIVE > PRODUCTIVITY), tandis que son concurrent *-ness* privilégie les bases finies par des suffixes germaniques (*par ex.* SAD > SADNESS). De plus, Lindsay & Aronoff (2013) identifient une niche morphologique favorable au suffixe *-ical* : les radicaux en *-olog* (*par ex.* MORPHOLOGY > MORPHOLOGICAL). À l'inverse, le suffixe *-ic* est associé à d'autres types de radicaux, ce qui constitue une distribution complémentaire. Enfin, s'agissant du groupe de conjugaison de la base verbale, Lapraye (2017), en étudiant les suffixes événementiels,

observe que le suffixe *-age* préfère les verbes du premier groupe, tandis que le suffixe *-ment* sélectionne les verbes des deuxième et troisième groupes.

Au niveau syntactico-sémantique, les recherches se concentrent principalement sur les propriétés sémantiques de la base verbale et sur la structure argumentale. Concernant le type de la base, Ferret et al. (2010) ont observé que les noms convertis sur le thème du participe présent féminin en *-ée* préfèrent les bases inaccusatives (*par ex.* MONTER > MONTEE), tandis que le suffixe *-age* tend à sélectionner des bases transitives (*par ex.* ASSEMBLER> ASSEMBLAGE). De même que Ferret et al. (2010), Fradin (2019) s'est focalisé sur la structure argumentale pour analyser les doublons en *-age* et en *-ment*. Il propose, bien que les suffixations en *-age* et *-ment* partagent un sens événementiel similaire, qu'elles exploitent des structures argumentales différentes du même verbe pour former des noms événementiels. Fradin (2019 : 78) illustre cette spécificité syntaxique à travers les doublons EMBALLAGE/EMBALLEMENT. Pour le verbe EMBALLER, il distingue trois lexèmes distincts selon leurs propriétés argumentales : « EMBALLER$_1$ est à la tête d'une construction agentive X EMBALLER Y et dénote un accomplissement, tandis que EMBALLER$_2$ est à la tête d'une construction spatiale X EMBALLER Y dénotant un état, où X est le moyen d'emballer Y. Quant à S'EMBALLER$_3$, il était à l'origine la tête d'une construction anticausative X s'EMBALLER, où X dénote un cheval ou, métaphoriquement, un moteur ou un système en mouvement rapide. » En ce qui concerne la nominalisation, le verbe EMBALLER$_1$ permet de dériver les noms EMBALLAGE/EMBALLEMENT, tandis que seule la forme EMBALLAGE est dérivé par le lexème EMBALLER$_2$. En revanche, s'EMBALLER$_3$ donne lieu à la nominalisation EMBALLEMENT et non EMBALLAGE. Nous reprenons des exemples de Fradin (2019 : 78) pour bien illustrer ce résultat.

(2) a. lois et réglages sur (. . .) l'emballage 1 des médicaments
b. L'emballement 1 des marchandises est terminé.
c. L'(emballage 2|*emballement 1) est déchiré.
d. l'emballement 3 des principales bourses

Le niveau sémantique représente un autre intérêt pour les morphologues. Ils s'intéressent particulièrement aux fonctions sémantiques des suffixes ainsi qu'aux propriétés sémantiques de la base du dérivé. Concernant les fonctions sémantiques, Lieber (2016) avance que les nominalisations sont généralement polysémiques et entrent en compétition en termes de fonctions sémantiques. Salvadori & Huyghe (2023), en analysant l'ensemble des nominalisations verbales en français, confirment les

résultats de Lieber (2016). Par exemple, les suffixes événementiels *-ment*, *-ade* et *-ure* partagent souvent certaines fonctions sémantiques. Toutefois, selon la similarité des fonctions dénotées, *-ment* se rapproche davantage de *-ade* que de *-ure*. En français, concernant le suffixe *-age* et la conversion, Missud & Villoing (2021) ont constaté que la conversion V>N construite sur le thème 12 (identique à la forme du participe passé sans porter la propriété ni grammaticale ni sémantique) présente un comportement très similaire à celui de la suffixation en *-age*. Néanmoins, la conversion V>N construite sur le thème 12 tend à dénoter plus fréquemment des noms de résultat ou d'événement (*par ex.* ARRIVER > ARRIVEE), ce qui marque une distinction sémantique entre les deux schémas morphologiques.

À partir de ces travaux, plusieurs morphologues et sémanticiens (Huyghe & Wauquier 2021 ; Guzmán Naranjo & Bonami 2023 ; Salvadori, Varvara & Huyghe 2024) ont proposé le concept de mode de compétition, distinguant notamment la compétition totale de la compétition partielle. Selon Huyghe et Varvara (2023 : 5), pour les suffixes polysémiques, « les différences sémantiques observées entre les affixes déterminent divers degrés de rivalité qui devraient être analysés et quantifiés en tant que tels ». En outre, Salvadori et al. (2024) ont adapté les méthodes de mesure de similarité proposées par Sørensen (1948) et Odum (1950), tout en y intégrant des indices de dissimilarité en tant que mesures complémentaires. L'objectif est d'évaluer la similarité et la dissimilarité des niches morphologiques du point de vue des propriétés sémantiques. Ils ont trouvé que la relation de compétition entre les suffixes est complexe. Pour analyser les similarités et les dissimilarité des niches, il faut prendre en compte à la fois les fonctions sémantiques partagées, les fonctions sémantiques non partagées, ainsi que la distribution et la fréquence d'utilisation de ces fonctions dans les dérivés.

A propos des propriétés sémantiques de la base du dérivé, en français, pour les suffixes événementiels *-age*, *-ment* et *-ion*, Martin (2010) a observé que le suffixe *-age* tend à dénoter un événement physique dans le domaine ontologique (cf. exemple 3a), tandis que le suffixe *-ion* favorise un événement abstrait dans le domaine ontologique (cf. exemple 3b). Contrairement à *-ion*, le suffixe *-ment* se distingue également par la base portant la propriété de discontinuité. Si le verbe de base représente une action composée de plusieurs étapes, cela favorise alors le choix du suffixe *-ion* pour former le nom d'action (cf. exemple 3c) .

(3) a. *un cassage de doigt, #une cassation de doigt*. (Martin 2010 : 133)

b. *la cassation d'une décision juridique, #le cassage d'une décision juridique*. (Martin 2010 : 133)

c. discontinuité : *Samira a alphabétisé Pierre en plusieurs étapes*. → *l'alphabétisation (!alphatisement) de Pierre en plusieurs étapes.* (Martin 2010 : 132)

A propos des approches synchroniques / diachroniques, les recherches menées dans une perspective synchronique sont plus nombreuses que celles adoptant une perspective diachronique, mais ces dernières produisent également des résultats remarquables. Tout d'abord, Uth (2010) s'appuie sur une perspective diachronique pour étudier les suffixes *-age* et *-ment*. Elle constate qu'en ce qui concerne l'interprétation événementielle, le suffixe *-age* connaît un véritable essor au XIXe siècle. Elle soutient qu'en français contemporain, des nuances sémantiques subsistent entre les deux suffixes : « le suffixe *-age* met l'accent sur la propriété des référents sujets dans l'événement désigné par le verbe de base, tandis que le suffixe *-ment* nominalise la propriété des arguments Thème à l'état résultant de l'événement de base » (Uth 2010 : 239). Des exemples sont fournis en (4). Dans l'exemple (4a), le sujet de la phrase est *l'avion*, qui joue le rôle d'agent dans la structure argumentale de la base verbale du lexème DECOLLAGE, à savoir *l'avion décolle*. En revanche, pour le lexème GRINCEMENT dans l'exemple (4b), le suffixe *-ment* nominalise le thème comme état résultant de l'événement de base, à savoir *le chariot grince*.

(4) a. *Vingt minutes après le décollage, l'avion fut abattu par un…* (Uth 2010 : 235)

b. *On entendait le grincement du chariot*. (Uth 2010 : 234)

En outre, Cartier & Huyghe (2021) s'intéressent aux préfixes de haut degré (ci-après PHD) en français : *archi-*, *extra-*, *hyper-*, *maxi-*, *méga-*, *super-*, *supra-* et *ultra-*. Ils observent que « la concurrence entre PHD ne se résout [...] pas en diachronie. La plupart des doublons lexicaux préfixés par *hyper-* et *ultra-* n'évoluent pas dans le sens d'une réduction des formes en diachronie » (Cartier & Huyghe 2021 : 20–21). Cela signifie que ces préfixes coexistent dans une certaine mesure longtemps.

En complément de ces études, j'ai moi-même travaillé sur ce sujet dans le cadre de Lin (2021). Mon travail porte sur la suffixation des noms d'instruments en adoptant une perspective ontologique. Il a mis en évidence que le principe d'exclusion compétitive présente certaines limites pour expliquer la compétition morphologique. Plus précisément, dans le cas des noms de récipients et d'ustensiles, dont la productivité reste faible, la compétition entre les suffixes concernés demeure irrésolue du XIIe siècle

jusqu'à aujourd'hui. À partir de ces observations, j'ai proposé trois modes de compétition, fondés sur une approche monofactorielle dans laquelle le facteur sémantico-ontologique guide la compétition : la compétition absolue, la compétition non-absolue et la compétition alternative. Dans le cas de la compétition absolue, un seul affixe reste dominant sur plusieurs siècles. A titre d'exemple, concernant la sous-classe ontologique Appareil, le suffixe *-eur* est plus dominant que les autres suffixes instrumentaux depuis le XIX siècle (Lin 2021 : 30). En revanche, la compétition alternative désigne une situation où deux ou trois affixes coexistent, les plus innovants surpassant progressivement les plus anciens en termes de productivité. Selon mon travail en 2021, la sous-classe ontologique Instrument de Captation est conforme à ce mode de compétition. Quant à son évolution, le suffixe *-oir* est plus dominant que les autres suffixes instrumentaux avant le XIXe siècle, tandis qu'après le XIXe siècle, le suffixe *-eur* dépasse *-oir* et devient le nouveau suffixe dominant dans cette sous-classe ontologique. Enfin, la compétition non-absolue correspond à un cas où aucun patron suffixal ne s'impose durablement au fil des siècles. Par exemple, selon mes propres travaux (Lin 2021 : 32), dans les sous-classes ontologiques Récipient et Ustensile, en raison d'une faible productivité, il n'existe aucun suffixe plus dominant que les autres du XIe au XXe siècle. Ce qui est conforme au mode de compétition non-absolue. Le premier mode de compétition valide le postulat de Lindsay & Aronoff (2013), tandis que les deux derniers en révèlent certaines limites. Par ailleurs, j'ai tenté d'expliquer ces résultats en mobilisant des facteurs socioculturels : l'évolution industrielle, notamment, joue un rôle important dans la productivité et la compétition des suffixes instrumentaux.

À travers les recherches examinées, les auteurs ont réussi à identifier des niches linguistiques distinctives pour les schémas morphologiques en adoptant une approche monofactorielle. Toutefois, plusieurs morphologues (Plag 1999 ; Arndt-Lappe 2014 ; Bonami & Thuilier 2019) soulignent que certains suffixes se distinguent par plus d'une propriété linguistique. Autrement dit, plusieurs facteurs peuvent interagir simultanément dans le choix des suffixes. Ces auteurs proposent donc une approche multifactorielle. Il est important de noter que, bien que notre recherche applique une approche multifactorielle pour modéliser les données et prédire les suffixes, les propriétés linguistiques identifiées dans le cadre de l'approche monofactorielle fournissent également des repères précieux pour la sélection des prédicteurs. Dans la partie suivante, nous présenterons cette approche novatrice en matière de compétition

morphologique.

2.3.2 Approche multifactorielle

À la différence des auteurs qui adoptent une approche monofactorielle (§2.3.1), certains morphologues (Plag 1999 ; Arndt-Lappe 2014 ; Bonami & Thuilier 2019) avancent l'hypothèse que, pour certains suffixes, plusieurs propriétés linguistiques exercent des effets interactifs sur la sélection des schémas morphologiques. Dans ce cadre, ces propriétés présentent des poids différenciés et s'organisent de manière hiérarchique lors de la compétition morphologique. Cette perspective a conduit à l'émergence d'une approche novatrice : l'approche multifactorielle. En outre, pour identifier les propriétés linguistiques discriminantes, les morphologues privilégient de plus en plus les méthodes quantitatives. Comme le soulignent Huyghe & Varvara (2023 : 10), « les méthodes quantitatives semblent bien adaptées à l'étude de la rivalité des affixes, car elles sont capables de modéliser des phénomènes graduels et d'évaluer l'influence relative de plusieurs facteurs. Grâce à l'analyse statistique, elles permettent de déterminer si certaines différences observées sont significatives ou simplement dues au hasard, et offrent un cadre objectif pour tester les hypothèses de recherche ». Cela ouvre de nouvelles perspectives aux morphologues pour explorer et formaliser la compétition morphologique.

Dans la littérature, on distingue généralement deux types de démarche. La première consiste à présenter des fréquences préférentielles pour chaque schéma morphologique ; la seconde recourt à des modélisations statistiques pour prédire le choix des suffixes à partir de propriétés linguistiques. Suivant cette logique, nous commencerons par les travaux fondés sur les statistiques descriptives, avant de présenter ceux utilisant des modélisations statistiques.

S'agissant des recherches fondées sur les statistiques descriptives, elles concernent notamment le niveau phonologique. Par exemple, Plag (1999), Koehl (2012) et Strnadová (2014) prennent en compte des propriétés telles que la taille du radical de la base du dérivé, la prosodie de ce radical et son segment final, afin d'identifier les préférences associées aux différents suffixes. Ainsi, en anglais, à propos des suffixes *-ify*, *-ize* et *-ate*, Plag (1999) observe que les radicaux des bases des dérivés en *-ify* sont souvent monosyllabiques (*par ex.* TECHNICAL > TECHNIFY), tandis que le suffixe *-ize* privilégie les radicaux dissyllabiques (*par ex.* TECHNICAL > TECHNICIZE). Parallèlement,

-ify préfère les bases iambiques[21] (*par ex.* KARST > KARSTIFY), tandis que *-ate* tend à sélectionner les bases trochaïques[22] (*par ex.* CANNULA > CANNULATE). Ces deux types de dérivés requièrent une alternance stricte de l'accentuation. En français, dans son étude sur les suffixes nominaux désadjectivaux, Koehl (2012) constate que les suffixations en *-ise* et *-erie* sont associées à des bases dissyllabiques ou trisyllabiques (*par ex.* MARCHAND > MARCHANDISE ; SAUVAGE > SAUVAGERIE), tandis que les suffixations en *-esse*, *-itude* et *-eur* préfèrent des bases monosyllabiques ou dissyllabiques (*par ex.* FAIBLE > FAIBLESSE ; ROND > RONDITUDE ; DOUX > DOUCENR). Par ailleurs, en ce qui concerne le segment final du radical, *-itude* sélectionne préférentiellement les consonnes /t/ et /ʁ/, tandis que *-ité* privilégie les consonnes /d/, /s/ et /l/.

Concernant les suffixes adjectivaux dénominaux, Strnadová (2014) relève que la suffixation en *-ique* est favorisée par les radicaux les plus longs (*par ex.* ACADEMIE > ACADEMIQUE), alors que la suffixation en *-u* privilégie les bases monosyllabiques (*par ex.* FEUILLE > FEUILLU). Pour ce qui est du segment phonologique final, le suffixe *-el* s'attache plus fréquemment aux noms se terminant par /ɔ̃/ (*par ex.* PROFESSION > PROFESSIONNEL), tandis que *-ique* favorise les bases se terminant par une consonne.

Par ailleurs, certains auteurs prennent en considération des propriétés issues de différents niveaux linguistiques pour prédire le choix des suffixes en recourant à des modélisations statistiques. Ainsi, Arndt-Lappe (2014) met l'accent sur les propriétés phonologiques et syntaxiques en utilisant le modèle analogique développé par Skousen & Stanford (2007), afin d'explorer la compétition morphologique entre les suffixes *-ness* et *-ity*. Contrairement à d'autres modèles statistiques fondés sur des propriétés linguistiques sélectionnées *a priori* par les chercheurs, ce modèle « est capable de prédire avec succès la distribution dans toutes les expériences, uniquement sur la base de la composition phonologique des deux syllabes finales de la base et de quelques informations syntaxiques élémentaires » (Arndt-Lappe 2014 : 500). Plus précisément, dans l'étude d'Arndt-Lappe (2014) sur la compétition entre *-ness* et *-ity* en anglais (*par ex.* ACCEPTABLE > ACCEPTABLENESS), la structure morphologique de la base du dérivé s'avère déterminante pour différencier les deux suffixes. L'auteure montre notamment que, du XVIIIe au XXe siècle, à l'exception du suffixe *-ive* (*par ex.* CONNECTABLE >

[21] Une base iambique désigne une base morphologique dont la structure prosodique suit le schéma faible–fort (iambique) (McMahon 2020 :137).
[22] Une base trochaïque désigne une base morphologique dont la structure prosodique suit le schéma fort–faible (S–W) (McMahon 2020 :137).

CONNECTABILITY ; ACCUMULATIVE > ACCUMULATIVENESS), les autres suffixes d'origine latine sont de plus en plus associés à *-ity*.

Sur la base des propriétés phonologiques et sémantiques prises en compte par Arndt-Lappe (2014), Bonami & Thuilier (2019) ajoutent des propriétés morphologiques comme variables prédictives pour modéliser la préférence entre les suffixes *-ifier* et *-iser* en français, en s'appuyant sur une régression logistique. Cette méthode permet de « modéliser la relation entre la variable dépendante et une ou plusieurs variables indépendantes, et d'effectuer des prédictions sur cette base » (Xu et al. 2023 : 267). Grâce à cet outil, les auteurs montrent que la structure morphologique et le nombre de syllabes du radical constituent des facteurs prépondérants pour distinguer les suffixations en *-ifier* et en *-iser*. Par exemple, les bases obtenues par conversion ou celles contenant le suffixe *-ique* tendent à sélectionner *-iser* (*par ex.* CARDINAL > CARDINALISER ; FANATIQUE > FANATISER), tandis que d'autres structures morphologiques préfèrent *-ifier* (*par ex.* AMPLE > AMPLFIER). Sur le plan phonologique, la présence d'une consonne sonore favorise également le choix de *-iser* (*par ex.* EGALE > EGALISER). Par ailleurs, plus le radical est long (en nombre de syllabes), moins il est probable qu'il sélectionne le suffixe *-ifier* (*par ex.* BON > BONIFIER).

En fonction des spécificités des langues et des suffixes étudiés, certains auteurs ajoutent encore d'autres propriétés linguistiques afin de mieux prédire le choix des suffixes, notamment à l'aide de l'arbre de décision. Bobkova (2023), par exemple, prend en compte des propriétés morphologiques, phonologiques, sémantiques et étymologiques pour étudier la suffixation des adjectifs dénominaux du russe, en particulier les suffixes *-n-*, *-sk-* et *-ov-*. Elle mobilise l'arbre de décision pour modéliser les données et montre que ces suffixes sont sensibles à des propriétés issues de différents niveaux linguistiques. Plus précisément, les propriétés sémantiques s'avèrent particulièrement discriminantes pour le suffixe *-sk-* dans sa distinction avec les deux autres. Sur le plan phonologique, le suffixe *-ov-* manifeste une préférence pour les bases contenant des consonnes vélaires ainsi que pour les noms monosyllabiques. Quant aux autres propriétés linguistiques (telles que la nature grammaticale du nom, la classe du phonème final, etc.), elles favorisent davantage le suffixe *-n-*, par exemple les noms communs se terminant par un phonème alvéolaire, dental ou labial.

A partir des propriétés retenues par Bonami & Thuilier (2019), Missud (2023) ajoute les propriétés syntaxiques pour analyser la compétition morphologique entre les suffixes des nominalisations événementielles déverbales en français. À l'aide de l'arbre

de décision, elle montre que, parmi les différents niveaux linguistiques, les propriétés morphologiques sont les plus significativement discriminantes pour distinguer les suffixes en question. Plus précisément : « la distinction la plus frappante s'observe sur le schéma à l'origine du verbe sélectionné par le suffixe, qui peut être une suffixation en *-iser* ou *-ifier*, une conversion d'adjectif à verbe ou de nom à verbe, ou une préfixation en *a-, en-* ou *é-*. Celui-ci constitue un facteur déterminant pour distinguer les suffixes : les nominalisations en *-ion* sont principalement construites sur des verbes suffixés, les nominalisations en *-age* sur des verbes convers, et celles en *-ment* sur des verbes préfixés » (Missud 2023 : 182). Selon l'auteure, ces résultats confirment l'hypothèse de Lindsay & Aronoff (2013), selon laquelle la compétition morphologique entre les suffixes événementiels en français obéit bien au principe d'exclusion compétitive (Gause 1934).

Hormis les propriétés linguistiques évoquées, certains auteurs intègrent également des propriétés extralinguistiques, en fonction de leurs objectifs et des suffixes analysés. À titre d'exemple, en utilisant la forêt aléatoire, Thuilier et al. (2023) ajoutent des variables géographiques pour modéliser la formation des gentilés. Partant de l'hypothèse que les suffixes concernés sont sensibles à des propriétés géographiques, telles que la ville, le pays ou la région, elles constatent que « quatre caractéristiques géographiques contribuent de manière significative à la discrimination des suffixes » (Thuilier et al. 2023 : 140), validant ainsi leur hypothèse initiale.

À travers les travaux présentés, il apparaît que les recherches sur la compétition morphologique foisonnent dans les deux approches. Grâce à des analyses empiriques rigoureuses, les auteurs ayant recours à l'approche multifactorielle valident leurs hypothèses et remettent en question l'approche monofactorielle. En particulier, selon les langues et les suffixes à étudier, les auteurs sélectionnent les propriétés linguistiques les plus pertinentes afin de modéliser les données et de prédire le choix des suffixes. Ces propriétés sont variées : elles couvrent les domaines de la phonologie, de la morphologie, de la syntaxe, de la sémantique, mais également des dimensions extralinguistiques. Cet ensemble constitue une base solide pour orienter notre propre recherche et justifier la sélection des propriétés les plus adéquates à intégrer dans notre modèle. Ainsi, dans cette thèse, nous adopterons une approche multifactorielle et sélectionnerons les propriétés les plus appropriées en fonction des cadres théoriques mobilisés et des suffixes analysés, afin de mener à bien notre étude. Jusqu'à présent, nous avons présenté respectivement les notions, les approches et les travaux de

recherche relatifs à la compétition morphologique dans les trois sections précédentes. Toutefois, ces trois volets présentent encore des problématiques et certaines limites, que nous allons examiner dans la section suivante.

2.4 Critiques

Cette section vise à formuler quelques remarques finales sur la compétition morphologique, afin de clore ce chapitre. Elle traite plus précisément des points suivants : l'applicabilité du principe d'exclusion compétitive à la morphologie, la notion de niche morphologique, les outils utilisés pour modéliser les données et visualiser les niches morphologiques, la mesure du chevauchement entre des niches morphologiques, ainsi que les propriétés retenues pour discriminer les suffixes en compétition.

Nous commençons par la notion de niche morphologique. Comme mentionné précédemment, cette notion a été introduite pour la première fois par Lindsay & Aronoff (2013). Selon eux, la compétition morphologique peut être interprétée de manière pertinente à travers le principe d'exclusion compétitive formulé par Gause (1934). Aronoff a d'ailleurs réaffirmé cette idée dans plusieurs publications ultérieures, soulignant notamment, en 2023, l'efficacité de ce principe pour expliquer la distribution des suffixes sans mentionner de limites théoriques majeures. Cependant, au cours de la décennie qui a suivi (2013–2023), Lindsay & Aronoff n'ont pas étendu leur analyse à un grand nombre de suffixes simultanément ; leurs recherches se sont principalement concentrées sur un nombre restreint de suffixes, et cela dans une perspective diachronique. Bien que certaines études synchroniques, comme celle de Missud (2023), viennent étayer leur postulat, elles se focalisent également sur un nombre limité de suffixes en compétition. Il semble donc légitime de s'interroger sur la validité généralisée de ce principe à la lumière d'études récentes, telles que celles de Lin (2021 : 33) et de Huyghe & Lombard (2022 : 44), qui en soulignent les limites.

Dans le cas des sous-classes ontologiques de noms d'instruments à productivité restreinte, nous pourrions formuler l'hypothèse que le principe d'exclusion compétitive ne suffit pas à rendre compte de l'évolution morphologique. Sur le long terme, aucun schéma morphologique ne s'impose durablement entre les suffixes en compétition. En outre, pour les sous-classes ontologiques plus productives, des facteurs extralinguistiques, notamment socioculturels, comme la révolution industrielle, ainsi

que la productivité actuelle des suffixes, induisent un mode de compétition alternatif. Ces deux observations ne s'accordent pas avec les prédictions du principe d'exclusion compétitive. Par conséquent, dans le cadre de cette thèse, il sera essentiel de réévaluer de manière approfondie le postulat de Lindsay & Aronoff (2013), en mobilisant des données à large échelle et des outils quantitatifs adaptés, afin d'en tester la portée explicative au-delà des cas déjà explorés.

De plus, bien que la notion de niche morphologique proposée par Lindsay & Aronoff (2013) ait été développée par Missud (2023), cette redéfinition présente néanmoins certaines limites. Comme mentionné dans la section 2.3.2, Missud parvient à identifier certaines propriétés linguistiques discriminantes permettant de différencier les suffixes étudiés. Toutefois, hormis cette fonction classificatoire, son analyse n'exploite pas pleinement les potentialités inhérentes à la notion de niche elle-même. En particulier, le chevauchement des niches n'est ni défini ni mis à profit. Or, comme l'ont montré les travaux en écologie (Hutchinson 1957 ; Ge 2008 ; Pocheville 2011), la différenciation entre les niches résulte souvent d'un chevauchement partiel entre niches. Ces observations invitent ainsi à remettre en question la définition formulée par Missud (2023) et à proposer une nouvelle conception des niches morphologiques, inspirée de Hutchinson (1957), mais adaptée aux spécificités du système morphologique du français. Cette nouvelle définition sera développée au chapitre 6.

Par ailleurs, parmi les outils statistiques mobilisés dans les travaux précédemment cités, figurent notamment le modèle analogique, la régression logistique, l'arbre de décision et la forêt aléatoire. Bien que ces outils s'y soient révélés efficaces, ils ne s'avèrent pas tous adaptés aux spécificités de notre recherche. Il convient de souligner que notre base de données intègre une diversité de propriétés linguistiques (phonologiques, morphologiques, syntaxiques et sémantiques), dont certaines présentent une forte colinéarité : autrement dit, certains phénomènes linguistiques affectent simultanément plusieurs propriétés. Cette double caractéristique a orienté notre choix méthodologique. D'une part, la prise en compte exclusive de propriétés phonologiques et syntaxiques rend le modèle analogique inadapté. D'autre part, la régression logistique se montre peu performante dans le traitement de données colinéaires, ce qui limite sa pertinence dans notre cas. Il subsiste alors deux options : l'arbre de décision et la forêt aléatoire. L'arbre de décision, bien qu'intéressant, est sensible aux variations des données et présente un risque élevé de sur-apprentissage (*overfitting*), ce qui en compromet la stabilité prédictive. En revanche, la forêt aléatoire

compense ces faiblesses en construisant un grand nombre d'arbres de décision en parallèle, ce qui permet de stabiliser les résultats et de mieux gérer la colinéarité des variables. Ainsi, nous avons opté pour la forêt aléatoire comme méthode de modélisation pour notre analyse (voir § 4.3.3).

En outre, bien que de nombreuses études s'intéressent aux niches morphologiques, elles ne cherchent généralement pas à les visualiser, mais plutôt à identifier les propriétés les plus discriminantes. Or, la visualisation de ces propriétés permettrait une compréhension plus globale et intuitive de la structuration des niches associées aux suffixes étudiés. Dans le cas de données caractérisées par des intersections complexes, deux outils de visualisation se distinguent : le diagramme de Venn et UpsetR. Le diagramme de Venn, bien que répandu, s'avère peu optimal lorsque le nombre de variables augmente, en raison du risque de surcharge visuelle et de perte de lisibilité. À l'inverse, UpsetR permet de représenter efficacement un grand nombre d'intersections sous forme de barres, ce qui facilite l'interprétation et la clarté des résultats. En tenant compte des caractéristiques de nos données, nous avons donc choisi d'utiliser UpsetR pour visualiser les niches morphologiques des suffixes liés à la dénomination des noms d'instruments (voir § 4.3.2).

En plus des points évoqués, il convient de souligner que, bien que Salvadori et al. (2024) adaptent les méthodes de mesure de similarité issues de Sørensen (1948) et Odum (1950), en y intégrant certains indices de dissimilarité à titre complémentaire, ces approches restent fondamentalement limitées dès lors qu'il s'agit de modéliser la compétition morphologique entre schémas morphologiques. En effet, les indices de similarité symétriques ne permettent pas de distinguer les relations directionnelles, c'est-à-dire les cas où un schéma chevauche partiellement ou totalement un autre sans que l'inverse soit vrai (Hurlbert 1978 : 69-70 ; Abrams & Rueffler 2009 : 812-813). Cela limite leur capacité à rendre compte des dynamiques compétitives entre schémas suffixaux. Ce constat justifie le recours à une mesure asymétrique. Dans la compétition morphologique, cette asymétrie du chevauchement est précisément ce qui rend possible l'analyse des rapports de la pression compétitive exercée par un schéma suffixal sur un autre. Ainsi, nous décidons d'utiliser la formule dite de taux de chevauchement (voir § 4.3.4), proposée par Hurlbert (1978) en vue de quantifier leur potentiel de compétition dans le lexique.

Enfin, à travers les recherches examinées, nous constatons également que l'approche multifactorielle apparaît plus pertinente que l'approche monofactorielle. En

effet, plusieurs morphologues (Plag 1999 ; Arndt-Lappe 2014 ; Bonami & Thuilier 2019) soulignent explicitement que la compétition morphologique ne peut être réduite à une propriété linguistique unique. Au contraire, elle résulte souvent d'une interaction complexe entre plusieurs propriétés linguistiques. L'approche monofactorielle, en ignorant cette complexité, se montre ainsi limitée et insuffisante pour identifier efficacement l'ensemble des contraintes linguistiques ou des facteurs sous-jacents à la sélection des suffixes.

Dans une optique multifactorielle, se pose alors la question suivante : parmi l'ensemble des propriétés linguistiques disponibles, quelles sont celles qui exercent une influence décisive dans la formation des noms d'instruments ? Autrement dit, tous les facteurs n'ont pas nécessairement le même poids explicatif, et certains peuvent s'avérer plus discriminants que d'autres. En ce sens, trois propriétés méritent une attention particulière, à savoir la sélection des thèmes (Bonami & Boyé 2003), la sous-catégorie ontologique sémantique (Lin 2021 ; Rainer 2021) et le profil sémantique assumé par la base dans la représentation sémantique du dérivé (voir § 3.1.3). Ces trois propriétés jouent-elles un rôle dans la prédiction du choix des suffixes dans la formation des noms d'instruments ? Il convient de soumettre cette hypothèse à vérification empirique.

En conclusion, cette section a mis en évidence plusieurs limites et points critiques concernant l'application du principe d'exclusion compétitive à la morphologie, la définition et l'usage de la notion de niche morphologique, les outils statistiques et visuels employés pour étudier cette notion, les approches fondées sur des facteurs explicatifs, ainsi que les propriétés linguistiques retenues pour distinguer les suffixes en compétition. Ces constats critiques nous conduisent à formuler les hypothèses suivantes :

- Les contraintes linguistiques interagissent dans la formation des noms d'instruments.
- Ces propriétés présentent des poids différenciés dans la prédiction du choix des suffixes instrumentaux et s'organisent selon une hiérarchie.
- Le principe d'exclusion compétitive (Gause 1934) présente certaines limites dans l'interprétation de la compétition morphologique entre suffixes instrumentaux. Certaines oppositions entre suffixes demeurent irrésolues dans le cadre d'un mode de compétition non absolu.

Ces hypothèses découlent directement des critiques formulées dans cette section.

Elles seront reprises et évaluées dans le chapitre 6. Dans le chapitre suivant (Chapitre 3), nous nous concentrerons sur les cadres théoriques fondamentaux sur lesquels repose cette thèse.

Chapitre 3 Cadres théoriques

Ce chapitre est consacré à la présentation des cadres théoriques mobilisés dans cette thèse. Celle-ci s'inscrivant dans la linguistique cognitive, dans une perspective onomasiologique, et relevant de la morphologie lexématique, nous avons retenu des théories en cohérence avec ces orientations. Dans la section 3.1, nous présenterons d'abord la Morphologie de Construction (Booij 2010), qui constitue le cadre théorique de départ (§3.1.1). Ce modèle offre une base théorique s'inscrivant dans la morphologie lexématique et la linguistique cognitive, en mettant l'accent sur les schémas morphologiques au sein du lexique. Toutefois, bien qu'il fournisse une base théorique pertinente, ce cadre ne permet pas de rendre compte de certains phénomènes centraux pour notre étude, tels que la sélection sémantique de la base, certains cas de dérivation impliquant la métonymie ou la métaphore, ou encore une représentation sémantique des lexèmes moins développée dans ce cadre.

Pour repenser ce cadre, nous nous appuierons sur plusieurs théories linguistiques d'inspiration (§3.1.2), telles que le *Modèle d'enracinement morphologique* (Schmid, 2016), le *Lexique génératif* (Pustejovsky 1996), ainsi que des approches issues de la linguistique cognitive : Lakoff & Johnson (1980) pour la métaphore, Lakoff (1987) pour la métonymie, et Fauconnier & Turner (2002) pour l'intégration conceptuelle. Ces apports permettent de mieux rendre compte des processus cognitifs impliqués dans la formation lexicale. C'est en repensant la Morphologie de Construction à la lumière de ces théories d'inspiration que nous proposons une nouvelle théorie, la Néo-Morphologie Cognitive (§3.1.3). Cette proposition vise à renouveler l'interprétation de la dérivation dans une perspective cognitive, notamment en rendant compte de la sélection cognitive de la base du dérivé, de la projection d'une trajectoire cognitive, ainsi que l'intégration morphosémantique nécessaire à l'activation des schémas morphologiques. Ensuite, dans la section 3.2, nous nous pencherons sur la définition et la classification des noms d'instruments proposées et utilisées dans cette étude. Nous commencerons par définir les noms d'instruments en identifiant les cas prototypiques à l'aide de tests linguistiques. À partir de ces sous-classes ontologiques prototypiques, nous proposerons une classification en soulignant leurs propriétés les plus saillantes et distinctives.

3.1 Fondements théoriques et nouveau modèle proposé

3.1.1 Cadre théorique de départ : *Morphologie de Construction*

3.1.1.1 Notions fondamentales

La Morphologie de Construction (*Construction Morphology*, Booij 2010) est une théorie morphologique développée à partir du cadre de la Grammaire de Construction (*Construction Grammar*, Goldberg 1995), au sein du courant de la linguistique cognitive. Elle repose sur une conception *usage-based* (Bybee 2006) des langues, où la grammaire se construit à partir des expériences d'usage récurrentes, et s'inscrit dans la morphologie lexématique (Aronoff 1994), où le lexème constitue l'unité minimale et centrale de la morphologie flexionnelle comme dérivationnelle. Les opérations morphologiques sont considérées comme des constructions stockées dans le lexique. Cette double orientation, à la fois cognitiviste et lexématique, rend la Morphologie de Construction (ci-après CM) particulièrement compatible avec la perspective que nous développons dans cette thèse. C'est pourquoi nous retenons la CM comme point d'ancrage théorique pour notre travail de recherche

Les notions de construction, schéma morphologique, hiérarchie et représentation lexicale (au sens de Booij) occupent une place centrale dans la CM. Elles structurent la majeure partie de cette théorie et permettent de mettre en lumière ses idées fondamentales ainsi que son fonctionnement interne. Selon ce fil conducteur, nous proposons dans les parties suivantes une présentation détaillée de la CM à travers l'analyse de ces notions centrales.

Nous commençons par la notion de « construction », qui joue un rôle primordial dans les sciences cognitives. Dans le cadre de la Grammaire de Construction (*Construction Grammar*, Goldberg 1995), « les constructions sont des appariements conventionnels forme-sens, qui sont reliés par des liens de correspondance symbolique »[23] (Masini & Audring 2019 : 366). Il convient également de noter que « les constructions peuvent varier en complexité et en spécificité »[24] (Goldberg 2009 : 94, repris par Booij 2010 : 15). Concrètement, les constructions peuvent être observées à différents niveaux linguistiques, allant des unités lexicales minimales telles que VERT,

[23] Le texte original : "constructions are signs, that is, conventionalized pairings of form and meaning."
[24] Le texte original : "constructions can vary in size and complexity."

PORTE, LIVRE dans la morphologie lexématique, jusqu'à des structures syntaxiques plus larges telles que SBJ be-TNS VERB-en by OBL. Cette organisation hiérarchique reflète un continuum syntaxe-lexique (Croft 2001, repris par Booij 2010 : 16). La CM, en se concentrant sur le niveau lexical, considère que les constructions s'organisent principalement autour du lexème. En outre, les constructions lexicales sont conçues comme des faisceaux complexes de propriétés. Comme l'indiquent Masini & Audring (2019 : 366), « le niveau de forme peut être divisé en un niveau phonologique (PHON) et un niveau morphosyntaxique, ce dernier contenant des informations morphologiques (MORPH) et syntaxiques (SYN), tandis que le niveau de sens est supposé inclure des informations sémantiques (SEM), pragmatiques (PRAG) et fonctions du discours (DISC). »[25] Par conséquent, « les éléments morphologiques (comme toutes les autres constructions) sont représentés comme des couches de structure indépendantes mais interconnectées. »[26] (Masini & Audring 2019 ibid.). Nous prenons le cas du lexème PRESENTATEUR comme un exemple pour bien illustrer la notion de « construction lexicale » dans le cadre de la CM.[27]

(1) <[[présenter]$V^{\alpha}{}_{j}$ *eur*]$_{Ni}$ ↔ [AGENT qui PRÉSENTER $_{j}$]$_{i}$ >

Dans l'exemple (1), le pôle de la forme et le pôle du sens sont représentés de part et d'autre de la flèche. Dans le pôle de la forme, sont indiquées les informations concernant la base et le dérivé, notamment par l'exposant α pour les propriétés morphologiques, et par les indices V et N pour les catégories syntaxiques. Quant au pôle du sens, il représente le sens du dérivé, auquel contribue la structure argumentale et la sémantique de la base. Hormis la construction PRESENTATEUR, les locuteurs généralisent à partir de séries de lexèmes comme DANSEUR, CONSTRUCTEUR,

[25] Le texte original : "the FORM level can be divided into a phonological (PHON) and a morphosyntactic tier, the latter containing morphological (MORPHO) and syntactic (SYN) features, whereas the MEANING level is assumed to include semantic (SEM), pragmatic (PRA), and discourse-functional (DISC) information."

[26] Le texte original : "morphological items (like all other constructions) are represented as independent but connected layers of structure."

[27] Nous avons adapté la manière de présenter la construction proposée par Masini & Audring (2019 : 369). Ainsi, dans la construction et le schéma morphologique (présenté après), les exposants sont les suivants : (i) Les exposants en lettres majuscules telles que <A>, <V>, <N>, etc., indiquent les catégories lexicales (A = adjectif ; N = nom ; V = verbe) ; (ii) Les indices en lettres minuscules telles que <i>, <j> et <k> servent à coindexer les variables linguistiques ; (iii) Les exposants en lettres grecques telles que <α>, <β> et <γ> désignent les traits morphologiques ; (iv) Les lettres non indices <x>, <y> et <z> représentent des variables, où : les lettres minuscules (<x, y, z>) désignent du matériel phonologique non spécifié, et les majuscules (<X, Y, Z>) une catégorie lexicale non spécifiée.

TRADUCTEUR et ACHETEUR, de manière à faire émerger le schéma illustré en (2), lequel leur permet de créer de nouveaux lexèmes (Booij 2010 : 2).

(2) $<[[X]_{V}{}^{\alpha}_{j}\ \textit{eur}]_{Ni} \leftrightarrow [\text{AGENT qui PRED}_{j}]_{i}>$

Concernant la notion « schéma morphologique », qui est définie comme « une construction abstraite, partiellement ou totalement, qui contient des variables. Ce type de schéma détermine la structure des lexèmes existants et permet la construction de nouveaux lexèmes »[28] (Masini & Audring 2019 : 366). Selon Booij (2010 : 17, repris par Masini & Audring 2019 : 369), dans le cadre de la CM, les affixes ne possèdent pas de sens autonome : leur contribution sémantique n'est accessible qu'à travers la construction morphologique dans laquelle ils sont intégrés. Cette conception reflète la pensée de la morphologie lexématique, selon laquelle le lien entre forme et sens dépend de la structure d'ensemble de la construction.

C'est précisément dans cette optique que la CM organise les schémas morphologiques selon différents niveaux d'abstraction. En effet, cette structuration hiérarchique permet de saisir comment les constructions concrètes dérivent progressivement de modèles abstraits à travers plusieurs niveaux intermédiaires. Selon Booij (2010 : 25-26), les schémas morphologiques peuvent être structurés en plusieurs niveaux de généralité, allant des constructions plus spécifiques (les constructions lexicales) à celles plus abstraites. Traugott (2008 : 236) introduit pour cela une tripartition fonctionnelle :

- les **macro-constructions**, qui représentent des modèles abstraits et généraux,
- les **méso-constructions**, qui forment des regroupements intermédiaires,
- et les **micro-constructions**, qui désignent des patrons formellement spécifiés, instanciés dans des lexèmes concrets.

Booij (2010 : 51) appelle les constructions de niveau intermédiaire des **sous-schémas**, soulignant leur rôle dans la structuration du système lexical entre les niveaux macro et micro. Comme le souligne Hilpert (2013 : 5), les méso-constructions peuvent elles-mêmes être organisées de façon hiérarchique, en intégrant d'autres méso-

[28] Le texte original : "Constructions can be partly or wholly abstract, which means that they contain variables. Such templates or schema motivate the structure of existing words and enable the coinage of new ones."

constructions moins schématiques. L'ensemble forme un réseau graduel, où chaque niveau reflète un degré différent d'abstraction. Cette structuration reflète la vision basée sur l'usage (Bybee 2006) propre à la CM, dans laquelle les régularités morphologiques émergent de l'usage et se schématisent progressivement à travers les niveaux de construction. La figure 3.1 illustre cette organisation hiérarchique des constructions dérivationnelles. En prenant le suffixe *-eur* comme fil conducteur, nous pouvons expliciter comment ces différents niveaux s'articulent pour construire les lexèmes tels que EPLUCHEUR ou AMPLIFICATEUR.

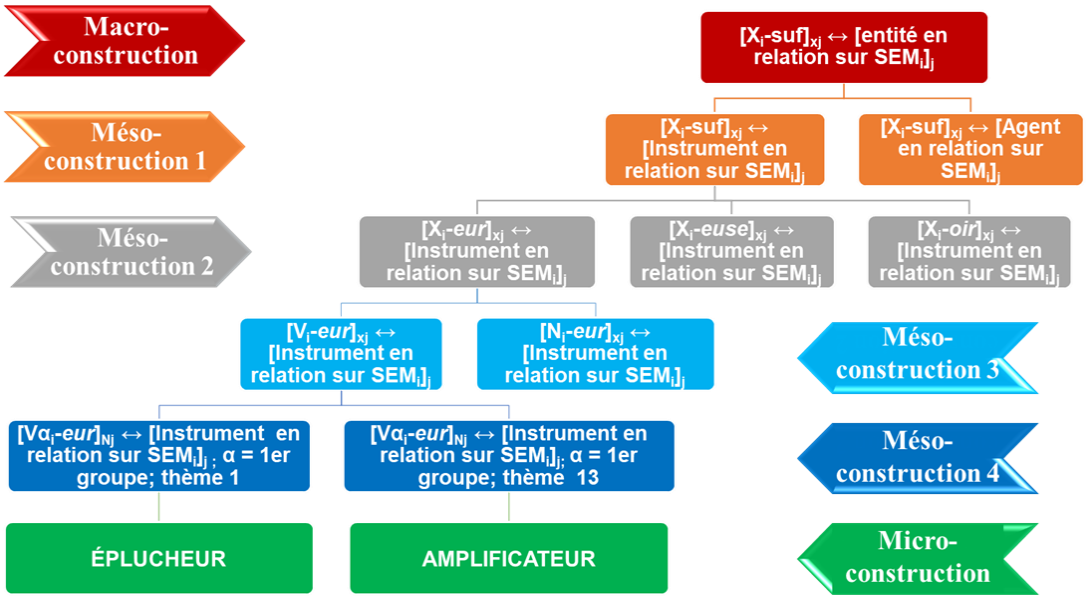

Figure 3.1 Organisation hiérarchique des constructions dérivationnelles, articulée en plusieurs niveaux d'abstraction

Une macro-construction générale associe une forme du type [X-suf] à une entité liée à une sous-classe ontologique ou un rôle sémantique non encore saturé. À partir de cette construction plus abstraite, les méso-constructions remplissent progressivement les variables, d'abord sur le plan sémantique (rôle sémantique ou sous-classe ontologique : instrument / agent), puis sur le plan morphologique (les suffixes). Pour les constructions instrumentales, le suffixe *-eur* est mobilisé à travers deux grands sous-schémas : un schéma déverbal et un schéma dénominal. Le schéma déverbal en *-eur* se subdivise lui-même en plusieurs constructions plus spécifiques, parmi lesquelles deux se révèlent particulièrement productives : l'une fondée sur le thème 1 du verbe de base (*par ex.* EPLUCHEUR), l'autre sur le thème 13 du verbe (*par ex.* AMPLIFICATEUR). Ces lexèmes correspondent au niveau de la micro-construction, où la construction est entièrement instanciée par une base donnée et un suffixe.

Enfin, dans cette organisation hiérarchique, la CM accorde un rôle central à la représentation des schémas, qui permet de rendre compte de la structuration interne des constructions morphologiques. Elle s'inspire notamment de l'architecture parallèle proposée par Jackendoff (2002 : 125), dans laquelle les niveaux linguistiques phonologique, syntaxique et sémantique sont représentés en parallèle et reliés par des correspondances bidirectionnelles, symbolisées par la flèche ↔. La représentation morphologique prend ici tout son sens. Elle formalise les régularités observées et explicite les variables linguistiques mobilisées dans chaque construction, qu'il s'agisse de propriétés phonologiques, morphologiques, syntaxiques et sémantiques. Cette formalisation rend visibles à la fois le fonctionnement propre à chaque niveau et les interactions qui les relient dans le système dérivationnel. C'est dans cette logique que chaque représentation de construction en CM peut être formulée sous forme de composante tripartite, intégrant :

- une composante phonologique (constituée des informations phonologiques),
- une composante morphosyntaxique (souvent spécifiée pour la catégorie syntaxique et les informations morphologiques),
- et une composante sémantique (incluant les variables liées aux sous-classes ontologiques ou aux rôles sémantiques).

Nous prenons à nouveau le lexème PRESENTATEUR afin d'illustrer plus précisément la représentation lexicale proposée par Booij (2010). La Figure 3.2 en présente l'architecture, qui se compose de trois parties principales. Précisément, la première partie, située à gauche de la représentation, correspond aux informations phonologiques. Le symbole « ω » désigne la forme phonologique du lexème, tandis que « σ » indique les syllabes. La forme phonologique /pʀezɑ̃tatœʀ/ comprend quatre syllabes : le radical /pʀezɑ̃tat/ en compte trois et se termine par la consonne /t/, tandis que le suffixe *-eur* forme une syllabe. De plus, la deuxième partie, située au centre de la représentation, concerne la composante morphosyntaxique. L'exposant α renvoie à l'allomorphe de base du verbe PRESENTER, identifié comme le Thème 13, à savoir /pʀezɑ̃tat/. Le V signale que la base appartient à la catégorie syntaxique verbe. Enfin, la troisième partie, située à droite, représente la composante sémantique. Plus précisément, le lexème PRESENTATEUR résulte de l'unification entre la base verbale PRESENTER et le schéma morphologique suffixal *-eur*, ce qui permet l'instanciation d'un nom à interprétation agentive en raison des propriétés événementielles de la base.

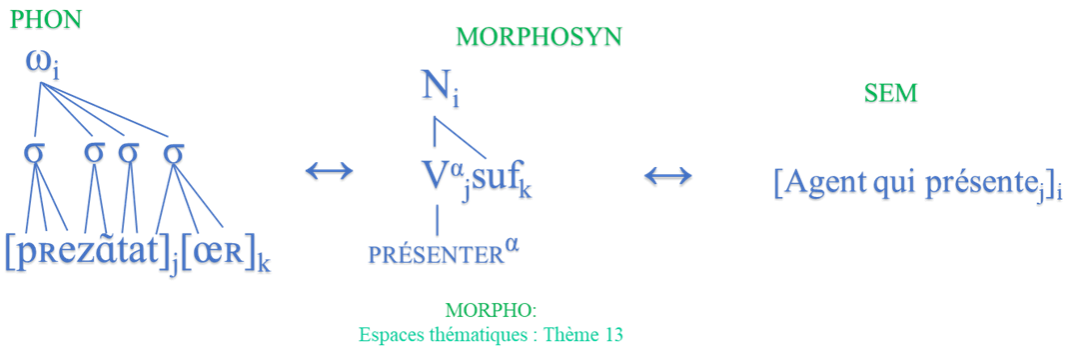

Figure 3.2 Représentation lexicale du lexème PRESENTATEUR

En définitive, la CM proposée par Booij constitue un fondement théorique pour les recherches en morphologie dans une perspective linguistique cognitive. Son cadre théorique a été employé et validé dans de nombreuses études. Par ailleurs, la représentation qu'il propose présente une certaine utilité méthodologique, dans la mesure où elle permet de formaliser plus clairement la construction lexicale ainsi que les schémas morphologiques. Malgré sa valeur descriptive et son utilité méthodologique, la CM présente plusieurs limites qui justifient une reconceptualisation partielle de son cadre théorique. Ces limites peuvent être regroupées en trois volets principaux : la sélection et les propriétés sémantiques des bases, la représentation sémantique, et l'articulation entre base et dérivé.

3.1.1.2 Limites

3.1.1.2.1 Sur la sélection et les propriétés sémantiques des bases

[1] La sélection cognitive de la base

Dans la CM, la base est considérée comme donnée, sans que cette théorie n'explique les mécanismes cognitifs ou sémantiques qui motivent sa sélection. Prenons l'exemple du lexème CAFETIERE, défini comme « récipient permettant de préparer une infusion de café ». Dans ce cadre, CAFE est simplement donné comme base du dérivé, sans que l'on s'interroge sur les raisons pour lesquelles ce nom est mobilisé dans la dérivation alors que d'autres lexèmes en lien avec le référent visé, tels que PREPARER, FABRIQUER ou INFUSER pourraient également être envisagés. L'absence d'explication concernant le processus de sélection de la base masque en réalité une étape essentielle de la dérivation : celle de la construction conceptuelle du référent et de l'identification de la base la plus saillante pour le représenter.

[2] L'hétérogénéité des bases verbales

La *Morphologie de Construction* ne distingue pas les types sémantiques de la base des noms déverbaux. Par exemple, les dérivations SONNER > SONNETTE et RASER > RASOIR reposent sur des verbes qui renvoient à des représentations conceptuelles différentes, mais sont traitées de la même manière dans CM. En effet, le nom SONNETTE est défini comme « petit instrument métallique qui sonne pour avertir », et sa base SONNER renvoie à un moyen de fonctionner, centré sur l'émission sonore en tant que mode opératoire. En revanche, RASOIR est défini comme « instrument à tranchant très fin servant à raser », et sa base RASER met en avant une fonctionnalité instrumentale, focalisée sur le but d'utiliser cet instrument. Ce contraste entre moyen opératoire et fonctionnalité révèle une hétérogénéité sémantique importante parmi les bases verbales, qui n'est ni identifiée ni prise en compte dans la CM.

3.1.1.2.2 Sur la représentation sémantique

[3] Traitement sémantique de la nominalisation

La CM accorde une attention privilégiée aux noms déverbaux sur le plan sémantique. En revanche, les noms dénominaux et désadjectivaux sont traités de manière schématique, sans formalisation précise des interprétations possibles. Le privilège accordé aux noms déverbaux cause ainsi un traitement sémantique asymétrique de la nominalisation dans son ensemble.

À titre d'exemple, dans l'exemple (1), le lexème PRESENTATEUR met en relief un procès événementiel de manière explicite dans sa représentation sémantique. En revanche, PIANISTE est défini comme « personne dont la profession est de jouer du piano ou qui joue du piano en amateur, mais avec un certain talent ». En adaptant le schéma proposé par Booij (2010 : 31), dans la partie sémantique de cette construction (cf. exemple 3), le verbe sous-jacent du PIANISTE est omis et il est remplacé par un lien ontologique entre PIANO et PIANISTE. Autrement dit, il est privé de la possibilité d'être analysé comme les noms d'agent. Cependant, d'un point de vue ontologique, PIANISTE peut être en partie classé dans cette catégorie, puisque le verbe JOUER permet de reconstruire un procès événementiel, où le PIANISTE occupe un rôle sémantique Agent. Cela souligne l'importance du verbe sous-jacent dans la représentation sémantique. Toutefois, la CM choisit généralement de recourir à des étiquettes ontologiques pour décrire les noms dénominaux et désadjectivaux, en omettant la structure événementielle

potentiellement associée. Cette omission entre ainsi en contradiction avec les présupposés sémantiques de la théorie elle-même, car elle cache la possibilité pour ces tels lexèmes d'être interprétés comme des noms d'agent.

(3) $<[[\text{piano}]_{N}{}^{\alpha}{}_{j}\ \text{iste}]_{Ni} \leftrightarrow [\text{Personne avec la compétence pour le } \text{PIANO}_{j}]_{i}>$

[4] Manque explicite en matière de métaphore et de métonymie

La CM ne propose aucun mécanisme pour traiter les opérations métaphoriques ou métonymiques intervenant dans la dérivation, alors même que ces projections cognitives jouent un rôle central dans de nombreuses constructions lexicales. Cette absence devient particulièrement problématique dans les cas où la relation entre base et dérivé n'est pas sémantiquement transparente, comme FENOUIL > FENOUILLETTE (métaphore) et PLUME > PLUMAIL (métonymie). Cette théorie n'explique pas comment une base conceptuellement éloignée peut être activée et interprétée comme la base du dérivé. Ce manque empêche de rendre compte des stratégies de dérivation mobilisées par les locuteurs dans la dérivation.

[5] Absence de traitement des restructurations métonymiques

La CM ne rend pas compte des cas où la base du dérivé est un adjectif, comme CHIMIQUE dans CHIMIQUIER, alors que la véritable base conceptuelle correspond à un syntagme implicite, produit chimique. Autrement dit, il est possible de construire un dérivé sur seulement un lexème dans le cas de base complexe (Corbin 1987). Ce type de dérivation repose sur une restructuration métonymique, par laquelle une entité complexe est réduite à une de ses propriétés plus saillantes. Dans ce cas, CHIMIQUE ne désigne pas une propriété autonome, mais métonymiquement l'ensemble produit CHIMIQUE. Ce glissement, qui sert de point d'ancrage cognitif à la construction de CHIMIQUIER, constitue une stratégie de dérivation que la CM n'aborde ni ne formalise.

[6] La précision et la formalisation de la représentation sémantique

La CM ne propose qu'une représentation fragmentaire du sens, centrée sur la base, sans offrir de description sémantique complète et explicite, que ce soit pour la base ou pour le dérivé. Prenons l'exemple du lexème PLAFONNIER : si l'on adopte la CM pour l'illustrer, sa construction correspond à l'exemple (4).

(4) <[[plafond]$_{N}{}^{\alpha}{}_{j}$ *ier*]$_{Ni}$ ↔ [Instrument fixé au PLAFOND$_j$]$_i$ >

Cependant, lorsqu'on examine la partie sémantique, on constate que de nombreuses informations essentielles restent implicites, en particulier celles qui justifient la catégorisation de PLAFONNIER comme instrument. De plus, la base PLAFOND et la catégorie ontologique visée par le dérivé relèvent en réalité de domaines conceptuels différents : la classe instrument est essentiellement liée au procès (voir la section 1.2), tandis que PLAFOND est issu d'un état statique, plutôt que d'un événement dynamique. Par conséquent, l'analyse fournie par CM cache une quantité importante d'informations implicites, ce qui rend difficile une interprétation fine des motivations sémantiques et des mécanismes cognitifs de construction du dérivé.

3.1.1.2.3 Sur l'articulation entre base et dérivé

[7] Mécanisme formel d'articulation morphosémantique

La CM ne formalise pas de manière explicite l'articulation entre les propriétés sémantiques de la base, les propriétés sémantiques du schéma morphologique et les propriétés sémantiques portées par le dérivé. Bien qu'elle place la base et le dérivé dans une même structure de construction, elle ne précise ni comment les propriétés sémantiques issues de la base et du dérivé sont activées, ni comment émerge une condition morphosémantique préalable à l'entrée dans le système morphologique, en vue de la construction du référent à nommer. Faute d'un espace d'intégration dynamique, l'opération morphosémantique nécessaire à l'activation du système morphologique demeure largement implicite.

En somme, les limites que nous avons identifiées dans la CM peuvent être regroupées en trois catégories principales. Premièrement, cette théorie ne permet pas totalement d'identifier et d'articuler les mécanismes cognitifs qui précèdent l'entrée dans le système morphologique, notamment en ce qui concerne la sélection de la base et sa trajectoire cognitive. Deuxièmement, elle n'aborde pas les cas de dérivation lexicale qui mobilisent des opérations métaphoriques ou métonymiques, alors que ces phénomènes jouent un rôle important dans la morphologie. Enfin, il lui manque un espace d'intégration morphosémantique permettant d'intégrer les informations linguistiques de la base du dérivé et les informations sémantiques du dérivé, et de faire émerger la condition morphosémantique nécessaire à l'activation des schémas morphologiques. En effet, c'est cet espace d'intégration qui fournit les informations

linguistiques nécessaires, lesquelles constituent des variables et des modalités permettant d'influer sur le choix des suffixes. Sans cet espace, nous ne pouvons pas savoir comment ces informations linguistiques finissent par entrer dans le système morphologique.

Dans la section suivante, nous proposerons de repenser la CM en y intégrant plusieurs théories issues de la linguistique cognitive et de la sémantique lexicale, afin de pallier les limites identifiées. Cette démarche servira de base à l'élaboration d'un nouveau cadre théorique, que nous désignons sous le nom de Néo-Morphologie Cognitive (Section 3.1.3).

3.1.2 Théories d'inspiration

Dans la section précédente, nous avons présenté les principes fondamentaux de la Morphologie de Construction et identifié trois grandes catégories de limites, au nombre de sept, que nous chercherons à dépasser en enrichissant la CM par l'intégration de plusieurs théories issues de la linguistique cognitive et de la sémantique. Il s'agit notamment du *Modèle d'enracinement morphologique* (Schmid 2016), du *Lexique génératif* (Pustejovsky 1996), des projections cognitives telles que la métaphore conceptuelle (Lakoff & Johnson 1980) et la métonymie conceptuelle (Lakoff 1987), ainsi que de la théorie de l'intégration conceptuelle (Fauconnier & Turner 2002). Suivant ce fil conducteur, nous présenterons brièvement les apports respectifs de ces cadres théoriques, en montrant comment chacun d'eux contribue à pallier les limites identifiées dans la CM. Ces apports permettent d'ouvrir la voie vers un nouveau cadre théorique, à même de rendre compte des mécanismes cognitifs impliqués dans la dérivation.

3.1.2.1 Modèle d'enracinement morphologique

Nous commençons par la théorie du *Modèle d'enracinement morphologique* (Schmid 2016). Cette théorie applique deux théories issues de la linguistique cognitive, à savoir la grammaire cognitive (*Cognitive Grammar*, Langacker 2008) et la sémantique des cadres (*Frame Semantics*, Fillmore 1968), au domaine de la morphologie. Cette approche vise à interpréter la formation lexicale à partir de scénarios conceptuels, où les rôles sémantiques sont distribués autour d'un verbe dans la représentation sémantique du dérivé. Concrètement, selon Schmid (2016), la construction d'un nouveau lexème part d'un scénario conceptuel du référent. Dans ce

scénario, un verbe — qu'il s'agisse de la base verbale ou d'un verbe sous-jacent — établit une structure argumentale. Au sein de cette structure argumentale, les éléments lexicaux susceptibles d'être participants dans un procès se voient attribuer des rôles sémantiques par ce verbe.

Il en résulte que les dénominaux et les noms désadjectivaux peuvent également être valorisés et traités, au même titre que les noms déverbaux, comme des nominalisations fondées sur des classes de noms. Cette approche repose sur une double orientation : d'une part, elle s'inscrit dans une morphologie morphématique, dans laquelle le morphème est considéré comme porteur direct de sens et comme unité minimale ; d'autre part, elle adopte une perspective onomasiologique, qui part des entités conceptuelles du monde réel pour expliquer leur encodage linguistique (Schmid 2016 : 92).

Dans cette approche, le concept central est le profilage (*profiling*), défini comme une opération cognitive de mise en saillance d'éléments conceptuels. L'auteur distingue trois types de profilage : le profilage conceptuel (*conceptual profiling*), le profilage interne / figure-fond (*internal / figure-ground profiling*), et le profilage par type conceptuel (*concept-type profiling*). Parmi ces trois types, le profilage conceptuel joue un rôle plus important dans cette thèse. Nous en présentons donc plus particulièrement les principes. Selon Schmid (2016), le profilage conceptuel se définit comme : pour les locuteurs, dans la formation des mots, le morphème sert de « profil », tandis que les autres éléments lexicaux du scénario, non sélectionnés dans la construction du lexèmes, constituent le fondement. (Schmid 2016 : 105) [29]. Cette analyse, directement inspirée de la grammaire cognitive, suppose une correspondance étroite entre forme morphémique et saillance conceptuelle : les morphèmes sont perçus comme des indices directs des composantes sémantiques mises en relief. Il convient toutefois de souligner que cette conception s'inscrit dans la morphologie morphématique où le profilage est opéré au niveau du morphème. Cependant, étant donné que notre thèse se positionne dans la morphologie lexématique, il est nécessaire d'adapter la définition de la notion de « profilage conceptuel ». Dans notre thèse, le profilage devrait porter non sur les morphèmes individuellement, mais sur les lexèmes, en tant que porteurs d'identités

[29] Le texte original en anglais : «The idea behind this kind of profiling is that concepts are highlighted by the very fact that they are explicitly encoded by morphemes. The morphemes that constitute a complex lexeme are perceived as profiled figures, while the ground is provided by the mental image or representation of a situation or state of affairs, which the speaker has in mind and wants the hearer to activate. »

conceptuelles autonomes. Cette différence n'est pas simplement formelle : elle traduit une autre manière d'aborder la contribution des lexèmes à la construction du sens, en s'appuyant non sur la segmentation morphémique mais sur des ensembles conceptuels portés par les lexèmes eux-mêmes. À titre d'exemple, pour le lexème CLIMATISEUR, la base est le verbe CLIMATISER. Selon l'approche morphématique proposée par Schmid (2016), après le profilage conceptuel, deux morphèmes sont identifiés comme particulièrement saillants pour la formation du lexème CLIMATISEUR : le nom CLIMAT et le suffixe *-iser*. Les conceptions de ces deux éléments lexicaux, ainsi que les informations linguistiques associées, contribuent à la formation du dérivé. En revanche, selon notre adaptation, c'est le lexème CLIMATISER dans son ensemble qui est considéré comme saillant pour former CLIMATISEUR : l'ensemble de la conception et les informations linguistiques portés par le verbe CLIMATISER participe à la dérivation, tandis que le suffixe *-iser* n'est qu'un schéma morphologique dépourvu de sens et de conception autonome.

Nous prenons le cas du lexème THEIERE pour illustrer de manière précise l'adaptation de Schmid (2016). Dans ce lexème, la base THE[30] désigne le contenu infusé (Patient dans le scénario d'infusion). Ce profil sémantique repose sur l'interprétation du lexème dans une scène construite autour du verbe INFUSER, ce qui implique une relation argumentale entre le contenu (THE) et le récipient (THEIERE), que seule la morphologie lexématique permet de représenter de manière cohérente.

À travers cette présentation, cette perspective permet de pallier, au moins partiellement, deux limites de la CM. D'une part, concernant la limite n°1, Schmid (2016) fournit un mécanisme cognitif explicite pour la sélection de la base en fonction de la saillance. D'autre part, quant à la limite n° 3, Schmid (2016) ouvre la possibilité d'envisager les bases non verbales, notamment nominales, comme appartenant elles aussi à des classes de noms, là où seules les bases déverbales étaient traditionnellement prises en compte.

Certes, cette perspective permet de répondre à certaines limites de la CM (Booij 2010), notamment en valorisant le morphème saillant durant le profilage conceptuel pour la sélection de la base, et en élargissant la prise en compte aux bases non-verbales. Toutefois, il existe les autres cinq limites qui demeurent non résolues.

[30] Il convient de préciser que le lexème THE présente une polysémie, puisqu'il peut désigner aussi bien les feuilles de thé que la boisson obtenue à partir de celles-ci. Dans le cadre de notre analyse, nous retenons le premier sens, à savoir « les feuilles de thé », afin d'illustrer notre propos.

Par exemple, d'une part, Schmid (2016) suppose que tout scénario conceptuel repose sur un verbe donné, mais il n'explicite pas quel verbe est identifié dans la représentation du dérivé pour construire ce scénario (limite n°2). Si, dans le cas de THEIERE, le verbe INFUSER s'impose assez naturellement (dans un scénario de fonctionnalité des instruments), pourquoi ne pas envisager d'autres verbes tels que PLACER ou VERSER ? Schmid (2016) ne précise toutefois ni les critères de sélection de ces verbes, ni leurs fonctions différentielles dans la représentation sémantique du dérivé.

D'autre part, la dérivation qui ne s'appuie pas sur des rôles sémantiques explicites, notamment dans les cas de projections cognitives telles que la métonymie ou la métaphore (limite n°4), n'est pas théorisée. Dans le cas du lexème FENOUILLETTE, qui désigne « petite pomme grise dont le parfum rappelle celui du fenouil » (Le Grand Robert, 2005), la base FENOUIL est mobilisée car le parfum de cette variété de pomme est similaire à celui du FENOUIL. Ici, le locuteur utilise la qualité du référent (le parfum) pour nommer l'objet ; c'est principalement une métaphore conceptuelle qui opère, en reliant deux champs sémantiques distincts entre la base et le dérivé. Étant donné que Schmid (2016) adopte les rôles sémantiques proposés par Fillmore (1968 ; 1982) pour interpréter la formation lexicale (*par ex.* Agent, Patient, Thème, Instrument, Lieu, Destinataire, Temps, etc.), une telle relation, qui ne relève pas des rôles sémantiques activés dans la structure argumentale des verbes (sous-jacents), n'est donc pas expliquée dans le cadre proposé par Schmid (2016).

En somme, malgré ses apports importants, le *Modèle d'enracinement morphologique* ne répond que partiellement aux limites de la CM. Les autres limites restent non résolues et nécessitent l'apport d'autres théories pour être pleinement prises en charge.

3.1.2.2 Lexique génératif

Le deuxième cadre théorique que nous introduisons est celui du *Lexique génératif* (Pustejovsky 1996). Il s'inscrit dans le cadre de la linguistique générative. Booij & Masini (2013) ont évoqué la possible compatibilité du *Lexique génératif* avec le cadre de la Morphologie de Construction, sans toutefois en proposer une intégration effective ou un développement systématique. Après analyse, nous souscrivons à cette hypothèse de compatibilité, pour deux raisons principales :

[1] Les deux approches opèrent dans le lexique et envisagent les unités

morphologiques comme des structures organisées et productives au sein de ce composant lexical ;

[2] La structure de qualia du *Lexique génératif* fournit une représentation sémantique riche et formalisée, particulièrement apte à renforcer la composante sémantique encore peu développée dans la CM.

Dans le cadre du *Lexique génératif* (Pustejovsky 1996), la représentation lexicale d'un mot repose sur quatre structures sémantiques principales : la structure d'argument[31], la structure événementielle, la structure de qualia et la structure d'héritage. Étant donné que notre modèle mobilise principalement la structure d'argument et la structure de qualia, nous nous limiterons ici à la présentation détaillée de ces deux structures.

Tout d'abord, selon Pustejovsky (1996, repris par Wang & Wu 2020 : 142), la structure d'argument est une structure positionnelle, définissant le nombre d'arguments qu'un prédicat exige ainsi que leur réalisation syntaxique). Pustejovsky (1996 : 63-64) propose quatre grandes catégories d'arguments intervenant dans la représentation lexicale des prédicats.

1) Les arguments explicites ("true arguments") sont les arguments réalisés syntaxiquement, directement requis par le prédicat dans sa structure argumentale, par exemple dans *Jean est arrivé à 10 heures*, où Jean constitue un argument syntaxique obligatoire.
2) Les arguments implicites ("default arguments") correspondent à des arguments attendus dans certaines structures sémantiques, notamment la structure de qualia, mais qui ne sont pas nécessairement exprimés dans la syntaxe de surface. À titre d'exemple, dans *Jean a épluché les pommes avec*

[31] Afin d'éviter toute ambiguïté terminologique, nous adopterons dans ce travail une distinction claire entre deux notions proches mais théoriquement distinctes :
– La structure d'argument, au sens du *Lexique génératif* (Pustejovsky 1995), désigne un module formel interne de la représentation sémantique du lexème, défini aux côtés de la structure événementielle, de la structure de qualia et de la structure d'héritage. Elle spécifie les positions argumentales attendues au niveau syntaxique pour un prédicat donné, mais ne détaille pas nécessairement les rôles sémantiques associés.
– La structure argumentale, dans son acception traditionnelle en syntaxe et sémantique, correspond à l'ensemble des relations qu'un prédicat entretient avec ses arguments en termes de nombre, de fonction syntaxique et de rôle sémantique (au sens des frames sémantiques).
Dans ce manuscrit, sauf indication contraire, toute mention d'une « structure argumentale » associée à un lexème renvoie à cette deuxième acception. Elle inclut la projection des rôles sémantiques nécessaires à la construction du référent ou de l'événement représenté, conformément à l'approche adoptée par Schmid (2016) ou Fillmore (1968).

un éplucheur, l'instrument est sémantiquement impliqué par le verbe EPLUCHER, bien que son expression syntaxique ne soit pas indispensable.

3) Les arguments latents ("shadow arguments") sont intégrés de manière non visible dans la sémantique du mot. Ils ne sont accessibles qu'au travers d'opérations interprétatives, sans être réalisables syntaxiquement de façon directe. Par exemple, dans *Louis a frappé le ballon avec son pied malade /* avec son pied*, *le pied* est incorporé dans le verbe FRAPPER, sans être syntaxiquement requis.
4) Les ajouts situationnels ("true adjuncts") désignent des modificateurs contextuels (temporels, spatiaux, etc.) qui enrichissent l'interprétation de l'événement, sans être liés à la structure lexicale du prédicat. À titre d'illustration, *Marie a fait le trajet jusqu'à Lyon un mercredi* contient une indication temporelle qui structure le contexte de l'action sans fonctionner comme un argument du verbe.

Ainsi, la structure d'argument du lexème α est représentée dans la figure 3. Cette unité comporte un certain nombre de structures argumentales de types distincts.

$$\begin{bmatrix} \alpha \\ \text{ARGSTR} = \begin{bmatrix} \text{ARG}_1 = \ldots \\ \text{ARG}_2 = \ldots \\ \text{D-ARG}_1 = \ldots \\ \text{S-ARG}_1 = \ldots \end{bmatrix} \\ \ldots \end{bmatrix}$$

Figure 3.3 Structure d'argument du lexème α (Pustejovsky 1996 : 67)

Ensuite, nous présentons la structure de qualia. Inspirée de la tradition aristotélicienne, cette structure proposée par Pustejovsky (1996) repose sur l'idée que les unités lexicales encodent des connaissances encyclopédiques qui participent à la construction du sens. L'auteur met en évidence l'importance de ces connaissances dans les interactions entre langage et objets du monde, ce qui rejoint certaines hypothèses fondamentales de la linguistique cognitive en sémantique lexicale. La structure de qualia se compose de quatre rôles fondamentaux, définis comme suit :

- Rôle constitutif : désigne la relation entre un objet et ses composantes, telles que sa matière, son poids ou ses différentes parties.

- Rôle formel : correspond à ce qui distingue l'objet au sein d'un domaine plus large, notamment par son orientation, sa taille, sa forme, ses dimensions, sa couleur ou sa position ; il s'agit souvent de son type.
- Rôle télique : renvoie à la fonction ou au but de l'objet.
- Rôle agentif : concerne les facteurs impliqués dans la création de l'objet, comme ses créateurs ou la chaîne causale ayant mené à son existence.

Prenons le lexème SONNETTE afin d'illustrer concrètement cette théorie. Les informations correspondant aux quatre rôles de la structure de qualia pour ce lexème sont les suivantes :

- Rôle constitutif : bouton, mécanisme sonore, boîtier…
- Rôle formel : la fonction essentielle est de **sonner** lorsqu'on l'actionne.
- Rôle télique : une sonnette sert à **avertir**.
- Rôle agentif : Le fabricant **fabrique** la sonnette.

Selon cet exemple, il apparaît que les locuteurs choisissent le verbe correspondant au rôle formel dans la structure de qualia comme la base morphologique pour former le dérivé servant à dénommer l'instrument destiné à avertir. L'identification des quatre rôles de la structure de qualia permet ainsi de préciser les types de verbes impliqués dans les constructions morphologiques, en fonction de leurs propriétés sémantiques et de leur articulation avec les connaissances encyclopédiques. Cela explique pourquoi SONNETTE est construit à partir du verbe SONNER, associé au rôle formel, plutôt qu'à partir du verbe AVERTIR, lié au rôle télique.

Après la combinaison du *Lexique génératif* (Pustejovsky 1996) et du *Modèle d'enracinement morphologique* (Schmid 2016) permettent de répondre à plusieurs limites de la CM (Booij 2010), notamment en proposant un cadre pour la sélection cognitive des bases (limite n°1), en élargissant la prise en compte des bases non verbales (limite n°3), en clarifiant les divergences conceptuelles entre plusieurs verbes candidats mobilisables dans la dérivation d'un même lexème au sein de sa représentation sémantique (limite n°2), et en approfondissant la formalisation des représentations sémantiques associées aux lexèmes (limite n°6).

Toutefois, certaines limites demeurent irrésolues. Tout d'abord, aucune de ces deux théories ne prend en compte les mécanismes de projection cognitive, tels que la métonymie ou la métaphore (limite n°4). De plus, ni Schmid (2016), ni le *Lexique*

génératif (Pustejovsky 1996) ne proposent de cadre permettant d'expliquer comment une base complexe comme *produit chimique* peut être réduite à CHIMIQUE dans la dérivation (limite n°5). Enfin, l'intégration des informations linguistiques portées par la base et les informations sémantiques par le dérivé, en vue de constituer une condition morphosémantique préalable, n'est pas abordée (limite n°7). Ces questions restent donc ouvertes et feront l'objet d'une proposition spécifique dans les sous-sections suivantes.

3.1.2.3 Projections cognitives

Comme évoqué dans les sous-sections précédentes, dans certains cas, la formation lexicale repose sur une propriété saillante qui ne peut être interprétée par des rôles sémantiques issus de la structure argumentale. Cela signifie que hormis les rôles sémantiques dans un procès événementiel participant à la dérivation, les projections cognitives y contribuent également. Ainsi, ce type de motivation nécessite l'introduction de deux mécanismes cognitifs spécifiques : la métaphore conceptuelle (Lakoff & Johnson 1980) et la métonymie conceptuelle (Lakoff 1987). Ces deux opérations, fondées sur des mécanismes cognitifs universels, interviennent directement dans la formation de nombreux lexèmes tout en s'intégrant sans contradiction aux cadres théoriques précédemment introduits.

Tout d'abord, la métaphore conceptuelle (Lakoff & Johnson 1980), correspond à un mécanisme cognitif qui établit un lien entre un concept source et un concept cible. Selon Ungerer & Schmid (2006 : 118 ; 127), il existe deux principaux types de correspondances (*mappings*) :

1) le *lean mapping*, entre deux concepts concrets spécifiques, sert à faire ressortir certains aspects particuliers du concept cible (*par ex.* l'attribution de la saleté à John dans « John est un cochon ») ;
2) le *rich mapping*, entre un concept source concret et un concept cible abstrait, voire un concept générique, sert à structurer des notions abstraites à l'aide d'images concrètes (*par ex. les idées peuvent être vendues (comme des objets)*).

Dans ce cadre, la formation du lexème FENOUILLETTE illustre un *lean mapping* : le parfum du FENOUIL (concept source) est transféré à une *petite pomme grise* (concept cible), à travers une analogie olfactive saillante. Cette projection permet de construire un nouveau lexème par activation métaphorique d'une propriété sensorielle.

Par ailleurs, concernant la métonymie conceptuelle (Lakoff 1987), elle intervient

également dans la formation de certains lexèmes. Contrairement à la métaphore, qui relie deux concepts appartenant à des cadres distincts, la métonymie repose sur une relation interne à un même domaine conceptuel : elle permet d'accéder à une entité à travers une autre, établie dans le même champ sémantique. Radden & Kövecses (1999, repris par Littlemore 2015 : 20) ont proposé une classification des métonymies selon six modèles cognitifs idéalisés (*Idealized Cognitive Models*, ou MCI), qui couvrent la majorité des cas représentés en sémantique cognitive :

1) Entités physiques : une partie d'une entité peut représenter le tout ou inversement ;
2) Échelles : l'extrémité d'une échelle peut servir à désigner l'ensemble ;
3) Constitution : un matériau peut désigner l'objet fabriqué à partir de ce matériau ;
4) Événements : une partie d'un événement peut représenter l'événement entier ;
5) Appartenance à une catégorie : un membre peut désigner la catégorie ;
6) Propriétés des catégories : une propriété saillante peut désigner la catégorie entière.

Par exemple, dans le cas de CHIMIQUIER, dérivé de CHIMIQUE, la motivation repose non pas sur une analogie inter-domaine, mais sur une opération de métonymie : une propriété catégorielle (le caractère chimique d'un produit) est mobilisée pour désigner un ensemble d'objets concrets (les produits chimiques transportés). Il s'agit ici d'une projection intra-domaine où une propriété saillante permet d'identifier la totalité de la catégorie, conformément au MCI métonymique « propriété pour catégorie ».

A partir de ces deux cas évoqués, ces projections cognitives, à savoir métaphore et métonymie, permettent ainsi d'expliquer la motivation cognitive dans la construction des lexèmes non prises en compte dans les cadres précédents, notamment lorsque la saillance ne repose ni sur un scénario verbal explicite, ni sur des rôles sémantiques dans une structure argumentale. Cela permet de résoudre les limites (n°4 et n°5) identifiées dans le cadre de la CM. Ainsi, en intégrant les trois théories évoquées jusqu'ici, nous constatons que six des sept limites formulées à l'encontre de la CM ont été levées. Il ne reste désormais qu'un seul point à traiter : l'intégration des informations linguistiques portées par la base avec les informations sémantiques du mot dérivé. Il semble donc nécessaire d'introduire une dernière théorie afin de combler cette limite finale.

3.1.2.4 Intégration conceptuelle

Comme évoqué précédemment, le dernier point à traiter parmi les limites identifiées dans la CM concerne l'articulation entre les informations linguistiques portées par la base et les informations sémantiques du dérivé. Face à cette limite, il nous paraît nécessaire d'introduire une théorie issue de la linguistique cognitive, capable d'assurer l'intégration et de faire émerger de nouvelles structures conceptuelles ou informationnelles au niveau cognitif. Deux théories issues de la linguistique cognitive répondent à ces critères : la métaphore conceptuelle (Lakoff & Johnson, 1980) et l'intégration conceptuelle (Fauconnier & Turner, 2002).

Cependant, comme le souligne Gréa (2002 : 109), par rapport à la métaphore conceptuelle (Lakoff & Johnson, 1980), l'intégration conceptuelle dépasse les limites de la métaphore conceptuelle en mettant « en œuvre un mécanisme d'intégration qui se fonde sur une approche beaucoup plus dynamique : régulation de contraintes multiples et propagation de l'information dans un réseau de relations, où l'état final du calcul correspond à la stabilisation du réseau sur une configuration qui n'épuise pas toutes les possibilités. » En raison de l'absence d'un espace d'intégration dans la théorie de la métaphore conceptuelle, celle-ci ne peut pas intégrer les informations morphosémantiques indispensables à la dérivation. Par conséquent, un cadre de type intégration conceptuelle (*blending*) s'avère nécessaire pour opérer cette mise en commun. En revanche, l'intégration conceptuelle propose un cadre cognitif dynamique, à entrées multiples et extensible, particulièrement apte à intégrer les informations linguistiques portées par la base et les informations sémantiques du dérivé. Elle se révèle ainsi plus puissante et explicative que le modèle de la métaphore. De plus, cette théorie repose sur la logique des espaces mentaux et s'appuie sur un mécanisme cognitif universel, ce qui garantit sa compatibilité avec les autres théories précédentes. Nous adoptons donc la théorie de l'intégration conceptuelle de Fauconnier & Turner (2002) comme cadre théorique pour intégrer les informations linguistiques portées par la base et les informations sémantiques du dérivé.

L'intégration conceptuelle (*Blending theory*, Fauconnier & Turner 2002) est une théorie développée à partir de la théorie des espaces mentaux (Fauconnier 1994). La partie la plus centrale de la théorie de l'intégration conceptuelle est le réseau d'intégration conceptuelle ("*Conceptual Blending Network*"). Il est constitué de quatre espaces mentaux : deux espaces d'entrée ("*Input Spaces*"), un espace générique ("*Generic Space*") et un espace intégrant ("*Blended Space*"), comme illustré dans la Figure 3.4 (Fauconnier & Turner 2002 : 46).

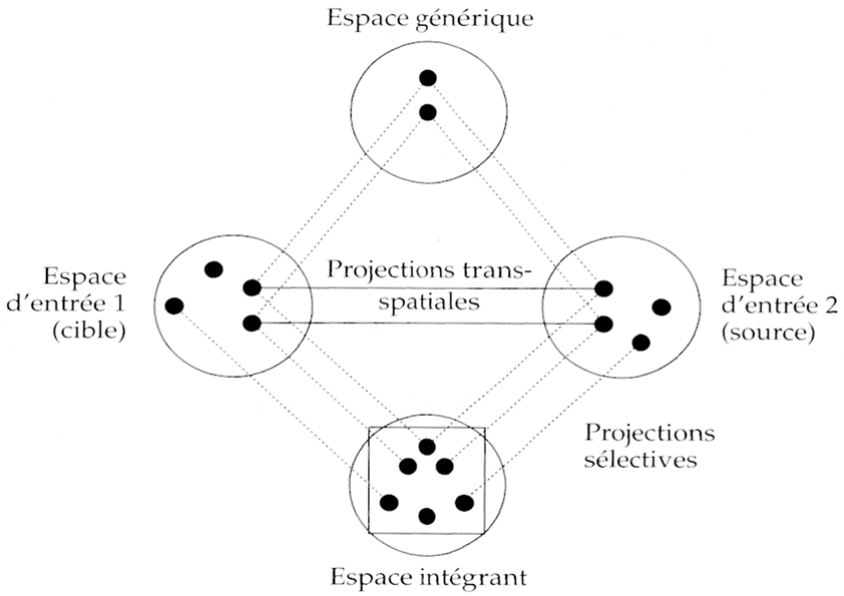

Figure 3.4 Réseau de l'intégration conceptuelle (Fauconnier & Turner 2002 : 46, repris par Gréa 2002 : 110)

Comme illustré dans la Figure 3.4, l'intégration conceptuelle repose sur l'interaction de quatre types d'espaces mentaux : deux espaces d'entrée, un espace générique et un espace intégrant. Selon Fauconnier & Turner (2002 : 47, repris par Liu & Li 2019 : 143), l'intégration conceptuelle repose sur un dispositif dynamique dans lequel plusieurs espaces mentaux interagissent. Les espaces d'entrée sont reliés par des correspondances conceptuelles croisées, permettant l'alignement des éléments pertinents à intégrer. Un espace générique extrait les informations communes qui fournissent la structure conceptuelle nécessaire à l'intégration. L'espace intégrant, quant à lui, opère une intégration sélective des éléments issus des espaces d'entrée, ce qui permet de faire émerger une configuration conceptuelle inédite, appelée structure émergente, qui ne préexiste dans aucun des espaces pris isolément. L'ensemble du réseau est soutenu par des liens implicites assurant la cohérence entre les informations issues des différents espaces. Il convient de noter que l'intégration conceptuelle repose sur une projection sélective : seules les informations jugées pertinentes dans les espaces d'entrée sont effectivement projetées dans l'espace intégrant. C'est précisément cette dynamique d'intégration qui ouvre une voie pertinente pour intégrer les informations linguistiques portées par la base et les informations sémantiques du dérivé, en faisant émerger la condition morphosémantique préalable nécessaire à l'activation des schémas morphologiques dans le système morphologique (pour le détail, voir le chapitre 7).

Ainsi, l'introduction de ces quatre théories linguistiques nous a permis d'enrichir la Morphologie de Construction en apportant des mécanismes explicatifs permettant de dépasser les sept limites identifiées. Parallèlement, l'ensemble de ces apports, provenant à la fois de la mise en perspective de la CM et de l'intégration des quatre théories employées, constitue désormais le socle théorique de notre propre modèle, la Néo-Morphologie Cognitive, dont la sous-section suivante présentera les principes et les articulations internes.

3.1.3 Propositions pour une Néo-Morphologie Cognitive

À travers les deux sous-sections précédentes, nous avons présenté le cadre théorique de la Construction Morphology (CM) et identifié ses principales limites. Face à ces limites et à la lumière des apports complémentaires des théories présentées ci-dessus, il devient nécessaire de dépasser le cadre initial et de proposer une approche plus adaptée aux objectifs de cette thèse. C'est dans cette perspective que la Néo-Morphologie Cognitive (ci-après NMC) est proposée dans le cadre de cette recherche doctorale : une proposition théorique, qui ne vise pas simplement à enrichir ou corriger la CM, mais à repenser l'analyse morphologique sur des principes cognitifs, orientés vers une interprétation fine et évolutive des mécanismes de dérivation. Ce cadre a vocation à être progressivement consolidé et affiné dans des travaux futurs. Dans cette sous-section, nous allons seulement présenter les grandes lignes et les notions fondamentales de la NMC. Pour le détail et les représentations lexicales de cette théorie, nous les présenterons dans le chapitre 7.

3.1.3.1 Positionnement épistémologique et originalité du modèle

En fonction de la présentation des théories mobilisées et des adaptations qui leur sont apportées, se dessine progressivement le positionnement épistémologique du cadre proposé dans cette thèse. Tout d'abord, à l'instar de la Morphologie de Construction (CM) et des théories de la linguistique cognitive telles que la métaphore conceptuelle (Lakoff & Johnson 1980), la métonymie conceptuelle (Lakoff 1987) et l'intégration conceptuelle (Fauconnier & Turner 2002), le modèle de la NMC s'inscrit dans le courant cognitiviste. Ensuite, comme la CM, la NMC adopte une perspective lexématique (Aronoff 1994), en ce sens que l'unité opératoire minimale y est le lexème, et que la relation entre la base et le dérivé constitue un enjeu central de l'analyse. Enfin, dans le prolongement des approches cognitives appliquées à la morphologie (Schmid

2016) ainsi que du *Lexique génératif* (Pustejovsky 1996), la NMC adopte une perspective onomasiologique : à rebours de la démarche sémasiologique, elle ne part pas de la forme linguistique, mais des éléments extralinguistiques du monde réel ou plutôt des concepts cognitifs (Schmid 2016 : 92). Les connaissances encyclopédiques y occupent un rôle fondamental dans la construction lexicale. Une question onomasiologique type pourrait être formulée comme suit : *quels schémas morphologiques sont mobilisés pour former les noms d'instruments ?* (Schmid 2016 : 92).

Ainsi, les idées originales de la NMC peuvent se résumer comme suit :

- La dérivation est envisagée comme un processus de construction lexicale motivé par des mécanismes cognitifs ;
- Le choix de la base n'est pas arbitraire, mais guidé par un profil sémantique et un processus de profilage conceptuel ;
- Le lexique est conçu non comme un simple répertoire formel, mais comme un réseau structuré d'unités cognitives ;
- Fondée sur l'unité de la construction, la démarche adoptée est à la fois onomasiologique et lexématique : elle part de l'objectif conceptuel à construire, et non de la forme linguistique donnée.

En ce qui concerne l'originalité du modèle, il se distingue nettement des approches existantes en morphologie dérivationnelle par une architecture modulaire intégrant explicitement les mécanismes cognitifs opératoires : profilage conceptuel, base morphologique, trajectoire cognitive, espaces d'intégration morphosémantique et système morphologique. Comparativement, la *Morphologie de Construction* (CM, Booij 2010) décrit des réseaux de constructions mais ne formalise ni la sélection cognitive des bases, ni leur projection sémantique. Parallèlement, la *Distributed Morphology* (DM, Halle & Marantz 1993) se concentre sur l'insertion syntaxique des formes, excluant toute structuration cognitive préalable à la dérivation. Le *Lexical Semantic Framework* (LSF, Lieber 2016) offre une décomposition sémantique lexicale riche, mais ne prévoit ni mécanisme procédural de sélection cognitive des bases, ni parcours sémantique explicite reliant base et suffixe. Quant à la *Relational Morphology* (RM, Audring & Jackendoff 2025), elle repose sur des liens lexicaux et sémantiques stockés, sans proposer de processus actif de projection ou d'interface dynamique morphosémantique. Enfin, les approches dites cognitives, comme celle de Schmid

(*Modèle d'enracinement morphologique*, 2016) et de Hamawand (*Cognitive Morphology*, 2011), restent descriptives et ne proposent aucune formalisation systématique des interfaces cognitives entre sélection des bases, projection sémantique et choix des suffixes. En somme, la Néo-Morphologie Cognitive se distingue par une formalisation explicite, modulaire et procédurale des mécanismes cognitifs impliqués dans la dérivation lexicale.

3.1.3.2 Fonctionnement global et Notions fondamentales

La Figure 3.5 présente l'organisation modulaire du modèle de Néo-Morphologie Cognitive, structuré en six étapes qui rendent compte du parcours complet de la dérivation lexicale.

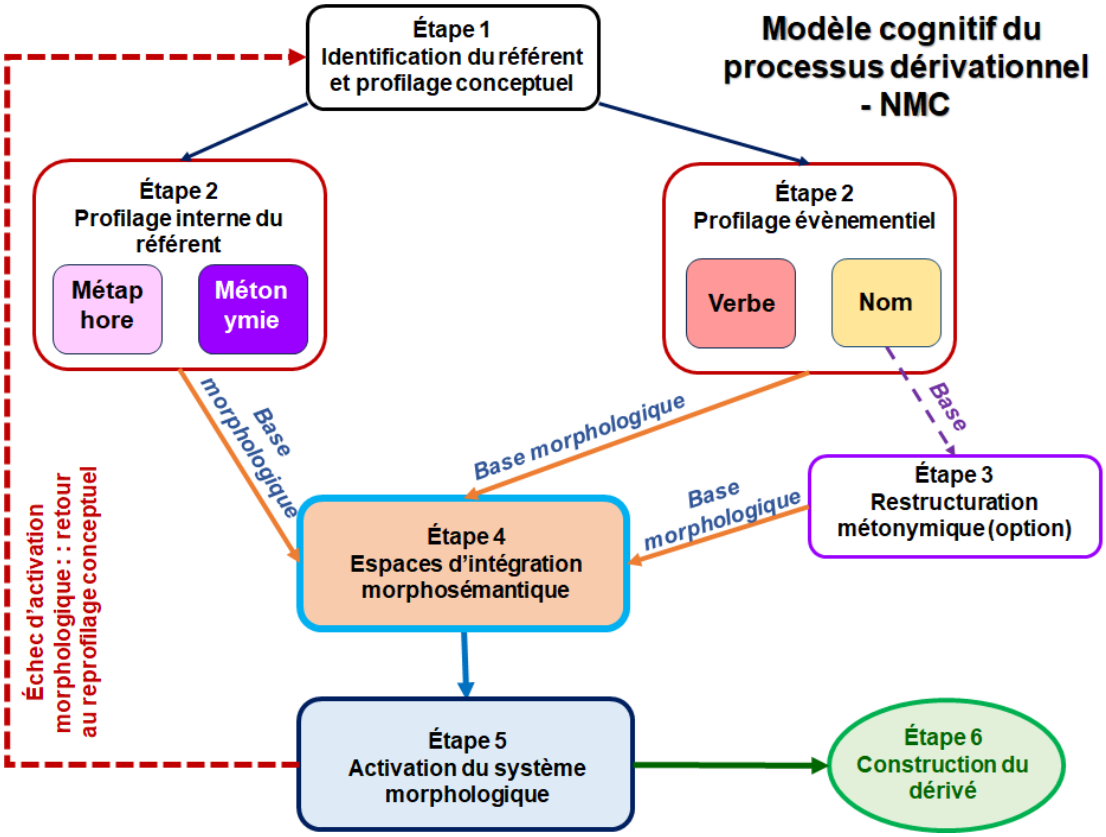

Figure 3.5 Schéma du modèle cognitif du processus dérivationnel de NMC

Le processus commence par l'identification d'un besoin de nomination, suivi d'un profilage conceptuel qui extrait, à partir de la représentation sémantique du dérivé, une base potentielle. Dans ce module, le résultat du profilage conceptuel est le profil sémantique assumé par la base dans la représentation du dérivé. Cette notion est proposée dans ce travail et s'inspire des travaux de Schmid (2016) et de Pustejovsky

(1996).

Précisément, le profil sémantique assumé par la base dans la représentation du dérivé désigne le ou les lexème(s) saillant(s) (en provenance du Verbe de qualia, du rôle sémantique ou de la projection cognitive), identifié(s) dans la structure de qualia du référent à dénommer et servant de critère de sélection pour la base lors du profilage conceptuel. Ces lexèmes saillants constituent la base potentielle, c'est-à-dire la base sélectionnée à l'issue du profilage conceptuel pour servir d'entrée dans le processus dérivationnel. Certaines bases potentielles peuvent devenir directement des bases morphologiques et être intégrées telles quelles dans le système morphologique, tandis que d'autres nécessitent une restructuration métonymique préalable lorsqu'elles dérivent de référents dont la conception est complexe.

Concernant le deuxième module, il s'agit de la restructuration métonymique. Celle-ci vise à dégager une base morphologique pour la dérivation à partir de bases potentielles qui dérivent de référents présentant une conception complexe. À l'issue de ces deux modules, émerge la base morphologique. Celle-ci constitue alors la base intégrée dans le système morphologique et servant finalement à construire les mots. Afin d'accéder au système morphologique, la base morphologique doit être dotée d'une trajectoire cognitive (flèche en orange dans la Figure 3.5). La trajectoire cognitive désigne la voie cognitive projetée par une base morphologique, véhiculant à la fois les informations linguistiques propres à cette base et les informations sémantiques du référent à dénommer.

Ces informations véhiculées par les trajectoires cognitives sont ensuite projetées dans l'espace d'intégration morphosémantique, qui constitue une véritable interface entre la sémantique et la morphologie. Dans cet espace, les informations linguistiques issues de la base et celles issues du dérivé sont intégrées, permettant ainsi l'émergence d'une condition morphosémantique préalable à l'activation des schémas morphologiques au sein du système morphologique. Enfin, lorsque les informations de la base et celles du dérivé accèdent au système morphologique, celui-ci active un mécanisme de compétition morphologique fondé sur quatre facteurs linguistiques : contraintes linguistiques (phonologiques, morphologiques, syntaxiques, sémantiques), effets de blocage, productivité morphologique et facteur socioculturel. La sélection finale des suffixes s'effectue en fonction du poids différentiel de ces facteurs selon la classe lexicale concernée. Une fois la compétition morphologique achevée, la base morphologique remplit le schéma morphologique sélectionné afin de construire le

dérivé. En cas d'échec, le système prévoit un retour vers le niveau conceptuel, autorisant ainsi un reprofilage conceptuel.

3.2 Définition et classification proposées des noms d'instruments

Dans cette section, nous nous concentrerons sur la définition et la classification des noms d'instruments prototypiques. Nous nous inspirerons des travaux en linguistique cognitive, notamment de la théorie des prototypes formulée par Rosch (1975) et Lakoff (1987), selon laquelle un prototype « correspond à une représentation mentale relativement abstraite, partageant les attributs clés du meilleur exemplaire d'une catégorie donnée. » (repris par Mei 2017). La section 3.2.1 sera consacrée à l'identification des caractéristiques prototypiques de ces noms dans une perspective ontologique, à partir des définitions existantes, afin de procéder à un premier tri des sous-classes ontologiques. Dans la section 3.2.2, nous adopterons une perspective linguistique, en mobilisant les tests proposés par Namer & Villoing (2008), pour affiner cette sélection. Sur la base des définitions analysées et des deux tris effectués, nous proposerons une nouvelle définition des noms d'instruments dans une perspective, que nous qualifierons d'"intégrative", intégrant l'approche ontologique et l'approche linguistique. Enfin, dans la section 3.2.3, nous classons les sous-classes ontologiques jugées prototypiques selon leurs traits distinctifs, dans le but de fournir une base rigoureuse au travail d'annotation mené dans cette thèse.

3.2.1 Dans la perspective ontologique

Nous commencerons par présenter les définitions existantes des noms d'instruments prototypiques dans une perspective ontologique, ainsi que les sous-classes ontologiques correspondantes. Nous définirons ensuite ces sous-classes et justifierons la sélection des références mobilisées. Enfin, nous procéderons à un premier tri des noms d'instruments prototypiques selon les caractéristiques prototypiques dans cette perspective.

3.2.1.1 Les définitions existantes et sous-classes ontologiques

Du point de vue ontologique, les noms d'instrument sont considérés comme « une sous-catégorie d'artefact caractérisée par la description d'une certaine forme de fonctionnalité » (Huyghe & Tribout 2015 : 106). Faber (2012 : 125, 162) définit

l'instrument comme un objet utilisé pour réaliser une action, délimité par des propriétés liées à la manipulation, à la fonction et au résultat. Wu & Yuan (2019 : 63) précisent que l'instrument est un objet dont dépend une action ou un comportement. En résumé, selon l'approche ontologique, comme une sous-classe ontologique des noms d'artefacts, un instrument est un objet manipulable, destiné à permettre l'exécution d'une action, et doté de fonctionnalités préalables. Nous pouvons ainsi identifier trois caractéristiques prototypiques : (1) sa manipulabilité, (2) son orientation vers l'exécution d'une action, et (3) la présence de fonctions préalables.

Les noms d'instruments regroupent eux-mêmes plusieurs sous-classes ontologiques. Comme mentionné dans la section 1.3.3, de nombreux auteurs ont proposé les sous-classes ontologiques des noms d'instrument : Instrument de captation, Machine, Appareil, Outil, Substance chimique, Moyen de transport, Récipient, Mécanisme, Pièce de vêtement et Ustensile (Dubois 1962, Sliwa 1990, Grevisse & Goosse 2016, Rainer 2021).

3.2.1.2 Sélection et définition des sous-classes ontologiques prototypiques

Trois types de ressources ont été mobilisés dans le cadre de cette étude : (i) *Le Grand Robert* (2005), qui constitue notre base de données principale ; (ii) les travaux de philosophes des sciences et de la technique, tels que Simondon (2001) et Vial (2012, 2017) ; (iii) des études lexico-sémantiques, comme Polguère (2003).

Le recours aux travaux philosophiques (Simondon 2001 et Vial 2012, 2017) se justifie pour deux raisons : d'une part, en tant que philosophes des sciences et de la technique, ils se sont attachés à distinguer différentes sous-catégories d'instruments, en fournissant des critères ontologiques précis ; d'une part, certaines de leurs définitions ou distinctions sont reprises dans *Le Grand Robert*, ce qui permet d'appuyer les analyses sémantiques sur des fondements à la fois lexicographiques et théoriques. Cette complémentarité permet ainsi d'établir des distinctions ontologiques robustes entre les différents noms d'instruments.

L'étape suivante consiste à préciser les définitions de chaque sous-classe ontologique. Nous commençons par l'outil, qui est plus basique que les sous-classes ontologiques. Selon Simondon (1969 : 114), celui-ci se définit comme « l'objet technique qui permet de prolonger et d'armer le corps pour accomplir un geste » À l'inverse, l'instrument de captation est défini comme « l'objet technique qui permet de

prolonger et d'adapter le corps pour obtenir une meilleure perception et s'oppose à outil en ce qu'il inclut les systèmes artificiels capables d'améliorer la perception » (Simondon, 2001 : 114).

Les autres sous-classes ne font pas l'objet de définitions aussi précises ou explicites dans son ouvrage. Concernant la Machine, Simondon (2001 : 113) a adapté la définition de l'Encyclopédie (Diderot et al. : 1900) : « les machines ne sont pas directement considérées comme des automates par les encyclopédistes, elles sont plutôt considérées comme un assemblage de dispositifs élémentaires. » A partir de cette définition, les caractéristiques de la machine et la relation entre la machine et les humains sont bien abordées. Or, pour mener à bien cette recherche, cette définition s'avère insuffisante, car elle reste trop générale et tend à confondre la machine avec l'appareil. Selon Rainer (2005 : 28), l'interprétation instrumentale des noms en *-eur* et *-euse* peut souvent être interprétée comme le résultat d'ellipses (omission du nom) dans les modèles appareil (nom masculin) + adjectif en *-eur* ou machine (nom féminin) + adjectif en *-euse*. Les différentes terminaisons adjectivales reflètent la distinction de genre entre les deux noms principaux. Ainsi, pour justifier ce postulat, nous avons besoin de distinguer machines et appareils. Pour ce faire, nous sélectionnons un autre philosophe, Stéphane Vial, qui fait bien cette distinction dans les approches technique et phénoménologique en abordant les instruments techniques (appareils et machines).

À partir du concept de système technique, Vial (2012 ; 2017) introduit les notions de machine et d'appareil. La machine, emblématique de la révolution industrielle, est définie de manière implicite comme un objet technique conçu pour accomplir une tâche donnée, à l'aide d'un système technique et de diverses formes d'énergie. Ces tâches sont en général mécanisées, bien que Vial ne précise pas toujours leur nature. Nous proposons ainsi de distinguer deux types de mécanisation opérés par les machines. D'une part, certaines machines mécanisent des tâches de main-d'œuvre, comme c'est le cas de EPLUCHEUSE, de la TONDEUSE ou du GAZEUR, qui remplacent l'intervention humaine dans des gestes répétitifs. D'autre part, d'autres machines mécanisent des phénomènes naturels, comme CONGELATEUR, FRIGO ou RADIATEUR, qui reproduisent artificiellement des effets naturels tels que le froid ou la chaleur à l'aide de systèmes techniques complexes. Par exemple, avant l'invention du CONGELATEUR, la GLACIERE permettait de conserver les aliments en utilisant de la glace naturelle. Le congélateur remplace ce processus par une production artificielle de froid, rendue possible par la transformation de l'électricité en énergie thermique. En somme, en prolongeant la

pensée de Vial (2012), nous définissons la machine comme un objet technique conçu pour accomplir une tâche donnée, en mobilisant un système technique et des énergies variées, soit pour remplacer le travail humain, soit pour reproduire artificiellement un phénomène naturel.

Concernant l'appareil, Vial (2012 ; 2016) prolonge l'approche phénoménologique de Gaston Bachelard, notamment à travers la notion de phénoménotechnique introduite par ce dernier. Selon Bachelard (1938 : 61), *« la science réalise ses objets, sans jamais les trouver tout faits. […] Un concept est devenu scientifique dans la proportion où il est devenu technique, où il est accompagné d'une technique de réalisation. »* La phénoménotechnique consiste ainsi à faire apparaître des phénomènes par le biais d'appareils techniques, sans lesquels ils seraient invisibles ou inexistants en tant que phénomènes. Dans cette lignée, Vial (2016) définit l'appareil comme ce qui permet la perception de réalités invisibles, notamment dans le cadre du système technique numérique. Il affirme que les *«* êtres numériques *»* contraignent l'humain à renégocier l'acte même de perception, en développant une nouvelle sensibilité à des objets inédits. Cette perception n'est ni naturelle ni spontanée, mais suppose un travail actif du sujet.

Vial (2016) insiste également sur le fait que les réalités scientifiques modernes comme celles de la physique quantique, n'existent pas en dehors des appareils capables de les révéler. En ce sens, *«* les phénomènes scientifiques […] ne commencent vraiment qu'au moment où l'on met en marche les appareils *»* (Bachelard, 1951 : 5). Ainsi, « pour devenir phénoménal », le monde microphysique « doit être manifesté techniquement par des appareils adaptés », tels que les accélérateurs de particules (Vial 2016).

Cependant, la définition proposée par Vial (2017) reste relativement restreinte, car elle se concentre sur des dispositifs numériques ou scientifiques complexes. Si l'on adopte cette définition stricte, il devient difficile de classer certains objets comme LAMPE, PLAFONNIER, ECOUTEUR, voire CHARGEUR. Ces objets ne relèvent clairement ni de la machine, ni pleinement de l'appareil au sens étroit. Pourtant, si des objets comme LAMPE, PLAFONNIER ou ECOUTEUR engagent une interaction directe avec la perception humaine, d'autres comme CHARGEUR n'ont pas de fonction perceptive. Plus précisément, le cas du CHARGEUR est particulièrement révélateur. Il est classé comme « appareil » dans *Le Grand Robert*, et sa fonction principale est de conduire l'électricité vers un autre dispositif. Son appartenance à la sous-classe ontologique « appareil » repose plutôt sur sa propriété à utiliser l'électricité son usage de l'électricité, son

fonctionnement non mécanique, et son intégration dans un système technique destiné à conduire de l'énergie.

Contrairement à la machine, l'appareil n'exécute aucune tâche mécanique, mais assure une fonction technique fondamentale : la conduite de l'énergie. Ce type d'objet, qui fonctionne de manière statique, partage donc davantage de caractéristiques avec l'appareil qu'avec la machine. Il met en évidence une autre dimension possible de l'appareil, non pas uniquement liée à la perception, mais également à la médiation énergétique. Ainsi, en nous appuyant sur Bachelard (1951) et Vial (2012 ; 2016), nous proposons d'élargir la définition de l'appareil. Nous le définissons comme un objet technique structuré par un système technique, destiné soit à observer, mesurer ou révéler des informations en lien avec la perception humaine, soit à conduire une énergie.

Suite aux machines et aux appareils, tournons-nous maintenant vers une autre catégorie ontologique : le récipient. Polguère (2003 : 63) définit le récipient comme un type particulier de contenant ayant l'utilisation proprement dite (stockage, traitement, etc. de contenu). Cette définition met en évidence deux traits fondamentaux : d'une part, le récipient appartient à la classe générale des contenants ; d'autre part, il se distingue par une finalité fonctionnelle, en ce qu'il permet non seulement de stocker, mais aussi d'affecter le contenu qu'il contient.

Après avoir présenté les définitions des principales sous-classes ontologiques issues de la littérature (telles que la machine, l'appareil ou encore le récipient), il reste à définir d'autres sous-classes ontologiques moins traitées dans les travaux de recherche. Faute d'analyses théoriques suffisantes à leur sujet, nous retenons les définitions proposées par *Le Grand Robert* (2005), qui offrent des repères lexicographiques utiles. L'ustensile est défini comme « objet ou accessoire d'usage domestique composé généralement d'une seule pièce et dont l'utilisation n'exige pas la mise en mouvement d'un mécanisme ». Le moyen de transport (ou véhicule) est un « engin muni de roues ou d'un moyen de propulsion, servant à transporter des personnes ou des marchandises ». Le mécanisme est une « combinaison, agencement de pièces, d'organes, montés en vue d'un fonctionnement d'ensemble ». Les substances (produits) chimiques sont des « corps obtenus à l'aide de procédés chimiques ». Enfin, le vêtement se définit comme « objets fabriqués pour couvrir le corps humain, le cacher, le protéger, le parer (coiffure, chaussures, linge, habits et accessoires) ».

Nous disposons désormais d'un ensemble de définitions pour toutes les sous-classes ontologiques évoquées. La question suivante est alors de déterminer lesquelles

d'entre elles peuvent être considérées comme des sous-classes des noms d'instruments prototypiques du point de vue ontologique. Ce premier tri sera effectué dans la sous-section 3.2.1.2.

3.2.1.2 Le premier tri pour les sous-classes ontologiques

Avant d'effectuer un premier tri, il convient d'examiner les étiquettes des sous-classes ontologiques, car les catégories évoquées présentent des phénomènes hétérogènes. Tout d'abord, en ce qui concerne la catégorie « moyen de transport » (ou véhicule), sa définition est la suivante : « engin muni de roues ou d'un moyen de propulsion, servant à transporter des personnes ou des marchandises ». Cette définition montre que les moyens de transport sont généralement constitués de machines. En effet, ils peuvent être considérés comme une forme particulière de machine. Ainsi, nous avons décidé de supprimer l'étiquette « moyen de transport » et de reclasser ces noms sous la catégorie « machine ».

Par ailleurs, selon Sliwa (1990 : 60) et *Le Grand Robert* (2005), les ustensiles sont souvent constitués de récipients (et, dans une moindre mesure, d'outils), et leur usage est généralement limité à la sphère domestique. Cette catégorie est donc moins homogène que d'autres. Par exemple, une poêle peut être classée à la fois dans la catégorie des « ustensiles » et dans celle des « récipients ». Afin d'assurer une meilleure homogénéité dans notre identification des sous-classes prototypiques et leur classification, il paraît nécessaire de reclasser certains noms d'instruments en supprimant les étiquettes ambiguës. Par conséquent, l'étiquette « ustensile » n'a pas été conservée, car elle recouvre des objets déjà désignés par les catégories plus précises de « récipient » ou d'« outil ».

Nous entamons maintenant le premier tri afin d'identifier, dans une perspective ontologique, les sous-classes qui correspondent au prototype des noms d'instruments. En premier lieu, nous avons constaté que les instruments de captation, les outils, les machines et les appareils remplissent toutes les caractéristiques identifiées des noms d'instruments prototypiques dans la perspective ontologique (voir la section 3.2.1.1), dans la mesure où les agents les manipulent ou interviennent activement pour effectuer une action. A titre d'exemple, le COUTEAU est une sorte d'outil servant à couper quelque chose. Il est entièrement manipulé par l'être humain pour couper un objet remplissant le rôle de Patient. Autre exemple, la LOUPE est une sorte d'instrument de captation, l'utilisateur la tient à la main pour voir plus clairement. Cette LOUPE agrandit l'objet à

voir et rend une image plus claire, agrandie et virtuelle à l'utilisateur. Par contre, les mécanismes qui composent les noms d'instruments (dits « dispositif ») ne doivent pas être envisagés comme des noms d'instruments parce qu'ils ne font que constituer une partie de l'objet instrument. Un agent manipule l'instrument entier pour effectuer une action et pas seulement les dispositifs séparés dont il est constitué. Et surtout, ils ne sont pas manipulés directement par l'agent. Ainsi, les mécanismes ne sont pas considérés comme des sous-classes ontologiques prototypiques.

En deuxième lieu, à partir des caractéristiques prototypiques des instruments (voir la section 3.2.1.1), une partie des récipients devrait être exclue. En effet, au sein de cette sous-classe, certains récipients remplissent deux fonctions distinctes : affecter le contenu ou simplement le conserver. À titre d'exemple, la CAFETIERE permet non seulement d'infuser le café, mais aussi de le contenir après sa préparation. Ce type de récipient devrait donc être considéré comme un nom d'instrument prototypique, car il satisfait à l'ensemble des caractéristiques identifiées : il est manipulable, il permet l'exécution d'une action, et le contenu (le café) est effectivement affecté. En revanche, les récipients dont la seule fonction est de conserver le contenu, sans l'affecter, ne devraient pas être considérés comme des instruments prototypiques. Ils devraient être classés parmi les noms de lieu. Comme le souligne Sliwa (1995 : 65), des récipients tels que SALIERE, ASSIETTE ou ENCRIER sont souvent assimilés à des objets locatifs, car ils sont associés à une fonction de localisation ou de stockage. Autrement dit, lorsqu'un récipient se limite à contenir sans affecter le contenu, il ne peut être considéré comme un véritable instrument au sens prototypique du terme. Parallèlement, étant donné que les récipients dont la fonction est uniquement de stocker sont exclus, les noms de meubles devraient également être écartés des noms d'instruments prototypiques. En effet, bien qu'ils soient destinés à conserver ou à déposer quelque chose, le contenu conservé subit éventuellement un changement de lieu.

En troisième lieu, en ce qui concerne les noms de substances chimiques, Sliwa (1990) & Rainer (2021) les considèrent comme des noms d'instruments. Mais, selon Pustejovsky et Batiukova (2019 : 197), ces noms ne relèvent pas de cette classe, car il s'agit de noms massifs ayant une fonction matérielle, et non instrumentale. Cette fonction matérielle consiste en sa nature physique ou chimique, au lieu d'une structure fonctionnelle conçue pour accomplir une tâche. Pour évaluer ces hypothèses, il convient de vérifier si ces noms satisfont aux caractéristiques prototypiques précédemment identifiées. Prenons le cas de LESSIVE : avec la lessive, nous pouvons nettoyer le linge.

Il s'agit d'un artefact manipulé par un agent en vue de produire un effet, et dans ce cas, le linge (en tant que patient) est effectivement affecté, puisqu'il devient propre. Ainsi, d'un point de vue ontologique, les substances chimiques remplissent les caractéristiques prototypiques précédemment identifiées des noms d'instruments. Nous les considérons donc comme une sous-classe prototypique, en nous appuyant sur les analyses de Sliwa (1990) et de Rainer (2021).

Enfin, en ce qui concerne les pièces de vêtement, ce sont des objets destinés à couvrir le corps humain. Cependant, une fois qu'ils ont été enfilés, aucune autre action n'est effectuée par leur intermédiaire. Autrement dit, bien que certaines parties du corps soient couvertes, les vêtements ne permettent pas l'exécution d'une action sur un patient ou un thème externe. En ce sens, ils ne satisfont pas à la caractéristique prototypique liée à l'orientation vers l'exécution d'une action. Ainsi, ils devraient être exclus des sous-classes ontologiques prototypiques.

A l'issue de ce premier tri, nous conservons les sous-classes ontologiques prototypiques dans la perspective ontologique comme suit : Instrument de captation, Outil, Machine, Appareil, Substance chimique ainsi que Récipient.

Dorénavant, nous devons vérifier que les noms issus de ces sous-classes ontologiques se comportent réellement comme des noms d'instruments d'un point de vue linguistique. Autrement dit, peuvent-ils être classés, sur le plan linguistique, comme des noms d'instruments ? Cela justifie une deuxième étape de l'analyse, fondée cette fois sur une perspective linguistique. Dans ce qui suit, nous présentons les définitions existantes de la notion de nom d'instrument selon cette perspective, en croisant les approches issues de la linguistique structurelle-générative (§1.2.1.2) et de la linguistique cognitive (§1.2.1.3).

3.2.2 Dans la perspective linguistique

Dans cette sous-section, nous présenterons tout d'abord les définitions existantes de la notion de nom d'instrument selon une perspective linguistique. Ensuite, nous procéderons à une validation linguistique des sous-classes ontologiques, à l'aide de tests linguistiques, afin d'identifier celles qui correspondent au prototype de l'instrument tel qu'il est défini à la fois sur le plan ontologique et linguistique (voir la Section 1.2).

3.2.2.1 Les définitions existantes

Fillmore (1968) constitue un point de départ fondamental, sa *Case Grammar* fournissant la base sémantique pour la définition des noms d'instruments. Il définit l'instrument comme « le cas d'une force ou d'un objet inanimé impliqué de manière causale dans l'action ou l'état désigné par le verbe »[32] (Fillmore 1968 : 46). Par la suite, Carnie (2007), Pustejovsky & Batiukova (2019) et Sliwa (1990) ont enrichi cette conception en y intégrant des considérations ontologiques, insistant sur le fait que l'instrument doit être une entité concrète, utilisée par un agent pour effectuer une action (Carnie 2007 ; Pustejovsky & Batiukova 2019 ; Sliwa 1990 : 59). Dans le domaine de la morphologie dérivationnelle, Namer & Villoing (2008) proposent une caractérisation plus fine des noms d'instruments, en soulignant deux propriétés spécifiques : « l'instrument possède deux propriétés : celle de provenir d'une création humaine, et celle d'être destiné à un usage particulier » (Namer & Villoing 2008 : 1542). Elles complètent cette approche par une série de tests linguistiques permettant d'identifier les noms d'instruments prototypiques, constituant ainsi un outil opératoire précieux pour l'analyse morphologique, et un support méthodologique essentiel pour le présent travail de thèse. Ces tests seront détaillés dans la section suivante.

Dans le cadre de la linguistique cognitive, plusieurs travaux s'inscrivent dans la lignée de Fillmore, en définissant l'instrument à partir de deux rôles sémantiques centraux, agent et patient, ainsi que d'une contrainte ontologique : celle d'être un objet inanimé. Selon Langacker (2014 : 216), « le rôle d'instrument est celui d'un objet inanimé manipulé par un agent pour agir sur un patient ; c'est par l'instrument que l'énergie est transmise de l'agent au patient. La présence d'un instrument qui médiatise ce transfert d'énergie de l'agent vers le patient constitue une caractéristique optionnelle des actions prototypiques. »[33] Sur cette base, Ungerer & Schmid (2002 : 184) élargissent la perspective en intégrant d'autres rôles sémantiques comme le thème et la cible, afin de préciser la définition des noms d'instruments. En résumé, selon cette approche, remplissant un rôle sémantique dans la structure argumentale du verbe issu

[32] Le texte original : "the case of the inanimate force or object causally involved in the action or state identified by the verb".

[33] Le texte original : "the instrument role is that of an inanimate object manipulated by an agent to affect a patient; it is through the instrument that energy is transmitted from the agent to the patient. The presence of an instrument mediating the transfer of energy from agent to patient is an optional feature of prototypical actions."

de la représentation sémantique du dérivé au niveau syntactico-sémantique, un instrument est une entité concrète, utilisée par un agent pour effectuer une action.

Cependant, à partir de cette définition, nous avons observé que la source de l'énergie est en provenance de l'agent, à savoir les êtres humains ou autre être animé. Mais avec l'essor de la technologie, les sources sont plus variées. Selon Rainer (2021), « les noms d'instruments ont connu une profonde transformation à la suite des révolutions scientifique et industrielle. Alors que les outils et ustensiles utilisés au foyer, dans l'agriculture, la chasse, le commerce, la guerre et d'autres domaines traditionnels de l'activité humaine étaient restés, pour l'essentiel, inchangés de l'époque romaine jusqu'au XVIIIe siècle, l'essor spectaculaire des sciences et des techniques à partir de ce siècle a fait apparaître une myriade de nouveaux instruments, parmi lesquels diverses machines autopropulsées. »[34] Ainsi, avec le développement de la technologie, les sources de l'énergie sont en provenance non seulement de l'être humain, mais aussi d'autres sources d'énergie, telles que le vent, l'électricité, la vapeur, etc. En effet, les noms d'instruments peuvent être associés à des sources d'énergie variées : ils peuvent être actionnés par des êtres humains ou bien par une énergie extérieure — qu'elle soit naturelle, électrique ou chimique — déclenchée par l'intervention d'un agent. Même si ces instruments fonctionnent grâce à des énergies différentes, ils peuvent toujours être considérés comme des noms d'instruments puisqu'un agent est toujours présent, même de manière indirecte, pour mettre en route le référent, lequel effectue alors une action. À ce titre, ces noms continuent d'assumer le rôle sémantique d'instrument au niveau syntaxico-sémantique. En fait, ces considérations d'ordre typiquement référentielles se reflètent partiellement dans leur rôle au niveau syntaxico-sémantique, dans la mesure où les propriétés conceptuelles d'instrument conditionnent leur rôle sémantique qu'ils occupent en langue.

Ainsi, selon des définitions évoquées, dans une perspective intégrative, les noms d'instrument prototypique doivent réunir l'ensemble des caractéristiques suivantes :

- Un artefact est utilisé pour mener une action relative à une entité. Plus précisément, si cette entité est considérée comme le Patient, son état est affecté. Si cette entité n'est pas affectée, il s'agit donc d'un thème. Si cette entité

[34] Le texte original : "Instrument nouns […] have suffered a major overhaul in the wake of the scientific and industrial revolutions. […] While tools and implements used in the household, in husbandry, hunting, trade, war, and other traditional areas of human activity remained largely unchanged from Roman times until the 18th century, the spectacular leap forward in science and technology since the 18th century has brought along a myriad of new instruments, including self-propelled machines of different kinds."

fonctionne comme cible, sa position est déplacée.

- L'utilisation ou l'intervention physique par l'agent. Concrètement, cela implique deux actions : soit les deux sont réalisées par l'agent, soit l'une est effectuée par l'agent et l'autre par les instruments.
- Le verbe concerné est issu dans la représentation sémantique de la base du dérivé.

En réunissant ces caractéristiques, la définition des noms d'instrument prototypiques dans une perspective intégrant la dimension ontologique et la dimension linguistique est la suivante : un instrument est un artefact qu'utilise un agent (directement ou indirectement via une énergie extérieure - naturelle, électrique, chimique) pour mener une action relativement à une entité. Cette entité peut être considérée comme un patient si elle est affectée, comme un thème si elle ne l'est pas (permettant ainsi d'engendrer une nouvelle entité relative), ou comme une cible si elle est déplacée. De plus, l'instrument joue un rôle sémantique « instrument » au niveau syntactico-sémantique en relation avec les verbes (que ce soit la base verbale ou le verbe dans un procès, c'est-à-dire le verbe sous-jacent).

3.2.2.2 Validation linguistique des sous-classes ontologiques avec les tests linguistiques

Afin d'identifier les noms d'instruments prototypiques, nous adaptons les tests linguistiques proposés par Namer & Villoing (2008 : 1543), qui sont aussi utilisés dans la recherche de Huyghe et Tribout (2015). Les tests (a) et (b) servent à tester l'instrumentalité. Ces deux tests constituent ainsi des critères minimaux pour identifier la nature instrumentale d'un nom.

Par ailleurs, le test (c), emprunté à Cadiot (1991), sert à identifier les noms d'instruments prototypiques (Namer & Villoing 2008 : 1542) en testant leur compatibilité avec des verbes dynamiques, par opposition aux verbes statifs. Finalement, le test (d) témoigne, en particulier, de la manière dont l'instrument se substitue parfois à l'agent en position sujet (Villoing & Namer 2008 : 1543). Dans le cadre de notre étude, nous mobilisons un aspect spécifique de la Grammaire Cognitive, le « trajector–landmark », proposé par Langacker afin d'expliquer pourquoi certains noms d'instruments peuvent occuper cette position, tandis que d'autres ne le peuvent

pas. En effet, le participant le plus saillant, appelé trajector (tr), est l'entité conçue comme étant localisée, évaluée ou décrite. Il constitue généralement le centre d'attention dans la relation profilée. Un autre participant peut être mis en relief de manière secondaire : c'est ce qu'on appelle le landmark (lm). En règle générale, dans une phrase simple, le trajector correspond au sujet et le landmark à l'objet. Toutefois, contrairement aux catégories grammaticales traditionnelles que sont le sujet et l'objet, la distinction trajector–landmark est plus générale et flexible, et peut s'appliquer à un plus grand nombre de situations linguistiques (Langacker 2008 : 70 ; Mei 2017). Autrement dit, si l'instrument est identifié comme trajector, il se réalise le plus souvent en position de sujet dans l'énoncé. Par rapport aux tests (a) et (b), la réussite aux tests (c) et (d) permet de distinguer entre une instrumentalité forte et une instrumentalité faible chez les noms d'instruments prototypiques. Autrement dit, les tests (a) et (b) constituent des épreuves fondamentales pour vérifier l'instrumentalité prototypique d'un nom, tandis que plus un nom réussit les tests (c) et (d), plus son degré d'instrumentalité peut être considéré comme fort.

Ce cadre permettra ainsi d'éclairer les résultats obtenus aux différents tests. En tout cas, tous les tests adaptés (Tableau 3.1) nous fournissent un bon outil pour identifier les noms d'instruments prototypiques. Il convient de préciser que, dans le cadre de ces tests, le verbe utilisé doit correspondre au rôle télique du nom d'instrument dans sa structure de qualia.

Tableau 3.1 Tests d'identification Instrument

TEST a)	NP0 Verbe au moyen de DET INST NP0 Verbe avec DET INST
TEST b)	NP0 Verbe avec DET INST ↔ NP0 utilise DET INST pour Vinf
TEST c)	NP0 Verbe avec DET INST ↔ NP0 Verbe au INST
TEST d)	NP0 Verbe NP1 avec DET INST ↔ DET INST Verbe NP1

En vue d'illustrer les quatre tests présentés dans le tableau 3.1, nous allons proposer quatre exemples représentatifs.

1) **TEST a) :** *Je hache ma viande avec un hachoir*. Dans ce test, le nom d'instrument HACHOIR est introduit par la préposition avec, ce qui confirme son interprétation comme complément à valeur instrumentale.

2) **TEST b) :** *Je hache ma viande avec un hachoir*. ↔ *J'utilise un hachoir pour hacher ma viande*. Ici, le nom HACHOIR occupe la position de complément d'objet du verbe UTILISER, marquant ainsi une interprétation instrumentale.
3) **TEST c) :** *Je hache ma viande avec un hachoir*. ↔ *Je hache ma viande au hachoir*. (Namer & Villoing 2008 : 1542) L'emploi de différentes prépositions teste le prototype instrumental dans la langue. À travers ces deux phrases, il apparaît que le nom HACHOIR accepte la variation prépositionnelle (avec / à), ce qui confirme son prototype instrumental.
4) **TEST d) :** *Je hache ma viande avec un hachoir*. ↔ *Le hachoir hache ma viande*. (Namer & Villoing 2008 : 1542). Ce test témoigne de la manière dont l'instrument peut parfois se substituer à l'agent en position sujet. Ainsi, le nom HACHOIR peut occuper la position de trajector dans la scène, et être conceptualisé comme l'élément qui initie ou cause l'action.

Nous procéderons désormais à la validation linguistique des sous-classes ontologiques précédemment identifiées. Cette validation sera menée une par une. Chaque sous-classe ontologique comprend cinq lexèmes sélectionnés pour les tests. Le choix de ces cinq lexèmes repose principalement sur leur fréquence d'usage en français, ou sur leur caractère représentatif de la sous-classe concernée.

Cette phase de validation linguistique poursuit deux objectifs principaux : d'une part, déterminer si la sous-classe ontologique examinée relève des noms d'instrument prototypiques ; d'autre part, identifier, au sein de chaque sous-classe, les lexèmes qui satisfont aux critères prototypiques des noms d'instruments et ceux qui s'en écartent. Selon Rainer (2021 : 1), certains lexèmes relevant de la classe des récipients présentent une ambiguïté catégorielle entre noms d'instruments et noms de lieu, rendant leur statut ontologique incertain. Cette démarche vise notamment à repérer, parmi les lexèmes situés à la frontière de deux classes ontologiques, ceux dont les propriétés justifient leur appartenance aux noms d'instruments.

3.2.2.2.1 Outils

Nous allons appliquer les tests aux noms [1] - [5] suivants :

[1] RASOIR : « Instrument à tranchant très fin servant à raser. »

[2] EPLUCHEUR : « Instrument (à couteaux, à disques abrasifs) pour éplucher les légumes. » (Le Grand Robert 2005)

[3] RACLOIR : « Outil servant à racler. » (Le Grand Robert 2005)

[4] COUTEAU : « Instrument tranchant servant à couper, composé d'une lame et d'un manche. » (Le Grand Robert 2005)

[5] PERÇOIR : « Outil pour percer. » (Le Grand Robert 2005)

- Test a) pour tester l'instrumentalité :

1) Je me rase avec un rasoir.
2) J'épluche des légumes avec un éplucheur.
3) Il racle la peinture écaillée avec un racloir.
4) Je coupe ce steak avec un couteau.
5) Il perce un mur avec un perçoir.

- Test b) pour tester l'instrumentalité :

6) J'utilise un rasoir pour me raser.
7) J'utilise un éplucheur pour éplucher des légumes.
8) Il utilise un racloir pour racler la peinture écaillée.
9) J'utilise un couteau pour couper le steak.
10) Il utilise un perçoir pour percer un mur.

Pour les Tests a) et b), les outils réussissent ces tests en raison de leur fonction directement liée à des actions à valeur de changement d'état.

- Test c) pour identifier les noms d'instruments prototypiques :

11) Je me rase au rasoir.
12) J'épluche des légumes à l'éplucheur.
13) Il racle la peinture écaillée au racloir.
14) Je coupe ce steak au couteau.
15) Il perce un mur au perçoir.

Concernant le Test c), nous avons constaté que le succès complet démontre une forte prototypicalité, indiquant une association naturelle et typique avec la préposition « à ».

- Test d) pour identifier les noms d'instruments prototypiques :

16) Un rasoir rase les poils.
17) Un éplucheur épluche les légumes.

18) Un racloir racle la peinture écaillée.
19) Un couteau coupe la viande.
20) Un perçoir perce les murs.

Enfin, en ce qui concerne le test (d), tous les noms d'outils ont obtenu des résultats positifs. Cela confirme non seulement leur instrumentalité élevée, mais aussi leur capacité à être conceptualisés comme trajectors dans la structuration des scènes linguistiques, selon la théorie trajector–landmark de Langacker. Leur prototypicalité tient précisément au fait que l'instrument, dans ce cas, constitue le centre de l'événement décrit.

Tableau 3.2 Résultat des tests pour les noms des Outils

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
RASOIR	+	+	+	+	4
EPLUCHEUR	+	+	+	+	4
RACLOIR	+	+	+	+	4
COUTEAU	+	+	+	+	4
PERÇOIR	+	+	+	+	4

Selon le tableau 3.2, tous les lexèmes (RASOIR, EPLUCHEUR, RACLOIR, COUTEAU, PERÇOIR) réussissent parfaitement tous les tests, indiquant une forte instrumentalité prototypique. Ils s'associent facilement aux verbes d'action et aux prépositions typiques (*à*), reflétant leur nature essentielle comme prolongement physique du geste (Simondon, 2001). Ainsi, la sous-classe Outil représente sans doute la sous-classe ontologique prototypique des noms d'instruments.

3.2.2.2.2 Instruments de captation

Concernant les instruments de captation, nous avons choisi les noms REGLE, LUNETTE, LORGNON, RAPPORTEUR, APPRECIATEUR comme les noms d'instruments que nous testons, comme nous le faisons pour les outils.

[1] REGLE : « Instrument constitué par une planchette allongée ou par une tige à arêtes rectilignes et qui sert à guider le crayon ou la plume quand on trace un trait, à mesurer une longueur, etc. » (Le Grand Robert 2005)

[2] LUNETTE : « Instrument d'optique composé d'une ou plusieurs lentilles et servant

soit à augmenter le diamètre apparent des objets, soit à rendre la vue plus nette et plus distincte. » (Le Grand Robert 2005)

[3] LORGNON : « Ensemble de deux lentilles et de leur monture sans branches, tenue à la main par une sorte de manche ou maintenu sur le nez par un ressort » (Le Grand Robert 2005)

[4] RAPPORTEUR : « Instrument à l'aide duquel on peut mesurer les angles ou construire (rapporter) un angle d'une mesure donnée. » (Le Grand Robert 2005)

[5] APPRECIATEUR : « Aréomètre estimant la quantité de pains que peut donner une farine, d'après sa composition. » (Le Grand Robert 2005)

A partir de ces cinq définitions de noms de l'instrument de captation, il est difficile de révéler leur comportement au niveau sémantique dans la langue. Ainsi, nous décidons de nous appuyer sur les tests linguistiques pour examiner leur comportement dans l'usage de la langue et identifier les noms d'instruments prototypiques dans le plan purement linguistique.

- Test a) pour tester l'instrumentalité :

21) Je mesure les objets avec une règle.
22) J'agrandis l'image des étoiles avec une lunette.
23) Il corrige sa vue avec un lorgnon.
24) Je mesure un angle avec un rapporteur.
25) Il estime la quantité de pains avec un appréciateur.

- Test b) pour tester l'instrumentalité :

26) J'utilise une règle pour mesurer des objets.
27) J'utilise une lunette pour agrandir l'image des étoiles.
28) Il utilise un lorgnon pour corriger sa vue.
29) J'utilise un rapporteur pour mesurer un angle.
30) Il utilise un appréciateur pour estimer la quantité de pain.

Pour les Tests a) et b), nous avons constaté que les instruments de captation réussissent car ils possèdent clairement une fonctionnalité explicite (MESURER, OBSERVER). Ils sont facilement associés à des contextes verbaux d'action ou de perception.

- Test c) pour identifier les noms d'instruments prototypiques :

31) Je mesure des objets à la règle.
32) J'agrandis les étoiles à la lunette.
33) ?Il corrige sa vue au lorgnon.
34) Je mesure un angle au rapporteur.
35) ?Il estime la quantité de pains à l'appréciateur.

- Test d) pour identifier les noms d'instruments prototypiques :

36) Une règle mesure des distances.
37) Une lunette agrandit l'image des étoiles.
38) Un lorgnon corrige la vue.
39) Un rapporteur mesure des angles.
40) Un appréciateur estime la quantité de pains.

Concernant le test (d), les résultats témoignent d'une réussite générale. Ce succès indique que les noms testés peuvent occuper la position de sujet dans la phrase. Dans le cadre de la théorie trajector–landmark de Langacker, cela signifie qu'ils sont conceptualisés comme trajectors, c'est-à-dire comme les entités saillantes dans la structuration de la scène. Autrement dit, c'est leur instrumentalité elle-même qui est mise au premier plan dans la scène décrite.

Tableau 3.3 Résultat des tests pour les noms des Instruments de captation

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
REGLE	+	+	+	+	4
LUNETTE	+	+	+	+	4
LORGNON	+	+	+	+	4
RAPPORTEUR	+	+	+	+	4
APPRECIATEUR	+	+	-	+	3

Selon le Tableau 3.3, nous avons constaté que la sous-classe ontologique des instruments de captation est une sous-classe prototypique, avec quelques limites dans les contextes des verbes statifs.

3.2.2.2.3 Appareil

Nous allons appliquer les tests aux noms [1] - [5] suivants :

[1] AGRANDISSEUR : « Appareil employé pour les agrandissements » (Le Grand Robert 2005)

[2] AMPLIFICATEUR : « Appareil destiné à augmenter l'amplitude d'un phénomène (oscillations électriques en particulier) et qui fournit une puissance utile de sortie supérieure à la puissance d'entrée. » (Le Grand Robert 2005)

[3] PLAFONNIER : « Appareil d'éclairage fixé au plafond sans être suspendu (à la différence du lustre, de la suspension). » (Le Grand Robert 2005)

[4] SCANNEUR : « Appareil électronique de photogravure produisant des clichés typographiques, des cylindres gravés d'héliogravure ou des films de demi-teinte ou tramés d'après des documents en noir et blanc ou en couleur, par balayage électronique. » (Le Grand Robert 2005)

[5] PROJECTEUR : « Appareil d'optique dans lequel les rayons d'une source lumineuse intense sont réfléchis et projetés en un faisceau parallèle. » (Le Grand Robert 2005)

- Test a) pour tester l'instrumentalité :

41) J'agrandis des photos avec un agrandisseur.
42) Elle amplifie les sons avec un amplificateur.
43) J'éclaire la pièce avec un plafonnier.
44) Il scanne des papiers avec un scanneur.
45) Il projette des images avec un projecteur.

- Test b) pour tester l'instrumentalité :

46) J'utilise un agrandisseur pour agrandir des photos.
47) Elle utilise un amplificateur pour amplifier les sons.
48) J'utilise un plafonnier pour éclairer la pièce.
49) Il utilise un scanneur pour scanner des papiers.
50) Il utilise un projecteur pour projeter des images.

Pour les Tests a) et b), tous les noms d'appareil réussissent ces deux tests. Cela signifie que ces appareils possèdent une instrumentalité facilement identifiable.

- Test c) pour identifier les noms d'instruments prototypiques :

51) J'agrandis des photos à l'agrandisseur.
52) Elle amplifie les sons à l'amplificateur.
53) ?J'éclaire la pièce au plafonnier.

54) Il scanne des papiers au scanneur.
55) Il projette des images au projecteur.

Cependant, pour le test c), il existe seulement un nom qui échoue à ce test. Parce que la préposition typique *à* employée avec ces appareils suppose un cadre d'action dynamique, associé à une manipulation physique directe et active, souvent incompatible avec certains verbes à valeur non-changement d'état, soit ECLAIRER.

- Test d) pour identifier les noms d'instruments prototypiques :

56) Un agrandisseur agrandit des photos.
57) Un amplificateur amplifie les sons.
58) Un plafonnier éclaire une pièce.
59) Un scanneur scanne des papiers.
60) Un projecteur projette des images.

Concernant le test (d), sa réussite révèle que les noms d'appareils peuvent occuper la position de sujet dans la phrase. Autrement dit, ils sont conceptualisés comme trajectors dans la scène, c'est-à-dire les entités mises au premier plan dans la structuration de l'événement. Cette possibilité repose sur le fait que leur instrumentalité est perçue comme centrale dans le déroulement de l'action décrite.

Tableau 3.4 Résultat des tests pour les noms des Appareils

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
AGRANDISSEUR	+	+	+	+	4
AMPLIFICATEUR	+	+	+	+	4
PLAFONNIER	+	+	-	+	3
SCANNEUR	+	+	+	+	4
PROJECTEUR	+	+	+	+	4

En bref, selon le tableau 3.4, cette sous-classe ontologique de noms (AGRANDISSEUR, AMPLIFICATEUR, PLAFONNIER, SCANNEUR, PROJECTEUR) réussit globalement bien les tests a, b et d, mais PLAFONNIER échoue au test c, car son verbe sous-jacent dans la représentation sémantique est statif. Ceci montre des limites de prototypicalité dans le cas des instruments basés sur des états. Pourtant, leur succès général aux tests confirme leur maintien au sein des instruments prototypiques.

3.2.2.2.4 Machine

Nous allons appliquer les tests aux noms [1] - [5] suivants :

[1] TONDEUSE : « Machine ou instrument destiné à couper court, et d'une manière égale, les poils de certains tissus (draps). » (Le Grand Robert 2005)

[2] PERCEUSE : « machine-outil utilisée pour percer des pièces métalliques, pour la finition de pièces, etc. » (Le Grand Robert 2005)

[3] SECHEUSE : « machine pour sécher (le linge) » (Le Grand Robert 2005)

[4] MEULEUSE : « Machine à meuler. » (Le Grand Robert 2005)

[5] VISSEUSE : « Appareil ou machine servant à visser. » (Le Grand Robert 2005)

- Test a) pour tester l'instrumentalité :

61) Je tonds les moutons avec une tondeuse.
62) Il perce des pièces métalliques avec une perceuse.
63) Je sèche le linge avec une sécheuse.
64) Il meule des pièces en fer avec une meuleuse.
65) L'électricien visse l'interrupteur avec une visseuse.

- Test b) pour tester l'instrumentalité :

66) J'utilise une tondeuse pour tondre les moutons.
67) Il utilise une perceuse pour percer des pièces métalliques.
68) J'utilise une sécheuse pour sécher mes chemises.
69) Il utilise une meuleuse pour meuler des pièces en fer.
70) L'électricien utilise une visseuse pour visser l'interrupteur.

Pour les Tests a) et b), tous les noms de cette sous-classe ontologique réussissent ces tests. Ceci indique qu'ils possèdent une instrumentalité frappante.

- Test c) pour identifier les noms d'instruments prototypiques :

71) Je tonds les moutons à la tondeuse.
72) Il perce des pièces métalliques à la perceuse.
73) Je sèche le linge à la sécheuse.
74) Il meule des pièces métalliques à la meuleuse.
75) L'électricien visse l'interrupteur à la visseuse.

Pour le test c), tous les noms de cette sous-classe ontologique réussissent ces tests. Cela indique que les noms appartenant à la sous-classe ontologique Machine sont plus

prototypiques.

- Test d) pour identifier les noms d'instruments prototypiques :

76) La tondeuse tond les moutons.
77) La perceuse perce des pièces métalliques.
78) La sécheuse sèche le linge.
79) La meuleuse meule des pièces métalliques.
80) La visseuse visse un interrupteur.

Concernant le test (d), tous les noms obtiennent un résultat positif. Autrement dit, ils peuvent occuper la position de sujet en tant que trajectors dans une scène.

Tableau 3.5 Résultat des tests pour les noms des Machines

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
TONDEUSE	+	+	+	+	4
PERCEUSE	+	+	+	+	4
SECHEUSE	+	+	+	+	4
MOULEUSE	+	+	+	+	4
VISSEUSE	+	+	+	+	4

Selon le Tableau 3.5, les noms de machine (TONDEUSE, PERCEUSE, SECHEUSE, MEULEUSE, VISSEUSE) réussissent tous les tests, illustrant une forte prototypicalité due à leur fonctionnalité explicite.

3.2.2.2.5 Récipient

Nous allons appliquer les tests aux noms [1] - [5] suivants :

[1] BOUILLOIRE : « Récipient métallique pansu, muni d'un bec et d'une anse, destiné à faire bouillir de l'eau. » (Le Grand Robert 2005)

[2] ARROSOIR : « Ustensile destiné à l'arrosage et qui se compose d'un récipient muni d'une anse et d'un long col (ou *queue*) terminé par une plaque percée de petits trous ou par un ajutage**. »** (Le Grand Robert 2005)

[3] CHOCOLATIERE : « Récipient où l'on verse le chocolat avant de le servir. » (Le Grand Robert 2005)

[4] CRISTALLISOIR : « Récipient en verre dans lequel on peut effectuer une

cristallisation. » (Le Grand Robert 2005)

[5] THEIERE : « Récipient dans lequel on fait infuser le thé » (Le Grand Robert 2005)

■ **Test a) pour tester l'instrumentalité :**

81) Je fais bouillir de l'eau avec une bouilloire.
82) J'arrose les fleurs avec un arrosoir.
83) Je sers du chocolat chaud avec une chocolatière.
84) Je cristallise le sel par dissolution avec un cristallisoir.
85) Je fais infuser du thé avec la théière.

■ **Test b) pour tester l'instrumentalité :**

86) J'utilise une bouilloire pour faire bouillir de l'eau.
87) J'utilise un arrosoir pour arroser les plantes.
88) J'utilise une chocolatière pour servir du chocolat chaud.
89) J'utilise un cristallisoir pour cristalliser le sel par dissolution.
90) J'utilise une théière pour faire infuser du thé.

En ce qui concerne les Tests a) et b) : Tous les récipients (BOUILLOIRE, ARROSOIR, CHOCOLATIERE, CRISTALLISOIR, THEIERE) réussissent ces tests, indiquant que ces référents possèdent clairement une instrumentalité liée à une action définie (BOUILLIR, ARROSER, SERVIR, CRISTALLISER, INFUSER).

■ **Test c) pour identifier les noms d'instruments prototypiques :**

91) Je fais bouillir de l'eau à la bouilloire.
92) Il arrose des plantes à l'arrosoir.
93) Je sers du chocolat chaud à la chocolatière.
94) Je cristallise le sel par dissolution au cristallisoir.
95) ?Je fais infuser du thé à la théière.

Pour le test (c), l'échec de ces noms de récipients s'explique par des contraintes de combinaison lexicale : certains verbes actifs n'acceptent tout simplement pas la syntagme prépositionnel « à + N », même si le nom en question a une instrumentalité potentielle. Ainsi, pour THEIERE, la sémantique du verbe sous-jacent INFUSER dans la représentation sémantique du dérivé n'est pas compatible avec le syntagme prépositionnel « à + N ».

■ **Test d) pour identifier les noms d'instruments prototypiques :**

96) Une bouilloire fait bouillir de l'eau.

97) ?Un arrosoir arrose des plantes.

98) *Une chocolatière sert du chocolat chaud.

99) Un cristallisoir cristallise le sel par dissolution.

100) ?Une théière fait infuser du thé.

Pour le test (d), trois sur cinq récipients ne peuvent pas être conceptualisés comme trajectors, soit ARROSOIR, CHOCOLATIERE et THEIERE. Autrement dit, ils ne peuvent pas occuper la position de sujet dans un scénario. Dans ce scénario, il est nécessaire d'introduire le rôle sémantique d'Agent, plutôt que de l'ellipser, afin de mettre en évidence les autres rôles sémantiques.

Tableau 3.6 Résultat des tests pour les noms des Récipients

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
BOUILLOIRE	+	+	+	+	4
ARROSOIR	+	+	+	-	3
CHOCOLATIERE	+	+	+	-	3
CRISTALLISOIR	+	+	+	+	3
THEIERE	+	+	-	-	2

Comme le tableau 3.6 l'illustre, les noms de récipients (BOUILLOIRE, ARROSOIR, CHOCOLATIERE, CRISTALLISOIR, THEIERE) présentent un bon succès global aux tests, sauf ARROSOIR, CHOCOLATIERE et THEIERE au test d, en raison de leurs verbes sous-jacents spécifiques. Selon Sliwa (1990), pour considérer un récipient comme instrument, il doit posséder une fonction de traitement du contenu, ce qui explique la meilleure instrumentalité de BOUILLOIRE et ARROSOIR.

3.2.2.2.6 Substances chimiques

Nous allons appliquer les tests aux noms [1] - [5] suivants :

[1] DURCISSEUR : « Produit ajouté à un autre, ou appliqué sur quelque chose pour durcir, pour augmenter la résistance. » (Le Grand Robert 2005)

[2] DENTIFRICE : « Préparation, le plus souvent pâteuse, propre à nettoyer et à blanchir les dents. » (Le Grand Robert 2005)

[3] EPAISSISSANT : « Substance capable de rendre plus épais un mélange liquide ou pâteux. » (Le Grand Robert 2005)

[4] PLASTIFIANT : « Se dit d'un composé (polyester) capable de rendre souple une matière plastique. » (Le Grand Robert 2005)

[5] DESHERBANT : « Qui désherbe, fait mourir la mauvaise herbe. » (Le Grand Robert 2005)

- **Test a) : pour tester l'instrumentalité**

101) Je durcis les ongles avec du durcisseur.

102) Je nettoie les dents avec du dentifrice.

103) J'épaissis la sauce avec de l'épaississant.

104) On plastifie le PVC avec du plastifiant.

105) Je désherbe les mauvaises herbes avec du désherbant.

- **Test b) : pour tester l'instrumentalité**

106) J'utilise du durcisseur pour durcir les ongles.

107) J'utilise du dentifrice pour nettoyer les dents.

108) J'utilise de l'épaississant pour épaissir la sauce.

109) On utilise du plastifiant pour plastifier le PVC.

110) J'utilise du désherbant pour désherber les mauvaises herbes.

Pour les tests a) et b), toutes les substances chimiques (DURCISSEUR, DENTIFRICE, EPAISSISSANT, LEVURE, DESHERBANT) réussissent ces tests, confirmant leur rôle instrumental explicite lié à leur fonction principale (DURCIR, NETTOYER, EPAISSIR, PLASTIFIER, DESHERBER).

- **Test c) pour identifier les noms d'instruments prototypiques :**

111) Je durcis les ongles au durcisseur.

112) Je nettoie les dents au dentifrice.

113) J'épaissis la sauce à l'épaississant

114) On plastifie le PVC au plastifiant.

115) Je désherbe au désherbant.

- **Test d) pour identifier les noms d'instruments prototypiques :**

116) Le durcisseur durcit les ongles.

117) Le dentifrice nettoie les dents.

118) L’épaississant épaissit la sauce.

119) Le plastifiant plastifie le PVC.

120) Le désherbant désherbe la mauvaise herbe.

Tableau 3.7 Résultat des tests pour les noms des Substances chimiques

Lexèmes	Test a)	Test b)	Test c)	Test d)	**Score**
DURCISSEUR	+	+	+	+	4
DENTIFRICE	+	+	+	+	4
EPAISSISSANT	+	+	+	+	4
PLASTIFIANT	+	+	+	+	4
DESHERBANT	+	+	+	+	4

En bref, selon le tableau 3.11, les substances chimiques (DURCISSEUR, DENTIFRICE, EPAISSISSANT, PLASTIFIANT, DESHERBANT) montrent une instrumentalité frappante.

En conclusion, à travers les tests proposés par Namer et Villoing (2008), nous avons validé six sous-classes ontologiques identifiées dans une perspective ontologique : instrument de captation, outil, machine, appareil, substance chimique et récipient. Ces sous-classes peuvent être considérées comme des sous-classes prototypiques des noms d'instruments, aussi bien d’un point de vue ontologique que linguistique. Dans la section suivante, nous proposerons une classification interne de ces six sous-classes, en mettant en évidence leurs caractéristiques prototypiques les plus saillantes, afin de mieux les différencier.

3.2.3 Classification des sous-classes ontologiques

Dans la sous-section précédente, nous avons validé, à l’aide de tests linguistiques, les sous-classes ontologiques prototypiques des noms d'instruments. Étant donné que les distinctions proposées entre ces sous-classes par le dictionnaire ne sont ni fiables ni facilement manipulables, cette section vise à identifier les caractéristiques les plus saillantes pour mieux les différencier. Nous examinerons ces distinctions et les illustrerons graphiquement afin de conclure cette partie.

Nous commencerons par la sous-classe des récipients, puisque celle-ci se distingue nettement des cinq autres, qui sont davantage similaires entre elles. Selon Polguère (2003), « un récipient est un type particulier de contenant destiné spécifiquement à une utilisation particulière, telle que le stockage ou le traitement d’un contenu ». D’un point

de vue technique, Balfet et al. (1983 : 7) divisent les récipients en deux sous-classes, ouverts et fermés, selon leurs caractéristiques dimensionnelles (diamètre maximal et d'ouverture). Ainsi, un récipient ouvert est caractérisé par l'absence de constriction, son diamètre maximal correspondant à son ouverture (*par ex.* ASSIETTE, ECUELLE, BOL). À l'inverse, un récipient fermé présente une ouverture d'un diamètre inférieur à son diamètre maximal (*par ex.* POT, BOUTEILLE, JARRE). En bref, les récipients sont des artefacts simples et creux, distincts des autres catégories ontologiques. La caractéristique la plus saillante est le caractère **« creux ».**

Continuons avec la sous-classe des substances chimiques. Selon Pustejovsky & Batiukova (2019 : 197), une substance chimique remplit une fonction matérielle, c'est-à-dire qu'elle agit en vertu de ses propriétés chimiques. Bien qu'elle soit un artefact manipulé par un agent en vue de produire un effet, l'énergie nécessaire à cette action provient d'une réaction chimique interne, et non directement de l'agent. Ainsi, la caractéristique la plus saillante de cette sous-classe est le caractère « action chimique ».

Il nous reste quatre autres sous-classes ontologiques : instrument de captation, outil, machine, appareil. Précisément, ils peuvent être regroupés en deux paires selon la perspective philosophique et technologique proposée par Rabardel (1995) et Vial (2012) : « instrument de captation/outil » et « machine/appareil ». Rabardel (1995 : 63) distingue ces paires en fonction de leur rôle et de leurs composantes. Les machines et appareils sont caractérisés par leurs divers organes et modalités d'action sur la matière. Vial (2012) insiste quant à lui sur le système technique, qui permet de différencier les machines (caractérisées par une mécanisation issue de la révolution industrielle) des appareils, associés à une technicité moderne impliquant l'électricité et l'informatique.

La première paire, « instrument de captation/outil », est précisée par Simondon (2001 : 114). L'instrument de captation est un artefact qui prolonge et adapte les sens humains pour améliorer la perception (*par ex.* LUNETTES, LOUPE, RAPPORTEUR). À l'inverse, l'outil est un objet fabriqué pour agir sur la matière, pour accomplir une action (*par ex.* RASOIR, CISEAUX, COUTEAU). L'outil se caractérise par sa « maniabilité pour effectuer une action », tandis que l'instrument de captation se distingue par « sa capacité à révéler des informations et améliorer la perception ».

Concernant la seconde paire, « machine/appareil », la distinction repose principalement sur la source d'énergie et le type de fonctionnement. D'après Canguilhem (1993 : 126-127), une machine est un artefact dont la fonction dépend d'un mécanisme transformant une énergie externe (animaux, eau, vent, gaz, électricité). Par

ailleurs, Vial (2016) précise que les appareils relèvent de l'ontophanie numérique, caractérisée par le calcul rapide, la fluidité des procédures et l'intégration dans des interfaces informatiques. Les appareils se distinguent par leur association avec la perception, l'information et la conduction d'énergie (*par ex.* SCANNEUR, ECOUTEUR, CHARGEUR), alors que les machines privilégient la mécanisation (*par ex.* TONDEUSE, LAVE-LINGE, SECHE-CHEVEUX). En résumé, bien que machines et appareils soient similaires, les caractéristiques « perception et conduction d'énergie » et « mécanisation » permettent de les différencier nettement.

Enfin, selon la source d'énergie, ces sous-classes peuvent être divisées en deux groupes présentés dans la figure 3.6. Ce schéma présente une classification des noms d'instruments fondée sur la source d'énergie mobilisée pour leur fonctionnement. Deux groupes sont distingués : d'une part, les instruments fonctionnant à l'énergie de l'agent, c'est-à-dire manipulés directement par un être humain ; d'autre part, ceux qui dépendent d'une énergie extérieure, telle qu'une force naturelle, électrique ou chimique. Dans le premier cas, nous distinguons notamment les récipients, caractérisés par leur creux, les outils, définis comme des artéfacts maniables pour effectuer une action, ainsi que les instruments de captation, destinés à faciliter la perception ou la collecte d'informations. Dans le second cas, la catégorie des machines correspond à la mécanisation d'actions humaines ou naturelles, les appareils remplissent des fonctions similaires aux instruments de captation mais via des procédés automatisés et la conduction de l'énergie, et les substances chimiques sont définies par leur capacité à produire une action par réaction interne. Ces caractéristiques prototypiques saillantes constituent ainsi des critères déterminants pour distinguer les différentes sous-classes ontologiques, et elles forment également la base théorique sur laquelle repose notre travail d'annotation.

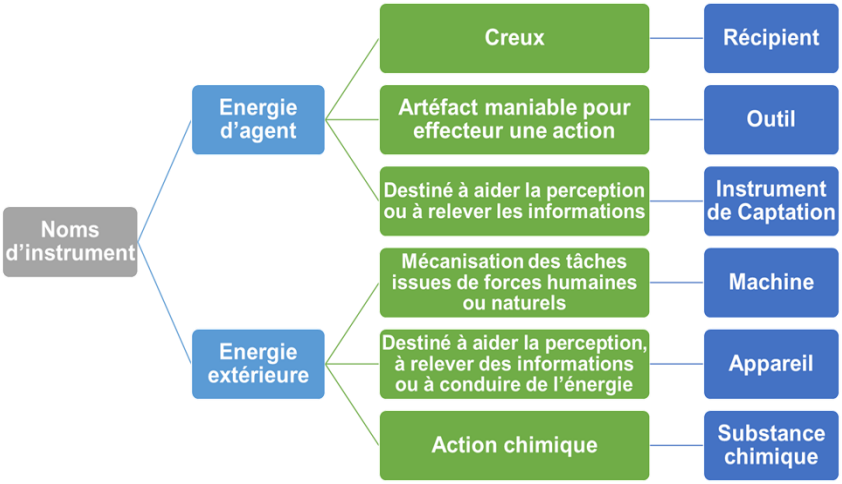

Figure 3.6 Arbre de classification ontologique des noms d'instruments

Chapitre 4 Méthodologies

Ce chapitre présente en détail les méthodologies adoptées dans ce travail. Les chapitres précédents ont permis de poser les bases théoriques de l'étude : une nouvelle définition des noms d'instrument a été proposée dans une perspective ontologique (chapitres 1 et 3) ; les contraintes linguistiques influençant le choix des suffixes ont été identifiées en nous appuyant sur des travaux antérieurs (chapitres 1 et 2) ; enfin, une revue critique des outils statistiques mobilisés dans les recherches existantes a été menée (chapitre 2). Sur la base des hypothèses avancées par plusieurs auteurs (Roché 1998 ; Plénat 2005 ; Bonami & Thuilier 2019, entre autres), nous avons identifié un ensemble de variables potentiellement déterminantes pour le choix des suffixes dans la construction des noms d'instruments en français. Il s'agit notamment de :

(1) la taille phonologique du radical de la base du dérivé

(2) le segment phonologique final du radical de la base du dérivé

(3) la structure morphologique de la base du dérivé

(4) la classe flexionnelle de la base verbale du dérivé

(5) le thème de la base du dérivé

(6) la catégorie syntaxique de la base du dérivé

(7) le profil sémantique assumé par la base dans la représentation sémantique du dérivé

(8) les sous-classes ontologiques du dérivé

(9) la datation des noms construits

Ces propriétés constituent le fondement des analyses théoriques et empiriques menées dans cette recherche : elles guident la constitution de la base de données, orientent l'annotation linguistique des lexèmes, et servent de variables explicatives dans les analyses et les modélisations statistiques. Ce chapitre est structuré en trois parties principales. Nous présentons d'abord la stratégie de constitution de la base de données. Ensuite, nous détaillons les théories et les propriétés linguistiques mobilisées pour l'annotation. Enfin, nous décrivons les outils statistiques utilisés pour l'analyse et la modélisation.

4.1 Méthodologie de constitution de la base des données

La présente recherche s'appuie sur une stratégie d'extraction lexicographique fondée sur l'exploitation des dictionnaires électroniques. Cette approche, mise en œuvre de manière manuelle ou à l'aide d'outils informatiques, permet de recueillir des données structurées à partir des entrées lexicographiques, incluant des informations phonologiques, morphologiques, syntaxiques, sémantiques et étymologiques. Ces types d'informations sont fréquemment mobilisés dans les études morphologiques, ainsi que dans la constitution de bases de données linguistiques. Concrètement, en phonologie, cette méthode est fréquemment employée pour mener des recherches phonotactiques (Frisch 2012 : 458-459). En morphologie, plusieurs études (Anshen & Aronoff 1999 ; Namer & Villoing 2008 ; Hilpert 2013 ; Arndt-Lappe 2014) ont recours à cette stratégie afin de renforcer leurs hypothèses théoriques à l'aide de données empiriques (Lieber 2014 : 87). A titre d'exemple, Anshen & Aronoff (1999) exploitent l'*Oxford English Dictionary* pour étudier la productivité diachronique des suffixes *-ity* et *-ment,* en s'appuyant sur les dates d'attestation et les informations étymologiques fournies.

Pour réaliser leurs études morphologiques, les auteurs mentionnés ont extrait manuellement des informations limitées à partir du dictionnaire électronique, en utilisant la fonction de recherche avancée proposée par l'interface du logiciel. Plus précisément, dans les recherches citées, Anshen et Aronoff (1999) ainsi que Hilpert (2013) se concentrent sur les informations étymologiques, tandis que Namer & Villoing (2008) s'appuient principalement sur les définitions pour l'analyse sémantique. Arndt-Lappe (2014) privilégie quant à elle les données phonologiques et syntaxiques pour prédire le choix des suffixes *-ity*/*-ness*, en s'appuyant sur *Oxford English Dictionary*.

Cependant, comme le souligne Plag (2018 : 172), les procédés morphologiques sont sensibles à une diversité de propriétés linguistiques, incluant des niveaux phonologiques, morphologiques, syntaxiques et sémantiques. Pour analyser les schémas morphologiques de manière plus approfondie, il apparaît donc nécessaire de prendre en compte ces différents niveaux linguistiques. Lorsque plusieurs types d'informations doivent être croisées, la collecte manuelle devient particulièrement laborieuse. La fonction de recherche avancée fournie par le logiciel reste en effet limitée : bien qu'elle permette de filtrer les entrées selon certains critères, elle ne facilite pas l'extraction d'informations plus spécifiques, ni la récupération systématique des

données lexicales. Dans cette optique, il convient de proposer une méthode plus efficace et adaptée pour optimiser le processus d'extraction et d'annotation des données dans le cadre de notre recherche.

4.1.1 Constitution des listes des entrées lexicales retenues

Nous présentons ici la méthode d'extraction des données, organisée en plusieurs étapes successives. La première consiste à rassembler les données brutes à partir du dictionnaire électronique *Le Grand Robert* de la langue française, en mobilisant deux versions complémentaires : l'édition 2023 (licence étudiant) et l'édition 2005 (licence étudiant). La version 2023 est utilisée pour effectuer des recherches ciblées via la fonction avancée de l'interface, ce qui permet d'obtenir des listes initiales de lexèmes potentiels. Toutefois, pour l'extraction systématique des articles complets nécessaires à la constitution de notre base de données, nous avons privilégié la version 2005. Celle-ci offre un accès direct aux fichiers HTML des entrées, dont la structure balisée permet d'identifier automatiquement les composantes internes des articles lexicographiques. Cette organisation en balises constitue un atout essentiel pour la mise en place d'un système d'annotation cohérent et reproductible au sein du corpus (voir Annexe 2 pour le détail des balises utilisées).

Bien que cette version soit légèrement postérieure, elle reprend l'ensemble des données de l'édition papier de 2001 tout en les rendant exploitables sous format numérique, ce qui combine la fiabilité documentaire de l'édition papier de 2001 avec les avantages techniques d'un format HTML structuré, compatible avec le traitement automatique des données. Avant toute exploitation, les données issues des deux versions doivent être harmonisées afin de constituer un ensemble cohérent. Les résultats issus des deux versions font ensuite l'objet d'un processus de dédoublonnage. Celui-ci permet d'identifier les entrées communes, d'éliminer les doublons, et de constituer une liste unifiée d'articles lexicaux, propre à être exploitée dans les étapes ultérieures d'annotation et d'analyse.

Par exemple, nous appliquons cette stratégie de recherche aux noms d'instruments dérivés par le suffixe *-oir*. Dans la version 2023 du *Grand Robert*, la fonction de recherche avancée permet d'extraire manuellement une liste d'entrées se terminant par *-oir*, incluant les noms potentiellement construits par suffixation. En combinant des filtres portant sur le graphème final et la catégorie grammaticale (Nom).

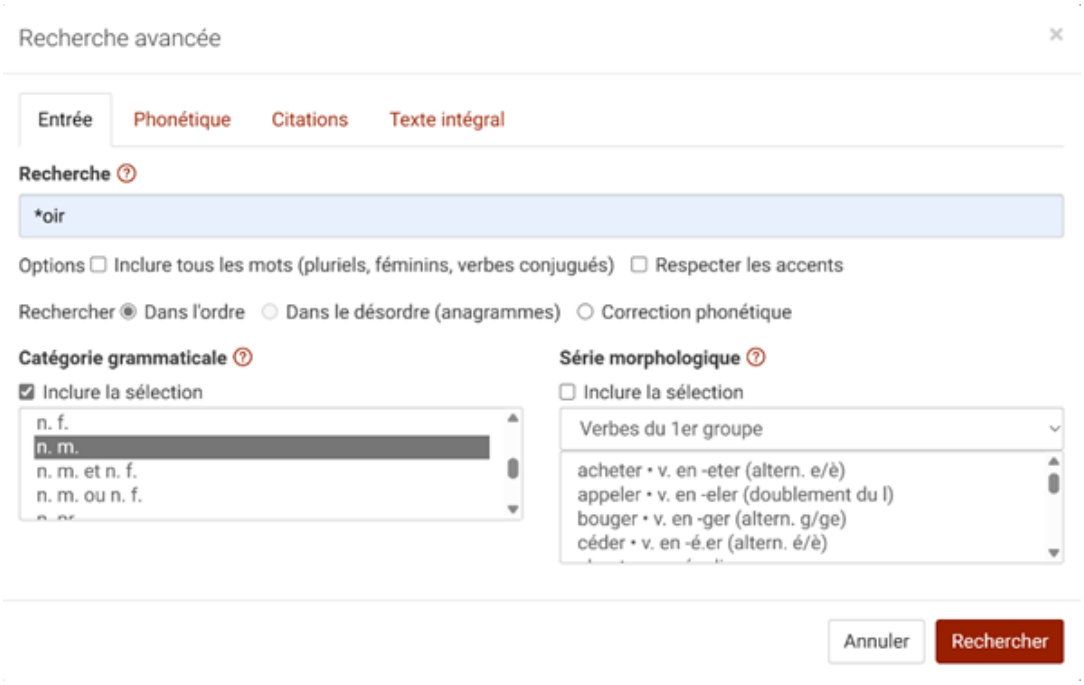

Figure 4.1 Recherche avancée dans *Le Grand Robert* 2023

Par la suite, nous compilons la liste des résultats obtenus et l'enregistrons dans un tableur en vue d'une analyse qualitative et morphologique, visant à identifier les noms d'instruments construits par suffixation en *-oir*. Il convient de préciser que toutes les entrées extraites ne correspondent pas à des noms d'instrument dérivés (*par ex.* NOIR, SOIR, VOULOIR). Ces entrées doivent donc être filtrées afin de ne retenir que les noms d'instruments pertinents. Nous avons d'abord repéré les lexèmes dont la définition comporte l'une des étiquettes ontologiques sélectionnées, puis nous avons vérifié manuellement les lexèmes ainsi filtrés, à l'aide de la définition des noms d'instruments prototypiques que nous proposons, afin d'identifier les noms d'instruments prototypiques construits morphologiquement par suffixation *-oir* (voir la Section 3.2), A l'issue de ces analyses, nous avons identifié un total de 206 entrées en *-oir*.

Pour les autres suffixes étudiés, les mêmes étapes ont été appliquées, depuis l'extraction initiale jusqu'à la validation des lexèmes dérivés. Ce traitement uniforme nous a permis d'identifier l'ensemble des noms d'instruments pertinents associés à chaque suffixe. Il convient de noter que certaines formes masculines et féminines apparaissent sous une seule entrée, généralement désignée par la forme masculine (*par ex.* LAVEUR ~ LAVEUSE). Des formes telles que LAVEUSE ne sont pas considérées ici comme de simples variantes de LAVEUR, mais comme des dérivés formés avec un suffixe distinct (*-euse*), relevant d'un suffixe autonome (cf. Chapitre 2). À ce titre, elles sont enregistrées séparément dans notre base de données, en tant que lexèmes

indépendants.

Le tableau 4.1 présente la synthèse quantitative de ces résultats, en distinguant le nombre d'entrées initiales et celui des lexèmes retenus pour chaque suffixe analysé.

Tableau 4.1 Nombre d'entrée totales et sélectionnées pour les suffixes étudiés

Suffixes	Nombre de lexèmes totaux	Nombre de lexèmes retenus
-eur	2401	455
-euse	1888	311
-ier	1249	71
-ière	915	63
-oir	363	206
-oire	135	27
-rice	751	11
-ant	1800	72
-ante	1713	3
-et	649	37
-ette	728	57
-on	1307	20
-ail	34	3
-aille	123	2

4.1.2 Constitution de la base de données

A partir de la liste ainsi constituée des noms d'instruments, nous avons procédé à l'extraction des articles correspondants dans la version 2005 du *Grand Robert*, en utilisant le mode « fichier » de l'interface du logiciel. La fonction « Exporter l'article » a permis de récupérer les données au format HTML, ensuite converties en fichiers texte (.txt). Ce format HTML, structuré par des balises imbriquées, permet de repérer automatiquement les segments pertinents et de les convertir en balises linguistiques adaptées à notre système d'annotation.

Dans un second temps, nous avons retravaillé les annotations métalexicographiques en définissant un ensemble de balises linguistiques pertinentes pour notre analyse. Les éléments superflus ont été supprimés, tandis que les balises correspondant aux cinq types d'informations linguistiques jugées pertinentes pour l'analyse ont été conservées. Cette étape a conduit à l'établissement de cinq catégories principales de balises (cf. Annexe 2).

Chaque balise est construite selon un système d'ancrage et de délimitation, garantissant une extraction précise des informations sans chevauchement entre

catégories (cf. Annexe 2). Une fois le balisage appliqué à l'ensemble du corpus, les données nécessaires à la constitution de notre base ont pu être extraites. Cette opération a été automatisée à l'aide du logiciel PowerGrep (*Just Great Software*, version 4), par le biais d'expressions régulières, permettant une extraction ciblée et efficace des informations pertinentes.

4.1.3 Organisation de la base de données

Bien que certaines informations linguistiques extraites puissent être intégrées directement dans la base de données (comme les données phonologiques et syntaxiques), d'autres nécessitent un traitement préalable, notamment les informations étymologiques et sémantiques. Il est donc nécessaire de réorganiser les données avant de procéder à l'annotation dans le tableur.

Par exemple, les informations étymologiques contiennent à la fois la date de première attestation du lexème et les procédés morphologiques qui lui sont associés. Ces deux éléments doivent être distingués, car ils relèvent de plans d'analyse différents. Sur cette base, nous avons extrait les propriétés linguistiques suivantes : « Base du dérivé », « Structure morphologique de la base du dérivé », « Classe flexionnelle de la base verbale » et « Date de la première attestation ». Ces données ont ensuite été organisées dans des colonnes distinctes du tableur. Le tableau ainsi structuré comprend l'ensemble des informations linguistiques suivantes (voir le tableau 4.2) :

- le lexème ;
- trois informations phonologiques : la forme phonologique du dérivé, la taille du radical en syllabe de la base du dérivé, et le segment phonologique final du radical ;
- quatre informations morphologiques : la base morphologique du dérivé, la structure morphologique de la base, le thème de la base, et la classe flexionnelle de la base verbale ;
- une information syntaxique : la catégorie syntaxique de la base du dérivé ;
- deux informations sémantiques : les sous-classes ontologiques du dérivé et le profil sémantique assumés par la base dans la représentation du dérivé ;
- une information étymologique : le siècle de la première attestation pour le sens instrumental.

Plus précisément, certaines de ces informations telles que la forme phonologique et la date d'attestation sont directement extraites du dictionnaire. En revanche, les autres

résultent d'un travail d'annotation réalisé à partir des données brutes.

Tableau 4.2 Exemple de la base de données annotées

Lexème	PHON	PHON_1_taille de rad. en syllabe	PHON_2_Segment final phono de rad.	MOR_base morphologique	MOR_1_structure morphologique de la base	MOR_2_Espace thématique	MOR_3_Classe flexionelle	SYN	SEM ONTO	Suffixe	ETYM	Profils sémantiques
ableret	ablaʁɛ	2	ʁ	able	Nom simple	N2	-	N	Outil	et	14	Patient
ablier	ablije	1	l	able	Nom simple	N2	-	N	Outil	ier	14	Patient
abluant	ablyɑ̃	2	y	ablue	verbe simple	V1	G1	V	Substance chimique	ant	19	Verbe Télique
absorbeur	apsɔʁbœʁ	2	b	absorber	verbe simple	V1	G1	V	Machine	eur	20	Verbe Télique
absorbeur	apsɔʁbœʁ	2	b	absorber	verbe simple	V1	G1	V	Substance chimique	eur	20	Verbe Télique
accélérateur	akselekatœʁ	4	t	accélérer	verbe simple	V13	G1	V	Machine	eur	20	Verbe Télique
accélérateur	akselekatœʁ	4	t	accélérer	verbe simple	V13	G1	V	Substance chimique	eur	19	Verbe Télique
accordeur	akɔʁdœʁ	2	d	accorder	verbe simple	V1	G1	V	Outil	eur	19	Verbe Télique
accordoir	akɔʁdwaʁ	2	d	accorder	conversion N/V	V1	G1	V	Outil	oir	17	Verbe Télique
accumulateur	akymylatœʁ	4	t	accumuler	verbe simple	V13	G1	V	Machine	eur	19	Verbe Télique
activateur	aktivatœʁ	3	t	activer	conversion A/V	V13	G1	V	Outil	eur	20	Verbe Télique
activateur	aktivatœʁ	3	t	activer	conversion A/V	V13	G1	V	Récipient	eur	20	Verbe Télique
activateur	aktivatœʁ	3	t	activer	conversion A/V	V13	G1	V	Substance chimique	eur	20	Verbe Télique
activeur	aktivœʁ	2	v	activer	conversion A/V	V1	G1	V	Substance chimique	eur	20	Verbe Télique
additionneuse	adisjɔnøz	3	n	additionner	conversion N/V	V1	G1	V	Appareil	euse	19	Verbe Télique
adoucissant	adusisɑ̃	3	s	adoucir	conversion A/V	V1	G2	V	Substance chimique	ant	17	Verbe Télique
adoucisseur	adusisœʁ	3	s	adoucir	préfixation a- V<A	V1	G2	V	Machine	eur	20	Verbe Télique
aérateur	aeʁatœʁ	3	t	aérer	verbe simple	V13	G1	V	Outil	eur	19	Verbe Télique
affiloir	afilwaʁ	2	l	affiler	préfixation a- V<N	V1	G1	V	Outil	oir	19	Verbe Télique
affiloire	afilwaʁ	2	s	affiler	préfixation a- V<N	V1	G1	V	Outil	oire	17	Verbe Télique
affinoir	afinwaʁ	2	n	affiner	préfixation a- V<A	V1	G1	V	Outil	oir	17	Verbe Télique
affronteur	afʁɔ̃tœʁ	2	t	affronter	préfixation a- V<N	V1	G1	V	Outil	eur	16	Verbe Télique
affûteuse	afytøz	2	t	affûter	préfixation a- V<N	V1	G1	V	Machine	euse	20	Verbe Télique
affûtoir	afytwaʁ	2	t	affûter	préfixation a- V<N	V1	G1	V	Outil	oir	19	Verbe Télique
agglomérant	aglɔmeʁɑ̃	3	ʁ	agglomérer	verbe simple	V1	G1	V	Substance chimique	ant	19	Verbe Télique
agitateur	aʒitatœʁ	3	t	agiter	verbe simple	V13	G1	V	Machine	eur	19	Verbe Télique
agrafeuse	agʁaføz	2	f	agrafer	conversion N/V	V1	G1	V	Outil	euse	20	Verbe Télique
agrandisseur	agʁɑ̃disœʁ	3	s	agrandir	préfixation a- V<A	V1	G2	V	Appareil	eur	19	Verbe Télique
aiguisoir	egizwaʁ	2	z	aiguiser	verbe simple	V1	G1	V	Outil	oir	15	Verbe Télique
alcoylant	alkɥilɑ̃	3	l	alcoyler	conversion N/V	V1	G1	V	Substance chimique	ant	20	Verbe Télique
aléseur	alezœʁ	2	z	aléser	verbe simple	V1	G1	V	Outil	eur	20	Verbe Télique
aléseuse	alezøz	2	z	aléser	verbe simple	V1	G1	V	Machine	euse	20	Verbe Télique
alésoir	alezwaʁ	2	z	aléser	verbe simple	V1	G1	V	Outil	oir	17	Verbe Télique

4.2 Méthodologie d'annotation de la base de données

Dans la section précédente, nous avons présenté la méthodologie de constitution de la base de données, ainsi que l'ensemble des noms d'instruments recensés en français par le dictionnaire *Le Grand Robert* et les propriétés linguistiques à annoter. La présente section est consacrée à la méthodologie d'annotation de cette base de données. Nous exposerons les théories mobilisées pour l'annotation des données, en fonction des différents niveaux linguistiques retenus. L'annotation repose sur quatre niveaux d'analyse distincts : le niveau phonologique (§4.2.1), le niveau morphologique (§4.2.2), le niveau syntaxique (§4.2.3) et le niveau sémantique (§4.2.4).

4.2.1 Au niveau phonologique

Comme indiqué au chapitre 1, la phonologie joue un rôle important dans la formation des noms (Roché 1998 ; Plénat 2005). Dans le cadre de cette thèse, deux propriétés phonologiques ont été retenues pour l'annotation :

(1) la taille du radical de la base du dérivé, en nombre de syllabes (Roché 1998 ; Plénat 2005, 2008) ;

(2) le segment phonologique final de la base du dérivé (Koehl 2012 ; Bonami & Thuilier 2019).

4.2.1.1 Taille du radical de la base du dérivé

La taille du radical de la base du dérivé correspond au nombre de syllabes constituant ce radical. Pour cette annotation, nous avons suivi la stratégie proposée par Koehl (2012) et Bonami & Thuilier (2019). Le comptage a été effectué automatiquement dans le tableur Excel (voir Annexe 3), en prenant le nombre de voyelles (de la transcription phonétique) comme indicateur du nombre de syllabes, chaque voyelle étant considérée comme une syllabe. Par exemple, le lexème REPAS /ʀəpɑ/ comporte deux voyelles (/ə/ et /ɑ/), ce qui correspond à deux syllabes.

4.2.1.2 Segment phonologique final de la base du dérivé

Le segment phonologique final du radical de la base du dérivé est défini comme le dernier phonème du radical, c'est-à-dire du lexème dérivé privé de son suffixe. Ce

principe s'applique également aux cas de radical d'origine latine (Bonami, Boyé & Kerleroux 2009 ; Grevisse & Goosse 2016). Par exemple, dans RASOIR /ʀazwaʀ/, le radical est /ʀaz/ et le suffixe /waʀ/ ; le segment final est donc /z/. De même, dans AERATEUR /aeʀatœʀ/, le radical est /aeʀat/ et le suffixe /œʀ/ ; le segment final est /t/.

Pour l'annotation de cette propriété, nous nous appuyons principalement sur la méthode décrite dans Koehl (2012), en complétant, pour les cas de radicaux latins, par les principes issus de Bonami, Boyé & Kerleroux (2009), conformément à la morphologie lexématique (Aronoff 1994).

4.2.2 Au niveau morphologique

Au niveau morphologique, il existe trois propriétés linguistiques à explorer. Tout d'abord, les recherches mentionnées dans les chapitres précédents ont mis en évidence la structure morphologique de la base du dérivé (Koehl 2012 ; Bonami & Thuilier 2019 ; Missud & Villoing 2020) ainsi que la classe flexionnelle de la base verbale du dérivé (Lapraye 2017), identifiées comme des propriétés morphologiques permettant soit de prédire le choix des suffixes, soit d'effectuer une analyse qualitative. Ensuite, selon Bonami, Boyé & Kerleroux (2009), certains thèmes présentent une affinité particulière avec certains suffixes. Par exemple, comparé au suffixe *-euse*, le suffixe *-rice* tend à sélectionner des radicaux latins. Ainsi, la sélection d'un des thèmes du verbe constitue une troisième propriété morphologique à étudier.

4.2.2.1 Structure morphologique de la base du dérivé

La structure morphologique de la base du dérivé renvoie à la structure interne de cette base, qu'elle soit un lexème simple ou formé par des procédés morphologiques tels que la préfixation, la suffixation, la conversion, la composition, etc. Pour annoter ces informations, nous nous sommes appuyés sur plusieurs travaux de morphologues (Corbin 1987 ; Namer 2009 ; Tribout 2010), ainsi que sur des ouvrages de morphologie et de grammaire (Huot 2005 ; Grevisse & Goosse 2016) (voir Annexe Guide d'annotation pour plus de détails). Le choix de ces références repose à la fois sur leur traitement systématique des schémas morphologiques concernés et sur leur prise en compte de structures moins étudiées.

Prenons l'exemple du lexème AGRANDISSEUR. Notre analyse morphologique montre que sa base est le verbe AGRANDIR, lequel entretient un lien étroit avec l'adjectif GRAND dans le cadre de la morphologie lexématique. Selon Namer (2009 : 208), le

schéma morphologique de la préfixation *a*-A > V correspond au sens « rendre (plus) A », où A est la base adjectivale. La base verbale d'AGRANDISSEUR est donc annotée comme issue du schéma *a*-A > V. Cet exemple met en évidence l'intérêt de mobiliser des références prenant en compte des schémas morphologiques moins étudiés. À la différence de la préfixation en ***a-***, peu de travaux de référence sont menés pour *-oyer*, donc nous utilisons les informations concernant ce suffixe contenues dans le *TLFi*.

4.2.2.2 Classe flexionnelle de la base verbale

La classe flexionnelle de la base verbale du dérivé exerce aussi des influences sur la formation des mots (Anscombre 2015). Pour annoter cette propriété, nous nous appuyons sur le *Littré Conjugaison* (Eluerd 2009), en tant que référence pour l'identification de la classe flexionnelle des bases verbales. Par exemple, pour le lexème DECLINATOIRE, la base verbale est DECLINER, qui appartient, selon Eluerd (2009), au premier groupe de conjugaison. Pour TONDEUSE, la base verbale TONDRE est annotée comme un verbe du troisième groupe.

4.2.2.3 Espaces thématiques

Selon Bonami (2014), les thèmes en français se distinguent selon la catégorie syntaxique du lexème : les thèmes verbaux et les thèmes nominaux et adjectivaux. Les travaux de Villoing (2002), Bonami, Boyé et Kerleroux (2009), Tribout (2010) et Koehl (2012) ont montré que certains suffixes sélectionnent certains thèmes particuliers. Pour l'annotation des données dans notre base de données, nous avons ainsi utilisé les thèmes verbaux décrits par Bonami (2014), les thèmes adjectivaux présentés par Koehl (2012), qui reprend les descriptions proposées par Bonami & Boyé (2005) et Bonami et al. (2009), ainsi que les thèmes nominaux décrits par Tribout (2010), reprenant Plénat (2008).

En ce qui concerne les thèmes verbaux, deux versions coexistent, numérotées différemment selon les auteurs : Bonami et Boyé (2003), Bonami, Boyé & Kerleroux (2009) et Bonami (2014). La version de Bonami & Boyé (2003), élaborée comme un premier cadrage descriptif, propose un inventaire de thèmes verbaux numérotés selon leurs comportements morphologiques, et a servi de référence dans plusieurs études ultérieures en morphologie constructionnelle (Fradin & Winterstein 2013 ; Strnadová 2014). Dans Bonami, Boyé & Kerleroux (2009), cet inventaire de thèmes a été développé par l'ajout d'un treizième thème, destiné à rendre compte de formes

construites sur un radical d'origine latine (issu du latin supin). De surcroît, Tribout (2010 : 199) a introduit le thème 0 pour les verbes du deuxième groupe, défini comme un radical dépourvu de la voyelle /i/ ou de la séquence /i/- /is/ caractéristique de ce groupe. Il est obtenu par la suppression du /i/ final du thème 3 et n'apparaît qu'en dérivation, afin de maintenir l'identité formelle entre le nom et le verbe (*par ex.* FLEUR / FLEURIR.). Par rapport aux versions antérieures sur lesquelles elle s'appuie, Bonami (2014) en propose une synthèse actualisée et systématique, plus cohérente et directement exploitable, que nous retenons ici pour l'annotation des thèmes verbaux dans notre base de données.

La Figure 4.2, reprise de Bonami (2014 : 51), répertorie les différents thèmes verbaux, à partir des formes fléchies uniquement, numérotés de 1 à 12. Pour illustrer cette classification, prenons l'exemple du verbe SORTIR : le thème 1 correspond à la forme *sort-* /sɔʀt/, utilisée aux première et deuxième personnes du pluriel au présent de l'indicatif, tandis que le thème 3 correspond à *sor-* /sɔʀ/, utilisé au singulier du présent de l'indicatif. Hormis les douze thèmes évoqués, le verbe PRESENTER comprend un autre thème, dit "caché", fondé sur un radical latin non représenté dans la conjugaison, *présentat*, que l'on retrouve dans des lexèmes comme PRESENTATEUR, PRESENTATION ou PRESENTATIF.

Tableau 4.3 Distribution des radicaux dans les paradigmes pour 8 verbes français (tiré et adapté de Bonami (2014 : 51) et inspiré de Tribout (2010))

N°	Utilisation	aimer	finir	rendre	sortir	prendre	savoir	conclure	aller
1	imp, prs 1/2 pl.	ɛm	finis	ʁɑ̃d	sɔʁt	pʁən	sav	kɔ̃kly	al
2	présent 3pl.	ɛm	finis	ʁɑ̃d	sɔʁt	pʁɛn	sav	kɔ̃kly	vɔ̃
3	présent sg.	ɛm	fini	ʁɑ̃	sɔʁ	pʁɑ̃	sɛ	kɔ̃kly	vɛ
4	participe prs	ɛm	fini	ʁɑ̃d	sɔʁt	pʁən	saʃ	kɔ̃kly	al
5	impératif 2sg.	ɛm	fini	ʁɑ̃	sɔʁ	pʁɑ̃	saʃ	kɔ̃kly	va
6	impératif pl.	ɛm	finis	ʁɑ̃d	sɔʁt	pʁən	saʃ	kɔ̃kly	al
7	subj. sg. 3pl.	ɛm	finis	ʁɑ̃d	sɔʁt	pʁɛn	saʃ	kɔ̃kly	aj
8	subj. 1/2 pl.	ɛm	finis	ʁɑ̃d	sɔʁt	pʁən	saʃ	kɔ̃kly	al
9	infinitif	ɛme	fini	ʁɑ̃d	sɔʁti	pʁɑ̃d	savwa	kɔ̃kly	ale
10	futur, cond.	ɛmə	fini	ʁɑ̃d	sɔʁti	pʁɑ̃d	so	kɔ̃kly	i
11	passé simple	ɛma	fini	ʁɑ̃di	sɔʁti	pʁi	sy	kɔ̃kly	ala
12	participe passé	ɛme	fini	ʁɑ̃dy	sɔʁti	pʁiz	sy	kɔ̃kly	ale

Après avoir abordé les thèmes verbaux, nous nous penchons désormais sur les thèmes nominaux et adjectivaux. Selon Boyé & Bonami (2005) repris par Koehl (2012 :

177), il existe cinq thèmes adjectivaux : le thème 1 (forme utilisée pour le masculin), le thème 2 (forme utilisée pour le féminin), le thème 3 (forme utilisée pour la liaison), le thème 4 (forme employée pour la formation des adverbes) et le thème L (forme latine ou savante). Parmi ces thèmes, seuls le thème 2 et le thème L sont utilisés pour la formation des noms par suffixation (Koehl 2012 : 178). Par exemple, pour l'adjectif ROND, son thème 1 est /ʀɔ̃/, son thème 2 est /ʀɔ̃d/, et son thème L est *rotond-* (Koehl 2012 : 178). Le nom dérivé ROTONDITE se forme en ajoutant le suffixe *-ité* au thème L *rotond-*.

Enfin, en ce qui concerne les thèmes nominaux, selon Plénat (2008) repris par Tribout (2010), il existe trois thèmes : le thème 1, le thème 2 et le thème L. Concrètement, le thème 1 est utilisé pour construire la forme du singulier, tandis que le thème 2 est employé pour la dérivation. Le thème L, quant à lui, sert à la dérivation de lexèmes savants (Tribout 2010 : 216). Selon Tribout (2010 : 217), l'étude des dérivés dénominaux révèle différents types de relation entre les thèmes nominaux, et en particulier quatre types de relations entre le thème 1 et le thème 2 : identité (cf. tableau 4.4 a), ajout d'une consonne finale (cf. tableau 4.4 b), opposition vocalique nasale/orale (cf. tableau 4.4 c), et autres différences (cf. tableau 4.4 d). Le thème L, quant à lui, n'apparaît qu'en tant que forme supplétive du thème 2.

Tableau 4.4 Thèmes nominaux et leur utilisation dans la dérivation (adapté par Tribout 2010 : 218)

Cas	Lexème	Thème 1	Thème 2	Thème L	Dérivé à partir du thème 2	Dérivé à partir du thème L
a	ARBRE	aʀbʀ	aʀbʀ	aʀbɔʀ	aʀbʀɛ	aʀbɔʀize
b	POINT	pwɛ̃	pwɛ̃t	pɔ̃kty	pwɛ̃te	pɔ̃ktɥɛl
c	NOM	nɔ̃	nɔm	nɔmin	nɔme	nɔminal
d	NŒUD	nø	nu	nɔdyl	nue	nɔdylø

4.2.3 Au niveau syntaxique : catégorie syntaxique de la base du dérivé

Concernant les propriétés syntaxiques, certains travaux relevant de la linguistique générative ont proposé d'analyser les suffixes en fonction de la structure argumentale de la base verbale, comme les verbes inchoatifs, causatifs ou accusatifs (Grimshaw 1990). Notre approche, fondée sur la linguistique cognitive, ne reprend pas cette

classification, dans la mesure où la structure argumentale est généralement déterminée par des facteurs sémantiques dans notre cadre théorique (Goldberg 1995).

Par conséquent, la seule propriété syntaxique retenue ici est la catégorie syntaxique de la base du dérivé (verbe, nom, adjectif), telle qu'elle est fournie par *Le Grand Robert* (2005). Par exemple, dans le cas d'AGRANDISSEUR, la base AGRANDIR est un verbe ; dans CHIMIQUIER, la base CHIMIQUE est un adjectif ; et dans THEIERE, la base THE est un nom.

4.2.4 Au niveau sémantique

Enfin, sur le plan sémantique, les deux premiers chapitres de ce travail doctoral présentent divers types de propriétés sémantiques, notamment le lien entre la dénotation du dérivé et des propriétés sémantiques de la base, le type sémantique du verbe ainsi que le type ontologique de la base du dérivé.

Étant donné que cette thèse s'inscrit principalement dans le cadre de la linguistique cognitive et s'appuie également sur le *Lexique génératif* (Pustejovsky 1996) pour l'analyse des lexèmes concernés, ces trois types de propriétés sémantiques ne s'avèrent pas pleinement compatibles avec les cadres théoriques adoptés, lesquels privilégient la structure de Qualia, les rôles sémantiques dans le cas des bases nominales, ainsi que les sous-classes ontologiques des noms. Par conséquent, nous n'utilisons pas les trois propriétés mentionnées précédemment, et nous retenons d'autres propriétés sémantiques plus en adéquation avec notre cadre théorique. En particulier, nous annotons le profil sémantique assumé par la base dans la représentation du dérivé.

De plus, à la suite de l'hypothèse formulée par Dubois (1962), Spence (1990) et confirmée dans mon travail (Lin 2021), selon laquelle certaines sous-classes ontologiques des noms d'instrument tendent à privilégier des suffixes spécifiques, les sous-classes ontologiques des dérivés sont également prises en compte en tant que propriétés sémantiques dans notre annotation, afin de tester cette hypothèse.

4.2.4.1 Profil sémantique assumé par la base dans la représentation sémantique du dérivé

En ce qui concerne le profil sémantique assumé par la base dans la représentation sémantique du dérivé, cette thèse prend en compte trois types d'informations à annoter. Il s'agit des verbes associés à la structure de qualia dans le cas des noms déverbaux,

des rôles sémantiques remplis par la base nominale du dérivé, ainsi que des projections cognitives mobilisées lorsque la base n'est pas verbale. Il convient de préciser l'usage du terme *rôle sémantique* tel qu'il est entendu ici. Un terme proche, *rôle thématique*, est également employé dans certaines traditions théoriques. Toutefois, une distinction mérite d'être soulignée : « le rôle thématique est principalement employé en grammaire générative pour désigner la fonction sémantique des arguments verbaux généralement obligatoires, tandis que le rôle sémantique peut s'appliquer aux ajouts ainsi qu'aux participants absents d'un événement »[35] (Pustejovsky & Batiukova 2019 : 166). Puisque cette étude s'inscrit dans le cadre de la linguistique cognitive, c'est la notion de rôle sémantique qui est ici privilégiée. Selon Pustejovsky & Batiukova (2019 : 393), le rôle sémantique désigne la fonction spécifique assumée par chaque participant dans un événement, telle qu'elle est encodée dans la structure argumentale du prédicat (*par ex.* Agent, Patient, Thème, Expérienceur, Instrument, etc.). Les rôles sémantiques peuvent être attribués non seulement aux arguments, mais aussi aux adjoints.

4.2.4.1.1 Bases verbales correspondant à la structure de Qualia

Conformément au cadre théorique choisi (cf. section 3.1.3), nous adoptons pour l'annotation de la base verbale la structure de Qualia (Pustejovsky 1996 ; Pustejovsky & Batiukova 2019), qui intègre les connaissances encyclopédiques. Nous identifions quatre rôles associés aux verbes dans cette structure : le rôle télique, le rôle formel, le rôle constitutif et le rôle agentif (pour des informations détaillées, se référer au chapitre 3 et à l'annexe intitulée "Guide d'annotation"). Par exemple, le lexème AGRANDISSEUR est défini comme « appareil employé pour les agrandissements ». Sa base, AGRANDIR, correspond à un rôle télique, car elle désigne la fonction ou le but de l'appareil. Ainsi, cette base verbale est annotée en tant que rôle télique. En revanche, dans l'exemple SONNER > SONNETTE, SONNER est associé à un rôle formel dans la structure qualia du dérivé. En effet, SONNETTE est défini comme « petit instrument métallique qui sonne pour avertir ». Ici, la fonction de l'instrument est d'avertir, tandis que SONNER constitue une caractéristique de l'outil, le distinguant des autres instruments utilisés pour alerter quelqu'un.

[35] Le texte original : "The terms semantic role and thematic role are not to be confused. The latter is mainly used in generative grammar to refer to the semantic function of typically obligatory verbal arguments, while the former can be applied to adjuncts, as well as absent participants of an event."

4.2.4.1.2 Rôles sémantiques occupés par la base nominale du dérivé

Selon Rainer (2021), la majorité des noms d'instruments dérivés par suffixation sont des noms déverbaux. Cependant, il existe aussi des noms dénominaux qui peuvent être classés comme des noms d'instruments selon une perspective ontologique (voir les chapitres 1 et 3). Bien qu'ils ne possèdent pas de base verbale, ces noms dérivés conservent un verbe sous-jacent dans leur représentation sémantique. Dans le procès événementiel construit à partir de ce verbe, la base peut remplir un rôle sémantique dans sa structure argumentale. Selon notre cadre théorique (cf. section 3.1.3), ce rôle sémantique participe également au processus de dérivation. Ainsi, nous avons choisi d'utiliser les rôles sémantiques pour annoter la base nominale en fonction de son rôle sémantique dans le procès événementiel du verbe sous-jacent, afin d'examiner s'il existe des différences dans le profilage au cours du processus de dérivation. Pour illustrer notre protocole d'annotation, prenons l'exemple du nom d'instrument CAFETIERE. Ce dérivé est formé sur la base nominale CAFE, à laquelle s'ajoute le suffixe *-ière*. Bien que cette base ne soit pas verbale, sa définition implique un verbe sous-jacent, à savoir INFUSER. Dans le procès événementiel correspondant, CAFE assume le rôle sémantique de patient, c'est-à-dire qu'il constitue l'élément sur lequel s'exerce l'action d'infusion. Conformément à notre cadre théorique (cf. section 3.1.3), nous annotons donc la base CAFE avec le rôle sémantique patient.

Enfin, il est important de noter qu'il existe différentes versions des rôles sémantiques (Fillmore 1968 ; Schmid 2016 ; Pustejovsky & Batiukova 2019). Nous avons retenu la version proposée par Pustejovsky & Batiukova (2019), en raison de son inventaire plus exhaustif, tel que présenté dans le Tableau 4.3, par rapport aux autres versions.

Tableau 4.5 Inventaire de rôles sémantiques et leur définition (tiré de Pustejovsky & Batiukova (2019 : 23))

Rôle sémantique	Définition
Agent	Initiateur ou acteur animé et volitif (mais pas nécessairement intentionnel) d'une action.
Patient/Thème	Entité subissant l'action et affectée par elle (mise en existence, détruite, déplacée, modifiée, etc.)
Expérienceur	Entité psychologiquement ou émotionnellement affectée par l'événement.
Bénéficiaire	Entité qui bénéficie de l'événement.
Lieu	Endroit où l'événement se déroule.
Source	L'entité ou le lieu d'où un événement de mouvement a lieu.
But	L'entité ou le lieu vers lequel un événement de mouvement se dirige.
Instrument	L'entité utilisée pour effectuer une action.
Cause	Initiateur non volitif d'une action.
Mesure	Étendue temporelle ou spatiale des événements et des entités.
Possesseur	Entité possédant quelque chose.
Manière	La façon dont une action est réalisée.

4.2.4.1.3 Projections cognitives

Bien que la structure de Qualia (Pustejovsky 1996 ; Pustejovsky & Batiukova 2018) et le cadre des rôles sémantiques (Fillmore 1986, repris par Pustejovsky & Batiukova 2018) permettent d'analyser la majorité des lexèmes impliqués dans la formation des noms d'instrument, certaines bases nominales ou adjectivales ne semblent pas correspondre à un rôle sémantique identifiable. Afin d'illustrer ce cas de figure, nous présentons ci-dessous deux exemples représentatifs.

(1) LUNE > LUNETTE

(2) CHIMIQUE > CHIMIQUIER

En ce qui concerne leurs définitions, la LUNETTE est décrite dans le *TLFi* comme un « instrument d'optique, généralement de forme cylindrique, comportant le plus souvent un objectif et un oculaire, servant à augmenter le diamètre apparent des objets éloignés et à en permettre une observation plus nette et plus distincte », tandis que le CHIMIQUIER est défini dans *Le Grand Robert* (2005) comme un « cargo transportant des produits chimiques ». Étant donné que *Le Grand Robert* ne fournit pas de description détaillée pour LUNETTE, nous avons privilégié le *TLFi* afin d'appuyer notre analyse sur une définition plus précise. Dans les deux cas analysés ici, LUNETTE et CHIMIQUIER, la base utilisée pour la formation du dérivé ne semble pas occuper de rôle sémantique argumental par rapport à un événement. Au lieu de cela, la relation entre la base et le dérivé repose sur des mécanismes cognitifs de projection, respectivement de nature métaphorique et métonymique, qui conditionnent la sélection des bases morphologiques pour la suffixation, à partir de traits sémantiques saillants.

Concrètement, dans le cas de LUNETTE, la base est analysée comme étant liée au dérivé par une métaphore conceptuelle fondée sur une ressemblance perceptive. La forme circulaire et bombée des lentilles évoque celle de la lune, ce qui active une correspondance cognitive basée sur un trait de forme. Ce type de correspondance perceptive transfère une propriété visuelle d'un référent source (LUNE) vers un instrument de captation (LUNETTE). Ainsi, du point de vue de la relation entre base et dérivé, nous annotons LUNETTE comme un cas de métaphore.

Dans le cas de CHIMIQUIER, la base CHIMIQUE n'entretient pas de relation directe avec le verbe sous-jacent TRANSPORTER. C'est plutôt le rôle sémantique Cible de ce verbe, à savoir produit chimique, qui est activé dans la représentation sémantique. Par le biais d'une métonymie conceptuelle, ce syntagme nominal est désigné indirectement

par l'adjectif CHIMIQUE, qui en devient la forme saillante mobilisée dans la formation du dérivé. La métonymie fonctionne ainsi comme un mécanisme de médiation cognitive entre le rôle sémantique du verbe sous-jacent et la base dérivationnelle (cf. Section 3.1.3). Ainsi, du point de vue de la relation entre base et dérivé, nous annotons CHIMIQUIER comme un cas de métonymie.

4.2.4.2 Sous-classes ontologiques

Comme mentionné au chapitre 3, et à partir des auteurs évoqués, nous avons identifié un ensemble de sous-classes ontologiques prototypiques pour les noms d'instruments, tout en cherchant à clarifier les distinctions conceptuelles qui les séparent. Pour rappel, six sous-classes ontologiques ont été retenues : l'instrument de captation, l'outil, l'appareil, la machine, le récipient et la substance chimique. La justification de cette classification a été présentée dans Section 3.2, où chaque sous-classe fait l'objet d'une définition précise. Pour des précisions concernant la procédure d'annotation, nous renvoyons à l'annexe intitulée « Guide d'annotation ».

4.2.5 Au niveau diachronique

L'annotation sur le plan diachronique s'appuie principalement sur les informations fournies par *Le Grand Robert* (2005). Lorsque cela s'avère nécessaire, ces données sont vérifiées ou complétées à l'aide du *Trésor de la langue française informatisé* (TLFi). Par exemple, pour le lexème CHIMIQUIER, selon l'information fournie par *Le Grand Robert*, il est attesté comme suit : « Années 1990 ; de chimique dans (produits) chimiques, et *-ier*, d'après pétrolier, vraquier, etc. » En conséquence, nous l'annotons comme étant formé au XXe siècle. Toutefois, si aucun des deux dictionnaires ne fournit d'information pour un lexème donné, aucune annotation n'est faite. À titre d'exemple, pour le lexème MAGASINIER, bien que *Le Grand Robert* et le *TLFi* indiquent une formation au XVIIe siècle, cette datation concerne exclusivement le sens « personne qui garde les marchandises déposées dans un magasin ». En revanche, pour le sens instrumental, aucun des deux dictionnaires ne fournit d'information. Dans ce cas, nous n'annotons pas d'information de datation.

4.3 Analyses et Modélisations statistiques

Après avoir présenté les méthodologies de constitution et d'annotation de données, cette section est consacrée à la présentation des outils statistiques utilisés pour l'analyse des données et la modélisation. Plus précisément, quatre outils principaux sont mobilisés dans cette thèse :

– les tests statistiques (test du chi carré et test exact de Fisher)
– la forêt aléatoire
– l'outil de visualisation UpsetR
– le formule pour calculer le chevauchement des niches

4.3.1 Tests de Chi carré et Test de Fisher

Dans le cadre de cette thèse, les outils statistiques ont pour fonction de mettre à l'épreuve l'existence d'éventuelles associations significatives entre les suffixes instrumentaux et les propriétés linguistiques identifiées. Autrement dit, il s'agit d'examiner si certaines propriétés linguistiques des bases du dérivé, ou des sous-classes ontologiques des dérivés, sont statistiquement plus fréquentes parmi les mots construits avec un suffixe donné. Pour réaliser cet objectif, nous utilisons le test du chi-carré d'indépendance et le test de Fisher.

Nous commençons par le test du chi-carré d'indépendance. Il est utilisé pour évaluer la relation entre deux variables qualitatives. Cet outil statistique permet de déterminer s'il existe un lien significatif entre les modalités de deux variables, autrement dit si la répartition des propriétés linguistiques dépend du suffixe employé dans la formation des noms d'instruments. Il compare les fréquences observées à celles attendues sous l'hypothèse d'indépendance et juge si l'écart peut s'expliquer par le hasard (Gries 2021 : 172). Le test utilise une table de contingence pour croiser les valeurs de deux variables catégoriques et applique la formule suivante (Wang 2019 : 158) :

$$\chi^2 = \sum_{i=1}^{n} \frac{(valeur\ observée_i - valeur\ attendue_i)^2}{valeur\ attendue_{\mathrm{i}}}$$

Pour cet outil statistique, nous utilisons la fonction {chisq.test} du package R

{stats}. Cette fonction calcule la statistique du chi carré afin d'évaluer la significativité des résultats et fournit directement la valeur *p* (le seuil de signification est généralement fixé à 0,05). Concrètement, il s'agit d'un test d'hypothèse. L'hypothèse nulle (H_0), à rejeter au seuil de 0,05, stipule que les deux variables observées sont indépendantes et que toute association entre elles est due au hasard. En rejetant H_0, on estime qu'il est probable d'affirmer que les deux variables en jeu sont effectivement liées. Si la valeur *p* est inférieure à 0,05, on rejette H_0, ce qui indique une association significative. Si la valeur *p* est supérieure à 0,05, on ne rejette pas H_0, ce qui signifie qu'il n'y a pas de preuve suffisante d'une différence significative (Xu 2013 : 92 ; Gries 2021 : 173). En outre, le test indique également les degrés de liberté (*ddl*), c'est-à-dire le nombre de valeurs indépendantes qui peuvent varier dans le calcul statistique sans violer une contrainte donnée (Gries 2021 : 173).

Par ailleurs, lorsque les effectifs sont faibles (inférieurs à cinq), le test exact de Fisher est recommandé en vue d'évaluer la signification des liens entre les variables (Wang 2019 : 166). Comme le test du chi-carré, il utilise également une table de contingence pour croiser les valeurs de deux variables catégoriques et repose sur une hypothèse nulle. Supposons que nous ayons le tableau de contingence 2×2 suivant. La valeur *p* unilatérale est obtenue à l'aide de la formule suivante, où le symbole "!" désigne la factorielle (Fisher 1935 : 49) :

$$p = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{(a!\,b!\,c!\,d!\,n!)}$$

	Groupe 1	Groupe 2	Total de la ligne
Catégorie 1	a	b	a+b
Catégorie 2	c	d	c+d
Total de la colonne	a+c	b+d	a+b+c+d = n

Concernant cet outil statistique, nous employons la fonction {fisher.test} du package R {stats}. A la différence du test du chi carré, cette fonction fournit uniquement la valeur *p*, le seuil de signification étant généralement fixé à 0,05. Elle fournit aussi les degrés de liberté (*ddl*), définis comme le nombre de fréquences qui peuvent varier indépendamment tout en maintenant les totaux marginaux constants (Gries 2021 : 173).

4.3.2 UpsetR

Dans cette recherche, à la lumière de Hutchinson (1957), nous proposons une nouvelle définition de la niche morphologique ainsi que de ses sous-niches (pour plus de détails, voir la section 6.2.1). Afin de visualiser la niche et les sous-niches des suffixes instrumentaux, nous utilisons le package R {UpSetR} (Conway et al., 2017 : 2938), qui repose sur une mise en page matricielle pour représenter les intersections entre ensembles ainsi que leur taille. Concrètement, UpSetR accepte trois formats d'entrée : « (i) un tableau dont les lignes représentent des éléments et dont les colonnes indiquent les appartenances aux ensembles ainsi que d'autres attributs ; (ii) des ensembles de noms d'éléments ; (iii) une expression décrivant la taille des intersections d'ensembles, selon le format du package venneuler. »[36] (Conway et al. 2017 : 2938).

En outre, Pour chaque ensemble participant à une intersection donnée, « un point solide rempli est placé dans la cellule de matrice correspondante. Si un ensemble ne fait pas partie de l'intersection, un point gris clair est affiché. Une ligne solide verticale relie le point solide le plus haut au point solide le plus bas dans chaque colonne pour souligner les relations basées sur les colonnes. La taille des intersections est montrée sous forme de graphique à barres placé au-dessus de la matrice, de sorte que chaque colonne s'aligne exactement avec une barre. Un second graphique à barres montrant la taille de chaque ensemble est affiché à gauche de la matrice. »[37] (Conway et al., 2017 : 2938–2939)

Afin d'illustrer ce mode de représentation, nous reproduisons la Figure 4.3 extraite du travail de Krassowski et al. (2022). Cette visualisation met en évidence, pour chaque intersection, un ensemble de points solides représentant les ensembles

[36] Le texte original : "(i) a table in which the rows represent elements and columns include set assignments and additional attributes; (ii) sets of elements names; and (iii) an expression describing the size of the set intersections as introduced by the venneuler package".

[37] Le texte original en anglais : "a black filled circle is placed in the corresponding matrix cell. If a set is not part of the intersection, a light gray circle is shown. A vertical black line connects the topmost black circle with the bottom most black circle in each column to emphasize the column-based relationships. The size of the intersections is shown as a barchart placed on top of the matrix so that each column lines up with exactly one bar. A second bar chart showing the size of the each set is shown to the left of the matrix."

impliqués, chacun étant associé à une propriété distinctive.

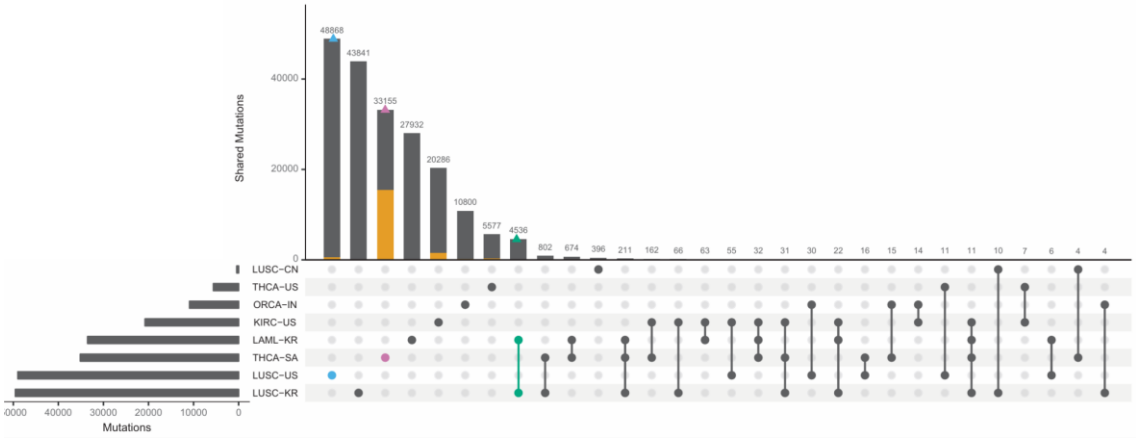

Figure 4.2 Exemple de la visualisation effectuée par UpsetR (Conway et al., 2017 : 2939)

4.3.3 Forêt aléatoire

Pour la modélisation statistique, nous utilisons la forêt aléatoire afin de modéliser les données et de prédire le choix du suffixe. Comme mentionné au chapitre 2, cette méthode s'avère particulièrement pertinente pour appréhender la notion de niche morphologique proposée par Missud (2023). Elle permet de prendre en compte simultanément plusieurs niveaux linguistiques et de traiter des variables à différentes échelles, y compris hiérarchisées.

Un arbre de décision est un algorithme d'apprentissage supervisé, représenté sous forme d'arbre, qui sert à classifier des observations et à effectuer des prédictions par régression. Cet arbre est constitué de nœuds et de branches, où chaque nœud représente un test sur une caractéristique de l'ensemble des données. L'arbre de décision débute au niveau du nœud initial (racine), où il sépare les individus en deux sous-groupes, qui forment des nœuds « filles ». Chaque sous-groupe est ensuite séparé à son tour, et ainsi de suite, dans le but de créer des sous-groupes aussi homogènes que possible par rapport à la variable à expliquer. À la fin de l'arbre, on trouve les feuilles contenant la prédiction (Xu et al. 2023 : 241-242). Concrètement, concernant son fonctionnement, afin de diviser les données en sous-groupes, cet algorithme commence en choisissant la caractéristique la plus informative et en cherchant à maximiser l'homogénéité des groupes. Ce processus est répété de manière itérative jusqu'à ce que le niveau de pureté ou de signification requis soit atteint (Zhang & Dong 2018 : 364).

Prenons l'exemple de la compétition entre les suffixes instrumentaux *-eur* et *-euse*. Comme nous le verrons au chapitre 5, les sous-classes ontologiques des dérivés

ainsi que les espaces thématiques des bases jouent un rôle déterminant dans leur différenciation. L'arbre de décision (cf. Figure 4.4) procède par des étapes successives de segmentation fondées sur des critères hiérarchisés.

Dans cet arbre de décision, l'embranchement initial porte sur la variable sémantique « Sous-classes ontologiques ». Si les lexèmes ne dénotent pas des noms de la sous-classe Machine (c'est-à-dire Appareil, Instrument de captation, Outil, Récipient ou Substance chimique), on aboutit à une feuille où le suffixe *-eur* est majoritaire (218 lexèmes sur 250). Sinon, pour les noms relevant de la sous-classe Machine, l'arbre examine d'abord la taille syllabique du radical, ce qui le divise en deux feuilles. Dans la première, qui regroupe les radicaux polysyllabiques, le sous-arbre se scinde selon le mode d'articulation : les radicaux à occlusive prennent presque toujours *-eur* (71/72), les fricatives restent majoritairement en *-eur* (31/44), tandis que les approximantes ou nasales privilégient plutôt *-euse* (16/24). Dans la seconde feuille, on trouve les radicaux dissyllabiques ou monosyllabiques. À ce niveau, l'étymologie (ETYM) devient déterminante : les noms formés avant l'ère scientifique conservent *-eur* (10/11), alors que les formations des XVIIIe–XXe siècles privilégient *-euse* (240/353).

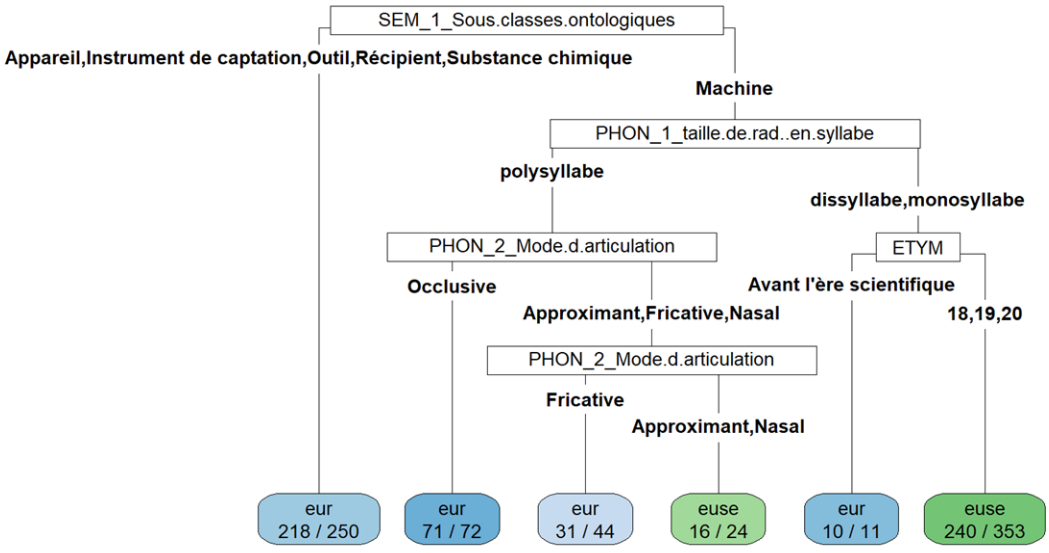

Figure 4.3 Exemple simple de l'arbre de décision

Cependant, l'arbre de décision présente un risque important de surapprentissage, c'est-à-dire une sensibilité excessive au bruit et aux valeurs aberrantes, ainsi qu'aux variations des données. Pour pallier ces limitations, Breiman (2001 : 11–14) a proposé une généralisation sous la forme de la forêt aléatoire. Concrètement, le processus

commence par la génération d'un échantillon *bootstrap*[38] à partir des données initiales, ce qui correspond à la phase d'apprentissage automatique. A chaque nœud de l'arbre, un sous-ensemble de variables est sélectionné aléatoirement, de manière indépendante à chaque division, puis le modèle recherche la meilleure segmentation possible à partir de ces variables. L'arbre est ensuite développé jusqu'à sa profondeur maximale, c'est-à-dire jusqu'à ce que chaque branche aboutisse à une feuille sans possibilité de division supplémentaire, sans procédure d'élagage. Ce processus est répété de nombreuses fois, chaque arbre étant ainsi construit sur un échantillon différent. Enfin, les résultats de tous les arbres de décision sont agrégés pour produire une prédiction finale : chaque arbre "vote", et la classe majoritaire est retenue pour l'observation concernée. Dans le cas d'une régression, la prédiction finale correspond à la moyenne des résultats fournis par l'ensemble des arbres.

Il existe principalement deux algorithmes utilisés dans les forêts aléatoires : CART (Classification and Regression Trees) et CTree (Conditional Inference Trees). CART, proposé par Breiman (2001), repose sur des critères d'impureté (comme l'indice de Gini) ou de variance, qui servent à déterminer, à chaque étape de la construction de l'arbre, le meilleur point de séparation entre les données. CTree, quant à lui, introduit des tests statistiques conditionnels à chaque nœud (Strobl et al., 2008), ce qui en fait une version optimisée de CART. Bien que l'algorithme CTree présente des avantages théoriques en matière de sélection non biaisée des variables, notamment lorsqu'il s'agit de variables corrélées ou de types hétérogènes (Strobl et al., 2008), nous avons choisi d'utiliser CART pour plusieurs raisons.

D'une part, le package R {caret} constitue une interface centrale pour l'implémentation du modèle de forêt aléatoire traditionnel (CART), éprouvé dans de nombreux travaux empiriques. Il permet notamment l'estimation d'indicateurs robustes tels que le Mean Decrease Accuracy (MDA), et la validation croisée. D'autre part, CTree demeure peu éprouvé dans le domaine de la linguistique, où son usage reste marginal comparé à d'autres approches plus établies. Son implémentation actuelle (notamment via le package {party}) présente certaines limites pratiques, en particulier en matière de documentation, de compatibilité avec des métriques standard comme le MDA, et de stabilité sur des corpus de grande taille. Ainsi, notre choix de CART repose

[38] Concernant le *bootstrap*, c'est une méthode statistique consistant à rééchantillonner de façon répétée, avec remise, les données d'origine afin d'estimer l'incertitude et la stabilité du modèle.

sur une combinaison de critères théoriques, techniques et pragmatiques visant à garantir à la fois la robustesse prédictive et la lisibilité interprétative du modèle.

L'importance des variables est estimée à l'aide de l'indicateur Mean Decrease Accuracy (MDA), qui mesure la baisse moyenne de la précision du modèle lorsqu'une variable est aléatoirement perturbée. Plus cette baisse est importante, plus la variable est considérée comme cruciale pour la qualité des prédictions (Xu et al. 2023 : 257). Cette estimation est réalisée dans le cadre de la forêt aléatoire par validation croisée 10 passes. Les performances globales du modèle sont évaluées principalement à l'aide de l'erreur OOB (Out-of-Bag), issue de l'implémentation du package {randomForest}, qui permet une estimation fiable sans recours à un jeu de test explicite. En complément, le package {caret} a été mobilisé pour l'exploration de paramètres internes (notamment mtry) et le calcul de mesures secondaires telles que l'AUC macro-moyenne, obtenue selon une stratégie one-vs-all [39] à l'aide du package {pROC} sur une partition 70/30 des données. Ces deux indicateurs offrent à la fois une robustesse dans l'évaluation et une lisibilité des performances du modèle.

Enfin, les paramètres ont été fixés selon des pratiques standards de modélisation, en fonction du nombre de prédicteurs et de la nature de la tâche (classification supervisée). Le paramètre *mtry*, qui détermine le nombre de variables sélectionnées aléatoirement à chaque nœud, a été fixé à la racine carrée du nombre de variables prédictives (√vp), conformément à l'usage courant pour les tâches de classification. Par exemple, pour un modèle intégrant 7 variables, mtry est fixé à 3. Le nombre d'arbres (*ntree*) a été fixé à 1000 dans l'ensemble des modèles, afin de garantir une convergence stable des résultats.

4.3.4 Formule pour mesurer le chevauchement des niches

Pour mesurer le chevauchement des niches que nous traitons ici, nous nous appuyons sur la formule dite de taux de chevauchement, empruntée aux études écologiques. Cette formule est initialement utilisée pour mesurer le degré de chevauchement entre les niches d'espèces différentes, c'est-à-dire la proportion des

[39] L'AUC macro-moyenne, obtenue selon une stratégie one-vs-all, consiste à calculer l'AUC séparément pour chaque classe en la comparant à l'ensemble des autres (one-vs-all), puis à faire la moyenne des valeurs obtenues pour chaque classe.

ressources partagées par rapport à l'ensemble des ressources utilisées par une espèce donnée (Hurlbert 1978). Dans notre thèse, nous l'appliquons aux modalités linguistiques activées par les schémas morphologiques : le chevauchement désigne ici la proportion de modalités linguistiques partagées entre deux schémas morphologiques, ce qui permet de quantifier leur potentiel de compétition dans le lexique. De plus, le calcul que nous proposons constitue une simplification binaire et directionnelle de la formule de chevauchement interspécifique proposée par Hurlbert (1978 : 70, eq. 11 et 14). En remplaçant les fréquences d'utilisation des ressources par des modalités linguistiques activées (valeurs 0/1) et en normalisant par le nombre total de modalités activées dans le schéma cible, nous obtenons un indicateur de recouvrement morphologique adapté à notre cadre linguistique. Plus précisément, concernant le taux de chevauchement, la formule est comme suit :

$$C_{\text{A sur B}} = \frac{|S_A \cap S_B|}{|S_B|}$$

Dans cette formule, $C_{\text{A sur B}}$, le taux de chevauchement de **A sur B**, c'est-à-dire la proportion du schéma B (S_B) qui se retrouvent aussi dans schéma A.

- Si $C_{\text{A sur B}} = 1$, alors $S_A \subseteq S_B$ (chevauchement total)
- Si $C_{\text{A sur B}} = 0$, alors $S_A \cap S_B = \emptyset$ (aucun chevauchement)

Plus la valeur est élevée, plus le chevauchement est important, et plus la compétition potentielle entre les deux schémas morphologiques est forte. Inversement, une valeur faible indique un chevauchement minimal, et donc une interaction compétitive limitée entre les deux schémas.

Chapitre 5 Construction des noms d'instruments

Dans ce chapitre, nous allons nous concentrer sur l'analyse des données. Comme cela a été évoqué dans les chapitres précédents, nous avons émis l'hypothèse selon laquelle la suffixation des noms d'instrument est sensible aux propriétés phonologiques (taille du radical de la base du dérivé et segment phonologique final du radical), morphologiques (classe flexionnelle du verbe de la base du dérivé, structure morphologique de la base du dérivé, espaces thématique), syntaxiques (catégorie syntaxique de la base du dérivé), ainsi que sémantiques (profils sémantiques assumés par la base du dérivé et les sous-classes ontologiques du dérivé). Parmi ces dimensions sémantiques, l'analyse des profils sémantiques assumés par la base sera explicitement interprétée à la lumière du cadre de la Néo-Morphologie Cognitive que nous proposons (NMC), conformément aux quatre étapes du processus de dérivation modélisées par ce cadre : identification du référent, profilage conceptuel, restructuration métonymique et les espaces d'intégration morphosémantique.

A cette fin, nous avons procédé à l'annotation des lexèmes concernés en tenant compte de ces différentes propriétés. En suivant ce fil conducteur, nous présenterons les résultats en fonction de ces propriétés. Plus précisément, dans les sections suivantes, nous exposerons de manière détaillée les résultats obtenus. Dans un premier temps, nous mettrons en évidence les proportions des schémas morphologiques recensés (les fréquences et les pourcentages) en fonction de diverses propriétés linguistiques. Nous mettrons en évidence les unités en vert, correspondant à des effectifs supérieurs à 5 et à des pourcentages dépassant 10 %, ainsi que les unités en rose, qui affichent des effectifs et des pourcentages nuls, afin de mieux mettre en lumière les résultats les plus intéressants. Par la suite, si cela s'avère nécessaire pour tester la différence significative entre les variables (schémas morphologiques), des analyses statistiques inférentielles seront réalisées. Nous nous efforcerons d'interpréter les résultats obtenus à l'aide des théories concernées et des liens entre les propriétés linguistiques prises en compte dans la thèse, en illustrant ces résultats par des exemples représentatifs. Enfin, la section 5.6 sera consacrée à une discussion comparative entre nos résultats et ceux des travaux antérieurs, en soulignant les apports de notre analyse.

5.1 Au niveau syntaxique

Selon plusieurs morphologues (Roché 1998 ; Dal 1999 ; Fradin 2003 ; Namer & Villoing 2008 ; Huyghe & Tribout 2015 ; Knittel 2017a ; Knittel 2017b), la contrainte syntaxique de la base du dérivé est déterminante dans la formation des mots. En ce qui concerne les noms d'instruments, comme l'indique le tableau 5.1, cette contrainte exerce une influence considérable sur leur construction. En effet, la base verbale s'avère plus productive, car les noms d'instruments sont fréquemment des noms déverbaux (Rainer 2021). Les suffixes tels que *-eur*, *-euse*, *-rice*, *-oir*, *-oire* et *-ant* s'attachent prototypiquement à une base verbale (cf. les exemples 1a-1f). De plus, les suffixes *-ier*, *-ière* s'attachent prototypiquement à une base nominale (cf. les exemples 2a-2b). En revanche, les suffixes *-et*, *-ette* et *-on* s'ajoutent à la fois à une base verbale (cf. les exemples 3a-3c) et à une base nominale (cf. les exemples 3d-3f). Cette observation valide les conclusions obtenues par les morphologues évoqués.

1) a. RAPPORTER $_{V}$ > RAPPORTEUR $_{N}$
   b. TONDRE $_{V}$ > TONDEUSE $_{N}$
   c. CALCULER $_{V}$ > CALCULATRICE $_{N}$
   d. FENDRE $_{V}$ > FENDOIR $_{N}$
   e. PASSER $_{V}$ > PASSOIRE $_{N}$
   f. ABLUER $_{V}$ > ABLUANT $_{N}$
2) a. PLAFOND $_{N}$ > PLAFONNIER $_{N}$
   b. CAFE $_{N}$ > CAFETIERE $_{N}$
3) a. SOUFFLER $_{V}$ > SOUFFLET $_{N}$
   b. SONNER $_{V}$ > SOUNETTE $_{N}$
   c. LORGNER $_{V}$ > LORGNON $_{N}$
   d. ABLE $_{N}$ > ABLERET $_{N}$
   e. LUNE $_{N}$ > LUNETTE $_{N}$
   f. BASSE $_{N}$ > BASSON $_{N}$

Tableau 5.1 Proportion des catégories syntaxiques par suffixes instrumentaux

Suffixes \ Catégories syntaxiques (pourcentage)	Verbe	Nom	Adjectif	Total
-eur	**444 (98%)**	11 (2%)	0 (0%)	455
-euse	**306 (98%)**	5 (2%)	0 (0%)	311
-rice	**11 (100%)**	0 (0%)	0 (0%)	11
-oir	**202 (98%)**	4 (2%)	0 (0%)	206
-oire	**27 (100%)**	0 (0%)	0 (0%)	27
-ant	**71 (97%)**	2 (3%)	0 (0%)	73
-ante	**3 (100%)**	0 (0%)	0 (0%)	3
-ier	4 (6%)	**66 (93%)**	1 (1%)	71
-ière	5 (8%)	**58 (92%)**	0 (0%)	63
-et	**25 (68%)**	**9 (24%)**	3 (8%)	37
-ette	**43 (75%)**	**14 (25%)**	0 (0%)	57
-ail	**12 (60%)**	**8 (40%)**	0 (0%)	20
-ante	2 (67%)	1 (33%)	0 (0%)	3
-aille	2 (100%)	0 (0%)	0 (0%)	2

5.2 Au niveau phonologique

Dans cette section, deux propriétés linguistiques principales sont prises en compte, à savoir la taille du radical de la base du dérivé et le segment phonologique final du radical de la base du dérivé.

5.2.1 Taille du radical de la base du dérivé

Selon les travaux de Plénat (2005) et Roché (2003), les contraintes phonologiques ne sont pas propres aux schémas morphologiques. En d'autres termes, elles ne concernent pas exclusivement un schéma mais s'imposent à l'ensemble du système morphologique. Le Tableau 5.2 montre la distribution de la taille syllabique du radical de la base du dérivé selon les schémas morphologiques.

Tableau 5.2 Distribution de la taille syllabique du radical de la base du dérivé selon les suffixes instrumentaux (en pourcentage)

Suffixes \ Taille (pourcentage)	1	2	3	4	5	6	Total
-eur	**83 (18%)**	**170 (37%)**	**146 (32%)**	**47 (10%)**	5 (1%)	4 (1%)	455
-euse	**129 (41%)**	**148 (48%)**	**32 (10%)**	2 (1%)	0 (0%)	0 (0%)	311
-rice	0 (0%)	1 (9%)	**9 (82%)**	1 (9%)	0 (0%)	0 (0%)	11
-oir	**103 (50%)**	**93 (45%)**	10 (5%)	0 (0%)	0 (0%)	0 (0%)	206
-oire	**12 (44%)**	**11 (41%)**	3 (11%)	1 (4%)	0 (0%)	0 (0%)	27
-ant	**10 (14%)**	**28 (39%)**	**28 (39%)**	4 (6%)	1 (1%)	1 (1%)	72
-ante	1 (33%)	1 (33%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)	3
-ier	**30 (42%)**	**38 (54%)**	3 (4%)	0 (0%)	0 (0%)	0 (0%)	71
-ière	**32 (51%)**	**28 (44%)**	3 (5%)	0 (0%)	0 (0%)	0 (0%)	63
-et	**28 (76%)**	**8 (22%)**	1 (3%)	0 (0%)	0 (0%)	0 (0%)	37
-ette	**36 (63%)**	**20 (35%)**	1 (2%)	0 (0%)	0 (0%)	0 (0%)	57
-on	**14 (70%)**	**6 (30%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20
-ail	1 (33%)	1 (33%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)	3
-aille	2 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2

Tout d'abord, selon le Tableau 5.2 nous avons observé que la majorité des schémas tendent à sélectionner des radicaux monosyllabiques et dissyllabiques. Des exemples illustrant chacun des schémas sont présentés en (4a-h).

4) a. *-euse* : TONDRE $_V$ > TONDEUSE $_N$; MOULINER $_V$ > MOULINEUSE $_N$

b. *-oir* : RASER $_V$ > RASOIR $_N$; ARROSER $_V$ > ARROSOIR $_N$

c. *-oire* : PASSER $_V$ > PASSOIRE $_N$; ATTRAPER $_V$ > ATTRAPOIRE $_N$

d. *-ier* : VRAC $_N$ > VRAQUIER $_N$; CHIMIQUE $_A$ > CHIMIQUIER $_N$

e. *-ière* : CAFE $_V$ > CAFETIERE $_N$; BALEINE $_N$ > BALEINIERE $_N$

f. *-et* : BLANC $_N$ > BLANCHET $_N$; PATOUILLE $_N$ > PATOUILLET $_N$

g. *-ette* : SONNER $_V$ > SONNETTE $_N$; OREILLE $_N$ > OREILLETTE $_N$

h. *-on* : BASSE $_N$ > BASSON $_N$; PALISSON $_N$ > PALISSON $_N$

En outre, les schémas *-eur, -rice, -ant* ont une propension à s'ajouter aux radicaux dissyllabiques et trisyllabiques (voir exemples 5a-5c). Il est important de souligner qu'en ce qui concerne l'effet de la taille des radicaux, certains lexèmes sont soumis à des conditions purement phonologiques (*par ex.* MAZOUTER $_V$ > MAZOUTEUR $_N$), tandis que d'autres dépendent de facteurs morphologiques. À titre d'exemple, en (5c), la

suffixation *-ifier* accroît la taille du radical de la base du dérivé de deux syllabes. En ce qui concerne l'exemple (5b), c'est le thème verbal 13 qui rallonge le radical d'une syllabe. Enfin, pour l'exemple (5a), la préfixation *a-* et le thème 1, sous la forme de l'adjonction de /is/, contribuent ensemble à l'augmentation de la taille phonologique du radical de deux syllabes.

5) a. *-eur* : AGRANDIR $_{V}$ > AGRANDISSEUR $_{N}$

b. *-rice* : CALCULER $_{V}$ > CALCULATRICE $_{N}$

c. *-ant* : ACIDIFIER $_{V}$ > ACIDIFIANT $_{N}$

En plus des trois types de syllabes mentionnés, le Tableau 5.2 met en évidence que les suffixations *-eur* et *-ant* peuvent s'ajouter aux radicaux de plus de trois syllabes. En ce qui concerne l'augmentation de la taille des radicaux, l'affixation et la sélection des thèmes jouent un rôle primordial. Nous prenons en exemple des noms avec la suffixation *-eur* (exemples 6) pour illustrer ces cas de manière précise. En effet, on observe dans l'exemple (6a), de même que dans (5b), que la sélection du thème 13 augmente la longueur du radical d'une syllabe. De plus, il est particulièrement intéressant de noter que, pour les exemples (6b) et (6c), les affixations et le thème verbal influent ensemble sur l'augmentation de la taille du radical. Plus précisément, dans l'exemple (6b), la suffixation *-ifier* et le thème verbal 13 augmentent la taille des radicaux de trois syllabes. Quant à l'exemple (6c), il s'agit de deux affixations (une préfixation et une suffixation) ainsi que de la sélection du thème verbal 13, qui contribuent tous au rallongement des radicaux.

6) a. thème verbal 13 : DENEBULER $_{V}$ > DENEBULATEUR $_{N}$

b. suffixation + thème verbal 13 : HUMIDIFIER $_{V}$ > HUMIDIFICATEUR $_{N}$

c. préfixation + suffixation + suffixation + thème verbal 13 : INSENSIBILISER $_{V}$ > INSENSIBILISATEUR $_{N}$; DESHUMIDIFIER $_{V}$ > DESHUMIDIFICATEUR $_{N}$

Il convient de souligner que la préférence pour certains types de radicaux est probablement liée à la différence de longueur entre les schémas. Selon Lignon (2013 : 116) et Bonami & Thuilier (2019 : 18), la longueur des schémas influence le choix des radicaux. En effet, les schémas comportant des segments plus longs ont tendance à choisir les radicaux plus courts (par exemple, les schémas *-oir/-oire*, qui comptent trois phonèmes). En revanche, les schémas plus courts s'ajoutent probablement à des radicaux plus longs (comme les schémas *-ant/-eur*, qui comptent moins de trois segments).

Du point de vue de la compétition morphologique, nous avons classé les suffixes concernés en deux groupes, en fonction de la taille phonologique de la base du dérivé, comme l'illustre le Tableau 5.3. Plus précisément, pour les radicaux monosyllabiques et dissyllabiques, les schémas *-eur, -euse, -oir, -oire, -ant, -ier, -ière, -et, -ette*, ainsi que *-on* se trouvent en concurrence. En ce qui concerne les radicaux comportant plus de syllabes, les schémas *-eur, -euse, -rice* et *-ant* sont en compétition. A partir de ces résultats, bien que certaines compétitions morphologiques soient clairement identifiées en fonction de la taille phonologique de la base du dérivé, certains suffixes sont encore en compétition. Par conséquent, nous nous appuierons sur d'autres propriétés linguistiques afin de mettre en relief la distinction entre ces schémas.

Tableau 5.3 Catégorisation des schémas en fonction des types de syllabes

Types de syllabes	schémas
Monosyllabe et dissyllabe	*-eur, -euse, -oir, -oire, -ant, -ier, -ière, -et, -ette, -on*
Trisyllabe et plus	*-eur, -euse, -rice, -ant*

5.2.2 Segment phonologique final du radical

Après avoir analysé la taille phonologique, nous nous penchons désormais sur le segment phonologique final du radical de la base du dérivé. Comme cela est indiqué au chapitre 1, selon Plénat (2005) et Roché (2003), les contraintes phonologiques influent sur le système morphologique au lieu de suffixes spécifiques. En effet, pour les segments phonologiques finaux du radical, les contraintes dissimilatives se manifestent de manière différente dans la formation des lexèmes, par exemple : base supplétive, décalage, infixation, épenthèse, etc. Toutefois, Koehl (2012 : 186) et Bonami & Thuilier (2019 : 19-20) soutiennent que certains schémas morphologiques favorisent des phonèmes spécifiques. Afin de vérifier ces deux hypothèses et d'approfondir l'analyse du segment phonologique final du radical dans le cadre de la suffixation des noms d'instruments, nous procèderons à une étude des lexèmes en la matière.

Tableau 5.4 Mode d'articulation des segments phonologiques finaux du radical

Suffixes \ Mode d'articulation (pourcentage)	Occlusive	Nasal	Fricative	Approximant	Voyelle	Total
-eur	**247 (54%)**	19 (4%)	**145 (32%)**	**44 (10%)**	0 (0%)	455
-euse	**75 (24%)**	**58 (19%)**	**113 (36%)**	**63 (20%)**	2 (1%)	311
-oir	**53 (26%)**	**25 (12%)**	**96 (47%)**	**32 (16%)**	0 (0%)	206
-ant	**15 (21%)**	4 (6%)	**31 (43%)**	**21 (29%)**	1 (1%)	74
-ier	**24 (34%)**	**17 (24%)**	**12 (17%)**	**17 (24%)**	1 (1%)	71
-ière	**31 (49%)**	**9 (14%)**	**17 (27%)**	4 (6%)	2 (3%)	63
-ette	2 (4%)	**17 (30%)**	**22 (39%)**	**16 (28%)**	0 (0%)	57
-et	**5 (14%)**	3 (8%)	**16 (43%)**	**13 (35%)**	0 (0%)	37
-oire	**8 (30%)**	3 (11%)	**10 (37%)**	**6 (22%)**	0 (0%)	27
-on	4 (20%)	1 (5%)	**10 (50%)**	4 (20%)	1 (5%)	20
-rice	**11 (100%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11
-ail	2 (67%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)	3
-ante	0 (0%)	2 (67%)	1 (33%)	0 (0%)	0 (0%)	3
-aille	0 (0%)	2 (100%)	0 (0%)	0 (0%)	0 (0%)	2

Selon le tableau 5.4, l'analyse du mode d'articulation des radicaux de la base du dérivé met en évidence plusieurs tendances selon les schémas. Le suffixe-*eur* montre une nette surreprésentation de radicaux terminant par un phonème occlusif, tandis qu'il évite des radicaux se terminant par une nasale et approximante. La suffixation en *-euse* montre une surreprésentation marquée dans la catégorie fricative. De plus, les suffixations en *-oir/-oire* sont fortement associés aux radicaux se terminant par une fricative et sous-représentés avec des radicaux se terminant par une consonne nasale. De manière complémentaire, *-ier* est surreprésenté avec des radicaux se terminant par une occlusive et sous-représenté dans la catégorie fricative. Et le suffixe *-ière* est surreprésenté avec des radicaux se terminant par une occlusive et faiblement représenté avec des radicaux se terminant par une approximante. Concernant le suffixe-*ant,* il est surreprésenté avec des radicaux se terminant par une fricative et évite des radicaux se terminant par une nasale.

En outre, les suffixes *-ette* et *-et* se distinguent également : *-ette* est sous-représenté avec des radicaux se terminant par une occlusive et surreprésenté avec des radicaux se terminant par une nasale et une approximante ; *-et* manifeste une sous-représentation dans la catégorie nasale et une surreprésentation avec des radicaux se terminant par une fricative. Enfin, *-on* est surreprésenté avec des radicaux se terminant par une fricative ; *-rice* suit une tendance comparable, avec une surreprésentation dans la catégorie occlusive. Ces résultats montrent que certains segments phonologiques exercent une influence déterminante sur le choix du schéma suffixal dans la formation des noms d'instruments.

Tableau 5.5 Segment phonologique final du radical

Suffixes \ Phonème (pourcentage)	b	d	e	f	g	ɥ	i	j	k	l	m	n	ɲ	p	ʁ	s	ʃ	t	u	v	w	y	z	ʒ	Total
-eur	5	11	0	4	4	0	0	22	4	21	1	17	1	7	15	60	8	216	0	7	1	0	42	9	455
	1%	2%	0%	1%	1%	0%	0%	5%	1%	5%	0%	4%	0%	2%	3%	13%	2%	47%	0%	2%	0%	0%	9%	2%	100%
-euse	5	13	0	1	1	0	0	29	4	34	11	43	4	5	28	36	16	47	2	5	0	0	19	8	311
	2%	4%	0%	0%	0%	0%	0%	9%	1%	11%	4%	14%	1%	2%	9%	12%	5%	15%	1%	2%	0%	0%	6%	3%	100%
-rice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	11
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	100%
-oir	2	14	0	5	0	0	0	10	4	22	8	13	4	5	8	36	18	28	0	6	0	0	16	7	206
	1%	7%	0%	2%	0%	0%	0%	5%	2%	11%	4%	6%	2%	2%	4%	17%	9%	14%	0%	3%	0%	0%	8%	3%	100%
-oire	0	2	0	0	0	0	0	2	0	4	2	1	0	1	0	10	0	5	0	0	0	0	0	0	27
	0%	7%	0%	0%	0%	0%	0%	7%	0%	15%	7%	4%	0%	4%	0%	37%	0%	19%	0%	0%	0%	0%	0%	0%	100%
-ant	1	2	0	0	0	1	0	14	0	6	3	1	0	3	7	6	2	9	0	1	0	1	15	0	72
	1%	3%	0%	0%	0%	1%	0%	19%	0%	8%	4%	1%	0%	4%	10%	8%	3%	13%	0%	1%	0%	1%	21%	0%	100%
-ante	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	3
	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	33%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
-ier	1	7	0	0	1	0	0	0	2	17	2	15	0	1	5	3	1	12	1	2	0	0	1	0	71
	1%	10%	0%	0%	1%	0%	0%	0%	3%	24%	3%	21%	0%	1%	7%	4%	1%	17%	1%	3%	0%	0%	1%	0%	100%
-ière	3	3	1	0	2	0	1	0	0	4	0	9	0	3	4	8	0	20	0	0	0	0	4	1	63
	5%	5%	2%	0%	3%	0%	2%	0%	0%	6%	0%	14%	0%	5%	6%	13%	0%	32%	0%	0%	0%	0%	6%	2%	100%
-et	1	0	0	0	1	0	0	2	3	9	1	2	0	0	6	0	6	0	0	3	2	0	0	1	37
	3%	0%	0%	0%	3%	0%	0%	5%	8%	24%	3%	5%	0%	0%	16%	0%	16%	0%	0%	8%	5%	0%	0%	3%	100%
-ette	0	0	0	0	0	0	0	6	2	9	2	14	1	0	4	8	3	0	0	2	1	0	4	1	57
	0%	0%	0%	0%	0%	0%	0%	11%	4%	16%	4%	25%	2%	0%	7%	14%	5%	0%	0%	4%	2%	0%	7%	2%	100%
-on	0	1	1	0	1	0	0	3	0	1	0	0	1	1	2	3	1	1	0	1	0	0	2	1	20
	0%	5%	5%	0%	5%	0%	0%	15%	0%	5%	0%	0%	5%	5%	10%	15%	5%	5%	0%	5%	0%	0%	10%	5%	100%
-ail	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	3
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	67%	0%	0%	0%	0%	0%	0%	100%
-aille	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%

Comme le Tableau 5.5 le montre, il nous semble que certains schémas s'adjoignent préférentiellement à des finales constituées de certains phonèmes. A partir de Tableau 5.6, nous classons les phonèmes préférentiels par suffixes instrumentaux.

Tableau 5.6 Les phonèmes préférentiels terminant les radicaux selon les suffixes

Phonèmes	Suffixes
/j/	*-on, -ant*
/s/	*-eur, -euse, -oir, -oire, -ière, -ette*
/t/	*-eur, -euse, -rice, -oir, -oire, -ant, -ier, -ière, -on*
/l/	*-euse, -oir, -ier, -ière, -et, -ette*
/n/	*-euse, -ier, -et, -ette*
/ʁ/	*-ant, -et, -ette*

A propos de ces préférences, il convient de noter que ces phonèmes préférentiels ne sont pas uniquement déterminés par des facteurs phonologiques, mais également par des facteurs morphologiques. Sur le plan morphologique, tout d'abord, le thème exerce une grande influence sur le segment final du radical. Par exemple, en ce qui concerne le suffixe *-rice*, celui-ci sélectionne exclusivement le phonème /t/. Cela s'explique par le fait que le phonème /t/ est systématiquement présent dans le thème verbal 13 (*par ex.* CALCULER $_{V}$ > CALCULATRICE $_{N}$). Étant donné que le suffixe *-rice* favorise le thème 13, le phonème /t/ s'avère privilégié. Une situation similaire se

présente aussi pour le suffixe -*eur* en relation avec le phonème /t/ (*par ex.* DESHUMIDIFIER $_{V}$ > DESHUMIDIFICATEUR $_{N}$). De plus, pour les verbes relevant du deuxième groupe de conjugaison, verbes souvent dérivés par la conversion et la préfixation, le thème 1 se distingue fréquemment par le phonème /s/ (*par ex*. AGRANDISSEUR). Dans notre corpus, on observe que pour un grand nombre de lexèmes se terminant par le schéma -*eur*, la base du dérivé est construite principalement par la conversion, avec le thème généralement identifié comme étant le thème verbal 1. C'est pourquoi le suffixe -*eur* favorise les radicaux terminant par le phonème /s/.

Par ailleurs, la base dérivée par la suffixation -*eur* joue également un rôle crucial. En ce qui concerne les bases se terminant par le phonème /z/, pour le suffixe -*eur*, 15 noms en -*eur* sont construits sur une base verbale en -*iser*, dont le thème 1 se termine en /z/ (*par ex*. CLIMATISEUR). Ainsi, dans le cas du suffixe -*eur* associé à des bases se terminant par le phonème /z/, une partie des lexèmes présente effectivement une motivation morphologique, notamment liée à la base suffixée. A l'opposé, pour le schéma -*euse*, la préférence du segment final du phonème /z/ est conditionnée plutôt phonologiquement que morphologiquement. Par rapport au schéma -*eur*, concernant la structure morphologique de la base du dérivé, moins de lexèmes en -*euse* sélectionnent la base suffixée par -*iser* (pour l'illustration de ce phénomène morphologique voir la section §5.2.2.2.2).

Nous examinerons ensuite les unités dont l'effectif et le pourcentage sont nuls. D'après le tableau 3, la majorité des suffixes n'optent pas pour les segments finaux suivants : /e/, /ɥ/, /i/, /u/, /w/, /y/. Ces phonèmes correspondent à des voyelles ou des semi-voyelles. En tant que contrainte phonologique, le hiatus constitue une restriction importante dans la formation des mots (Plénat 2005), car il est généralement évité par le biais de diverses stratégies, telles que l'utilisation d'interfixes ou le décalage. Hormis les phonèmes vocaliques mentionnés, d'autres phonèmes consonantiques ne sont pas sélectionnés par certains suffixes, comme nous le montrons dans le tableau 5.7.

Tableau 5.7 Les phonèmes finaux des radicaux qui ne sont pas choisis par les suffixes

Phonèmes	Suffixes
/b/	*-oire, -ette*
/d/	*-et, -ette*
/f/	*-oire, -ant, -ier, -ière, -et, -ette*
/g/	*-oir, -oire, -ant, -ette*
/k/	*-oir, -oire, -ière*
/ɲ/	*-oire, -ant, -ier, -ière, -et*
/p/	*-et, -ette*
/ʀ/	*-oire*
/ʃ/	*-oire, -ière*
/t/	*-et, -ette*
/v/	*-oire, -ière*
/z/	*-oire, -et*
/ʒ/	*-oire, -ant, -ier*

En effet, pour ces phonèmes préférentiels des suffixes instrumentaux, certaines bases sont conditionnées par des facteurs phonologiques, tandis que d'autres le sont par des facteurs sémantiques. En fonction de ces types de conditions, nous les classons, comme illustré dans le tableau 5.8. À titre d'exemple, pour les phonèmes conditionnés phonologiquement, les schémas *-oir* et *-oire* ne peuvent pas être ajoutés au phonème /g/, car selon Klausenburger (1970 : 67), la consécution /gw/ est rare dans le français moderne selon la théorie phonotactique (*par ex.* GOINFRE et ses dérivés, GOITRE et son dérivé). C'est pourquoi il n'existe pas de noms se terminant par *-oir* ou *-oire* pour le phonème /g/. En revanche, de nombreux schémas n'apparaissent pas après certains phonèmes pour d'autres raisons, en particulier des contraintes sémantiques. Par exemple, en ce qui concerne les paires entre le phonème /ɲ/ et les schémas *-oire*, *-ant* et *-et*, elles sont possiblement conditionnées par des facteurs sémantiques. En effet, nous pouvons lister des lexèmes tels que BAIGNOIRE, ENSEIGNANT et POIGNET. Cela signifie que ces consécutions sont acceptables en français moderne. Ainsi, pour ces cellules dont le contenu est nul, il convient de noter qu'il n'existe pas de noms d'instruments dont le radical de la base du dérivé se termine par les phonèmes mentionnés.

Tableau 5.8 Classification des schémas selon les types de conditions linguistiques

Types de conditions	Phonèmes	Suffixes
Phonologique	/g/	*-oir, -oire* (Klausenburger 1970)
	/ɲ/	*-ier, -ière* (Roché 1998)
Sémantique	/b/	*-oire, -ette*
	/d/	*-et, -ette*
	/f/	*-oire, -ant, -ier, -ière, -et, -ette*
	/g/	*-ant, -ette*
	/k/	*-oir, -oire, -ière*
	/ɲ/	*-oire, -ant, -et*
	/p/	*-et, -ette*
	/ʀ/	*-oire*
	/ʃ/	*-oire, -ière*
	/t/	*-et, -ette*
	/v/	*-oire, -ière*
	/z/	*-oire, -et*
	/ʒ/	*-oire, -ant, -ier*

Enfin, nous abordons la question de la compétition morphologique. A partir du Tableau 5.6, il est possible d'observer que sept phonèmes finaux des bases se révèlent plus rentables pour les suffixes instrumentaux. Parmi ces phonèmes, certains facilitent la résolution des compétitions entre les schémas. Par exemple, les radicaux se terminant par les phonèmes /l/ et /n/ sont majoritairement associés au suffixe *-euse* plutôt qu'au suffixe *-eur*, ce qui contribue à différencier la distribution de ces deux schémas suffixaux. D'autres schémas demeurent en concurrence. A titre d'exemple, les schémas *-eur* et *-euse* tendent à privilégier tous les deux les bases se terminant par le phonème /s/.

5.3 Au niveau morphologique

Dans cette section, nous examinerons les données au niveau morphologique. Comme mentionné dans le chapitre 1, nous avons identifié trois propriétés morphologiques à étudier, à savoir la classe flexionnelle du verbe et les thèmes de la base du dérivé sélectionnés par les suffixes, ainsi que sa structure morphologique. En

suivant ce fil conducteur, nous procèderons à l'analyse de ces propriétés morphologiques dans les sous-sections suivantes.

5.3.1 Classe flexionnelle du verbe de la base du dérivé

D'après les recherches de Lapraye (2017), la classe flexionnelle du verbe de la base du dérivé exerce une influence sur le choix des suffixes événementiels (*-age* et *-ment*). Certains suffixes sont associés à des classes flexionnelles plus spécifiques. Comme l'indique le Tableau 5.9, le premier groupe de conjugaison s'avère plus rentable pour la majorité des schémas. Ce constat n'est guère surprenant, car selon Eluerd (2009 : 20), le premier groupe de conjugaison est plus productif en français par rapport aux deuxième et troisième groupes, englobant environ 10 000 verbes. C'est pourquoi la grande proportion du premier groupe se présente avec ces schémas.

Tableau 5.9 Distribution de la classe flexionnelle préférentielle du verbe de la base du dérivé selon les suffixes

Suffixes \ Classes flexionnelles (pourcentage)	1	2	3	Total
-eur	**391 (88%)**	28 (6%)	25 (6%)	444
-euse	**277 (91%)**	15 (5%)	14 (5%)	306
-rice	**9 (82%)**	0 (0%)	2 (18%)	11
-oir	**184 (91%)**	7 (3%)	11 (5%)	202
-oire	**23 (85%)**	3 (11%)	1 (4%)	27
-ant	**64 (90%)**	3 (4%)	4 (6%)	71
-ante	3 (100%)	0 (0%)	0 (0%)	3
-ier	4 (100%)	0 (0%)	0 (0%)	4
-ière	**5 (100%)**	0 (0%)	0 (0%)	5
-et	**25 (100%)**	0 (0%)	0 (0%)	25
-ette	**39 (91%)**	1 (2%)	3 (7%)	43
-on	**12 (100%)**	0 (0%)	0 (0%)	12
-ail	2 (100%)	0 (0%)	0 (0%)	2
-aille	1 (50%)	0 (0%)	1 (50%)	2

De plus, selon le Tableau 5.10, nous pouvons observer aussi que pour les noms construits par les suffixes *-eur* et *-euse*, leurs bases verbales appartiennent au deuxième groupe (voir exemples 7a–7b) et au troisième groupe (voir exemples 7c–7d), tandis qu'un certain nombre de noms suffixés par *-oir* sont assez productifs avec des verbes du troisième groupe (voir exemples 7e).

7) a. AGRANDIR $_V$ > AGRANDISSEUR $_N$

b. APLANIR $_V$ > APLANISSEUSE $_N$

c. BATTRE $_V$ > BATTEUR $_N$

d. TONDRE $_V$ > TONDEUSE $_N$

e. FENDRE $_V$ > FENDOIR $_N$

Ensuite, sous l'angle de la compétition morphologique entre les suffixes *-eur*, *-euse*, et *-oir*, ils se révèlent en effet plus productifs dans au moins deux classes flexionnelles. Afin de tester s'il existe des différences significatives entre ces suffixes, nous appliquons le test du chi-carré. Les résultats de ce test (χ^2 = 2,76, df = 4, p > 0,05) indiquent qu'il n'y a pas de différence significative entre les trois schémas. Par conséquent, nous concluons que ces schémas sont en concurrence dans les trois classes flexionnelles de la base verbale.

5.3.2 Structure morphologique de la base du dérivé

Comme évoqué dans les chapitres précédents, plusieurs morphologues (Koehl 2012 ; Bonami & Thuilier 2019 ; Missud & Villoing 2022 ; Missud 2023) mettent en évidence le fait que les suffixations sont influencées par la structure morphologique de la base du dérivé. Certains schémas semblent privilégier des structures morphologiques spécifiques de la base du dérivé, ce qui leur permet de servir de prédicteurs pour le choix des suffixes. Dans cette section, nous procéderons à une analyse des résultats obtenus en fonction des types de structure morphologiques des bases, en distinguant les lexèmes simples, les dérivés et les composés.

Avant d'entamer l'analyse détaillée de chaque catégorie spécifique, nous présentons la proportion des schémas selon trois catégories. Comme l'indique le tableau 5.10, les bases simples et dérivées se révèlent plus fréquentes que les bases composées pour la formation des noms d'instruments. De plus, d'après le tableau 5.10, il est possible de classer les schémas en trois groupes distincts. Le premier groupe comprend les suffixes *-euse* et *-ant*, qui tendent à favoriser des bases verbales dérivées (exemples 8a-8b). Le deuxième groupe inclut les suffixes *-rice*, *-oire*, *-ier*, *-ière*, *-et*, *-ette*, ainsi que *-on*, caractérisés par une préférence pour les bases simples (exemples 9a-9b). Enfin, le dernier groupe regroupe les suffixes-*eur* et *-oir*, qui se distinguent par leur productivité dans la préfixation *a-* et le verbe simple (exemples 10a-10d).

8) a. la préfixation *a-* : PLAN $_A$ > APLANIR $_V$ > APLANISSEUSE $_N$

b. la suffixation *-ifier* : ACIDE $_{A}$ > ACIDIFIER $_{V}$ > ACIDIFIANT $_{N}$

9) a. le verbe simple : TRADUIRE $_{V}$ > TRADUCTRICE $_{N}$

b. le nom simple : THE $_{N}$ > THEIERE $_{N}$

10) a. la préfixation *a-* : GRAND $_{A}$ > AGRANDIR $_{V}$ > AGRANDISSEUR $_{N}$

b. la préfixation *a-* : FIL $_{N}$ > AFFILER $_{V}$ > AFFILOIR $_{N}$

c. le verbe simple : EPLUCHER $_{V}$ > EPLUCHEUR $_{N}$

d. le verbe simple : RASER $_{V}$ > RASOIR $_{N}$

Tableau 5.10 Distribution de la structure morphologique des bases selon les suffixes formant des noms d'instrument

Suffixes \ Structure morphologique (pourcentage)	Lexème simple	Dérivation	Composition	Total
-eur	**214 (47%)**	**237 (52%)**	4 (1%)	455
-euse	**83 (27%)**	**225 (72%)**	3 (1%)	311
-rice	**9 (82%)**	2 (18%)	0 (0%)	11
-oir	**83 (40%)**	**123 (60%)**	0 (0%)	206
-oire	**16 (59%)**	**11 (41%)**	0 (0%)	27
-ant	**20 (28%)**	**50 (69%)**	2 (3%)	72
-ante	1 (33%)	2 (67%)	0 (0%)	3
-ier	**67 (94%)**	3 (4%)	1 (1%)	71
-ière	**60 (95%)**	3 (5%)	0 (0%)	63
-et	**23 (62%)**	**14 (38%)**	0 (0%)	37
-ette	**36 (63%)**	**21 (37%)**	0 (0%)	57
-on	**19 (95%)**	1 (5%)	0 (0%)	20
-ail	2 (67%)	1 (33%)	0 (0%)	3
-aille	1 (50%)	1 (50%)	0 (0%)	2

5.3.2.1 Lexème simple

En ce qui concerne le lexème simple, nous nous engageons dans une classification des structures morphologiques des bases en trois catégories, à savoir le verbe simple, le nom simple et l'adjectif simple, en fonction de leur catégorie syntaxique. D'après le tableau 5.11, les suffixes *-eur*, *-euse*, *-rice*, *-oir*, *-ant* et *-ette* sont principalement associés aux verbes simples (exemples 9a, 10d et 11a-11d). Les schémas *-ier* et *-ière* semblent, quant à eux, privilégier les noms simples (exemples 12 et 9b). En revanche, les suffixes *-et* et *-on* sélectionnent les bases verbales et les bases nominales (exemples 13a-13d).

11) a. EPLUCHER $_{V}$ > EPLUCHEUR $_{N}$

b. TONDRE $_{V}$ > TONDEUSE $_{N}$

c. ABLUER $_{V}$ > ABLUANT $_{N}$

d. LORGNER $_{V}$ > LORGNETTE $_{N}$

12) THON $_{N}$ > THONIER $_{N}$

13) a. SOUFFLER $_{V}$ > SOUFFLET $_{N}$

b. FRISER $_{V}$ > FRISON $_{N}$

c. ABLE $_{N}$ > ABLERET $_{N}$

d. BASSE $_{N}$ > BASSON $_{N}$

Pour interpréter les résultats obtenus, nous avons décidé d'ajouter les bases construites dans le tableau 5.11 et d'examiner ensemble leurs proportions. Selon le tableau 5.12, en ce qui concerne les bases verbales, à l'exception des suffixes *-eur*, *-et* et *-ette*, qui ne montrent pas de préférence marquée pour les types de bases verbales, les autres schémas présentent des préférences distinctes. Concrètement, les suffixes-*euse*, *-oir*, *-ant* favorisent les bases verbales construites. En revanche, les schémas *-rice* et *-oire* privilégient les bases verbales simples. Quant aux suffixes qui ont tendance à sélectionner les bases nominales, les cinq suffixes favorisent les bases nominales simples. Ainsi, il apparaît que, concernant les schémas déverbaux, tous ne privilégient pas les bases verbales simples, mais ils montrent des préférences distinctes. Cependant, étant donné que la productivité des bases verbales est plus élevée, une petite proportion de bases verbales simples peut entraîner un grand nombre de cas (*par ex.* les schémas *-euse* et *-oir*).

Tableau 5.11 Proportion des différentes catégories de lexèmes simples selon les schémas

Suffixes \ structure morphologique (pourcentage)	Verbe simple	Nom simple	Adjectif simple	Total
-eur	**205 (96%)**	9 (4%)	0 (0%)	214
-euse	**81 (96%)**	3 (4%)	0 (0%)	84
-rice	**9 (100%)**	0 (0%)	0 (0%)	9
-oir	**79 (95%)**	4 (5%)	0 (0%)	83
-oire	**16 (100%)**	0 (0%)	0 (0%)	16
-ant	**20 (100%)**	0 (0%)	0 (0%)	20
-ante	1 (100%)	0 (0%)	0 (0%)	1
-ier	2 (3%)	**65 (97%)**	0 (0%)	67
-ière	2 (3%)	**58 (97%)**	0 (0%)	60
-et	**11 (48%)**	**9 (39%)**	3 (13%)	23
-ette	**22 (61%)**	**14 (39%)**	0 (0%)	36
-on	**11 (58%)**	**8 (42%)**	0 (0%)	19
-ail	1 (50%)	1 (50%)	0 (0%)	2
-aille	1 (100%)	0 (0%)	0 (0%)	1

Tableau 5.12 Répartition des suffixes instrumentaux selon les catégories syntaxiques et les structures morphologiques des lexèmes

Suffixes \ Catégorie syntaxique et Procédé morphologique (pourcentage)	Verbe		Nom		Adjectif		
	simple	**construit**	**simple**	**construit**	**simple**	**construit**	**Total**
-eur	**205 (45%)**	**239 (53%)**	9 (2%)	2 (0%)	0 (0%)	0 (0%)	455
-euse	**81 (26%)**	**225 (72%)**	3 (1%)	2 (1%)	0 (0%)	0 (0%)	311
-rice	**9 (82%)**	2 (18%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	11
-oir	**79 (38%)**	**123 (60%)**	4 (2%)	0 (0%)	0 (0%)	0 (0%)	206
-oire	**16 (59%)**	**11 (41%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)	27
-ant	**20 (28%)**	**51 (71%)**	0 (0%)	1 (1%)	0 (0%)	0 (0%)	72
-ante	1 (33%)	2 (67%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3
-ier	2 (3%)	2 (3%)	**65 (88%)**	1 (1%)	3 (4%)	1 (1%)	74
-ière	2 (3%)	3 (5%)	**58 (92%)**	0 (0%)	0 (0%)	0 (0%)	63
-et	**11 (30%)**	**14 (38%)**	**9 (24%)**	0 (0%)	3 (8%)	0 (0%)	37
-ette	**22 (39%)**	**21 (37%)**	**14 (25%)**	0 (0%)	0 (0%)	0 (0%)	57
-on	**11 (55%)**	1 (5%)	**8 (40%)**	0 (0%)	0 (0%)	0 (0%)	20
-ail	1 (33%)	1 (33%)	1 (33%)	0 (0%)	0 (0%)	0 (0%)	3
-aille	1 (50%)	1 (50%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2

Enfin, sous l'angle de la compétition morphologique, il ressort que, bien que les schémas *-euse* et *-oir* privilégient les bases verbales construites, leur grande fréquence engendre une certaine compétition entre les schémas déverbaux. Ainsi, pour la base verbale simple, les suffixes *-eur*, *-euse*, *-rice*, *-oir*, *-oire*, *-et*, *-ette* et *-on* sont en concurrence. De même, en ce qui concerne la base nominale simple, cinq schémas sont en compétition : *-ier*, *-ière*, -et, *-ette*, et *-on*. Dans les sous-sections suivantes, nous examinerons les bases construites plus en détail.

5.3.2.2 Dérivation

Comme indiqué dans le tableau 5.12, en plus du lexème simple, la base construite représente une structure morphologique essentielle dans la formation des noms d'instruments en français. Selon le tableau 5.13, quatre types de procédés morphologiques de dérivation sont utilisés pour les noms d'instruments : la préfixation, la suffixation, la conversion et la régression[40]. Parmi elles, la préfixation et la conversion sont plus fréquentes que la suffixation et la régression. Cependant, par rapport à la suffixation, la régression s'avère moins productive dans le cadre des noms d'instruments. Ainsi, dans les sous-sections suivantes, nous nous concentrons principalement sur la préfixation, la suffixation ainsi que la conversion.

Par ailleurs, en fonction du pourcentage et du nombre de schémas, nous pouvons classer les schémas instrumentaux en deux catégories distinctes. Le premier groupe comprend les schémas *-euse*, *-oir*, *-et*, *-ette* et *-on*, caractérisés par la productivité élevée dans les bases dérivées par la conversion et la préfixation. Concernant le second groupe, il inclut les schémas *-eur* et *-ant*, qui se distinguent par l'utilisation de l'ensemble des trois types de procédés morphologiques. En général, les schémas instrumentaux tendent à privilégier les bases dérivées par la préfixation et la conversion. Ce phénomène linguistique est principalement lié à l'étymologie. En effet, hormis les schémas *-eur* et *-oire*, la majorité des schémas morphologiques en jeu sont des schémas populaires. Selon Huot (2005 : 65), ces schémas morphologiques populaires ont une tendance à choisir des bases populaires, lesquelles sont souvent dérivées par la conversion et la préfixation. Ainsi, la plupart des suffixes instrumentaux favorisent les bases préfixées et converties, car les suffixations verbales productives *-ifier* et *-iser* sont considérées comme des suffixes savants.

[40] La régression désigne la création d'une forme plus courte à partir d'une forme complexe, par inférence analogique, lorsque le locuteur reconstruit un radical supposé absent dans le lexique, par exemple : DELINQUANT > DELINQUER (Dal 2003 : 12-13).

Tableau 5.13 Proportion des procédés morphologiques de dérivation

Suffixes \ Structure morphologique (pourcentage)	Préfixation	Suffixation	Conversion	Régression	Total
-eur	**77 (32%)**	**41 (17%)**	**117 (49%)**	2 (1%)	237
-euse	**74 (33%)**	8 (4%)	**141 (63%)**	1 (0%)	224
-rice	1 (50%)	0 (0%)	1 (50%)	0 (0%)	2
-oir	**57 (46%)**	1 (1%)	**65 (53%)**	0 (0%)	123
-oire	3 (27%)	1 (9%)	**7 (64%)**	0 (0%)	11
-ant	**17 (33%)**	**17 (33%)**	**18 (35%)**	0 (0%)	52
-ante	1 (50%)	0 (0%)	1 (50%)	0 (0%)	2
-ier	0 (0%)	1 (33%)	2 (67%)	0 (0%)	3
-ière	0 (0%)	0 (0%)	3 (100%)	0 (0%)	3
-et	0 (0%)	0 (0%)	**14 (100%)**	0 (0%)	14
-ette	**5 (24%)**	2 (10%)	**14 (67%)**	0 (0%)	21
-on	0 (0%)	0 (0%)	1 (100%)	0 (0%)	1
-ail	1 (100%)	0 (0%)	0 (0%)	0 (0%)	1
-aille	0 (0%)	0 (0%)	1 (100%)	0 (0%)	1

5.3.2.2.1 Préfixation

Comme il a été souligné dans le chapitre 2, la recherche menée par Missud & Villoing (2022) indique que certains schémas morphologiques présentent une préférence pour des préfixes particuliers. Comme le montre le tableau 5.14, les résultats obtenus renforcent la conclusion de Missud & Villoing (2022). Plus précisément, on observe que les bases dérivées par la préfixation *a*-A>V privilégient la suffixation en *-eur* (*par ex.* AGRANDIR $_{V}$ > AGRANDISSEUR $_{N}$), tandis que les bases dérivées par la préfixation *en*-N>V privilégient le suffixe *-euse* (ENFOURNER $_{V}$ > ENFOURNEUSE $_{N}$).

Tableau 5.14 Proportion des préfixations par suffixes instrumentaux

suffixe \ préfixation (pourcentage)	a-A>V	a-N>V	contre-	dé-V>V	dé-N>V	dé-A>V	é-N>V	é-V>V	é-A>V	en-N>V	en-V>V	en-A>V	in-N>N	inter-	re-V>V	sous-	sur-	trans-	Total
-eur	**8**	2	0	**9**	**16**	0	**10**	0	3	5	2	1	0	0	**14**	0	2	5	77
	10%	3%	0%	**12%**	**21%**	0%	**13%**	0%	4%	6%	3%	1%	0%	0%	**18%**	0%	3%	6%	100%
-euse	1	2	1	**7**	**16**	1	**18**	1	0	**14**	0	0	0	1	**10**	1	1	0	74
	1%	3%	1%	**9%**	**22%**	1%	**24%**	1%	0%	**19%**	0%	0%	0%	1%	**14%**	1%	1%	0%	100%
-rice	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	100%
-oir	3	3	0	**6**	**7**	0	**23**	1	1	3	2	0	0	0	**8**	0	0	0	57
	5%	5%	0%	**11%**	**12%**	0%	**40%**	2%	2%	5%	4%	0%	0%	0%	**14%**	0%	0%	0%	100%
-oire	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3
	33%	33%	0%	0%	0%	0%	33%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
-ant	1	0	0	**7**	**5**	1	0	0	0	0	0	0	1	0	2	0	0	0	17
	6%	0%	0%	**41%**	**29%**	6%	0%	0%	0%	0%	0%	0%	6%	0%	12%	0%	0%	0%	100%
-ante	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
-ail	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
-ette	0	1	0	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	5
	0%	20%	0%	20%	0%	0%	40%	0%	0%	0%	0%	0%	0%	0%	20%	0%	0%	0%	100%

De plus, nous avons trouvé qu'il existe quatre préfixes préférentiels pour les schémas instrumentaux, à savoir *dé*-V>V, *dé*-N>V, *é*-N>V et *re*-V>V. Plus précisément, les suffixes *-eur*, *-euse*, *-oir* et *-ant* ont tendance à choisir des bases préfixées par les schémas dé-V>V (voir exemples 14a-14d) et *dé*-N>V (voir exemples 15a-15d). En revanche, les suffixes *-eur*, *-euse* et *-oir* semblent privilégier des bases préfixées par les schémas *é*-N>V (voir exemples 16a-16c) et *re*-V>V (voir exemples 17a-17c).

14) a. *-eur* : BOUCHER $_{V}$ > DEBOUCHER $_{V}$ > DEBOUCHEUR $_{N}$
    b. *-euse* : CATIR $_{V}$ > DECATIR $_{V}$ > DECATISSEUSE $_{N}$
    c. *-oir* : CHAUSSER $_{V}$ > DECHAUSSER $_{V}$ > DECHAUSSOIR $_{N}$
    d. *-ant* : GRIPPER $_{V}$ > DEGRIPPER $_{V}$ > DEGRIPPANT $_{N}$
15) a. *-eur* : BROUSSE $_{N}$ > DEBROUSSER $_{V}$ > DEBROUSSEUR $_{N}$
    b. *-euse* : BROUSSAILLE $_{N}$ > DEBROUSSAILLER $_{V}$ > DEBROUSSAILLEUSE $_{N}$
    c. *-oir* : CRASSE $_{N}$ > DECRASSER $_{V}$ > DECRASSOIR $_{N}$
    d. *-ant* : FANE $_{N}$ > DEFANER $_{V}$ > DEFANANT $_{N}$
16) a. *-eur* : BULLE $_{N}$ > EBULLER $_{V}$ > EBULLEUR $_{N}$
    b. *-euse* : BOUE $_{N}$ >EBOUER $_{V}$ > EBOUEUSE $_{N}$
    c. *-oir* : GOUTTE $_{N}$ > EGOUTTER $_{V}$ > EGOUTTOIR $_{N}$
17) a. *-eur* : AFFINER $_{V}$ > RAFFINER $_{V}$ > RAFFINEUR $_{N}$
    b. *-euse* : AMASSER $_{V}$ > RAMASSER $_{V}$ > RAMASSEUSE $_{N}$
    c. *-oir* : POUSSER $_{V}$ > REPOUSSER $_{V}$ > REPOUSSOIR $_{N}$

Enfin, dans le cadre de la compétition morphologique, les deux résultats obtenus précédemment indiquent que les préfixations *a*-A>V et *en*-N>V favorisent la différenciation des suffixes instrumentaux, car les bases préfixées par ces suffixes ont une préférence pour les schémas *-eur* et *-euse*. En revanche, les bases préfixées par *dé*-V>V, *dé*-N>V, *é*-N>V et *re*-V>V sont partagées par plusieurs suffixes instrumentaux, ce qui n'est pas efficace pour résoudre la compétition entre eux.

5.3.2.2.2 Suffixation

À l'instar de la préfixation, plusieurs études montrent que les bases dérivées par la suffixation exercent aussi une influence considérable sur le choix des suffixes (Koehl

2012 ; Strnadová 2014 ; Missud & Villoing 2022). Certains ont une propension à choisir des bases suffixées pour construire les dérivés. Pour les noms d'instruments, comme le tableau 5.15 le montre, les suffixes instrumentaux se distinguent par une caractéristique spécifique en la matière. Concrètement, la plupart des instrumentaux résistent aux bases suffixées. A l'opposé, les *-eur* et *-ant* permettent de s'ajouter aux bases suffixées. Le suffixe *-eur* préfère les suffixations N-*ifier*>V, A-*iser*>V, ainsi que N-*iser*>V (voir les exemples 18a-18c). Quant au *-ant*, il a tendance à sélectionner les bases construites par les suffixations A-*iser*>V et N-*iser*>V (voir les exemples 19a-19b). À la lumière des résultats obtenus, nous pouvons interpréter ces tendances en se référant à la théorie de l'ordre des suffixes (Saarinen & Hay 2014) : les *-ant* et *-eur* présentent une propension marquée à sélectionner des bases déjà suffixées, tandis que d'autres suffixes ont tendance à les éviter. La préférence observée dans la sélection des bases, déjà mise en lumière dans mon travail précédent (Lin, soumis (a)), révèle des combinaisons suffixales spécifiques, soutenant ainsi l'autonomie du système morphologique à ce niveau (Aronoff 1994).

Tableau 5.15 Proportion des suffixations par suffixes instrumentaux

suffixe \ suffixation (pourcentage)	N-ique>A	V-et>N	N-et>N	A-ifier>V	N-ifier>V	N-iller>V	A-iser>V	N-iser>V	V-oter>V	A-oyer>V	N-oyer>V	Total
-eur	0	0	0	4	5	1	13	16	0	1	1	41
	0%	0%	0%	10%	12%	2%	32%	39%	0%	2%	2%	100%
-euse	0	1	0	1	0	1	2	2	1	0	0	8
	0%	13%	0%	13%	0%	13%	25%	25%	13%	0%	0%	100%
-ier	1	0	0	0	0	0	0	0	0	0	0	1
	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
-oir	0	0	0	0	0	0	0	1	0	0	0	1
	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	100%
-oire	0	0	0	1	0	0	0	0	0	0	0	1
	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	100%
-ant	0	0	0	2	0	0	7	6	1	1	0	17
	0%	0%	0%	12%	0%	0%	41%	35%	6%	6%	0%	100%
-ette	0	0	2	0	0	0	0	0	0	0	0	2
	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%

18) a. N-*ifier*>V : VIN $_N$ > VINIFIER $_V$ > VINIFICATEUR $_N$

b. A-*iser*>V : STABLE $_A$ > STABILISER $_V$ > STABILISATEUR $_N$

c. N-*iser*>V : VAPEUR $_N$ > VAPORISER $_V$ > VAPORISATEUR $_N$

19) a. A-*iser*>V : STERILE $_A$ > STERILISER $_V$ > STERILISANT $_N$

b. N-*iser*>V : CRISTAL $_N$ > CRISTALLISER $_V$ > CRISTALLISATEUR $_N$

Néanmoins, bien que nous ayons constaté que les *-eur* et *-ant* peuvent souvent être ajoutés aux bases suffixées, cela ne signifie pas que les autres schémas évitent les bases

suffixées. Selon mon travail sur l'ordre des suffixes (Lin, soumis (a)), les suffixes *-ier*/*-ière* sélectionnent de temps en temps des bases suffixées en français moderne, notamment pour les noms de profession (agent), ils pourraient sélectionner les bases suffixées par *-ure* (*par ex.* PEINTURE $_{N}$ > PEINTURIER $_{N}$; BROCHURE $_{N}$ > BROCHURIERE $_{N}$). Il est donc important de souligner que des effectifs faibles dans les bases dérivées par la suffixation ne signifient pas que ces schémas ne peuvent pas s'ajouter aux bases suffixées en français. Par exemple, les effectifs faibles observés dans les bases dérivées par la suffixation pour des *-ier*/*-ière* ne relèvent que du cas des noms d'instrument du français et ne peuvent pas être généralisés à l'ensemble des noms en français. C'est la sémantique qui joue un rôle important pour ce phénomène linguistique.

De plus, il a été observé que les bases suffixées par *-ifier* ont davantage tendance à prendre *-eur* plutôt que *-euse*. Ce phénomène morphologique semble être davantage lié à l'étymologie. Selon mon travail (Lin soumis (b)) portant sur l'influence de l'étymologie sur l'ordre des suffixes et les propriétés morphologiques des suffixes *-eur*/*-euse*/*-rice*, bien que la majorité des suffixes français ne soient plus influencés par l'étymologie lors de la formation des mots, l'effet étymologique persiste pour *-euse*. En effet, du point de vue diachronique, le *-euse* est un schéma populaire, tandis que le schéma *-rice* est un schéma savant (d'origine latine). Donc, sous l'influence de l'étymologie, il est peu fréquent de rencontrer les combinaisons suffixales *-iseuse* ou *-ifieuse*, les formes privilégiées étant *-isatrice* et *-ificatrice*. La combinaison suffixale *-iseuse* n'a été acceptée qu'au XXe siècle et uniquement pour un lexème : RECTIFIEUSE.

Enfin, partant de la perspective de la compétition morphologique, il semble que le comportement du *-eur* soit similaire à celui du *-ant*. Autrement dit, il est possible que les deux schémas soient en compétition. Un test exact de Fisher ($p > 0,05$) ne met pas en évidence de différence significative entre les deux. Cela signifie que les *-ant* et *-eur* sont complètement en compétition parmi ces trois suffixations. Afin de différencier ces deux schémas, il faudrait avoir recours aux autres propriétés linguistiques.

5.3.2.2.3 Conversion

Tout comme la préfixation et la suffixation, selon certaines recherches plus spécialisées, la base construite par la conversion constitue une contrainte cruciale dans la formation des noms du français (Missud & Villoing 2022 ; Missud 2023 ; Villoing & Gueye 2024).

Par exemple, le schéma *-age* tend à privilégier une base verbale convertie. En ce qui concerne les noms d'instruments, comme l'indique le Tableau 5.16, nous avons constaté que la majorité des suffixes favorisent les bases construites par la conversion N/V (voir les exemples 20a-20g). Il convient de noter que les suffixes *-oire* et *-et* atteignent un taux de 100 % lorsqu'il s'agit de sélectionner la base convertie N/V. En revanche, seuls trois schémas permettent de s'ajouter fréquemment à la base convertie A/V, à savoir les *-eur*, *-oir* et *-ant* (voir les exemples 21a-21c). En outre, cette observation confirme l'hypothèse avancée par les auteurs évoqués (Missud et Villoing 2022 ; Missud 2023 ; Villoing et Gueye 2024), selon laquelle certains schémas ont une propension à la base convertie. En d'autres termes, la conversion pourrait constituer la propriété morphologique essentielle permettant de distinguer les schémas.

20) a. *-eur* : ANALYSE $_{N}$ > ANALYSER $_{V}$ > ANALYSEUR $_{N}$

b. *-euse* : MOULIN $_{N}$ > MOULINER $_{V}$ > MOULINEUSE $_{N}$

c. *-oir* : ACCORD $_{N}$ > ACCORDER $_{V}$ > ACCORDOIR $_{N}$

d. *-oire* : BASSIN $_{N}$ > BASSINER $_{V}$ > BASSINOIRE $_{N}$

e. *-ant* : ANESTHESIE $_{N}$ > ANESTHESIER $_{V}$ > ANESTHESIANT $_{N}$

f. *-et* : BLOC $_{N}$ > BLOQUER $_{V}$ > BLOQUET $_{N}$

g. *-ette* : CLAC $_{N}$ > CLAQUER $_{V}$ > CLAQUETTE $_{N}$

21) a. *-eur* : LISSE $_{A}$ > LISSER $_{V}$ > LISSEUR $_{N}$

b. *-oir* : PLAN $_{A}$ > PLANER $_{V}$ > PLANOIR $_{N}$

c. *-ant* : EMULSIF $_{A}$ > EMULSIFIER $_{V}$ > EMULSIFIANT $_{N}$

Tableau 5.16 Proportion des conversions par suffixes instrumentaux

Suffixes \ conversion (pourcentage)	N/V	A/V	Total
-eur	**97 (83%)**	**20 (17%)**	117
-euse	**135 (96%)**	6 (4%)	141
-rice	1 (100%)	0 (0%)	1
-oir	**58 (89%)**	**7 (11%)**	65
-oire	**7 (100%)**	0 (0%)	7
-ant	**11 (61%)**	**7 (39%)**	18
-ante	1 (100%)	0 (0%)	1
-ier	2 (100%)	0 (0%)	2
-ière	2 (67%)	1 (33%)	3
-et	**14 (100%)**	0 (0%)	14
-ette	**13 (93%)**	1 (7%)	14
-on	1 (100%)	0 (0%)	1
-aille	1 (100%)	0 (0%)	1

Enfin, sous l'angle de la compétition morphologique, il apparaît que les bases construites par la conversion N/V engendre une forte concurrence entre les schémas instrumentaux, ce qui s'avère inefficace pour les distinguer et résoudre cette compétition. En raison de sa spécificité à certains schémas, la conversion A/V pourrait, dans une certaine mesure, contribuer à atténuer cette concurrence.

5.3.2.3 Composition et Régression

Hormis les deux types de procédés morphologiques de dérivation mentionnés, peu de noms d'instruments prennent pour base des verbes construits par composition et régression. Plus précisément, concernant la composition, dix schémas néoclassiques sont pris en compte dans la formation des noms d'instruments. Comme l'indique le Tableau 5.17, dans nos données, quatre suffixes s'ajoutent aux bases composées, à savoir les schémas *-eur*, *-euse* et *-ier*. Nous fournissons quelques exemples pour illustrer ce point (voir exemples 22).

22) a. *-eur* : VISER $_V$ > TELEVISER $_V$ > TELEVISEUR $_N$

b. *-euse* : COPIER $_V$ > POLYCOPIER $_V$ > POLYCOPIEUSE $_N$

c. *-ier* : [[TERRE]$_N$-[NEUVE]$_A$]$_N$ > TERRE-NEUVIER $_N$

Tableau 5.17 Proportion des compositions par suffixes instrumentaux

suffixe \ composition (pourcentage)	chél-	électro-	photo-	poly-	télé-	NA	thermo-	Total
-eur	1	1	0	0	2	0	0	4
	25%	25%	0%	0%	50%	0%	0%	100%
-euse	0	0	1	1	0	0	1	3
	0%	0%	33%	33%	0%	0%	33%	100%
-ier	0	0	0	0	0	1	0	1
	0%	0%	0%	0%	0%	100%	0%	0%

Enfin, en ce qui concerne la régression, il existe trois lexèmes construits par la régression, à savoir INTERCEPTEUR, SURVOLTEUR, TREFILEUSE. Nous prenons le lexème SURVOLTEUR pour bien illustrer ce phénomène. Le lexème SURVOLTEUR est formé sur la base verbale SURVOLTER, qui elle-même est dérivée rétroactivement du nom SURVOLTAGE. Il s'agit donc d'un processus de régression, dans lequel un verbe est reconstruit à partir d'un nom existant, afin de servir de base à une nouvelle dérivation. Comme le souligne Dal (2003 : 13), cette stratégie de régression permet de rétablir une base verbale à partir d'un nom d'action lorsque celle-ci est absente du lexique, afin de poursuivre la dérivation.

5.3.3 Espaces thématiques

Dans cette sous-section, nous abordons la propriété morphologique des « espaces thématiques », qui revêtent une importance primordiale pour la morphologie constructionnelle. D'après les recherches menées par plusieurs morphologues (Villoing 2009 ; Bonami, Boyé & Kerleroux 2009 ; Tribout 2010 ; Fradin & Winterstein 2013), certains schémas ont une préférence pour des thèmes spécifiques. Plus précisément, en ce qui concerne les espaces thématiques, on peut distinguer deux catégories : les thèmes verbaux et les thèmes nominaux et adjectivaux. En suivant ce fil conducteur, nous nous attacherons à analyser les thèmes préférentiels associés aux schémas instrumentaux.

5.3.3.1 Thèmes verbaux

En ce qui concerne les noms d'instruments, comme l'illustre le tableau 5.18, la majorité des schémas instrumentaux préfèrent le thème 1 du verbe, c'est-à-dire celui que l'on retrouve à la forme de la première personne du pluriel au présent (voir les exemples 23a-23h). Ce dernier se distingue comme le patron parmi les quatre thèmes présentés dans le tableau 5.18. Par ailleurs, le thème 13 (radical latin) occupe la deuxième position. Les schémas *-eur*, *-rice* et *-oire* s'ajoutent à des radicaux relevant du thème 13 (radical latin) (voir les exemples 24a-24c). Enfin, en dehors des thèmes 1 et 13 (thèmes servant fréquemment à la dérivation, identifiés par les morphologues évoqués), nous avons également identifié le thème 8[41] (radical de la forme au subjonctif 1ère et 2e personne du pluriel) pour les noms en *-eur* (cf. exemple 25a) et le thème 12 (radical de la forme au participe passé) pour les noms en *-euse* (cf. exemple 25b). Notons en particulier que concernant le lexème BOUILLISSEUR, il est irrégulier de sélectionner le thème 8 comme base de la suffixation en *-eur* qui sélectionne de façon régulière les thèmes 1 et 13.

23) a. *-eur* : RAPPORTER $_V$ > RAPPORTEUR $_N$
    b. *-euse* : MOULINER $_V$ > MOULINEUSE $_N$
    c. *-oir* : RASER $_V$ > RASOIR $_N$
    d. *-oire* : PASSER $_V$ > PASSOIRE $_N$
    e. *-ant* : ANESTHESIER $_V$ > ANESTHESIANT $_N$
    f. *-et* : BLOQUER $_V$ > BLOQUET $_N$
    g. *-ette* : SONNER $_V$ > SONNETTE $_N$
    h. *-on* : CASSER $_V$ > CASSON $_N$
24) a. *-eur* : AMPLIFIER $_V$ > AMPLIFICATEUR $_N$
    b. *-rice* : PERFORER $_V$ > PERFORATRICE $_N$
    c. *-oire* : VAPORISER $_V$ > VAPORISATEUR $_N$
25) a. *-eur* : BOUILLIR $_V$ > BOUILLISSEUR $_N$
    b. *-euse* : FRIRE $_V$ > FRITEUSE $_N$

[41] Ici, il convient de signaler que pour le lexème BOUILLISSEUR, la base est BOUILLIR et le radical utilisé pour la dérivation est *bouilliss-*. Ce radical n'est pas explicitement répertorié comme thème dans les espaces thématiques proposés par Bonami (2014), mais sa forme le rapproche du « thème 8 », puisqu'une base en *-iss-* n'apparaît que dans l'imparfait du subjonctif de BOUILLIR. Pour cette raison, nous l'annotons comme relevant du thème 8.

Tableau 5.18 Proportion des thèmes verbaux par suffixes instrumentaux

Suffixes \ Thèmes verbaux (pourcentage)	Thème 1	Thème 8	Thème 12	Thème 13	Total
-eur	**287 (65%)**	1 (0%)	0 (0%)	**156 (35%)**	444
-euse	**305 (100%)**	0 (0%)	1 (0%)	0 (0%)	306
-rice	0 (0%)	0 (0%)	0 (0%)	**11 (100%)**	11
-oir	**199 (99%)**	0 (0%)	0 (0%)	3 (1%)	202
-oire	**24 (89%)**	0 (0%)	0 (0%)	3 (11%)	27
-ant	**71 (100%)**	0 (0%)	0 (0%)	0 (0%)	71
-ante	3 (100%)	0 (0%)	0 (0%)	0 (0%)	3
-ier	4 (100%)	0 (0%)	0 (0%)	0 (0%)	4
-ière	**5 (100%)**	0 (0%)	0 (0%)	0 (0%)	5
-et	**25 (100%)**	0 (0%)	0 (0%)	0 (0%)	25
-ette	**43 (100%)**	0 (0%)	0 (0%)	0 (0%)	43
-on	**12 (100%)**	0 (0%)	0 (0%)	0 (0%)	12
-ail	2 (100%)	0 (0%)	0 (0%)	0 (0%)	2
-aille	2 (100%)	0 (0%)	0 (0%)	0 (0%)	2

Concernant cette préférence, nous nous attachons à l'examiner sous l'angle de l'étymologie. Selon Lieber (2021 : 200), les schémas en français se classent en deux catégories, à savoir les schémas populaires et les schémas savants. Plus précisément, « les premiers proviennent du latin ayant subi l'ensemble des changements phonétiques auxquels le lexique français a été soumis. Les seconds, quant à eux, dérivent du latin savant par le biais d'emprunts ultérieurs dans l'histoire de la langue française. »[42] En effet, d'après les recherches de Huot (2005) et Lieber (2021), il apparaît que les schémas d'origine populaire tendent à s'associer à des radicaux ou schémas également populaires, tandis que les schémas d'origine savante privilégient les radicaux ou schémas savants. C'est pourquoi comme l'illustre le tableau 5.19, les schémas *-eur*, *-euse*, *-oir*, *-oire*, *-ant*, ainsi que d'autres schémas, ont tendance à s'ajouter aux radicaux d'origine populaire, car ces schémas sont considérés comme populaires. En revanche, les schémas *-eur*, *-rice* et *-oire* peuvent également se lier à des radicaux d'origine savante,

[42] Le texte original : "French suffixes can be divided into those that are called 'popular' (in French 'populaire') and those that are called 'learned' (in French 'savant'). The former have descended from Latin undergoing all the sound changes that the vocabulary of French has been subject to. The latter come from scholarly Latin by borrowing later in the history of French. Popular suffixes typically attach to popular roots, and learned suffixes to learned roots."

étant donné qu'ils sont issus d'emprunts au latin. Cela valide la conclusion tirée par les auteurs évoqués (Huot 2005, Lieber 2021, Lin (soumis a)).

De plus, il est important de souligner que sur le plan étymologique, deux schémas se présentent sous les formes *-eur* et *-oire*. Selon Grevisse & Goosse (2016 : 711), le premier est hérité du latin et devient ensuite le suffixe français, tandis que le second réintroduit à l'époque classique via emprunt au latin (*-tor*). Au fil du temps, avec l'évolution de la langue française et l'emprunt de schémas latins, le schéma *-eur*, tant français que latin, a convergé vers une forme unique. De même, le schéma *-oire* en français représente la forme féminine du schéma *-oir*, alors que le schéma *-oire* en latin dérive à la fois des adjectifs et des noms. Tout comme le schéma *-eur*, les deux suffixes *-oire* (*-oria* en forme populaire et *-orius(m)* en forme latine) (TLFi ; Fradin & Winterstein 2013 : 2 ; Grevisse & Goosse 2016 : 184) ont également évolué pour se fondre en une seule forme. C'est pourquoi les schémas *-eur* et *-oire* permettent d'associer des radicaux d'origine à la fois populaire (thème 1) et savante (thème 13).

Enfin, dans le cadre de la compétition morphologique, ces résultats montrent que par rapport au thème 1, le thème 13 est plus efficace pour différencier les schémas instrumentaux, notamment entre les schémas *-euse* et *-rice*. En effet, le thème 13 est plus spécifique aux schémas savants, tandis que la plupart des schémas relèvent des schémas populaires. Sous l'effet de la contrainte morphologique (étymologique), de nombreux noms d'instruments sont en concurrence dans le thème 1, ce qui n'est pas efficace pour résoudre la compétition morphologique.

5.3.3.2 Thèmes nominaux et adjectivaux

Après avoir analysé les thèmes verbaux, venons-en aux thèmes nominaux et adjectivaux sélectionnés par les schémas formateurs de noms d'instruments. Comme mentionné dans le chapitre 4 de cette thèse, selon Tribout (2010) et Koehl (2012), concernant les thèmes nominaux et adjectivaux, le thème 2 (forme identique à celle féminine des noms et des adjectifs, qui ne portent aucune propriété ni grammaticale ni sémantique) et le thème L (la forme du latin) servent à dériver les noms en français. Comme le tableau 5.19 le montre, la plupart des suffixes instrumentaux choisissent le thème 2 (cf. les exemples 26a-26f). Comme nous l'avons argumenté dans la sous-section §5.3.3.1, l'étymologie influe beaucoup sur le choix des thèmes. Etant donné que

la plupart des suffixes instrumentaux sont les suffixes populaires, ils sélectionnent prototypiquement le thème nominal 2 et le thème adjectival 2.

26) a. *-eur* : DOSSE $_N$ > DOSSEUR $_N$

b. *-ier* : PLAFOND $_N$ > PLAFONNIER $_N$

c. *-ière* : CHOCOLAT $_N$ > CHOCOLATIERE $_N$

d. *-et* : CISEAU $_N$ > CISELET $_N$

e. *-ette* : FOURCHE $_N$ > FOURCHETTE $_N$

f. *-on* : BASSE $_N$ > BASSON $_N$

Tableau 5.19 Proportion des thèmes nominaux et adjectivaux par suffixes instrumentaux

Suffixes \ Thèmes nominaux et adjectivaux (pourcentage)	Thème 2	Thème L	Total
-eur	**9 (82%)**	2 (18%)	11
-euse	**5 (100%)**	0 (0%)	5
-oir	**4 (100%)**	0 (0%)	4
-ant	1 (33%)	0 (0%)	3
-ier	**67 (100%)**	0 (0%)	67
-ière	**55 (95%)**	3 (5%)	58
-et	**10 (100%)**	0 (0%)	10
-ette	**13 (93%)**	1 (7%)	14
-on	**8 (100%)**	0 (0%)	8
-ail	1 (100%)	0 (0%)	1

Il convient de noter que certains thèmes L sont sélectionnés par un nombre restreint de suffixes. Cependant, ce phénomène n'est pas déterminé morphologiquement, mais phonologiquement. À titre d'exemple, pour le lexème NAZIERE, sa base nominale est NEZ. Toutefois, pour le suffixe *-ière*, il choisirait le thème 2 en raison de son étymologie (populaire). Cependant, la forme phonologique attendue *NEZIERE, aurait conduit à une consécution de phonèmes trop proches. Sous cette contrainte phonologique, il recourt au thème latin *naz-*. Cela valide l'idée de Roché (1998), selon laquelle concernant l'évitement de la consécution de phonèmes trop proches, il serait préférable d'utiliser des radicaux supplétifs, comme dans le cas de PLUIE $_N$ > PLUVIER $_N$.

Étant donné que la majorité des schémas privilégient le thème nominal 2 pour former des noms d'instruments, cela accentue la compétition morphologique. Ainsi, il nous apparaît que les thèmes nominaux et adjectivaux ne sont pas des critères déterminants pour différencier les suffixes instrumentaux.

5.4 Au niveau sémantique

Cette sous-section se concentre sur les propriétés sémantiques de la base des dérivés. Comme cette thèse s'inscrit dans le cadre de la linguistique cognitive, laquelle met un accent particulier sur les connaissances encyclopédiques, nous dirigerons notre attention vers le profil sémantique assumé par la base du dérivé ainsi que sur les sous-classes ontologiques du dérivé.

En ce qui concerne le profil sémantique assumé par la base du dérivé, nous adoptons les propositions que nous avons avancées dans les sections précédentes (cf. ce que nous avons appelé le cadre de la Néo-Morphologie Cognitive (NMC)). Pour les noms déverbaux, l'analyse repose sur l'identification sémantique activée par le profilage conceptuel, c'est-à-dire le rôle que la base verbale assume dans la trajectoire cognitive menant à la formation du dérivé. Pour les noms dénominaux, nous considérons non seulement le rôle sémantique que la base occupe dans la structure évènementielle potentielle, mais aussi les mécanismes cognitifs impliqués, tels que la métaphore conceptuelle ou la métonymie, qui peuvent motiver la sélection de la base lorsque celle-ci n'est pas directement liée à un argument du verbe sous-jacent. Enfin, nous analyserons les sous-classes ontologiques du dérivé afin d'explorer celles qui sont préférentiellement associées aux schémas instrumentaux. Dans le cadre de la NMC, ces sous-classes sont issues du rôle formel de la structure de qualia lors du profilage conceptuel, puis projetées dans le système morphologique au cours de la formation du dérivé.

5.4.1 Profil sémantique assumé par la base du dérivé

Conformément à l'architecture du modèle, cette analyse sera structurée autour des deux principaux types de trajectoires cognitives identifiées : d'une part, la trajectoire cognitive verbale, qui concerne les noms déverbaux ; d'autre part, la trajectoire cognitive nominale, qui concerne les noms dénominaux et les noms désadjectivaux. Chaque partie met en lumière les mécanismes de sélection, de profilage

et de projection sémantique propres à chaque type de base dans la suffixation des noms d'instruments.

5.4.1.1 Trajectoire cognitive verbale

Selon la Structure de Qualia (Pustejovsky, 1996), il existe quatre types de verbes correspondant à quatre rôles : le verbe constitutif, le verbe formel, le verbe agentif et le verbe télique (voir la section 3.1.2). D'après Faber (2012 : 155), pour les noms d'instruments, le verbe télique est plus productif que le verbe formel pour désigner les référents de ces noms. Comme l'illustre le tableau 5.20, la majorité des schémas instrumentaux privilégient les verbes téliques comme bases verbales (voir exemples 27a à 27h). En revanche, un certain nombre de noms construits avec les suffixes *-et* et *-ette* utilisent le verbe formel pour former des noms d'instruments (voir 28a et 28b). Cette observation confirme les résultats de Faber (2012 : 155).

Tableau 5.20 Proportion des verbes dans la Structure de Qualia par suffixes instrumentaux

Suffixe \ Verbe de Qualia (Pourcentage)	Verbe Télique	Verbe Formel	Total
-eur	**423 (95%)**	21 (5%)	444
-euse	**295 (96%)**	11 (4%)	306
-rice	**10 (91%)**	1 (9%)	11
-oir	**197 (98%)**	5 (2%)	202
-oire	**25 (93%)**	2 (7%)	27
-ant	**68 (97%)**	3 (4%)	71
-ante	2 (67%)	1 (33%)	3
-ier	3 (75%)	1 (25%)	4
-ière	**5 (100%)**	0 (0%)	5
-et	**19 (76%)**	**6 (24%)**	25
-ette	**35 (81%)**	**8 (19%)**	43
-on	**11 (92%)**	1 (8%)	12
-ail	2 (100%)	0 (0%)	2
-aille	2 (100%)	0 (0%)	2

Dans le cadre de la NMC, cette répartition s'explique par les trajectoires cognitives verbales différentes. Lors de la construction d'un nom d'instrument à partir d'une base verbale, le rôle télique tend à être cognitivement le plus saillant et par conséquent privilégié comme base dans la dérivation. À l'inverse, certains schémas suffixaux comme ceux en *-et* et *-ette* privilégient une trajectoire cognitive différente, en profilant le rôle formel de la structure de qualia, c'est-à-dire en mettant l'accent sur le moyen de fonctionner plutôt que sur sa finalité.

Nous prenons les exemples 27 et 28 pour illustrer ces phénomènes. L'exemple 27e BOUILLOIRE désigne un « récipient métallique pansu, muni d'un bec et d'une anse, destiné à faire bouillir de l'eau. » (Le Grand Robert, 2005). Ici, le verbe BOUILLIR incarne la fonction télique, orientée vers l'objectif d'utilisation de l'instrument : la base verbale BOUILLIR est sélectionnée comme la base morphologique et projetée dans le système morphologique *via* la trajectoire cognitive fondée sur le rôle télique. À l'opposé, l'exemple 28b SONNETTE désigne un « petit instrument métallique qui sonne pour avertir. » (Le Grand Robert, 2005). Dans ce cas, en tant que la base morphologique, le verbe SONNER représente une propriété distinctive de fonctionnement, soit le rôle formel, et la base est sélectionnée en fonction de cette caractéristique, conformément à une trajectoire cognitive fondée sur le rôle formel.

27) a. *-eur* : ALESER $_V$ > ALESEUR $_N$
    b. *-euse* : AGRAFER $_V$ > AGRAFEUSE $_N$
    c. *-rice* : TRADUIRE $_V$ > TRADUCTRICE $_N$
    d. *-oir* : ALLUMER $_V$ > ALLUMOIR $_N$
    e. *-oire* : BOUILLIR $_V$ > BOUILLOIRE $_N$
    f. *-et* : CLAQUER $_V$ > CLAQUET $_N$
    g. *-ette* : LORGNER $_V$ > LORGNETTE $_N$
    h. *-on* : HARPER $_V$ > HARPON $_N$
28) a. *-et* : NOUER $_V$ > NOUET $_N$
    b. *-ette* : SONNER $_V$ > SONNETTE $_N$

Il convient de noter que ce résultat n'est pas surprenant, car, selon les définitions des noms d'instruments dans une perspective ontologique (Flaux & Van de Velde, 2000), la fonctionnalité constitue la caractéristique la plus importante des artefacts. Les noms d'instruments représentent une sous-classe des noms d'artefacts. C'est pourquoi le verbe télique est plus rentable pour la formation des noms d'instruments surtout pour les noms déverbaux.

Étant donné que la majorité des bases verbales sont des verbes téliques, ce type de verbe s'avère peu efficace pour différencier les suffixes instrumentaux. Néanmoins, étant donné que les suffixes *-et* et *-ette* privilégient la base qui est le verbe formel, il est

possible d'affirmer que ce type de verbe est dans une certaine mesure utile pour distinguer les suffixes *-et*/*-ette* des autres suffixes instrumentaux.

5.4.1.2 Trajectoires cognitives nominale et adjectivale

Après avoir analysé les noms déverbaux, nous nous penchons désormais sur les noms dénominaux et désadjectivaux. D'après le tableau 5.21, six rôles sémantiques et deux projections cognitives sont pris en considération, à savoir : Patient, Cible, Thème, Lieu, Bénéficiaire, Moyen, Métaphore et Métonymie. Parmi ces rôles sémantiques et ces projections cognitives, le Patient, la Cible, la Métaphore et la Métonymie jouent un rôle prépondérant dans la formation des noms d'instruments.

Tableau 5.21 Proportion des rôles sémantiques et des projections cognitives par suffixes instrumentaux

Suffixes \ Rôle sémantique & Projections cognitives (pourcentage)	Patient	Cible	Thème	Lieu	Bénéficiaire	Moyen	Métaphore	Métonymie	Total
-eur	**8**	0	0	0	0	0	0	3	11
	73%	0%	0%	0%	0%	0%	0%	27%	100%
-euse	3	0	0	0	0	1	0	1	5
	60%	0%	0%	0%	0%	20%	0%	20%	100%
-oir	1	0	0	0	0	0	0	3	4
	25%	0%	0%	0%	0%	0%	0%	75%	100%
-ant	0	0	0	0	0	0	0	1	1
	0%	0%	0%	0%	0%	0%	0%	100%	100%
-ier	**39**	8	1	3	0	0	0	**16**	67
	58%	12%	1%	4%	0%	0%	0%	**24%**	100%
-ière	**40**	5	0	1	0	0	0	**12**	58
	69%	9%	0%	2%	0%	0%	0%	**21%**	100%
-et	3	0	0	0	0	0	0	**9**	12
	25%	0%	0%	0%	0%	0%	0%	**75%**	100%
-ette	2	0	0	2	1	0	1	**8**	14
	14%	0%	0%	14%	7%	0%	7%	**57%**	100%
-on	3	0	1	0	0	0	1	3	8
	38%	0%	13%	0%	0%	0%	13%	38%	100%
-ail	0	0	0	0	0	0	0	1	1
	0%	0%	0%	0%	0%	0%	0%	100%	100%

Plus précisément, les suffixes *-eur*, *-ier* et *-ière* privilégient la base nominale correspondant au Patient, comme le montrent les exemples 29a à 29c. De plus, nous avons observé que le suffixe *-ier* a une légère préférence pour la base nominale correspondant à la Cible (exemple 30a). Selon la NMC, cela signifie que les suffixes *-eur* et *-ière* préfèrent les bases portées par la trajectoire nominale fondée sur le Patient, tandis que le suffixe *-ier* peut sélectionner des bases portées tant par une trajectoire fondée sur le Patient que par celle fondée sur la Cible, la première restant

cependant majoritaire. Enfin certains suffixes se rencontrent fréquemment avec la base nominale ayant subi une métonymie : en particulier, *-et* et *-ette* (31c-d), et dans une moindre mesure *-ier* et *-ière* (31a-b). Nous présentons un exemple représentatif dans chaque cas pour illustrer ces phénomènes.

Le DOSSEUR (exemple 29a) désigne un « Machine pour traiter les dosses. » (Le Grand Robert, 2005). D'après cette définition, le verbe TRAITER représente la fonctionnalité de DOSSEUR. Par rapport au verbe télique, DOSSE occupe le rôle de Patient dans l'événement. Ce rôle, mis en saillance lors du profilage conceptuel, conduit à la sélection de DOSSE comme base morphologique. Cette base est ensuite projetée vers le module d'intégration morphosémantique, où elle est intégrée aux informations linguistiques. Ce n'est qu'après cette intégration que la base peut être prise en charge par le système morphologique, qui réalise la dérivation du lexème DOSSEUR.

Pour le lexème CHIMIQUIER (31a), il s'agit d'un « cargo transportant des produits chimiques. » (Le Grand Robert, 2005). Le verbe télique associé est TRANSPORTER, et le Patient correspondant est les produits chimiques. Il est important de souligner que l'adjectif CHIMIQUE représente la caractéristique principale de ce référent et constitue la qualité déterminante pour le classer parmi les sous-classes des noms d'instruments. Cela illustre une trajectoire cognitive nominale fondée sur une restructuration métonymique (métonymie indirecte), conformément au cadre de la NMC. La base morphologique CHIMIQUE est ainsi sélectionnée et projetée vers le module d'intégration morphosémantique jusque dans le système morphologique, qui réalise la dérivation du lexème CHIMIQUIER.

29) Patient :

a. *-eur* : DOSSE $_{N}$ > DOSSEUR $_{N}$ (Machine pour traiter les dosses.) (Le Grand Robert 2005)

b. *-ier* : HOMARD $_{N}$ > HOMARDIER $_{N}$ (Bateau pour la pêche au homard.) (Le Grand Robert 2005)

c. *-ière* : PELAMIDE $_{N}$ > PELAMIDIERE $_{N}$ (Filet à petites mailles pour la pêche à la pélamide.) (Le Grand Robert 2005)

30) Cible :

a. *-ier* : PETROLE $_{N}$ > PETROLIER $_{N}$ (Navire-citerne construit pour le transport du pétrole en vrac.) (Le Grand Robert 2005)

31) Métonymie :

a. *-ier* : CHIMIQUE $_{A}$ > CHIMIQUIER $_{N}$ (Cargo transportant des produits chimiques (métonymie par propriété saillante : chimique désigne les produits chimiques)) (Le Grand Robert 2005)

b. *-ière* : COQUE $_{N}$ > COQUETIERE $_{N}$ (Ustensile où l'on met à cuire les œufs à la coque (métonymie par propriété saillante : la coque désigne l'œuf à la coque)) (Le Grand Robert 2005)

c. *-et* : LONG $_{A}$ > LONGUET $_{N}$ (Marteau long et fin du facteur de pianos (métonymie par propriété saillante : longuet désigne un objet caractérisé par sa forme longue)) (Le Grand Robert 2005)

d. *-ette* : PETROLE $_{N}$ > PETROLETTE $_{N}$ (Petite voiture à pétrole (métonymie par propriété saillante : pétrolette désigne la petite voiture fonctionnant au pétrole)) (TLFi)

Dans la perspective de la compétition morphologique, à partir du tableau 5.22, nous avons constaté que le rôle de Patient et la projection cognitive de métonymie sont plus déterminants pour différencier les deux groupes de suffixes *-eur*/*-ier*/*-ière* et *-et*/*-ette*. Notamment, la métonymie s'avère plus efficace pour distinguer les deux groupes de suffixes *-eur* et *-ier*/*-ière*/*-et*/*-ette*. Toutefois, en ce qui concerne les profils sémantiques évoqués, certains schémas demeurent en concurrence. Il serait donc nécessaire de recourir à d'autres propriétés linguistiques pour résoudre cette compétition.

Enfin, comme le montrent les tableaux 5.21 et 5.22, la suffixation des noms d'instruments révèle deux grandes tendances : les noms déverbaux privilégient le profil sémantique verbe télique, tandis que les noms dénominaux se répartissent majoritairement entre la projection métonymique et le rôle Patient. Dans le cadre du modèle de Néo-Morphologie Cognitive (NMC), cela indique que, parmi les deux grands types de profilage conceptuel mobilisés par les locuteurs, le profilage interne du référent se manifeste majoritairement par une projection métonymique, tandis que le profilage événementiel privilégie le verbe télique et le Patient. Ce constat révèle une convergence dans la manière dont les locuteurs natifs procèdent à la dénomination des instruments : ils tendent à sélectionner, comme base, les caractéristiques du référent qui véhiculent le plus d'information et de saillance, c'est-à-dire l'attribut ou la propriété la plus marquante du référent (projection métonymique), la fonctionnalité instrumentale

(verbe télique), ainsi que l'entité affectée par l'action (Patient). L'ensemble de ces stratégies cognitives constitue ainsi la motivation principale du choix de la base dans la suffixation des noms d'instruments en français. Cet ensemble met également en lumière ce qui distingue la classe ontologique des noms d'instruments des autres classes ontologiques des noms suffixés, en soulignant la spécificité de leurs mécanismes de dénomination d'un point de vue morphosémantique.

5.4.2 Sous-classes ontologiques du dérivé

Concernant les sous-classes ontologiques du dérivé, selon notre approche (cf. la NMC), l'information ontologique relative à la sous-classe provient de la représentation sémantique du dérivé. Dès la phase d'identification des référents, cette information ontologique est portée jusqu'à la phase finale de compétition morphologique, où elle peut influencer le choix du suffixe. Concrètement, comme il a été souligné dans le chapitre 1, plusieurs auteurs (Dubois 1962 ; Spence 1990 ; Sliwa 1990 ; Rainer 2021 ; Lin 2021) insistent sur le fait que les sous-classes ontologiques sont propres à certains suffixes. En d'autres termes, certains schémas tendent à dériver des sous-classes spécifiques. Comme le tableau 5.22 le montre, hormis l'instrument de captation, les autres cinq sous-classes ontologiques sont propres à certains suffixes. En vue de mieux illustrer les sous-classes ontologiques des noms d'instruments préférentielles par schémas instrumentaux, nous présentons les résultats les plus intéressants dans le tableau 5.23.

Tableau 5.22 Proportion des rôles sémantiques et des projections cognitives par suffixes instrumentaux

suffixe \ Sous-classe ontologique (pourcentage)	Instrument de capatation	Outil	Appareil	Machine	Récipient	Substance chimique	Total
-eur	12	102	61	**234**	17	29	455
	3%	22%	13%	**51%**	4%	6%	100%
-euse	0	22	6	**276**	7	0	311
	0%	7%	2%	**89%**	2%	0%	100%
-rice	0	0	1	**10**	0	0	11
	0%	0%	9%	**91%**	0%	0%	100%
-oir	0	**175**	0	18	13	0	206
	0%	**85%**	0%	9%	6%	0%	100%
-oire	1	**18**	0	1	6	1	27
	4%	**67%**	0%	4%	22%	4%	100%
-ant	1	2	1	0	1	**67**	72
	1%	3%	1%	0%	1%	**93%**	100%
-ante	1	0	0	1	1	0	3
	33%	0%	0%	33%	33%	0%	100%
-ier	1	**26**	1	**35**	8	0	71
	1%	**37%**	1%	**49%**	11%	0%	100%
-ière	0	**29**	0	12	**22**	0	63
	0%	**46%**	0%	19%	**35%**	0%	100%
-et	2	**32**	0	2	1	0	37
	5%	**86%**	0%	5%	3%	0%	100%
-ette	5	**35**	2	12	3	0	57
	9%	**61%**	4%	21%	5%	0%	100%
-on	2	**14**	1	0	2	1	20
	10%	**70%**	5%	0%	10%	5%	100%
-ail	0	3	0	0	0	0	3
	0%	100%	0%	0%	0%	0%	100%
-aille	0	2	0	0	0	0	2
	0%	100%	0%	0%	0%	0%	100%

Tableau 5.23 Sous-classes ontologiques préférentielles par suffixes instrumentaux

Sous-classes ontologiques	Suffixes
Outil	*-eur, -oir, -oire, -ier, -ière, -et, -ette, -on*
Appareil	*-eur*
Machine	*-eur, -euse, -rice, -oir, -ier, -ière, ette*
Récipient	*-oire, -ier, -ière*
Substance chimique	*-ant*

Selon le Tableau 5.23, tout d'abord, les suffixes *-eur* et *-ant* préfèrent respectivement dériver les noms appartenant aux sous-classes ontologiques Appareil et Substance chimique (cf. 32a-32b). Il y a en a d'autres mais en petit nombre (quoique 29 substances chimiques sont dérivées en *-eur*, soit environ 30% des substances chimiques, ce qui n'est pas négligeable). En revanche, concernant d'autres sous-classes ontologiques, elles sont plutôt portées par plusieurs suffixes (cf. 32c-32e).

32) a. Appareil : SCANNER $_V$ > SCANNEUR $_N$

b. Substance chimique : ACIDIFIER $_V$ > ACIDIFIANT $_N$

c. Outil :

-eur : EPLUCHER $_V$ > EPLUCHEUR $_N$

-oir : RASER $_V$ > RASOIR $_N$

-ier : ALOSE $_N$ > ALOSIER $_N$

-ière : BOULON $_N$ > BOULONNIERE $_N$

-et : FAUCHET $_V$ > FAUCHET $_N$

-ette : SONNER $_V$ > SONNETTE $_N$

-on : BASSE $_N$ > BASSON $_N$

d. Machine :

-eur : REFRIGERER $_V$ > REFRIGERATEUR $_N$

-euse : EPANDRE $_V$ > EPANDEUSE $_N$

-rice : PERFORER $_V$ > PERFORATRICE $_N$

-oir : FILER $_V$ > FILOIR $_N$

-ier : HOMARD $_N$ > HOMARDIER $_N$

-ière : FACTURE $_N$ > FACTURIERE $_N$

-ette : PETROLETTE $_N$ > PETROLETTE $_N$

e. Récipient :

-oire : BOUILLIR $_V$ > BOUILLOIRE $_N$

-ier : GARGOUSSE $_N$ > GARGOUSSIER $_N$

-ière : CAFE $_N$ > CAFETIERE $_N$

Cependant, il est important de souligner que, concernant les sous-classes ontologiques Instrument de captation et Récipient, bien que le pourcentage associé au suffixe *-eur* demeure relativement faible, leur occurrence est suffisamment fréquente pour surpasser celle d'autres schémas. Par conséquent, il est nécessaire d'adopter une démarche plus prudente dans l'interprétation de ces résultats (voir les tableaux 5.24 et 5.25). D'après ces deux tableaux, le suffixe *-eur* se révèle être particulièrement productif dans les deux sous-classes ontologiques. Cela indique que le schéma *-eur* englobe l'ensemble des sous-classes ontologiques, ce qui en fait le schéma le plus productif pour la formation de noms d'instruments en français.

Tableau 5.24 Proportion des noms d'Instrument de captation par suffixes

Suffixes	Effectif	Pourcentage
-eur	12	44%
-ette	5	19%
-ant	3	11%
-et	2	7%
-on	2	7%
-oire	1	4%
-ante	1	4%
-ier	1	4%

Tableau 5.25 Proportion des noms de récipient par suffixes

Suffixes	Effectif	Pourcentage
-ière	22	27%
-eur	17	21%
-oir	13	16%
-ier	8	10%
-euse	7	9%
-oire	6	7%
-ette	3	4%
-on	2	2%
-et	1	1%
-ant	1	1%
-ante	1	1%

Enfin, nous abordons la compétition morphologique entre les suffixes instrumentaux dans les sous-classes ontologiques des dérivés. Il apparaît que, hormis le suffixe *-eur*, nous avons identifié des sous-classes ontologiques qui sont propres à certains suffixes. Toutefois, hormis les sous-classes ontologiques Appareil et Substance chimique, d'autres suffixes continuent de rivaliser au sein de plusieurs sous-classes ontologiques. Cette situation complique la résolution de la compétition morphologique entre les suffixes instrumentaux. Par ailleurs, selon mon travail (Lin 2021), la compétition morphologique des suffixes instrumentaux en lien avec les sous-classes ontologiques est fortement influencée par des facteurs socioculturels. Il serait donc nécessaire de recourir à des analyses diachroniques afin d'obtenir une vue d'ensemble de cette compétition au fil du temps et de mieux comprendre les modalités de cette compétition.

5.5 Au niveau diachronique

A la suite des analyses synchroniques menées dans les sections précédentes, nous passons maintenant à l'analyse diachronique des suffixes instrumentaux. Nous avons étudié leur évolution quantitative au fil des siècles en examinant la productivité morphologique des suffixes instrumentaux, en décrivant dans un premier temps les grandes tendances de variation. Cette étape descriptive a été suivie d'une tentative d'interprétation, fondée sur des facteurs à la fois morphologiques et socioculturels. L'analyse des dynamiques diachroniques nous a permis de regrouper les schémas en cinq catégories, selon trois critères principaux : le niveau de productivité morphologique, la tendance évolutive et la répartition temporelle dominante. Cette catégorisation vise à fournir une vue d'ensemble synthétique de leurs évolutions diachroniques respectives.

Tableau 5.26 Répartition diachronique de la productivité des suffixes instrumentaux (XIIe–XXe siècle)

suffixes \ siècles (pourcentage)	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	Total par suffixes
-eur	0	0	1	0	5	14	19	**205**	**207**	451
	0%	0%	3%	0%	13%	15%	15%	**45%**	**40%**	
-euse	0	0	0	0	2	0	3	**110**	**188**	303
	0%	0%	0%	0%	5%	0%	2%	**24%**	**36%**	
-oir	2	7	9	12	9	**38**	**59**	**50**	18	204
	25%	33%	26%	48%	23%	**40%**	**46%**	**11%**	3%	
-ant	0	0	0	0	0	2	6	19	43	70
	0%	0%	0%	0%	0%	2%	5%	4%	8%	
-ier	0	1	3	3	3	13	11	11	25	70
	0%	5%	9%	12%	8%	14%	9%	2%	5%	
-ière	2	2	3	2	4	6	10	26	8	63
	25%	10%	9%	8%	10%	6%	8%	6%	2%	
-ette	0	5	4	1	3	7	5	18	14	57
	0%	24%	12%	4%	8%	7%	4%	4%	3%	
-et	2	6	7	3	5	4	7	1	2	37
	25%	29%	21%	12%	13%	4%	5%	0%	0%	
-oire	0	0	5	2	4	6	4	4	2	27
	0%	0%	15%	8%	10%	6%	3%	1%	0%	
-on	1	0	1	0	3	4	5	3	2	19
	13%	0%	3%	0%	8%	4%	4%	1%	0%	
-rice	0	0	0	0	0	0	0	2	9	11
	0%	0%	0%	0%	0%	0%	0%	0%	2%	
-ante	0	0	0	0	0	0	0	2	1	3
	0%	0%	0%	0%	0%	0%	0%	0%	0%	
-ail	0	0	0	2	1	0	0	0	0	3
	0%	0%	0%	8%	3%	0%	0%	0%	0%	
-aille	1	0	1	0	0	0	0	0	0	2
	13%	0%	3%	0%	0%	0%	0%	0%	0%	
TOTAL par siècles	**8**	**21**	**34**	**25**	**39**	**96**	**129**	**451**	**520**	**1312**

Nous commençons par mesurer la productivité diachronique des quatorze suffixes instrumentaux du français à l'aide de la méthode *realised productivity* (Aronoff &

Lindsay 2014 : 76), dans le but d'analyser leur évolution au fil des siècles. Comme l'indique le tableau 5.26, les valeurs du TOTAL par siècle permettent de diviser approximativement la période allant du XIIe au XXe siècle en trois phases : le XIIe–XVIe siècle, le XVIIe–XVIIIe siècle, et enfin le XIXe–XXe siècle.

Au cours de la première phase (XIIe–XVIe siècle), le nombre de dérivés instrumentaux formés par suffixation reste globalement faible. Les schémas les plus représentés à cette période sont *-oir* et *-et*, tandis que *-ette*, *-oire* et *-eur* apparaissent de manière plus ponctuelle. Ce faible rendement peut s'expliquer par plusieurs facteurs. Premièrement, la couverture incomplète des corpus pour les périodes les plus anciennes limite l'observation des dérivés. Certains peuvent ne pas avoir été conservés ou documentés. Deuxièmement, cette phase correspond au Moyen Âge et au début de la Renaissance française, où les activités scientifiques restent limitées, réduisant ainsi les besoins de dénomination des noms d'instruments.

La deuxième phase (XVIIe–XVIIIe siècle) connaît une augmentation notable du nombre de dérivés instrumentaux. Deux schémas se distinguent particulièrement : la suffixation en *-oir*, déjà productif à l'époque précédente, et celle en *-eur*, qui émerge à partir du XVIe siècle. Le suffixe-*ier* participe également à la dérivation d'un certain nombre de noms. Cette évolution peut être mise en relation avec la révolution scientifique et les débuts de la révolution industrielle : les sciences pénètrent dans la société, engendrant de nouveaux besoins de dénomination instrumentale.

Enfin, la troisième phase (XIXe–XXe siècle) se caractérise par une augmentation marquée du nombre de dérivés instrumentaux, avec 451 lexèmes recensés pour le XIXe siècle et 520 pour le XXe siècle. Les dominants sont *-eur* et *-euse*, tandis que *-oir* subit un déclin progressif. Cette évolution s'inscrit dans le contexte socioculturel façonné par les deux révolutions industrielles. Le plein essor des sciences et des technologies a conduit à la prédominance croissante des sous-classes ontologiques Appareil et Machine, au détriment de la sous-classe ontologique Outil, progressivement reléguée à une position secondaire (Rainer 2021 : 9).

A l'issue de cette analyse diachronique, il est possible de regrouper ces schémas en cinq catégories selon leur profil de productivité et leur évolution diachronique. Cette catégorisation repose sur trois critères principaux :

(1) le niveau de productivité morphologique, évalué en fonction du nombre de dérivés au fil des siècles ;

(2) la tendance évolutive observée sur le plan diachronique, qu'elle soit ascendante, stable, déclinante ou sporadique ;
(3) la répartition temporelle dominante, qui distingue les schémas modernes, historiques, ou à distribution étendue sur plusieurs siècles.

Tableau 5.27 Catégorisation diachronique des suffixes instrumentaux selon leur productivité et leur évolution diachronique

Catégories	Description	Suffixes
Suffixes modernes à productivité élevée	Suffixes très productifs, particulièrement représentés aux XIXe et XXe siècles.	*-eur, -euse*
Suffixes émergents à productivité moyenne	Suffixes à productivité modérée, en progression nette à l'époque moderne.	*-ant*
Suffixes persistants à productivité moyenne et modeste	Suffixes moins représentatifs mais persistants au fil des siècles, avec une tendance relativement stable.	*-ier, -ière, -ette, -et, -oire, -on*
Suffixes historiques à productivité déclinante	Suffixes historiquement productifs, à productivité déclinante à l'époque moderne.	*-oir*
Suffixes éteints ou récents à productivité sporadique	Suffixes à très faible productivité, soit anciens et éteints, soit récents mais sporadiques.	*-rice, -ante, -ail, -aille*

Le tableau 5.27 présente les cinq types de suffixes identifiés à partir de l'analyse diachronique. Ces cinq catégories reflètent la diversité de l'évolution des suffixes instrumentaux : certaines catégories sont associées à une productivité élevée contemporaine, d'autres à un comportement ancien ou discontinu, tandis que quelques catégories présentent une stabilité diachronique relative.

Prenons quelques suffixes comme exemples pour bien illustrer cette catégorisation. Concernant les suffixes *-eur* et *-oir*, leur distinction est bien claire. Le suffixe *-eur*, relativement peu productif avant le XIXe siècle, connaît une productivité élevée à l'époque contemporaine, notamment dans la dénomination des noms d'instruments plus techniques modernes. Le suffixe *-oir*, en revanche, est historiquement très actif, avec un pic aux XVIe–XVIIIe siècles, mais décline nettement depuis le XIXe siècle. Ce sont précisément ces évolutions opposées qui justifient leur appartenance à des catégories distinctes. À titre d'exemple complémentaire, les suffixes *-aille* et *-ail* se révèlent très peu productifs : ils ont dérivé leurs derniers noms d'instruments respectivement aux XIVe et XVIe siècles, et ne sont plus employés à cette fin depuis. Nous les avons donc classés parmi les suffixes éteints.

5.6 Discussion

Ce chapitre se consacre à l'analyse des noms d'instruments construits par les suffixes instrumentaux du français aux niveaux linguistiques (phonologique, morphologique, syntaxique, sémantique), ainsi qu'au niveau de la sélection de la base morphologique dans le cadre de la Néo-Morphologie Cognitive, et à leur analyse diachronique.

Au niveau phonologique, deux variables linguistiques sont prises en compte : la taille du radical de la base du dérivé et le segment phonologique final de ce radical. Concernant **la taille du radical**, la dérivation semble privilégier, de manière idéale, les radicaux dissyllabiques (Plénat 2009 : 63). Cependant, comme nous l'avons constaté dans la section 5.2.1, de nombreux suffixes semblent au contraire favoriser les radicaux monosyllabiques, souvent en raison de contraintes phonologiques. Par ailleurs, les suffixes *-eur*, *-rice* et *-ant* présentent une tendance à s'ajouter à des radicaux trisyllabiques. Plus particulièrement, les suffixes *-eur* et *-ant* peuvent s'ajouter à des radicaux de plus de trois syllabes, ce qui dépasse également la taille idéale attendue d'un radical pour la dérivation en français (Plénat 2009). La divergence principale entre nos résultats et ceux de Plénat (2009) tient au fait que son étude ne prend en compte que quelques suffixes tels que *-oche*, *-ette* ou *-esque*, et n'inclue pas de suffixes majeurs en français. Par exemple, des suffixes comme *-eur* ou *-ant*, qui tendent à s'ajouter à des radicaux plus longs, n'ont pas été intégrés dans son analyse. Par conséquent, les conclusions qu'il en tire présentent certaines limites.

En ce qui concerne **le segment phonologique final du radical**, nos résultats confirment l'existence de segments préférentiels dans la suffixation des noms d'instruments. Par exemple, le schéma *-eur* privilégie les radicaux se terminant par une occlusive, tout en évitant ceux qui se terminent par une nasale ou une approximante. Nos résultats contredisent l'hypothèse de Plénat (2005) et Roché (2003) selon laquelle les contraintes phonologiques relèvent exclusivement du système morphologique et ne varient pas d'un suffixe spécifique à l'autre. En revanche, nos résultats corroborent ceux d'autres auteurs, tels que Koehl (2012 : 186) ou Bonami & Thuilier (2019 : 19-20), qui montrent que certains schémas suffixaux favorisent effectivement certains segments phonologiques finaux. Cela montre que l'hypothèse de Plénat (2005) et Roché (2003) présente des limites, notamment dans les cas spécialisés. Comme pour la

taille du radical, ces divergences s'expliquent principalement par le fait que leurs études se concentrent sur moins de suffixes, tels que les suffixes *-on* et *-ette*, sans recours à une analyse à grande échelle. Leurs conclusions s'en trouvent donc nécessairement limitées.

Au niveau morphologique, nous avons pris en considération trois variables : la classe flexionnelle du verbe servant de base au dérivé, la structure morphologique de cette base, ainsi que les thèmes du lexème de base. Concernant **la classe flexionnelle du verbe de la base du dérivé**, comme l'indique le tableau 5.10, la majorité des suffixes sélectionne les bases verbales appartenant au premier groupe de conjugaison, c'est-à-dire qu'il n'existe pas de classe flexionnelle préférentielle entre le suffixes en jeu. Nos résultats contredisent les recherches de Lapraye (2017), selon lesquelles la classe flexionnelle du verbe servant de base au dérivé exercerait une influence sur le choix des schémas événementiels, certains suffixes étant associés à des classes spécifiques. Cette divergence s'explique, d'une part, par le fait que, selon Eluerd (2009 : 20), le premier groupe de conjugaison est de loin le plus productif en français, regroupant environ 10 000 verbes, ce qui le distingue nettement des deuxième et troisième groupes. D'autre part, il convient de souligner que les schémas événementiels et les schémas instrumentaux relèvent de catégories distinctes et mobilisent des bases verbales différentes, ce qui se reflète également dans la classe flexionnelle du verbe servant de base au dérivé.

En ce qui concerne **la structure morphologique de la base du dérivé**, les résultats obtenus dans la section 5.3.2 montrent que certains suffixes semblent privilégier des structures morphologiques spécifiques. Par exemple, le schéma morphologique *-ant* tend à favoriser des bases verbales suffixées, en particulier celles construites par les suffixations A-*iser*>V et N-*iser*>V. Ce constat renforce, d'une part, les résultats obtenus par plusieurs auteurs (Koehl 2012 ; Bonami & Thuilier 2019 ; Missud et Villoing 2022 ; Missud 2023), selon lesquels le choix des suffixes est influencé par la structure morphologique de la base du dérivé. D'autre part, ce résultat pourrait étayer celui obtenu dans mon propre travail (Lin, soumis (a)) dans le cadre de l'ordre des suffixes : le schéma *-ant* a une propension à sélectionner les bases suffixées et occupe fréquemment la position de suffixe externe dans une combinaison suffixale.

En ce qui concerne **les thèmes du lexème de base**, selon les travaux de plusieurs morphologues (Bonami, Boyé et Kerleroux 2009 ; Tribout 2010 ; Fradin et Winterstein 2013), certains schémas instrumentaux déverbaux présentent une préférence pour des

thèmes spécifiques. Concrètement, le suffixe *-oir* ainsi que le suffixe *-ant* privilégient le thème 1 du verbe, tandis que les suffixes *-eur*/*-rice* préfèrent le thème 13. Le suffixe *-oire* aurait quant à lui deux thèmes préférentiels : le thème 1 et le thème 13 (Fradin et Winterstein 2013 ; Grevisse 2016). Ces résultats sont confirmés dans le présent travail. De plus, nous avons identifié que les suffixes *-euse*, *-ière*, *-et*, *-ette* et *-on* privilégient le thème 1, tandis que le suffixe *-eur*, comme *-oire*, privilégie à la fois le thème 1 et le thème 13. Ces observations renforcent l'hypothèse d'une influence du facteur étymologique dans le choix du thème 13, certains lexèmes construits avec les suffixes *-eur* et *-oire* étant empruntés au latin. D'ailleurs, hormis les sept thèmes identifiés (les thèmes 1 et 13 pour la suffixation, le thème 3 pour la composition, les thèmes 0, 12 et 13 pour la conversion ; deux thèmes nominaux et adjectivaux : le thème 2 et le thème L.) dans la constructions des lexèmes par plusieurs morphologues (Plénat 2008 ; Villoing 2009 ; Bonami, Boyé et Kerleroux 2009 ; Tribout 2010 ; Koehl 2012 ; Fradin et Winterstein 2013), nous avons également identifié le thème 8 (forme du subjonctif imparfait), qui participe à la construction de certains lexèmes, comme BOUILLIR > BOUILLISSEUR. Cela valide l'hypothèse formulée par Aronoff (1994) et par Boyé et Bonami (2003), selon laquelle les radicaux impliqués dans la conjugaison peuvent également participer à la dérivation.

Par ailleurs, comme mentionné dans le chapitre 1, la fonction du genre grammatical des suffixes ne fait pas l'objet d'un consensus parmi les morphologues. En effet, certains auteurs (Sliwa 1990 ; Namer & Villoing 2008 ; Rosenberg 2008 ; Fradin & Winterstein 2013) considèrent les formes masculines et féminines des suffixes comme relevant d'un même suffixe, les formes féminines étant perçues comme des variantes des formes masculines. D'autres, en revanche, soutiennent que la forme féminine constitue un suffixe autonome permettant de dériver des noms d'instruments (Spence 1990 et Wauquier 2020 pour *-eur, -euse et -rice*). Les résultats obtenus dans ce chapitre confirment l'analyse de Spence (1990) et de Wauquier (2020). À titre d'exemple, pour les suffixes *-eur* et *-euse*, le suffixe *-eur* préfère les radicaux plus longs, ceux se terminant par une occlusive, les bases suffixées, les thèmes 1 et 13, ainsi que les dérivés appartenant à la sous-classe appareil et machine, tandis que le suffixe *-euse* privilégie les radicaux plus courts, ceux se terminant par une fricative, les bases construites par conversion, le thème 1, ainsi que les dérivés appartenant à la sous-classe machine. De même, en ce qui concerne les suffixes *-ier* et *-ière*, bien qu'ils partagent la plupart des propriétés linguistiques préférentielles, ils se distinguent par la sous-

classe ontologique de leurs dérivés. Plus précisément, le suffixe *-ier* dérive davantage de noms de machine, tandis que le suffixe *-ière* construit préférentiellement des noms de récipient. Ces deux observations montrent que, par l'affinage des propriétés linguistiques en jeu, il est possible d'identifier des différences systématiques entre la forme masculine et la forme féminine d'un suffixe. Autrement dit, chacune des formes peut effectivement constituer un suffixe autonome pour la dérivation de noms d'instruments, ce qui ouvre la voie à une application élargie à d'autres types de noms.

Au niveau syntaxique, nous n'abordons ici que **la catégorie syntaxique de la base du dérivé**. Comme mentionné dans le Chapitre 1, selon plusieurs auteurs (Roché 1998 ; Dal 1999 ; Anscombre 2015 ; Namer & Villoing 2008 ; Roy et Soare 2012 ; Huyghe et Tribout 2015 ; Grevisse et Goosse 2016 ; Knittel 2017a ; Knittel 2017b), les suffixes *-eur*, *-euse*, *-rice*, *-oir*, *-oire*, *-ant*, *-ante*, ainsi que *-et* et *-ette* préfèrent les bases verbales, tandis que les suffixes *-ier* et *-ière* ont une prédilection pour les bases nominales. Les résultats obtenus dans la section 5.1 confirment ces observations. De plus, nous avons également identifié que le suffixe *-on* privilégie la base verbale. Ces résultats valident aussi le postulat proposé par Rainer (2021 : 9), selon lequel les noms déverbaux sont plus représentatifs parmi les noms d'instruments.

Au niveau sémantique, nous avons analysé deux variables linguistiques : le profil sémantique assumé par la base et la sous-classe ontologique du dérivé. Concernant **le profil sémantique assumé par la base**, il occupe une place importante dans la NMC, car il est étroitement lié au profilage conceptuel et à la trajectoire cognitive (voir section 3.1.3). Comme indiqué dans la section 5.4.1, nous avons observé que trois profils sémantiques se dégagent : le verbe télique, le Patient et la Métonymie. Plus précisément, le verbe télique est le plus représenté dans la suffixation des noms d'instruments, ce qui valide le postulat avancé par Faber (2012 : 155), selon lequel, dans la conceptualisation de l'instrument, le rôle télique est essentiel. En outre, nous avons également identifié que les profils sémantiques Patient et Métonymie sont plus saillants dans les cas de noms dénominaux et désadjectivaux. Ainsi, ces trois profils constituent les profils sémantiques prototypiques et représentatifs dans la suffixation des noms d'instruments en français.

Concernant **la sous-classe ontologique du dérivé,** plusieurs auteurs (Dubois 1962 ; Spence 1990 ; Roché 1998 ; Devos & Taeldeman 2003 ; Rainer 2005 ; Knittel 2017a ; Lin 2021 ; Rainer 2021, etc.) soulignent que certains schémas suffixaux instrumentaux tendent à dériver des sous-classes ontologiques spécifiques (voir tableau

1.3). À titre d'exemple, selon Spence (1990) et Lin (2021), le schéma *-eur* privilégie les noms d'instruments de captation et les noms d'appareil, tandis que le schéma *-euse* préfère dériver les noms de machine. Rainer (2005 : 28) précise que les suffixes *-eur* et *-euse*, dans leur usage instrumental, seraient issus de l'ellipse d'expressions telles que « appareil + adjectif en *-eur* » et « machine + adjectif en *-euse* ». Par ailleurs, le suffixe *-oir* tend à dériver des noms d'outil et des noms d'instruments de captation, tandis que les suffixes *-ier* et *-ière* privilégient les noms de récipient, et le suffixe *-ant* les noms de substances chimiques. Selon les résultats obtenus dans la section 5.4.2, certaines de ces observations sont confirmées, tandis que d'autres ne le sont pas. Les suffixes *-eur* et *-oir* montrent des divergences. Le suffixe *-eur* dérive non seulement des noms appartenant à plusieurs sous-classes telles que les outils, les instruments de captation et les appareils, mais aussi des noms de machines et de substances chimiques. En revanche, le suffixe *-oir* ne dérive que des noms d'outil. La principale cause de ces divergences réside dans le manque de critère homogène pour la classification des sous-classes ontologiques. En effet les auteurs utilisent des critères différents pour classer les noms d'instruments.

Au niveau diachronique, dans la section 5.5, nous avons mis en relief **l'évolution des suffixes instrumentaux au fil du temps**. Comme mentionné dans le chapitre 1, parmi les suffixes instrumentaux analysés, les suffixes *-eur*, *-ant* et *-oir*/*-oire* ont particulièrement attiré l'attention des morphologues. Dubois (1962 : 40-44) souligne que le suffixe *-eur* dénote des noms d'instruments depuis le XIXe siècle, en raison de la révolution industrielle. Toutefois, Spence (1990 : 30) ne partage pas cette opinion. Il a en effet observé que le suffixe *-eur* était déjà utilisé pour former des noms d'instruments avant la révolution industrielle. En revanche, selon Rosenberg (2008 : 158), c'est le suffixe *-ant* qui devient plus productif pour la dénomination des noms d'instruments à partir du XIXe siècle. En outre, selon Spence (1990 : 29), les suffixes *-oir*/*-oire* ont commencé à perdre en productivité à la même période, au moment de l'essor du schéma *-eur*. Les résultats obtenus dans notre étude (voir tableau 5.27) valident l'ensemble de ces observations, à l'exception de celle proposée par Dubois. Nous avons également constaté que, comme le suffixe *-ant*, le suffixe *-euse* est devenu productif pour former des noms d'instruments à partir du XIXe siècle. Ces résultats montrent que la révolution industrielle a eu une influence frappante sur les schémas suffixaux instrumentaux. En d'autres termes, le choix des suffixes pour la formation des noms d'instruments est sensible à des facteurs socioculturels. Par ailleurs,

la divergence entre les conclusions de Dubois (1962) et celles d'autres morphologues, ainsi que de notre propre étude, peut s'expliquer par les conditions techniques de l'époque : à cette période, les ressources telles que les corpus à grande échelle ou les bases lexicographiques n'étaient pas encore disponibles. Il est donc compréhensible que certains cas rares n'aient pas été identifiés.

Enfin, concernant les préférences des propriétés linguistiques et l'évolution diachronique des suffixes des noms d'instruments en français, le présent travail présente certaines limites. Plus précisément, elle s'appuie principalement sur *Le Grand Robert* (2005) comme source de données, ce qui exclut les lexèmes construits les plus récents du XXIe siècle. Il en résulte une vision partielle des tendances actuelles en matière de préférences linguistiques et d'évolution des suffixes. Des recherches futures pourraient s'appuyer sur des corpus ou dictionnaires plus extensifs et plus récents pour approfondir cette perspective.

Chapitre 6 Compétition morphologique entre schémas suffixaux instrumentaux

Dans le chapitre 5, nous avons exploré, d'une part, l'influence des contraintes linguistiques sur la construction des noms d'instruments en français et, d'autre part, les mécanismes dérivationnels dans une perspective de linguistique cognitive. À travers ces analyses, nous avons mis en évidence les propriétés linguistiques préférées par les schémas instrumentaux du français. Nous avons constaté que certaines bases présentant des propriétés spécifiques privilégient certains suffixes instrumentaux. Par exemple, les bases dont le radical appartient au thème 13 tendent à privilégier les suffixes *-eur* et *-rice* plutôt que les suffixes *-oir* et *-euse*. Nous avons également observé qu'en ce qui concerne certaines propriétés, plusieurs suffixes peuvent entrer en compétition. Ces observations suggèrent que les suffixes instrumentaux occupent des niches morphologiques, certaines propriétés jouant un rôle plus discriminant que d'autres dans la différenciation entre suffixes. Ce constat ouvre ainsi la voie à une exploration systématique de la compétition morphologique dans la suffixation des noms d'instruments en français.

Dans ce chapitre, nous nous concentrons sur la compétition morphologique. Selon notre perspective (la NMC), la compétition morphologique joue un rôle important au sein du module de système morphologique. Elle se déclenche sous l'effet de l'interaction entre plusieurs propriétés linguistiques et d'autres facteurs (voir section 3.1.3). C'est cette interaction qui conduit à la sélection du schéma « gagnant », résultat optimal de la compétition. Il apparaît donc nécessaire d'étudier plus en détail la compétition morphologique. Concernant le plan de ce chapitre, tout d'abord, nous procédons aux prétraitement et regroupement des modalités pour chaque variable en vue d'éviter une distribution trop éparpillée des données (§ 6.1). Dans la section 6.2, nous dresserons un panorama des niches morphologiques des suffixes instrumentaux. Nous analyserons ces niches morphologiques ainsi que leurs sous-niches, ce qui nous permettra également de mieux comprendre les niches plus denses des noms d'instruments en français. De plus, dans la section 6.3, nous réaliserons une analyse diachronique. Plus précisément, nous étudierons l'évolution des propriétés les plus distinctes au fil du temps, ce qui

pourrait révéler leur mode de compétition (voir Lin 2021). Ensuite, dans la section 6.4, nous procèderons à des modélisations statistiques visant à mettre en évidence les propriétés les plus distinctives permettant de prédire le choix des suffixes instrumentaux. Avec ces modèles, nous serons en mesure d'identifier les propriétés les plus déterminantes pour différencier les suffixes instrumentaux. Enfin, nous discuterons des résultats obtenus dans notre thèse en les comparant aux conclusions des travaux antérieurs présentés dans les chapitres 1 et 2, afin d'analyser les divergences (§ section 6.5).

6.1 Prétraitement et regroupement des modalités

Afin de faciliter l'analyse descriptive des niches morphologiques des suffixes instrumentaux, nous allons, dans cette section, procéder au prétraitement des données et au regroupement des modalités en jeu. Concernant les variables en jeu, plusieurs modalités initialement très fragmentées ont été regroupées selon des critères phonologiques, morphologiques, étymologiques, afin de faciliter l'interprétation des variables dans les modèles et d'éviter une trop grande dispersion des données. Il convient de noter que ce travail de regroupement s'est révélé également pertinent dans un second temps pour la modélisation prédictive (section 6.4), en assurant une représentation cohérente et interprétable des variables dans les modèles.

Les principes de regroupement sont les suivants :

Phonologie 1 : Types syllabiques de radical de la base du dérivé

Selon les résultats analysés dans la section 5.2.1, six longueurs de radicaux selon le nombre de syllabes ont été identifiées dans la suffixation des noms d'instruments en français. Toutefois, les radicaux comportant plus de trois syllabes se révèlent peu fréquents, ce qui justifie un regroupement visant à limiter la dispersion des données. Nous retenons ainsi trois modalités pour ce niveau phonologique, en regroupant les radicaux de plus de deux syllabes sous une même modalité dite polysyllabique.

- Monosyllabique
- Dissyllabique
- Polysyllabique (incluant de 3 syllabes à 6 syllabes)

Phonologie 2 : Mode d'articulation des segments finaux du radical de la base du dérivé

Les 24 segments finaux du radical de la base du dérivé ont été réorganisés selon leur mode d'articulation, en se fondant sur des classifications phonologiques proposées dans les travaux morphologiques (Plénat 2009 ; Koehl 2012 ; Roché & Plénat 2016; Thuilier et al. 2023). En particulier, les catégories voyelle et approximante ont été regroupées, car elles partagent la même caractéristique articulatoire, à savoir la propriété vocalique. De plus, dans la perspective quantitative, elles se caractérisent par la faible fréquence dans notre base de données. Les quatre nouvelles modalités sont donc :

- Occlusives : /p/, /b/, /t/, /d/, /k/, /g/
- Nasales : /m/, /n/, /ɲ/
- Fricatives : /f/, /v/, /s/, /z/, /ʃ/, /ʒ/, /ʀ/
- Approximantes (voyelles + approximantes) : /l/, /j/, /ɥ/, /w/, /e/, /y/, /i/, /u/

Morphologie 1 : Procédés morphologiques

Les 43 structures morphologiques de la base du dérivé, qui sont présentées dans le chapitre 5, ont été regroupées en cinq modalités, en fonction des procédés morphologiques. Tout d'abord, les modalités verbe simple, nom simple et adjectif simple ont été regroupées sous une même modalité appelée simplex. D'une part, ces schémas morphologiques partagent une même propriété morphologique, à savoir la base simple. D'autre part, il est possible d'identifier leur structure morphologique spécifique à partir de leur catégorie syntaxique : par exemple, un verbe simple correspond à la modalité simplex au niveau MOR1, combinée à la catégorie verbe au niveau SYN. En outre, en raison de leur faible fréquence (cf. tableau 5.13, section 5.3.2), composition et régression ont été regroupées sous une même modalité, intitulée "Autres procédés morphologiques", afin d'éviter une dispersion excessive des données.

- Simplex : regroupement des simplex verbaux, nominaux et adjectivaux
- Préfixation
- Suffixation
- Conversion
- Autres procédés morphologiques : composition et régression

Morphologie 2 : Thèmes

Concernant les thèmes des bases sélectionnés par les suffixes, dans la section 5.3.3, nous avons identifié six thèmes verbaux servant à la construction par suffixation des noms d'instruments en français. Afin d'éviter certaines dispersions au sein de certains thèmes, tels que V8 et V12, nous avons choisi de regrouper ces six thèmes en nous appuyant sur leur étymologie ainsi que sur leur distribution lexicale. Ce reclassement nous a permis de dégager quatre grandes modalités correspondant à ce niveau, que nous présentons comme suit :

- Thèmes d'origine latine et grecque : regroupement de V13 et NL
- Autres thèmes verbaux : V8 et V12 (fréquences marginales, regroupées pour stabilité)
- Thèmes de forme féminine : N2 et A2
- Thème verbal 1 : le radical le plus fréquent dans les lexèmes construits en jeu

Sémantique 2 : Profil sémantique assumé par la base du dérivé

En outre, comme pour les autres niveaux linguistiques abordés précédemment, une forte dispersion est également observée au niveau des profils sémantiques de la base du dérivé. Nous avons donc appliqué le même principe de regroupement, en conservant les modalités les plus fréquentes, telles que verbes téliques, verbes formels, patient et métonymie, et en regroupant les autres modalités moins représentées, telles que thème, lieu, métaphore, etc. Plus précisément, les rôles sémantiques lieu, thème, cible, destinataire et moyen ont été regroupés sous une nouvelle modalité intitulée *autres rôles sémantiques*. De même, les modalités métaphore et métonymie, relevant toutes deux d'un mécanisme de projection cognitive, ont été regroupées sous la modalité *projection cognitive*.

- Verbe télique
- Verbe formel
- Patient
- Autre rôles sémantique : lieu, thème, cible, bénéficiaire et moyen
- Projections cognitives

Datation : Siècles d'apparition

Comme mentionné dans la section 5.5, et d'après les informations recensées dans les deux grands dictionnaires du français, *Le Grand Robert* et le *TLFi*, nous avons identifié des noms d'instruments formés entre le XIIe et le XXe siècle. Afin de renforcer la cohérence de l'analyse, ces neuf siècles d'émergence lexicale ont été regroupés en quatre grandes modalités, en prenant pour critère les principaux tournants socioculturels, tels que les révolutions industrielle et scientifique. Comme l'indique le tableau 5.27, les noms d'instruments apparus avant le XVIIIe siècle sont moins fréquents et présentent une distribution relativement dispersée. C'est au XVIIIe siècle que la France entre véritablement dans l'ère scientifique, notamment grâce à la publication de *l'Encyclopédie* dirigée par Diderot et d'Alembert (1751), et sous l'influence du mouvement des Lumières. Cette dynamique intellectuelle et technologique entraîne, à partir de ce siècle, une forte augmentation du nombre de noms techniques, y compris ceux d'instruments (Gohin 1931 : 312-313 ; Wang & Gai 2013 : 57). Ainsi, nous regroupons la période allant du XIIe au XVIIe siècle « avant l'ère scientifique », et réservons les trois siècles suivants à la période influencée par les révolutions industrielles, sans pour autant opérer une division stricte fondée sur ces révolutions. Ce choix s'explique principalement par les incertitudes liées aux périodes de transition, notamment pour certains lexèmes dont la date d'apparition n'est pas précisément documentée.

- Avant l'ère scientifique : regroupe les siècles XII à XVII
- XVIIIe siècle
- XIXe siècle
- XXe siècle

6.2 Panorama des niches morphologiques

Dans cette section, nous dresserons un panorama des niches morphologiques associées aux suffixes instrumentaux en français. En premier lieu, nous proposons de nouvelles définitions de niche morphologique et de sous-niche morphologique dans cette thèse. Ensuite, nous mesurerons le chevauchement des niches morphologiques des schémas morphologiques dénotant les noms d'instruments. Enfin, nous aurons recours à l'outil UpSetR pour représenter visuellement les niches morphologiques des schémas

instrumentaux, à travers les différentes modalités linguistiques activées dans la construction des noms d'instruments.

Comme nous l'avons expliqué dans le chapitre 2, nous nous sommes inspirés de Missud (2023), selon laquelle chaque niche est constituée d'un ensemble de propriétés linguistiques réparties sur plusieurs niveaux linguistiques. Dans cette optique, nous avons retenu huit variables linguistiques pour analyser la suffixation des noms d'instrument :

(1) la taille phonologique du radical de la base du dérivé (PHON 1) ;

(2) le mode d'articulation des segments finaux de radical de la base du dérivé (PHON 2) ;

(3) la structure morphologique de la base du dérivé (MOR 1) ;

(4) les thèmes de la base du dérivé (MOR 2) ;

(5) la classe flexionnelle de la base verbale (MOR 3) ;

(6) la catégorie syntaxique de la base du dérivé (SYN) ;

(7) les sous-classes ontologiques du dérivé (SEM 1) ;

(8) Profil sémantique assumé par la base dans la représentation sémantique du dérivé (SEM 2).

6.2.1 Définitions de niche et de sous-niche morphologiques proposées

Comme mentionné au chapitre 2, nous avons analysé les définitions de la niche morphologique proposées par Lindsay & Aronoff (2013) et Missud (2023), et en avons souligné les limites. La problématique chez Lindsay & Aronoff (2013) réside dans une vision réductrice : la niche y est conçue comme étant définie par une seule variable linguistique. Quant à Missud (2023), bien qu'elle propose une approche multifactorielle en enrichissant la définition initiale formulée par Lindsay & Aronoff (2013), sa propre définition reste limitée à un ensemble de propriétés distinctives, sans intégrer la perspective hypervolumineuse, telle qu'elle a été formulée dans le cadre écologique par Hutchinson (1957). En particulier, la question cruciale du chevauchement entre les niches n'est pas prise en compte. Ainsi, nous nous inspirons de Hutchinson (1957) pour proposer une nouvelle définition de la niche morphologique, visant à rendre compte de la répartition fonctionnelle des schémas morphologiques ainsi que de leurs relations de compétition.

Pour mieux introduire la nouvelle définition de la niche morphologique dans cette étude, il faudrait avant tout présenter toutes les modalités associées à chaque variable linguistique, activées par les schémas morphologiques en jeu. Autrement dit, ces modalités sont mobilisées pour construire l'ensemble des lexèmes dérivés par ces suffixes. Nous prenons un exemple fictif, illustré à la Figure 6.1, représentant une situation avec cinq variables distinctes, chacune possédant plusieurs modalités qui sont effectivement activées dans la dérivation. Cet exemple ne reflète pas directement les variables et modalités analysées dans ce travail, mais sert uniquement d'illustration. Ces modalités, propres à chaque variable, sont susceptibles de constituer une composante essentielle de la niche morphologique d'un des schémas morphologiques en jeu.

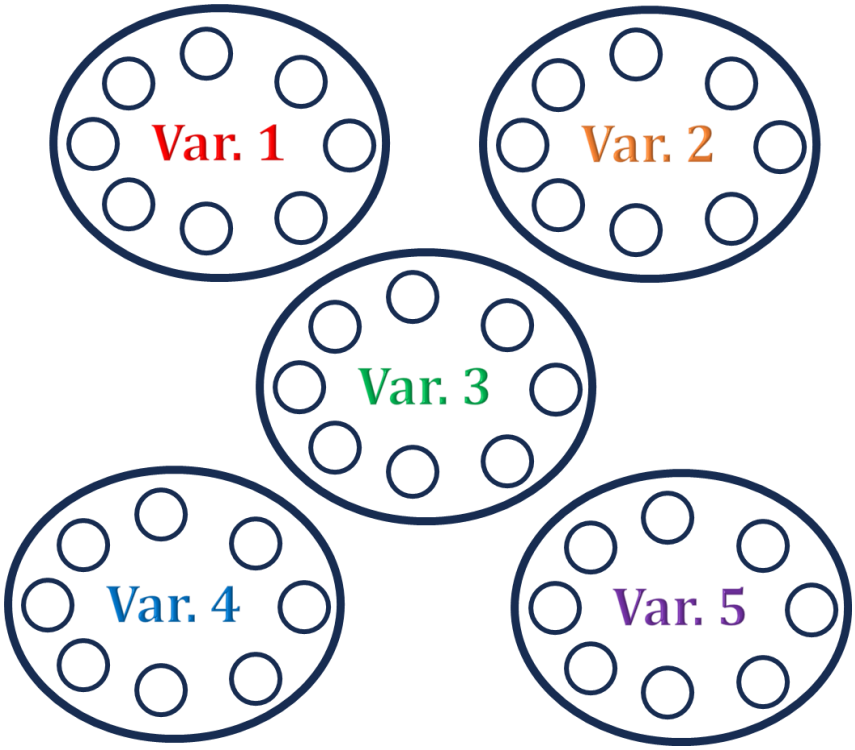

Figure 6.1 Schéma représentant toutes les variables linguistiques avec leurs modalités possibles

Après avoir présenté, à la lumière de la Figure 6.1, les cinq variables ainsi que l'ensemble des modalités qui peuvent être activées dans la dérivation, nous pouvons désormais introduire la notion de niche morphologique. La Figure 6.2 en fournit une première illustration : elle montre comment le schéma morphologique fictif n°1 active un ensemble spécifique de modalités dans chacune des variables, dans au moins un lexème construit par ce schéma morphologique. C'est cet ensemble particulier de modalités activées, représentées en bleu et réparties sur les différentes variables, qui

délimite la niche propre à ce schéma morphologique n°1. Ainsi, la niche morphologique désigne l'ensemble particulier de modalités activées dans au moins un lexème construit par ce schéma morphologique, réparties sur les différentes variables en jeu.

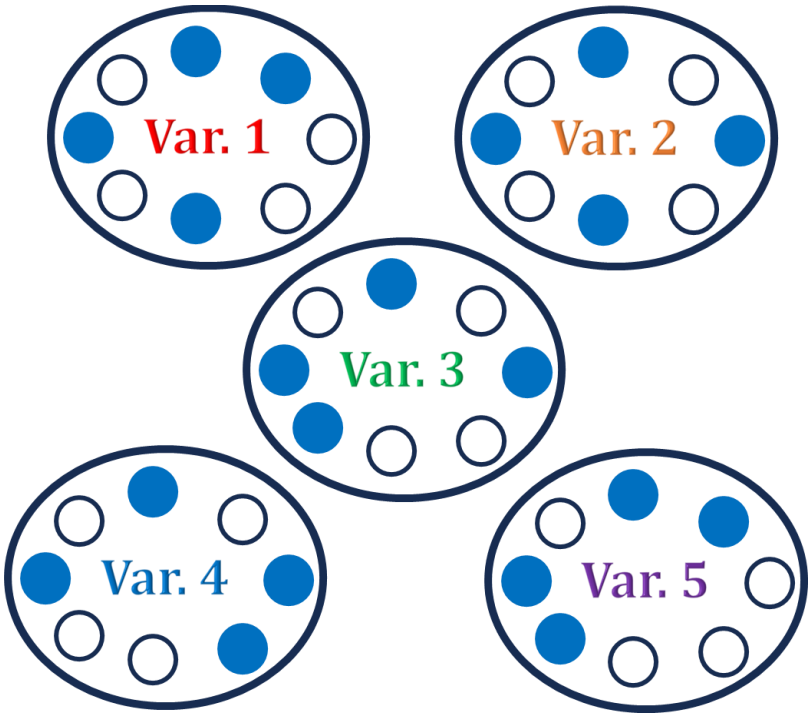

Figure 6.2 Niche morphologique d'un hypothétique schéma morphologique n°1 (représenté en bleu)

Selon cette même logique, nous pouvons à présent identifier la niche d'un schéma morphologique hypothétique n°2, illustrée dans la Figure 6.3. Les modalités activées par ce schéma sont représentées en vert clair, et réparties elles aussi sur les cinq variables évoquées.

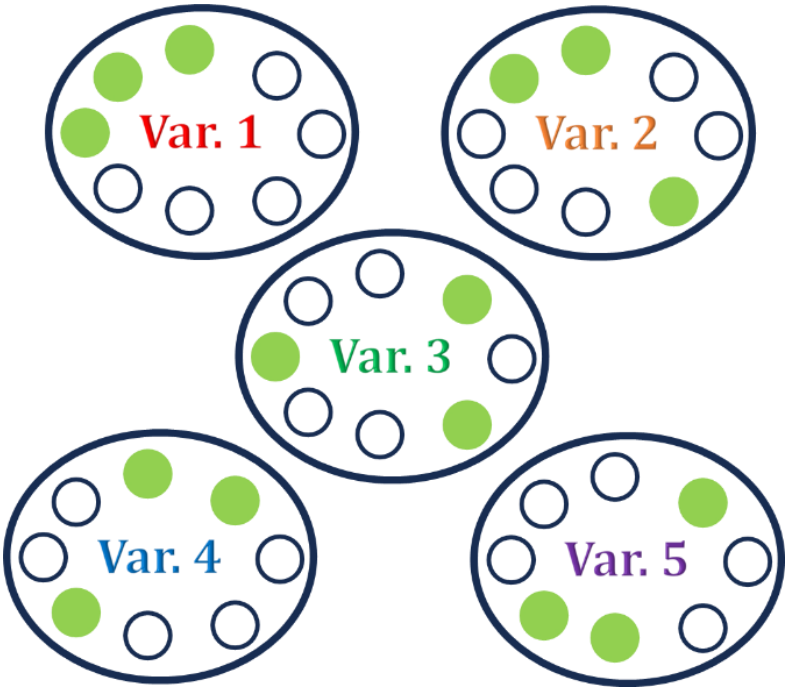

Figure 6.3 Niche morphologique du schéma morphologique n°2 (représenté en vert clair)

En comparant la Figure 6.2 et la Figure 6.3, nous observons que, parmi l'ensemble des modalités activées par tous les schémas en jeu, certaines sont propres au schéma morphologique n°1 (en bleu), d'autres au schéma morphologique n°2 (en vert clair), tandis que plusieurs modalités sont partagées par les deux schémas, représentées en rouge (voir la Figure 6.4). Nous considérons ces modalités communes comme constituant le chevauchement des niches. Par chevauchement des niches, nous entendons ici l'ensemble des modalités activées qui sont partagées par plusieurs schémas morphologiques.

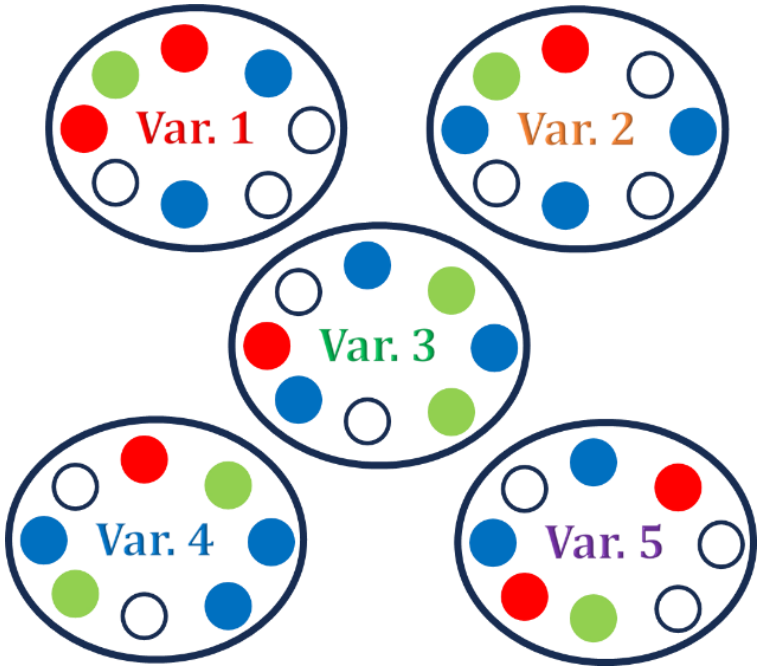

Figure 6.4 Chevauchement des niches morphologique n°1 et n°2 (représenté en rouge)

Il convient de noter que, étant donné que, dans cette thèse, nous examinons principalement la suffixation des noms d'instrument, ce sont les schémas suffixaux dénotant les noms d'instrument qui font l'objet de notre étude, et non les schémas de plus haut niveau, tels que le schéma morphologique *-eur*, dont la sémantique n'est pas encore précisée. Autrement dit, nous analysons essentiellement, dans cette thèse, des niches de schémas morphologiques situés au niveau méso-constructionnel (voir Section 3.1.1).

Après avoir introduit la notion de niche morphologique telle que nous l'entendons dans le cadre de cette thèse, nous allons présenter une autre notion importante : celle de la sous-niche morphologique. En effet, la niche morphologique constitue un concept abstrait : toutes les modalités activées qui la composent proviennent en réalité des différentes sous-niches observables au sein des lexèmes construits par un même schéma morphologique. L'étude des sous-niches permet ainsi de mieux comprendre comment la niche morphologique d'un schéma se construit à partir de modalités activées dans des lexèmes construits, et permet d'identifier les constructions représentatives au niveau inférieur (voir la section 3.1). Concrètement, concernant la sous-niche morphologique, elle désigne une combinaison de modalités, composée d'une seule modalité activée pour chaque variable linguistique, et effectivement réalisée dans au moins un lexème construit par un même schéma morphologique. Deux sous-niches sont

considérées comme distinctes dès lors qu'elles diffèrent par la modalité activée sur au moins une des variables.

À titre d'exemple, les modalités activées dans les lexèmes DANSEUR et PRESENTATEUR sont issues de la niche morphologique du schéma morphologique en *-eur* dénotant les noms d'agents. Ces modalités incluent par exemple la base verbale, une base de type verbe simple, les thèmes sélectionnés par la suffixation (thème 1 et thème 13), le mode d'articulation des segments finaux du radical (occlusive et fricative), etc. Étant donné que certaines modalités, telles que les thèmes ou les modes d'articulation, diffèrent entre ces deux lexèmes, ils relèvent en effet de deux sous-niches morphologiques distinctes. Toutefois, l'ensemble de ces modalités activées contribue à la constitution de schéma morphologique en *-eur* dénotant les noms d'agent.

Enfin, il convient de noter que dans cette thèse, les suffixes instrumentaux étudiés renvoient à un schéma morphologique spécifiquement dénotant les noms d'instruments. Dans ce schéma, hormis l'information sémantique concernant la classe de noms dénotée, les autres informations linguistiques (*par ex.* phonologiques ou morphologiques) ne sont pas remplies. Ainsi, la notion de niche morphologique telle qu'employée ici correspond à la niche d'un suffixe pour la formation des noms d'instruments, et non à la niche complète de ce suffixe en français. Par exemple, le suffixe *-oir* peut dériver à la fois des noms d'instruments et des noms de lieux (Namer & Villoing 2008). Cependant, ce qui est étudié dans le présent travail, c'est son schéma morphologique pour la dérivation des noms d'instruments, et non l'ensemble de sa niche incluant la formation des noms de lieu.

6.2.2 Chevauchement de niches morphologiques

Avant de mesurer le chevauchement de niches des suffixes instrumentaux, nous présentons tout d'abord le nombre de modalités activées par chaque suffixe instrumental.

Tableau 6.1 Classement de nombre de modalités activées selon les schémas morphologiques

Classement	Schéma	Nombre de modalités activées
1	*-eur*	31
2	*-euse*	29
2	*-ette*	29
4	*-ant*	28
5	*-ier*	27
6	*-oir*	26
7	*-et*	24
7	*-oire*	24
9	*-on*	23
10	*-ière*	21
11	*-ante*	16
12	*-ail*	15
13	*-rice*	14
14	*-aille*	10

Le tableau 6.1 présente 14 suffixes morphologiques instrumentaux, classés selon le nombre de modalités linguistiques activées, ce qui indique l'étendue des modalités mobilisées par chaque schéma morphologique dans le lexique. Le schéma *-eur*, avec 31 modalités activées, occupe la première position. Il est suivi de *-euse* et *-ette*, qui en activent chacune 29. Des suffixes en position intermédiaire comme *-ant*, *-ier* ou *-oir* présentent également une large distribution, à savoir de nombreuses modalités activées. En revanche, des suffixes tels que *-ail*, *-rice* et *-aille* activent un nombre plus restreint de modalités, respectivement 15, 14 et 10, ce qui révèle une distribution limitée dans le lexique.

Tableau 6.2 Chevauchement directionnel des niches morphologiques entre les suffixes morphologiques instrumentaux du français

SA \ SB	*-eur*	*-euse*	*-ette*	*-ant*	*-ier*	*-oir*	*-et*	*-oire*	*-on*	*-ière*	*-ante*	*-ail*	*-rice*	*-aille*
-eur		0.97	0.97	**1.00**	0.93	**1.00**	0.96	**1.00**	0.96	0.95	**1.00**	**1.00**	**1.00**	**1.00**
-euse	0.90		0.93	0.89	0.93	0.96	0.88	0.88	0.91	0.95	0.94	**1.00**	0.93	**1.00**
-ette	0.90	0.93		0.93	0.93	**1.00**	0.96	0.96	0.96	**1.00**	1.00	**1.00**	1.00	**1.00**
-ant	0.90	0.86	0.90		0.85	0.92	0.88	0.96	0.91	0.86	0.94	**1.00**	0.93	**1.00**
-ier	0.81	0.86	0.86	0.82		0.85	0.96	0.79	0.96	0.95	0.94	0.93	0.79	0.90
-oir	0.84	0.86	0.90	0.86	0.81		0.92	0.92	0.83	0.95	0.94	**1.00**	0.93	**1.00**
-et	0.74	0.72	0.79	0.75	0.85	0.85		0.79	0.87	0.95	0.94	0.93	0.79	0.90
-oire	0.77	0.72	0.79	0.82	0.70	0.85	0.79		0.74	0.76	**1.00**	0.80	0.93	**1.00**
-on	0.71	0.72	0.76	0.75	0.81	0.73	0.83	0.71		0.86	0.81	0.87	0.64	0.90
-ière	0.65	0.69	0.72	0.64	0.74	0.77	0.83	0.67	0.78		0.75	0.87	0.71	0.80
-ante	0.52	0.52	0.55	0.54	0.56	0.58	0.63	0.67	0.57	0.57		0.67	0.71	0.80
-ail	0.48	0.52	0.52	0.54	0.52	0.58	0.58	0.50	0.57	0.62	0.63		0.57	0.80
-rice	0.45	0.45	0.48	0.46	0.41	0.50	0.46	0.54	0.39	0.48	0.63	0.53		0.60
-aille	0.32	0.34	0.34	0.36	0.33	0.38	0.38	0.42	0.39	0.38	0.50	0.53	0.43	

Le tableau 6.2 illustre le chevauchement directionnel des niches morphologiques entre les suffixes instrumentaux du français. Les cellules indiquent le degré de chevauchement des niches (*SA* représentant le suffixe A, *SB* le suffixe B) (voir la section 4.3.4), conformément à la notion de chevauchement directionnelle proposée dans cette thèse. Le rouge correspond à un chevauchement élevé, tandis que le bleu indique un chevauchement faible.

Précisément, pour les indices de chevauchement analysées, tout d'abord, le schéma *-eur* se distingue par un chevauchement élevé, recouvrant presque totalement les niches des schémas morphologiques en jeu comme *-ant* (1,00), *-ette* (1,00), *-rice* (1,00), et *-aille* (1,00). Cette dominance indique une force compétitive du schéma *-eur* dans le lexique du français. Cependant, cette relation entre les suffixes instrumentaux demeure asymétrique. A titre d'exemple, bien que le suffixe *-eur* recouvre la niche de *-rice* et de *-aille*, ces derniers ne parviennent à recouvrir que très partiellement la niche du suffixe *-eur* (respectivement 0.45 et 0.32). Cette asymétrie révèle que certains suffixes aux modalités réduites sont fragiles face à la dominance compétitive des schémas morphologiques comme *-eur*.

Par ailleurs, certaines relations présentent des chevauchements plus symétriques, par exemple entre *-eur* et *-euse* (0.97 contre 0.90), indiquant une compétition élevée entre ces deux schémas morphologiques. De même, les suffixes *-ette* et *-ant* (1.00 contre 0.90) montrent une compétition étroite, indiquant une similarité forte. Enfin, le

tableau 6.2 montre que des suffixes ayant un nombre limité de modalités, comme *-rice* (14 modalités) et *-aille* (10 modalités), sont très souvent entièrement chevauchés par des suffixes dominants. Cela souligne soit une position fragile, soit une spécialisation plus marquée dans le lexique. En définitive, cette analyse quantitative du chevauchement morphologique met en lumière la répartition des niches morphologiques occupées par les schémas morphologiques instrumentaux en français. C'est précisément l'ampleur de ces zones de chevauchement qui constitue le cœur même de leur concurrence réciproque. Dans la section 6.4, nous nous attacherons donc à identifier, parmi les modalités activées au sein de chaque niche, celles qui s'avèrent les plus discriminantes pour départager les suffixes en compétition.

Une fois cette mesure effectuée, il devient possible de visualiser les niches et sous-niches morphologiques, ce qui permet de mieux rendre compte du panorama des niches morphologiques associées aux schémas instrumentaux du français.

6.2.3 Visualisation de niche et de sous-niche morphologiques

Comme mentionné dans le chapitre 2, les niches morphologiques se caractérisent par un hypervolume, rendant leur visualisation directe difficile. Pour surmonter cette difficulté, nous analysons les dérivés en construisant des sous-niches morphologiques : chacune d'elles correspond à une combinaison spécifique de propriétés phonologiques, morphologiques, syntaxiques et sémantiques, partagée par un ou plusieurs dérivés. Nous utilisons ensuite le package UpSetR pour construire les intersections entre ces propriétés linguistiques, chaque sous-niche étant représentée par l'ensemble des points reliés par une ligne (le point représente la modalité activée). Cette méthode permet ainsi une visualisation indirecte de la structure hyperdimensionnelle des niches morphologiques associées aux suffixes instrumentaux.

Tableau 6.3 Le nombre de sous-niches et le pourcentage selon les schémas instrumentaux

	Nombre de sous-niches	Pourcentage	Nombre total de dérivés	Moyenne dérivé s/sous-niches
eur	186	32.5%	455	2.45
euse	87	15.2%	311	3.57
oir	61	10.6%	206	3.38
ier	47	8.2%	71	1.51
ette	44	7.7%	57	1.30
ière	35	6.1%	63	1.80
ant	35	6.1%	72	2.06
et	24	4.2%	37	1.54
oire	23	4.0%	27	1.17
on	18	3.1%	20	1.11
rice	5	0.9%	11	2.20
ante	3	0.5%	3	1.00
ail	3	0.5%	3	1.00
aille	2	0.3%	2	1.00
Total	573	100.0%	1338	2.34

Le tableau 6.3 présente, pour chacun des 14 schémas instrumentaux étudiés, une série d'indicateurs quantitatifs relatifs à la manière dont chaque suffixe se distribue sur l'ensemble des sous-niches morphologiques identifiées. La colonne "Nombre de sous-niches" indique combien de sous-niches composent la niche morphologique du suffixe concerné, et le pourcentage correspondant exprime leur part relative dans l'ensemble des 573 sous-niches recensées. De plus, la colonne "Nombre total de dérivés" indique combien de lexèmes ont été construits avec ce schéma dans notre base de données. Enfin, la moyenne dérivé / sous-niches permet d'évaluer la densité écologique du suffixe, c'est-à-dire le nombre moyen de dérivés partageant une même sous-niche. En moyenne, la densité des suffixes en jeu est de 2,34, ce qui signifie que chaque sous-niche accueille en général un peu plus de deux dérivés (cf. Tableau 6.3). Cette valeur servira de point de référence dans l'analyse comparative qui suit, notamment pour évaluer si la niche d'un suffixe peut être considérée comme relativement dense ou non.

Il convient de souligner que le ratio élevé signifie que plusieurs dérivés occupent la même sous-niche (forte densité), tandis qu'un ratio proche de 1 reflète une spécialisation écologique : chaque dérivé occupe alors une sous-niche propre. Par

exemple, le suffixe -*euse* (87 sous-niches pour 311 dérivés) présente une densité élevée (3,57), témoignant d'un regroupement de dérivés dans certaines sous-niches, tandis que -*ette* ou -*oire*, avec une moyenne inférieure à 1,5, montrent une spécialisation plus marquée dans l'espace des niches morphologiques. Il ne s'agit pas ici d'une mesure de la productivité morphologique, mais d'un indicateur de répartition écologique : dans quelle mesure les sous-niches sont-elles densément occupées ou au contraire associées de façon exclusive à un seul dérivé.

Étant donné le nombre important de sous-niches morphologiques recensées pour chaque schéma, nous proposons une visualisation graphique permettant de représenter, pour chacun d'eux, la manière dont sa niche morphologique se répartit entre les sous-niches observées dans ses dérivés. Par ailleurs, nous ne retenons dans cette série que les suffixes pour lesquels au moins une sous-niche regroupe plus de cinq dérivés. Cette série de figures permet de mieux comprendre comment les dérivés se répartissent au sein de la niche de chaque suffixe, en fonction des sous-niches qu'ils occupent.

Nous commençons par le suffixe -*eur*. Selon le tableau 6.3, ce suffixe appartient à la catégorie des suffixes dont la niche morphologique est à la fois dense et relativement spécialisée (186 sous-niches pour 455 dérivés, avec une densité de 2,45). Selon la Figure 6.5, deux sous-niches morphologiques plus denses peuvent être identifiées. La première est constituée par les modalités linguistiques activées suivantes : {polysyllabe, plosive, simplex, thème 13, G1, verbe, machine, verbe télique}. Cette sous-niche regroupe 38 lexèmes, tels que CONGELATEUR, INSPIRATEUR, et ASPIRATEUR. La deuxième sous-niche dense repose sur un ensemble très proche de variables : {polysyllabe, plosive, simplex, thème 13, G1, verbe, outil, verbe télique}, et elle est partagée par 21 lexèmes construits, comme DILATATEUR, PERFORATEUR, EXPLORATEUR. Ces deux sous-niches se distinguent par la sous-classe ontologique du dérivé, à savoir machine pour la première et outil pour la seconde.

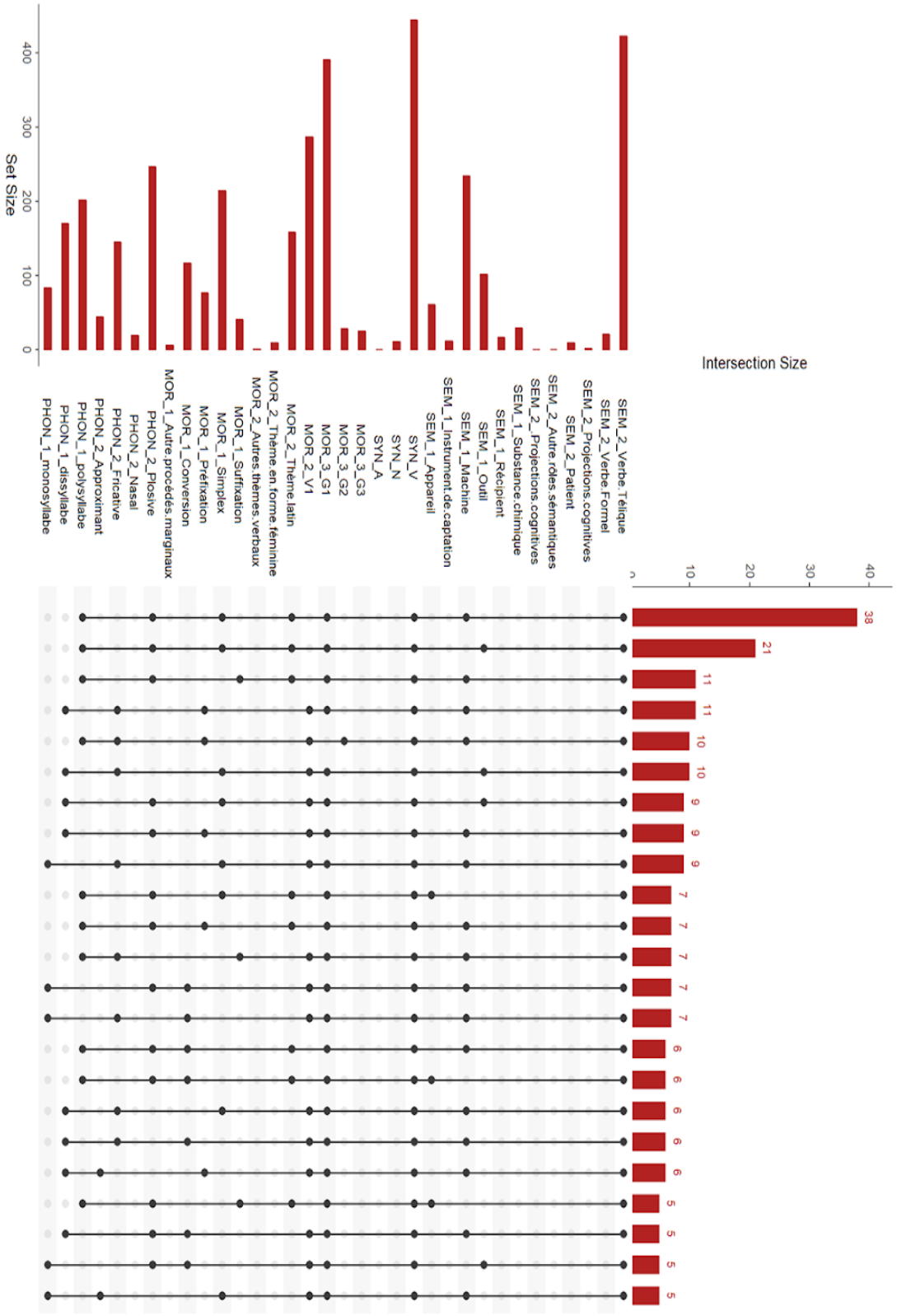

Figure 6.5 La niche morphologique du suffixe *-eur* et ses sous-niches (au moins 5 individus)

Nous poursuivons avec le suffixe *-euse*. Selon le tableau 6.3, ce suffixe appartient à la catégorie des suffixes dont la niche morphologique est très dense (87 sous-niches pour 311 dérivés, avec une densité de 3,57). D'après la Figure 6.6, deux sous-niches

morphologiques plus denses peuvent être observées. La première est constituée par les variables linguistiques suivantes : {monosyllabe, Approximant, conversion, Thème 1, G1, Verbe, Machine, Verbe Télique}, et elle regroupe 23 lexèmes, parmi lesquels COLLEUSE, NIVELEUSE et CRIBLEUSE. La seconde sous-niche dense repose sur un ensemble similaire : {monosyllabe, Fricative, conversion, Thème 1, G1, Verbe, Machine, Verbe Télique}, partagé par des lexèmes tels que CIREUSE, FAUCHEUSE et LISSEUSE. Ces deux sous-niches se différencient par le mode d'articulation des segments finaux du radical de la base du dérivé : Approximant pour la première, et Fricative pour la seconde. Par ailleurs, en plus de ces deux sous-niches les plus denses, la niche morphologique du suffixe *-euse* comporte encore 7 autres sous-niches contenant chacune entre 10 et 19 dérivés. Ce total de 9 sous-niches relativement peuplées, contre 6 pour le schéma *-eur*, renforce notre observation basée sur le tableau 6.2 concernant la densité écologique élevée de *-euse*.

Nous abordons à présent le suffixe *-oir*. D'après le tableau 6.3, ce suffixe appartient à la catégorie des suffixes dont la niche morphologique est très dense (61 sous-niches pour 206 dérivés, avec la densité 3,38). La Figure 6.7 révèle plusieurs sous-niches bien peuplées. La plus dense regroupe 21 dérivés et active des modalités telles que {monosyllabe, Fricative, conversion, Thème 1, G1, Verbe, Outil, Verbe Télique}, illustrée par des lexèmes comme HACHOIR, FICHOIR ou POCHOIR. Une autre sous-niche rassemble 17 dérivés, fondée sur un ensemble proche de modalités activées, notamment {monosyllabe, Fricative, simplex, Thème 1, G1, Verbe, Outil, Verbe Télique}, incluant par exemple GRAVOIR, PERÇOIR ou RASOIR. Ces deux sous-niches se différencient notamment par les structures morphologiques de la base du dérivé : la conversion pour la première, et le simplex pour la seconde. Au total, la niche morphologique du suffixe *-oir* compte 7 sous-niches contenant chacune entre 10 et 21 dérivés, ce qui renforce notre observation basée sur le tableau 6.1 concernant la densité élevée de ce schéma.

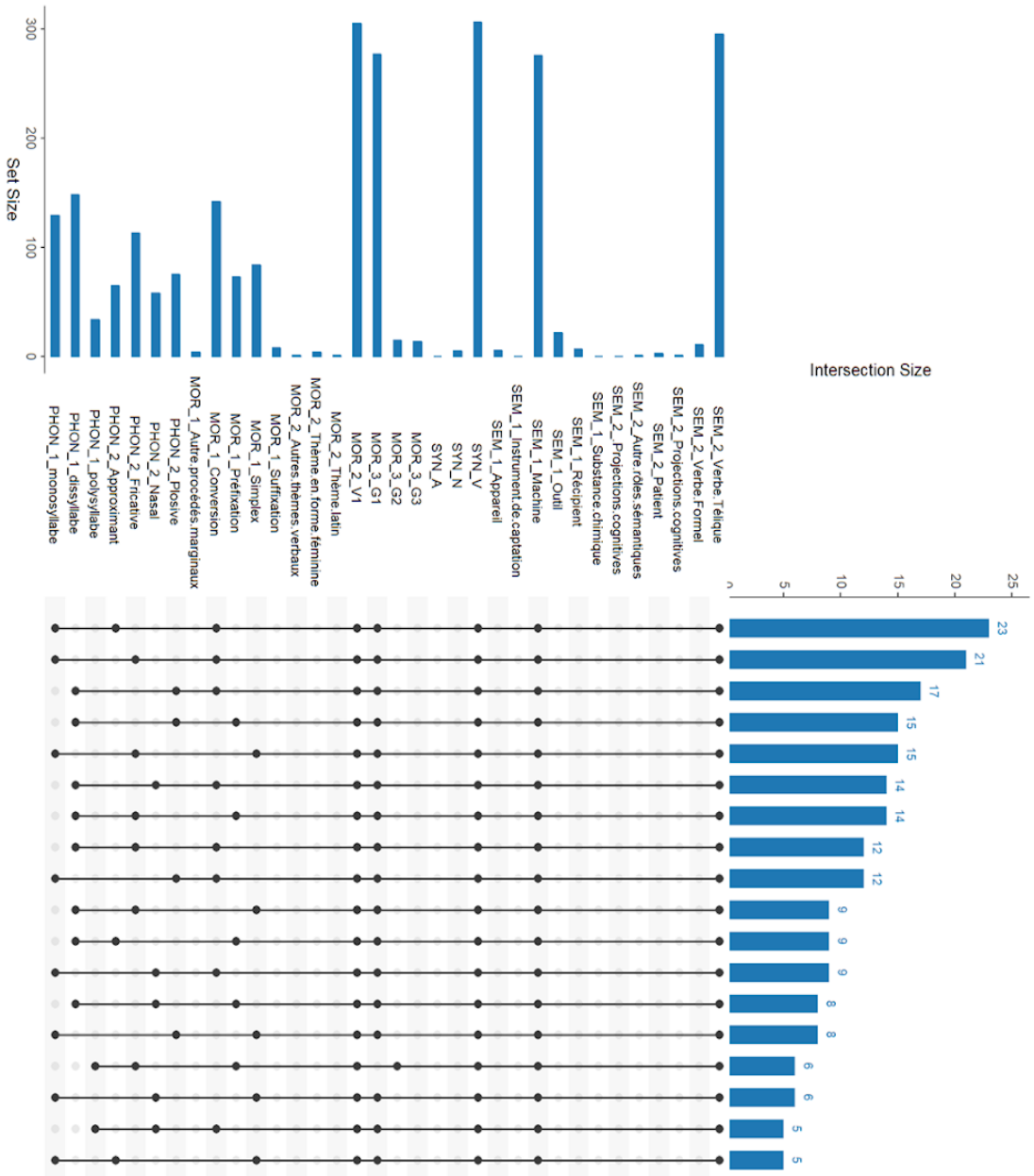

Figure 6.6 La niche morphologique du suffixe -*euse* et ses sous-niches (au moins 5 dérivés)

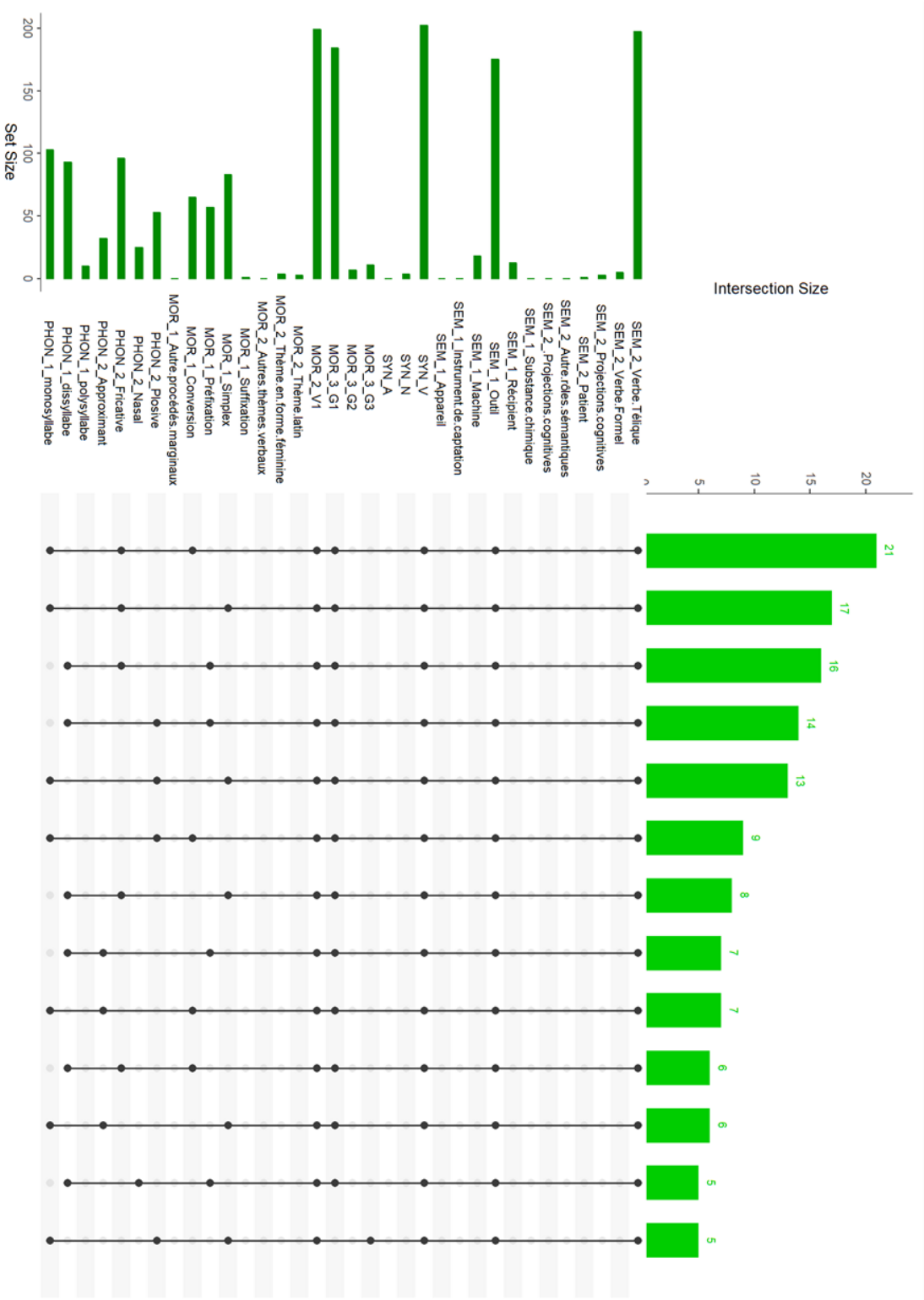

Figure 6.7 La niche morphologique du suffixe -*oir* et ses sous-niches (au moins 5 dérivés)

Nous examinons à présent le suffixe -*ier*, d'après le tableau 6.3, ce schéma appartient à la catégorie des schémas dont la niche morphologique est peu dense (47

sous niches pour 74 dérivés, avec une densité de 1,51). La Figure 6.8 confirme cette faible densité par la présence de plusieurs sous-niches à densité limitée, sans qu'aucune ne domine nettement. Il existe trois sous-niches, qui regroupent plus de 5 lexèmes. Plus précisément, la sous-niche la plus dense regroupe 6 dérivés, construite sur les modalités suivantes : {dissyllabe, Plosive, simplex, Thème 2, nom, Machine, Patient}, illustrée par des exemples comme CREVETTIER, HOMARDIER ou HARENGUIER. Deux autres sous-niches regroupent 5 dérivés chacune, tandis que la majorité des combinaisons activées ne sont partagées que par un ou deux lexèmes. La dispersion des dérivés parmi de nombreuses sous-niches à faible densité, sans schéma dominant clairement identifiable, confirme notre observation issue du tableau 6.3 : le suffixe *-ier* présente une niche morphologique peu dense.

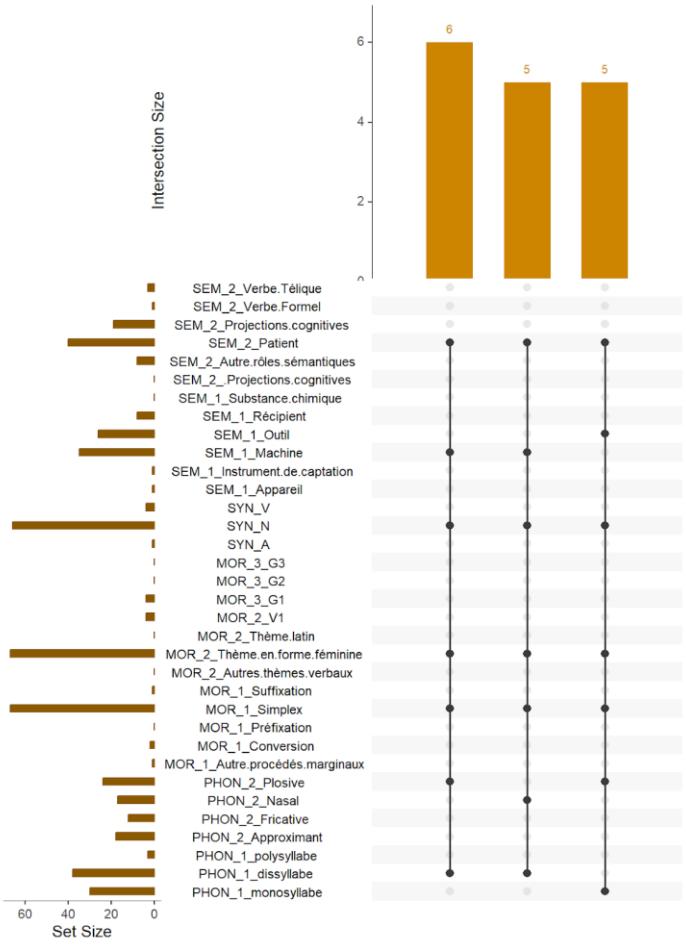

Figure 6.8 La niche morphologique du suffixe *-ier* et ses sous-niches (au moins 5 dérivés)

Concernant le suffixe *-ière*, le tableau 6.3 montre qu'il appartient à la catégorie des suffixes dont la niche morphologique est moyennement dense (35 sous niches pour

63 dérivés, avec une densité de 1,80). Précisément, il existe seulement une sous-niche, qui regroupe 5 dérivés. Concrètement, cette niche est constituée par les modalités suivantes : {monosyllabe, Plosive, simplex, Thème 2, Nom, Outil, Patient}, illustrée par des lexèmes tels que BOGUIERE, GOUTTIERE, JANTIERE, MULETIERE et RATIERE. La présence de quelques sous-niches à densité moyenne confirme notre observation du tableau 6.3 : par rapport au suffixe *-eur* et *-ier*, la niche morphologique du suffixe *-ière* présente une densité intermédiaire.

Comme le suffixe *-ière*, le suffixe *-ant* appartient à la catégorie des suffixes dont la niche morphologique est moyennement dense (35 sous niches pour 74 dérivés, avec une densité de 2,06). Concrètement, il existe seulement deux sous-niches qui regroupent plus de 5 dérivés. Parmi elles, la sous-niche la plus dense regroupe 13 dérivés. Cette sous-niche est constituée par les modalités linguistiques suivantes : {polysyllabe, Fricative, suffixation, Thème 1, G1, Verbe, Substance chimique, Verbe Télique}, et elle est illustrée par des lexèmes tels que FERTILISANT, FLEGMATISANT et IMPERMEABILISANT. Cette sous-niche se distingue notamment des autres sous-niches de suffixe *-ant*, beaucoup moins peuplées, ce qui explique que la densité globale du suffixe *-ant* reste moyenne malgré une forte concentration sur une seule sous-niche.

Hormis les suffixes évoqués, les autres suffixes instrumentaux, tels que *-et, -ette, -oire, -on, -rice, -ante*, *-ail* et *-aille*, présentent une distribution particulièrement dispersée. Pour chacun de ces suffixes, chaque sous-niche est occupée en moyenne par peu de lexèmes. Autrement dit, le nombre de sous-niches est exactement égal au nombre de lexèmes construits. Dans la section 6.3, nous poursuivrons l'analyse diachronique, afin de mieux examiner le mode de compétition morphologique.

6.3 Mode de compétition morphologique

Selon les résultats présentés dans les sections 5.4.2 et 5.5, certains suffixes préfèrent dériver les noms appartenant à certaines sous-classes ontologiques, et montrent par ailleurs des évolutions diachroniques différentes. À partir de ces observations, plusieurs questions importantes émergent : la relation entre un suffixe et une sous-classe ontologique donnée est-elle historiquement stable, ou varie-t-elle selon les époques sous l'effet de dynamiques de compétition ? Autrement dit, la manière dont certains suffixes, comme *-eur* ou *-oir*, s'associent à des sous-classes ontologiques telles

qu'outil ou appareil est-elle restée constante au fil du temps, ou a-t-elle évoluée en fonction des mécanismes de compétition morphologique ? Par ailleurs, le principe d'exclusion compétitive (cf. Chapitre 2) formulé par Lindsay & Aronoff (2013) est-il réellement applicable de façon générale au système morphologique ? En d'autres termes, l'émergence d'un schéma implique-t-elle nécessairement la disparition d'un autre ? Ou bien observe-t-on, au contraire, des formes de coexistence partielle rendues possibles par une différenciation des niches, notamment à travers la spécialisation sémantique ou le redéploiement fonctionnel ? La coexistence prolongée entre schémas concurrents est-elle vraiment impossible ? Ces interrogations nécessitent une réflexion approfondie ainsi qu'une vérification empirique.

Pour répondre à ces questions, nous menons dans cette section une analyse diachronique systématique, fondée sur l'évolution de la productivité des schémas dans les six principales sous-classes ontologiques des noms d'instrument identifiées (cf. section 5.4.2). Chaque figure présente l'évolution de la productivité suffixale du XIIe au XXe siècle et permet d'observer les modes de compétition morphologique propres à chaque sous-classe ontologique des noms d'instrument suffixés. Il convient de noter que, pour la définition de la productivité morphologique, nous nous appuyons sur la notion de productivité réalisée (*realized productivity*), qui consiste principalement à quantifier le nombre de néologismes construits à partir d'un affixe spécifique pendant une période définie, tels qu'ils sont répertoriés dans un dictionnaire (Aronoff & Lindsay, 2014 : 76).

Après avoir analysé et visualisé la productivité des schémas instrumentaux dans les six sous-classes ontologiques des noms d'instrument, et conformément à notre travail précédent (Lin 2021), nous proposons de classer ces dynamiques selon une typologie articulée autour de trois modes de compétition :

- **Compétition absolue**, mode de concurrence où un schéma devient dominant sur une longue période, les autres restant marginalement productifs ou absents ;

- **Compétition alternative**, mode de compétition marqué par une évolution diachronique dans la répartition des schémas, où un patron prend progressivement le relais d'un autre, sans nécessairement entraîner sa disparition ;

- **Compétition non absolue**, mode de compétition où plusieurs schémas coexistent durablement, sans qu'un patron unique ne s'impose de manière claire ou exclusive.

Cette démarche vise à mieux comprendre les mécanismes évolutifs de la compétition morphologique en français, tout en clarifiant le rôle structurant des propriétés sémantiques dans l'évolution diachronique des dérivés. Dans ce qui suit, en suivant le fil conducteur des trois modes de compétition, nous détaillerons chaque mode ainsi que les sous-classes ontologiques et les schémas qui leur sont associés. Il convient de souligner que les schémas très faiblement représentés, notamment ceux ne présentant en moyenne qu'un ou deux dérivés par siècle, ont été écartés de l'analyse approfondie, afin de concentrer l'interprétation sur les schémas les plus représentatifs en termes de productivité, c'est-à-dire ceux dont la fréquence permet d'observer des tendances exploitables. Seule la sous-classe Instrument de captation a été conservée, malgré une productivité très faible, en raison de son intérêt diachronique particulier.

6.3.1 Compétition absolue

Nous commençons par le mode de compétition absolue, principalement représenté par la sous-classe ontologique Appareil. Concernant la sous-classe Appareil, les résultats de la section 5.4.2 montrent que les noms appartenant à cette sous-classe ontologique principalement construits avec sept schémas instrumentaux du français, parmi lesquels le suffixe *-eur* se révèle le plus productif. Comme l'illustre la Figure 6.9, ces observations viennent confirmer les résultats présentés dans la section 5.4.2.

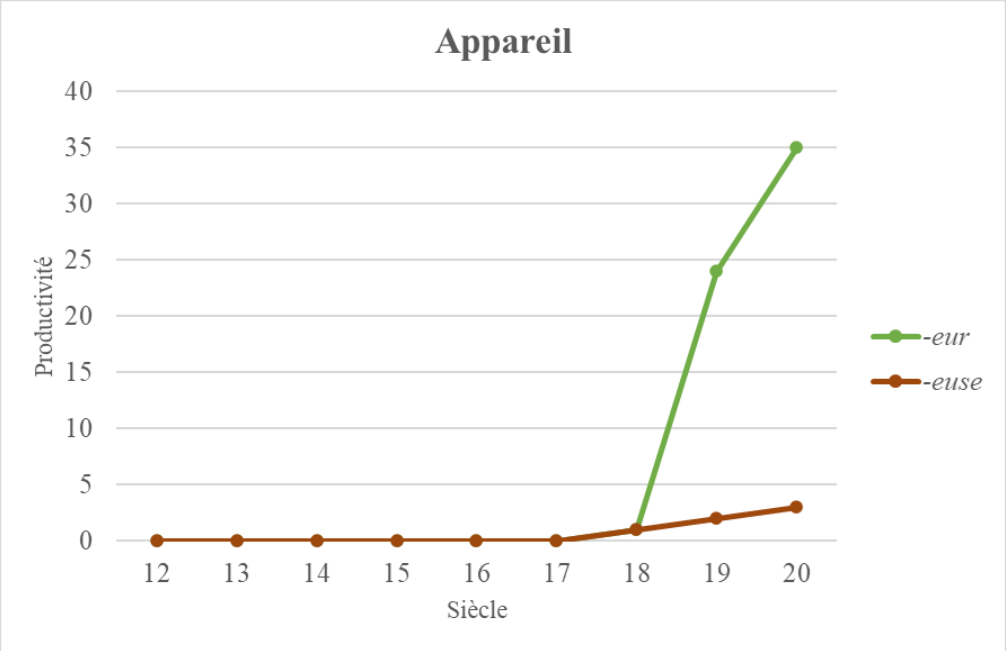

Figure 6.9 Compétition absolue dans la sous-classe ontologique Appareil

Concrètement, la Figure 6.9 met en évidence une dynamique claire de compétition morphologique absolue, illustrée par la domination prépondérante du suffixe *-eur* dans la sous-classe ontologique Appareil. Il semble que peu de noms d'appareils soient dérivés entre le XIIe et le XVIIIe siècle, ce qui peut refléter une productivité limitée dans cette catégorie à cette période. Cela s'explique par le fait que la sous-classe ontologique Appareil est historiquement peu représentée avant la seconde révolution industrielle, période à laquelle ce type d'objet technique commence à se diffuser largement (cf. Section 3.2). À partir du XIXe siècle, sa productivité augmente de manière spectaculaire, atteignant un sommet au XXe siècle avec 35 dérivés. Le suffixe *-euse* ne parvient pas à surpasser le suffixe *-eur*, qui s'impose comme le patron exclusif et stable au sein de cette sous-classe ontologique.

De plus, il convient de souligner que l'augmentation soudaine de la productivité du suffixe *-eur* à partir du XIXe siècle peut être mise en relation avec les évolutions socio-culturelles de l'époque. La révolution industrielle, marquée par l'essor de l'électricité et des progrès scientifiques, a profondément transformé les pratiques sociales et les besoins techniques. L'émergence de nouveaux objets techniques nécessitant une dénomination a entraîné un recours accru à la dérivation. Le suffixe *-eur*, déjà bien ancré dans la suffixation des noms d'instruments, s'est ainsi imposé naturellement au sein de la sous-classe Appareil. Cette dynamique est notamment

observable dans des domaines comme la médecine et les technologies de l'information, où sont apparus des termes tels que SCANNEUR ou DETECTEUR, illustrant la tendance à nommer de nouveaux appareils par la dérivation.

6.3.2 Compétition alternative

Après avoir analysé les cas relevant du mode de compétition absolue, nous présentons maintenant le deuxième mode : la compétition alternative. Celle-ci se manifeste de manière particulièrement claire dans les sous-classes Outil et Machine. Contrairement aux cas précédents, où un schéma s'impose de façon durable et exclusive, la dynamique observée ici repose sur une succession diachronique de deux schémas concurrents.

Commençons par la sous-classe ontologique Outil, dont l'existence précède largement les révolutions industrielles. Représentant l'un des marqueurs fondamentaux de la civilisation humaine, cette catégorie se distingue des autres sous-classes analysées par son ancienneté historique : des noms d'outil apparaissent en effet dès le XIIe siècle, bien avant l'essor technologique des XIXe et XXe siècles.

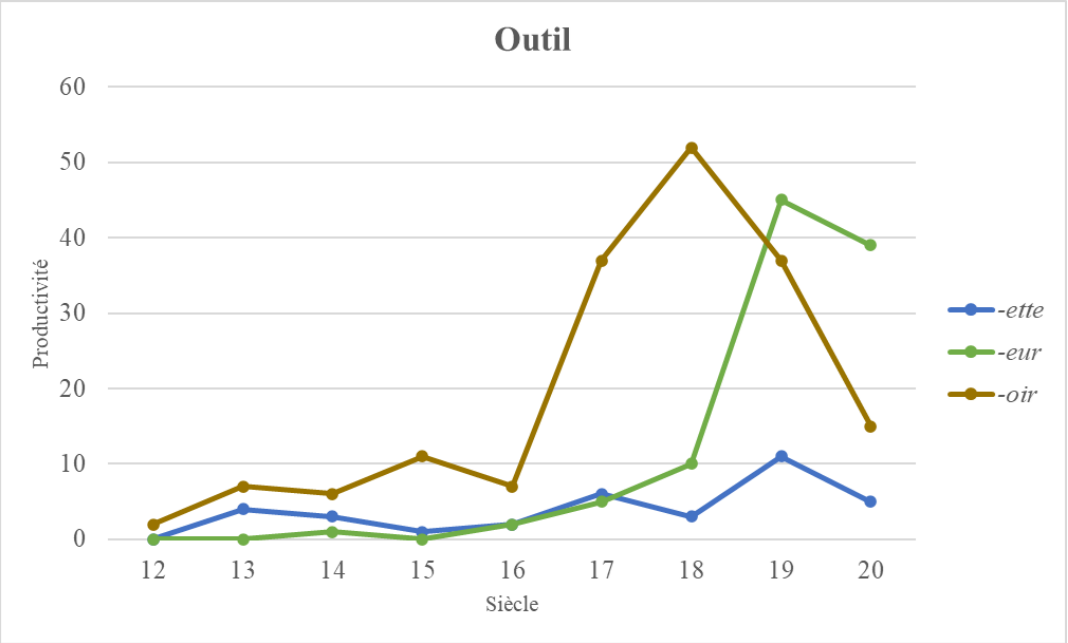

Figure 6.10 Compétition alternative dans la sous-classe ontologique Outil

Nous observons, à travers la Figure 6.10, une alternance marquée entre les schémas *-oir* et *-eur* dans la formation des noms d'outils, après le XVIIIe siècle. Ce cas illustre un mode de compétition alternative, caractérisé par la succession temporelle de deux schémas dominants. Le suffixe *-oir* prédomine entre le XVIe et le XVIIIe siècle,

atteignant un pic d'environ 50 lexèmes, puis est progressivement surpassé par *-eur* à partir du XIXe siècle, jusqu'à l'avènement de celui-ci comme suffixe principal dans cette sous-classe ontologique. Selon Lin (2023), cette succession progressive ne s'explique pas par un déclin de la sous-classe Outil à cette époque, mais plutôt par le fait que le suffixe *-eur* possédait une productivité morphologique particulièrement élevée, que ce soit dans la formation des noms d'instruments ou dans celle des noms d'agents. Précisément, c'est cette productivité élevée qui a favorisé son emploi dans la sous-classe Outil. En revanche, au XXe siècle, nous constatons un recul notable du schéma du schéma *-oir*, tandis que *-eur* connaît une légère diminution, probablement en raison des effets du développement technologique : les outils cessent alors d'être au cœur des dénominations instrumentales, au profit des machines.

Après avoir analysé la dynamique de compétition suffixale dans la sous-classe Outil, il convient désormais de s'intéresser à une autre sous-classe relevant du même mode de compétition alternative : la sous-classe Machine. Cette sous-classe résulte de la transformation d'anciens outils, qui désormais intègrent des mécanismes complexes ou des systèmes techniques.

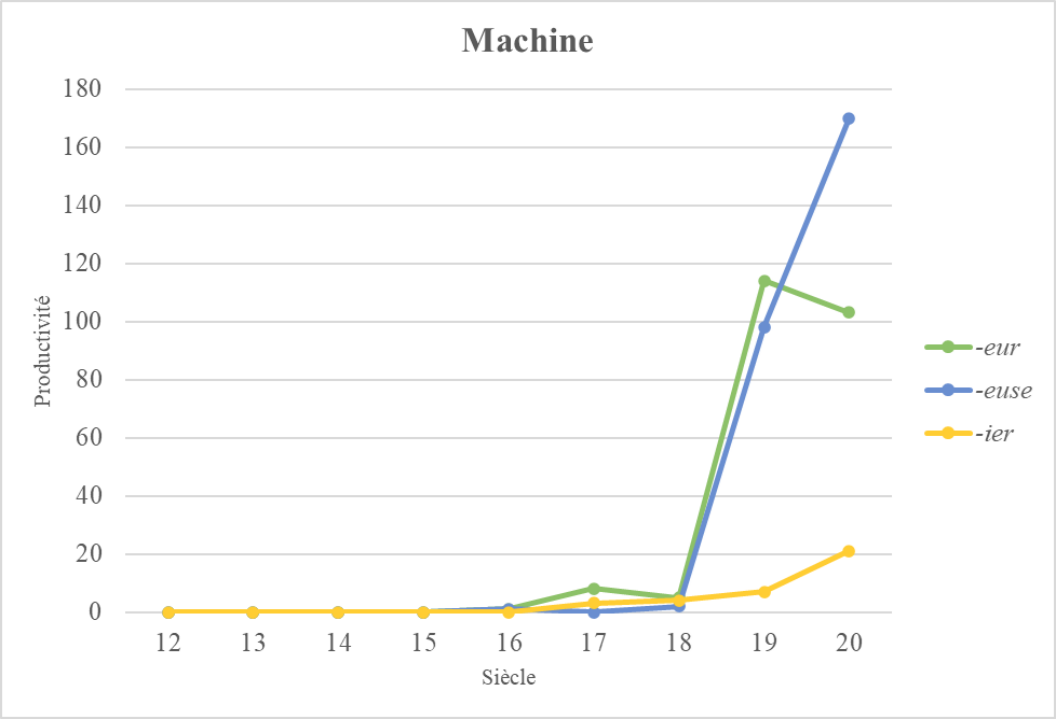

Figure 6.11 Compétition alternative dans la sous-classe ontologique Machine

La Figure 6.11 illustre l'évolution de la productivité des suffixes dans la formation des noms désignant des machines, entre le XIIe et le XXe siècle. Nous observons deux

phases successives bien distinctes, correspondant à une dynamique de compétition alternative entre les suffixes *-eur* et *-euse*. Au XIXe siècle, les suffixes *-eur* et *-euse* connaissent une forte productivité, avec chacun plus d'une centaine de dérivés, sans qu'un seul ne domine clairement. Cependant, cette tendance a changé au XXe siècle, *-euse* connaissant une progression encore plus marquée et culminant à 170 dérivés. Le suffixe *-ier* demeure marginal sans jamais menacer la position dominante des deux schémas principaux. Ce mode de compétition met en évidence une alternance dans la position dominante des schémas au fil du temps, aucun ne parvenant à s'imposer durablement comme suffixe prépondérant dans la sous-classe ontologique Machine.

La productivité élevée des suffixes liés aux noms de machines s'explique en grande partie par le contexte socio-culturel. Ce sont les deux révolutions industrielles qui ont exercé une influence déterminante dans ce domaine. Plus précisément, la généralisation de la vapeur au XIXe siècle, puis de l'électricité au XXe, ont donné lieu à la création de nombreux noms de machines (*par ex.* RECHAUFFEUR, LAVEUSE). A chaque période, de nouveaux types de machines ont vu le jour, ce qui a exigé leur dénomination. Aujourd'hui, les noms de machines occupent une place centrale dans les noms d'instruments, lesquels sont fréquemment dérivés par les schémas *-eur* et *-euse*.

6.3.3 Compétition non-absolue

Hormis les modes de compétition analysés précédemment, nous présenterons un autre mode de compétition : la compétition non absolue. Ce phénomène se retrouve notamment dans trois sous-classes : les instruments de captation, les récipients ainsi que les substances chimiques. Toutefois, les deux premières sous-classes sont très peu représentées dans les données, ce qui limite fortement la possibilité de tirer des conclusions solides. Nous observons néanmoins la coexistence de plusieurs suffixes sur une longue période, sans qu'un seul ne s'impose définitivement comme patron.

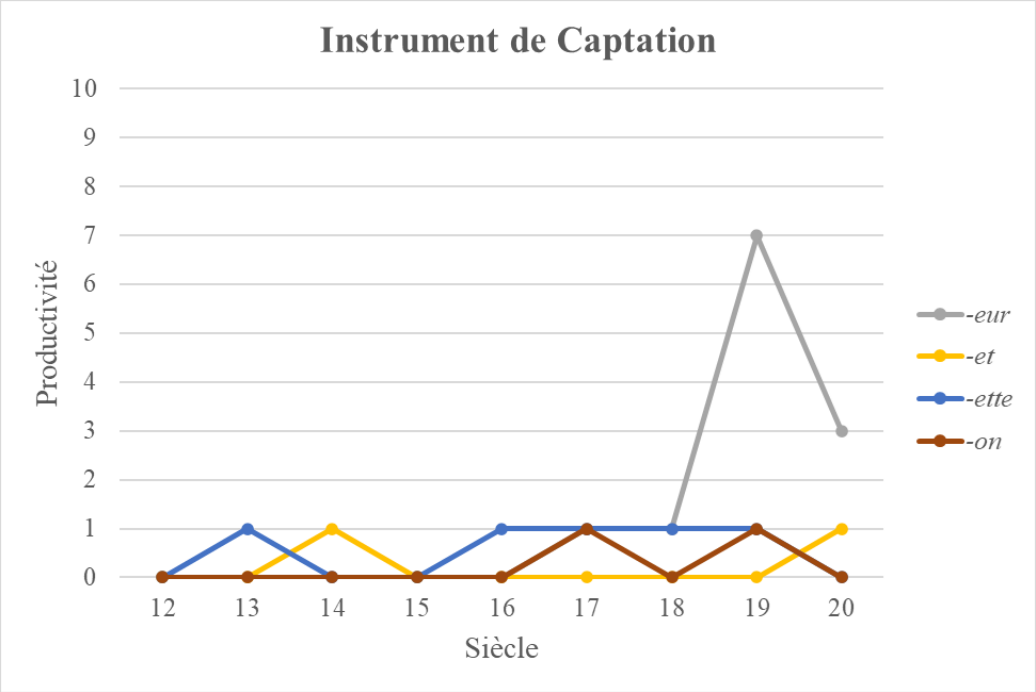

Figure 6.12 Compétition non-absolue dans la sous-classe ontologique Instrument de captation

Nous commencerons par la sous-classe ontologique Instrument de captation. Selon la Figure 6.12, nous observons que la majorité des suffixes, tels que *-ette*, *-et*, *-on*, présentent une productivité quasi-nulle au fil du temps, souvent limitée à un ou deux dérivés par siècle. Le suffixe *-eur* connaît certes un pic au XIXe siècle, atteignant 7 dérivés, avant de chuter au XXe siècle. Toutefois, compte tenu du faible niveau de productivité générale, le suffixe *-eur* ne se distingue pas des autres schémas et ne peut pas être considéré comme le patron de cette sous-classe ontologique.

La productivité faible observée dans cette sous-classe s'explique notamment par deux raisons principales. En premier lieu, une partie des instruments de mesure relevant des instruments de captation tend à être dénommée par la formation de noms composés comprenant l'élément lexical *-mètre* (*par ex.* GRAPHOMETRE, HELIOMETRE), en privilégiant la composition à la dérivation. En second lieu, avec le développement de la technologie à partir du XIXe siècle, de nombreux appareils ont vu le jour. Ces nouveaux instruments ont progressivement pris en charge des fonctions auparavant occupées par certains instruments de captation, contribuant ainsi à leur recul en termes de productivité (Rainer 2021 : 9).

Une fois les instruments de captation analysés, nous abordons une autre sous-classe ontologique relevant du même mode de compétition : la sous-classe Récipient.

Selon la Figure 6.13, nous observons que depuis le XVIIe siècle, plusieurs schémas apparaissent dans cette sous-classe, sans qu'aucun ne s'impose durablement comme patron. Le suffixe *-eur* semble légèrement plus fréquent au XIXe siècle, avec quelques dérivés concentrés à cette période, mais cette tendance reste difficile à interpréter de manière fiable en raison du très faible nombre de dérivés recensés. Le suffixe *-ière* atteint son maximum au XIXe siècle avec 7 dérivés, avant de connaître une baisse. En parallèle, le suffixe *-euse* commence à se développer au XXe siècle, atteignant un niveau de productivité comparable, voire supérieur au suffixe *-ière*. D'autres suffixes, comme *-et*, *-ette*, *-ier* ou *-on*, restent moins productifs.

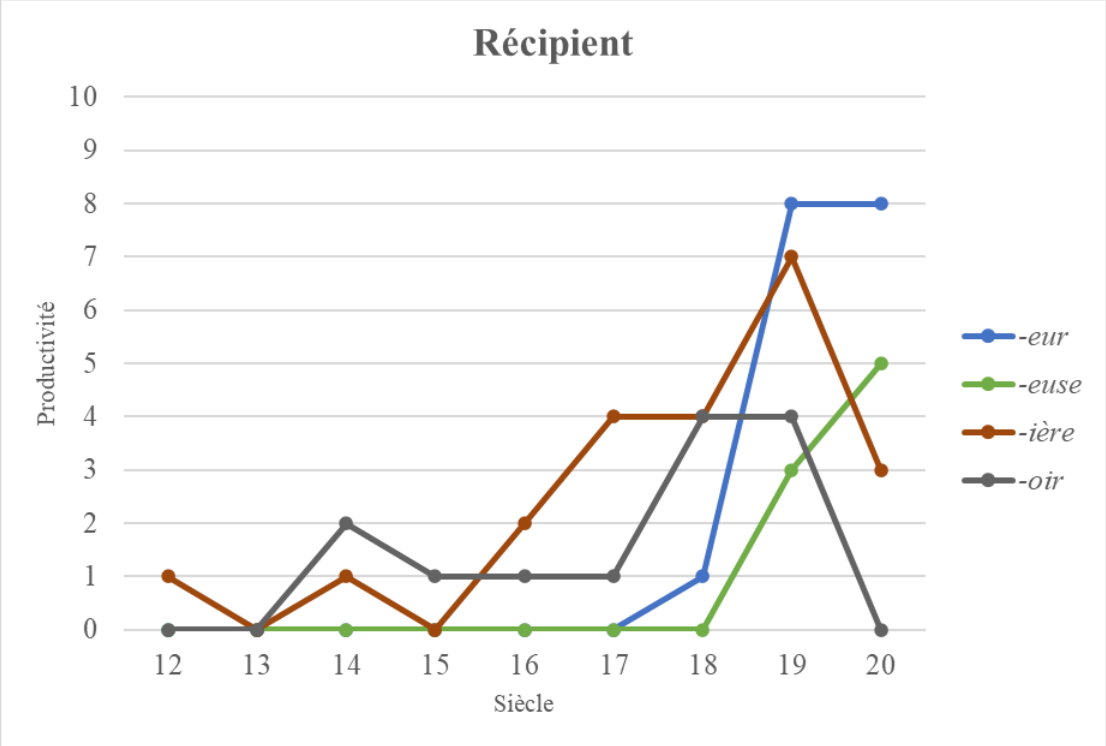

Figure 6.13 Compétition non-absolue dans la sous-classe ontologique Récipient

Dans l'ensemble, bien que les suffixes *-ière* et *-eur* connaissent des pics ponctuels, aucun ne domine réellement cette sous-classe ontologique. Cette absence de spécialisation claire peut s'expliquer par une productivité globalement limitée, qui ne favorise pas l'émergence d'un patron morphologique stable. Cette faible productivité semble en partie liée à la nature même de la sous-classe, le nombre de récipients à nommer étant relativement restreint.

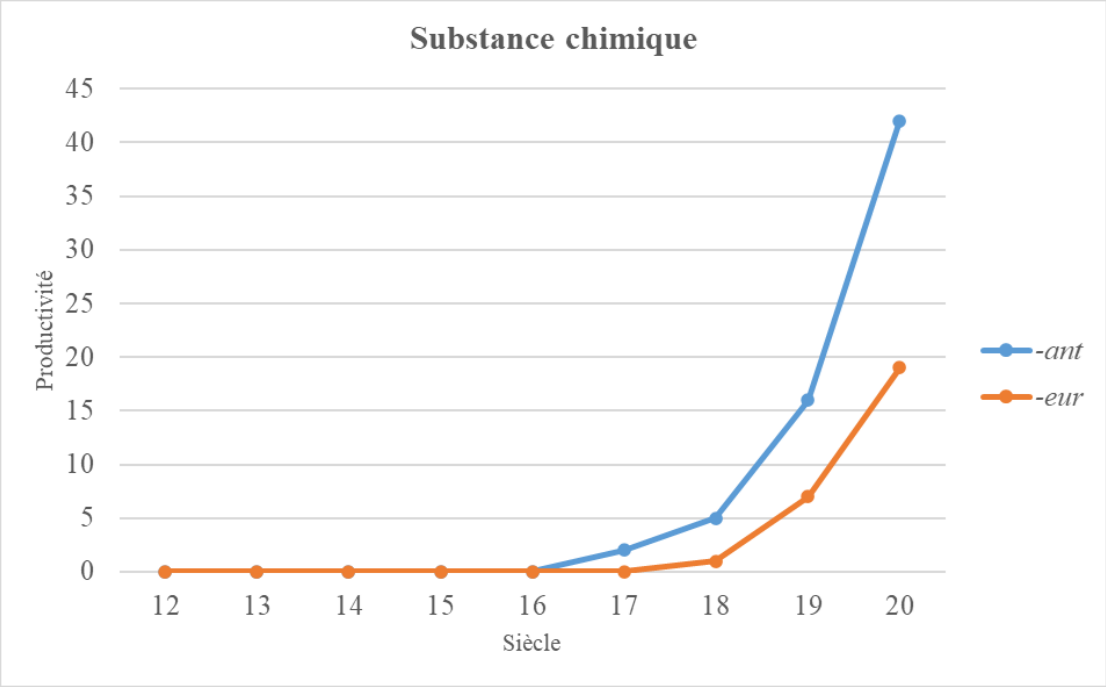

Figure 6.14 Compétition absolue dans la sous-classe ontologique Substance chimique

Venons-en à une autre sous-classe ontologique relevant du mode de compétition non absolue : la sous-classe Substance chimique. La Figure 6.14 montre une évolution importante dans la formation des noms désignant cette sous-classe ontologique, avec une augmentation régulière de la productivité à partir du XIXe siècle, portée principalement par les suffixes *-ant* et *-eur*. Avant cette période, la productivité reste très faible, ce qui s'explique par l'émergence tardive des objets extralinguistiques désignés, souvent relevant des appareils ou des machines, dont la diffusion massive n'intervient qu'à partir de la seconde révolution industrielle. A partir du XIXe siècle, la productivité augmente de manière régulière, portée principalement par les suffixes *-ant* et *-eur*. Le suffixe *-ant* est utilisé environ deux fois plus souvent que *-eur*, mais tous deux restent productifs tout au long de la période. Il s'agit donc d'un cas de compétition non absolue, caractérisé par une coexistence durable entre deux suffixes dont l'un est plus productif que l'autre, sans pour autant exclure son concurrent. Cette dynamique s'inscrit également dans le contexte de la révolution industrielle, qui a fortement stimulé la découverte et la désignation de nouvelles substances chimiques (*par ex.* AZURANT, DESHERBANT, PLASTIFIANT).

À travers l'analyse diachronique des différentes sous-classes ontologiques, nous constatons que la relation entre un suffixe et une sous-classe ontologique donnée n'est

pas toujours stable. Hormis le mode de compétition absolue, les deux autres types de dynamique révèlent une évolution des rapports de compétition entre schémas. Cette évolution est souvent influencée par divers facteurs, notamment des facteurs socio-culturels ainsi que la productivité morphologique des suffixes à une époque spécifique.

A titre d'exemple, la compétition morphologique entre les suffixes *-eur* et *-oir* dans la sous-classe Outil illustre clairement cette évolution : un suffixe peut dominer pendant une certaine période, puis être progressivement remplacé par un autre. Ces observations remettent en question la validité générale du principe d'exclusion compétitive proposé par Lindsay & Aronoff (2013), selon lequel l'apparition d'un nouveau schéma entraîne inévitablement la disparition d'un autre, ou oblige ce dernier à différencier sa niche pour pouvoir coexister. Si ce principe semble partiellement confirmé dans les cas de compétition absolue identifiés dans notre étude, les deux autres modes, à savoir la compétition alternative et la compétition non absolue, en constituent une remise en cause directe.

Plus précisément, d'une part, dans le mode de compétition alternative, le déclin du « perdant » ne résulte pas nécessairement de la montée du « gagnant », mais peut refléter un affaiblissement de sa productivité morphologique à l'époque spécifique. Par exemple, le recul du suffixe *-oir* semble moins lié à la montée de *-eur* qu'à un affaiblissement progressif de sa propre productivité morphologique, une dynamique déjà mise en évidence dans une étude antérieure (Lin 2023), qui montre que ce suffixe entre dans une phase de désaffection à partir du XXe siècle, indépendamment des trajectoires concurrentes. Il en va de même pour les suffixes *-eur* et *-euse* dans la sous-classe ontologique Machine. Même si le suffixe *-euse* dépasse *-eur* dans la sous-classe Machine, il n'y a pas de raison d'affirmer que *-eur* ait été évincé ou ait changé d'orientation compétitive.

D'autre part, dans le mode de compétition non absolue, la multiplicité des schémas et leur faible productivité individuelle rendent difficile l'émergence d'un schéma durablement dominant. Nous observons plutôt une alternance de schémas concurrents, donnant lieu à une coexistence prolongée sous une forme de faible intensité compétitive. En conclusion, ces deux modes de compétition (mode de compétition non-absolu et mode de compétition alternative) montrent que, dans certains cas, la compétition entre schémas n'aboutit pas toujours à l'exclusion de l'un d'eux, mais peut prendre la forme d'une alternance dynamique ou d'une coexistence durable.

6.4 Modélisations statistiques

Dans le chapitre 5, nous avons observé que certaines propriétés linguistiques sont propres à des suffixes instrumentaux spécifiques du français. À partir de l'hypothèse formulée dans la section 6.2, selon laquelle la répartition des schémas suffixaux constitue un phénomène multifactoriel, nous constatons que de nombreuses niches morphologiques se chevauchent largement, certains suffixes allant même jusqu'à recouvrir l'ensemble de la niche de l'autre schéma. Dès lors, une question essentielle se pose : dans quelle mesure la combinaison des propriétés linguistiques analysées dans ce chapitre permet-elle de prédire, de façon fiable, cette distribution malgré ces zones de chevauchement ?

Afin de répondre à cette question, nous avons développé une stratégie de modélisation basée sur l'algorithme de la forêt aléatoire, utilisant des arbres de décision (CART) pour la classification (voir la section 4.3.3). Dans cette section, nous commencerons par décrire le modèle général de forêt aléatoire, puis nous aborderons la mise en place de sous-modèles spécifiques pour les différents types de schémas instrumentaux.

6.4.1 Modèle Général

Avant de modéliser les données, un prétraitement de la base de données est nécessaire afin de prédire avec précision le choix des suffixes tout en minimisant le bruit. Selon le tableau 6.4, notre base de données apparaît éparpillée. Certains suffixes, tels que *-eur, -euse, -oir*, etc., sont plus productifs que d'autres, comme *-ette, -et, -oire*, etc. Pour améliorer la prédiction, nous avons conservé uniquement les suffixes les plus productifs et représentatifs : *-eur, -euse, -oir* et *-ant* pour les noms déverbaux, et *-ier* et *-ière* pour les noms non-déverbaux. En définitive, nous retenons six suffixes instrumentaux que nous intégrerons dans le modèle.

Afin de comprendre quelles propriétés linguistiques influencent le choix des suffixes instrumentaux, nous avons d'abord entraîné une forêt aléatoire (ci-après **Modèle Général**) basée exclusivement sur des propriétés linguistiques communes aux noms déverbaux, dénominaux et désadjectivaux. Cette restriction s'explique par l'hétérogénéité entre ces trois types de noms, qui ne permettent pas d'intégrer l'ensemble des variables disponibles dans un même modèle. Concrètement, les noms

dénominaux et désadjectivaux n'ont pas d'informations relatives à la classe flexionnelle et ne partagent pas les mêmes thèmes que les bases verbales.

Tableau 6.4 Le nombre de lexèmes et leur pourcentage selon les schémas

Suffixe	Nombre de lexèmes	Pourcentage
-eur	**455**	**34%**
-euse	**311**	**23%**
-oir	**206**	**15%**
-ant	**72**	**5%**
-ier	**71**	**5%**
-ière	**63**	**5%**
-ette	57	4%
-et	37	3%
-oire	27	2%
-on	20	1%
-rice	11	1%
-ail	3	0%
-ante	3	0%
-aille	2	0%

Dans le modèle, le suffixe est considéré comme la variable dépendante, tandis qu'un ensemble de variables indépendantes, correspondant à des variables linguistiques spécifiques issues de différents niveaux linguistiques (*par ex.* phonologique, morphologique, syntaxique) est pris en compte. Ces variables ont été intégrées comme prédicteurs distincts dans le modèle, avec les paramètres suivants : *mtry* = 3, *ntree* = 1000, conformément aux pratiques standards détaillées dans la section 4.3.3. Parmi les six variables linguistiques prises en compte, deux variables appartiennent au niveau phonologique :

- PHON1 : la taille phonologique du radical de la base du dérivé, comprenant trois modalités (monosyllabique, dissyllabique, polysyllabique) ;
- PHON2 : le mode d'articulation des segments finaux du radical de la base du dérivé, comprenant 4 modes d'articulation (occlusive, nasal, fricative, approximante).

Une variable morphologique est également incluse :

- MOR1 : la structure morphologique de la base du dérivé, comprenant 5 modalités (simplex, préfixation, suffixation, conversion, autre procédés marginaux).

Nous ajoutons une variable syntaxique, une variable sémantique et une variable diachronique :

- SYN : la catégorie syntaxique de la base du dérivé (verbe, nom, adjectif) ;
- SEM1 : les sous-classes ontologiques du dérivé, qui englobent six sous-classes (outil, machine, appareil, instrument de captation, récipient et substance chimique.).
- Datation : les siècles d'apparition des lexèmes construits, comprenant quatre phases (Avant l'ère scientifique, XVIIIe siècle, XIXe siècle et XXe siècle)

Concernant l'évaluation du Modèle Général, la Figure 6.15 présente l'importance des variables en jeu, mesurée par le « mean decrease in accuracy » (voir la section 4.3.3). En premier lieu, nous constatons que la variable la plus importante est SEM1, c'est-à-dire les sous-classes ontologiques. Les modalités se distinguent comme étant les plus prédictives : par exemple, le suffixe *-ant* est associé aux substances chimiques, *-eur* aux appareils et aux machines, *-oir* aux outils, et *-euse* aux machines (cf. Section 5.4.2). De plus, la variable syntaxique (SYN) apparaît également comme influente dans la prédiction du choix des schémas : la base nominale ou verbale permet notamment d'opposer les suffixes déverbaux (*-eur*, *-euse*, *-oir*, *-ant*) aux suffixes dénominaux (*-ier, -ière*) (cf. Section 5.1).

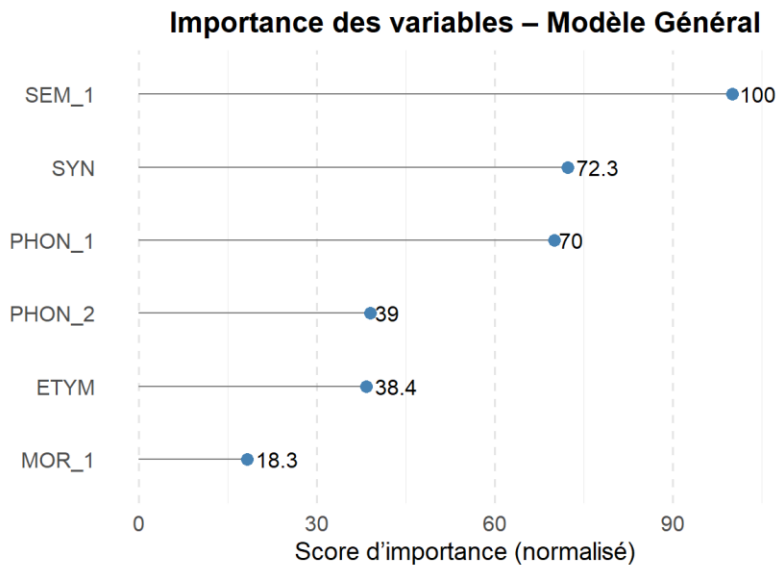

Figure 6.15 Importance des variables dans le modèle de forêt aléatoire (**MODÈLE GÉNÉRAL**), selon la mesure de *mean decrease in accuracy*

Les deux variables phonologiques se classent respectivement en troisième et quatrième position parmi les plus discriminantes. Concrètement, pour la variable PHON1, le suffixe *-oir* préfère les radicaux monosyllabiques. À l'inverse, les suffixes *-eur* et *-ant* ont tendance à sélectionner des radicaux polysyllabiques, lesquels sont rejetés par des suffixes tels que *-euse*, *-oir*, *-ier* et *-ière*. En ce qui concerne la variable PHON2, le suffixe *-eur* montre une nette préférence pour les segments finaux occlusifs, tout en rejetant les segments nasaux et approximants. Il convient de souligner que cette préférence du suffixe *-eur* ne relève pas uniquement du facteur phonologique, mais renvoie aussi à un facteur morphologique. En effet, comme indiqué en Section 5.3.3, *-eur* tend à sélectionner le thème verbal 13, caractérisé, lorsqu'il est régulier, par une finale en /t/. En revanche, *-euse* est significativement associé à des radicaux se terminant par les segments nasaux, et évite les occlusives. Le suffixe *-oir* privilégie les radicaux se terminant par les segments fricatifs, tandis que *-ier* montre une nette préférence pour les radicaux se terminant par les segments nasaux, et une sous-représentation significative dans la catégorie fricative. Le suffixe *-ant* privilégie des radicaux se terminant par un segment approximant, alors que *-ière* est significativement sous-représenté dans cette même catégorie (cf. Section 5.2.2).

Ensuite, la variable diachronique s'avère relativement importante. Comme mentionné dans la Section 5.5, avant l'ère scientifique, les suffixes *-oir*, *-ier* et *-ière* dominent, tandis que *-eur* et *-euse* sont presque absents. Au XVIIIe siècle, *-oir* reste très

productif, alors que *-euse* demeure rare et *-eur* n'apparaît encore qu'à la marge. Ce contraste met en évidence une opposition nette entre des suffixes historiques et d'autres plus modernes. Concrètement, le XIXe siècle ouvre une nouvelle phase : la productivité de *-eur* devient très forte, celle de *-euse* augmente, tandis que *-oir* commence à perdre du terrain. Cette réorganisation s'accentue au XXe siècle, où *-eur*, *-euse* et *-ant* s'imposent, alors que des schémas plus anciens comme *-oir*, *-ier* et *-ière* connaissent un recul marqué.

Enfin, la variable morphologique MOR1 (procédés morphologiques) se classe en dernière position, ce qui indique qu'elle exerce une influence relativement limitée dans la prédiction du choix des schémas. Toutefois, certaines structures morphologiques des bases contribuent tout de même à distinguer certains suffixes instrumentaux (cf. Section 5.3.2). Par exemple, les suffixes *-eur* et *-ant* tendent à privilégier les bases verbales suffixées, alors que *-oir* et *-euse* ont tendance à l'éviter (voir les sections 5.2.1 et 5.3.2). Cette contrainte morphologique s'explique, d'une part, par l'autonomie que possède la morphologie en tant que système indépendant. L'une de ses manifestations est l'ordre des suffixes : par exemple, *-eur* et *-ant* s'attachent fréquemment à une base déjà suffixée, apparaissant comme des suffixes externes ; en revanche, *-oir* et *-euse* tendent à éviter de s'attacher à des bases déjà suffixées.

D'autre part, selon la théorie de l'ordre des suffixes (Hay & Plag 2004 ; Plag & Baayen 2009 ; Saarinen & Hay 2014), les suffixes et bases (radicaux) ayant la même origine étymologique se combinent plus facilement pour former un dérivé. Ce phénomène a été confirmé en français (Huot 2005 : 65 ; Lieber 2021 : 200). Par ailleurs, dans Lin (soumis b), j'ai précisé que cette influence se manifeste de manière plus marquée dans la suffixation verbale que dans les autres types de suffixation. Du point de vue étymologique, les principaux procédés morphologiques pour former des verbes en français incluent la suffixation, la conversion et la préfixation. Parmi eux, concernant la suffixation, les suffixes *-iser* et *-ifier*, sont empruntés au latin (Grevisse & Goosse 2016 : 187). La conversion et la plupart des préfixations sont héritées du latin et deviennent des procédés indispensables dans le système morphologique du français (Grevisse & Goosse 2016 : 188 - 192). Parmi les quatre schémas analysés, seul *-eur* présente une double nature : d'un côté, il a été hérité du latin et adapté aux mécanismes dérivationnels du français ; de l'autre, il a également été réemprunté directement depuis le latin à une époque ultérieure. Ces deux formes se sont fusionnées dans le système du français pour former un schéma unique (Grevisse & Goosse 2016 : 180). C'est pourquoi

-eur peut s'attacher aussi bien à une base (radical) populaire qu'à une base (radical) savant(e). Le suffixe *-ant*, quant à lui, provient de la forme du participe présent, et a été progressivement intégré au système dérivationnel par le biais de la grammaticalisation (Grevisse & Goosse 2016 : 178 ; Knittel 2017a ; Knittel 2017b). Ainsi, bien qu'il soit considéré comme un suffixe français, *-ant* peut néanmoins se combiner avec des bases (radical) savant(e)s (pour les définitions, voir la section 5.3.2). En revanche, *-oir* et *-euse* sont hérités du latin et deviennent des suffixes populaires en français, et ont donc tendance à s'associer à des bases construites par la conversion ou la préfixation, qui sont de nature populaire (Lin, soumis b).

Comme indiqué dans la section 4.3.3, l'évaluation du modèle de forêt aléatoire repose principalement sur deux indicateurs : la précision de classification et la valeur de l'AUC. Le modèle atteint une précision de 0,707 et une AUC de 0,949, ce qui témoigne d'un pouvoir de discrimination globalement satisfaisant. Cela confirme que les variables sémantiques, phonologiques, syntaxiques et morphologiques participent à la sélection des schémas instrumentaux et que le facteur étymologique exerce une influence sur le choix des schémas. Toutefois, la performance n'est pas homogène pour tous les schémas.

Le Tableau 6.5 présente la matrice de confusion du modèle final de la forêt aléatoire Modèle Général (c'est-à-dire le modèle optimisé obtenu après une validation croisée en 10 passes de la forêt aléatoire). Le suffixe *-ière* présente des taux d'erreur d'environ 44%, suggérant de fortes confusions avec *-ier*. En revanche, pour les cinq autres schémas, entre 21% et 33% des noms d'instruments ont été incorrectement classés. Plus précisément, les principales sources d'erreur peuvent être identifiées comme suit. D'une part, les noms d'instruments en *-ière* tendent à être classés parmi ceux en *-ier*, tandis que l'inverse est également observé, mais de manière moins fréquente. D'autre part, les noms d'instruments en *-eur* sont fréquemment classés parmi ceux en *-oir, -euse* et *-ant*, bien que cette tendance soit moins marquée dans l'autre sens. Ces types d'erreurs peuvent s'expliquer par la similarité de distribution entre *-ière* et *-ier*, d'une part, et entre *-eur*, *-oir* et *-ant*, d'autre part, notamment au regard de certaines propriétés linguistiques décrites au Chapitre 5, telles que la catégorie syntaxique de la base du dérivé, ainsi que la taille phonologique du radical (monosyllabique ou dissyllabique) pour les dérivés concernés. Autrement dit, ces schémas entrent, dans une certaine mesure, en situation de forte compétition.

Tableau 6.5 Matrice de confusion du modèle final de la forêt aléatoire Modèle Général basée sur les propriétés linguistiques. Les vrais positifs sont surlignés en vert

		Prédiction							
		ant	*eur*	*euse*	*ier*	*ière*	*oir*	Total	Class.Error
Référence	*ant*	57	12	0	0	0	3	72	0.208
	eur	16	304	98	4	3	26	451	0.326
	euse	0	57	235	4	1	6	303	0.224
	ier	0	3	2	43	20	2	70	0.386
	ière	0	3	1	22	35	2	63	0.444
	oir	0	37	15	2	2	148	204	0.275
Total		73	416	351	75	61	187	1163	

6.4.2 Sous-modèles

Dans la section 6.4.1, nous avons élaboré le Modèle Général afin de prédire le choix des suffixes instrumentaux en français à partir des propriétés linguistiques communes aux noms déverbaux et non déverbaux. Bien que ce modèle ait produit des résultats satisfaisants, des améliorations demeurent envisageables, notamment par l'intégration de propriétés linguistiques supplémentaires susceptibles de raffiner la prédiction. Comme présenté dans la section 6.4.1, hormis les variables déjà mobilisées dans le Modèle Général, d'autres facteurs morphologiques et sémantiques pertinents peuvent être considérés, tels que les thèmes morphologiques et la classe flexionnelle de la base verbale du dérivé, ainsi que le profil sémantique assumé par la base du dérivé. Compte tenu de l'hétérogénéité de la structure morphologique des trois types de noms dérivés étudiés, nous avons opté pour une division en deux groupes distincts : les noms déverbaux, d'une part, et les noms non déverbaux (dénominaux et désadjectivaux), d'autre part. Chaque groupe fait ainsi l'objet d'un modèle spécifique intégrant des variables linguistiques adaptées à ses propriétés morphologiques et à ses profils sémantiques. Dans cette sous-section, nous commencerons par l'analyse des noms déverbaux, avant d'aborder, dans un second temps, les noms non déverbaux.

6.4.2.1 Sous-modèle 1 : modèle destiné à prédire les schémas déverbaux

Dans le but de prédire de manière plus fine le choix des schémas déverbaux, nous avons conçu un premier modèle spécifique, intitulé Sous-modèle 1. À l'instar du Modèle Général, ce modèle actuel repose sur un prétraitement rigoureux des données, destiné à améliorer la précision de la prédiction tout en réduisant le bruit. Sur la base des tableaux 6.4 et 5.1 (cf. section 5.1), seuls les schémas déverbaux les plus productifs et représentatifs ont été conservés, à savoir : *-eur, -euse, -oir* et *-ant*. Dans ce modèle, le suffixe constitue la variable dépendante, tandis que huit variables linguistiques indépendantes ont été retenues (ce qui implique de fixer *mtry* à 3). Il convient de noter que la variable syntaxique (SYN, catégorie de la base) utilisée dans le Modèle Général a ici été exclue, puisque toutes les bases sont verbales. En définitive, ces variables couvrent quatre dimensions linguistiques :

- Phonologique :
    - PHON1 : la taille phonologique du radical de la base du dérivé, comprenant trois modalités (monosyllabique, dissyllabique, polysyllabique) ;
    - PHON2 : le mode d'articulation des segments finaux de radical de la base du dérivé, comprenant 4 modes d'articulation (occlusive, nasal, fricative, approximant).
- Morphologique :
    - MOR1 : la structure morphologique de la base du dérivé, comprenant 5 modalités (simplex, préfixation, suffixation, conversion, autre procédés marginaux).
    - MOR2 : les thèmes, comprenant 2 modalités (thème verbal (V1) et thème verbal latin (V13)).
    - MOR3 : la classe flexionnelle de la base verbale, comprenant 3 modalités (Groupe 1, Groupe 2, Groupe 3)
- Sémantique :
    - SEM 1 : les sous-classes ontologiques du dérivé, comprenant 6 modalités (outil, machine, appareil, instrument de captation, récipient et substance chimique).
    - SEM 2 : le profil sémantique assumé par la base dans la représentation du dérivé, comprenant 2 modalités (verbe télique et verbe formel).

- Datation :
    - ETYM : les siècles d'apparition des lexèmes construits, comprenant quatre phases (Avant l'ère scientifique, XVIIIe siècle, XIXe siècle et XXe siècle)

Le **Sous-modèle 1** repose sur une forêt aléatoire, où le schéma constitue la variable dépendante et les huit variables linguistiques, les prédicteurs. Les paramètres du modèle sont les mêmes que précédemment : *mtry* = 3, *ntree* = 1000, et validation croisée à 10 passes. L'importance des variables issues du modèle **Sous-modèle 1** est illustrée à la Figure 6.16.

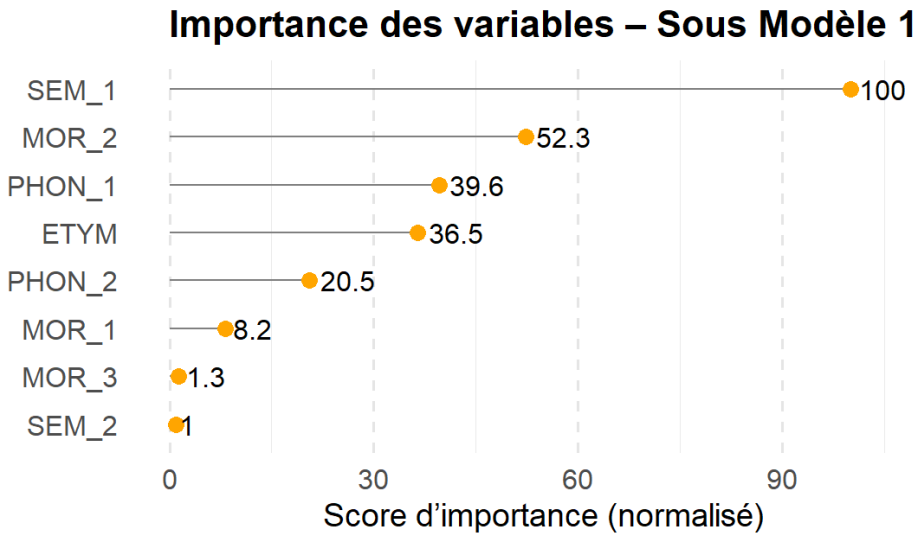

Figure 6.16 Importance des variables dans le modèle de forêt aléatoire (Sous-modèle 1), selon la mesure de *mean decrease in accuracy*

Conformément aux résultats de la Figure 6.16, la sous-classe ontologique du dérivé (SEM1) s'avère être la variable la plus déterminante dans la compétition des schémas déverbaux, comme déjà observé dans le **Modèle Général** (cf. section 6.4.1). Les thèmes de la base (MOR2) se classent en deuxième position : le thème verbal 13 (radical latin) distingue notamment certains dérivés en *-eur* de ceux en *-euse*, *-oir* et *-ant*. Plus précisément, *-euse* et *-ant* privilégient le thème verbal 1, tandis que *-eur* peut apparaître avec les thèmes 1 et 13 (cf. section 5.3.2). En outre, la taille phonologique du radical (PHON1) et la datation (ETYM) interviennent également comme des variables non négligeables. Le mode d'articulation (PHON2) occupe la cinquième

position en termes d'importance, exerçant une influence modérée sur la prédiction des schémas. Il est toutefois surprenant de constater que, bien que la section 5.3.2 ait mis en évidence l'existence de structures morphologiques préférentielles de la base du dérivé (MOR1) selon les suffixes, cette variable n'apparaît pas comme un prédicteur significatif dans le modèle (cf. Figure 6.16). Nous supposons que cela pourrait être dû à des effets de colinéarité, en particulier avec la variable PHON1. En effet, la préfixation, tout comme la suffixation, tend à augmenter la taille phonologique des formes, ce qui pourrait atténuer l'effet propre de MOR1 dans le modèle. Finalement, la classe flexionnelle (MOR3) et le profil sémantique assumé par la base du dérivé (SEM2) semblent jouer un rôle moins primordial, ce qui corrobore les observations faites dans les sections 5.3.1 et 5.4.1.

Il convient de souligner que le profil sémantique assumé par la base du dérivé (SEM2) ne joue pas un rôle déterminant dans la prédiction de tous les schémas déverbaux. Cette observation suggère que, dans le cadre de la Néo-Morphologie cognitive, la relation sémantique entre la base et le schéma, notamment telle qu'elle est représentée dans la Structure de Qualia, ne constitue pas un élément structurel intrinsèque du système morphologique. Dans le cas des schémas déverbaux, c'est à travers le profilage cognitif que les profils sémantiques de la base sont mises en évidence, ce qui permet d'initier la projection sémantique dans le système dérivationnel. Toutefois, ces schémas tendent à privilégier des trajectoires cognitives similaires, notamment celles associées à des verbes à rôle télique dans la Structure de Qualia. Cette convergence des préférences limite la capacité de cette variable à jouer un rôle discriminant dans la compétition morphologique entre les suffixes déverbaux instrumentaux.

Concernant l'évaluation de ce modèle actuel, il atteint une précision de 0,725 et une AUC de 0,935, ce qui témoigne d'un pouvoir de discrimination globalement satisfaisant. Cela confirme que les variables sémantiques, phonologiques, syntaxiques et morphologiques participent à la sélection des suffixes instrumentaux et que la datation exerce une influence sur le choix des suffixes. Toutefois, la performance n'est pas homogène pour tous les suffixes. Ainsi, nous procédons à une analyse plus fine des performances du modèle, laquelle révèle des disparités selon les suffixes. Comme l'indique le Tableau 6.6, le suffixe *-eur* affiche un taux d'erreur particulièrement élevé, atteignant près de 35 %. Cela signifie que le suffixe *-eur* présente une distribution similaire à celle des autres suffixes instrumentaux étudiés, ce qui témoigne d'une forte

situation de compétition. Les propriétés linguistiques identifiées ne permettent pas de prédire de manière fiable le choix du suffixe *-eur*. Cette difficulté pourrait s'expliquer par une dynamique diachronique, dans laquelle *-eur* aurait progressivement occupé certaines niches morphologiques auparavant associées à d'autres suffixes concurrents.

Tableau 6.6 Matrice de confusion du modèle final **(Sous-modèle 1)**, fondé sur les propriétés linguistiques. Les vrais positifs sont surlignés en vert

		Prédiction					
		ant	*eur*	*euse*	*oir*	Total	Class.Error
Référence	*ant*	55	11	0	3	69	0.203
	eur	13	286	107	34	440	0.350
	euse	0	47	244	7	298	0.181
	oir	0	40	15	145	200	0.275
Total		68	384	366	189	1007	

Enfin, pour chaque suffixe, nous avons réalisé une analyse visant à identifier les propriétés linguistiques les plus déterminantes en fonction de leur contribution à l'amélioration de la précision du modèle. La Figure 6.17 met en évidence les prédicteurs les plus influents pour chaque suffixe dans le sous-modèle 1. Tout d'abord, nous avons observé que la sous-classe ontologique (SEM1) constitue la variable linguistique la plus déterminante pour prédire le choix des suffixes. En particulier, la modalité principale de chaque suffixe instrumental reflète les préférences ontologiques spécifiques associées au dérivé, ce qui corrobore les observations présentées dans la section 5.4.2. Par ailleurs, du point de vue des variables linguistiques, les quatre suffixes se distinguent non seulement par la nature de leurs modalités pertinentes, mais aussi par leur nombre. Cette diversité correspond aux propriétés linguistiques préférentielles analysées au chapitre 5. Par exemple, pour le suffixe *-ant*, quatre modalités sont pertinentes en SEM2, deux en MOR2, trois en PHON2, ainsi qu'une en MOR1. En SEM2, à l'exception de la sous-classe Substance chimique, le suffixe *-ant* tend à éviter les sous-classes Machine, Outil et Appareil. Dans le cadre de MOR2, les thèmes 1 et 13 sont associés à *-ant*, mais celui-ci privilégie clairement le thème 1. Sur le plan phonologique (PHON2), *-ant* montre une préférence des bases dont le radical se termine par le mode d'articulation Approximant, tandis que les deux autres modalités sont préférées par d'autres suffixes (*-oir* et *-eur*). Enfin, en ce qui concerne MOR1, les bases construites par préfixation jouent un rôle plus important pour le suffixe *-ant*. Comme

mentionné dans la section 5.3.2, ce procédé morphologique dérivant les bases verbales est particulièrement préféré par les suffixes *-oir* et *-euse*, ce qui constitue un critère discriminant pertinent entre eux.

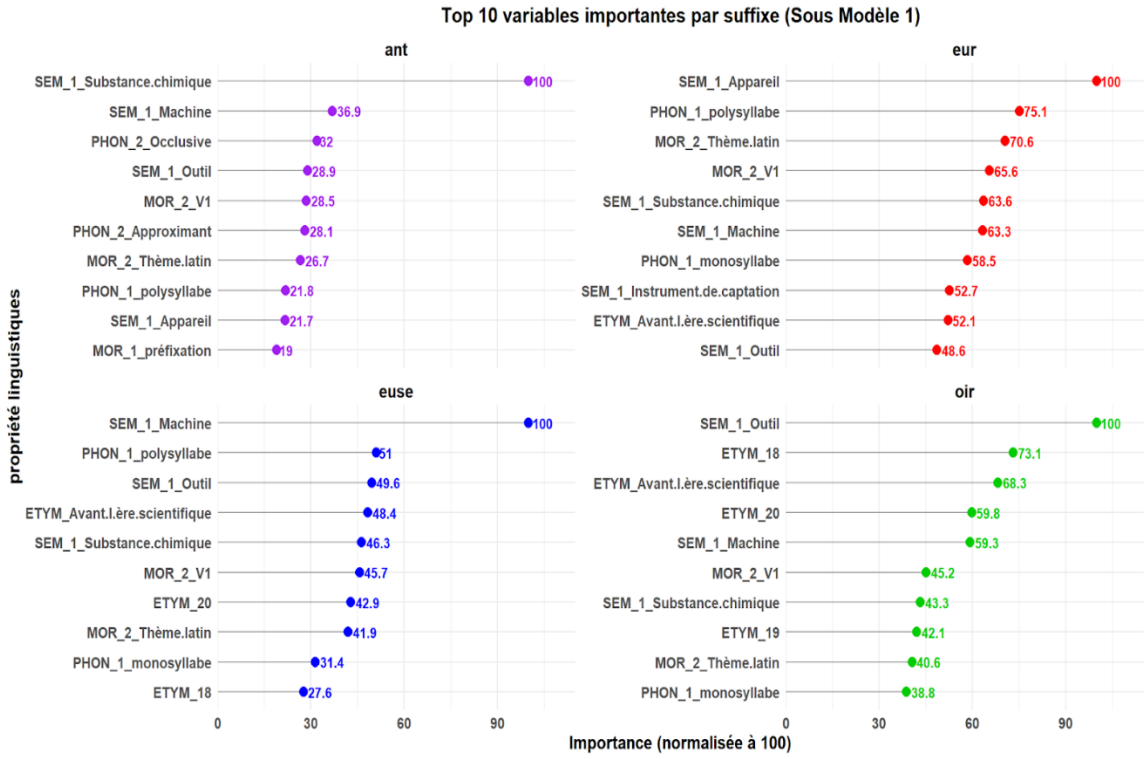

Figure 6.17 Importance des variables linguistiques dans la prédiction des quatre suffixes instrumentaux (Sous-modèle 1). Les barres indiquent la contribution moyenne à l'amélioration de la précision (*Mean Decrease Accuracy*)

6.4.2.2 Sous-modèle 2 : modèle destiné à prédire les suffixes non-déverbaux

Après avoir entraîné un modèle dédié aux suffixes déverbaux, nous procédons à la mise en place d'un second modèle **Sous-modèle 2**, destiné à prédire les suffixes propres aux noms non déverbaux. Comme pour les modèles précédents, un prétraitement de la base de données s'avère nécessaire afin d'optimiser la précision prédictive et de limiter le bruit statistique. Conformément aux observations des tableaux 6.4 et 5.1 (cf. section 5.1), nous avons retenu les suffixes non déverbaux les plus productifs et représentatifs, à savoir *–ier* et *–ière*.

Dans ce sous-modèle, le suffixe constitue la variable dépendante, tandis que huit variables indépendantes sont mobilisées (ce qui implique de fixer le paramètre *mtry* à 3). A noter que la variable, la classe flexionnelle de la base verbale (MOR3), mobilisée dans le Sous-modèle 1, est ici écartée, les bases nominales et adjectivales n'étant pas soumises à la conjugaison verbale. Les variables indépendantes sont donc les suivantes :

- Phonologique :
    - PHON1 : la taille phonologique du radical de la base du dérivé, comprenant trois modalités (monosyllabique, dissyllabique, polysyllabique) ;
    - PHON2 : le mode d'articulation des segments finaux du radical de la base du dérivé, comprenant 4 modes d'articulation (occlusive, nasal, fricative, approximante).
- Morphologique :
    - MOR1 : la structure morphologique de la base du dérivé, comprenant 5 modalités (simplex, préfixation, suffixation, conversion, autre procédés marginaux).
    - MOR2 : les thèmes de la base du dérivé, comprenant 3 modalités (thème 1, thème 2 et le thème L).
- Syntaxique :
    - SYN : la catégorie syntaxique de la base du dérivé, comprenant 2 modalités (nom et adjectif).
- Sémantique :
    - SEM 1 : les sous-classes ontologiques du dérivé, comprenant 6 modalités (outil, machine, appareil, instrument de captation, récipient et substance chimique).
    - SEM 2 : le profil sémantique assumé par la base dans la représentation du dérivé, comprenant 3 modalités (Patient, autre rôles sémantiques et projections cognitives).
- Datation :
    - ETYM : les siècles d'apparition des lexèmes construits, comprenant quatre phases (Avant l'ère scientifique, XVIIIe siècle, XIXe siècle et XXe siècle).

Le Sous-modèle 2 repose sur une forêt aléatoire où le suffixe est prédit à partir des huit variables mentionnées. Le modèle est paramétré de manière identique à celui du Sous-modèle 1 (*mtry* = 3, *ntree* = 1000). L'importance relative des variables prédictives est illustrée dans la Figure 6.18. Les résultats indiquent que, comme dans les modèles précédents, les sous-classes ontologiques (SEM 1) du dérivé demeurent la variable prédictive la plus influente (cf. section 5.4.1). Cette variable est suivie par la datation (ETYM). De plus, deux variables phonologiques, à savoir le mode d'articulation des segments finaux du radical (PHON2) et la taille phonologique (PHON1), occupent respectivement les troisième et quatrième rangs. À la différence du Sous-modèle 1, le profil sémantique assumé par la base du dérivé (SEM2) exerce ici une certaine influence sur la sélection des suffixes non déverbaux. En revanche, les espaces thématiques (MOR2) ont une influence plus limitée sur la prédiction de ces suffixes. Enfin, les variables MOR1 et SYN exercent une influence très faible dans ce modèle en ce qui concerne le choix des suffixes non déverbaux.

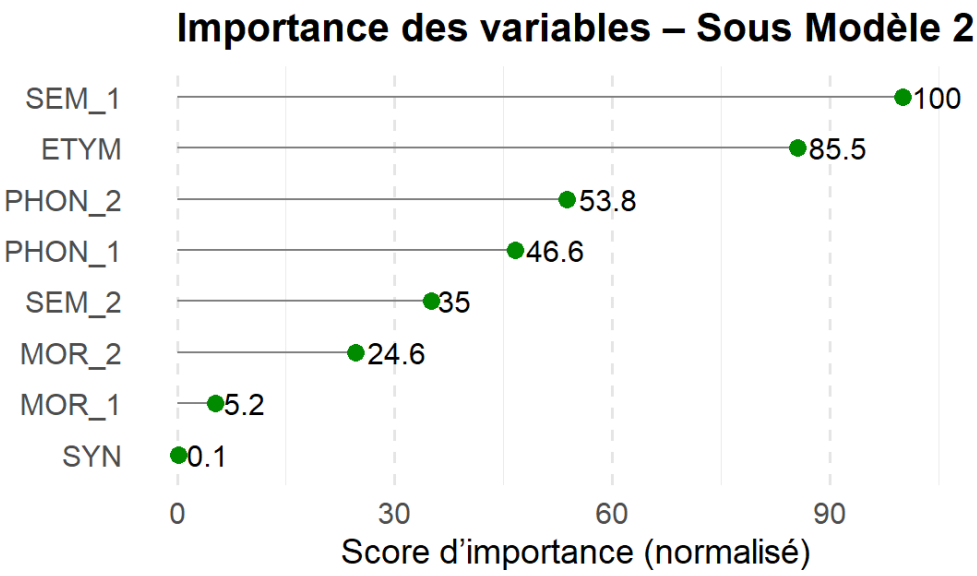

Figure 6.18 Importance des variables dans le modèle de forêt aléatoire (Sous-modèle 2), selon la mesure de *mean decrease in accuracy*

De plus, selon la Figure 6.18, parmi l'ensemble des variables intégrées dans le modèle, la sous-classe ontologique et le facteur étymologique apparaissent comme les prédicteurs les plus discriminants pour les noms non déverbaux. Elles permettent de distinguer de manière nette les choix suffixaux observés dans les données, bien plus

que les autres variables linguistiques. En outre, ce résultat, bien qu'en deçà des attentes, révèle la complexité inhérente à la dérivation des noms non déverbaux. Il confirme l'hypothèse selon laquelle la compétition suffixale ne résulte pas exclusivement de contraintes linguistiques internes, mais dépend également de facteurs socio-culturels, diachroniques ou cognitifs.

Il convient de souligner que par rapport au Sous-modèle 1, le profil sémantique assumé par la base dans la représentation sémantique du dérivé (SEM 2) présente une un certain poids dans la prédiction de tous les schémas non-déverbaux, mais moins que les classes ontologiques (SEM1) ou la datation (ETYM). Comme nous l'avons souligné dans la Section 3.1.3, dans le cadre de la Néo-Morphologie Cognitive, la relation sémantique entre la base et le suffixe, telle qu'elle est représentée dans la Structure de Qualia, ne constitue pas un élément structurel intrinsèque du système morphologique. Elle est mobilisée dans le cadre du profilage cognitif, qui identifie un profil sémantique saillant et active une trajectoire cognitive. Cette trajectoire oriente la construction du sens du dérivé et conduit à l'espace d'intégration morphosémantique, où se forment les conditions sémantiques préalables à l'activation d'un schéma suffixal, permettant ainsi l'accès au système dérivationnel. Plus précisément, dans le cas des noms d'instruments, le profil sémantique issu du profilage cognitif ne constitue pas une variable suffisamment pertinente pour prédire le choix des suffixes, en raison de la similarité des trajectoires cognitives activées dans la majorité des cas. À titre d'exemple, les suffixes *-ier* / *-ière* privilégient notamment les bases occupant le rôle Patient pour la formation des noms d'instruments dénominaux. En revanche, les autres modalités, c'est-à-dire les autres rôles sémantiques et les projections cognitives, présentent une distribution similaire entre les deux schémas.

Pour évaluer le Sous-modèle 2, les résultats indiquent une précision de classification de 0,645 et une valeur d'AUC de 0,887, ce qui suggère que, dans le cas des schémas instrumentaux non déverbaux, elles constituent des prédicteurs assez discriminants. Cela confirme que les variables sémantiques, phonologiques, syntaxiques et morphologiques participent à la sélection des schémas instrumentaux et que la datation exerce une influence sur le choix des schémas. Cependant, la performance n'est pas homogène pour tous les schémas. Ainsi, nous procédons finalement à une analyse détaillée des performances, qui révèle une forte instabilité dans la classification. Le Tableau 6.7, présentant la matrice de confusion du modèle final (forêt aléatoire avec validation croisée à 10 passes), indique que 38 % des dérivés

en *-ier* et 33 % des dérivés en *-ière* sont mal classés. Cette évolution souligne que les suffixes *-ier* et *-ière* demeurent difficiles à prédire, à l'instar du suffixe *-eur*. Cette faiblesse discriminatoire témoigne d'une forte compétition morphologique entre ces deux schémas et confirme la difficulté persistante du modèle à les différencier de manière fiable.

Tableau 6.7 Matrice de confusion du modèle final de la forêt aléatoire **Sous-modèle 2** basée sur les propriétés linguistiques. Les vrais positifs sont surlignés en vert

		Prédiction			
		ier	*ière*	Total	Class.Error
Référence	*ier*	41	25	66	0.379
	ière	19	39	58	0.328
Total		60	64	124	

Cette instabilité dans la classification et la compétition marquée entre certains suffixes, notamment *-ière* et *-ier*, ne peut être entièrement expliquée par les seules propriétés linguistiques. Elle semble plutôt indiquer une convergence partielle de leurs préférences distributionnelles sur certaines propriétés linguistiques, traduisant un certain chevauchement de niches morphologiques (cf. Section 6.2). Ce phénomène se manifeste également dans les cas où *-eur* entre en compétition avec *-oir* ou *-euse*.

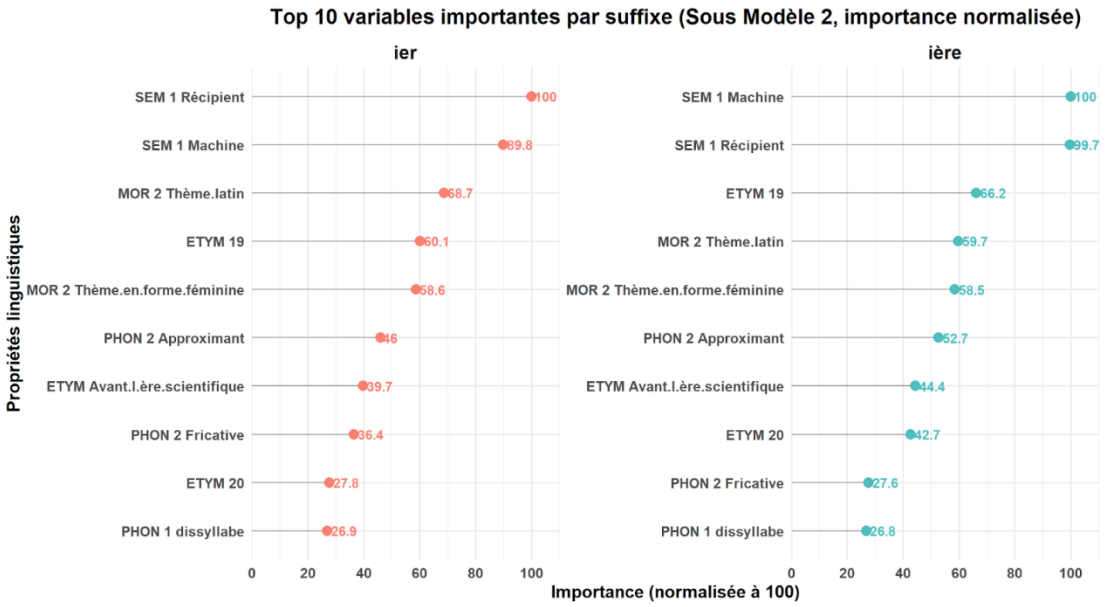

Figure 6.19 Importance des variables linguistiques dans la prédiction des quatre schémas instrumentaux, selon le Sous-modèle 2. Les barres indiquent la contribution moyenne à l'amélioration

de la précision (*Mean Decrease Accuracy*)

Enfin, pour les suffixes *-ier*/*-ière*, nous avons réalisé une analyse visant à identifier les propriétés linguistiques les plus déterminantes en fonction de leur contribution à l'amélioration de la précision du modèle. La Figure 6.19 met en évidence les prédicteurs les plus influents pour chaque suffixe dans le Sous-modèle 2. Concernant ces deux suffixes, la sous-classe ontologique (SEM1) constitue la variable prépondérante pour prédire le choix de ces suffixes. Concrètement, selon le test du chi-carré, la distribution des sous-classes ontologiques Outil, Machine et Récipient varie de manière statistiquement significative entre les suffixes **-***ier* et **-***ière* ($\chi^2 = 17{,}72$, $p < 0{,}05$, *ddl* = 2). Ce résultat corrobore les deux premières modalités importantes dans la Figure 6.19. Par ailleurs, pour ces deux suffixes, les thèmes de la base du dérivé sélectionnés par le suffixe nominalisateur (MOR 2) contribuent également à leur différenciation. Dans le cas du suffixe **-***ière*, certaines formations utilisent le thème L (radical latin) afin d'éviter les contraintes dissimilatives d'ordre phonologique, comme l'illustre l'exemple NAZIERE (voir la section 5.2.1). En revanche, un tel phénomène demeure beaucoup plus rare dans le cas du schéma *-ier*.

Hormis les variables évoquées, la datation (ETYM) joue également un rôle dans la différenciation des suffixes. Comme mentionné en section 5.5, *-ier* est plus productif au XXe siècle qu'au XIXe, tandis que *-ière* suit une tendance inverse. Ainsi, cette dynamique temporelle constitue une autre variable distinctive. De plus, le mode d'articulation des segments finaux du radical (PHON2) apparaît comme une variable discriminante : les radicaux se terminant par une consonne approximante sont davantage associés à *-ier*, tandis que ceux se terminant par une nasale tendent à favoriser *-ière*, ce qui confirme les résultats de la section 5.2.2. Enfin, la taille phonologique du radical (PHON1) influence également le choix du suffixe. On observe que *-ier* présente une légère préférence pour les bases dissyllabiques, alors que *-ière* montre une légère préférence pour les bases monosyllabiques (cf. section 5.2.1).

6.5 Discussion

Dans ce chapitre, nous avons procédé à l'analyse de la compétition morphologique de la suffixation des noms d'instruments en français. Plus précisément, nous avons

abordé ce sujet sous trois aspects : la niche morphologique, les modes de compétition et la modélisation statistique.

Tout d'abord, concernant **la niche morphologique**. Comme mentionné dans le chapitre 2, Lindsay & Aronoff (2013) ont emprunté le concept de niche à la biologie pour l'appliquer à la compétition morphologique. Partant d'une approche monofactorielle, une modalité linguistique discriminante peut constituer une niche. Par la suite, Missud (2023) a développé cette définition dans une approche multifactorielle, en proposant qu'une niche d'un schéma morphologique soit composée de modalités discriminantes provenant de différents niveaux linguistiques. Cependant, nous avons constaté que cette définition ne permet pas de rendre compte des mécanismes de compétition entre les schémas morphologiques, notamment en raison de facteurs comme le chevauchement. Ainsi, dans ce chapitre, inspiré par Hutchinson (1957), nous proposons une nouvelle définition de la « niche » et de la « sous-niche », sur lesquels nous avons ensuite calculé le chevauchement des niches morphologiques. Concrètement, nous avons pris le chevauchement comme un indicateur important de la compétition des niches. Lors du calcul du chevauchement, nous n'avons pas utilisé la méthode de mesure de similarité de Salvadori & Huyghe (2024), basée sur celles de Sørensen (1948) et Odum (1950), car la similarité ne reflète pas toujours de manière adéquate la compétition directionnelle (Hurlbert 1978 : 69-70 ; Abrams & Rueffler 2009 : 812-813). Nous avons donc adapté la formule du taux de chevauchement (Hurlbert 1978 : 69-70), proposé une nouvelle formule et calculé le chevauchement des niches de différents schémas morphologiques. Nous avons constaté que le chevauchement entre les suffixes *-eur*, *-euse*, *-ette* et *-ant* était particulièrement élevé. La niche du suffixe *-eur* recouvre, par exemple, des suffixes comme *-ail* et *-rice*, qui contiennent moins de modalités activées.

En outre, nous avons utilisé UpsetR pour visualiser certains schémas. Ces visualisations nous ont permis de mieux comprendre les sous-niches de chaque suffixe ainsi que le nombre de lexèmes construits par les modalités activées dans chaque sous-niche. Bien que certains suffixes aient un nombre limité de lexèmes dans leurs sous-niches, la visualisation des niches des suffixes comme *-eur*, *-euse*, *-oir* et *-ier* a montré quelles sous-niches étaient les plus représentatives dans la dérivation. Ainsi, par rapport aux définitions de la niche morphologique proposées par Lindsay & Aronoff (2013) et Missud (2023), notre nouvelle définition est plus proche du concept de niche en écologie et donne des voies pour la mesure des niches morphologiques. De plus,

comparé à la méthode de Salvadori et al. (2024), notre formule permet d'appréhender indirectement la largeur des niches et leur chevauchement, ce qui nous aide à mieux comprendre la répartition des modalités en jeu de ces schémas dans le lexique. En somme, nous avons proposé une nouvelle définition de la niche morphologique ainsi que des méthodes de mesure et de visualisation qui pourront être appliquées dans de futures recherches.

En ce qui concerne **les modes de compétition**, selon Lindsay & Aronoff (2013) et Aronoff (2019 ; 2023), le principe d'exclusion compétitive permet de mieux expliquer la compétition morphologique. Pour vérifier cette hypothèse, nous avons effectué une analyse diachronique de six sous-classes ontologiques pour les suffixes en jeu. Nous avons identifié trois modes de compétition, à savoir le mode de compétition absolue (Appareil), le mode de compétition alternative (Outil, Machine), le mode de compétition non-absolue (Récipient, Instrument de Captation, Substance Chimique). Parmi eux, hormis le mode de compétition absolue, les deux autres modes ne sont pas conformes au principe d'exclusion compétitive. Cela signifie que bien que le principe d'exclusion compétitive puisse expliquer certains phénomènes de compétitions morphologiques, il présente néanmoins certaines limites. En particulier, lorsque la productivité est faible, la compétition ne suffit quelquefois pas à éliminer les suffixes moins compétitifs, qui parviennent malgré tout à se maintenir durablement dans le système du lexique. Ainsi, ces résultats rejoignent les conclusions que j'avais déjà proposées dans Lin (2021), qui a mis en évidence des cas de coexistence durable entre plusieurs schémas au sein d'une même niche, sans exclusion ni spécialisation fonctionnelle nette. Huyghe & Lombard (2022) ont ensuite conceptualisé ce phénomène sous le nom *overabondance*, en montrant que le système dérivationnel peut tolérer, voire entretenir, des redondances de formes, ce qui rend le principe d'exclusion compétitive peu pertinent dans le domaine de la suffixation.

Enfin, nous avons réalisé une modélisation par forêts aléatoires afin de vérifier si la compétition morphologique relève effectivement d'un mécanisme multifactoriel. Les résultats montrent clairement que la compétition morphologique entre suffixes instruments est de nature multifactorielle, ce qui révèle les limites d'une explication monofactorielle et confirme ainsi la supériorité de l'approche multifactorielle.

Trois modèles ont été construits (modèle général, sous-modèle 1, sous-modèle 2). Nous avons constaté que, dans ces trois modèles, la variable la plus discriminante est la sous-classe ontologique du dérivé, ce qui souligne son importance particulière dans

le choix des suffixes instrumentaux. Dans notre cadre théorique de la NMC, cette propriété sémantique intervient dès l'identification du référent et prend sa trajectoire cognitive jusqu'à entrer dans le système morphologique, où elle joue finalement un rôle décisif, en tant que contrainte sémantique, lors de la compétition morphologique.

Outre la sous-classe ontologique du dérivé, d'autres variables linguistiques se sont révélées également importantes. Concernant le modèle général, la variable syntaxique (catégorie syntaxique de la base du dérivé) et deux variables phonologiques (taille syllabique du radical et segment final du radical) jouent également un rôle discriminant dans la prédiction du choix des schémas suffixaux. À titre d'exemple, pour la variable PHON1 (taille syllabique du radical), le suffixe *-oir* privilégie des radicaux monosyllabiques, tandis que les suffixes *-eur* et *-ant* tendent à sélectionner des radicaux polysyllabiques, lesquels sont rejetés par des suffixes tels que *-euse*, *-oir*, *-ier* et *-ière*. Quant à la variable PHON2, le suffixe *-eur* présente une préférence nette pour les segments finaux occlusifs et rejette les segments nasaux et approximants. En ce qui concerne le sous-modèle 1, en dehors des variables déjà mentionnées, la variable MOR2 (espaces thématiques) s'avère également primordiale pour prédire le choix des schémas instrumentaux. Par exemple, le thème 13 permet une distinction claire entre certains noms construits par les suffixes *-eur/-rice* et ceux construits par *-euse*/*-oir*.

Ainsi, par rapport aux modèles élaborés dans les travaux de Bonami & Thuilier (2019), Missud (2023), Bobkova (2023) et Thuilier et al. (2023), nous avons ajouté les variables « sous-classe ontologique du dérivé » et « espaces thématiques » comme variables explicatives pour prédire le choix des suffixes instrumentaux, contribuant ainsi à une meilleure interprétation du mécanisme de compétition morphologique. Plus précisément, l'importance élevée de la variable « sous-classe ontologique du dérivé » indique que certaines propriétés sémantiques exercent une influence réelle sur le choix des suffixes, soulignant que l'interaction entre les êtres humains et les entités du monde réel est liée à la nature même de la construction des lexèmes. Dans la philosophie générale de la linguistique cognitive, la formation des mots n'est pas uniquement conditionnée par des contraintes internes à la langue. En effet, selon cette approche, le système linguistique reflète nécessairement les interactions concrètes entre les êtres humains et leur environnement objectif. Autrement dit, le processus de dérivation est conçu comme une véritable co-construction cognitive, fondée sur l'expérience humaine (perceptive, sensorimotrice, etc.) du monde réel (adapté de Langacker 1987). En outre, l'importance de la variable espaces thématiques rejoint l'observation de Bauer et al.

(2013 : 615) concernant le sujet de l'ordre des suffixes, selon lequel même si certaines études estiment que l'influence de l'étymologie sur le processus de construction lexicale actuel est limitée, celle-ci conserve néanmoins une existence latente au niveau du traitement psycholinguistique dans le cadre de la compétition morphologique.

Il convient également de noter que la variable « profil sémantique » constitue précisément l'un des points centraux traités par la théorie NMC : d'où provient la base du dérivé, et quel est son lien cognitif avec le dérivé ? Comme indiqué dans les résultats obtenus dans la section 6.4, bien que cette variable ne soit pas prédictive, cela ne signifie pas qu'elle n'influence pas le choix des schémas instrumentaux une fois projetée au système morphologique *via* une trajectoire cognitive.

Deux raisons expliquent pourquoi cette variable n'est pas fortement discriminante : premièrement, l'existence d'une colinéarité entre variables ; deuxièmement, dans le cadre NMC (à voir le chapitre 7), concernant la suffixation des noms d'instrument, le locuteur, dans l'étape de profilage conceptuel, est souvent encadré par des profils sémantiques précis (patient, métonymie, verbe télique). Par conséquent, la plupart des schémas instrumentaux se concentrent sur un nombre restreint de profils sémantiques, ce qui limite leur discriminabilité. Cependant, cette observation ne diminue en rien leur importance ; au contraire, elle montre que ces profils sémantiques sont précisément les plus saillants pour les noms d'instruments. Ainsi, ce résultat reflète, dans une certaine mesure, la stratégie cognitive de nomination générale adoptée par les locuteurs lors du processus morphologique de suffixation.

Enfin, une limite de la présente étude réside dans le manque d'une analyse approfondie des propriétés sémantiques de la base verbale. Une piste de recherche future consistera à examiner si les verbes comportant une valeur privative ou de changement d'état influencent le choix des schémas suffixaux, et comment ces caractéristiques pourraient être formalisées dans le cadre théorique NMC.

Chapitre 7 Mécanismes et Représentations lexicales de la Néo-Morphologie Cognitive

Dans la section 3.1.3, nous avons présenté les grandes lignes et les notions fondamentales de la Néo-Morphologie Cognitive (ci-après NMC). Dans ce chapitre, nous allons développer cette nouvelle théorie à l'aide d'exemples représentatifs. Concrètement, la section 7.1 se concentre sur le détail de ce modèle. Ensuite, nous présenterons les représentations lexicales et prendrons trois noms d'instruments représentatifs, à savoir EPLUCHEUR, CAFETIERE, CHIMIQUIER, pour illustrer comment fonctionne la NMC afin d'interpréter leur formation dans la perspective cognitive.

7.1 Principes architecturaux du modèle

Comme indiqué dans la section 3.1.3, le cadre de la NMC repose sur une réélaboration critique de la *Morphologie de Construction* (CM), enrichie par des apports empruntés à d'autres théories cognitives. Toutefois, ces contributions ne forment pas un assemblage juxtaposé. Autrement dit, elles s'intègrent dans une architecture modulaire rigoureusement structurée. Chaque module ou sous-module issu de ces apports théoriques (profilage conceptuel, espaces d'intégration morphosémantique, restructuration métonymique, compétition morphologique, etc.) joue un rôle spécifique au sein de l'ensemble. Ces modules interagissent selon des logiques conditionnelles et rétroactives : le fonctionnement du modèle repose sur une dynamique opératoire, et non sur une linéarité rigide. En effet, certains modules peuvent être facultatifs, d'autres récurrents, et des opérations comme l'échec ou le retour à une étape antérieure font partie intégrante du modèle. Cette souplesse permet à la NMC de formaliser des trajectoires de dérivation complexes, conditionnées à la fois par les informations linguistiques de la base, les informations du dérivé, ainsi que par les contraintes linguistiques propres aux schémas morphologiques.

Concrètement, la NMC repose sur quatre modules principaux ainsi qu'une trajectoire cognitive qui structure le déroulement du processus dérivationnel. Plus précisément, les quatre modules sont : le profilage conceptuel, la restructuration métonymique, l'intégration morphosémantique et le système morphologique. La

trajectoire cognitive se décline en trois types selon la nature de la base : trajectoire nominale, trajectoire adjectivale et trajectoire verbale. Selon cette architecture, nous présenterons successivement chacun de ces modules et chacune de ces trajectoires cognitives.

7.1.1 Module de profilage conceptuel

Nous commençons par le module de profilage conceptuel, qui est constitué de plusieurs théories évoquées, telles que le *Modèle d'enracinement morphologique* (Schmid 2016), le *Lexique génératif* (Pustejovsky 1996), la métaphore conceptuelle (Lakoff & Johnson 1980), ainsi que la métonymie conceptuelle (Lakoff 1987). Concrètement, le profilage conceptuel vise à déterminer le référent ou la propriété le plus saillant dans une scène conceptuelle donnée (Schmid 2016 : 105). Cette opération de profilage s'accompagne de l'activation de sa représentation sémantique, formalisée dans le cadre de cette étude par la structure de qualia (Pustejovsky 1996). Le profilage conceptuel a pour fonction de sélectionner, parmi les quatre rôles définis dans la structure de qualia du dérivé (agentif, télique, formel, constitutif), le rôle le plus saillant et pertinent du point de vue de la dérivation, afin de déterminer une base lexicale appropriée. Cette base lexicale appropriée peut, après le profilage conceptuel ou de restructuration métonymique (mécanismes détaillés par la suite), devenir une base morphologique, c'est-à-dire le lexème effectivement intégré dans le système morphologique. Dès le départ du profilage conceptuel, deux types de profilage peuvent être distingués selon l'étendue du champ conceptuel mobilisé et la structure sémantique activée : d'une part, le profilage interne du référent, fondé sur la saillance locale d'un constituant ou d'une propriété ; d'autre part, le profilage événementiel, fondé sur une structure de procès impliquant plusieurs entités. Dans ce qui suit, nous procèderons à une analyse détaillée de chacun de ces deux types de profilage conceptuel.

[1] Le profilage interne du référent

Ce type de profilage conceptuel s'applique dans le cadre du référent lui-même, sans activation d'un événement complexe. Il repose sur la mise en saillance d'un élément interne du référent à construire, tel qu'un composant matériel, une propriété sémantique ou une image perceptive, qui constitue un ancrage cognitif pour la formation du dérivé. Selon la nature de la projection activée, deux types de projections

sont possibles :

- **Projection métonymique (Lakoff 1987)**

Lorsque l'élément saillant appartient au rôle constitutif dans la structure de qualia du référent et peut le représenter de manière conventionnelle ou motivée, on observe une métonymie conceptuelle. Le constituant devient alors base lexicale du dérivé (*par ex.* PLUME > PLUMAIL), sans changement de domaine sémantique (c'est-à-dire l'espace conceptuel de connaissances et d'expériences qu'un mot mobilise pour être interprété.).

- **Projection métaphorique (Lakoff & Johnson 1980)**

Lorsque l'élément saillant évoque un autre concept perçu comme analogue sur le plan sémantique ou perceptif, une association analogique inter-domaines est déclenchée. Cette correspondance conceptuelle produit une métaphore cognitive, qui sert de base à la formation lexicale (*par ex.* FENOUIL > FENOUILLETTE). Bien que motivée par une saillance interne, cette trajectoire repose sur une reconfiguration du référent *via* un changement de domaine.

[2] **Le profilage événementiel (Schmid 2016)**

Ce type de profilage concerne un rôle identifié dans la structure de qualia (agentif, télique, ou formel), susceptible d'activer une structure argumentale au sein de laquelle se distribuent des rôles sémantiques au sens des scénarios (Fillmore 1968, Schmid 2016). Ces rôles sémantiques contribuent à la détermination d'un profil sémantique, au sens du modèle de la NMC, c'est-à-dire le rôle ou la contribution que la base assume dans la structure de qualia du référent à construire. Le profilage événementiel vise ainsi à projeter une trajectoire cognitive à partir d'un rôle dans la structure de qualia, enrichi de rôles sémantiques spécifiques, vers un profil sémantique intégré. Dans ce cadre, la base sélectionnée peut relever de deux types distincts :

- **Si la base est un verbe,** il correspond généralement au verbe impliqué dans la structure argumentale, lequel relève de l'un des quatre rôles de la structure de qualia (rôle télique, rôle formel, rôle constitutif et rôle agentif). Ce verbe constitue le moteur sémantique principal du rôle en question et sert de point de départ à la construction du dérivé. Il convient de souligner que ce verbe ne

coïncide pas nécessairement avec le prédicat exprimé au niveau syntactico-sémantique du dérivé (*par ex.* RASER > RASOIR), mais peut également provenir d'un autre rôle dans la structure de qualia, jouant ainsi le rôle d'un verbe sous-jacent dans la représentation sémantique (*par ex.* SONNER > SONNETTE).

- **Si la base est un nom,** celui-ci peut remplir un des rôles sémantiques dans la structure argumentale du verbe, lequel est lui-même issu d'un des rôles définis dans la structure de qualia (*par ex.* CAFE > CAFETIERE). Dans ce cas, la base nominale est dérivée non pas d'un composant physique du référent, mais d'un participant conceptuel impliqué dans un scénario fonctionnel ou télique.

À l'issue de ce premier module, une partie des bases sélectionnées devient des bases morphologiques et accède aux modules suivants jusqu'au système morphologique, tandis qu'une autre partie ne le devient pas. (cf. la Figure 3.5 dans la section 3.1.3) C'est précisément ce second module que nous allons maintenant examiner : le module de restructuration métonymique.

7.1.2 Module de restructuration métonymique

Certaines bases potentielles issues du profilage conceptuel ne peuvent pas directement devenir des bases morphologiques, et par conséquent accéder au système morphologique, en raison de leur structure conceptuelle complexe ou compositionnelle. C'est notamment le cas : (i) lorsque le référent profilé correspond à un syntagme dont la sémantique repose sur une composition ou une hiérarchisation de propriétés ; (ii) lorsque le rôle sémantique issu du profilage événementiel fait l'objet d'une métonymie supplémentaire, mettant en avant une propriété saillante pour désigner le référent. Ces bases nécessitent alors une opération intermédiaire de restructuration métonymique (Lakoff 1987), qui constitue le second module de notre modèle.

Dans ce module, la dérivation ne s'appuie pas sur l'ensemble du référent profilé, mais sur l'extraction d'une propriété saillante. Par exemple, dans le lexème CHIMIQUIER, le référent profilé est "produit chimique", mais la base effectivement utilisée est CHIMIQUE, résultat d'une restructuration métonymique qui extrait la propriété la plus saillante (la catégorie de ce produit). Cette opération permet d'activer une base intégrable tout en maintenant la cohérence sémantique avec le référent de départ.

7.1.3 Trajectoire cognitive

Une fois les deux premiers modules appliqués, à savoir le profilage conceptuel et la restructuration métonymique, certaines bases accèdent au statut de base morphologique et peuvent dès lors être traitées dans les modules suivants. Mais comment les informations linguistiques des bases morphologiques et les informations sémantiques du dérivé sont-elles transmises vers les modules suivants ? C'est ce passage que prend en charge le troisième module de notre modèle : la trajectoire cognitive. Précisément, la trajectoire cognitive de la base morphologique désigne la voie cognitive projetée par une base morphologique, portant à la fois les informations sémantiques et formelles de cette base, ainsi que les informations sémantiques du référent à construire (*par ex.* la sous-classe ontologique). Elle est nommée selon le profil sémantique que la base réalise dans la structure de qualia du lexème à construire. Elle se décline en trois grands types :

[1] Trajectoire nominale : Voies sémantiques associées à des bases nominales, classées selon leur profil sémantique :

- **participant d'une structure événementielle** (*par ex.* CAFÉ > CAFETIÈRE) ;
- **projection métonymique** :
  - **directe** (*par ex.* PLUME > PLUMAIL) ;
  - **restructurée** (*par ex.* PRODUIT EN VRAC > VRAC) ;
- **projection métaphorique** (*par ex*. FENOUIL > FENOUILLETTE).

[2] Trajectoire adjectivale : Voies sémantiques associées à des bases adjectivales, classées selon leur profil sémantique :

- **projection métonymique** :
  - **directe** (*par ex.* BLANC > BLANCHET) ;
  - **restructurée** (*par ex.* PRODUIT CHIMIQUE > CHIMIQUE) ;

[3] Trajectoire verbale : Voies sémantiques associées à des bases verbales issues de trois rôles spécifiques dans la structure de qualia :

- **rôle télique** (*par ex.* VERSER > VERSEUR) ;
- **rôle agentif** (*par ex.* PARLER > PARLEUR) ;
- **rôle formel** (*par ex.* SONNER > SONNETTE).

En définitive, les informations linguistiques des bases morphologiques et les informations sémantiques des dérivés sont conjointement transmises par les trajectoires cognitives différentes. Mais où ces informations sont-elles effectivement traitées et

intégrées ? C'est ce que nous allons examiner dans le module suivant : les espaces d'intégration morphosémantique.

7.1.4 Espaces d'intégration morphosémantique

Les espaces d'intégration morphosémantique s'inspirent de la théorie de l'intégration conceptuelle (Fauconnier & Turner 2002). Ce module repose sur trois espaces mentaux, à savoir l'espace de la base morphologique, l'espace du dérivé, ainsi que l'espace intégrant. Il se situe en amont du système morphologique, assurant la jonction entre le système sémantique et le système morphologique.

De plus, selon la Figure 7.1, les informations cognitives fournies par la trajectoire cognitive sont séparées et projetées dans deux espaces distincts, à savoir l'espace de la base morphologique et l'espace du dérivé. Ensuite, certaines informations issues de ces deux espaces sont projetées ensemble dans l'espace intégrant. C'est l'espace intégrant qui distingue d'abord les informations linguistiques de la base morphologique pour la dérivation et les informations sémantiques pertinentes du dérivé (telles que le rôle sémantique ou la classe ontologique : agent, instrument, lieu, etc.). Ensuite, il organise conjointement ces informations pour faire émerger la sémantique du schéma morphologique, c'est-à-dire la condition morphosémantique préalable, qui sert d'interface avec le système des schémas morphologiques.

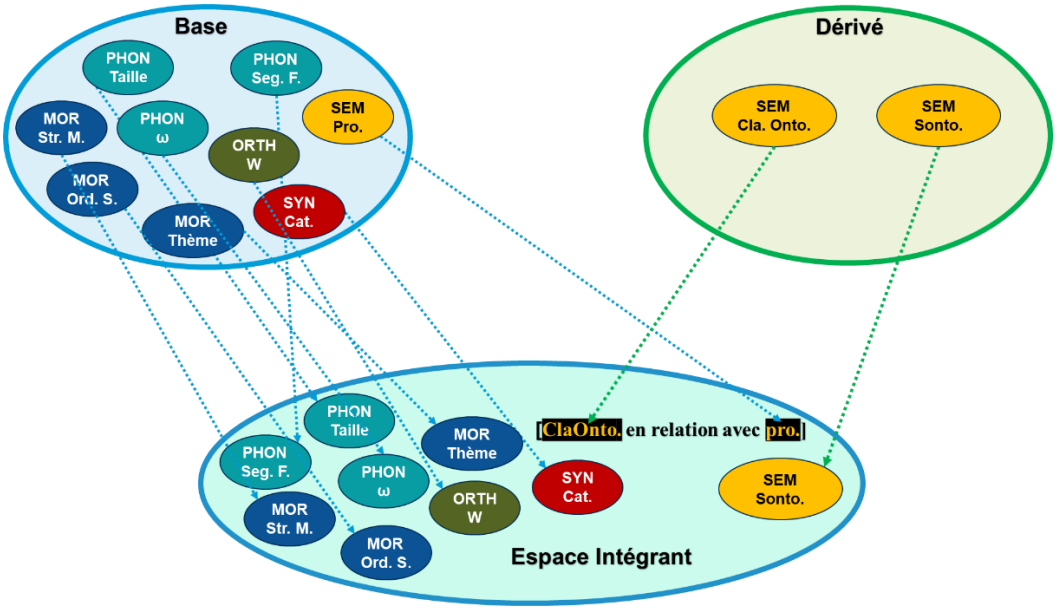

Figure 7.1 Espaces d'intégration morphosémantique

7.1.5 Système morphologique

Une fois les informations pertinentes intégrées dans les espaces morphosémantiques, celles mobilisables pour la dérivation accèdent au système morphologique. Conformément au positionnement proposé par les morphologues (Aronoff 1994 ; Booij 2010), nous considérons que la morphologie est autonome dans le système linguistique, en ce sens qu'elle repose sur des principes et des unités propres, distincts de ceux de la syntaxe ou de la sémantique. C'est pourquoi le système morphologique est, dans NMC, envisagé comme un module indépendant.

C'est dans ce module que s'exerce le mécanisme de compétition morphologique, que nous formalisons de manière inédite. Dans le chapitre 2, nous avons présenté les recherches existantes sur la compétition morphologique. Cependant, ces travaux, bien qu'ils décrivent de nombreux faits empiriques, n'ont pas formalisé la compétition comme un véritable mécanisme intégré au système dérivationnel. Elle reste généralement abordée comme un simple phénomène, sans faire l'objet d'une modélisation explicite de son fonctionnement interne (Nagano et al. 2024 : 3).

Dans notre modèle, nous proposons au contraire de définir la compétition morphologique comme un mécanisme propre au système morphologique. Nous considérons cette modélisation qui vise à combler une lacune théorique dans la description des dynamiques internes du système dérivationnel. Ce mécanisme intervient après l'étape des espaces d'intégration morphosémantique, lorsque plusieurs schémas suffixaux ont été validés sur le plan sémantique, et qu'il s'agit de déterminer lequel sera effectivement sélectionné pour la construction du lexème. Il repose sur quatre facteurs fondamentaux : 1) les contraintes linguistiques (Bauer et al. 2013 ; Plag 2018) ; 2) le blocage (Aronoff 1976, 1994) ; 3) la productivité (Bauer 2001 ; Bauer et al. 2013 ; Dixon 2014) ; 4) les facteurs socioculturels (Bauer 2001 ; Bauer et al. 2013 ; Lin 2021).

Parmi ces quatre facteurs, les contraintes linguistiques méritent une présentation plus fine, car ce sont elles qui déterminent le plus souvent le schéma morphologique finalement sélectionné. Ces contraintes se déclinent en plusieurs types (Booij 2010 ; Plag 2018) :

- Contraintes phonologiques : taille syllabique, segment final de la base du dérivé, prosodie, etc.
- Contraintes morphologiques : structure morphologique de la base du dérivé,

thèmes de base du dérivé, classes flexionnelles, ordre des affixes, etc.

- Contrainte syntaxique : catégorie syntaxique de la base du dérivé, etc.
- Contraintes sémantiques : profil sémantique assumé par la base dans la représentation du dérivé, sous-classe ontologique du dérivé, etc.

Parmi ces contraintes, hormis les contraintes morphologiques proprement dites, les autres relèvent d'interfaces avec les niveaux phonologique, syntaxique ou sémantique. En effet, les contraintes morphologiques sont les seules à constituer des opérations internes au système morphologique lui-même, c'est-à-dire qu'elles opèrent sur ses propres unités. Dans cette perspective, nous proposons d'intégrer l'ordre des affixes (Bauer et al. 2013 ; Saarinen & Hay 2014 ; Lieber 2021) au sein des contraintes morphologiques déjà recensées par Booij (2010) et Plag (2018). Parce que même si la structure morphologique de la base coïncide parfois avec l'ordre des affixes, il existe des cas où ces deux dimensions divergent, avec des conséquences remarquables sur les possibilités de la construction des lexèmes. C'est pourquoi nous considérons que l'ordre des affixes doit être pris en compte comme une contrainte morphologique à part entière, en complément de la structure morphologique de la base.

A titre d'exemple, dans le lexème SENSIBILISATION, la base est SENSIBILISER, dont la structure morphologique correspond à la suffixation en *-iser* sur une base adjectivale. De même, concernant l'ordre des affixes, le suffixe interne est également *-iser*. En revanche, dans DESENSIBILISATION, la base est DESENSIBILISER, dont la structure morphologique est fondée sur la préfixation *dé-*. Ici, le suffixe interne désigne, dans une combinaison suffixale, le suffixe qui est plus proche du radical. En revanche, le suffixe externe désigne le suffixe qui est plus loin du radical et s'ajoute au suffixe interne. Toutefois, selon l'ordre des affixes, le suffixe interne est *-iser*, car ce dernier est plus proche du suffixe externe *-ion*. Cette divergence reflète en partie la hiérarchie des contraintes morphologiques activées lors de la dérivation. Comme le montrent Villoing & Gueye (2024), le suffixe *-ment* tend à préférer une base construite par la préfixation *dé-* à valeur privative. Toutefois, certains lexèmes s'écartent de cette tendance : les bases en *dé*-X-*iser* privilégient souvent *-ion*, alors que *-ion* ne montre pas cette préférence pour les bases en *dé*-X sans *-iser*. En d'autres termes, c'est le suffixe *-iser*, et non le préfixe *dé-*, qui conditionne la sélection des suffixes. La contrainte morphologique l'emporte donc, *a priori*, sur la contrainte sémantique. Cela suggère que dans certains cas, l'ordre des affixes prévaut sur la structure morphologique

de la base : certains suffixes se montrent plus sensibles au suffixe interne qu'à la structure morphologique de la base du dérivé.

Par ailleurs, le blocage (Aronoff 1976, 1994), la productivité (Bauer 2001 ; Bauer et al. 2013 ; Dixon 2014), ainsi que les facteurs socioculturels (Bauer 2001 ; Bauer et al. 2013 ; Lin 2021) interviennent également dans le choix des suffixes. Par exemple, le blocage[43] peut empêcher la création d'un dérivé pourtant formellement possible, lorsqu'un mot concurrent existe déjà dans le lexique (Aronoff & Lindsay 2014 : 70) (*par ex. went* bloque **goed*).

Tous ces facteurs interviennent en réalité dans le cadre du mécanisme de compétition morphologique (Lindsay & Aronoff 2013 ; Nagano et al. 2024), qui conditionne la sélection finale du schéma morphologique. Leur poids respectif varie selon les classes lexicales concernées, influençant directement le choix du suffixe et, par conséquent, la forme finale du dérivé.

7.2 Représentation lexicale de NMC

Après avoir présenté les modules de Néo-Morphologie Cognitive, il convient maintenant d'examiner les représentations lexicales en vue de bien illustrer le résultat de tout le processus cognitif pour construire les lexèmes.

7.2.1 Architecture de la représentation lexicale en NMC

La représentation lexicale proposée dans la Néo-Morphologie Cognitive reprend, dans une certaine mesure, les composantes formelles issues de la CM, notamment les niveaux phonologique, morphosyntaxique et sémantique. Toutefois, à partir de ces composantes initiales, plusieurs extensions ont été apportées. D'une part, nous avons introduit explicitement un niveau orthographique (ORTHO), et d'autre part, nous avons dissocié la composante morphologique en tant que module autonome, distinct de la composante morphosyntaxique. Cette décision s'appuie sur plusieurs travaux récents : selon Pustejovsky & Batiukova (2019 : 11) ainsi qu'Audring & Jackendoff (2025 : 103), le niveau orthographique est intégré dans la représentation lexicale. Par ailleurs, Aronoff (2019) souligne que la niche orthographique permet de différencier des

[43] Le blocage se définit comme « La tendance d'un mot déjà existant à empêcher la dérivation d'un autre mot ayant le même sens » (Lieber 2021 : 252).

suffixes en compétition, ce qui renforce la pertinence de ce niveau dans notre modèle. Quant au niveau morphologique, comme développé en détail dans la sous-section 3.1.3, la morphologie obéit à des principes propres et constitue un niveau distinct, non réductible au morphosyntaxique. Enfin, sur le plan sémantique, étant donné la représentation moins développée de ce niveau dans la CM, nous avons enrichi la représentation sémantique en y intégrant la structure de qualia ainsi que la structure d'argument. De plus, concernant la structure argumentale de chaque rôle dans la structure de qualia, chaque argument est doté des rôles sémantiques.

Il convient de noter que, contrairement à la CM qui conçoit la représentation lexicale comme un simple triplet (forme, catégorie, sens), la Néo-Morphologie Cognitive propose une structure stratifiée, capable d'encoder les traces du processus dérivationnel et de conserver les paramètres cognitifs mobilisés lors de la formation du dérivé. D'un point de vue graphique, cette représentation fait apparaître les liens cognitifs entre la base et le dérivé au moyen de flèches directionnelles, tandis que le profil sémantique activé est souligné afin d'en marquer la saillance au sein de la structure. Dans les sous-sections suivantes, nous allons prendre trois lexèmes représentatifs de trois stratégies, afin d'illustrer plus clairement l'architecture de la représentation lexicale en Néo-Morphologie Cognitive.

7.2.2 Stratégies fondées sur les bases verbales

Concernant les stratégies fondées sur les bases verbales, il s'agit des éléments saillants qui sont des verbes dans la structure de qualia du dérivé après le profilage conceptuel. Nous prenons la représentation du lexème EPLUCHEUR comme exemple pour illustrer ce type de stratégie. Comme le montre la figure 7.2, après l'identification du référent « éplucheur » et le profilage conceptuel, dans la représentation sémantique du dérivé, le verbe EPLUCHER, issu du rôle télique dans la structure de qualia, est plus saillant et est choisi comme base. La représentation lexicale de ce verbe établit, via le profilage événementiel, la relation et le lien avec la représentation sémantique du dérivé. Ce verbe devient alors la base morphologique pour construire le dérivé.

Ensuite, les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont réunies, et elles sont transmises par la trajectoire cognitive « verbe télique » vers l'espace d'intégration morphosémantique, où les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont projetées respectivement dans l'espace de la base morphologique et

dans l'espace du dérivé. Ensuite, certaines informations de la base morphologique EPLUCHEUR et du dérivé sont projetées dans l'espace d'intégration morphosémantique, où, en plus des informations linguistiques de ces deux éléments, la condition morphosémantique émerge : [Instrument servant à éplucher] → [Instrument servant à PRED]. Cette condition morphosémantique active le module du système morphologique, et avec les informations linguistiques portées par EPLUCHER et les informations sémantiques portées par le dérivé, cette condition entre dans le système morphologique. Dans ce module, sous l'activation de la condition morphosémantique, les schémas suffixaux instrumentaux sont activés, tels que les suffixes *-eur*, *-euse*, *-oir*. Dès que les informations linguistiques de la base morphologique EPLUCHER et les informations sémantiques du dérivé (*par ex.* la sous-classe ontologique Outil) sont définies, la compétition morphologique se déclenche. À l'issue de cette compétition, le suffixe *-eur* l'emporte. Voilà le lexème construit : EPLUCHEUR.

Figure 7.2 La représentation du lexème EPLUCHEUR

7.2.3 Stratégies fondées sur les bases non-verbales

Après avoir présenté les stratégies fondées sur les bases verbales, nous allons présenter celles fondées sur les bases non verbales. Concrètement, il s'agit des profils sémantiques qui sont des noms ou des adjectifs dans la structure de qualia du dérivé, après le profilage conceptuel (ou la restructuration métonymique pour les syntagmes). Selon les théories présentées dans les sous-sections 7.1, nous pouvons identifier deux grandes catégories de stratégies fondées sur les bases non verbales : le profilage événementiel, comme dans les stratégies fondées sur les bases verbales, et le profilage interne (métonymie directe ou indirecte – c'est-à-dire une restructuration métonymique – et métaphore).

Nous commençons par le profilage événementiel et prenons la représentation du lexème CAFETIERE comme exemple pour illustrer ce type de stratégie. Comme l'illustre la figure 7.3, après l'identification du référent « cafetière » et le profilage conceptuel, dans la représentation sémantique du dérivé, le nom CAFE_1 est identifié comme un rôle sémantique de patient dans la structure argumentale d'un verbe issu du rôle télique dans la structure de qualia du référent. Ce nom devient ainsi le profil sémantique et est choisi comme base. Par contre, le deuxième sens de CAFE_2 n'est pas sélectionné. La représentation lexicale de ce nom établit, via le profilage événementiel, la relation et le lien avec la représentation sémantique du dérivé. Ce nom devient alors la base morphologique pour construire le dérivé.

Ensuite, les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont réunies, et elles sont transmises par la trajectoire cognitive « Patient » vers l'espace d'intégration morphosémantique, où les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont projetées respectivement dans l'espace de la base morphologique et dans l'espace du dérivé. Ensuite, certaines informations de la base morphologique CAFE et du dérivé sont projetées dans l'espace d'intégration morphosémantique, où, en plus des informations linguistiques de ces deux éléments, la condition morphosémantique émerge : [Instrument sert à infuser du café] → [Instrument sert à *verbe sous-jacent* BASE]. Cette condition morphosémantique active le module du système morphologique, et avec les informations linguistiques portées par CAFE et les informations sémantiques portées par le dérivé, cette condition entre dans le système morphologique. Dans ce module, sous l'activation de la condition morphosémantique,

les schémas suffixaux instrumentaux sont activés, tels que les suffixes *-ier*, *-ière*, *-eur*, etc. Dès que les informations linguistiques de la base morphologique CAFE et les informations sémantiques du dérivé (*par ex.* la sous-classe ontologique Récipient) sont définies, la compétition morphologique se déclenche. À l'issue de cette compétition, le suffixe *-ière* l'emporte, car la contrainte syntaxique influe beaucoup sur le choix des suffixes. Voilà le lexème construit : CAFETIERE.

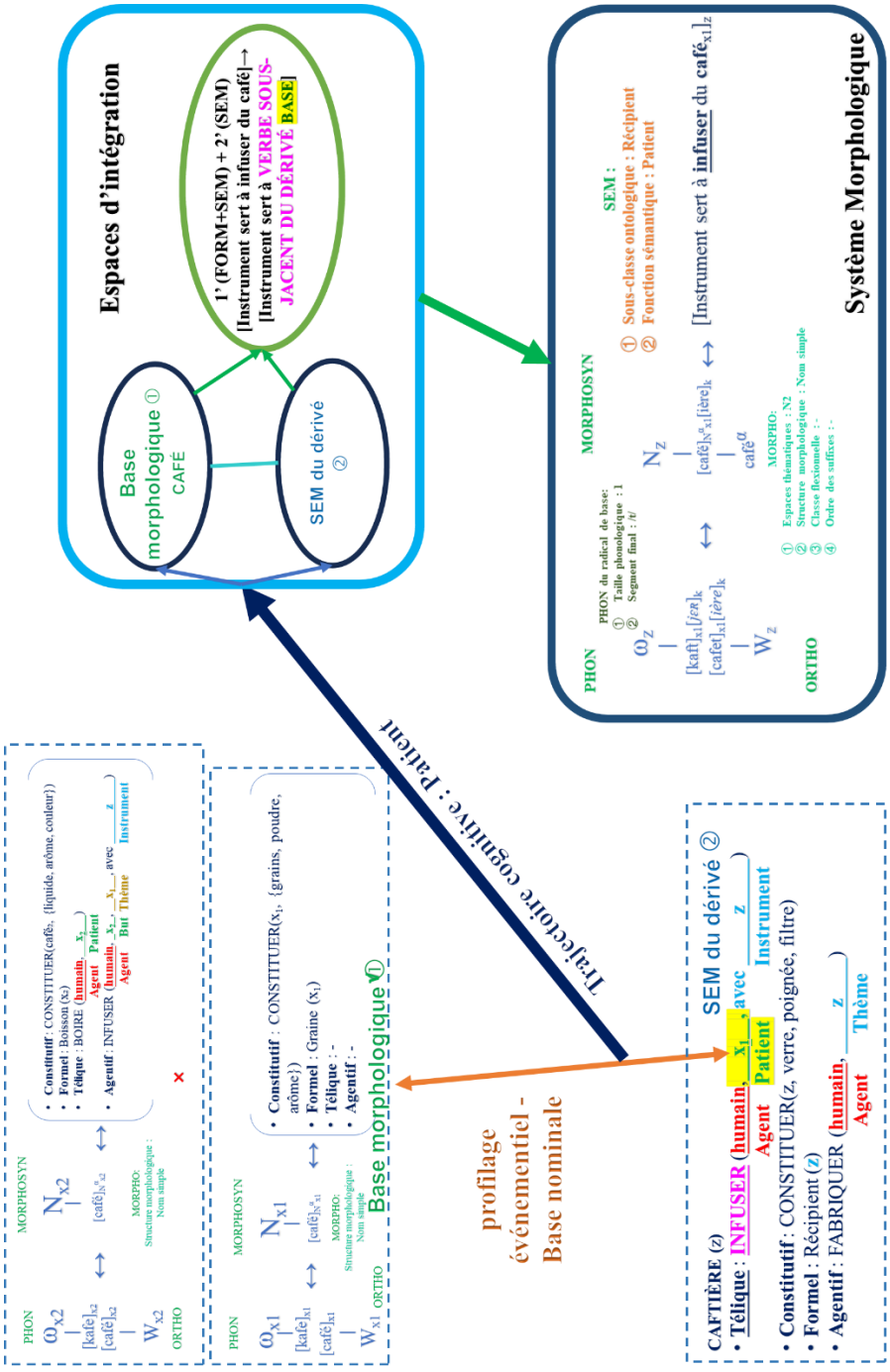

Figure 7.3 La représentation du lexème CAFETIERE

Nous terminons la section 7.2 par le profilage interne et prenons la représentation

du lexème CHIMIQUIER comme exemple pour illustrer ce type de stratégie. Comme l'illustre la figure 7.4, après l'identification du référent « chimiquier » et le profilage conceptuel, dans la représentation sémantique du dérivé, le syntagme *produit chimique* est identifié comme un rôle sémantique de thème dans la structure argumentale d'un verbe issu du rôle télique dans la structure de qualia du référent. La représentation lexicale de ce nom établit, *via* le profilage événementiel, la relation et le lien avec la représentation sémantique du dérivé. Cependant, ce syntagme ne devient pas la base morphologique. Il doit passer par le module de restructuration métonymique pour permettre la substitution de la base. À travers ce module, sous l'effet d'une métonymie conceptuelle, la catégorie chimique, en tant qu'élément le plus saillant, en vient à représenter l'ensemble « produit chimique ». D'où la base morphologique CHIMIQUE. Il convient de noter que cette base morphologique CHIMIQUE équivaut, sur le plan sémantique, à l'expression « produit chimique ». En revanche, sur le plan formel, elle porte les informations linguistiques du lexème CHIMIQUE. C'est pourquoi nous rendons son indice « x » en italique et l'annotons sous la forme x', au moyen d'un prime, afin d'indiquer sa nature linguistique particulière.

Ensuite, les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont réunies, et elles sont transmises par la trajectoire cognitive « Métonymie » vers l'espace d'intégration morphosémantique, où les informations linguistiques portées par la base et les informations sémantiques portées par le dérivé sont projetées respectivement dans l'espace de la base morphologique et dans l'espace du dérivé. Ensuite, certaines informations de la base morphologique CHIMIQUE et du dérivé sont projetées dans l'espace d'intégration morphosémantique, où, en plus des informations linguistiques de ces deux éléments, la condition morphosémantique émerge : [Instrument sert à transporter les produits chimiques] → [Instrument sert à *verbe sous-jacent* des entités liées à BASE]. Cette condition morphosémantique active le module du système morphologique, et avec les informations linguistiques portées par CHIMIQUE et les informations sémantiques portées par le dérivé, cette condition entre dans le système morphologique. Dans ce module, sous l'activation de la condition morphosémantique, les schémas suffixaux instrumentaux sont activés, tels que les suffixes *-ier*, *-ière*, *-ette*, etc. Dès que les informations linguistiques de la base morphologique CHIMIQUE et les informations sémantiques du dérivé (*par ex.* la sous-classe ontologique Machine) sont définies, la compétition morphologique se déclenche. À l'issue de cette compétition, le suffixe

-ier l'emporte. Voilà le lexème construit : CHIMIQUIER.

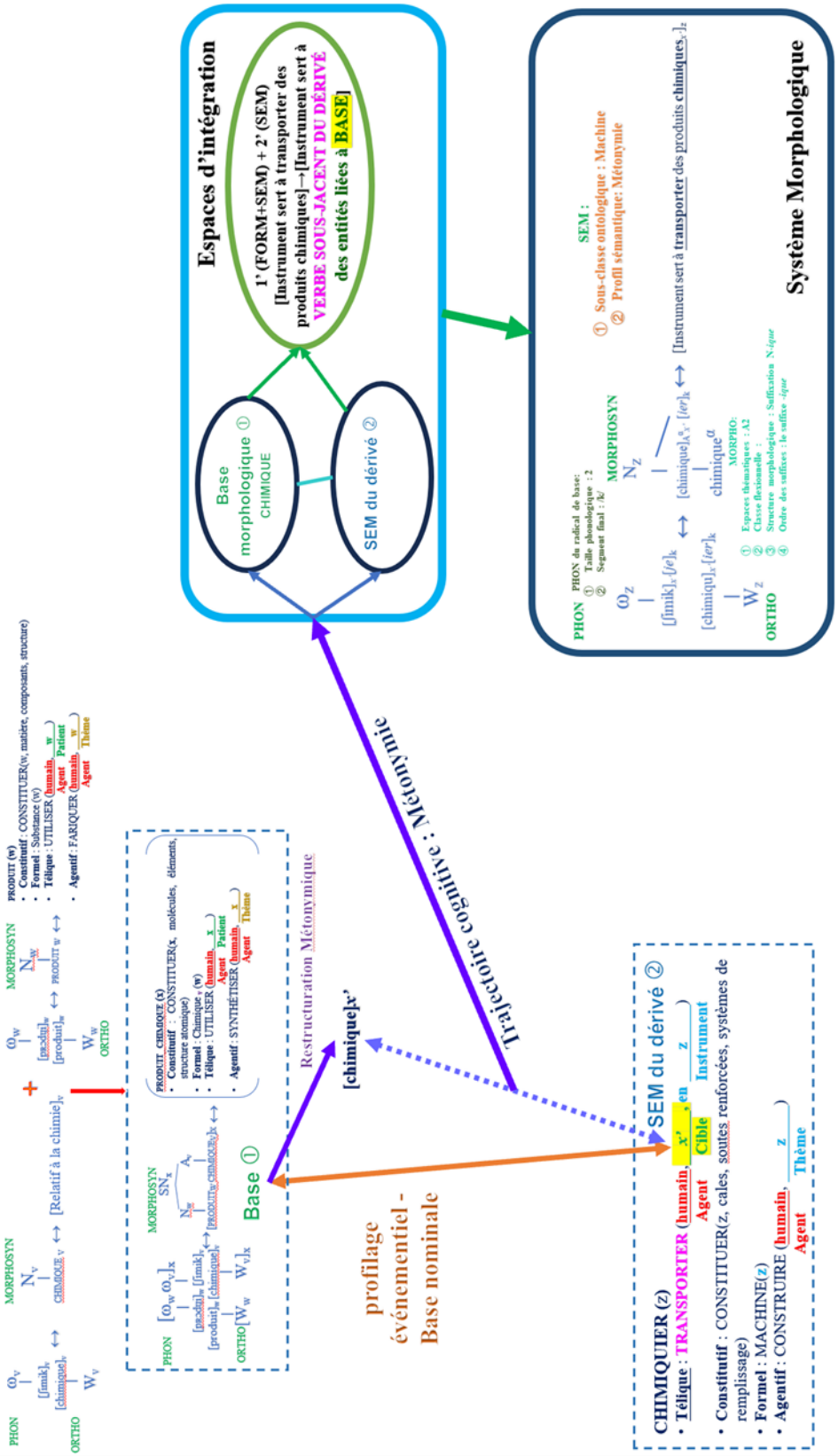

Figure 7.4 La représentation lexicale du lexème CHIMIQUIER

Jusqu'ici, nous avons présenté l'ensemble des éléments constitutifs du modèle de

la NMC : la *Morphologie de Construction* comme point de départ, les cadres complémentaires intégrés pour pallier ses limites, et surtout l'architecture propre à la *Néo-Morphologie Cognitive*. La NMC constitue ainsi une nouvelle proposition théorique, en ce qu'elle ne se limite pas à enrichir les cadres existants, mais vise à refonder l'analyse morphologique sur des principes cognitifs explicites, en articulant les dimensions conceptuelles, sémantiques et morphologiques dans un système procédural cohérent. Ce cadre, encore en construction, a vocation à être progressivement consolidé dans des travaux ultérieurs.

Conclusion générale et perspectives

Conclusion générale

Cette thèse a pour objectif d'étudier la formation des noms d'instruments par suffixation en français et la compétition entre quatorze suffixes spécialisés dans cette formation (*-eur*, *-euse*, *-rice*, *-oir*, *-oire*, *-ier, -ière*, *-et*, *-ette*, *-on*, *-ant*, *-ante*, *-ail*, *-aille*), dans la perspective de la linguistique cognitive, et plus particulièrement à travers le cadre théorique de la Néo-Morphologie Cognitive (NMC), élaboré au cours de ce travail. Pour ce faire, nous avons constitué une base de données à partir du *Grand Robert* (2005), comprenant 1338 noms d'instruments construits par ces suffixes et annotés selon neuf variables linguistiques réparties sur cinq niveaux : phonologique, morphologique, syntaxique, sémantique et diachronique. L'analyse a été conduite selon une méthode mixte, combinant la méthode qualitative et la méthode quantitative afin de croiser les interprétations linguistiques et les résultats empiriques.

Nous avons d'abord mené une étude descriptive complétée par des tests du chi-carré pour mettre en évidence les propriétés préférentielles associées à chaque suffixe. Nous avons ensuite proposé un indicateur original pour mesurer le chevauchement des niches morphologiques, dont la visualisation a été réalisée à l'aide de l'outil UpsetR. Une analyse diachronique a également été menée, ce qui nous a permis d'identifier trois modes de compétition suffixale. Enfin, une modélisation par forêt aléatoire a été mise en œuvre afin d'identifier les variables et modalités les plus discriminantes, susceptibles de prédire le choix des suffixes instrumentaux. Dans ce qui suit, nous présenterons les résultats majeurs de cette recherche doctorale ainsi que les contributions qu'elle apporte à l'étude de la dérivation morphologique et à la linguistique cognitive.

1. Néo-Morphologie Cognitive (NMC)

Dans le cadre de cette recherche doctorale, nous avons proposé un nouveau cadre théorique : la Néo-Morphologie Cognitive (NMC). Ce modèle s'inscrit dans la continuité critique de la Morphologie de Construction (CM), tout en remettant en question plusieurs postulats. En examinant les principes de la CM, nous avons identifié sept limites en trois volets principaux : la sélection et les propriétés sémantiques des

bases, la représentation sémantique, et l'articulation entre base et dérivé. Pour surmonter ces limites, nous avons intégré plusieurs apports issus des différentes théories linguistiques : le Modèle d'enracinement morphologique (Schmid 2016), le Lexique génératif (Pustejovsky 1996), la théorie de la métaphore (Lakoff & Johnson 1980) et de la métonymie conceptuelles (Lakoff 1987), la théorie des espaces d'intégration (Fauconnier & Turner 2002). Ce sont des théories en linguistique cognitive ou des théories qui valorisent le facteur ontologique dans les études linguistiques.

La NMC propose d'analyser la dérivation comme un processus de construction lexicale fondé sur des mécanismes cognitifs. Elle considère que le choix de la base n'est pas arbitraire : il est guidé par un profil sémantique et un profilage conceptuel, tandis que le lexique est envisagé comme un réseau structuré d'unités cognitives. Le positionnement adopté par la NMC est onomasiologique, constructionnelle et lexématique. Par ailleurs, l'architecture de la NMC s'organise autour de quatre modules principaux — le profilage conceptuel, la restructuration métonymique, l'intégration morphosémantique et le système morphologique, où la compétition morphologique constitue un mécanisme central — ainsi que des trajectoires cognitives, qui encadrent le déroulement du processus dérivationnel selon la nature de la base (nominale, adjectivale ou verbale). De façon générale, la plupart des dérivations lexicales doivent mobiliser la majorité de ces modules pour parvenir à la formation d'un nouveau lexème. Ce cadre offre une interprétation dynamique et cognitive de la dérivation en français et constitue ainsi une contribution originale à la théorie morphologique contemporaine.

2. Les noms d'instruments prototypiques et sous-classes ontologiques

Dans une perspective intégrée (ontologique et linguistique), nous avons proposé une nouvelle définition des noms d'instruments prototypiques. Concrètement, le nom d'instrument prototypique se définit comme « un artefact qu'utilise un agent (directement ou indirectement via une énergie extérieure - naturelle, électrique, chimique) pour mener une action relativement à une entité. Selon le rôle sémantique, cette entité peut être considérée comme un patient si elle est affectée (*par ex.* DOSSE > DOSSEUR), comme un thème si elle ne l'est pas (*par ex.* SCANNER > SCANNEUR), ou comme une cible si elle est déplacée. (*par ex.* CHIMIQUE > CHIMIQUIER) » De plus,

l'instrument joue un rôle sémantique « instrument » au niveau syntactico-sémantique en relation avec les verbes (que ce soit la base verbale ou le verbe dans un procès, c'est-à-dire le verbe sous-jacent). »

Nous inspirant des tests linguistiques proposés par Namer & Villoing (2008), nous avons identifié les sous-classes ontologiques prototypiques suivantes : Instrument de Captation, Machine, Appareil, Outil, Substance chimique et Récipient. Sur la base de la source d'énergie mobilisée pour leur fonctionnement, ces sous-classes ont été regroupées en deux grandes catégories fonctionnelles.

- D'une part, les instruments fonctionnant à l'énergie de l'agent, c'est-à-dire manipulés directement par un être humain :
  - Instrument de Captation : aider la perception ou à relever les informations ;
  - Outil : maniable pour effectuer une action ;
  - Récipient : creux.
- D'autre part, les instruments qui dépendent d'une énergie extérieure, telle qu'une force naturelle, électrique ou chimique :
  - Appareil : aider la perception et à relever des informations ou à conduire de l'énergie ;
  - Machine : mécanisation des tâches issues de forces humaines ou naturels ;
  - Substance chimique : action chimique.

3. Constructions des noms d'instruments

Nous avons mené des analyses sur les propriétés linguistiques préférentielles par les schémas suffixaux instrumentaux en français et leur évolution diachronique. Dans cette thèse, nous avons pris en compte quatre niveaux linguistiques : phonologie, morphologie, syntaxe ainsi que sémantique. Nous avons constaté que les schémas présentent des préférences linguistiques marquées, avec des dynamiques évolutives contrastées au fil des siècles. Plus précisément, **au niveau phonologique**, pour le variable « taille du radical de la base du dérivé », une tendance générale se dessine : les suffixes instrumentaux privilégient les bases courtes, mais certains schémas comme *-eur* et *-ant* s'appliquent aussi à des radicaux longs, allant jusqu'à trois syllabes ou plus. Concernant la variable « segment phonologique final du radical », l'analyse phonologique a démontré que les suffixes instrumentaux présentent des préférences

systématiques selon le segment final du radical, révélant l'existence de contraintes phonologiques qui influencent la construction des noms d'instruments.

Au niveau morphologique, les résultats révèlent plusieurs préférences remarquables. D'une part, la majorité des schémas suffixaux privilégient les verbes du premier groupe de conjugaison comme base. D'autre part, les bases simples ainsi que celles dérivées par préfixation ou conversion sont nettement plus fréquentes que les bases composées dans la construction des noms d'instruments. Cela reflète le lexique du français et n'est pas spécifique à la suffixation formant des noms d'instruments. Enfin, en ce qui concerne les thèmes verbaux, les schémas déverbaux se concentrent généralement sur le thème 1, mais les suffixes *-eur* et *-oire* permettent d'atteindre des champs plus spécifiques, tels que le thème verbal 13. Par ailleurs, concernant les thèmes nominaux et adjectivaux, les suffixes instrumentaux (*par ex*. *-ier* et *-ière*) manifestent une préférence pour le thème 2 des noms et des adjectifs.

Au niveau syntaxique, l'analyse montre qu'à l'exception des suffixes en *-ier* et *-ière*, la majorité des schémas suffixaux privilégient les bases verbales.

Sur le plan sémantique, deux variables ont été examinées : le profil sémantique assumé par la base du dérivé et les sous-classes ontologiques du dérivé. Les profils sémantiques telles que le verbe télique, le patient et la métonymie apparaissent comme prototypiques, révélant les stratégies cognitives les plus fréquemment mobilisées dans la dénomination des noms d'instruments. Quant aux sous-classes ontologiques, le suffixe *-eur* se distingue par sa forte productivité dans plusieurs catégories, notamment celle des appareils, tandis que le suffixe *-ant* est davantage associé à la catégorie des substances chimiques. En revanche, les autres sous-classes tendent à être partagées entre plusieurs suffixes, ce qui reflète une certaine flexibilité dans leur réalisation morphologique.

Enfin, **sur le plan diachronique**, l'étude révèle une évolution structurée de la suffixation instrumentale en français. Trois grandes phases se dégagent : une productivité limitée avant l'ère scientifique, une expansion aux XVIIe–XVIIIe siècles portée par *-oir* et *-eur*, puis une généralisation aux XIXe–XXe siècles dominée par *-eur* et *-euse*, en lien avec les évolutions industrielles. À partir de ces observations, une classification des schémas suffixaux a été établie selon trois critères principaux : le niveau de productivité morphologique (nombre de dérivés par siècle), la tendance évolutive (ascendante, stable, déclinante ou sporadique) et la répartition temporelle dominante (moderne, historique ou étendue). Sur cette base, cinq profils de suffixes ont

été dégagés : certains suffixes, comme *-eur*, connaissent une productivité croissante à l'époque contemporaine ; d'autres, comme *-oir*, montrent un déclin marqué après un pic historique ; tandis que des suffixes tels que *-aille* ou *-ail* s'éteignent.

4. Compétition morphologique

En ce qui concerne la compétition morphologique, avant tout inspiré par Hutchinson (1957), cette thèse a proposé une nouvelle définition de la notion de niche morphologique, centrée sur les interactions multifactorielles des contraintes linguistiques. Précisément, une niche morphologique se définit comme « l'ensemble particulier de modalités activées dans au moins un lexème construit par ce schéma morphologique, réparties sur les différentes variables, qui constitue la niche propre de ce schéma morphologique. ». Et la sous-niche désigne « une combinaison de modalités, composée d'une seule modalité activée pour chaque variable linguistique, et effectivement réalisée dans au moins un lexème construit par un même schéma morphologique. ». À partir de cette définition enrichie, nous avons élaboré une formule originale permettant de mesurer quantitativement le chevauchement entre niches morphologiques, dont la visualisation avec l'outil UpsetR a révélé des structures concurrentielles inédites entre les suffixes instrumentaux. Les résultats ont mis en évidence la dominance du suffixe *-eur*, capable de recouvrir entièrement les niches de plusieurs autres suffixes (*par ex. -ant*, *-ette*, *-rice*, *-aille*), tout en révélant l'asymétrie de la compétition : des suffixes motivant moins de modalités se trouvent dans une certaine mesure fragile ou marginale dans leur capacité compétitive. Ensuite, la visualisation des niches et sous-niches, à l'aide de l'outil UpsetR, a permis d'identifier la densité et la distribution des dérivés dans le lexique. Bien que certains suffixes disposent d'un nombre restreint de lexèmes dans leurs sous-niches, des suffixes comme *-eur*, *-euse*, *-oir* et *-ier* ont montré quels sous-schémas étaient les plus représentatifs dans la suffixation.

Par ailleurs, l'analyse diachronique des sous-classes ontologiques a permis d'identifier trois modes de compétition morphologique : le mode de compétition absolue (Appareil), le mode de compétition alternative (Outil, Machine) et le mode de compétition non-absolue (Récipient, Instrument de Captation, Substance Chimique). Parmi ces trois, seuls les deux derniers s'écartent du principe d'exclusion compétitive,

soulignant ainsi les limites de ce principe pour interpréter la compétition morphologique.

Enfin, en ce qui concerne la modélisation statistique, une modélisation par forêt aléatoire a permis de construire trois modèles (un modèle général et deux sous-modèles), dont les résultats convergent sur un point central : la variable la plus discriminante dans le choix des suffixes instrumentaux est la sous-classe ontologique du dérivé. Selon le cadre de la Néo-Morphologie Cognitive, cette propriété sémantique constitue une contrainte décisive dans la phase de compétition morphologique. En particulier, certains schémas suffixaux tendent à privilégier des sous-classes ontologiques spécifiques : le suffixe *-ant* privilégie les substances chimiques, *-eur* préfère les appareils et les machines, *-oir* tend à se spécialiser dans les outils, tandis que *-euse* montre une préférence marquée pour les machines. Ces préférences permettent de distinguer les suffixes en compétition. Cela souligne que, dans la suffixation des noms d'instruments en français, grâce à la conceptualisation, les connaissances ontologiques issues de l'interaction entre l'humain et le monde réel sont intégrées au système linguistique, rendant ainsi cette dimension cruciale dans le choix du suffixe. À l'inverse, les propriétés linguistiques traditionnelles s'avèrent insuffisantes pour prédire ou interpréter cette compétition morphologique.

D'autres variables linguistiques jouent également un rôle important. Dans le modèle général, la catégorie syntaxique de la base ainsi que deux variables phonologiques (la taille syllabique et le segment final du radical) se révèlent discriminantes. Par exemple, *-eur* et *-ant* préfèrent des radicaux polysyllabiques, tandis que *-oir* privilégie des formes monosyllabiques. Le suffixe *-eur* montre une préférence marquée pour les radicaux terminés par une occlusive, et rejette les nasales et approximantes. Dans le sous-modèle 1, la variable espace thématique permet également de distinguer les schémas, notamment grâce au thème 13, qui oppose certains dérivés en *-eur*/*-rice* à ceux en *-euse*/*-oir*.

Perspectives socio-cognitives et humanistes

À la lumière des conclusions présentées, il convient d'envisager les prolongements possibles en termes de perspectives socio-cognitives et humanistes. Dans cette thèse, les résultats obtenus confirment que la dérivation instrumentale n'est jamais un

processus réduit à une opération linguistique purement formelle : elle est profondément liée à la cognition humaine et aux dynamiques socioculturelles. Cette observation soulève des enjeux épistémologiques majeurs pour la linguistique contemporaine. A notre sens, dans un contexte où les grands modèles de langage tendent à occulter la dimension humaine du langage au profit d'approches purement statistiques, la Néo-Morphologie Cognitive, ancrée dans le méta-paradigme de la linguistique cognitive, apporte une réponse théorique à ce défi. Elle rappelle que l'étude de la morphologie, et plus largement de la langue, doit rester fondée sur l'analyse de la cognition propre à l'humain et ne peut se réduire à une simple extraction de régularités depuis les corpus. Ainsi, la NMC offre une voie pour renouveler la réflexion linguistique à l'ère de l'intelligence artificielle. Elle réaffirme le rôle essentiel de la cognition et de la conceptualisation, ouvrant des perspectives nouvelles pour une linguistique qui n'abandonne pas son ancrage dans l'esprit humain et pour les recherches futures en linguistique.

Trois axes de recherche se dessinent.

I. Sur le plan psycholinguistique : tester expérimentalement le coût métonymique ou l'inhibition suffixale prédits par la NMC.
II. Sur le plan morphologique macroscopique : examiner la concurrence entre différents procédés morphologiques dans la formation des noms d'instruments, en incluant la conversion V>N et la composition VN (cf. Tribout & Villoing 2014), les profils sémantiques associés à chacun, ainsi que les régularités transversales.
III. Une approche lexico-culturelle pourrait retracer, du XIIe siècle à nos jours, les trajectoires culturelles de la dérivation des noms d'instruments dans l'histoire du français.

Ces pistes renforcent une vision du langage comme phénomène cognitif, culturel et dynamique.

Annexes

1. Guide d'annotation

Ce guide d'annotation a été élaboré dans le cadre d'une recherche portant sur la suffixation des noms d'instrument en français. Cette recherche est basée sur l'annotation détaillée d'un large échantillon de noms déverbaux et dénominaux tirées du corpus lexicographique Le *Grand Robert de la langue française* (version 2005). Huit groupes de propriétés linguistiques qui sont annotées :

1. Taille du radical du lexème de base en nombre de syllabe ;
2. Segment final phonologique du radical de la base du dérivé ;
3. Structure morphologique du lexème de base ;
4. Thème de la base du dérivé sélectionné par la suffixation ;
5. Classe flexionnelle de la base verbale du dérivé ;
6. Catégorie syntaxique de la base du dérivé ;
7. Profil sémantique assumé par la base du dérivé ;
8. Sous-classes ontologiques du dérivé ;

1.1 Niveau phonologique

1.1.1 Taille phonologique du radical de la base du dérivé

Le nombre des syllabes est annoté automatiquement sur un tableur Excel format xlsm. Pour compter automatiquement la taille du radical, nous avons inventé cet outil. (voir Annexe 3) Concernant le détail de cet outil, nous le présenterons dans « Outil de comptage automatique de la taille phonologique du radical » de l'Annexe. Concrètement, dans ce tableur, nous avons créé deux fichiers. L'un est destiné à entrer les formes phonologiques à compter. L'autre est destiné à enregistrer les macros, qui comprennent toutes les voyelles du français. Lorsque nous avons mis toutes les formes phonologiques à compter, nous pouvons exécuter la fonction de macro pour compter automatiquement la taille phonologique. Après avoir exécuté la fonction de macro, la taille phonologique est apparue à droite de la forme phonologique.

Critère : Nombre de syllabes dans le radical du lexème de base

Etiquette : /PHON1_nb_syl/

Exemple :

RASOIR – <ras> / ʀaz / ➞ 1

RAPPORTEUR – <rapport> / ʀapɔʀt / ➞ 2

AGRANDISSEUR – <agrandiss> / agʀɑ̃dis / ➞ 3

ACCELERATEUR– <accélérat> / akseleʀat / ➞ 4

DESHUMIDIFICATEUR – < déshumidificat > / dezymidifikat / ➞ 6

1.1.2 Segment final phonologique du radical de la base du dérivé

Concernant les segments finaux phonologiques du radical de la base du dérivé, nous rendons compte de toutes les variables de phonèmes finaux, car nous avons émis l'hypothèse que certains segments finaux ne sont pas conditionnés purement par la phonologie, mais par la morphologie. En outre, pour annoter les lexèmes en jeu en la matière, nous avons utilisé les fonctions de formules fournies par le tableur Excel en vue de récupérer le segment final phonologique de radical. Concrètement, pour l'automatisation de la récupération du segment final phonologique de radical, nous aurions recours à la fonction LEFT et à celle RIGHT. Plus précisément, concernant la fonction LEFT, elle fait référence à l'extraction d'un nombre spécifié de caractères à partir du premier caractère sur le côté gauche du texte. La fonction RIGHT, quant à elle, est l'opposé de la fonction LEFT, qui renvoie un nombre spécifié de caractères à partir du côté droit du texte spécifié. En combinant les deux fonctions, il est plus efficace de récupérer le segment final phonologique du radical de la base du dérivé.

Critère : Segment final phonologique de radical

Etiquette : /PHON2_seg/

Exemple :

RAPPORTEUR – / ʀapɔʀtœʀ / ➞ t

DRAINEUSE – / dʀɛnøz / ➞ n

THEIERE – / tejɛʀ / ➞ e

1.2 Niveau morphologique

1.2.1 Structure morphologie de la base du dérivé

Critère : structure morphologie de la base dans le français contemporain
Etiquette : /MORPH1_struc/
Exemple :

1. a. AGRANDIR > AGRANDISSEUR ➞ / MORPH1_struc_préfixation_*a*- A>V/
   b. DRAINER > DRAINEUR ➞ / MORPH1_struc_conversion_N>V/
   c. CUEILLIR > CUEILLEUR ➞ / MORPH1_struc_simple_V/

1.2.1.1 Préfixation

1.2.1.1.1 Préfixe *a*-

Vu qu'il y a deux procédés morphologiques (à savoir le procédé dénominal et le désadjectival), nous les considérons comme homonymes, nous les traitons séparément comme deux schémas morphologiques.

- **Préfixe *a*-A>V**

Critère : base nominale préfixée morphologiquement par le préfixe *a*- pour dériver un lexème verbal
Sémantique et fonction : Rendre (plus) A (Namer 2009 : 208)
Etiquette : /MORPH1_struc_préfixation_a-A>V /
Exemple :

2. a. $[(a)_{\text{préfixe}}\,[\text{DOUX}]_{A}\,]_{V}$ > ADOUCIR $_{V}$ > ADOUCISSEUR $_{N}$
   b. $[(a)_{\text{préfixe}}\,[\text{GRAND}]_{A}\,]_{V}$ > AGRANDIR $_{V}$ > AGRANDISSEUR $_{N}$
   c. $[(a)_{\text{préfixe}}\,[\text{ROND}]_{A}\,]_{V}$ > ARRONDIR $_{V}$ > ARRONDISSEUR $_{N}$

- **Préfixe *a*-N>V**

Critère : base nominale préfixée morphologiquement par le préfixe *a-* pour dériver un lexème verbal

Sémantique et fonction : la direction, le but à atteindre, ou le passage d'un état à un autre (TLFi ; Villoing & Gueye 2024 : 9)

Etiquette : /MORPH1_struc_préfixation_*a*-N>V /

Exemple :

3. a. $[(a)_{\text{préfixe}}\,[\text{LIGNE}]_{N}\,]_{V}$ > ALIGNER $_{V}$ > ALIGNOIR $_{N}$
   b. $[(a)_{\text{préfixe}}\,[\text{FUT}]_{N}\,]_{V}$ > AFFUTER $_{V}$ > AFFUTEUSE $_{N}$
   c. $[(a)_{\text{préfixe}}\,[\text{MASSE}]_{N}\,]_{V}$ > AMASSER $_{V}$ > AMASSETTE $_{N}$

1.2.1.1.2 Préfixe *contre-*

Critère : base préfixée morphologiquement par le préfixe *contre-*.

Sémantique et fonction : marque l'inversion ou l'opposition. (Amiot et Montermini 2009 : 136)

Etiquette : /MORPH1_struc_préfixation_contre-V>V/

Exemple :

4. $[(contre)_{\text{préfixe}}\,[\text{COLLER}]_{V}\,]_{V}$ > CONTRECOLLER $_{V}$ > CONTRECOLLEUSE $_{N}$

1.2.1.1.3 Préfixe *dé-*

Puisqu'il y a deux procédés morphologiques (à savoir le procédé dénominal et le déverbal), nous les considérons comme homonymes, nous les traitons séparément comme deux schémas morphologiques.

- **Préfixe *dé*-N >V**

Critère : base nominale préfixée morphologiquement par le préfixe *dé-* pour dériver un lexème verbal

Sémantique et fonction : (Enlever-(Faire) sortir de) N (Namer 2009 : 209)

Etiquette : /MORPH1_struc_préfixation_dé-N>V /

Exemple :

5. a. $[(dé)_{\text{préfixe}}\,[\text{BROUSSE}]_{N}\,]_{V}$ > DEBROUSSER $_{V}$ > DEBROUSSEUR $_{N}$

b. [(*dé*)$_{\text{préfixe}}$ [CAPSULE]$_{N}$]$_{V}$ > DECAPSULER $_{V}$ > DECAPSULEUR $_{N}$

c. [(*dé*) $_{\text{préfixe}}$ [LAIT]$_{N}$]$_{V}$ > DELAITER $_{V}$ > DELAITEUSE $_{N}$

- **Préfixe *dé*-V1 > V2**

Critère : base verbale préfixée morphologiquement par le préfixe *dé-* pour dériver un lexème verbal

Sémantique et fonction : 1) une entité cible a se retrouve à l'extérieur de l'espace (concret ou abstrait) défini par une entité site b de telle sorte que cette extériorité affecte l'état de l'entité sur laquelle porte le verbe. (cf. 5c.) ; 2) Annuler le résultat V1 (cf. 5a-5b.) (Namer 2009 : 209 ; Jalenques 2014 ; Jalenques 2016)

Etiquette : /MORPH1_struc_préfixation_dé-V1 > V2 /

Exemple :

6. a. [(*dé*)$_{\text{préfixe}}$ [CHARGER]$_{V}$]$_{V}$ > DECHARGER $_{V}$ > DECHARGEUR $_{N}$

   b. [(*dé*)$_{\text{préfixe}}$ [ROULER]$_{V}$]$_{V}$ > DEROULER $_{V}$ > DEROULEUR $_{N}$

   c. [(*dé*) $_{\text{préfixe}}$ [COUPER]$_{V}$]$_{V}$ > DECOUPER $_{V}$ > DECOUPOIR $_{N}$

1.2.1.1.4 Préfixe *é-*

Il existe trois schémas morphologiques différents.

- **Préfixe *é*-N > V**

Critère : base nominale préfixée morphologiquement par le préfixe *é-* pour dériver un lexème verbal

Sémantique et fonction : Priver de N (Namer 2009 : 209)

Etiquette : /MORPH1_struc_préfixation_é-N > V /

Exemple :

7. a. [(*é*)$_{\text{préfixe}}$ [GOUTTE]$_{N}$]$_{V}$ > EGOUTTER $_{V}$ > EGOUTTOIR $_{N}$

   b. [(*é*)$_{\text{préfixe}}$ [GRAPPE]$_{N}$]$_{V}$ > EGRAPPER $_{V}$ > EGRAPPOIR $_{N}$

   c. [(*é*) $_{\text{préfixe}}$ [MIETTE]$_{N}$]$_{V}$ > EMIETTER $_{V}$ > EMETTEUR $_{N}$

- **Préfixe *é*-A > V**

Critère : base adjective préfixée morphologiquement par le préfixe *é-* pour dériver un lexème verbal

Sémantique et fonction : Rendre (plus) A (Namer 2009 : 208)

Etiquette : /MORPH1_struc_préfixation_é-A > V /

Exemple :

8. a. [(*é*)$_{\text{préfixe}}$ [LARGE]$_{A}$]$_{V}$ > ELARGIR $_{V}$ > ELARGISSEUR $_{N}$

   b. [(*é*)$_{\text{préfixe}}$ [MINCE]$_{A}$]$_{V}$ > EMINCER $_{V}$ > EMINCEUR $_{N}$

   c. [(*é*)$_{\text{préfixe}}$ [PUR]$_{A}$]$_{V}$ > EPURER $_{V}$ > EPURATEUR $_{N}$

1.2.1.1.5 Préfixe *en-*

Concernant le préfixe *en-*, il existe trois schémas morphologiques différents comme pour le préfixe *dé-*.

- **Préfixe *en*-N > V**

Critère : base nominale préfixée morphologiquement par le préfixe *é-* pour dériver un lexème verbal

Sémantique et fonction : (Munir de – mettre dans le) N (Namer 2009 : 210)

Etiquette : /MORPH1_struc_préfixation_ en-N > V/

Exemple :

9. a. [(*en*)$_{\text{préfixe}}$ [PATE]$_{N}$]$_{V}$ > EMPATER $_{V}$ > EMPATEUR $_{N}$

   b. [(*en*)$_{\text{préfixe}}$ [GRAIN]$_{N}$]$_{V}$ > ENGRENER $_{V}$ > ENGRENEUR $_{N}$

   c. [(*en*) $_{\text{préfixe}}$ [BOUTEILLE]$_{N}$]$_{V}$ > EMBOUTEILLER $_{V}$ > EMBOUTEILLEUSE $_{N}$

- **Préfixe *en*-A > V**

Critère : base adjective préfixée morphologiquement par le préfixe *en-* pour dériver un lexème verbal

Sémantique et fonction : (Rendre – devenir) (plus) A (Namer 2009 : 210)

Etiquette : /MORPH1_struc_préfixation_en-A > V /

Exemple :

10. [(*en*)$_{\text{préfixe}}$ [RICHE]$_{A}$]$_{V}$ > ENRICHIR $_{V}$ > ENRICHISSEUR $_{N}$

- **Préfixe *en*-V > V**

Critère : base verbale préfixée morphologiquement par le préfixe *en-* pour dériver un lexème verbal

Sémantique et fonction : à valeur factitive (Huot 2005 : 119)

Etiquette : /MORPH1_struc_préfixation_*en* V<V /

Exemple :

11. a. [(*en*)$_{\text{préfixe}}$ [BROUILLER]$_{V}$]$_{V}$ > EMBROUILLER $_{V}$ > EMBROUILLEUR $_{N}$

    b. [(*en*)$_{\text{préfixe}}$ [ROULER]$_{V}$]$_{V}$ > ENROULER $_{V}$ > ENROULEUR $_{N}$

1.2.1.1.6 Préfixe *inter*-V>V

Critère : base verbale préfixée morphologiquement par le préfixe *inter-*

Sémantique et fonction : Pour indiquer que l'action est réciproque. (Huot 2006 : 116 ; Grand Robert)

Etiquette : /MORPH1_struc_préfixation_inter /

Exemple :

12. [(*inter*)$_{\text{préfixe}}$ [CLASSER]$_{V}$]$_{V}$ > INTERCLASSER $_{V}$ > INTERCLASSEUSE $_{N}$

1.2.1.1.7 Préfixe *re*-V>V

Critère : base verbale préfixée morphologiquement par le préfixe *re-*

Etiquette : /MORPH1_struc_préfixation_*re* /

Sémantique et fonction : 1) à valeur réitérative, retour à un ancien état (cf. 12 b-12c) ;) simple renforcement (cf. 12a) (Huot 2001 : 97 ; Grevisse 2016 : 195)

Exemple :

13. a. [(*re*)$_{\text{préfixe}}$ [AFFINER]$_{V}$]$_{V}$ > RAFFINER $_{V}$ > RAFFINEUR $_{N}$

    b. [(*re*)$_{\text{préfixe}}$ [FRAICHIR]$_{V}$]$_{V}$ > RAFRAICHIR $_{V}$ > RAFRAICHISSEUR $_{N}$

    c. [(*re*)$_{\text{préfixe}}$ [BOUILLIR]$_{V}$]$_{V}$ > REBOUILLIR $_{V}$ > REBOUILLEUR $_{N}$

1.2.1.1.8 Préfixe *sous*-V>V

Critère : base verbale préfixée morphologiquement par le préfixe *sous-*

Sémantique et fonction : suffixe au sens locatif ou temporel « en bas » ou aux valeurs évaluatives marquant l'infériorité quantitative objective (Amiot 2004 : 7 ; Amiot et Montermini 2009 : 136)

Etiquette : /MORPH1_struc_préfixation_*sous* /

Exemple :

14. [(*sous*)$_{\text{préfixe}}$ [TIRER]$_{V}$]$_{V}$ > SOUTIRER $_{V}$ > SOUTIREUSE $_{N}$

1.2.1.1.9 Préfixe *sur*-V>V

Critère : base verbale préfixée morphologiquement par le préfixe *sur*-

Sémantique et fonction : suffixe au sens locatif ou temporel « en haut » ou aux valeurs évaluatives marquant l'excès ou la répétition d'un premier procès (Amiot 2004 : 7 ; Amiot et Montermini 2009 : 136)

Etiquette : /MORPH1_struc_préfixation_*sur* /

Exemple :

15. a. [(*sur*)$_{\text{préfixe}}$ [LIGNER]$_{V}$]$_{V}$ > SURLIGNER $_{V}$ > SURLIGNEUR $_{N}$

    b. [(*sur*)$_{\text{préfixe}}$ [GELER]$_{V}$]$_{V}$ > SURGELER $_{V}$ > SURGELATEUR $_{N}$

1.2.1.1.10 Préfixe *trans*-V

Critère : base verbale préfixée morphologiquement par le préfixe *trans*-

Sémantique et fonction : à travers (Huot 2001 : 95)

Etiquette : /MORPH1_struc_préfixation_trans /

Exemple :

16. a. [(*trans*)$_{\text{préfixe}}$ [CODER]$_{V}$]$_{V}$ > TRANSCODER $_{V}$ > TRANSCODEUR $_{N}$

    b. [(*trans*)$_{\text{préfixe}}$ [FORMER]$_{V}$]$_{V}$ > TRANSFORMER $_{V}$ > TRANSFORMATEUR $_{N}$

1.2.1.1.11 Préfixe *in*-N

Critère : base nominale préfixée morphologiquement par le préfixe *in*-

Sémantique et fonction : à valeur négative (Huot 2001 : 97)

Etiquette : /MORPH1_struc_préfixation_in /

Exemple :

17. [(*in*)$_{\text{préfixe}}$ [CAPACITE]$_{N}$]$_{N}$ > INCAPACITE $_{V}$ > INCAPACITANT $_{N}$

1.2.1.2 Suffixation

1.2.1.2.1 Suffixe N-*ique* > A

Critère : la base nominale du dérivé est suffixée morphologiquement par le suffixe *-ique*

Sémantique et fonction : relatif aux noms de disciplines ou de la terminologie scientifique et technique (Grevisse 2016 : 182 ; Strnadová 2017)

Etiquette : /MORPH1_struc_suffixation_N-*ique* > A/

Exemple :

18. [[CHIMIE]$_{N}$ (*ique*)$_{suffixe}$]$_{A}$ > CHIMIQUE $_{A}$ > CHIMIQUIER $_{N}$

1.2.1.2.2 Suffixe N-*et* > N

Critère : la base nominale du dérivé est suffixée morphologiquement par le suffixe -*et*

Sémantique et fonction : valeur diminutive, parfois dépréciative et tirant parti des informations portées par le nom dérivé (Dal 1997 ; Fradin 1999 ; Fradin 2003 ; Huot 2001 : 79 ; Corbin -et)

Étiquette : /MORPH1_struc_suffixation_*et* /

Exemple :

19. [[ŒIL]$_{N}$ (*et*)$_{suffixe}$]$_{N}$ > ŒILLET $_{N}$ > ŒILLETEUSE $_{N}$

1.2.1.2.3 Suffixe -*ifier*

- **Suffixe A-*ifier* > V**

Critère : la base adjectivale du dérivé est suffixée morphologiquement par le suffixe verbal -*ifier*

Sémantique et fonction : Rendre/devenir (plus) A (Namer 2009 : 208)

Etiquette : /MORPH1_struc_suffixation_A-*ifier* > V /

Exemple :

20. a. [[AMPLE]$_{A}$ (*ifier*)$_{suffixe}$]$_{V}$ > AMPLIFIER $_{V}$ > AMPLIFICATEUR $_{N}$

    b. [[HUMIDE]$_{A}$ (*ifier*)$_{suffixe}$]$_{V}$ > HUMIDIFIER $_{V}$ > HUMIDIFICATEUR $_{N}$

- **Suffixe N-*ifier* > V**

Critère : la base nominale du dérivé est suffixée morphologiquement par le suffixe verbal -*ifier*

Sémantique et fonction : (Transformer en – Munir de) N (Namer 2009 : 210)

Etiquette : /MORPH1_struc_suffixation_N-*ifier* > V /

Exemple :

21. a. [[GAZ]$_{N}$ (*ifier*)$_{suffixe}$]$_{V}$ > GAZEIFIER $_{V}$ > GAZEIFICATEUR $_{N}$

    b. [[LIQUIDE]$_{N}$ (*ifier*)$_{suffixe}$]$_{V}$ > LIQUEFIER $_{V}$ > LIQUEFACTEUR $_{N}$

1.2.1.2.4 Suffixe *-iser*

- **Suffixe N-*iser* > V**

Critère : la base nominale du dérivé est suffixée morphologiquement par le suffixe verbal *-iser*

Sémantique et fonction : Transformer en (cf. 22b) / Munir de (cf. 22a) / Mettre dans N (Namer 2009)

Etiquette : /MORPH1_struc_suffixation_N-*iser* > V /

Exemple :

22. a. [[AROMATE]$_N$ (*iser*)$_{suffixe}$]$_V$ > AROMATISER $_V$ > AROMATISANT $_N$

    b. [[CRISTAL]$_N$ (*iser*)$_{suffixe}$]$_V$ > CRISTALLISER $_V$ > CRISTALLISATEUR $_N$

- **Suffixe A-*iser* > V**

Critère : la base adjectivale du dérivé est suffixée morphologiquement par le suffixe verbal *-iser*

Sémantique et fonction : Rendre (plus) A (Namer 2009)

Etiquette : /MORPH1_struc_suffixation_A-*iser* > V /

Exemple :

23. a. [[HOMOGENE]$_A$ (*iser*)$_{suffixe}$]$_V$ > HOMOGENEISER $_V$ > HOMOGENEISATEUR $_N$

    b. [[PARALLELE]$_A$ (*iser*)$_{suffixe}$]$_V$ > PARALLELISER $_V$ > PARALLELISEUR $_N$

1.2.1.2.5 Suffixe V-*iller* > V

Critère : la base verbale du dérivé est suffixée morphologiquement par le suffixe *-iller*

Sémantique et fonction : suffixe évaluatif aux valeurs diminutive, dépréciative et itérative (Amiot et Stosic 2011)

Etiquette : /MORPH1_struc_suffixation_V-*iller* > V /

Exemple :

24. a. [[POINT]$_N$ (*iller*)$_{suffixe}$]$_V$ > POINTILLER $_V$ > POINTILLEUR $_N$

    b. [[TOUPIE]$_N$ (*iller*)$_{suffixe}$]$_V$ > TOUPILLER $_V$ > TOUPILLEUSE $_N$

1.2.1.2.6 Suffixe *-oyer*

- **Suffixe A-*oyer* > V**

Critère : la base adjectivale du dérivé est suffixée morphologiquement par le suffixe *-oyer*

Sémantique et fonction : une valeur factitive (TLF)

Etiquette : /MORPH1_struc_suffixation_A-*oyer* > V /

Exemple :

25. $[[\text{NET}]_{A}\,(\textit{oyer})_{\text{suffixe}}]_{V}$ > NETTOYER $_{V}$ > NETTOYEUR $_{N}$

- **Suffixe N-*oyer* > V**

Critère : la base nominale du dérivé est suffixée morphologiquement par le suffixe *-oyer*

Sémantique et fonction : une valeur factitive (TLFi)

Etiquette : /MORPH1_struc_suffixation_N-*oyer* > V /

Exemple :

26. $[[\text{JOINT}]_{N}\,(\textit{oyer})_{\text{suffixe}}]_{V}$ > JOINTOYER $_{V}$ > JOINTOYEUR $_{N}$

1.2.1.3 Conversion

Quant à la conversion, il existe un cas de conversion Nom / Verbe et celui Adjectif / Verbe (Tribout 2010).

- **Conversion Nom / Verbe**

Critère : la base du dérivé est convertie morphologiquement Nom / Verbe

Sémantique et fonction : (Faire-Dire-Utiliser) un/des N (cf. 27b) – Se composer comme N- Donner l'apparence de N-Mettre dans N

Etiquette : /MORPH1_struc_conversion_V/ N /

Exemple :

27. a. $[[\text{ACTION}]_{N\ \text{conversion}}]_{V}$ > ACTIONNER $_{V}$ > ACTIONNEUR $_{N}$

b. $[[\text{AMORCE}]_{N\ \text{conversion}}]_{V}$ > AMORCER $_{V}$ > AMORCEUR $_{N}$

c. $[[\text{GRENADE}]_{N\ \text{conversion}}]_{V}$ > GRENADER $_{V}$ > GRENADEUR $_{N}$

- **Conversion** Adjectif / Verbe

Critère : la base du dérivé est convertie morphologiquement Adjectif / Verbe

Sémantique et fonction : (Rendre-devenir) (plus) A

Etiquette : /MORPH1_struc_conversion_V/A /

Exemple :

28. [[SEC]$_{\text{A CONVERSION}}$]$_{\text{V}}$ > SECHER $_{\text{V}}$ > SECHEUR $_{\text{N}}$

1.2.1.4 Composition néoclassique

Certains lexèmes à étudier dans cette recherche sont construits par des schémas de composition néoclassique. Ainsi, nous allons annoter ces éléments comme pour les parties précédentes.

1.2.1.4.1 *photo-* NV>V, N = photo

Critère : la base du dérivé est construite morphologiquement par l'élément *photo-*

Sémantique et fonction : lumière. (Grevisse 2016 : 211)

Etiquette : /MORPH1_struc_composition_photo /

Exemple :

29. [COMPOSER]$_{\text{V}}$ > PHOTOCOMPOSER $_{\text{V}}$ > PHOTOCOMPOSEUSE $_{\text{N}}$

1.2.1.4.2 *poly-* AV>V

Critère : la base du dérivé est construit morphologiquement par l'élément *poly-*

Sémantique et fonction : plusieurs. (Apothéloz 2002 : 79 ; Grevisse 2016 : 212 ; Dumarest & Morsel 2017 : 37)

Etiquette : /MORPH1_struc_composition_ poly- AV>V /

Exemple :

30. [COPIER]$_{\text{V}}$ > POLYCOPIER $_{\text{V}}$ > POLYCOPIEUSE $_{\text{N}}$

1.2.1.4.3 *télé-* composition ADV V>V

Critère : la base du dérivé est construit morphologiquement par l'élément *télé-*

Sémantique et fonction : au loin. (Corbin 1987 : 597 ; Grevisse & Goosse 2016 : 212)

Etiquette : /MORPH1_struc_composition_télé- composition ADVV>V /

Exemple :

31. [COPIER]$_{\text{V}}$ > TELECOPIER $_{\text{V}}$ > TELECOPIEUR $_{\text{N}}$

1.2.1.4.4 *thermo-* N+N > N

Critère : la base du dérivé est construit morphologiquement par l'élément *thermo-*

Sémantique et fonction : chaleur (Grevisse & Goose 2016 : 211)

Etiquette : /MORPH1_struc_composition_ thermo- N+N> N /

Exemple :

32. [COPIE]$_N$ > THERMOCOPIE $_N$ > THERMOCOPIEUSE $_N$

1.2.1.5 Lexème simple

Il reste encore à étudier les lexèmes dont la base du dérivé est un lexème simple. Nous les classons selon leur catégorie syntaxique, à savoir nom ou verbe. Un fait non négligeable : certains lexèmes pourraient être analysés ou segmentés comme complexes en diachronie, mais cette recherche se cantonne à la perspective synchronique. Ainsi, si les lexèmes ne sont pas susceptibles d'être construits ou analysés par les locuteurs natifs de la langue française contemporaines, nous les considérons comme des lexèmes simples plutôt que complexes. Nous prenons deux lexèmes NETTOYER et FARDEAU pour bien illustrer notre propos. Le lexème NETTOYER est suffixé morphologiquement en 1175 par le suffixe *-oyer* à partir de la base NET. Les locuteurs natifs le considèrent comme un simplex au lieu de complexe. Cependant, les lexèmes en *-oyer* construits anciennement seraient réanalysés par les locuteurs natifs, surtout au niveau sémantique pour la relation étroite entre la base et le dérivé. Ainsi, il est probable que le lexème soit analysé comme un lexème complexe. En revanche, ce n'est pas le même cas pour le lexème FARDEAU, qui est construit par FARDE en 1205. Vu que le lexème FARDEAU est dérivé par le sens ancien de FARDE, pour les locuteurs natifs du français moderne, il est difficile de réanalyser tel lexème comme le lexème complexe. Ils préfèrent le traiter comme un nom simple.

1.2.1.5.1 Nom simple

Critère : la base du dérivé est un nom simple

Etiquette : /MORPH1_struc_simple_ns /

Exemple :

33. a. ALOSE $_N$ > ALOSIER $_N$

    b. GELATINE $_N$ > GELATINEUR $_N$

c. CREPE $_N$ > CREPIER $_N$

1.2.1.5.2 Verbe simple

Critère : la base du dérivé est un verbe simple

Etiquette : /MORPH1_struc_simple_vs /

Exemple :

34. ALLUMER $_V$ > ALLUMOIR $_N$

1.2.2 Espace thématique

Concernant l'espace thématique, qui se concentre sur l'allomorphie, il y a plusieurs thèmes utilisés pour construire les noms d'instrument. Plus précisément, les verbes comptent 13 thèmes (Bonami, Boyé & Kerleroux 2009 : 121) ou 14 thèmes, en incluant le thème zéro proposé par Tribout (2010 : 200) (Bonami 2014 : 50). En revanche, pour les thèmes nominaux et adjectivaux, il y a deux thèmes destinés à dériver les mots (Tribout 2010 : 217 ; Koehl 2012 : 178).

1.2.2.1 Thèmes verbaux

- **Thème verbal 1**

Critère : Thème verbal 1 (correspondant à la forme qui apparaît à la première personne du pluriel au présent) présent dans le radical de la base verbale du dérivé

Etiquette : /MORPH2_ET_V1 /

Exemple :

35. a. BOUILLOIRE $_N$: *bouill-oire* $_N$ → V1

b. RASOIR $_N$: *ras-oir* $_N$ → V1

c. EPLUCHEUR $_N$: *épluch-oir* $_N$ → V1

- **Thème verbal 8**

Critère : Thème verbal 8 (correspondant à la forme qui apparaît à la première personne du pluriel au subjonctif imparfait) présent dans le radical de la base verbale du dérivé. Sur la base de Bonami (2014), nous incluons également dans le thème verbal 8 les radicaux en *-iss-* qui n'apparaissent qu'à l'imparfait du subjonctif, forme précisément attestée dans le radical du verbe BOUILLIR.

Etiquette : /MORPH2_ET_V8 /

Exemple :

36. BOUILLISSEUR $_{N}$: *bouilliss -eur* $_{N}$ → V8

- **Thème verbal 12**

Critère : Thème verbal 12 présent (correspondant à la forme qui apparaît au participe passé) dans le radical de la base verbale du dérivé

Etiquette : /MORPH2_ET_V12 /

Exemple :

37. FRITEUSE $_{N}$: *frit -euse* $_{N}$ → V12

1.2.2.1.4 Thème verbal 13

Critère : Thème latin (ou radical latin) sert à la base du dérivé (Bonami, Boyé & Kerleroux 2009)

Etiquette : /MORPH2_ET_V13 /

Exemple :

38. a. EVAPORATEUR $_{N}$: *évaporat -eur* $_{N}$ → V13
    b. INTEGRATEUR $_{N}$: *intégrat -eur* $_{N}$ → V13
    c. INTERRUPTEUR $_{N}$: *interrupt -eur* $_{N}$ → V13

1.2.2.2 Thèmes nominaux

- **Thème nominal 2**

Critère : Thème 2 (correspondant à la forme féminin des adjectifs ou aux formes lexicalisées, sans en porter la propriété ni sémantique ni grammaticale) sert à la base du dérivé (Tribout 2010 : 217 ; Koehl 2012 : 178)

Etiquette : /MORPH2_ET_N2 /

Exemple :

39. a. DOSSEUR $_{N}$: *dosse -oire* $_{N}$ → N2
    b. BETONNEUSE $_{N}$: *béton -euse* $_{N}$ → N2
    c. PINARDIER $_{N}$: *pinard -ier* $_{N}$ → N2

1.2.2.2.2 Thème nominal L

Critère : Thème L (radical latin ou radical savant) sert à la base du dérivé (Tribout 2010 : 217 ; Koehl 2012 : 178)

Etiquette : /MORPH2_ET_NL /

Exemple :

40. a. SALIERE $_{N}$: *sal* « SEL » *–ière* $_{N}$ → NL

b. NASIERE $_{N}$: *nas* « NEZ » *–ière* $_{N}$ → NL

1.2.3 Classe flexionnelle de la base verbale

Concernant la classe flexionnelle de la base verbale, en français il y a trois classes flexionnelles du verbe : le premier groupe, le deuxième groupe ainsi que le troisième groupe (Eluerd 2009 : 18-19)

1.2.3.1 Le premier groupe

Critère : verbe en *-er* (sauf les verbes, tels que ALLER) (Eluerd 2009 : 18)

Etiquette : / MORPH3_CF_G1 /

Exemple :

41. a. PASSOIRE $_{N}$ < PASSER → G1

b. RASOIR $_{N}$ < RASER → G1

c. EPLUCHEUR < EPLUCHER → G1

1.2.3.2 Le deuxième groupe

Critère : Verbe en *-ir*/*-issant* (Eluerd 2009 : 19)

Etiquette : / MORPH3_CF_G2 /

Exemple :

42. a. AGRANDISSEUR $_{N}$ < AGRANDIR → G2

b. ADOUCISSEUR $_{N}$ < ADOUCIR → G2

c. ASSOUPLISSANT $_{N}$ < ASSOUPLIR → G2

1.2.3.3 Le troisième groupe

Critère : Verbe en *-ir/-oir/-re*, etc…. (Eluerd 2009 : 19)

Etiquette : / MORPH3_CF_G3 /

Exemple :

43. a. FENDOIR $_{N}$ < FENDRE $_{V}$ → G3

b. BOUILLOIRE $_{N}$ < BOUILLIR $_{V}$ → G3

c. EXTRACTEUR < EXTRAIRE $_{V}$ → G3

1.3 Niveau Syntaxique

1.3.1 Verbe

Etiquette : /SYN_V/

Exemple :

44. a. BOUILLOIRE $_{N}$ < BOUILLIR → V
    b. RASOIR $_{N}$ < RASER → V
    c. EPLUCHEUR < EPLUCHER → V

1.3.2 Nom

Etiquette : /SYN_N/

Exemple :

45. a. FOURCHETTE $_{N}$ < FOURCHE → N
    b. ALOSIER $_{N}$ < ALOSE → N
    c. ABLERET < ABLE → N

1.3.3 Adjectif

Etiquette : /SYN_A/

Exemple :

46. a. CHIMIQUIER $_{N}$ < CHIMIQUE → A
    b. BLANCHET $_{N}$ < BLANC → A
    c. LONGUET $_{N}$ < LONG → A

1.4 Niveau Sémantique

1.4.1 Profil sémantique assumé par la base dans la représentation du dérivé

Pour le profil sémantique assumé par la base dans la représentation du dérivé, nous avons annoté deux catégories. Pour la première catégorie, il s'agit de la base verbale du dérivé pour les noms déverbaux, laquelle correspond aux quatre rôles de la Structure de Qualia (Pustejovsky 1996). Concernant la deuxième catégorie, nous avons

annoté les rôles sémantiques joués par les bases nominales (Fillmore 1968) dans la représentation du verbe sous-jacent, et les projections cognitives (métaphore conceptuelle et métonymie conceptuelle).

1.4.1.1 Les verbes correspondant aux quatre rôles de la Structure de Qualia

Pour les noms d'instruments, nous avons recensé deux types de verbes aux deux rôles de la Structure de Qualia, à savoir le verbe télique et le verbe formel.

1.4.1.1.1 Le verbe télique

Critère : il s'agit de la fonction ou le but de l'objet

Etiquette : /SEM1_FS_VT/

Exemple :

47. a. BOUILLOIRE $_{N}$ < BOUILLIR → VT [télique bouillir]
    b. RASOIR $_{N}$ < RASER → VT [télique raser]
    c. EPLUCHEUR < EPLUCHER → VT [télique éplucher]

1.4.1.1.2 Le verbe formel

Critère : ce qui distingue l'objet d'un domaine plus large (orientation, taille, forme, dimension, couleur, position) : souvent son type

Etiquette : /SEM1_FS_VF/

Exemple :

48. a. NOUET $_{N}$ < NOUER → VF [formel nouer] (Linge noué (moyen de fonctionner), dans lequel on a placé une substance (médicamenteuse, aromatique) pour la faire infuser, bouillir (la fonction de cet objet, les verbes téliques).) (Grand Robert 2005)
    b. SONNETTE $_{N}$ < SONNER → VF [formel sonner] (Petit instrument métallique qui *sonne* (*moyen de fonctionner*) pour avertir (fonction, les verbes téliques).)
    c. PASSOIRE $_{N}$ < PASSER → VF [formel passer] (Récipient percé de trous et utilisé pour écraser ou égoutter (fonction, les verbes téliques) des aliments, pour filtrer (fonction, les verbes téliques) sommairement des liquides.) (Une passoire est un filtre qui revêt la forme d'un récipient percé de nombreux orifices. Ceux-ci permettent le *passage (moyen de fonctionner)* du liquide et retiennent les aliments de taille suffisante.)

1.4.1.2 Les rôles sémantiques joués par les bases nominales

Nous avons recensé sept rôles sémantiques et un rôle sémantique composé.

1.4.1.2.1 Patient

Critère : entité subissant l'action et d'une certaine manière affectée par celle-ci (Pustejovsky & Batiukova 2019 : 29).

Etiquette : /SEM1_FS_Patient/

Exemple :

49. a. HOMARDIER $_{N}$ < HOMARD → Patient (Bateau pour la pêche [télique **pêcher**] au **homard**, pourvu d'un vivier qui permet de transporter les crustacés vivants.)
    b. PELAMIDIERE $_{N}$ < PELAMID → Patient (Filet à petites mailles pour la pêche [télique **pêcher**] à la **pélamide**.)
    c. DOSSEUR < DOSSE → Patient (Machine servant à traiter [télique] les **dosses**.)

1.4.1.2.2 Thème

Critère : entité non-affectée localisé dans le temps ou l'espace, auquel est attribuée une propriété (Pustejovsky & Batiukova 2019 : 29 ; Abeillé et al. 2021 : 189)

Etiquette : /SEM1_FS_Thème/

Exemple :

50. a. ŒILLET $_{N}$ < ŒIL → Thème (Appareil ou machine effectuant la pose [télique **poser**] des **œillets**.)

1.4.1.2.3 Cible

Critère : entité subissant l'action et d'une certaine manière affectée par celle-ci (amenée à être déplacée) (Pustejovsky & Batiukova 2019 : 27).

Etiquette : /SEM1_FS_Cible/

Exemple :

51. a. MAZOUTEUR $_{N}$ < MAZOUT → Cible (Navire spécialisé dans le **transport** [télique **transporter**] du **mazout**.)
    b. GARGOUSSIER $_{N}$ < GARGOUSSE → Cible (Boîte utilisée pour **transporter** [télique] les **gargousses**.)
    c. PINARDIER < PINARD → Cible (Navire-citerne destiné au transport [télique **transporter**] du **vin (Vin ordinaire : pinard)**.)

1.4.1.2.4 Bénéficiaire

Critère : participant au profit ou au détriment duquel la situation existe (Abeillé et al. 2021 : 188).

Etiquette : /SEM1_FS_ Destinataire/

Exemple :

52. a. SERINETTE $_{N}$ < SERIN → Destinataire (Petit orgue mécanique dont on joue en actionnant une manivelle et avec lequel on **apprend** [Verbe télique] un **air** [Thème] aux **serins** [Destinataire], aux oiseaux chanteurs.)

1.4.1.2.5 Lieu

Critère : Lieu où l'événement a lieu (Pustejovsky & Batiukova 2019 : 29)

Etiquette : /SEM1_FS_ Lieu/

Exemple :

53. a. MAGASINIER $_{N}$ < MAGASIN → Lieu (Chariot élévateur pour la manutention [Verbe télique maintenir] des **marchandises** [Thème] en **magasin** [Lieu].)

    b. OREILLETTE $_{N}$ < OREILLE → Lieu (Petit récepteur (destiné à recevoir [Verbe télique] des impulsions électriques ou ondes électromagnétiques.) qui s'adapte à l'**oreille** [Lieu] (utilisé en télévision.)

    c. PLAFONNIER $_{N}$ < PLAFOND → Lieu (Appareil d'**éclairage** [Verbe télique **éclairer**] fixé au **plafond** [Lieu] sans être suspendu.)

1.4.1.2.6 Moyen

Critère : la manière dont une action est réalisée (Pustejovsky & Batiukova 2019 : 27)

Etiquette : /SEM1_FS_Moyen/

Exemple :

54. a. THERMOCOPIEUSE $_{N}$ < THERMOCOPIE → Moyen (Machine reproduisant [Verbe télique] des documents par thermocopie [Moyen].)

1.4.1.3 Les projections cognitives

1.4.1.3.1 Métaphore

Critère : la similitude entre les deux concepts

Etiquette : /SEM1_FS_ Métaphore /

Exemple :

55. a. LUNETTE $_{N}$ < LUNE → Métaphore (Instrument d'optique, généralement de forme **cylindrique**, comportant le plus souvent un objectif et un oculaire, servant à augmenter le diamètre apparent des objets éloignés et à en permettre une observation plus nette et plus distincte.) (objet en forme de lune) (TLFi)

1.4.1.3.2 Métonymie

Critère : la substitution d'une entité pour une autre entité avec une relation de contiguïté

Etiquette : /SEM1_FS_ Métonymie/

Exemple :

56. a. CHIMIQUIER $_{N}$ < CHIMIQUE → Métonymie (Cargo transportant des produits chimiques [properties of categories]) (chimique est la propriété la plus saillante pour les produits chimiques)

    b. VRAQUIER $_{N}$ < VRAC → Métonymie (Navire transportant des produits en vrac [properties of categories].) (en vrac est la propriété la plus saillante pour les produits en vrac)

    c. BOULONNIERE $_{N}$ < BOULON → Métonymie (Tarière à percer les trous des boulons [properties of categories] .) (BOULON est la propriété plus saillante pour la propriété saillante de l'outil)

1.4.2 Sous-classes ontologiques du dérivé

1.4.2.1 Instrument de Captation

Critère : « l'objet technique qui permet de prolonger et d'adapter le corps pour obtenir une meilleure perception et s'oppose à outil en ce qu'il inclut les systèmes artificiels capables d'améliorer la perception. » (Simondon 1969 : 114)

Etiquette : /SEM2_Instrument de captation/

Exemple :

57. a. RAPPORTEUR $_{N}$ → Instrument de captation [mesurer + la vue]

    b. LUNETTE $_{N}$ → Instrument de captation [augmentation du diamètre apparent des objets + la vue]

    c. COMPTEUR → N Instrument de captation [évaluation du temps + la vue]

1.4.2.2 Outil

Critère : Objet technique qui permet de prolonger et d'armer le corps pour accomplir un geste.

Etiquette : / SEM2_Outil /

Exemple :

58. a. FOURCHETTE $_N$ → Outil [piquer est un geste sans impliquer la perception]
    b. SONNETTE $_N$ → Outil [appuyer est un geste sans impliquer la perception]
    c. RASOIR $_N$ → Outil [raser est un geste sans impliquer la perception]

1.4.2.3 Appareil

Critère : Objet ayant un système technique, qui sert à percevoir et manifester techniquement les phénomènes du monde pour obtenir l'expérience des entités perçues (en adaptant de Vial 2012 ; 2017)

Etiquette : / SEM2_Appareil/

Exemple :

59. a. SCANNEUR $_N$ → Appareil [perception + système technique + expérience]
    b. AGRANDISSEUR $_N$ → Appareil [perception + système technique + expérience]
    c. INDICATEUR $_N$ → Appareil [perception + système technique + expérience]

1.4.2.4 Machine

Critère : « Objet fabriqué avec un système technique en vue d'accomplir une tâche donnée (en mécanisant les tâches de main-œuvre et les énergies naturelles) en utilisant des énergies variées. » (adaptation de Vial 2012)

Etiquette : / SEM2_Machine /

Exemple :

60. a. TONDEUSE $_N$ → Machine [mécanisation des tâches de main-œuvre + sans perception]
    b. REFRIGERATEUR $_N$ → Machine [mécanisation des phénomènes naturels + sans perception]
    c. EPLUCHEUSE → Machine [mécanisation des tâches de main-œuvre + sans perception]

1.4.2.5 Récipient

Critère : les récipients sont des objets destinés à stocker quelque chose et en même temps, servir à affecter le contenu, et se caractérisent par des rapports simples de profondeur et de diamètres (maximal et d'ouverture), et secondairement de dimensions. (en adaptant Balfet et al. (1985 : 7) et Polguère (2003 : 67))

Etiquette : / SEM2_Récipient /

Exemple :

61. a. CAFETIERE $_{N}$ → Récipient [stockage + infusion + des rapports simples de profondeur et de diamètres]

    b. THEIERE $_{N}$ → Récipient [stockage + infusion + des rapports simples de profondeur et de diamètres]

    c. PASSOIR $_{N}$ → Récipient [stockage + filtrer + des rapports simples de profondeur et de diamètres]

1.4.2.6 Substance chimique

Critère : corps obtenus à l'aide de procédés chimiques ; matières premières industrielles de l'industrie chimique. (*Le Grand Robert*)

Etiquette : / SEM2_ Substance chimique /

Exemple :

62. a. ACIDIFIANT $_{N}$ → Substance chimique

    b. DEBROUSSAILLANT $_{N}$ → Substance chimique

    c. DEGRAISSANT $_{N}$ → Substance chimique

2. Approche technique systématique

Dans le cadre de notre recherche, nous avons élaboré une **approche technique systématique** pour l'extraction et l'adaptation des données métalexicographiques issues du *Grand Robert* (édition 2005), à des fins d'analyse morphologique. Plus précisément, nous avons utilisé la fonction « Exporter l'article » du dictionnaire électronique pour extraire les entrées au format HTML (voir Figure A.2.1), puis les avons converties en fichiers *.txt*. Ce format favorise une adaptation efficace des annotations métalexicographiques aux balises linguistiques, permettant ainsi l'extraction ciblée des informations grâce à un logiciel textuel et facilitant leur classification automatique dans un tableur en fonction de leurs propriétés linguistiques.



Figure A.2.1 Interface du logiciel *Le Grand Robert* 2005

Tous les fichiers *txt* extraits ont été regroupés dans un seul fichier final. À partir de ce corpus unifié, nous avons procédé à l'adaptation des annotations métalexicographiques, en éliminant celles qui ne contribuent pas à notre objectif, tout en conservant celles utiles pour l'annotation linguistique. Nous avons ainsi identifié cinq catégories de balises linguistiques pertinentes (voir tableau A.2.1). Par exemple, l'annotation phonologique est signalée dans le code source par , ce qui permet de l'exploiter lors de l'extraction automatisée des informations phonologiques.

Tableau A.2.1 Notice de l'annotation des données

Balises	Fonctions	Propriétés linguistiques concernées

ENTREE	Indicateur nécessaire pour extraire les données et calculer le nombre d'entrées ou de lexèmes à étudier	La forme graphémique
PHON	Indicateur pour extraire les informations phonologiques	Phonologie
CAT	Indicateur pour extraire les informations sur la catégorie syntaxique	Syntaxe
ETYM	Indicateur pour extraire les informations étymologiques	L'étymologie, la datation et le processus de formation morphologique
DEF	Indicateur pour extraire les définitions	Sémantique

Pour la structuration des fichiers finaux, chaque étiquette linguistique (sauf « ENTRÉE ») est encadrée par deux balises : l'une marquant le début (ancrage), l'autre la fin (frontière). Par exemple, l'information phonologique est encadrée par <PHON> et </PHON>, ce qui permet de distinguer clairement les différentes catégories d'informations linguistiques (voir figure 3). Grâce à ce système d'étiquetage, nous pouvons extraire les informations complètes ou partielles des entrées selon les besoins de notre analyse.

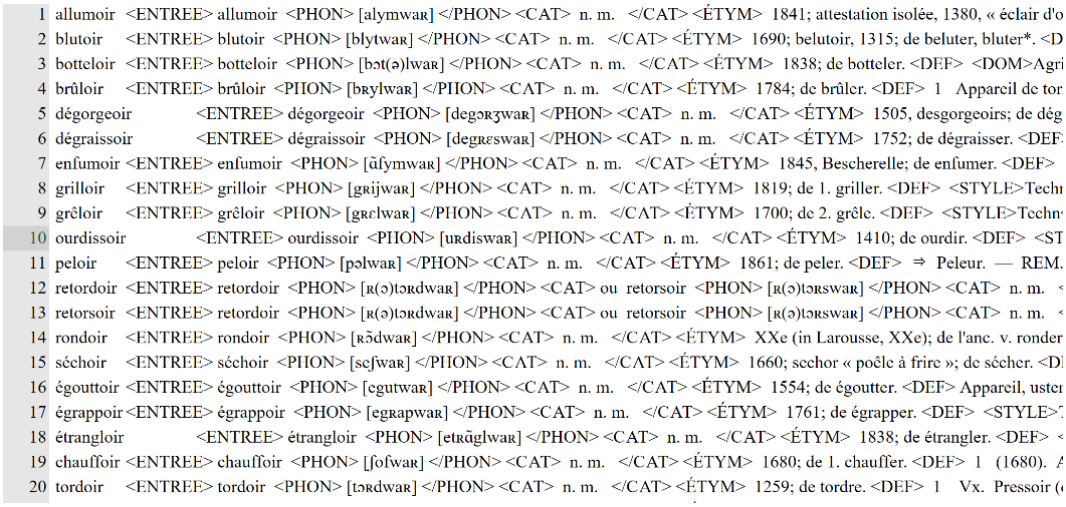
1 allumoir <ENTREE> allumoir <PHON> [alymwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1841; attestation isolée, 1380, « éclair d'o
2 blutoir <ENTREE> blutoir <PHON> [blytwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1690; belutoir, 1315; de beluter, bluter*. <D
3 botteloir <ENTREE> botteloir <PHON> [bɔt(ə)lwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1838; de botteler. <DEF> <DOM>Agri
4 brûloir <ENTREE> brûloir <PHON> [bʀylwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1784; de brûler. <DEF> 1 Appareil de tor
5 dégorgeoir <ENTREE> dégorgeoir <PHON> [degɔʀʒwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1505, desgorgeoirs; de dég
6 dégraissoir <ENTREE> dégraissoir <PHON> [degʀɛswaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1752; de dégraisser. <DEF
7 enfumoir <ENTREE> enfumoir <PHON> [ɑ̃fymwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1845, Bescherelle; de enfumer. <DEF>
8 grilloir <ENTREE> grilloir <PHON> [gʀijwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1819; de 1. griller. <DEF> <STYLE>Techn
9 grêloir <ENTREE> grêloir <PHON> [gʀɛlwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1700; de 2. grêle. <DEF> <STYLE>Techn
10 ourdissoir <ENTREE> ourdissoir <PHON> [uʀdiswaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1410; de ourdir. <DEF> <ST
11 peloir <ENTREE> peloir <PHON> [pəlwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1861; de peler. <DEF> ⇒ Peleur. — REM.
12 retordoir <ENTREE> retordoir <PHON> [ʀ(ə)tɔʀdwaʀ] </PHON> <CAT> ou retorsoir <PHON> [ʀ(ə)tɔʀswaʀ] </PHON> <CAT> n. m.
13 retorsoir <ENTREE> retordoir <PHON> [ʀ(ə)tɔʀdwaʀ] </PHON> <CAT> ou retorsoir <PHON> [ʀ(ə)tɔʀswaʀ] </PHON> <CAT> n. m.
14 rondoir <ENTREE> rondoir <PHON> [ʀɔ̃dwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> XXe (in Larousse, XXe); de l'anc. v. ronder
15 séchoir <ENTREE> séchoir <PHON> [seʃwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1660; sechor « poêle à frire »; de sécher. <D
16 égouttoir <ENTREE> égouttoir <PHON> [egutwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1554; de égoutter. <DEF> Appareil, uste
17 égrappoir <ENTREE> égrappoir <PHON> [egʀapwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1761; de égrapper. <DEF> <STYLE>
18 étrangloir <ENTREE> étrangloir <PHON> [etʀɑ̃glwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1838; de étrangler. <DEF>
19 chauffoir <ENTREE> chauffoir <PHON> [ʃofwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1680; de 1. chauffer. <DEF> 1 (1680).
20 tordoir <ENTREE> tordoir <PHON> [tɔʀdwaʀ] </PHON> <CAT> n. m. </CAT> <ÉTYM> 1259; de tordre. <DEF> 1 Vx. Pressoir (

Figure A.2.2 Extrait de fichiers finaux annotés par les étiquettes linguistiques

Une fois le corpus source établi, nous avons utilisé le logiciel PowerGrep pour extraire les données linguistiques pertinentes. Ce logiciel permet l'extraction (avec ou sans expressions régulières) et la modification des données. À titre illustratif, l'extraction des informations phonologiques des lexèmes en *-oir* a été effectuée en appliquant une expression régulière ciblée (voir figure 4). Les résultats montrent que les lexèmes et leurs annotations phonologiques ont été correctement extraits (figure 5).

Les balises ont ensuite été supprimées, et les données résiduelles ont été organisées en deux colonnes dans un tableur, intitulées respectivement « lexème » et « PHON (phonologie) ».

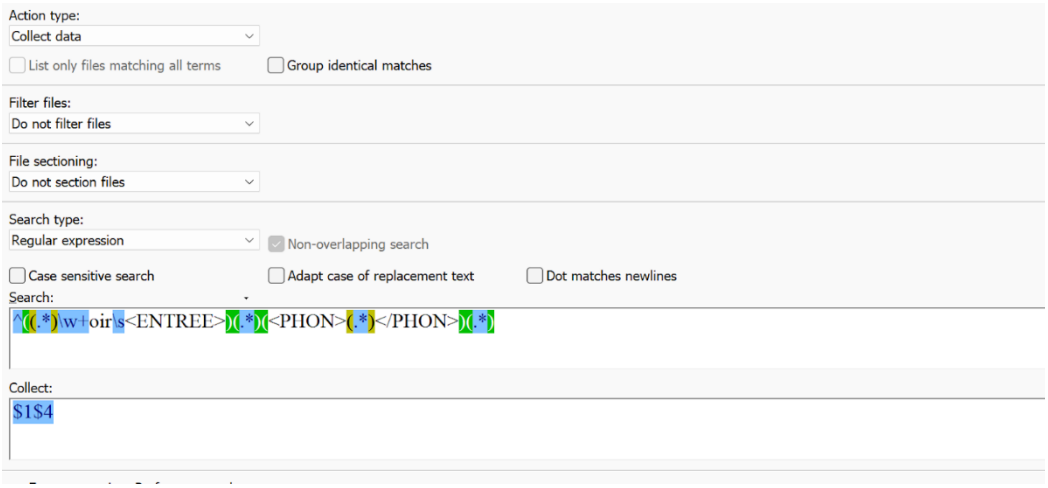

Figure A.2.3 Extraction des informations phonologiques avec PowerGrep 4

1	allumoir	<ENTREE><PHON> [alymwaʀ] </PHON>
2	blutoir	<ENTREE><PHON> [blytwaʀ] </PHON>
3	botteloir	<ENTREE><PHON> [bɔt(ə)lwaʀ] </PHON>
4	brûloir	<ENTREE><PHON> [bʀylwaʀ] </PHON>
5	dégorgeoir	<ENTREE><PHON> [degɔʀʒwaʀ] </PHON>
6	dégraissoir	<ENTREE><PHON> [degʀɛswaʀ] </PHON>
7	enfumoir	<ENTREE><PHON> [ɑ̃fymwaʀ] </PHON>
8	grilloir	<ENTREE><PHON> [gʀijwaʀ] </PHON>
9	grêloir	<ENTREE><PHON> [gʀɛlwaʀ] </PHON>
10	ourdissoir	<ENTREE><PHON> [uʀdiswaʀ] </PHON>
11	peloir	<ENTREE><PHON> [pəlwaʀ] </PHON>
12	retordoir	<ENTREE><PHON> [ʀ(ə)tɔʀswaʀ] </PHON>
13	retorsoir	<ENTREE><PHON> [ʀ(ə)tɔʀswaʀ] </PHON>
14	rondoir	<ENTREE><PHON> [ʀɔ̃dwaʀ] </PHON>
15	séchoir	<ENTREE><PHON> [seʃwaʀ] </PHON>
16	égouttoir	<ENTREE><PHON> [egutwaʀ] </PHON>
17	égrappoir	<ENTREE><PHON> [egʀapwaʀ] </PHON>
18	étrangloir	<ENTREE><PHON> [etʀɑ̃glwaʀ] </PHON>
19	chauffoir	<ENTREE><PHON> [ʃofwaʀ] </PHON>
20	tordoir	<ENTREE><PHON> [tɔʀdwaʀ] </PHON>

Figure A.2.4 Les résultats des informations phonologiques

Cette approche, conçue par nous, constitue une méthode originale et rigoureuse pour exploiter les ressources du *Grand Robert* (2005) dans le cadre d'une recherche en morphologie. De plus, avec sa flexibilité et à sa simplicité de mise en œuvre, cette

méthode peut aisément être adaptée à d'autres dictionnaires électroniques et à divers types de recherches linguistiques ou morphologiques.

3. Outil technique

Afin de calculer rapidement le nombre de syllabes en français, nous avons développé un programme Excel (fichier xlsm) basé sur un tableau des voyelles, conçu pour le calcul automatique du nombre de syllabes des lexèmes ou des radicaux.

3.1 Processus d'innovation

La création de ce programme s'appuie sur une analyse systématique du système vocalique du français. Nous avons d'abord recensé l'ensemble des phonèmes vocaliques du français et les avons normalisés selon la transcription de l'Alphabet Phonétique International (API). Chaque phonème a ensuite été saisi individuellement dans la feuille principale « VowDicSheet » (voir la Figure A.3.1), chaque ligne correspondant à une voyelle, afin de garantir une structure de données simple, normalisée et facilement exploitable par le programme. L'ensemble du processus s'est appuyé sur le manuel de l'API (International Phonetic Association 1999) afin d'assurer la rigueur scientifique et la représentativité des voyelles sélectionnées.

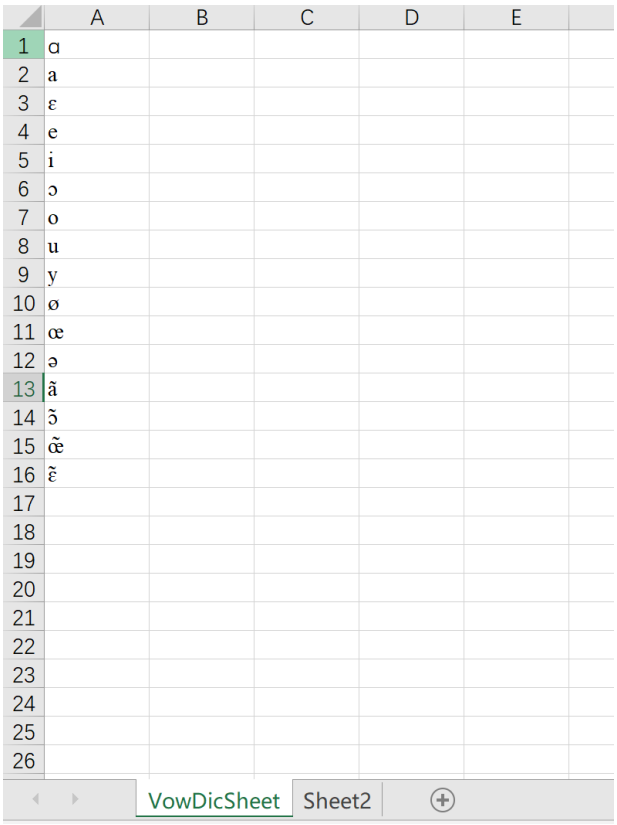

Figure A.3.1 Feuille principale « VowDicSheet »

3.2 Processus d'utilisation

Ce programme utilise une macro interne d'Excel qui lit directement les voyelles de la feuille « VowDicSheet » afin de réaliser le calcul automatique et en lot du nombre de syllabes. Concrètement, il suffit à l'utilisateur d'entrer les lexèmes ou radicaux à analyser dans la colonne A de la feuille « Sheet2 ». Ensuite, *via* le menu « Affichage » → « Macro » → « Afficher les macros », il faut sélectionner et exécuter la macro concernée : le nombre de syllabes pour chaque lexème ou radical sera automatiquement affiché dans la colonne B. Ce processus permet d'effectuer rapidement le comptage syllabique et d'automatiser le traitement des données phonologiques.

Bibliographie

Abeillé, A. et al. (2021). *La Grande Grammaire du français* (2 vols.). Paris, France : Actes Sud.

Abrams, P. A., & Rueffler, C. (2009). Coexistence and limiting similarity of consumer species competing for a linear array of resources. *Ecology*, *90*(3), 812–822.

Amiot, D. (2004). Préfixes ou prépositions ? Le cas de sur-, sans-, contre- et les autres. *Lexique*, 16, 67–83.

Amiot, D., & Montermini, F. (2009). Affixes et mots grammaticaux. In B. Fradin, F. Kerleroux, & M. Plénat (Eds.), *Aperçus de morphologie du français*. Saint-Denis : Presses Universitaires de Vincennes, 127-141.

Anshen, F., & Aronoff, M. (1981). Morphological productivity and phonological transparency. *Canadian Journal of Linguistics*, *26*(1), 63–72.

Anshen, F., & Aronoff, M. (1999). Using dictionaries to study the mental lexicon. *Brain and Language, 68*(1-2), 16-26.

Anscombre, J.-C. (2001). À propos des mécanismes sémantiques de formation des noms d'agent en français et en espagnol. *Langages*, *143*, 28–48.

Anscombre, J-C. (2015). Pour une approche morphosémantique des noms d'action en -tion, -ment et -age. *Revue de sémantique et pragmatique. 35-36* : 49-80.

Aristote. (1840). *La Métaphysique* (A. Pierron & C. Zévort, Trans.). Paris: Ébrard.

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

Aronoff, M. (1994). *Morphology by itself: Stems and inflectional classes*. MIT Press.

Aronoff, M. (2016). Competition and the lexicon. In A. Elia, C. Iacobini, & M. Voghera (Eds.), *Livelli di Analisi e fenomeni di interfaccia: Atti del XLVII congresso internazionale della Società di Linguistica Italiana*. Roma: Bulzoni Editore, 39-52.

Aronoff, M. (2019). Competitors and alternants in linguistic morphology. In F. Rainer, F. Gardani, W. Dressler, & H. C. Luschützky (Eds.), *Competition in Inflection and Word Formation*. Berlin: Springer, 39-66.

Aronoff, M. (2023). Three ways of looking at morphological rivalry. *Word Structure*, 16(1), 49–62.

Aronoff, M., & Lindsay, M. (2014). Productivity, blocking, and lexicalization. In S. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Derivational Morphology*. Oxford: Oxford University Press, 95-112.

Aronoff, M., & Lindsay, M. (2016). *Partial organization in languages: La langue est un système où la plupart se tient*. In *Proceedings of Décembrettes 8*. New York : Stony Brook University, 1-14.

Audring, J., & Jackendoff, R. (2025). Construction morphology and relational morphology. In M. Fried & K. Nikiforidou (Eds.), *The Cambridge Handbook of Construction Grammar* (pp. 101–128). Cambridge : Cambridge University Press.

Bachelard, G. (1938), *La Formation de l'esprit scientifique*, Paris, Vrin, 1993.

Bachelard, G. (1951), *L'Activité rationaliste de la physique contemporaine*, Paris, PUF, 1965.

Balfet, H., Fauvet-Berthelot, M.-F., & Monzon, S. (1983). *Pour la normalisation de la description des poteries*. Paris : Éditions du Centre national de la recherche scientifique (CNRS).

Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford : Oxford University Press.

Benveniste, E. (1975). *Problèmes de linguistique générale* (Vol. 2). Paris: Gallimard.

Bobkova, N. (2023). *La concurrence suffixale dans la construction des adjectifs dénominaux en russe : analyse des suffixes -n-, -sk- et -ov-*. Thèse de doctorat, Université Toulouse 2 – Jean Jaurès.

Bonami, O. (2014). *La structure fine des paradigmes de flexion* (Habilitation à diriger des recherches, Université Paris Diderot). Paris : Université Paris Diderot.

Bonami, O., & Boyé, G. (2003). Supplétion et classes flexionnelles dans la conjugaison du français. *Langages*, (152), 102–126.

Bonami, O., & Boyé, G. (2005). Construire le paradigme d'un adjectif. *Recherches linguistiques de Vincennes*, *34*, 77–98.

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., Boyé, G., & Kerleroux, F. (2009). L'allomorphie radical et la relation flexion-construction. In Fradin B, Kerleroux G & Plénat M. (éds). *Aperçus de morphologie du français,* 103-126.

Booij, G. (2010). *Construction Morphology*. Oxford : Oxford University Press.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5-32.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Cadiot, P. (1991). À la hache ou avec la hache ? Représentation mentale, expérience située et donation du référent. *Langue française, 91*, 7–23.

Carnie, A. (2007). *Syntax: A generative introduction* (2nd ed.). Malden, MA: Blackwell Publishing.

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives.* London: Routledge.

Cartier, E., & Huyghe, R. (2021). La concurrence affixale en diachronie : Le cas des préfixes de haut degré en français. *Linx*, *82*, 1–15.

Canguilhem, G. (1993). *La connaissance de la vie* (2e éd.). Paris : Vrin.

Centre National de Ressources Textuelles et Lexicales. (n.d.). *Trésor de la langue française informatisé* (TLFi). CNRTL.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpsetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics, 33*(18), 2938-2940.

Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer.

Corbin, D., & Temple, M. (1994). Le monde des mots et des sens construits : catégories sémantiques, catégories référentielles. *Cahiers de lexicologie, 65*(2), 5–28.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford : Oxford University Press.

Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics* (1st ed.). Oxford : Oxford University Press.

Dal, G. (1999). Suffixation par –et(te) et bases verbales. In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché (Eds.), *La morphologie des dérivés évaluatifs, Silexicales* 2, 37-47.

Dal, G. (2003). Analogie et lexique construit : Quelles preuves ? *Cahiers de Grammaire*, *28*, 9–30.

Darmesteter, A. (1877). *La vie des mots étudiée dans leurs significations*. Paris : Librairie Ch. Delagrave.

Darwin, C. R. (1872). *The expression of the emotions in man and animals* (p. 85). London : John Murray.

Devos, F., & Taeldeman, J. (2003). Deverbal nouns and the agentive dimension across languages. In D. Willems, B. Defrancq, T. Colleman, & D. Noël (Eds.), *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison*. London : Palgrave Macmillan, 155-171.

Diderot, D., d'Alembert, J. le R. (Éds.). (1900). *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*. Paris : Librairie Classique Eugène Belin, 1751-1772.

Dirven, R. (1999). Conversion as a conceptual metonymy of event schemata. In Klaus-Uwe Panther and Günter Radden (Eds.) *Metonymy in Language and Thought*, 275-288.

Dixon, R. M. W. (2014). *Making New Words: Morphological Derivation in English*. Oxford : Oxford University Press.

Dubois, J. (1962). *Étude sur la dérivation suffixale*. Paris : Larousse.

Eluerd, L. (2009). *Littré Conjugaison*. Paris : Librairie Garnier.

Faber, P. (2012). *A cognitive linguistics view of terminology and specialized language*. De Gruyter Mouton.

Fauconnier, G. (1994). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge, MA : MIT Press.

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York : Basic Books.

Ferret, K. (2013). Semantic propertie(s) of the base verbs and locative or instrumental interpretation of *N-oir(e)*. Paper presented at the *9th Mediterranean Morphology Meeting*, University of Zagreb, Dubrovnik, Croatia.

Ferret, K., Soare, E., & Villoing, F. (2010). Rivalry between French *-age* and *-ée*: The role of grammatical aspect in nominalization. In M. Aloni, H. Bastiaanse, T. De Jager, & K. Schultz (Eds.), *Lecture Notes in Computer Science: Vol. 6042. Logic, language and meaning. 17th Amsterdam Colloquium, The Netherlands, December 2009, Revised selected papers*. Berlin : Springer, 284-295.

Ferret, K., & Villoing, F. (2015). French N-age instrumentals: Semantic properties of the base verb. *Morphology*, *25*(4), 473–496.

Fisher, R. (1935). *The Design of Experiments.* Edinburgh : Oliver & Boyd.

Fillmore, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York : Holt, Rinehart and Winsto, 1-88.

Fillmore, C. J. (1977). The case for case reopened. In P. Cole & J. Sadock (Eds.), *Syntax and semantics: Vol. 8. Grammatical relations*. New York : Academic Press, 59-81.

Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morning calm*. Seoul : Hanshin Publishing Co, 111-137.

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

Fradin, B. (2003). *Nouvelles approches en morphologie*. Presses Universitaires de France.

Fradin, B. (2005). On a semantically grounded difference between derivation and compounding. In G. Booij, G. Ralli, & S. Scalise (Eds.). Amsterdam/Philadelphia : John Benjamins.

Fradin, B. (2019). Competition in inflection and word-formation. In *Morphology and its demarcations: Selected papers from the 11th Morphology meeting*. Vienna: Austrian Committee for Morphology, 67-93.

Fradin, B., Hathout, N., & Meunier, F. (2003). La suffixation en -ET et la question de la productivité. *Langue française*, *140*(4), 56–78.

Fradin, B., & Winterstein, G. (2013). Tuning agentivity and instrumentality: deverbal nouns in -oir revisited, 1-18 (article non publié).

Frisch, S. A. (2012). Phonotactic patterns in lexical corpora, in Abigail C. Cohn, Cécile Fougeron, and Marie K. Huffman (Eds), *The Oxford Handbook of Laboratory Phonology.* 458-470.

Gause, G. F. (1934). *The struggle for existence*. Baltimore, MD : Williams & Wilkins.

Ge, F. (2008). *Modern Ecology* (2nd Edition) 【现代生态学 第二版】. Beijing : Science Press.

Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford : Oxford University Press.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago : University of Chicago Press.

Goldberg, A. E. (2009). The nature of generalization in language. *Cognitive. Linguistics, 20,* 93-127.

Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions.* New Jersey : Princeton University Press.

Grevisse, M., & Goosse, A. (2016). *Le bon usage* (16ème éd.). Louvain-la-Neuve : De Boeck Supérieur.

Gréa, P. (2002). Intégration conceptuelle et métaphore filée. *Langue française*, *134*, 109–123.

Gries, S. T. (2021). *Statistics for linguistics with R: A practical introduction* (3rd ed.). Berlin : De Gruyter.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA : MIT Press.

Grinnell, J. (1917). The niche-relationships of the California Thrasher. *Auk*, *34*(4), 427–433.

Grzegorczykowa, R., & Puzynina, J. (1984). Problemy ogólne słowotwórstwa. Słowotwórstwo rzeczowników. In *Gramatyka współczesnego języka polskiego. Morfologia* (pp. 307–407). Warszawa : PWN.

Guzmán Naranjo, M., & Bonami, O. (2023). Distributional evidence for derivational paradigms. In S. Kotowski & I. Plag (Eds.), *The semantics of derivational morphology: Theory, methods, evidence*. Berlin/Boston : De Gruyter, 219-258.

Haas, P., Barque, L., Huyghe, R., & Tribout, D. (2022). Pour une classification sémantique des noms en français appuyée sur des tests linguistiques. *Journal of French Language Studies, 32*(1), 1–30.

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20*. Cambridge, MA : MIT Press, 117-176.

Hamawand, Z. (2011). *Morphology in English: Word formation in cognitive grammar*. London : Continuum.

Hardin, G. (1960). The competitive exclusion principle. *Science*, *131*(3409), 1292–1297.

Hay, J., & Plag, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, *22*(3), 565–596.

Hilpert, M. (2013). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge University Press.

Hurlbert, S. H. (1978). The measurement of niche overlap and some relatives. *Ecology, 59*(1), 67–77.

Huot, H. (2001). *Morphologie : forme et sens des mots du français* (1re éd.). Paris : Armand Colin.

Huot, H. (2005). *La Morphologie : forme et sens des mots du français* (2e éd.). Paris : Armand Colin.

Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 415-427.

Huyghe, R., & Lombard, A. (2022). Les néologismes en *-age* en français contemporain : Héritage verbal et polysémie. *Journal of French Language Studies, 32*(1), 25–47.

Huyghe, R., & Tribout, D. (2015). Noms d'agents et noms d'instruments : le cas des déverbaux en -eur. *Langue française, 185*(1), 99-112.

Huyghe, R., & Varvara, R. (2023). Affix rivalry: Theoretical and methodological challenges. *Word Structure*, *16*(1), 1–23.

Huyghe, R., & Wauquier, M. (2021). What's in an agent?: A distributional semantics approach to agent nouns in French. *Word Structure*, *14*(3), 354–391.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford : Oxford University Press.

Jalenques, P. (2014). Proposition d'analyse synchronique de certains emplois idiomatiques de verbes préfixés en DÉ- avec une base verbale, séminaire TOPE (Théorie des Opérations Prédicatives et Enonciatives d'Antoine Culioli), Journée d'étude La compositionnalité des verbes préfixés, Université de Paris 7.

Jalenques, P. (2015). Le problème de l'opacité sémantique dans les verbes préfixés à base verbale : Pour une approche sémantique constructiviste. L'exemple du préfixe DÉ- en français actuel, séminaire de recherche du laboratoire ATILF, Université de Nancy.

Ježek, E. (2016). *The lexicon: An introduction*. Oxford University Press.

Jin, L. (2011). *What is linguistic typology* [什么是语言类型学]. Shanghai : Shanghai Foreign Language Education Press.

Jin, C., Wu, C., Ding, Y., Yuan, W., Zhu, J., Huang, Y., Jiao, J., Zang, R., & Jiang, B. (2021). The functional groups and succession characteristics of dominant populations in an evergreen secondary broad-leaved forest of Wuchao Mountain. *Acta Ecologica Sinica, 41*(8) : 3053-3066**.**

Klausenburger, J. (1970). *French prosodics and phonotactics: An historical typology.* Berlin, New York : Max Niemeyer Verlag.

Koehl, A. (2012). *La construction morphologique des noms désadjectivaux suffixés en français.* Université de Lorraine.

Knittel, M. (2017a). French deverbal nominals in *–ant*: semantic properties, ISMo 2017, Université de Lille, Lille, France.

Knittel, M. (2017b). French derived nominals in*–ant*: a semantic typology, JENOM 7, Université de Fribourg, Fribourg, Suisse.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Volume 1. Theoretical prerequisites*. Stanford, CA : Stanford University Press.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. Oxford : Oxford University Press.

Langacker, R. W. (2014). Settings, participants, and grammatical relations. In S. L. Tsohatzidis (Ed.), *Meanings and prototypes*. London : Routledge, 213-238.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago : University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago : University of Chicago Press.

Lapraye, A. (2017). *Une approche statistique de la concurrence entre procédés constructionnels la dérivation en -age et en -ment en français*. Mémoire de Master, Université Paris Diderot.

Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax–lexical semantics interface*. Cambridge, MA : MIT Press.

Lex, A., & Gehlenborg, N. (2014). Points of view: Sets and intersections. *Nature Methods, 11*(8), 779.

L'Homme, M.-C. (2020). *Lexical Semantics for Terminology An introduction*. Amsterdam / Philadelphia : John Benjamins Publishing Company.

Lieber, R. (2014). Theoretical approaches to derivation. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Derivational Morphology*. Oxford : Oxford University Press, 52-69.

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

Lieber, R. (2021). *Introducing Morphology 3rd edition*. Cambridge : Cambridge University Press.

Lignon, S. (2013). '-ISER and -IFIER suffixation in French: Verifying data to 'verize' hypotheses'. In N. Hathout, F. Montermini, and J. Tseng (eds.), *Morphology in Toulouse, Selected prodeedings of Décembrettes 7*. Munich : Lincom Europa, 119–132.

Lin, Y. (2021). Recherche sur la compétition morphologique des suffixes instrumentaux du français sous le principe d'exclusion compétitive. *Etude Francophone, 26* (3), 26-34.

Lin, Y. (2023). Les contraintes et la productivité morphologiques des suffixes *-oir* et *-oire*. *Journée de doctorant 2023,* Nanterre.

Lin, Y. (soumis a). The morphological productivity of French suffixes from the perspective of suffix ordering.

Lin, Y. (soumis b). A Quantitative Study of French Suffix Ordering: Morphological Constraints and Etymological Influence.

Lindsay, M. (2012). Rival suffixes: Synonymy, competition, and the emergence of productivity. *Proceedings of Mediterranean Morphology Meetings*, *8*, 192–203.

Lindsay, M., & Aronoff, M. (2013). Natural selection in self-organizing morphological systems. In F. Montermini, G. Boyé, & J. Tseng (Eds.), *Morphology in Toulouse: Selected Proceedings of Décembrettes 7*. Munich : Lincom Europa, 133-153.

Littlemore, J. (2015). *Metonymy: Hidden shortcuts in language, thought and communication*. Cambridge : Cambridge University Press.

Liu, Z, & Li, Y. (2019). *Dix conférences sur la linguistique cognitive*. Shanghai : Éditions de l'éducation en langues étrangères de Shanghai.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge : Cambridge University Press.

Martin, F. (2010). The semantics of eventive suffixes in French. In A. Alexiadou & M. Rathert (Eds.), *The semantics of nominalizations across languages and frameworks*. Berlin : Mouton de Gruyter, 109-140.

Masini, F., & Audring, J. (2019). Construction morphology. In J. Audring & F. Masini (Eds.), *The Oxford handbook of morphological theory*. Oxford : Oxford University Press, 365-389.

McMahon, A. (2020). *An introduction to English phonology (2nd edition)*. Edinburgh : Edinburgh University Press.

Mei, D. (2017). *Encyclopedia of Linguistics and Applied linguistics*. Beijing : Université de Beijing.

Meyer-Lübke, W. (1895). *Grammaire des langues romanes : Morphologie* (Vol. 2). Paris : C. Klincksieck.

Missud, A. (2023). *Modélisation de la compétition entre suffixations nominalisatrices : Productivité et contraintes morphologiques.* Thèse de doctorat, Université Paris Nanterre.

Missud, A., & Villoing, F. (2021). Investigating the distributional properties of rival -age suffixation and verb-to-noun conversion in French. *Verbum*, *43* (1), 41–68.

Missud, A., & Villoing, F. (2022). The morphology of rival -ion, -age and -ment selected verbal bases. *Lexique, 26,* 29-52.

Monod, J. (1970). *Le hasard et la nécessité: Essai sur la philosophie naturelle de la biologie moderne*. Paris : Éditions du Seuil.

Nagano, A., Bagasheva, A., & Renner, V. (2024). Towards a competition-based word-formation theory: Core research questions and major hypotheses. In A. Bagasheva, A. Nagano, & V. Renner (Eds.), *Competition in word-formation*. Amsterdam : John Benjamins, 1-32.

Namer, F. (2009). *Morphologie, lexique et traitement automatique des langues*. Cachan : Hermès-Lavoisier.

Namer, F., & Villoing, F. (2008). Interpréter les noms déverbaux : quelle relation avec la structure argumentale du verbe de base? Le cas des noms en–*oir(e)* du français. Dans Durand J. Habert B., Laks B. (Eds.), *Actes du Congrès Mondial de Linguistique Française 2008.* Paris : Institut de Linguistique Française, 1539-1557.

Nagórko, A. (1981). Motywacja derywatów rzeczownikowych we współczesnej polszczyźnie. *Prace Filologiczne*, *21*, 183–258.

Oczkowska, B. (1983). Formalno-syntaktyczne wyznaczniki serbskochorwackiej kategorii nomen instrumenti. *Roczniki Slawistyczne*, *44* (1), 39–46.

Odum, H. T. (1950). *The biogeochemistry of strontium: With discussion on the ecological integration of elements*. Thèse de doctorat, Yale University.

Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Berlin : Mouton de Gruyter.

Plag, I. (2018). *Word-formation in English (2nd ed.)*. Cambridge : Cambridge University Press.

Plag, I., & Baayen, H. (2009). Suffix ordering and morphological processing. *Language*, *85*(1), 109–152.

Plénat, M. (2005). Rosinette, cousinette, putinette, starlinette, chipinette Décalage, infixation et épenthèse devant *–ette*. In I. Choï Jonin, M. Bras, A. Dagnac & M. Rouquier (éds). *Questions de classification en linguistique : méthodes et descriptions. Mélanges offerts au Professeur Christian Molinier.* Berne : Peter Lang, 275-298.

Plénat, M. (2008). Le thème L de l'adjectif et du nom. Dans J. Durand, B. Habert & B. Laks (Éds.), *Actes du Congrès Mondial de Linguistique Française 2008.* Paris : Institut de Linguistique Française, 1613-1626.

Plénat, M. (2009). Les contraintes de taille. In Fradin B, Kerleroux G & Plénat M. (éds). *Aperçus de morphologie du français,* Saint-Denis : Presses Universitaires de Vincennes, 47-65.

Pocheville, A. (2011). La niche écologique : histoire et controverses récentes. In T. Heams, P. Huneman, G. Lecointre & M. Silberstein (Eds), *Les Mondes darwiniens. L'évolution de l'évolution*, Paris : éditions matériologiques, 897-933.

Polguère, A. (2003). Étiquetage sémantique des lexies dans la base de données DiCo. *Traitement Automatique des Langues*, *44*(2), 39–68.

Polguère, A. (2016). *Lexicologie et sémantique lexicale: Notions fondamentales* (3e éd.). Montréal : Les Presses de l'Université de Montréal.

Pustejovsky, J. (1996). *The generative lexicon*. Cambridge : MIT Press.

Pustejovsky, J., & Batiukova, O. (2019). *The Lexicon: An introduction*. New York : Cambridge University Press.

Rabardel, P. (1995). *Les hommes et les technologies : approche cognitive des instruments contemporains*. Paris : Armand Colin.

Radden, G., & Kövecses, Z. (1999). *Towards a theory of metonymy*. In K.-U. Panther & G. Radden (Eds.), *Metonymy in Language and Thought*. Amsterdam/Philadelphia : John Benjamins, 17-59.

Rainer, F. (2005). Typology, diachrony, and universals of semantic change in word formation: A Romanist's look at the polysemy of agent nouns. In G. Booij, A. Ralli, S. Scalise, & S. Sgroi (Eds.), *Proceedings of the Fifth Mediterranean Morphology Meeting*. Catania : University of Catania, Faculty of Letters.

Rainer, F. (2021). Instrument and place nouns in the Romance languages. In Mark Aronoff (Ed.), *Oxford Research Encyclopedias: Linguistics*. Oxford : Oxford University Press, 1-27.

Riemer, N. (2010). *Introducing semantics*. Cambridge : Cambridge University Press.

Rey, A. (Ed.). (2005). *Le Grand Robert de la langue française* (Nouvelle édition). Paris : Le Robert.

Rey, A., & Rey-Debove, J. (2022). *Le Petit Robert de la langue française* (Éd. 2022). Paris : Le Robert.

Robert, P., & Rey, A. (Éds.). (2023). *Le Grand Robert de la langue française* (9e éd.). Paris : Le Robert.

Roché, M. (1998). *Deux études sur la dérivation en-ier (e)*. Université de Toulouse-Le Mirail, Équipe de recherche en syntaxe et sémantique.

Roché, M. (2003). De la « bonne formation » des dérivés en *–on*. *Cahiers de Grammaire*, *28*, 91–112.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 1). London : Academic Press, 1–49.

Rosenberg, M. (2008). La formation agentive en français : les composés [VN/A/Adv/P]N/A et les dérivés V-ant, V-eur et V-oir(e). Thèse de doctorat. Stockholm University.

Rosenberg, M. (2011). The polysemy of agent nouns: Diachronic, synchronic and contrastive evidence from French and Swedish. *STUF – Language Typology and Universals, 64(1)*, 53–64.

Roy, I., & Soare, E. (2012). L'enquêteur, le surveillant et le détenu : Les noms déverbaux de participants aux événements, lectures événementielles et structure argumentale. *Lexique*, *20*, 207–231.

Saarinen, P., & Hay, J. (2014). Affix ordering in derivation. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology*. Oxford : Oxford University Press, 370–383.

Salvadori, J., & Huyghe, R. (2023). Affix polyfunctionality in French deverbal nominalizations. *Morphology, 33*, 1–39.

Salvadori, J., Varvara, R., & Huyghe, R. (2024). Measuring affix rivalry as a gradient relationship. In A. Bagasheva, A. Nagano, & V. Renner (Eds.), *Competition in word-formation*. Amsterdam : John Benjamins, 104-138.

Schmid, H.-J. (2016). *English Morphology and Word-formation: An Introduction 3rd edition*. Erich Schmidt Verlag.

Schlesinger, I. M. (1981). Semantic assimilation in the development of relational categories. In W. Deutsch (Ed.), *The child's construction of language*. London : Academic Press, 223-243.

Simondon, G. (2001). *Du mode d'existence des objets techniques* (éd. revue et corrigée). Paris : Aubier.

Śliwa, D. (1990). Étude sémantique des nomina instrumenti en *-oir*. *Roczniki Humanistyczne, 37(5)*, 57–70.

Skousen, R., & Stanford, T. (2007). *Analogical modeling and morphological change: The case of the adjectival negative prefix in English*. In R. Skousen, D. Lonsdale, & D. B. Parkinson (Eds.), *Analogical modeling of language*. Amsterdam : John Benjamins, 157-179.

Spence, J. (1990). Instrumental "-Eur/-Ateur/-Euse/-Atrice". *The Modern Language Review, 85*(1), 29-35.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, *9*, 307.

Stefanowitsch, A., & Gries, S. T. (2005). Covarying collexemes. *Corpus Linguistics and Linguistic Theory, 1*(1), 1-43.

Strnadová, J. (2014). *Les réseaux adjectivaux. Sur la grammaire des adjectifs dénominaux en français*. Thèse de doctorat, Université Paris Diderot – Paris 7; Univerzita Karlova, Prague.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter, Kongelige Danske Videnskabernes Selskab*, *5*, 1–34.

Taylor, J. R. (2003). *Linguistic categorization* (3rd ed.). Oxford : Oxford University Press.

Traugott, E. C. (2008). *Constructionalization and constructional changes*. Oxford : Oxford University Press.

Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français*. Thèse de doctorat, Université Paris Diderot.

Thuillier, J., Tribout, D., & Wauquier, M. (2023). Affix rivalry in French demonym formation: The role of linguistic and non-linguistic parameters. *Word Structure, 16*(1), 115–146.

Uth, M. (2010). The rivalry of French -ment and -age from a diachronic perspective. In A. Alexiadou & M. Rathert (Eds.), *The semantics of nominalizations across languages and frameworks*. Berlin : Mouton de Gruyter, 215-244.

Ungerer, F., & Schmid, H.-J. (2006). *An introduction to cognitive linguistics* (2nd ed.). Harlow : Pearson Longman.

Van de Velde, D. (1995). *Le spectre nominal: Des noms de matières aux noms d'abstractions*. Louvain : Peeters.

Van Valen, L. (1965). Morphological Variation and Width of Ecological Niche. *The American Naturalist 99*, 377–390.

Vial, S. (2012). *La structure de la révolution numérique : Philosophie de la technologie.* Thèse de doctorat, Université René Descartes – Paris V.

Vial, S. (2016). Voir et percevoir à l'ère numérique : Théorie de l'ontophanie. In *Vivre par(mi) les écrans*. Dijon : Les Presses du Réel, 63-85.

Villoing, F. (2002). *Les mots composés [VN] N/A du français : réflexions épistémologiques et propositions d'analyse*. Thèse de doctorat, Université Paris Nanterre.

Villoing, F. (2009). Les mots composés VN. In B. Fradin, F. Kerleroux, & M. Plénat (Eds.), *Aperçus de morphologie du français*. Saint-Denis : Presses Universitaires de Vincennes, 175–198.

Villoing, F., & Gueye, M. (2024). La nominalisation néologique en *-ment* du français : structures morphologiques des bases verbales. Dans F. Neveu, S. Prévost, A. Montébran, A. Steuckardt et G. Bergounioux (Eds.). *SHS Web of Conferences* **191**, 08005, 1-16.

Wang, J. (2019). *Statistical Methods in Language Research with R* [基于 R 的语言学统计方法]. Beijing : Foreign Language Teaching and Research Press.

Wang, W., & Wu, J. (2020). *Lexical semantics* [词汇语义学]. Beijing : Foreign Language Teaching and Research Press.

Wauquier, M. (2020). *Confrontation des procédés dérivationnels et des catégories sémantiques dans les modèles distributionnels*. Thèse de doctorat, Université Toulouse 2 - Jean Jaurès.

Wilkinson, L. (2012). Exact and approximate area-proportional circular Venn and Euler diagrams. *IEEE Trans. Vis. Comput. Graph.*, 18(2), 321–331.

Xu, H. (2013). *Learning Statistics from Examples of Second Language Research* [第二语言研究中的统计案例分析]. Beijing : Foreign Language Teaching and Research Press.

Xu, J. (2023). *Research Methods in Corpus Linguistics* [语料库研究方法]. Beijing : Foreign Language Teaching and Research Press.

Zhang, W., & Dong, W. (2018). *Advanced SPSS Statistical Analysis Tutorial (3rd Edition* [SPSS 统计分析高级教程（第 3 版）]. Beijing : China Higher Education Press.