

Justine Salvadori

L'ambiguïté des noms déverbaux en français : une étude quantitative du sens construit



DOI : <https://doi.org/10.51363/unifr.lth.2025.044>

Justine Salvadori  0000-0002-4994-1675

© Justine Salvadori, 2025



Cet ouvrage est publié sous une licence Creative Commons Attribution 4.0 International (CC BY 4.0) : <https://creativecommons.org/licenses/by/4.0/>

L'ambiguïté des noms déverbaux en français

Une étude quantitative du sens construit

Thèse de Doctorat

présentée devant la Faculté des lettres
et des sciences humaines de
l'Université de Fribourg

par

Justine SALVADORI
originaire de Savièse (VS)

sous la direction du
Prof. Richard HUYGHE

Approuvé par la Faculté des lettres et des sciences humaines
sur proposition des professeurs
Richard HUYGHE (premier rapporteur)
Olivier BONAMI (deuxième rapporteur) et
Fiammetta NAMER (troisième rapporteure).

Fribourg, le 18 décembre 2024. Le Doyen Prof. Dominik SCHÖBI.

L'ambiguïté des noms déverbaux en français

Une étude quantitative du sens construit

Les noms dérivés de verbes (p. ex. *construction*, *couveuse*, *emballage*, *gravure*, *séchoir* en français) constituent un terrain d'exploration privilégié pour l'étude des relations entre forme et sens au sein du lexique. Le présent travail porte sur l'ambiguïté de ces noms, un champ de recherche comportant encore de nombreuses zones d'ombre. À ce jour, une description systématique et à grande échelle des configurations d'ambiguïté que présentent les nominalisations reste en effet à établir, et l'origine des cas observés demeure sujette à débat. La polysémie du nom *construction*, par exemple, qui peut désigner tant l'action de construire quelque chose que le résultat de cette action, soulève la question de savoir si le sens résultatif est le fait du verbe *construire* et de la suffixation en *-ion*, ou s'il découle d'une métonymie suivant un schéma de polysémie régulière de type événement → résultat. Par ailleurs, la possibilité d'une double motivation morphologique et sémantique conduisant à la construction de noms déverbaux ambigus n'est que très rarement envisagée.

Pour pallier ces lacunes, cette thèse propose une analyse sémantique approfondie de 5'274 noms déverbaux du français et des verbes apparentés. Noms et verbes sont analysés manuellement et décrits dans leur multiplicité sémantique selon une double classification (ontologique et relationnelle) permettant de distinguer entre ce qui relève de la nature des référents et de la relation que les noms entretiennent avec les bases verbales. Trois méthodes quantitatives sont ensuite appliquées aux données recueillies : l'analyse de règles d'association classiques, qui sert à mettre au jour des associations non aléatoires entre types sémantiques ; l'analyse de règles d'association séquentielles, qui rend possible l'identification de motifs d'association récurrents entre types sémantiques nominaux et propriétés morphosémantiques (procédés morphologiques et propriétés sémantiques des lexèmes-bases) ; et l'analyse de réseau, qui permet de décrire les types sémantiques en tant qu'entités interagissant au sein d'une structure globale. Notre étude se distingue ainsi par la combinaison inédite d'analyses qualitatives fines établies à partir de vastes échantillons représentatifs de l'ensemble du système étudié et de généralisations statistiques.

Les résultats mettent en lumière des configurations d'ambiguïté précédemment non décrites et montrent comment l'extension sémantique et la construction morphologique, en tant qu'outils de création de sens, participent à la construction de formes déverbaux ambiguës. Bien que ces mécanismes puissent fonctionner indépendamment l'un de l'autre, ils collaborent aussi fréquemment sous différentes formes de contributions conjointes. Sur le plan systémique, l'ascendant morphologique verbal paraît déterminant dans la structuration de l'ambiguïté des noms déverbaux, souvent articulée autour du domaine de l'action et des participants impliqués dans les procès. Dans le prolongement de ces apports, notre travail ouvre de nouvelles perspectives sur des thèmes connexes, tels que la compétition morphologique, la colexification et la parenté structurelle entre l'ambiguïté des mots et des affixes. Plus largement, cette recherche illustre la fécondité des études situées à l'interface de la sémantique lexicale et de la morphologie dérivationnelle.

The Ambiguity of Deverbal Nouns in French

A Quantitative Study of Constructed Meaning

Nouns derived from verbs (e.g., *construction* ‘building’, *couveuse* ‘brooder’/‘incubator’, *emballage* ‘packaging’, *gravure* ‘engraving’, *séchoir* ‘drying room’/‘airer’ in French) provide a fertile ground for exploring the relationships between form and meaning within the lexicon. This study is devoted to the ambiguity of these lexical units, a research area that is still largely unexplored. A systematic, large-scale description of the ambiguity configurations exhibited by nominalizations has yet to be established, and the origin of the observed cases remains a matter of debate. For instance, the polysemy of the noun *construction*, which can refer to both the action of building something and the result of this action, raises the question of whether the resultative meaning stems from the verb *construire* and the suffixation in *-ion*, or whether it results from a metonymic extension following a regular polysemy pattern (event → result). Moreover, the potential for a dual motivation account of ambiguity, involving both morphological and semantic factors, is rarely considered.

To address these issues, this thesis offers an in-depth semantic analysis of 5,274 French deverbal nouns and their related verbs. Nouns and verbs are manually analyzed and described using a dual ontological and relational classification that distinguishes between what pertains to the nature of the referents and the relationship that the nouns maintain with their bases. Three quantitative methods are applied to the collected data : classical association rule analysis, which reveals non-random associations between semantic types ; sequential association rule analysis, which identifies recurring association patterns between nominal semantic types and morphosemantic properties (morphological processes and the semantic properties of the base lexemes) ; and network analysis, which describes semantic types as entities interacting within a global structure. The research is unique in that it combines, in an unprecedented way, detailed qualitative analyses based on large, representative samples of the entire system under study and statistical generalizations.

The results highlight novel configurations of ambiguity and demonstrate how semantic extension and morphological construction, as tools to create meaning, contribute to the formation of ambiguous deverbal forms. While these mechanisms can operate independently, they also frequently interact in various ways. On a systemic level, the verbal morphological origin emerges as a key factor in structuring the ambiguity of deverbal nouns, often centered around the domain of action and its potential participants. Building on these contributions, my work offers new perspectives on such topics as morphological competition, colexification, and structural similarity between word and affix ambiguity. More broadly, this research illustrates the relevance of studies at the interface of lexical semantics and derivational morphology.

Remerciements

Alors que ce chapitre de ma vie se clôt, il m'est difficile de trouver les mots justes pour exprimer toute la reconnaissance que je ressens envers les nombreuses personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse.

Ma plus profonde gratitude va évidemment à Richard Huyghe. La patience, la gentillesse et la rigueur scientifique dont il a fait preuve durant ces dernières années m'ont été inestimables. Mille mercis de m'avoir guidée avec autant de bienveillance ; je n'aurais pu rêver meilleur directeur.

Je remercie aussi chaleureusement les membres de mon jury pour l'expertise et le temps consacrés à l'évaluation de ce travail. Leurs commentaires et suggestions ne manqueront pas d'améliorer considérablement mon manuscrit.

Cette thèse n'aurait jamais vu le jour sans l'engagement des chercheur·euses du projet « La sémantique des noms déverbaux », financé par le Fonds national suisse. Un grand merci à Lucie, Pauline, Richard, Alizée, Matthieu, Sandra, Delphine, Rossella et Marine pour leurs efforts inestimables.

Mon parcours doctoral a été grandement enrichi par l'accueil chaleureux de l'équipe du Département de français de l'Université de Fribourg. Un merci tout particulier à Suzanne pour ses promesses de fondue qui ont accompagné chaque chapitre rédigé, à Matthieu pour son humour et son art de la conversation, à Rossella pour nos discussions toujours bienvenues sur la gastronomie et les chats, à Alizée pour avoir accepté de danser sur de l'indie pop jusqu'au petit matin au milieu d'Anglais·es en furie, à Sandra pour m'avoir initiée aux joies de la programmation, ainsi qu'à Christelle, Gilles et Megan dont la gentillesse et la présence rassurante ont été des piliers tout au long de cette aventure.

Ma reconnaissance s'étend également à d'autres chercheur·euses que j'ai eu la chance de rencontrer à l'occasion d'écoles doctorales ou de conférences : Mathis, Katya, Baptiste, Donia, Cézane, Aylin, Simone, Céline, Linda, Loanne et Peter.

Parce que le diable est dans les détails, je tiens à remercier Suzanne, Matthieu, Bérangère et Valentin pour leur œil attentif lors de la relecture de ce manuscrit.

À mes ami·es, du Valais comme de l'Université, je dois des moments de répit précieux qui m'ont permis de garder la tête hors de l'eau ces quatre dernières années. Un merci spécial à Bérangère pour la vie de château, les gâteaux les plus délicieux de l'univers et nos messages vocaux/podcasts, à Valentin et ses impôts fédéraux, ainsi qu'à Judith, Marie C., André, Simren, Julia, Charles, Massimo, Hadrien, Marie G., Méline, Sarah et tous les membres de la galaxie Foucoupe, sans qui *Wuthering Heights* de Kate Bush ne résonnerait pas de la même manière.

Je n'oublie pas ma famille : Maria, Didier, Gérardo, Murielle, Estelle, Thomas, Albino, Alexandra, Césarine, Victor, Alissa, Denise, Sylvie, Roland et Crabi. Un merci tout particulier à mes grands-parents adorés, Dora, Albertine, Francis et Marcellino, dont les nombreux sacrifices m'ont permis d'entreprendre des études universitaires. Enfin, du fond du cœur, merci à Jacqueline, Silvano, Arnaud et Romeo pour leur amour et leur soutien inconditionnels. Cette thèse leur est dédiée.

Charmed as I am
by the sputter of bacon,
and the eye-opening properties
of eggs,
it's the coffee
that's really sacramental.

(Dorothea Grossman, «I allow myself», *Poetry*, Mars 2010)

Table des matières

Liste des figures	xiii
Liste des tableaux	xv
Liste des abréviations	xix
Types relationnels	xix
Types ontologiques	xx
Introduction	1
1 Contextualisation théorique	7
1.1 La sémantique des noms déverbaux	8
1.1.1 Pour une description systématisée de la diversité sémantique des nominalisations	9
1.1.2 Corrélats sémantiques de l'opération dérivationnelle	13
1.2 L'ambiguïté des noms déverbaux	17
1.2.1 Regard sémantique	21
1.2.2 Regard morphologique	27
1.3 Bilan	32
2 Méthodologie	35
2.1 Échantillonnage	36
2.1.1 Extraction automatique	36
2.1.2 Sélection manuelle	39
2.2 Principes d'analyse	42
2.2.1 Type sémantique des noms	42
2.2.2 Aspect lexical	48
2.2.3 Rôles sémantiques	50

2.2.4	Ambiguïté	53
2.3	Procédure d'analyse	56
2.3.1	Évaluation du modèle	57
2.3.2	Annotation	58
2.4	Bilan	59
3	Profil sémantique des noms déverbaux et des verbes de base	61
3.1	Informations générales	61
3.1.1	Noms	62
3.1.2	Verbes	68
3.2	Associations entre propriétés	74
3.2.1	Variables catégorielles	74
3.2.2	Loi sens-fréquence	79
3.3	Configurations d'ambiguïté	82
3.3.1	Bases multiples	83
3.3.2	Bases uniques	84
3.4	Bilan	89
4	Associations sémantiques régulières et réseau d'ambiguïté	91
4.1	Relations sémantiques observées	93
4.1.1	Types ontologiques	93
4.1.2	Types relationnels	96
4.1.3	Types combinés	98
4.2	Règles d'association	101
4.2.1	Extraction	102
4.2.2	Résultats et discussion	107
4.3	Réseau général	119
4.3.1	Constitution du réseau	120
4.3.2	Importance des types	122
4.3.3	Structures fréquentes	126
4.4	Bilan	131
5	L'ambiguïté nominale au prisme de la construction morphologique	133
5.1	Procédés morphologiques	134
5.1.1	Premières considérations	135

5.1.2	Étude de cas : les réseaux d'associations des procédés les plus poly-fonctionnels	142
5.2	Propriétés sémantiques des verbes de base	156
5.2.1	L'extraction de règles séquentielles	157
5.2.2	Aspect lexical	162
5.2.3	Rôles sémantiques des arguments	165
5.3	Collaboration entre attributs	169
5.3.1	Extraction des règles	169
5.3.2	Résultats et discussion	170
5.4	Bilan	175
	Conclusion	177
	Bibliographie	185

Liste des figures

1	Exemple de réseau représentant des associations orientées entre types sémantiques	5
1.1	Schémas constructionnels envisageables pour le nom <i>préchauffage</i>	8
1.2	Continuum d’ambiguïté lexicale	20
1.3	Trois configurations d’ambiguïté considérant l’appariement entre sens nominaux (N_X) et verbaux (V_X)	30
3.1	Distribution des noms annotés en fonction du nombre de sens qu’ils comportent	62
3.2	Fréquence d’association entre types ontologiques et relationnels en nombre de lexèmes	65
3.3	Traits aspectuels des noms d’éventualité annotés ($n = 4'162$)	67
3.4	Classe aspectuelle des noms d’éventualité annotés ($n = 4'162$)	67
3.5	Rôles sémantiques assignés par les lexèmes dotés d’une structure argumentale ($n = 4'452$)	69
3.6	Distribution des verbes annotés en fonction du nombre de sens qu’ils comportent	70
3.7	Traits aspectuels des verbes annotés	71
3.8	Classe aspectuelle des verbes annotés	71
3.9	Rôles sémantiques assignés par les verbes annotés	73
3.10	Associations entre propriétés nominales selon le τ de Goodman et Kruskal .	75
3.11	Associations entre propriétés verbales selon le τ de Goodman et Kruskal . .	78
3.12	Distribution des mots en fonction de leur fréquence dans le corpus FR-COW16A et du nombre de sens qu’ils comportent dans la base de données .	81
3.13	Configurations générales d’ambiguïté	82
3.14	Configurations d’ambiguïté à bases multiples	83
3.15	Configurations d’ambiguïté à base unique	86

3.16	Première forme de codérivation. V_1 , N_1 et N_2 entretiennent tous un lien conceptuel fort	87
3.17	Deuxième forme de codérivation. N_2 entretient un lien conceptuel clair avec N_1 mais incertain avec V_1	87
3.18	Troisième forme de codérivation. N_2 entretient un lien conceptuel plus fort avec V_1 qu'avec N_1	88
3.19	Quatrième forme de codérivation. Le lien conceptuel entre N_1 et N_2 est indirect	88
3.20	Sens événementiel N_0 mettant en relation N_1 et N_2	88
4.1	Exemple de réseau de règles d'association orientées entre types sémantiques	92
4.2	Associations bipartites entre types ontologiques	95
4.3	Associations bipartites entre types relationnels	97
4.4	Associations bipartites entre types combinés	100
4.5	Contributions convergentes. Une extension sémantique ($N_1 \rightarrow N_2$) et un procédé morphologique ($V_1 \rightarrow N_2$) sont appliqués simultanément pour former un même lexème	113
4.6	Contributions enchâssées. La construction morphologique et l'extension sémantique collaborent pour former deux lexèmes liés sémantiquement dont les sens sont conditionnés par la forme ($V_1 \rightarrow N_1 + N_2$)	113
4.7	Contributions cloisonnées. Deux procédés morphologiques sont appliqués de manière indépendante pour former deux lexèmes ($V_1 \rightarrow N_1$, $V_1 \rightarrow N_2$). Un lien sémantique entre les lexèmes ainsi formés peut ensuite se dessiner ($N_1 \rightarrow N_2$)	113
4.8	Réseau général d'associations de types sémantiques pour les noms déverbaux ambigus	121
4.9	Corrélations de rang (test τ de Kendall) entre les différentes mesures de centralité	124
4.10	Résultats d'un test CUG pour la centralisation de degré	127
4.11	Résultats d'un test CUG pour la centralisation de vecteur propre	127
4.12	Triades pouvant être observées au sein d'un réseau orienté	129
5.1	Types sémantiques combinés par procédé morphologique	136
5.2	Rapport entre richesse sémantique et nombre moyen de sens par vocable pour 34 procédés morphologiques	139
5.3	Réseaux d'associations obtenus pour les différents procédés morphologiques	149

Liste des tableaux

1.1	Alternances régulières instanciées par les noms déverbaux	23
2.1	Procédés morphologiques examinés	38
2.2	Nombre de paires verbe-nom par procédé morphologique dans la base de données finale	41
2.3	Types ontologiques	45
2.4	Types relationnels	47
2.5	Classes aspectuelles et combinaison de valeurs correspondant à chacune d'elles	51
2.6	Analyse sémantique complète du nom ambigu <i>vol</i> et du verbe apparenté <i>voler</i> . Trois sens nominaux et deux sens verbaux ont été distingués et appariés (<i>vol</i> ₁ / <i>voler</i> ₁ , <i>vol</i> ₂ / <i>voler</i> ₂ , <i>vol</i> ₃ / <i>voler</i> ₂)	55
2.7	Taux d'accord entre annotateur-ices par propriété	57
2.8	Deux échelles proposées pour l'interprétation du κ de Cohen	57
3.1	Répartition des noms et des sens en fonction du procédé morphologique .	64
3.2	Fréquence des noms et verbes analysés dans le corpus FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015)	79
3.3	Résultats des analyses de corrélation (test τ de Kendall) entre le nombre de sens et la fréquence des noms et verbes apparentés	80
3.4	Distribution des noms ambigus et sens apparentés en fonction des configurations d'ambiguïté	83
3.5	Distribution des paires verbe-nom où un seul lexème-base est lié à plusieurs lexèmes nominaux en fonction du nombre de sens qu'elles comportent . .	85
4.1	Associations complètes les plus fréquentes entre types ontologiques	94
4.2	Associations complètes les plus fréquentes entre types relationnels	97
4.3	Associations complètes les plus fréquentes entre types combinés	99
4.4	Extrait de la base de données d'origine	102

4.5	Matrice booléenne exprimant la (non-)réalisation de types sémantiques par certains vocables	103
4.6	Exemples de « transactions » pouvant être extraites de la base de données .	103
4.7	Propriétés des 42 règles extraites	105
4.8	Règles d'association finales	110
4.9	Taux de réalisation autonome des types sémantiques constituant l'antécédent ou le conséquent de règles d'association	115
4.10	Scores de centralité des types sémantiques examinés	124
4.11	Recensement de triades et résultats de tests CUG basés sur 2'000 répliques	130
5.1	Extrait de l'input (propriétés morphosémantiques) et de l'output (types sémantiques uniques ou ensembles de types sémantiques) d'une extraction de règles séquentielles	134
5.2	Analyse des différents procédés morphologiques en termes de nombre de vocables, nombre de sens, ambiguïté moyenne au niveau nominal et nombre de types sémantiques combinés réalisés au moins deux fois	139
5.3	Associations entre types sémantiques les plus fréquentes pour une sélection de procédés morphologiques	141
5.4	Synthèse de l'extraction des règles d'association pour une sélection de procédés morphologiques	144
5.5	Principales propriétés des réseaux constitués à partir des règles d'association extraites pour chaque procédé morphologique	148
5.6	Types sémantiques présentant les scores de centralité les plus élevés pour chaque procédé morphologique examiné	153
5.7	Centralisation de degré des différents procédés morphologiques	154
5.8	Centralisation de vecteur propre des différents procédés morphologiques .	154
5.9	Taux de réciprocité des différents procédés morphologiques	155
5.10	Extrait d'un tableau de données préparé pour l'extraction de règles séquentielles	160
5.11	Sélection de motifs fréquents extraits dans le cadre de l'analyse de l'ensemble des propriétés morphosémantiques (procédés morphologiques, aspect lexical et rôles sémantiques assignés par les verbes de base)	161
5.12	Extrait des règles séquentielles obtenues dans le cadre de l'analyse de l'ensemble des propriétés morphosémantiques (procédés morphologiques, aspect lexical et rôles sémantiques assignés par les verbes de base)	161
5.13	Sélection de règles extraites pour l'analyse des propriétés aspectuelles des bases verbales	164
5.14	Sélection de règles extraites pour l'analyse des rôles des bases verbales . . .	168

5.15	Synthèse des règles extraites pour chaque association sémantique	171
5.16	Sélection de règles extraites pour l'analyse des procédés morphologiques, des traits d'aspect lexical et des rôles des arguments	173
5.17	Propriétés morphosémantiques favorisant chaque association de types sé- mantiques	174

Liste des abréviations

Les différents types sémantiques utilisés dans le présent travail ont été initialement présentés en anglais (Huyghe, [2021](#); Salvadori et al., [2023](#)). Nous proposons ici une traduction française de ces catégories, tout en conservant les abréviations d'origine.

Types relationnels

Abréviation	Version originale en anglais	Traduction française
AGT	AGENT	AGENT
BEN	BENEFICIARY	BÉNÉFICIAIRE
CAU	CAUSE	CAUSE
DES	DESTINATION	DESTINATION
EXP	EXPERIENCER	EXPÉRIENCEUR
EXT	EXTENT	ÉTENDUE
FIG	FIGURATIVE	FIGURATIF
INS	INSTRUMENT	INSTRUMENT
LOC	LOCATION	LIEU
MAN	MANNER	MANIÈRE
PAT	PATIENT	PATIENT
PTH	PATH	CHEMIN
PVT	PIVOT	PIVOT
RES	RESULT	RÉSULTAT
SRC	SOURCE	SOURCE
STI	STIMULUS	STIMULUS
THM	THEME	THÈME
TPC	TOPIC	TOPIQUE
TSP	TRANSPOSITION	TRANSPOSITION

Types ontologiques

Abréviation	Version originale en anglais	Traduction française
Anm	Animate	Entité animée
Art	Artefact	Artéfact
Art*Cog	Artefact*Cognitive	Artéfact*Cognitif
Art*Ist	Artefact*Institution	Artéfact*Institution
Cog	Cognitive	Objet cognitif
Cog*Evt	Cognitive*Event	Événement*Cognitif
Coll	Collective	Collectif
Dis	Disease	Maladie
Dom	Domain	Domaine
Evt	Event	Événement
Evt*Fin	Event*Financial	Événement*Financier
Evt*Nat	Event*Natural	Événement*Naturel
Evt*Phn	Event*Phenomenon	Événement*Phénomène
Evt*Sta	Event*State	Événement*État
Fin	Financial	Objet financier
Ins	Institution	Institution
N/A	N/A	N/A
Nat	Natural	Objet naturel
Phn	Phenomenon	Phénomène
Ppt	Property	Propriété
Qua	Quantity	Quantité
Sta	State	État
Tim	Time	Temps

Introduction

La nominalisation déverbale, à en croire Steven Pinker, consisterait essentiellement à prendre un verbe bien sous tous rapports et à lui ajouter un suffixe pour en faire un nom terne et sans vie : « The nominalization rule takes a perfectly spry verb and embalms it into a lifeless noun by adding a suffix like *-ance*, *-ment*, *-ation*, or *-ing* » (Pinker, 2015, p. 50). Rien n'est pourtant moins sûr. L'ampleur et la variété des études s'intéressant à la sémantique des noms déverbaux dans diverses langues (Booij, 1986; Namer & Villoing, 2008; F. Martin, 2010; Fradin, 2012, 2016; Lieber, 2016; Varvara et al., 2021; Lieber & Plag, 2022; Plag et al., 2023; Kawaletz, 2023; *inter alia*) attestent au contraire du fait que ces dérivés constituent un terrain d'exploration particulièrement fertile pour quiconque souhaite sonder les relations entre forme et sens dans le lexique.

Le présent travail porte sur l'ambiguïté de ces unités lexicales, définie *a priori* comme le fait qu'elles puissent avoir différents sens¹. L'ambiguïté des noms déverbaux présente en effet des régularités frappantes tant sur le plan morphologique que sémantique, un phénomène qu'illustrent bien les trois exemples suivants :

- (1) a. *construire*_V 'fabriquer' → *construction*_N 'action de construire' / 'objet construit'
- b. *créer*_V 'donner l'existence' → *création*_N 'action de créer' / 'objet créé'
- c. *réaliser*_V 'concevoir' → *réalisation*_N 'action de réaliser' / 'objet réalisé'

Les nominalisations *construction*, *création* et *réalisation* sont toutes formées avec le suffixe *-ion*, comportent deux lectures à première vue unies par une métonymie, et investissent qui plus est des catégories de sens identiques (événement et résultat).

Motivation

La littérature s'efforce depuis longtemps de lever le voile sur la forme et l'origine de telles configurations d'ambiguïté dans le lexique déverbal. En sémantique lexicale, différents travaux (Pustejovsky, 1995; Ježek, 2008; Barque et al., 2014; *inter alia*) ont notamment porté sur les patrons de polysémie régulière (cf. Apresjan, 1974) instanciés par les nominalisations :

1. Une mise au point théorique sur les notions liées à l'ambiguïté lexicale (p. ex. polysémie, homonymie, métaphore, métonymie) est proposée dans le chapitre 1.

Introduction

événement/résultat comme le montrent les exemples ci-dessus, mais aussi agent/instrument (*accordeur, découpeuse*), événement/agent (*direction, gouvernement*), événement/lieu (*passage, croisement*) ou encore instrument/lieu (*séchoir, saloir*). Dans le domaine de la morphologie, ce sont principalement la variété des catégories sémantiques associées aux affixes nominalisateurs (*-ion*, mais aussi *-age, -ment* ou *-eur* en français) et le rôle tenu par la sémantique des bases qui ont fait l'objet d'un intérêt particulier (Scott, 2010; Melloni, 2011; Müller, 2011; Ferret & Villoing, 2015; Fradin, 2016; Lieber, 2016; *inter alia*).

Il reste que, malgré ces efforts, de nombreuses zones d'ombre subsistent. Les études existantes n'offrent par la force des choses qu'une vision fragmentaire du phénomène, n'étant souvent dévolues qu'à quelques patrons de polysémie régulière (p. ex. événement/résultat, agent/instrument) ou à une poignée de procédés morphologiques² (p. ex. *-age, -ment* et *-ion* en français). Il s'ensuit qu'une description systématique et à large échelle de l'ambiguïté des noms déverbaux dans une langue comme le français manque encore. Par ailleurs, aucun consensus n'a pour l'heure été atteint s'agissant de l'origine des nominalisations ambiguës. Dans les exemples en (1) ci-dessus, les sens résultatifs constituent-ils par exemple des extensions métonymiques des interprétations actionnelles, ou sont-ils plutôt le fait des verbes et du suffixe *-ion*? Force est de constater que les études confrontant sans ambages ces deux hypothèses – sémantique et morphologique – restent peu nombreuses (mais voir p. ex. Ferret & Villoing, 2015; Bauer, 2017; Huyghe & Lombard, 2022). La possibilité d'une double motivation morphologique et sémantique à la construction de noms déverbaux ambigus n'est en outre que très rarement envisagée.

Objectifs

Deux objectifs principaux se dessinent pour nous à l'aune de ces différents éléments. Nous souhaitons d'une part proposer une description étendue et systématique de l'ambiguïté des noms déverbaux du français en synchronie, prenant en compte l'ensemble des procédés morphologiques employés pour la nominalisation de verbes ainsi que les configurations d'appariement entre sens verbaux et nominaux. D'autre part, nous visons à examiner la contribution respective et la combinaison des procédés d'extension sémantique (métonymie, métaphore) et de construction morphologique (procédés morphologiques, contraintes sémantiques sur les bases verbales) dans la formation des noms déverbaux ambigus.

De manière plus précise, il s'agira :

- (i) de constituer une grille d'analyse sémantique adaptée aux nominalisations et à leurs bases;
- (ii) de l'appliquer de manière rigoureuse à un large échantillon de noms déverbaux et de verbes apparentés représentant l'ensemble des procédés nominalisateurs du français;

2. Nous regroupons sous l'expression « procédé morphologique » les suffixes et les formes de conversion.

- (iii) d'examiner de quelle manière lexèmes nominaux et lexèmes verbaux s'associent au sein de configurations d'ambiguïté, si l'on part du principe que le processus de dérivation s'applique à des éléments spécifiés sémantiquement (cf. Mel'čuk, 1993; Fradin & Kerleroux, 2003; Fradin, 2023);
- (iv) de quantifier et de décrire les différentes associations de types sémantiques investies par les noms déverbaux, qu'elles soient bien établies ou encore inconnues de la littérature;
- (v) de déterminer de quelle manière les principes cognitifs généraux façonnant le lexique nominal ambigu (cf. Xu et al., 2020; Karjus et al., 2021; Brochhagen & Boleda, 2022; *inter alia*) s'accommodent des idiosyncrasies du lexique déverbal;
- (vi) d'évaluer dans quelle mesure et de quelle façon la construction morphologique (propriétés sémantiques des bases et procédés morphologiques) peut conditionner l'ambiguïté des noms déverbaux.

Plusieurs retombées sont envisageables pour les domaines de la sémantique lexicale et de la morphologie. Sur le plan méthodologique, ce travail propose une approche intégrée qui conjugue une description manuelle fine des interprétations des dérivés avec des généralisations statistiques mettant en jeu des méthodes d'analyse quantitative pas ou peu employées en linguistique. Ce genre d'approche duale reste rare dans les études sur la sémantique de la dérivation – à notre connaissance du moins. D'un point de vue théorique, cette étude aspire à offrir de nouvelles perspectives sur l'ambiguïté des déverbaux, ainsi que sur les noms ambigus et les noms dérivés de verbes de manière plus générale. Elle peut en corollaire faire avancer la compréhension de phénomènes connexes, notamment : la compétition morphologique (Gardani et al., 2019; Nagano et al., 2024; *inter alia*), s'il s'avère que les différents procédés sont associés à des configurations d'ambiguïté distinctes; la colexification (François, 2008; List, 2023; *inter alia*), en faisant intervenir la question du rôle de la construction morphologique dans les associations de sens régulières; ou encore la relation *one form-many meanings* (Plag et al., 2023; Haspelmath, 2023; *inter alia*), en comparant ambiguïté nominale et polyfonctionnalité affixale.

Moyens d'investigation

Ce travail s'inscrit dans le cadre d'un projet de recherche financé par le Fonds national suisse et portant sur la sémantique des noms déverbaux³. Fondé sur l'analyse sémantique approfondie d'une grande quantité de nominalisations ($n = 5'274$) et de verbes apparentés, il met à profit une **double classification sémantique** – ontologique (p. ex. Événement, Propriété, Artéfact, Entité animée) et relationnelle (p. ex. TRANSPOSITION, RÉSULTAT, INSTRUMENT, AGENT) – permettant de distinguer entre ce qui relève respectivement de la nature des

3. Un descriptif est disponible à l'adresse <https://data.snf.ch/grants/grant/188782>. Le projet a impliqué les chercheur-euses suivant-es (par ordre alphabétique) : Lucie Barque, Pauline Haas, Richard Huyghe, Alizée Lombard, Matthieu Monney, Justine Salvadori, Sandra Schwab, Delphine Tribout, Rossella Varvara et Marine Wauquier.

Introduction

référents et de la relation que les noms entretiennent avec les bases. Comme le montrent les exemples suivants, un nom déverbal qui dénote un artéfact peut par exemple exprimer différentes relations par rapport à l'éventualité dénotée par le verbe de base (2), tandis qu'un nom déverbal qui exprime le résultat d'une éventualité peut dénoter différents types de référents (3) :

- (2) a. *bâtir* → *bâtiment* [Artéfact-RÉSULTAT]
b. *raser* → *rasoir* [Artéfact-INSTRUMENT]
c. *garer* → *garage* [Artéfact-LIEU]
- (3) a. *bâtir* → *bâtiment* [Artéfact-RÉSULTAT]
b. *créer* → *créature* [Entité animée-RÉSULTAT]
c. *énervé* → *énervement* [État-RÉSULTAT]

Les types sémantiques ainsi identifiés tiennent lieu de base de travail pour trois formes d'analyse quantitatives, dont certaines sont à notre connaissance très peu employées en linguistique fondamentale. Proposée à l'origine pour l'étude de paniers d'achats en marketing, l'**analyse de règles d'association** (Agrawal et al., 1993) sert dans le cadre de cette étude à la détection d'associations orientées et non aléatoires entre types sémantiques. Par exemple, la règle d'association suivante indique que si un nom instancie le type Artéfact-INSTRUMENT, il y a de fortes chances qu'il investisse également le type Entité animée-AGENT (p. ex. *accordeur*, *arroiseur*, *découpeuse*) :

- (4) {Artéfact-INSTRUMENT} ⇒ {Entité animée-AGENT}

Une **version séquentielle des règles d'association** (Agrawal & Srikant, 1995) nous permet quant à elle de mettre au jour des relations logiques et non aléatoires entre attributs morphosémantiques (procédés morphologiques, propriétés sémantiques des bases verbales) et catégories de sens dans le domaine nominal. La règle suivante indique à titre d'exemple qu'un verbe requérant un sujet agentif et décrivant une action qui dure dans le temps, lorsqu'il est allié au suffixe *-age*, peut donner lieu à un nom ambigu réalisant conjointement les types Événement-TRANSPOSITION et Artéfact-RÉSULTAT (p. ex. *gribouiller*/*gribouillage*, *griffonner*/*griffonnage*, *mouler*/*moulage*) :

- (5) {Verbe à sujet agentif, Verbe duratif, Suffixe *-age*} ⇒ {Événement-TRANSPOSITION, Artéfact-RÉSULTAT}

Enfin, différents outils d'**analyse de réseau** (cf. Wasserman & Faust, 1994; Newman, 2018; Zweig, 2016) sont mobilisés pour examiner comment l'ambiguïté des noms déverbaux se structure d'un point de vue systémique. À titre d'illustration, le schéma suivant représente des types sémantiques (p. ex. Événement-TRANSPOSITION, Artéfact-RÉSULTAT) dont d'autres dépendent plus ou moins régulièrement, et inversement, des catégories (p. ex.

Artéfact-INSTRUMENT, Propriété-RÉSULTAT) dont la présence est conditionnée plus ou moins fortement à celle d'autres types.

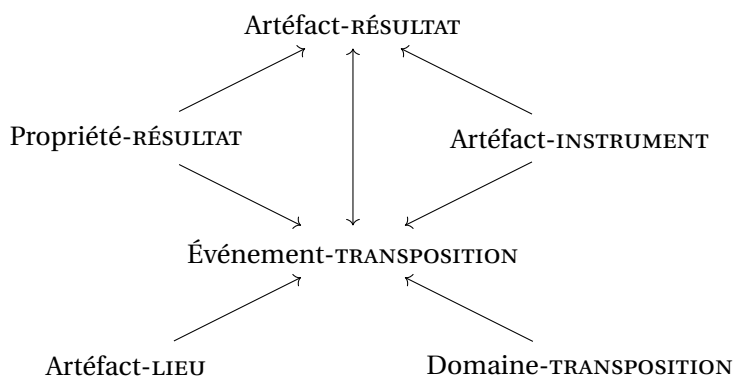


FIGURE 1 – Exemple de réseau représentant des associations orientées entre types sémantiques

Plan du travail

Notre travail est structuré comme suit.

Le **chapitre 1** pose les bases théoriques guidant le reste de l'étude. Abordant différents sujets clés de la littérature sur la sémantique des noms déverbaux, il rappelle que la question de la place qu'occupent respectivement la construction morphologique et l'extension sémantique dans la formation de nominalisations ambiguës n'a toujours pas trouvé réponse. La nécessité d'une description à la fois qualitative et quantitative de l'ambiguïté des noms déverbaux en français est également mise en exergue. Une telle description permettrait non seulement de vérifier certains postulats théoriques sur la sémantique des nominalisations, mais pourrait aussi faire émerger des phénomènes jusque-là négligés.

Le **chapitre 2**, à visée méthodologique, décrit la collecte et l'analyse sémantique fine d'un très large échantillon de nominalisations et de verbes apparentés. Les propriétés considérées comprennent les catégories de sens investies par les noms, l'aspect lexical et les rôles sémantiques des arguments des verbes. Une attention particulière est accordée à la manière dont l'ambiguïté est traitée dans les domaines verbal et nominal, et la procédure d'analyse suivie est décrite en détail (p. ex. calcul de scores d'accord entre annotateur-ices).

Le **chapitre 3** établit le profil sémantique des noms déverbaux et des verbes apparentés qui ont été analysés. Il décrit leur comportement sémantique en fonction de différentes propriétés (nombre de sens, types sémantiques, aspect lexical, rôles sémantiques), met en évidence l'existence de relations significatives entre ces dernières, vérifie l'application de la loi sens-fréquence (Zipf, 1945) aux données observées, puis examine de quelle façon lexèmes verbaux et lexèmes nominaux peuvent s'associer les uns aux autres en cas d'ambiguïté.

Introduction

Le **chapitre 4** s'attelle à l'examen des facteurs sémantiques œuvrant à la construction de noms déverbaux ambigus. Sur la base d'une analyse de règles d'association (Agrawal et al., 1993), il identifie et quantifie dans un premier temps différentes associations orientées et non aléatoires entre types sémantiques pouvant s'apparenter à des patrons de polysémie régulière (cf. Apresjan, 1974; Barque, 2008). Il propose dans un second temps une analyse de réseau (cf. Wasserman & Faust, 1994; Zweig, 2016; Newman, 2018) fondée sur les règles d'association précédemment identifiées et visant à une compréhension holistique du rapport entre forme et sens dans le lexique déverbal.

Le **chapitre 5** examine les ressorts et l'ampleur de la contribution morphologique à l'ambiguïté des déverbaux. Il s'attache dans un premier temps à analyser la seule implication des procédés morphologiques. Dans ce but, une étude de cas portant sur les plus polyfonctionnels d'entre eux (-age, -ment, -ion, -erie, -ure, conversion sur thèmes 0 et 12) et fondée sur la même méthodologie que celle employée dans le chapitre 4 est proposée. Le chapitre aborde par la suite l'implication des propriétés sémantiques des bases dans l'ambiguïté de leurs dérivés grâce à une analyse de règles séquentielles (Agrawal & Srikant, 1995). Enfin, il propose un examen de la contribution conjointe des procédés morphologiques et des propriétés sémantiques des verbes, lui aussi basé sur des règles d'association séquentielles.

La **conclusion** synthétise nos principaux résultats, propose un commentaire général, revient sur les limites de cette étude, et avance enfin plusieurs pistes de recherche pouvant alimenter de futurs travaux.

Disponibilité des données

La base de données exploitée dans ce travail, le guide d'annotation employé pour la description sémantique des verbes et des noms (Salvadori et al., 2023), ainsi que les scripts R⁴ constitués pour les analyses statistiques sont accessibles à l'adresse https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2.

4. Les analyses statistiques ont été réalisées avec R version 4.4.0 (R Core Team, 2024).

Chapitre 1

Contextualisation théorique

On admet volontiers que *rasoir*, *déménageur* ou encore *accouchement* constituent des noms « déverbaux » en français. Il est en effet entendu que ces derniers sont construits à partir de verbes, moyennant l'intervention d'affixes nominalisateurs tels que *-oir*, *-eur* et *-ment* :

- (6) a. *raser* → *rasoir*
- b. *déménager* → *déménageur*
- c. *accoucher* → *accouchement*

Ces trois exemples illustrent bien une perspective répandue en morphologie dérivationnelle, qui veut que tout lexème complexe soit dérivé d'une seule « base », c'est-à-dire d'un autre lexème à la fois conceptuellement antérieur et formellement plus simple (cf. Iacobini, 2000; Stump, 2019; Bonami & Guzmán Naranjo, 2023; Copot & Bonami, 2024; *inter alia*). Aussi le bien nommé nom « déverbal » *rasoir* ne peut avoir d'autre origine que le verbe *raser* selon cette conception.

Bien qu'elle paraisse ainsi relativement intuitive de premier abord, la notion de nom « déverbal » touche pourtant à un certain nombre de problèmes, dont les tenants et aboutissants hantent la littérature consacrée depuis plusieurs décennies déjà. Que faire de *préchauffage* en (1.1), par exemple, un nom d'apparence déverbale que deux bases — l'une effectivement verbale, l'autre nominale — peuvent motiver en raison de sa double affixation? À ce défi épineux que constitue la multimotivation (Corbin, 1976; Hathout & Namer, 2014b; Strnadová, 2014; *inter alia*) s'ajoute celui de l'orientation de la relation entre bases et dérivés. Le cas des rétroformations (Becker, 1993; Štekauer, 2015; *inter alia*) est à cet égard bien connu. Namer (2012), par exemple, s'est penchée sur les verbes de forme [YX]_V (p. ex. *neurostimuler*, *thermocoller*) et les noms apparentés de forme [YXSfx]_N (p. ex. *neurostimulation*, *thermocollage*), où Y symbolise un composant nominal d'origine grecque ou latine, X un composant verbal, et Sfx un suffixe déverbal permettant de former des noms d'événement ou d'état. Ses résultats montrent que les noms examinés ne sont pas dérivés de verbes, mais résultent

1.1. La sémantique des noms déverbaux

plutôt d'une composition savante. Les verbes [YX]_V sont ensuite créés à partir de ces noms par un processus de rétroformation.

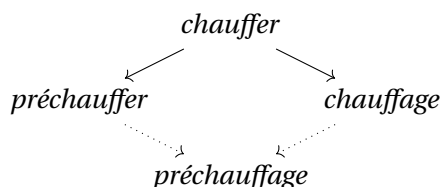


FIGURE 1.1 – Schémas constructionnels envisageables pour le nom *préchauffage*

Face aux problèmes relevés, différentes approches théoriques sont envisagées. Certaines cherchent à amender le modèle binaire orienté (cf. Booij, 2019; Jackendoff & Audring, 2020; *inter alia*), tandis que d'autres préconisent des modèles paradigmatiques (Bauer, 1997; Štekauer, 2014; Bonami & Strnadová, 2019; Namer & Hathout, 2020; *inter alia*). L'objectif de ce travail n'est pas de discuter en profondeur des cadres théoriques proposés, qu'ils soient d'obédience lexématique ou paradigmatique. Il n'est pas non plus de revenir en détail sur des points chauds qui ont déjà fait couler beaucoup d'encre, si l'on considère du reste que l'importance statistique des cas de multimotivation ou de possible rétroformation reste difficile à évaluer. Notre intérêt pour l'ambiguïté dérivationnelle, plus modeste, repose fondamentalement sur les relations sémantiques qu'entretiennent verbes et noms apparentés. Parce qu'il est toujours nécessaire de définir une base de travail, on se contentera donc d'appeler « déverbal » tout nom suffixé à partir d'un verbe, auquel il est lié sémantiquement en synchronie. Dans le cadre de cette étude, sont donc déverbaux *rasoir*, *déménageur* et *accouchement*, mais aussi *préchauffage*, *neurostimulation* et *thermocollage*. Une opérationnalisation de cette définition – tout compte fait assez inclusive – sera proposée lors de la présentation de la méthode d'analyse dans le chapitre 2. Dans l'intervalle, ce chapitre servira à exposer une sélection d'éléments théoriques qui aiguilleront le reste de notre étude. Nous reviendrons dans un premier temps sur diverses questions liées à la sémantique des noms déverbaux en général (section 1.1), puis à leur polysémie plus spécifiquement (section 1.2).

1.1 La sémantique des noms déverbaux

La sémantique n'a pas toujours constitué une priorité pour les linguistes s'intéressant aux nominalisations. À la suite des travaux fondateurs de Lees (1960) et de Chomsky (1970), la recherche s'est en effet longtemps concentrée sur les propriétés syntaxiques de ces noms, et en particulier sur leur structure argumentale (Grimshaw, 1990; Alexiadou, 2001; Borer, 2003; *inter alia*). Plusieurs études à l'interface de la sémantique lexicale et de la morphologie dans les décennies 1980, 1990 et 2000 (Booij, 1986; Beard, 1990, 1995; Temple, 1996; Jurafsky, 1996; Gaeta, 2000; Lieber, 2004; Namer & Villoing, 2008; *inter alia*) ont toutefois ouvert la voie à des travaux plus centrés sur la question du sens des noms dérivés de verbes. Différents axes de

recherche ont ainsi été développés. Parmi eux, relevons le transfert de propriétés sémantiques verbales à travers la nominalisation (Haas et al., 2008; Fábregas & Marín, 2012; *inter alia*), la variété sémantique des dérivés et leur ambiguïté (Ježek, 2008; Melloni, 2011; Bauer et al., 2013; *inter alia*), ou encore la concurrence sémantique entre suffixes nominalisateurs (Kelling, 2001; F. Martin, 2010; *inter alia*). La diversité des orientations théoriques peut également être soulignée, qu'il s'agisse du *Lexical Semantic Framework* développé par Lieber (2004, 2016), des cartes sémantiques de Schulte (2015a, 2015b), de l'approche *frame-based* qu'exploite Kawaletz (2023), ou du modèle analogique exploré par Plag et ses collègues (2023).

Détour nécessaire pour pouvoir aborder librement la question de l'ambiguïté des nominalisations dans la prochaine section, cette partie se concentre sur les principes généraux d'analyse du sens des noms déverbaux à travers deux aspects fondamentaux : leur mode de description sémantique (section 1.1.1) et leur relation aux bases verbales (section 1.1.2).

1.1.1 Pour une description systématisée de la diversité sémantique des nominalisations

On doit aux études syntaxiques mentionnées ci-dessus une distinction fondamentale traversant encore aujourd'hui les recherches de tous bords sur les nominalisations : celle entre dérivés dits « événementiels » d'une part, et dérivés dits « référentiels » (ou « résultatifs »)⁵ d'autre part. Sur le plan syntaxique, les premiers (*exécution*, *énervement*) se distinguent des seconds (*arrosoir*, *promeneur*) par le fait qu'ils peuvent accepter un programme argumental similaire à celui de leur verbe de base. Dans ce qui suit, le verbe *exécuter* et son dérivé *exécution* impliquent par exemple tous deux un agent et un patient :

- (7) a. *Le soldat **exécute** le prisonnier.*
b. *l'**exécution** du prisonnier par le soldat*

Au plan sémantique, les dérivés « événementiels » constituent en principe des éventualités au sens de Bach (1981), décrivant le même type de situation que leur base (p. ex. l'action d'exécuter dans le cas d'*exécution*) et possédant de ce fait des propriétés temporelles. Les dérivés « référentiels » renvoient généralement à un élément lié à la situation décrite par le verbe (p. ex. l'instrument servant à arroser dans le cas du nom *arrosoir*).

Les études morphosémantiques se sont depuis longtemps emparées de cette dichotomie, cherchant à l'affiner d'un point de vue sémantique et à distinguer ainsi plusieurs formes de dérivés événementiels (p. ex. action, état) et référentiels (p. ex. instrument, résultat). Différentes méthodes ont pour ce faire été proposées dans la littérature. Les approches lexicographiques, bien que couvrantes, peuvent manquer de systématisation et ne sont sans

5. La tradition anglo-saxonne parle respectivement d'*event* et de *referential* (ou *result*) *nominals*. Voir Ježek (2008, p. 253) et Melloni (2011, p. 5-6) pour une discussion des problèmes terminologiques associés à la catégorie référentielle.

1.1. La sémantique des noms déverbaux

doute pas les plus prisées dans les travaux sur les nominalisations. Une exception notable pour le français est l'étude de Jacquy (2013), qui développe une méthodologie reposant sur l'extraction et la synthétisation du contenu de définitions du *Trésor de la Langue Française informatisé* (ATILF - CNRS & Université de Lorraine, s. d.). Plus populaires, les outils de sémantique distributionnelle tels que *Word2Vec* (Mikolov et al., 2013), *GloVe* (Pennington et al., 2014) et *BERT* (Devlin et al., 2019) sont mobilisés depuis plusieurs années en morphologie dérivationnelle (Marelli & Baroni, 2015; Padó et al., 2016; Wauquier, 2020; Varvara et al., 2021; Denistia et al., 2022; Stupak & Baayen, 2022; Bonami & Guzmán Naranjo, 2023; Guzmán Naranjo & Bonami, 2023; *inter alia*). Ces outils, fondés sur des vecteurs employés comme des représentations de sens enregistrant les contextes dans lesquels un mot donné apparaît, offrent des avantages indéniables : automatisation et rapidité du processus, travail sur une large quantité de données, et objectivation des analyses. Toutefois, ils ne permettent pas de décrire directement et d'identifier un contenu sémantique; seulement de comparer les capacités sémantiques de différents mots sur la base de leurs distributions. Certains modèles peinent en outre à décrire la polysémie et l'homonymie (Wauquier, 2022).

Nous choisissons par conséquent une troisième option pour cette étude, le typage sémantique, qui consiste à attribuer des catégories de sens à grain plus ou moins fin (ACTION, AGENT, INSTRUMENT, LIEU, etc.) à des mots donnés suivant les traits sémantiques qu'ils possèdent (Dixon, 2005, p. 7; Haas et al., 2023, p. 54). L'attribution des différentes classes repose généralement sur des définitions et/ou des tests linguistiques distributionnels. Par exemple, on peut considérer que *déménageur*, *surveillant* et *forgeron* appartiennent à la classe des noms d'AGENT en français, car ils dénotent des entités animées « réalisant intentionnellement l'action décrite par [leur] verbe de base » (Huyghe & Tribout, 2015, p. 102), respectivement *déménager*, *surveiller* et *forger*. Contrairement à *rasoir*, *tatouage* et *énervement*, ils peuvent par ailleurs être insérés dans les structures syntaxiques en (8) et (9) sans pour autant générer d'impression d'agrammaticalité (Huyghe & Tribout, 2015, p. 103) :

- (8) Det N_{eur} {a décidé / a choisi} de Vb (NP₁)
Le {déménageur / surveillant / forgeron} a décidé de manger son déjeuner.
*Le {rasoir / tatouage / énervement} a décidé de manger son déjeuner.
- (9) Det N_{eur} (de NP) Vb {volontairement / délibérément / consciemment / intentionnellement} (NP₁)
Le {déménageur / surveillant / forgeron} a volontairement renversé cette chaise.
*Le {rasoir / tatouage / énervement} a volontairement renversé cette chaise.

Le typage sémantique peut être réalisé manuellement ou de manière (semi-)automatique. Si le typage (semi-)automatique présente des avantages indéniables en termes de gain de temps et de traitement de grandes quantités de données, la granularité de l'analyse reste par la force des choses relativement large, et certaines distinctions sémantiques peuvent dès lors être difficiles à capturer. Dans la base lexicale morphologique du français Démonette-1.2 (Hathout & Namer, 2014a), les types sémantiques sont par exemple calculés à partir des

propriétés morphologiques des mots. Les noms masculins suffixés en *-eur* reçoivent ainsi l'étiquette @AGM désignant un agent masculin ou un instrument, deux types sémantiques courants pour ces dérivés. Cette étiquette générale ne permet cependant pas de distinguer les cas d'ambiguïté où les deux types, agent et instrument, sont applicables (p. ex. *accordeur* 'personne qui accorde' / 'instrument servant à accorder') des cas de monosémie où seul l'un des deux types est pertinent (p. ex. agent pour *cambricoleur* 'malfaiteur', instrument pour *aspirateur* 'appareil électroménager'). En raison de ce genre de limite, nous privilégierons dans ce travail le typage sémantique manuel qui, bien que chronophage, permet une analyse plus fine des sens des noms dérivés et des configurations d'ambiguïté qu'ils instancient.

Le choix de la typologie à appliquer reste crucial. De nombreuses classifications ont été proposées pour diverses langues, souvent dans le cadre d'un effort plus large de caractérisation de la sémantique de la dérivation (Lieber, 2004; Schulte, 2015a, 2015b; *inter alia*). Nous identifions deux écoles principales dans la littérature. Une première série d'études et de ressources recourent d'une part à des catégories ontologiques plus ou moins larges, établies sur la base d'une partition du réel de référence (voir p. ex. Huguin et al., 2023 pour la base de données dérivationnelle du français Démonette-2 de Namer et al., 2023; Lieber, 2004 et Schulte, 2015a, 2015b pour l'anglais; Ježek, 2008 et Melloni, 2011 pour l'italien). Sont ici mobilisées à la fois des distinctions générales éprouvées en sémantique lexicale comme CONCRET *vs* ABSTRAIT (*rasoir vs devinette*) (cf. Van de Velde, 1995; Flaux & Van de Velde, 2000; *inter alia*) ou SITUATION *vs* ENTITÉ (*confrontation vs marinade*) (cf. Godard & Jayez, 1996; Huyghe, 2015; *inter alia*), mais aussi des sous-catégories plus fines telles que QUANTITÉ (*pin-cée*), DROIT (*essartage*), GROUPE HUMAIN (*immigration*), OBJET INFORMATIONNEL (*préférence*), etc. L'attribution des catégories se fait optionnellement par l'examen de traits sémantiques distinctifs. Le *Lexical Semantic Framework* de Lieber (2004, p. 23-24), par exemple, fait appel aux oppositions [\pm matériel] et [\pm dynamique] pour distinguer respectivement entre noms abstraits et noms concrets d'une part, et entre noms d'événement et noms d'état d'autre part. Un autre volet de travaux cherche quant à eux à caractériser les noms à travers la relation qu'ils entretiennent avec les verbes de base (voir p. ex. Fradin, 2012 pour le français; Bauer et al., 2013, Kawaletz & Plag, 2015, Lieber, 2016 et Kawaletz, 2023 pour l'anglais), en sollicitant essentiellement des catégories s'apparentant à des rôles sémantiques comme THÈME (*roulotte*), AGENT (*chercheur*), INSTRUMENT (*rasoir*), RÉSULTAT (*enjolivure*) ou MANIÈRE (*allure*). Comme le relève Laca (2001, p. 1222), l'identification se fait souvent au moyen de formules paraphrastiques contenant *a minima* une variable correspondant à la base et permettant de cette façon d'explicitier le rapport entre verbe et nom. En français, l'interprétation de LIEU peut ainsi être assignée au nom *plantation*, car celui-ci exprime le « lieu où X V » – littéralement le lieu où l'on plante quelque chose (Fradin, 2012, p. 2).

On peut s'interroger sur les enseignements à tirer de cet aperçu. Nonobstant les cadres théoriques variables, nous suivons ici Buljan (2018, p. 189) lorsqu'elle note que si certaines catégories reviennent régulièrement (QUALITÉ, ÉTAT, CONDITION, ACTION, etc.), on ne trouve que peu de cohérence dans leur emploi d'une étude à l'autre. L'étiquette COLLECTIF constitue à ce titre un exemple parlant. Alors que Lieber (2016, p. 18-19) la met sur un pied d'égalité

1.1. La sémantique des noms déverbaux

avec la catégorie ABSTRAIT dans sa typologie, Schulte (2015b, p. 44) considère que les noms collectifs peuvent être associés à des entités concrètes comme abstraites, et Fradin (2012) lui fait correspondre la glose ‘personnes qui V’ (p. ex. *administration*). On peut par conséquent formuler une série de recommandations dont nous tenterons de tenir compte dans le cadre de cette étude (cf. chapitre 2).

Distinction entre plans d’analyse Comme le relèvent Laca (2001, p. 1222) et Huyghe (2021, p. 21), la description du sens des mots construits devrait théoriquement s’articuler autour de deux plans d’analyse distincts : (i) ce qui relève d’une part de la *sémantique lexicale*, i.e. la sémantique des noms conventionnalisés dans le lexique, qu’ils soient simples ou construits; (ii) et ce qui se rapporte d’autre part à la *sémantique de la dérivation*, i.e. les opérations sémantiques associées aux procédés morphologiques. Dans un effort de clarté, cette étude s’appuiera donc sur une double typologie, réunissant au passage les deux écoles descriptives mentionnées ci-dessus : une typologie ontologique, reposant sur la nature des référents désignés et pouvant en principe s’appliquer à n’importe quel item du lexique nominal (p. ex. *artéfact*, *entité animée*, *état*); et une typologie relationnelle, servant à caractériser les relations entre bases verbales et dérivés (p. ex. *résultat*, *instrument*, *lieu*).

Classifications générales Même si certaines catégories restent très marginales, l’analyse sémantique des dérivés devrait si possible reposer sur une classification générale des noms, de façon à garantir l’application la plus étendue qui soit. Nous privilégierons par conséquent des typologies aspirant à une couverture lexicale complète, en nous basant sur l’étude de Haas et ses collègues (2023) pour le plan d’analyse ontologique, et les listes de rôles sémantiques présentées par Kipper-Schuler (2005) (*VerbNet*) et Petukhova et Bunt (2019) (*LIRICS*) pour le plan d’analyse relationnel.

Grain descriptif fin Un problème parfois relevé dans la littérature (Lieber, 2004, p. 4; Buljan, 2018, p. 189; Kawaletz, 2023, p. 45) tient au choix de la granularité de la typologie à employer. La question du degré de spécification sémantique des procédés dérivationnels n’étant toujours pas résolue, nous opterons par précaution pour un grain d’analyse relativement fin et conforme au niveau de description habituellement choisi dans les études de morphologie dérivationnelle.

Critères transparents pour l’attribution des catégories Enfin, d’un point de vue pratique, Kawaletz (2023, p. 45) constate à raison que certaines catégories sémantiques sont souvent mobilisées sans être définies au préalable, ce qui peut entraîner des malentendus et des interprétations erronées. C’est pourquoi nous chercherons à éviter toute étiquette sémantique « impressionniste » (Plag et al., 2023, p. 105), pour au contraire proposer des définitions et des critères linguistiques explicites. Une attention particulière sera par ailleurs portée à la validation empirique du schéma d’analyse, via une campagne d’annotation test en double aveugle avec évaluation de l’accord entre annotateur-ices.

1.1.2 Corrélats sémantiques de l'opération dérivationnelle

Les schémas descriptifs proposés par Fradin (2012), Lieber (2016) et Kawaletz (2023) que nous venons d'évoquer présentent l'avantage de caractériser les liens sémantiques qu'entretiennent verbes et noms. En faisant un pas de côté, on peut à partir de là vouloir se questionner plus largement sur les corrélats sémantiques de l'opération dérivationnelle ou, en d'autres termes, chercher à déterminer ce qui se passe « sémantiquement parlant » lorsqu'un verbe se voit transformé en nom.

Rappelons pour commencer que le passage d'une catégorie grammaticale à une autre implique nécessairement des différences structurelles entre sémantique de la base et sémantique du dérivé. Dans le domaine de l'action, verbes et noms événementiels – qui possèdent à la différence des dérivés référentiels des propriétés temporelles – sont ainsi pourvus d'ensembles de traits aspectuels distincts⁶. Selon Huyghe (2011), seuls les seconds sont notamment à même de décrire l'opposition entre actions occurrentielles (p. ex. *cambriolage, manifestation*) et non occurrentielles (p. ex. *jonglage, natation*), suivant leur caractère respectivement comptable ou massif (cf. Langacker, 1991; Van de Velde, 1995; Kleiber, 1997; Nicolas, 2002; *inter alia*). L'ambiguïté lexicale, objet de ce travail, se manifeste aussi différemment dans les domaines verbal et nominal. Gentner (1981) relève par exemple que les verbes sont plus ambigus que les noms, quel que soit leur niveau de fréquence. S'agissant des figures sémantiques en jeu, plusieurs travaux s'accordent également sur le fait que la métaphore et la métonymie n'interviennent pas de la même manière dans les différentes parties du discours. King et Gentner (2022) ont montré dans une étude récente que les verbes (p. ex. *fleurir*) étaient plus susceptibles que les noms d'étendre leur signification à l'aide de métaphores (10), et que les noms (p. ex. *construction*) se prêtaient à l'inverse plus facilement que les verbes à la métonymie (11). Vicente (2018, p. 958) partage cette perspective, affirmant que la métonymie affecte principalement les noms, tandis que les extensions métaphoriques s'appliquent non seulement aux noms, mais aussi à d'autres catégories grammaticales (verbes, adjectifs, prépositions).

- (10) a. *Des milliers de roses **fleurissent** dans la ville.* (web)
- b. *Les idées **fleurissent** dans votre esprit.* (web)
- (11) a. *La **construction** de la nouvelle mosquée a duré près d'une décennie.* (web)
- b. *La **construction** mesure plus de deux mètres de haut.* (web)

Au-delà de ces différences structurelles, on peut s'interroger sur la contribution sémantique globale du processus constructionnel en cas de dérivation. Assiste-t-on en effet à un ajout, à une suppression et/ou à un maintien de la matière sémantique entre mots de base et mots dérivés? Selon certains travaux, l'adjonction au lexème-base d'un élément formel

6. Nous faisons la distinction entre (i) aspect « lexical » – aussi appelé « *Aktionsart* » depuis Agrell (1908) – qui correspond aux traits aspectuels inhérents des verbes et des noms, et (ii) aspect « grammatical », qui s'exprime par des variations de temps verbal.

1.1. La sémantique des noms déverbaux

se traduit naturellement par un apport sémantique. La *Monotonicity Hypothesis* formulée par Koontz-Garboden (2007) (voir aussi Kiparsky, 1982) indique ainsi que les opérations dérivationnelles peuvent ajouter du sens, mais n'en suppriment jamais. Bauer et ses collègues (2013, p. 29) suivent un chemin similaire en affirmant que, contrairement à la flexion, « la dérivation enrichit le sens de la base ⁷ ». Un exemple bien connu, cité par Haspelmath et Sims (2013), est celui des noms d'arbre fruitier dérivés de noms de fruit en français (12), où la suffixation en *-ier* coïncide avec une complexification sémantique impliquant à la fois une idée de production et un rapport de contenant à contenu (cf. Roché, 1998, p. 59) :

- (12) a. *cerise* → *cerisier*
b. *pomme* → *pommier*
c. *prune* → *prunier*

D'autres auteur·ices restent plus circonspect·es. Anderson (1994, p. 7), notamment, soutient que certaines formes dérivées peuvent se défaire d'une partie du contenu sémantique de leur base, et donne comme exemple de non-monotonie une série de verbes suffixés en *-st* en islandais (13), où la « couche causative » originelle disparaît à travers le processus dérivationnel :

- (13) a. *kvelja_V* 'torturer' → *kveljast_V* 'souffrir'
b. *heyra_V* 'entendre' → *heyrast_V* 'être entendu'
c. *gleðja_V* 'rendre heureux' → *gleðjast_V* 'se réjouir'

On peut enfin envisager, comme Beard (s. d.), la possibilité du *status quo*, ou de la simple « transposition ». Selon lui, la nominalisation désadjectivale générant en anglais les mots *importance* 'importance' et *creativity* 'créativité' n'ajoute aucun contenu sémantique ni ne modifie une quelconque caractéristique inhérente de l'adjectif sous-jacent (respectivement *important* 'important' et *creative* 'créatif') : elle le transforme « simplement » en nom.

Ces dissonances dans l'analyse incitent naturellement à se pencher d'un peu plus près sur les propriétés sémantiques qui seraient, sinon ajoutées ou supprimées, du moins maintenues à travers la nominalisation. Il convient une fois encore de rappeler ici la distinction fondamentale existant entre dérivés événementiels et dérivés référentiels mentionnée dans la section précédente. Comme nous le présentons dans les sous-points suivants, les deux groupes ne donnent en effet pas lieu aux mêmes analyses du fait de leur rapport différent aux bases verbales, les premiers étant considérés comme sémantiquement plus proches des verbes que les seconds (Roché, 2009; Wauquier, 2020; *inter alia*).

7. [En anglais dans le texte] « Inflection does not add significantly to the meaning of the base, but contextualizes that meaning; derivation adds to the meaning of the base ».

Noms événementiels

Les études portant sur le transfert de propriétés sémantiques entre verbes et noms événementiels s'intéressent principalement à (i) l'aspect lexical (ou *Aktionsart*) et à (ii) l'attribution des rôles sémantiques.

S'agissant du point (i), on présuppose souvent que les nominalisations dénotant des éventualités conservent les valeurs aspectuelles de leur base, qui correspondent en règle générale aux oppositions [\pm dynamique], [\pm délimité] et [\pm duratif] (Haas et al., 2008). Cette idée a été explicitement formulée par Fábregas et ses collègues (2012, p. 162) sous la forme de l'*Aspect Preservation Hypothesis*. Une campagne d'annotation en corpus à large échelle (Balvet et al., 2011) a toutefois montré que verbes et noms événementiels dérivés pouvaient diverger de ce point de vue. Par exemple, le verbe *imaginer* (14) peut être considéré comme dynamique, car il s'accommode bien de la périphrase *être en train de*. Son dérivé *imagination* (15a), en revanche, tend plutôt vers la stativité. Contrairement à *atterrissage* (15b) et *éboulement* (15c), qui dénotent des actions dynamiques, il accepte en effet difficilement d'être employé comme sujet des verbes *se produire* et *avoir lieu* (Haas, 2009).

- (14) *Avec les changements climatiques, l'Isère est en train « d'imaginer ce que peut être la station de ski de demain. »* (web)
- (15) a. **L'imagination de la station de ski de demain {a eu lieu/s'est produite} ce matin.*
 b. *Certains pensent que l'atterrissage a eu lieu sur la terre ferme.* (web)
 c. *L'éboulement s'est produit à 3500 mètres, dans un secteur peu fréquenté par les alpinistes.* (web)

Au regard du point (ii), le postulat de base met également en avant le maintien de la structure argumentale à travers la nominalisation. Considérant le nombre d'arguments que possèdent verbes et noms apparentés, Condette et ses collègues (2012, p. 849) vérifient par exemple largement leur hypothèse de départ, à savoir que la structure argumentale des noms est « héritée ou dérivée de celle des verbes de base ». Seul un nombre très marginal de décalages est en effet relevé dans leurs données (à peine 7 % des 339 paires verbe-nom étudiées). Comme pour l'aspect lexical, néanmoins, ce tableau peut encore être nuancé dans la mesure où les rôles sémantiques associés aux verbes et noms apparentés diffèrent parfois. Le nom *glissade* implique par exemple nécessairement un effectuateur animé contrairement à sa base *glisser* :

- (16) *En pleine séance, son pull a glissé sur ses épaules.* (web)
- (17) a. **En pleine séance, son pull a fait une glissade sur ses épaules.*
 b. *Il a volé le tabouret de la maîtresse et il a fait une glissade à plat ventre dessus! C'est interdit!* (web)

1.1. La sémantique des noms déverbaux

Noms référentiels

Examinons maintenant les noms référentiels. Leur intérêt central consiste plutôt cette fois dans la focalisation d'arguments syntaxiques et/ou sémantiques du verbe.

D'un point de vue syntaxique d'abord, il est généralement admis que les dérivés référentiels peuvent lexicaliser le sujet, l'objet ou l'oblique de leur verbe de base. Alors que Fradin (2003) parle d'argument « distingué », la littérature anglo-saxonne (Booij, 1986; Booij & Lieber, 2004; Bauer et al., 2013; Lieber, 2016; *inter alia*) recourt quant à elle aux expressions *subject-* vs. *object-oriented* ou encore *subject-* vs. *object-referencing* pour désigner des noms lexicalisant respectivement un argument externe ou interne. On trouve facilement des exemples de ces différents types en français. Un *catalyseur* est quelque chose qui *catalyse* (sujet), une *sucette* est quelque chose que l'on *suce* (objet), et un *légataire* est quelqu'un à qui on *lègue* quelque chose (oblique). Pour ce qui concerne les arguments sémantiques, Plag et ses collègues (2018, p. 469) font remarquer que « les nominalisations déverbales peuvent non seulement lexicaliser l'événement dénoté par le verbe ou les arguments syntaxiques du verbe, mais aussi d'autres entités qui font partie de la représentation sémantique du verbe de base ⁸ ». Les dérivés dénotant des instruments, par exemple, sont de fait souvent apparentés à des verbes requérant un participant instrumental dans leur représentation sémantique (Plag et al., 2023, p. 124) : une *binette* demeure toujours bien utile pour *biner*, comme l'est une *muselière* pour *museler*, ou encore une *écumoire* pour *écumer*.

Toujours est-il que, comme les dérivés événementiels, les nominalisations référentielles peuvent présenter des écarts sémantiques plus ou moins forts par rapport à leur base. Certains dérivés ne semblent ainsi correspondre à aucun argument sémantique ou syntaxique du verbe apparenté : la nominalisation locative *buvette*, notamment, ne trouve pas d'équivalent parmi les participants impliqués dans le sémantisme du verbe *boire*, qui requiert tout au plus un sujet agentif et un objet patientif. Un autre cas de figure souvent relevé tient à une perte du sens compositionnel attendu. Pour l'allemand, Marzo (2015, p. 984) (voir aussi Barz, 1982) oppose ainsi *Zuschauer* 'spectateur' à *Bäcker* 'boulangier', respectivement dérivés de *zuschauen* 'regarder' et de *backen* 'faire cuire (au four)'. Tandis que le premier nom peut référer à n'importe quelle personne effectuant l'action dénotée par *zuschauen*, le second comporte selon elle une dimension professionnelle absente de sa base. Ce genre de décalage peut être relevé en français avec la paire *casser/casseur*, où le dérivé renvoie nécessairement à un effectuateur animé et intentionnel (18a), contrairement au sujet du lexème-base (18b) :

- (18) a. *Inspirés par les émeutes en France, des **casseurs** s'attaquent au Flon.* (web)
b. *La tempête **a cassé** et déraciné des arbres, les projetant sur la route et sur des bâtiments, et arraché des pylônes et des toits de bâtiments.* (web)

8. [En anglais dans le texte] « deverbal nominalizations can not only lexicalize the event denoted by the verb or the verb's syntactic arguments, but also other entities that are part of the semantic representation of the base verb ».

On retiendra de cet aperçu que le transfert de sens entre verbes et noms constitue un processus complexe et non linéaire. De fait, il paraît naturel de s'interroger sur les facteurs sous-tendant les variations sémantiques observées. La lexicalisation est souvent invoquée pour expliquer ces phénomènes, notamment en raison de sa propension à engendrer des effets de démotivation et d'opacification (Corbin, 1991; Brinton & Traugott, 2005; *inter alia*). Comme le fait remarquer Aronoff (1976, p. 18), « les mots, une fois formés, persistent et changent; ils développent des idiosyncrasies, avec pour résultat qu'ils ne peuvent bientôt plus être générés par un simple algorithme de quelque généralité que ce soit⁹ ». La question se pose dès lors de savoir comment neutraliser les potentiels effets de la lexicalisation dans l'analyse sémantique des mots construits. Une parade employée dans de nombreux travaux consiste à étudier les néologismes (Plag et al., 2018; Dal et al., 2018; *inter alia*), lesquels offrent en principe la possibilité d'approcher les processus de dérivation en synchronie sans se soucier d'éventuels effets annexes. Bien que l'on puisse se demander si certains néologismes ne dévient pas d'emblée des patrons dérivationnels, particulièrement en présence d'un impératif onomasiologique important, il est raisonnable de supposer qu'ils restent dans l'ensemble plus transparents que les mots attestés de longue date. Une autre solution, plus holistique, peut être de chercher à identifier des régularités sémantiques au sein des mots construits, qu'ils soient néologiques ou non, et de considérer qu'il y a potentiellement un patron dérivationnel des cas de construction observés à partir d'une certaine récurrence. Nous retenons cette seconde option pour le présent travail.

1.2 L'ambiguïté des noms déverbaux

Nous venons de le voir, les travaux sur la sémantique des noms déverbaux s'intéressent volontiers à la construction du sens nominal à partir du sens verbal. Considérons maintenant un autre pan de recherche ayant fait couler beaucoup d'encre : celui de l'ambiguïté de ces noms. Il est en effet bien connu que de nombreuses nominalisations peuvent se voir attribuer plusieurs interprétations (Rainer, 1996, 2014; Melloni, 2011; Lieber, 2018; *inter alia*). Les dérivés *accordeur* (19), *maquillage* (20) et *rédaction* (21) illustrent parfaitement ce point :

- (19) a. AGENT : *L'accordeur est venu, mardi, s'occuper de l'orgue de Sablé pour que l'instrument sonne juste pour la messe de Noël.* (web)
- b. INSTRUMENT : *Vous souhaitez acheter un **accordeur** pour votre guitare, votre basse, votre ukulélé ou un autre type d'instrument de musique?* (web)
- (20) a. ÉVÉNEMENT : *Comment bien préparer la peau avant de procéder au **maquillage**?* (web)
- b. INSTRUMENT : *Quel est le meilleur endroit pour acheter du **maquillage** bio?* (web)

9. [En anglais dans le texte] « Words, once formed, persist and change; they take on idiosyncrasies, with the result that they are soon no longer generable by a simple algorithm of any generality ».

1.2. L'ambiguïté des noms déverbaux

- c. RÉSULTAT : *Tandis que le cortège nuptial s'avance, le héraut efface son **maquillage** de bouffon.* (web)
- (21) a. ÉVÉNEMENT : *Quelles précautions faut-il prendre lors de la **rédaction** de la clause bénéficiaire?* (web)
- b. RÉSULTAT : *Tom a déchiré la **rédaction** de sa sœur.* (web)
- c. AGENT : *La **rédaction** a décidé de donner la parole à plusieurs activistes issus de différentes régions de Suisse romande.* (web)
- d. LIEU : *Donald Trump s'est rendu à la **rédaction** du New York Times, le 22 novembre 2016.* (web)

Cet état de choses conduit à s'interroger à la fois sur la forme et l'origine de l'ambiguïté des nominalisations. Quelles sont les catégories sémantiques mobilisées et comment celles-ci s'associent-elles? À quel point l'ambiguïté des déverbaux est-elle comparable à celle observée dans le lexique simple? Comment se construit cette ambiguïté et quelle est l'implication des procédés morphologiques et des bases? De même, quelle importance attribuer aux différents types d'extension de sens (métaphore, métonymie, etc.), dans la mesure où ils peuvent *a priori* exercer leur influence sur l'ensemble du lexique nominal?

Ces questions – qui ne sont certainement pas nouvelles – ont été abordées suivant plusieurs angles de recherche dans la littérature. Les sémanticien·nes, d'une part, ont étudié les types d'ambiguïté que peuvent manifester les nominalisations, ainsi que les associations régulières observées entre leurs différents sens (voir p. ex. Pustejovsky, 1995 pour l'anglais; Ježek, 2008 pour l'italien; Barque et al., 2014 pour le français). Les morphologues, d'autre part, ont sondé la multiplicité des significations associées aux affixes déverbaux et le rôle joué par la sémantique des bases (voir p. ex. Müller, 2011 pour l'allemand; Lieber, 2016 pour l'anglais; Fradin, 2016; Ferret & Villoing, 2015 pour le français). Cette section cherchera à faire le point sur ces deux approches, en examinant la problématique de l'ambiguïté des noms déverbaux à travers un prisme sémantique (sous-section 1.2.1), puis morphologique (sous-section 1.2.2).

Mais avant toute chose, un bref détour terminologique inspiré des travaux de Bréal (1897), Kleiber (1999), Ravin et Leacock (2000), Nerlich et Clarke (2003), Falkum et Vicente (2015), Vicente et Falkum (2017) et Valera (2020). Nous utiliserons dans ce qui suit le terme « ambiguïté » pour référer à la propriété d'un mot d'avoir plusieurs significations distinctes¹⁰. Le dérivé *peinture*, par exemple, peut dénoter une action (22a), un instrument (22b) ou un résultat (22c) :

- (22) a. *Nous avons procédé à la **peinture** de la partie supérieure des murs, en tôle, en appliquant l'enduit thermoplastique.* (web)

10. On distingue habituellement plusieurs types d'ambiguïté : ambiguïté de mot *vs* ambiguïté de construction syntaxique (p. ex. *Il a avoué son allergie à la presse*, cf. Fuchs, 1996); et ambiguïté de mot en langue *vs* ambiguïté de mot en discours (p. ex. *Le sandwich au jambon est assis à la table 20*, cf. Nunberg, 1979). Nous nous intéressons ici à l'ambiguïté de mot en langue.

- b. *Mathilde a acheté de la **peinture** jaune et a repeint sa chambre.* (web)
- c. *On a vendu une **peinture** de Mai-Thu 537 000 euros en 2021.* (web)

À l'inverse, les noms *boursicoteur* et *crachin* ne semblent posséder qu'un seul sens : 'personne qui boursicote' (23a) et 'pluie fine' (23b). Le *Petit Robert* (s. d.-a) et le *Wiktionnaire* (s. d.) ne comportent au reste qu'une seule entrée pour ces deux mots¹¹.

- (23) a. *40 % des **boursicoteurs** interrogés se disent optimistes pour l'avenir de l'économie française.* (web)
- b. *Ce matin-là, un **crachin** tenace a dissuadé les visiteurs d'arpenter le cimetière national de Gettysburg.* (web)

L'ambiguïté englobe à la fois la polysémie et l'homonymie : les sens examinés sont liés sémantiquement dans la première situation, mais ne le sont pas dans la seconde. En français, on peut par exemple estimer que, contrairement à *peinture* en (22), le nom déverbal *carotteur* constitue un homonyme. Ses deux sens – 'personne qui escroque' (24a) et 'instrument servant à prélever un échantillon cylindrique' (24b) – sont en effet relativement distincts :

- (24) a. *Un des **carotteurs** témoigne qu'il aurait réussi à vendre une plaquette de Kit Kat pour 460 euros.* (web)
- b. *Réaliser l'ouverture d'installation selon l'épaisseur de la paroi avec une fraise couronne en diamant de 82 mm ou avec un **carotteur**.* (web)

S'agissant de la polysémie, il convient encore de distinguer entre les différents types de relation sémantique en jeu, en principe au nombre de quatre (cf. Geeraerts, 2017) : métaphore et métonymie d'une part; et extension et restriction de sens d'autre part. Les polysèmes construits par métonymie, comme *peinture* en (22), reposent sur une relation de contiguïté référentielle. Ceux construits par métaphore, comme *lacet* en (25), instancient pour leur part une relation d'analogie :

- (25) a. *Le sportif attache ses **lacets** pour se préparer à l'entraînement.* (web)
- b. *À la sortie du village de Vissoie, le premier **lacet** de la route qui mène à Saint-Luc doit faire l'objet d'une réfection.* (web)

Extension et restriction de sens se distinguent quant à elles selon que l'interprétation seconde est respectivement plus générale ou plus spécifique que le sens premier (cf. Barque, 2008).

11. De manière plus générale, nous laisserons ici de côté les variations de sens s'apparentant à ce que la tradition anglo-saxonne nomme *vagueness* (cf. Tuggy, 1993; M. L. Murphy, 2010; *inter alia*), i.e. des fluctuations sémantiques qui ne peuvent être interprétées comme des significations distinctes et qui s'apparentent ainsi plus à de la monosémie qu'à de l'ambiguïté à proprement parler. Des exemples souvent cités en anglais sont des noms non spécifiés du point de vue du genre, tels que *neighbor* 'voisin-e', *cousin* 'cousin-e' ou encore *teacher* 'enseignant-e'.

1.2. L'ambiguïté des noms déverbaux

Le verbe *adopter* en (26), par exemple, témoigne d'une extension de sens, la lecture initiale 'prendre pour enfant' (26a) s'étendant à 'accorder une protection' (26b) :

- (26) a. "On ne s'en remet jamais": ils s'aperçoivent 33 ans plus tard qu'ils **ont adopté** un bébé volé. (web)
- b. La salle a cartonné, les gens l'**ont** tout de suite **adoptée** et il s'est passé des choses extraordinaires. (web)

Une restriction de sens peut à l'inverse être observée avec le verbe *s'abstenir* en (27), où le sens premier 'ne pas faire une action' (27a) se restreint à 'ne pas voter' (27b) :

- (27) a. *L'expert **s'est abstenu** de donner sa préférence pour l'un ou l'autre des modèles.*
(web)
- b. *Sept pays ont voté contre et 32 **se sont abstenus**.* (web)

Soulignons enfin qu'à l'heure actuelle, monosémie, polysémie et homonymie sont rarement envisagées comme des phénomènes discrets. Comme l'illustre le schéma 1.2, plusieurs recherches soutiennent en effet que la monosémie et l'homonymie constituent les extrémités d'un continuum sur lequel viennent se placer polysémies métaphorique et métonymique (Lipka, 1990; Tuggy, 1993; Cruse, 1995; Klepousniotou et al., 2012; Yurchenko et al., 2020; *inter alia*). Les raisons en sont multiples. Au-delà des difficultés que peut poser l'application de tests permettant de distinguer l'une ou l'autre catégorie (cf. Geeraerts, 1993), on peut parfois s'interroger sur la présence d'un lien (p. ex. *jalousie* 'sentiment'/'persienne'), ou à l'inverse questionner l'existence d'une différence réelle entre deux interprétations proches (p. ex. *amour* 'sentiment amoureux'/'amitié') (Barque & Haas, 2022, p. 6). L'impulsion donnée ces dernières décennies par des travaux en psycholinguistique venus étayer, mais aussi affiner la distinction entre les différents types d'ambiguïté (Frisson & Pickering, 1999; Klepousniotou et al., 2012; Lopukhina et al., 2018; Lombard et al., 2023; *inter alia*) n'est pas non plus à négliger de ce point de vue. Le chapitre 2 reviendra plus en détail sur les critères de différenciation appliqués dans le cadre de cette étude.

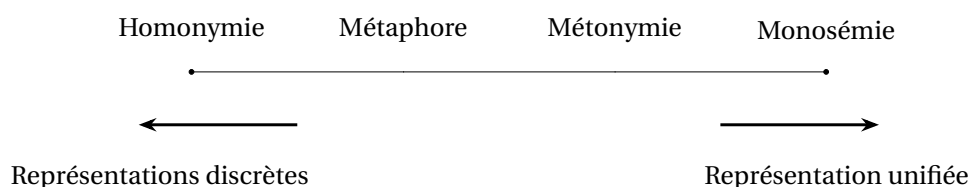


FIGURE 1.2 – Continuum d’ambiguïté lexicale

1.2.1 Regard sémantique

Nous considérons dans ce qui suit deux sujets de recherche abordés dans les travaux de sémantique sur l'ambiguïté des noms déverbaux, à savoir les alternances de sens régulières que ces noms réalisent et les différentes formes d'ambiguïté pouvant être observées.

Alternances régulières

Nous avons vu précédemment (cf. section 1.1.1) que certaines interprétations (ACTION, AGENT, INSTRUMENT, LIEU, etc.) sont récurrentes dans le lexique déverbal. Beaucoup de ces lectures se retrouvent aussi fréquemment associées au sein de dérivés donnés. Un certain nombre de noms déverbaux en français, comme *couveuse*, *calculatrice* ou encore *navigateur*, peuvent par exemple se voir attribuer les types sémantiques AGENT et INSTRUMENT :

- (28) a. AGENT : *La femelle pond deux ou trois fois dans l'année (avec à chaque ponte entre dix et quinze œufs); c'est une bonne **couveuse**.* (web)
b. INSTRUMENT : *Charles et Alexandre, en DUT Science et Génie des Matériaux, ont construit une **couveuse** à poussins pour les élèves âgés de cinq et six ans d'une maternelle de Virey-le-Grand.* (web)
- (29) a. AGENT : *Celle qui est évidemment la mère (et divine déesse) de Lucifer est aussi une **calculatrice** hors pair, qui ne perd pas une seconde pour semer le chaos (tout en finesse) une fois arrivée sur Terre [...].* (web)
b. INSTRUMENT : *Je tourne ce problème dans tous les sens, et à part une envie de casser ma **calculatrice**, je n'arrive à rien de valable!* (web)
- (30) a. AGENT : *Parti de Saint-Malo (France), en 1535, le **navigateur** arriva dans le Golfe après une traversée de presque deux mois.* (web)
b. INSTRUMENT : *Si vous cherchez un **navigateur** pour remplacer Chrome, plusieurs options s'offrent à vous.* (web)

Les déverbaux ne sont de toute évidence pas les seuls noms du lexique à présenter ce genre de comportement. Les noms dénotant un animal, par exemple, peuvent aussi servir à référer à la viande qui en est tirée (p. ex. *veau*, *agneau*, *saumon*), tout comme les noms renvoyant à un contenant peuvent être employés pour désigner la quantité mesurée au moyen de cet objet (p. ex. *tasse*, *verre*, *bol*) (Dölling, 2020). Tous ces cas correspondent en réalité à un phénomène bien identifié dans les études de sémantique lexicale : celui de la polysémie régulière. Apresjan (1974, p. 16) le définit comme suit dans son étude pionnière :

« La polysémie d'un mot A comportant les sens a_i et a_j peut être considérée comme régulière si, dans la langue en question, il existe au moins un autre mot B

1.2. L'ambiguïté des noms déverbaux

comportant les sens b_i et b_j qui se distinguent sémantiquement l'un de l'autre de la même manière que a_i et a_j , et si a_i et b_i , et a_j et b_j ne sont pas synonymes¹² ».

Quelques commentaires à ce sujet. La polysémie régulière s'oppose naturellement à la polysémie irrégulière (ou idiosyncrasique), qui s'observe lorsque la distinction sémantique qui tient entre deux sens d'un mot donné ne se trouve nulle part ailleurs dans le lexique (p. ex. *planteur* 'personne qui plante quelque chose' / 'cocktail à base de rhum'). On identifie à l'opposé des alternances dites « systématiques », où le fait pour un nom de présenter un sens X implique nécessairement de posséder un sens Y (p. ex. l'alternance *CONTENANT*/*CONTENU* mentionnée plus haut). Entre ces deux extrêmes, la littérature s'accorde sur le fait que les alternances peuvent correspondre à des degrés divers de régularité (cf. Barque & Haas, 2022), les facteurs à l'origine de cette gradabilité pouvant être discutés. Apresjan (1974) estime par exemple que les différences de régularité observées ne sont pas étrangères aux mécanismes d'extension sémantique, dans la mesure où la métonymie est selon lui plus régulière que la métaphore. On trouve la même idée chez Vicente (2018, p. 958), qui explique cette situation par le fait que les différents sens de polysèmes métaphoriques entretiennent une relation de similitude, et que « la similitude dépend en grande partie de la façon dont l'observateur la perçoit¹³ ». Nous suivons toutefois Lombard et ses collègues (2023) dans ce travail, pour qui figures sémantiques et régularité polysémique doivent être décorréliées en tant que plans d'analyse. En effet, l'hypothèse d'une corrélation entre figure et régularité n'a pas été démontrée de façon conclusive, en grande partie parce que l'évaluation du degré de régularité d'un patron reste difficile à mettre en œuvre pratiquement.

Un phénomène linguistique intéressant dont les ramifications touchent à la polysémie régulière est celui de la « colexification » (cf. François, 2008), qui se manifeste « lorsque plusieurs concepts sont exprimés dans une langue avec le même mot¹⁴ » (Di Natale et al., 2021, p. 99). Là où le français et l'anglais emploient par exemple des mots différents pour les concepts *DOIGT* et *ORTEIL*, l'espagnol les colexifie sous un seul terme, *dedo*. Les études sur la colexification paraissent particulièrement éclairantes pour l'analyse des schémas d'ambiguïté régulière instanciés par les nominalisations, dans la mesure où elles cherchent à comprendre pourquoi certaines associations conceptuelles sont plus fréquentes que d'autres dans les langues du monde, et à lever ainsi le voile sur les mécanismes cognitifs traversant l'ambiguïté lexicale. Xu et ses collègues (2020) considèrent par exemple que la variation interlinguistique dans la fréquence de colexification n'est pas arbitraire, mais reflète un principe général d'économie cognitive : les concepts les plus souvent colexifiés sont ceux dont la mise en relation nécessite un effort cognitif moindre. Leurs résultats suggèrent ainsi

12. [En anglais dans le texte] « Polysemy of the word A with the meanings a_i and a_j is called regular if, in the given language, there exists at least one other word B with the meanings b_i and b_j , which are semantically distinguished from each other in exactly the same way as a_i and a_j and if a_i and b_i , a_j and b_j are nonsynonymous ».

13. [En anglais dans le texte] « Metaphor-based polysemy is typically *irregular* (in the Apresjan's sense) and rather *idiosyncratic*. The relationship between the different senses is one of similarity, and similarity is in a good part up to the eye of the beholder ».

14. [En anglais dans le texte] « Colexification is a linguistic phenomenon that occurs when multiple concepts are expressed in a language with the same word. »

que l'associativité conceptuelle constitue une contrainte importante dans le développement du lexique. Plus récemment, Brochhagen et Boleda (2022) ont examiné l'interaction entre la pression cognitive vers la simplicité linguistique et la pression concurrente pour une transmission précise de l'information. Leurs résultats soutiennent l'hypothèse selon laquelle la colexification suit le principe de « Boucles d'or » (*Goldilocks principle*), équilibrant ces deux pressions : les concepts sont plus susceptibles d'être colexifiés s'ils sont liés à un degré optimal – i.e. ni trop, ni trop peu. Il reste que la colexification et la polysémie régulière constituent des phénomènes distincts. La colexification est une propriété des concepts ; l'ambiguïté et la polysémie sont des propriétés des expressions linguistiques. Par ailleurs, les études sur la colexification et la polysémie régulière n'adoptent pas la même perspective. Les premières portent majoritairement sur des concepts spécifiques (p. ex. DOIGT, ORTEIL) ; les secondes, sur des catégories sémantiques générales (p. ex. partie du corps).

Alternance	Exemples en français
AGENT/INSTRUMENT	<i>accordeur, découpeuse</i>
ÉTAT/STIMULUS	<i>appréhension, obsession</i>
ÉVÉNEMENT/AGENT	<i>gouvernement, rébellion</i>
ÉVÉNEMENT/ARTÉFACT	<i>construction, reproduction</i>
ÉVÉNEMENT/DOMAIN D'ACTIVITÉ	<i>recherche, randonnée</i>
ÉVÉNEMENT/ÉTAT	<i>humiliation, satisfaction</i>
ÉVÉNEMENT/INSTRUMENT	<i>éclairage, équipement</i>
ÉVÉNEMENT/LIEU	<i>passage, croisement</i>
ÉVÉNEMENT/MOYEN	<i>revêtement, consolation</i>
ÉVÉNEMENT/NOURRITURE	<i>déjeuner, goûter</i>
ÉVÉNEMENT/OBJECT INFORMATIONNEL	<i>description, analyse</i>
ÉVÉNEMENT/PROPRIÉTÉ	<i>détermination, éducation</i>
ÉVÉNEMENT/RÉSULTAT	<i>création, gribouillage</i>
INSTRUMENT/LIEU	<i>saloir, séchoir</i>

TABLEAU 1.1 – Alternances régulières instanciées par les noms déverbaux

Toutes ces considérations nous amènent aux questions suivantes : quelles sont au juste les associations sémantiques investies par les noms déverbaux, et comment les sens concernés se lient-ils les uns aux autres ? L'association sémantique sans doute la plus discutée correspond à la distinction entre interprétations événementielles et référentielles déjà mentionnée dans la section 1.1.1, les secondes dérivant en principe des premières suivant une extension de type métonymique (Anscombe, 1986 ; Grimshaw, 1990 ; Pustejovsky, 1995 ; Jacquey, 2006 ; Melloni, 2011 ; Lieber, 2016 ; Alexiadou, 2019 ; *inter alia*). De nombreuses alternances impliquant un sens d'éventualité et diverses sous-catégories de sens référentiel ont ainsi été observées, telles que ÉVÉNEMENT/RÉSULTAT, ÉVÉNEMENT/ARTÉFACT, ÉVÉNEMENT/OBJECT INFORMATIONNEL, ÉVÉNEMENT/INSTRUMENT, ÉVÉNEMENT/MOYEN, ÉVÉNEMENT/LIEU, ÉVÉNEMENT/AGENT ou encore ÉTAT/STIMULUS (Bierwisch, 1990-1991 ; Godard & Jayez, 1996 ; Osswald,

1.2. L'ambiguïté des noms déverbaux

2005; Jacquey, 2006; Bisetto & Melloni, 2007; Ježek, 2008; Scott, 2010; Melloni, 2011, 2015; Fradin, 2012; Barque et al., 2012; Barque et al., 2014; Huyghe, 2014; Ferret & Villoing, 2015; Lieber, 2016; *inter alia*). Des patrons impliquant uniquement des sens référentiels (p. ex. AGENT/INSTRUMENT, INSTRUMENT/LIEU) ou des éventualités (p. ex. ÉVÉNEMENT/PROPRIÉTÉ, ÉVÉNEMENT/ÉTAT) ont également été relevés, mais restent toutefois plus marginaux (Booij, 1986; Luschützky & Rainer, 2011, 2013; Huyghe & Tribout, 2015; Montermini, 2015; *inter alia*). Le tableau 1.1 propose un aperçu des alternances repérées dans la littérature, sans distinction quant aux cadres d'analyse, qui varient en toute logique d'une étude à l'autre (cf. section 1.1.1). Nous notons au passage que ces associations peuvent naturellement se combiner entre elles pour donner lieu à des patrons tripartites (Ježek, 2008; Lieber, 2018; *inter alia*). En français, les noms *traduction*, *addition* et *enregistrement* dénotent par exemple un événement, un objet informationnel et un objet matériel.

Quelques observations s'imposent quant aux associations évoquées pour conclure cette section. Nous relevons tout d'abord la surreprésentation de la catégorie ÉVÉNEMENT, qui semble corroborer l'hypothèse d'une relation privilégiée entre les noms déverbaux et le domaine de l'action – du moins d'un point de vue qualitatif. En effet, 11 des 14 alternances présentées dans le tableau 1.1 comportent un type événementiel. Il convient ensuite de souligner la prédominance des liens de contiguïté s'agissant des types de relations sémantiques en jeu. Ce constat, révélateur du fort potentiel métonymique de la description événementielle fondée sur les relations entre éventualités et participants, doit toutefois être nuancé pour les dérivés qui ne réalisent pas le type ÉVÉNEMENT. Certain-es auteur-ices (Meyer-Lübke, 1890, p. 272; Booij, 2005, p. 221; *inter alia*) estiment par exemple que l'alternance AGENT/INSTRUMENT repose sur une extension métaphorique plutôt que métonymique. Par ailleurs, le lien « métonymique » reliant les interprétations de certains dérivés est parfois ténu, même lorsque ces lectures entretiennent une relation avec le sens actionnel de la base (p. ex. *saloir* 'pièce où l'on sale' et *saloir* 'instrument servant à saler')¹⁵. Enfin, concernant l'orientation entre interprétations, on notera le passage fréquent de l'abstrait au concret (cf. Rainer, 1996), un phénomène qui paraît une fois encore lié à la prévalence du type événementiel. Signalons cependant que l'interprétation ÉVÉNEMENT n'est pas systématiquement première, comme on pourrait le penser à première vue : la lecture 'visite' du nom *parloir*, par exemple, semble plutôt dériver du sens référentiel 'local' que l'inverse (Salvadori & Huyghe, 2022).

Formes d'ambiguïté

Abordons maintenant une autre dimension des noms déverbaux ambigus : leur forme. À l'instar de nombreux noms simples comme *lapin* 'animal'/'viande' et *chêne* 'arbre'/'bois', les dérivés ambigus présentent souvent des interprétations mutuellement exclusives, i.e. qui renvoient à des situations ou à des entités distinctes. Tel est le cas, par exemple, du nom

15. La situation décrite n'est pas sans rappeler ce que Barque (2008, p. 134-138) appelle la métonymie « faible » par opposition à la métonymie « forte » (p. ex. *construction* 'action de construire'/'résultat de l'action de construire'), et ce que R. Martin (1983, p. 70-71) décrit sous le nom de « polysémie lâche ».

accordeur et de ses deux lectures ‘personne qui accorde’ (31a) et ‘objet qui sert à accorder’ (31b). Cette situation correspond à ce que l’on appelle communément la polysémie disjointe.

- (31) a. *Au Royaume-Uni, un **accordeur** de piano a retrouvé l’équivalent de 600 000 € en pièces d’or daté de la fin du XIXe siècle.* (web)
 b. *Vous pouvez acheter un **accordeur** électronique ou utiliser une appli pour smartphone.* (web)

Plusieurs études (Pustejovsky, 1995; Jacquy, 2006; Ježek & Melloni, 2009; Melloni, 2011; *inter alia*) ont cependant montré que les noms déverbaux peuvent aussi comporter une structure sémantique hybride, caractérisée par des interprétations distinctes mais sémantiquement compatibles. Comme le montrent les phrases suivantes, les lectures ‘objet concret’ et ‘contenu cognitif’ du dérivé *circulaire* peuvent par exemple être isolées (32a)-(32b) ou conciliées (32c) :

- (32) a. *Objet concret : Susie centralise les inscriptions et distribue la **circulaire** à ceux qui n’étaient pas présents.* (web)
 b. *Contenu cognitif : Je précise que je n’ai pas lu la **circulaire**.* (web)
 c. *Objet concret et contenu cognitif : On parle de la **circulaire** et de chasse à l’ours avec Enrico!* (web)

Ce phénomène, qui a fait l’objet de nombreux travaux, reçoit le nom de « facettes sémantiques » (Cruse, 1995; Croft & Cruse, 2004), de « multitypage lexical » (Godard & Jayez, 1996), de « types pointés » (Pustejovsky, 1995; Asher, 2011), de « métonymie intégrée » (Kleiber, 1999), de « métonymie dense » (Nunberg, 1995), ou encore « d’ambivalence sémantique » (Milicevic & Polguère, 2010) dans la littérature.

Il y a lieu de réfléchir à la manière de distinguer entre noms à facettes et polysèmes « standards ». Trois éléments peuvent être convoqués ici. La caractéristique sans doute la plus révélatrice des facettes sémantiques est que, contrairement aux sens de polysèmes standards, elles acceptent la coprédication, pouvant en effet être utilisées conjointement en contexte sans créer d’effet de zeugme (Cruse, 1986; Liebesman & Magidor, 2017; Ortega-Andrés & Vicente, 2019; E. Murphy, 2021; Vicente, 2021; *inter alia*). En (33), les deux interprétations du nom déverbal *certificat* (← *certifier*) sont par exemple convoquées ensemble : le verbe *exhiber* sélectionne celle associée à l’objet matériel, tandis que la tournure *selon lequel* appelle celle liée au contenu informationnel.

- (33) *Il a bénéficié d’une relaxe sur ce point après avoir exhibé un **certificat** selon lequel il était bien assuré le jour de l’accident.* (web)

Il en va de même pour le dérivé *dispensaire* (← *dispenser*) en (34), où le verbe *diriger* renvoie à une institution, et la préposition *au-dessus de*, à un objet construit :

1.2. L'ambiguïté des noms déverbaux

- (34) À Arzila, le drapeau espagnol flotte au-dessus d'un **dispensaire** que dirige un médecin militaire en uniforme. (web)

Ce genre d'association est par contraste impossible pour les sens 'personne qui accorde' et 'objet servant à accorder' du nom ambigu *accordeur* (← *accorder*), incompatibles dans la phrase suivante :

- (35) **L'accordeur qui s'est occupé de mon piano a été fabriqué à l'étranger.*

Une deuxième propriété marquante des noms à facettes tient au fait qu'elles supportent volontiers la sous-spécification contextuelle, puisqu'on peut en effet les convoquer ensemble au moyen d'un seul prédicat (Cruse, 1995, 2004; Nunberg, 1995). La phrase (36), par exemple, décrit une situation où l'objet physique et le contenu informationnel rattachés au nom *certificat* se trouvent ensemble dans le salon :

- (36) Le **certificat** se trouve dans le salon.

À l'inverse, les noms polysémiques comme *accordeur* génèrent automatiquement de l'ambiguïté lorsqu'ils sont employés dans le même genre de contexte. En (37), se trouve dans le salon soit une personne, soit un objet, mais pas les deux à la fois :

- (37) *L'accordeur* se trouve dans le salon.

Une dernière caractéristique des noms à facettes réside dans l'absence d'ordre logique entre les différentes interprétations (Cruse, 2004). La polysémie standard repose en règle générale sur la création d'un sens second à partir d'un sens premier, bien souvent au moyen d'une métaphore ou d'une métonymie. Les noms à facettes, en revanche, ne présentent pas une orientation clairement définie entre leurs interprétations, malgré la relation évidente entre celles-ci. Il semble ainsi difficile de déterminer si le sens premier de *certificat* correspond à l'objet physique ou au contenu informationnel.

Nous terminerons ici en mentionnant que le multitypage lexical est parfois considéré comme un cas particulier de polysémie régulière, en vertu de la régularité de certaines associations entre facettes. La littérature recourt ainsi fréquemment aux termes de polysémie « logique » ou « inhérente » pour référer à ce premier phénomène (Asher, 2011; Pustejovsky, 1995; Arapinis & Vieu, 2015; Ortega-Andrés & Vicente, 2019; *inter alia*). Cet amalgame est toutefois mis en doute (Cruse, 2000; Liebesman & Magidor, 2017; Vicente, 2018; Dölling, 2020; Huyghe et al., 2024; *inter alia*). Selon toute vraisemblance, les noms à facettes s'apparentent en effet davantage à des monosèmes qu'à de véritables polysèmes, un point dont nous tiendrons compte dans le chapitre 2.

1.2.2 Regard morphologique

Passons à présent aux thèmes de recherche les plus débattus dans les études de morphologie qui abordent la question de l'ambiguïté des déverbaux. Cette section s'attardera sur trois éléments en particulier : les similarités et différences entre noms ambigus simples et complexes, l'implication des propriétés des verbes de base dans la formation de dérivés ambigus, ainsi que l'apport des procédés morphologiques (affixes et formes de conversion).

Comparaison avec le lexique simple

L'absence de consensus quant à la manière dont l'ambiguïté des dérivés se compare à celle des noms simples est manifeste dans la littérature. Un certain nombre d'auteur·ices, dont Apresjan (1974), estiment d'une part que les lexiques simple et construit adoptent le même comportement. Lieber (2004, p. 11), par exemple, fait remarquer que « les types de polysémie que nous trouvons dans le lexique simple devraient également se retrouver parmi les mots dérivés ¹⁶ ». Cette perspective trouve aussi des échos favorables chez Bierwisch (1990-1991, p. 53), qui considère que la variation sémantique associée à l'alternance ÉVÉNEMENT/RÉSULTAT ne doit en aucun cas être considérée « comme une spécificité de la nominalisation événementielle, ni d'ailleurs de la morphologie dérivationnelle », puisqu'elle reflète simplement « le phénomène global de la polysémie ¹⁷ ». Différents travaux contestent néanmoins cette position. Scott (2010, p. 713) soutient notamment que les dérivés polysémiques en *-ung* de l'allemand (p. ex. *Bewachung* 'action de garder' → 'personnes qui gardent') contredisent « l'hypothèse largement répandue selon laquelle le sens du changement sémantique va du concret à l'abstrait ¹⁸ ». Melloni (2011, p. 163), à rebours de Bierwisch (1990-1991), envisage également la polysémie ÉVÉNEMENT/RÉSULTAT comme le propre des noms déverbaux, ce qu'elle explique par leur structure morphologique, i.e. la présence du verbe de base ¹⁹.

On peut émettre des doutes au sujet de cette dernière affirmation, puisqu'il semble exister en français des noms simples comme *méchoui* instanciant le patron ÉVÉNEMENT/RÉSULTAT :

- (38) a. ÉVÉNEMENT : *L'association Amitiés douzennes a rassemblé de nombreux sympathisants lors du méchoui, qui à cause du mauvais temps a eu lieu au foyer municipal.* (web)

16. [En anglais dans le texte] « the sorts of polysemy we find in the simplex lexicon should also be found in derived words ».

17. [En anglais dans le texte] « the type of variation considered here now must in no way be taken as a specificity of event nominalization, or derivational morphology, for that matter. The variation simply reflects the overall phenomenon of polysemy ».

18. [En anglais dans le texte] « The direction of semantic change in these words clearly goes from abstract to concrete, contradicting the widely held assumption that the direction of semantic shift is concrete to abstract ».

19. [En anglais dans le texte] « I wish to emphasize that the morphological structure of derived nominals, i.e. the presence of the base verb, has influence not only at the morpho-syntactic level but also at the semantic level, determining the peculiar event-result polysemy we do not find in the simple lexicon ».

1.2. L'ambiguïté des noms déverbaux

- b. RÉSULTAT : *Tous les ans avant la fin des vacances d'été, l'association La Fraternelle invite les habitants de Nottonville et leurs amis à venir déguster un **méchoui** et passer une bonne soirée.* (web)

Il n'est cependant pas exclu que des études axées sur la fréquence d'instanciation des schémas de polysémie puissent révéler des différences notables entre lexiques simple et construit. Quand bien même les alternances en jeu se trouveraient être identiques, il est probable que celles impliquant un sens événementiel se manifestent moins fréquemment parmi les noms simples comparativement aux noms déverbaux, ces derniers dénotant *a priori* plus souvent des événements.

Propriétés des lexèmes-bases

Examinons maintenant plus avant les liens entre structure morphologique et ambiguïté nominale, en commençant par le rôle joué par les bases. L'idée qu'il existe un lien entre les propriétés de ces dernières et la polysémie des dérivés apparentés n'est certainement pas nouvelle (voir p. ex. Gamillscheg, 1921), et découle logiquement du constat général selon lequel la sémantique des nominalisations demeure étroitement liée à celle des verbes (cf. section 1.1.2). Mais de quelles propriétés parle-t-on exactement ?

Une série d'études se sont premièrement penchées sur l'aspect lexical (ou *Aktionsart*) des lexèmes-bases. La télélicité des verbes, i.e. leur capacité à décrire une éventualité comportant un point culminant – le *telos* – qui actualise l'action décrite (cf. Garey, 1957; Vendler, 1967; Rothstein, 2004; Demonte & McNally, 2012; Filip, 2012; *inter alia*), apparaît ici comme un point d'intérêt fort. S'agissant des noms déverbaux en allemand, Osswald (2005, p. 259) affirme par exemple que seuls les verbes téléliques peuvent donner lieu à une double lecture événementielle et résultative, en partant de l'idée que cette dernière est favorisée par la culmination de l'action. La proposition est séduisante, et les exemples nombreux en français : *construire/construction*, *gribouiller/gribouillage*, *imprimer/impression*, *graver/gravure*, *broder/broderie*, etc. Osswald (2005) reconnaît toutefois la télélicité comme une condition nécessaire mais non suffisante, dans la mesure où certains noms dérivés de verbes téléliques ne comportent qu'une lecture événementielle (p. ex. *fertigstellen* 'achever' / *Fertigstellung* 'achèvement'). Ce constat nuancé est corroboré par les résultats de Huyghe et Lombard (2022) qui ont montré, à partir de l'analyse de 139 néologismes suffixés en *-age* en français, que la seule télélicité des lexèmes-bases ne suffit pas à prédire la double nominalisation d'action et de résultat : celle-ci doit en effet être couplée à la présence d'un patient dans la structure argumentale des verbes. Melloni (2015) relève en outre – *contra* Osswald (2005) – que les verbes d'activité peuvent eux aussi donner lieu à des lectures résultatives, tout en estimant aussi que les verbes d'accomplissement et d'achèvement demeurent les candidats privilégiés pour la formation de noms résultatifs. On peut ainsi mentionner pour le français la paire *macérer/macération*, où le nom instancie bien l'alternance ÉVÉNEMENT/RÉSULTAT (39), mais où le verbe dénote une activité.

- (39) a. La **macération** des baies a duré environ trois semaines. (web)
 b. Alternativement, vous pouvez acheter ces **macérations** toutes prêtes [...]. (web)

Les types de procès décrits par les bases verbales sont aussi fréquemment examinés. Bierwisch (1990-1991, p. 53) relève notamment l'impossibilité, pour des noms comme *Sprung* 'saut', *Wurf* 'lancer' ou encore *Ritt* 'sortie à cheval' en allemand, d'instancier une lecture résultative, car leurs bases *springen* 'sauter', *werfen* 'lancer' et *reiten* 'monter à cheval' dénotent des situations qui n'amènent pas à un résultat pouvant être distingué de l'occurrence de l'événement lui-même²⁰. Pour l'italien, Melloni (2011, p. 239) soutient que la polysémie ÉVÉNEMENT/PRODUIT (p. ex. *costruzione* 'action de construire'/'résultat de l'action de construire', *traduzione* 'action de traduire'/'résultat de l'action de traduire') n'apparaît que si le verbe de base peut être envisagé comme décrivant un événement de création²¹. Enfin, Ferret et Villoing (2015) ont montré que la lecture instrumentale au sein de l'alternance ÉVÉNEMENT/INSTRUMENT dépend de la réalisation des propriétés [+causation externe] et [+manière] par les bases dans le cas des dérivés en *-age* du français (p. ex. *camouflage*, *éclairage*, *embrayage*).

Nous terminerons ici avec une troisième propriété des verbes qu'il paraît pertinent d'examiner : leur ambiguïté. En effet, si l'on part du principe que les procédés de dérivation s'appliquent à des éléments spécifiés sémantiquement (Mel'čuk, 1993; Fradin & Kerleroux, 2003; *inter alia*), différentes configurations d'appariement entre sens verbaux et nominaux peuvent être envisagées. Huyghe et Lombard (2022, p. 40) notent ainsi que :

« Deux acceptions d'une même forme nominale peuvent être soit liées à la même acception verbale (p. ex. *entassage*₁ 'fait d'entasser des objets' et *entassage*₂ 'objets entassés', tous deux associés à *entasser* 'mettre en tas'), soit liées à des acceptions verbales distinctes (p. ex. *empruntage*₁ 'fait d'obtenir à titre de prêt' et *empruntage*₂ 'fait de suivre une voie', associés respectivement à *emprunter*₁ 'obtenir à titre de prêt' et *emprunter*₂ 'suivre une voie'). »

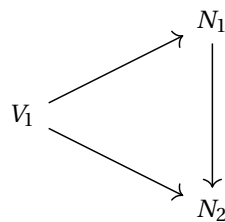
Ce second scénario est également relevé par Brdar (2017, p. 169-170), Barque et ses collègues (2018, p. 101) et Polguère (2018, p. 815-816) qui, de manière intéressante, décrivent même deux sous-cas possibles. Le premier sous-cas correspond aux situations où les acceptions du verbe ne sont pas liées sémantiquement et donnent donc lieu à deux interprétations nominales indépendantes l'une de l'autre (Barque et al., 2018). Les lectures 'rendre mat' (40a) et 'réprimer' (40b) du verbe *mater* n'entretiennent par exemple pas de lien sémantique, et il en va de même des interprétations 'action de rendre mat' (41a) et 'action de réprimer' (41b) du dérivé *matage* :

20. [En anglais dans le texte] « the events do not lead to a result distinguishable from the occurrence of the event itself ».

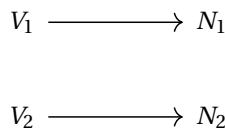
21. [En anglais dans le texte] « in order to have a nominal with polysemous E [event] and PRODUCT reading, the base verb must be conceived of as describing a creation event ».

1.2. L'ambiguïté des noms déverbaux

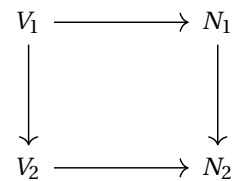
- (40) a. *La fraise est pourvue de particules de diamant et permet ainsi, par exemple, de **mater** le verre et d'effectuer d'autres travaux de finition avec un minimum d'effort.* (web)
 b. *La police **a maté** la mutinerie en usant de grenades lacrymogènes.* (web)
- (41) a. *Le **matage** du verre est réalisé par projection de microgranulés.* (web)
 b. *Il semble avoir eu peu d'intérêt pour ce travail, et guère plus pour les diverses campagnes militaires auxquelles il participa, que ce soit le **matage** de la révolte de Abaza ou le siège infructueux de Bagdad, tombée aux mains des Séfévides.* (web)



(a) Sens nominaux associés à une seule acception verbale



(b) Sens nominaux associés à deux acceptions verbales indépendantes sémantiquement



(c) Sens nominaux associés à deux acceptions verbales liées sémantiquement

FIGURE 1.3 – Trois configurations d'ambiguïté considérant l'appariement entre sens nominaux (N_X) et verbaux (V_X)

Le second sous-cas correspond aux situations où les interprétations verbales entretiennent un lien sémantique et où ce lien est également instanciés par leurs dérivés (Brdar, 2017; Polguère, 2018). Par exemple, les sens 'passer le balai' (42a) et 'passer sur quelque chose' (42a) du verbe *balayer* sont liés par une métaphore, tout comme les lectures 'action de passer le balai' (43a) et 'action de passer sur quelque chose' (43b) du nom ambigu *balayage* :

- (42) a. *Aujourd'hui, c'est Amina qui **balaie** la cour.* (web)
 b. *Le radar **balaie** constamment l'horizon et émet des impulsions électromagnétiques qui se propagent dans l'atmosphère.* (web)
- (43) a. *Si le travail s'est révélé ingrat et gigantesque, les appelés avaient néanmoins apprécié ces corvées, plus gratifiantes que le **balayage** de la cour de la caserne.* (web)

- b. Si le ciel est un objet d'observation accessible à tous, encore faut-il savoir comment l'observer, car le **balayage** du ciel doit respecter certaines règles. (web)

Nous illustrons en (1.3) les trois configurations d'ambiguïté relevées, en nous inspirant du schéma que Polguère (2018) propose pour l'adjectif polysémique *applicable* et sa base *appliquer*.

Procédés morphologiques

Un constat faisant sans doute l'unanimité dans les études sur la sémantique de la dérivation veut que la polysémie des dérivés soit intimement liée non seulement aux lexèmes-bases, mais aussi aux procédés morphologiques en jeu. Un nombre important de travaux portent ainsi sur les différentes alternances sémantiques associées à des formes de conversion ou à des affixes particuliers. Pour l'allemand, Scott (2010) soutient par exemple que le suffixe *-ung* constitue le seul procédé productif permettant de former des noms d'action déverbaux dont les dérivés reçoivent régulièrement une seconde interprétation (résultative ou autre)²². En comparaison, il observe que les conversions infinitives n'acquièrent une lecture résultative qu'à titre exceptionnel, tandis que les dérivés en *-erei* tendent à favoriser une double interprétation ÉVÉNEMENT/LIEU. Schulte (2015a) relève aussi, dans le cadre de son étude des suffixes *-age* et *-ery* en anglais, que les dérivés formés avec ces deux procédés se voient attribuer des types sémantiques très similaires, mais que les catégories en jeu ne s'associent cependant pas de manière équivalente selon le suffixe : l'alternance ACTION → COLLECTIF est davantage privilégiée par les dérivés en *-age* que par les nominalisations en *-ery*. Suivant l'hypothèse que bases et affixes peuvent interagir (cf. Lieber, 2004; Plag et al., 2018; *inter alia*), bon nombre d'études prennent par ailleurs en considération les caractéristiques des lexèmes-bases en complément des procédés morphologiques. Nous avons déjà mentionné l'étude de Ferret et Villoing (2015) à propos du patron ÉVÉNEMENT/INSTRUMENT instancié par certains dérivés en *-age* en français, qui requiert des propriétés sémantiques précises dans le domaine verbal. Pour le néerlandais, Booij (1986) argumente quant à lui que l'alternance tripartite AGENT PERSONNEL/AGENT IMPERSONNEL/INSTRUMENT qu'instancient certains noms en *-er* (p. ex. *sender* 'personne qui envoie'/'station radio ou TV'/'émetteur') peut être expliquée par le schéma d'extension suivant quand ils sont dérivés de verbes à agent sujet : AGENT PERSONNEL > AGENT IMPERSONNEL > INSTRUMENT.

Il reste que bon nombre d'études s'intéressant à la diversité sémantique permise par les procédés morphologiques n'établissent pas toujours une distinction claire entre polysémie lexicale d'une part, et ce que la littérature appelle « polysémie affixale²³ » de l'autre (cf. Plag

22. [En anglais dans le texte] « *-ung* [...] is the only productive means of forming deverbal action nouns whose derivatives regularly receive a secondary (result or other concrete) interpretation. The products of infinitival conversion [...] only exceptionally acquire a result interpretation [...] ».

23. Il convient de relever que les différentes notions habituellement réservées à l'étude de l'ambiguïté lexicale (p. ex. polysémie, homonymie, métonymie) sont non seulement mobilisées pour l'analyse d'affixes, mais aussi des relations entre bases morphologiques et noms dérivés. Valera et Ruz (2021), par exemple, passent en revue

1.3. Bilan

et al., 2023; Kawaletz, 2023; *inter alia*), i.e. le fait qu'un affixe donné puisse être utilisé pour former des noms de types sémantiques différents. Comme le font remarquer Kotowski et Plag (2023, p. 2), en effet, « on ne sait pas ce qui doit être considéré comme polysémique : l'affixe en tant que tel ou les dérivés dans lesquels il apparaît²⁴ ». Les deux phénomènes ne sont pourtant pas équivalents. Le fait qu'un affixe génère plusieurs sens n'implique pas nécessairement que ses dérivés les possèdent tous (cf. Scott, 2010, p. 726). À l'inverse, le fait qu'un dérivé présente plusieurs sens n'implique pas que l'affixe avec lequel il est formé génère nécessairement toutes ces lectures. L'interprétation nominale seconde peut en effet être acquise par une extension métaphorique ou métonymique indépendante des processus morphologiques. Considérons à titre d'exemple le cas du suffixe *-er* en anglais, sans doute l'un des procédés morphologiques de nominalisation les plus étudiés (cf. Ryder, 1999; Panther & Thornburg, 2002; Bauer et al., 2013; Bauer, 2017; *inter alia*). Il semble difficilement concevable qu'un seul dérivé donné puisse instancier l'ensemble des types sémantiques attribués au suffixe *-er*, connu pour former des noms d'agent humain (*writer* 'écrivain'), d'agent non humain (*warbler* 'fauvette'), d'expérimenteur (*beholder* 'observateur'), d'instrument (*amplifier* 'amplificateur'), de lieu (*smoker* 'fumoir') ou encore de patient (*scratcher* 'güiro'). À contrario, comme le relève Bauer (2017, p. 6), l'interprétation 'lieu où l'on mange' du dérivé *diner* (← *dine*) pourrait tout à fait constituer le résultat d'une extension métonymique à partir du sens 'personne qui mange' du même mot, plutôt qu'une lecture distincte du suffixe *-er*²⁵. Nous examinerons plus avant les liens entre polysémies lexicale et affixale dans le chapitre 5, en cherchant au passage à nous informer sur la manière dont la langue traite la distribution du sens dans différents types de formes ambiguës.

1.3 Bilan

Les sections précédentes ont exposé une série d'éléments théoriques destinés à guider le reste de notre étude. Résumons brièvement trois d'entre eux faisant consensus dans la littérature :

Récurrence de certaines catégories et associations sémantiques Les noms déverbaux instancient fréquemment certains types sémantiques (p. ex. ÉVÉNEMENT, RÉSULTAT, AGENT, INSTRUMENT), qui se retrouvent par ailleurs souvent associés au sein de dérivés donnés (p. ex. ÉVÉNEMENT/RÉSULTAT, AGENT/INSTRUMENT). À l'instar des noms simples, les nominalisations peuvent ainsi s'inscrire dans des schémas de polysémie régulière.

les différentes analyses ayant été proposées dans la littérature pour caractériser les liens entre lexèmes dans le cadre la conversion (p. ex. homonymie, polysémie, extension lexicale). Pour la dérivation marquée formellement, Janda (2011, 2014) affirme quant à elle que les dérivés constituent des extensions « métonymiques » des bases. Cette position est toutefois critiquée (Brdar & Brdar-Szabó, 2014; Brdar, 2017; *inter alia*).

24. [En anglais dans le texte] : « it is unclear what is to be taken as polysemous : an affix as such or the derivatives it occurs in ».

25. [En anglais dans le texte] : « *Diner* in the sense of 'location where one eats' could arise by metonymy from *diner* 'a person who eats', rather than a separate meaning of *-er* ».

Formes d’ambiguïté variables La polysémie des noms déverbaux n’est pas homogène d’un point de vue structurel. Elle peut être de forme conjointe (p. ex. *certificat*) ou disjointe (p. ex. *accordeur*), selon que les sens concernés peuvent ou non apparaître conjointement dans l’interprétation.

Influence des bases verbales et des procédés morphologiques L’ambiguïté des noms déverbaux peut être conditionnée par les lexèmes-bases, de même que par les procédés morphologiques en jeu. Une interaction entre ces deux éléments n’est pas à exclure.

Malgré ces points d’accord, le sujet de l’ambiguïté des nominalisations n’est certainement pas épuisé. Nous identifions deux lacunes en particulier, au-delà de l’absence fréquente d’analyses quantitatives dans ce domaine :

Description étendue Il manque d’une part une description systématisée de l’ambiguïté des noms déverbaux du français en synchronie, qui tiendrait compte non seulement de l’ensemble des procédés morphologiques pouvant être employés dans le cadre de la nominalisation de verbe à nom, mais aussi des configurations d’appariement entre sens verbaux et nominaux.

Origine de l’ambiguïté Se pose aussi la question de l’origine de l’ambiguïté des noms déverbaux. En particulier, il faudrait déterminer dans quelle mesure et de quelle façon extension sémantique et construction morphologique peuvent concourir pour former des dérivés ambigus (cf. Ferret & Villoing, 2015; Huyghe & Lombard, 2022). Certains cas semblent au demeurant clairs : s’agissant des noms présentant deux lectures dont l’une n’entretient aucun lien sémantique avec la base (p. ex. *lacer* → *lacet* ‘objet servant à lacer’ / ‘tournant d’une route’), le fait que le second sens dérive du premier par une figure sémantique fait peu de doute. Il reste que pour de nombreux dérivés, les hypothèses d’une extension figurative et d’une construction morphologique entrent en concurrence. Par exemple, le sens résultatif de *construction* constitue-t-il une métonymie de l’interprétation actionnelle du même nom suivant un patron de polysémie régulière ÉVÉNEMENT → RÉSULTAT, ou dérive-t-il plutôt directement du verbe *construire*?

Nos prochains chapitres sont dédiés à ces deux problématiques. Le chapitre 2 décrit la collecte et l’analyse sémantique approfondie d’un vaste échantillon de nominalisations et de verbes apparentés. Le chapitre 3 établit le profil sémantique de ces noms et de ces verbes, et décrit notamment comment leurs différents sens peuvent s’apparier. Le chapitre 4 sonde l’existence d’associations non aléatoires entre types sémantiques, potentiellement révélatrices de relations conceptuelles privilégiées. Enfin, le chapitre 5 explore l’influence des bases et des procédés morphologiques dans la formation de noms déverbaux ambigus.

Chapitre 2

Méthodologie

Le présent travail s’attache à la description et à l’analyse de l’ambiguïté des noms déverbaux en français. Nous l’avons vu précédemment, différents facteurs liés à ce phénomène ont été identifiés dans la littérature, qu’il s’agisse de l’*Aktionsart* des lexèmes-bases ou encore des rôles sémantiques associés à leurs programmes de complémentation. Fait toutefois encore défaut une étude approfondie et à grande échelle de l’implication de ces éléments dans la construction de l’ambiguïté des nominalisations.

Les outils de sémantique distributionnelle et le typage sémantique (semi-)automatique n’offrent pas, à première vue, la précision nécessaire pour une analyse aussi fine. Pour pallier cette lacune, nous avons donc entrepris l’annotation manuelle d’un large échantillon de paires verbe-nom, permettant d’identifier les multiples sens en jeu et de poser ainsi les bases pour des analyses statistiques ultérieures. Ce chapitre détaille la méthodologie adoptée pour cette collecte et analyse sémantique. Les données rassemblées s’inscrivent dans le cadre d’un projet plus vaste sur la sémantique des noms déverbaux, financé par le Fonds national suisse (FNS)²⁶ et au sein duquel la description de l’ambiguïté des déverbaux tient une place importante. Elles sont le fruit d’une annotation réalisée en collaboration par plusieurs chercheur-euses dont l’auteurice de cette thèse, qui a joué un rôle central dans le travail de constitution et d’homogénéisation des données.

Le chapitre est structuré comme suit. La section 2.1 introduit la méthode d’échantillonnage employée pour la collecte des différents noms et verbes associés. La section 2.2 présente ensuite les principes d’analyse suivis pour l’étude des types sémantiques nominaux (2.2.1), ainsi que de l’aspect lexical (2.2.2), des rôles sémantiques (2.2.3) et de l’ambiguïté (2.2.4) des verbes et des noms. La section 2.3 revient enfin sur la procédure d’analyse appliquée par les annotateur-ices.

26. De plus amples informations sont disponibles à l’adresse suivante : <https://data.snf.ch/grants/grant/188782>.

2.1 Échantillonnage

L'objectif général de la collecte effectuée consistait à obtenir un très large échantillon de paires verbe-nom (i.e. plus de 4'000 paires) dans lequel :

- (i) verbes et noms sont apparentés formellement et sémantiquement en synchronie;
- (ii) un maximum de procédés morphologiques permettant la nominalisation de verbes en français sont inclus;
- (iii) les procédés morphologiques rares (p. ex. *-aille*, *-asse*, *-is*) sont surreprésentés en fréquence par rapport aux procédés morphologiques très productifs (p. ex. *-age*, *-eur*, *-ment*), de façon à garantir leur présence en nombre suffisant et la possibilité d'une généralisation des observations.

Cette sélection s'est faite en deux étapes : une extraction formelle automatique des paires verbe-nom d'abord, puis une sélection manuelle sur la base de différentes conditions. Nous les détaillons dans les sections 2.1.1 et 2.1.2.

2.1.1 Extraction automatique

La première étape de l'échantillonnage a consisté à extraire de manière automatique tous les noms et verbes formellement apparentés du corpus FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015), un vaste corpus web du français contemporain contenant 10.8 milliards de tokens. Cette ressource présente l'avantage d'inclure des noms bien lexicalisés et donc répertoriés dans les dictionnaires de langue courante (p. ex. *lancée*, *tricherie*, *vieillessement*), ainsi que des néologismes (p. ex. *noctambulation*, *pendouillement*, *roucoulage*). Elle permet en outre d'extraire des familles morphologiques relativement complètes (p. ex. *arrosage*, *arrosment*, *arrosoir*, *arroseur*, *arroseuse*), pouvant notamment renseigner sur la répartition des types sémantiques entre les procédés morphologiques. Différents cas d'allo-morphie régulière, de base ou affixale, ont été pris en compte lors de cette extraction (p. ex. *certifier/certification*, *comprendre/compréhension*, *décroître/décroissance*).

Au total, 42 suffixes préalablement identifiés comme formateurs de noms déverbaux dans la littérature ont été considérés pour la sélection formelle (voir p. ex. Uth, 2008 pour *-age* et *-ment*; Aliquot-Suengas, 2003 pour *-ade*; Plénat, 1999 pour *-aille*; Ferret & Villoing, 2015 pour *-age*; Schnedecker & Aleksandrova, 2016 pour *-aire*; Burdy, 2013 pour *-aison*; Dal & Namer, 2010 pour *-ance/-ence*; Glaser, 1910 pour *-ard* et *-arde*; Dal, 1999 pour *-et* et *-ette*; Huyghe & Tribout, 2015 pour *-eur*; Vendryes, 1946 pour *-is*; Namer & Villoing, 2008 pour *-oir* et *-oire*; Zellmer, 1935 pour *-ure*; *inter alia*). Par souci d'exhaustivité, nous avons ajouté aux formes suffixées les conversions verbo-nominales, notamment lorsque celles-ci comportent un exposant verbal permettant d'orienter la dérivation (Tribout, 2010, 2012). Plus précisément, 3 formes de conversion marquées ont été prises en compte : la conversion à partir du thème 12 du verbe, qui permet de former les participes passés (cf. Bonami & Boyé, 2003); la conversion à partir du thème 13 du verbe, caché à la flexion et servant de base à la suffixation « savante » (cf.

Kerleroux, 2007; Bonami et al., 2009); et les « infinitifs substantivés », bien que leur statut de convertis soit discuté (cf. Kerleroux, 1996). Certains cas de conversion sur le thème 0 (Tribout, 2010, 2012) ont également été ajoutés²⁷, dans des conditions restreintes que nous détaillons plus loin (cf. section 2.1.2). Ce sont au final 46 procédés de construction morphologique qui ont été retenus, présentés en intégralité dans le tableau 2.1 ci-dessous.

Procédé	Exemple
-ade	<i>promener / promenade</i>
-age	<i>atterrir / atterrissage</i>
-ail	<i>épouvanter / épouvantail</i>
-aille	<i>sonner / sonnaillie</i>
-ain	<i>écrire / écrivain</i>
-aire (fém.)	<i>mandater / mandataire</i>
-aire (masc.)	<i>commenter / commentaire</i>
-aison	<i>livrer / livraison</i>
-ance/-ence	<i>provenir / provenance</i>
-ant/-ent	<i>militer / militant</i>
-ante/-ente	<i>diriger / dirigeante</i>
-ard	<i>piller / pillard</i>
-arde	<i>fêter / fêtarde</i>
-asse	<i>lier / liasse</i>
-eau	<i>traîner / traîneau</i>
-elle	<i>sauter / sauterelle</i>
-er	<i>surfer / surfer</i>
-er/-ier	<i>hériter / héritier</i>
-ère/-ière	<i>conseiller / conseillère</i>
-eresse	<i>vendre / venderesse</i>
-erie	<i>cajoler / cajolerie</i>
-et	<i>siffler / sifflet</i>
-ette	<i>renifler / reniflette</i>
-eur (fém.)	<i>valoir / valeur</i>
-eur (masc.)	<i>assurer / assureur</i>
-eure	<i>gouverner / gouverneure</i>
-euse	<i>photocopier / photocopieuse</i>
-in	<i>pétrir / pétrin</i>
-ine	<i>téter / tétine</i>
-ing	<i>brieffer / briefing</i>

27. Tribout (2012, p. 121-122) estime que les cas ordinaires de conversion de verbe à nom sont problématiques lorsque le verbe de base appartient au deuxième groupe de conjugaison, ce qui nécessite selon elle de considérer un thème supplémentaire, appelé thème 0. Celui-ci est identique au thème 3, sauf lorsque le verbe appartient au deuxième groupe, auquel cas le thème 0 est identique au thème 3 moins /i/. Si ce thème supplémentaire est par exemple repris par Missud et Villoing (2021), Bonami (2014, p. 72) questionne quant à lui le gain analytique d'un tel ajout à l'espace thématique des verbes. Nous retenons ici par commodité l'étiquette « thème 0 », bien qu'elle puisse être discutée.

2.1. Échantillonnage

Procédé	Exemple
-ion	<i>indiquer/indication</i>
-is	<i>dégueuler/dégueulis</i>
-ise	<i>hanter/hantise</i>
-ment	<i>chuchoter/chuchotement</i>
-oir/-oire	<i>mourir/mouroir</i>
-oire	<i>écumer/écumoire</i>
-on	<i>jurer/juron</i>
-onne	<i>écrivain/écrivainne</i>
-ot	<i>brûler/brûlot</i>
-ote/-otte	<i>juger/jugeote</i>
-rice	<i>sculpter/sculptrice</i>
-ure	<i>craqueler/craquelure</i>
Conversion - Thème 0	<i>ébaucher/ébauche</i>
Conversion - Thème 12	<i>fumer/fumée</i>
Conversion - Thème 13	<i>postuler/postulat</i>
Conversion - Infinitif	<i>sourire_V/sourire_N</i>

TABLEAU 2.1 – Procédés morphologiques examinés

Le choix de ces différents procédés morphologiques appelle quelques commentaires. On remarque premièrement que certains suffixes parfois cités comme déverbaux (p. ex. *-iste*) ont été exclus de l'échantillonnage. Cette décision se justifie par l'impossibilité, dans leur cas, de garantir l'existence d'un patron constructionnel de type $V \rightarrow N-x$, déterminé par la présence dans le lexique d'au moins deux cas de $N-x$ strictement analysables comme dérivés d'un verbe. Par exemple, bien que certains noms en *-iste* tels que *copiste* ou *cymbaliste* soient formellement et sémantiquement apparentés à des verbes (respectivement *copier* et *cymbaler*), des incertitudes demeurent toutefois quant à leur possible construction sur base nominale (*copie* \rightarrow *copiste*, *cymbale* \rightarrow *cymbaliste*). Il a par ailleurs été décidé de regrouper et d'analyser conjointement les cas de variation orthographique ou allomorphique (p. ex. *-ier* et *-er*, *-ance* et *-ence*, *-ote* et *-otte*). En revanche, les suffixes différant selon le genre du dérivé ont été traités séparément, dans la mesure où nous ne postulons pas d'équivalence sémantique entre les formes masculines et féminines. Cette distinction concerne les affixes suivants : *-ant/-ent* et *-ante/-ente*; *-ard* et *-arde*; *-eau* et *-elle*; *-er/-ier* et *-ère/-ière*; *-et* et *-ette*; *-eur*, *-eure*, *-euse* et *-rice*; *-in* et *-ine*; *-oir/-oire* (masculin) et *-oire* (féminin); *-on* et *-onne*. Deux suffixes – l'un masculin, l'autre féminin – ont également été postulés pour la forme *-aire*, même si les dérivés concernés ne semblent varier en genre que dans le cas de la dénotation humaine.

2.1.2 Sélection manuelle

Au final, 59'353 paires verbe-nom candidates ont été extraites automatiquement. Les données étant très bruitées à ce stade de l'analyse, une sélection contrôlée des paires pertinentes a dû être effectuée. Cette seconde étape de l'échantillonnage a été réalisée manuellement par 6 personnes membres du projet sur la sémantique des noms déverbaux, dont l'autrice du présent travail ²⁸. Elle s'est déroulée en plusieurs temps :

- (i) un tri exhaustif des listes brutes obtenues lors de l'extraction automatique pour 37 procédés morphologiques relativement peu représentés dans le corpus FRCOW16A (i.e. comptabilisant moins de 950 items), tels que *-aille*, *-ot* et *-ise*;
- (ii) la complétion des amorces de familles morphologiques ainsi obtenues à l'aide de noms formés avec les 9 procédés restants (i.e. représentant 950 items ou plus dans le corpus), tels que *-age*, *-eur* et *-ment*;
- (iii) enfin, l'ajout de paires verbe-nom sélectionnées aléatoirement pour les 9 procédés morphologiques considérés à l'étape (ii), de manière à augmenter leur représentation dans la base de données. Des familles complètes de noms déverbaux (i.e. comprenant des noms formés avec les 9 procédés considérés) ont également été constituées à ce stade.

Diverses conditions ont été appliquées tout au long de cette seconde sélection afin de garantir la bonne qualité de l'échantillon. Deux d'entre elles concernent l'ensemble des noms :

C1 - Analysabilité sémantique de relation La première condition veut qu'une relation sémantique puisse être postulée, dans l'état actuel de la langue, entre au moins un sens du nom et un sens du verbe apparenté. Par exemple, la paire *visionner/visionneuse* est conservée, car le nom *visionneuse* dénote un instrument servant à visionner quelque chose. À l'inverse, la paire *peigner/peignoir* est exclue, le nom ayant en effet perdu toute motivation par rapport à son lexème-base, bien qu'il y soit lié historiquement.

C2 - Régionalismes, archaïsmes et termes de spécialité La deuxième condition stipule que les noms examinés ne doivent être ni des régionalismes, ni des archaïsmes, ni des termes de spécialité. Si elle sert d'une part à éviter une surabondance de ces derniers dans l'échantillon, elle vise surtout à garantir la qualité globale de l'évaluation sémantique effectuée par les annotateur·ices. En effet, un manque d'intuition face à un sens inconnu peut rendre son évaluation délicate. Cette condition est vérifiée sur la base des indications fournies dans les dictionnaires *Le Petit Robert* (Éditions Le Robert, s. d.-a) et le *Trésor de la Langue Française informatisé* (ATILF - CNRS & Université de Lorraine, s. d.) pour les noms lexicalisés, et des informations glanées sur la Toile pour les néologismes. Les lexèmes-bases ne sont en principe pas concernés

28. Les personnes suivantes ont été impliquées (par ordre alphabétique) : Lucie Barque, Pauline Haas, Richard Huyghe, Justine Salvadori, Delphine Tribout et Marine Wauquier.

2.1. Échantillonnage

par cette deuxième condition, à l'exception des verbes de spécialité, qui sont retirés de l'échantillon suivant le même critère que pour les noms.

Deux autres conditions concernent spécifiquement les noms suffixés multimotivés, i.e. les formes pour lesquelles plusieurs lexèmes sont admissibles comme bases (cf. chapitre 1) :

C3 - Double analysabilité portant sur l'ordre d'affixation Cette troisième condition s'applique spécifiquement aux noms à la fois préfixés et suffixés pour lesquels plusieurs schémas de construction sont envisageables. Il est difficile de dire, en effet, si la base de *préfertilisation* est le verbe *préfertiliser* ou le nom *fertilisation*, et si celle de *coexécuteur* est *coexécuter* ou *exécuteur*. Par défaut, ces paires sont conservées dans l'échantillon si elles satisfont à la condition **C1** (analysabilité sémantique de relation), i.e. si un lien sémantique peut être postulé entre le verbe et le nom.

C4 - Double analysabilité portant sur la base La quatrième condition vise deux types de noms en particulier. Elle se rapporte d'une part aux lexèmes formés avec des procédés morphologiques pouvant s'adjoindre à des bases autres que des verbes. Par exemple, il peut être difficile de déterminer si certains noms d'instrument suffixés en *-ette* (p. ex. *moulinette*) sont dérivés d'un verbe (p. ex. *mouliner* / *moulinette* 'objet servant à mouliner') ou d'un nom (p. ex. *moulin* / *moulinette* 'petit moulin'). De nombreux dérivés peuvent d'autre part être obtenus par conversion à partir d'adjectifs. Sont notamment concernés certains noms d'agent suffixés en *-aire* comme *suicidaire* et *protestataire*, répertoriés à la fois comme des adjectifs et des noms dans le *Petit Robert* (Éditions Le Robert, s. d.-a) et dans le *Trésor de la Langue Française informatisé* (ATILF - CNRS & Université de Lorraine, s. d.). Pour tous ces noms, la condition de sélection retenue stipule qu'il doit exister un patron constructionnel de type $V \rightarrow N-x$, garanti par l'existence d'au moins deux cas de $N-x$ strictement analysables comme dérivés d'un verbe, mono-sémiques, et du même type sémantique que les cas examinés. L'identification préalable de ces patrons de construction s'est faite sur la base des informations disponibles dans la littérature (Dubois, 1962; Thiele, 1987; Dubois & Dubois-Charlier, 1999; Salvadori & Huyghe, 2023) et la lexicographie (*Le Robert méthodique*, Rey-Debove, 1985; *Le Petit dictionnaire des suffixes du français*, Éditions Le Robert, s. d.-b). Les noms d'instrument en *-ette* doublement analysables comme *moulinette* ('objet servant à mouliner') sont ainsi conservés, car il existe les noms *allumette* ('objet servant à allumer') et *chaufferette* ('objet servant à chauffer') pour lesquels une base nominale est exclue. Il en va de même pour les noms d'agent en *-aire* comme *suicidaire* et *protestataire*, du fait de l'existence de noms d'agent comme *signataire* et *locataire*, identifiés comme des substantifs mais non comme des adjectifs dans les dictionnaires.

Enfin, une série de conditions sont spécifiquement destinées aux formes converses :

C5 - Formes infinitives Les infinitifs substantivés (p. ex. *sourire_V* / *sourire_N*, *déjeuner_V* / *déjeuner_N*) sont conservés d'office.

C6 - Thème 0 Une contrainte sémantique a été appliquée pour les convertis de forme simple : ils sont conservés par défaut s'ils dénotent une éventualité dans un de leurs sens au moins (p. ex. *chasser/chasse, danser/danse, marcher/marche*), et exclus s'ils renvoient à une entité (p. ex. *clouer/clou, poivrer/poivre, singer/singe*).

C7 - Thèmes 12 et 13 Les conversions sur thèmes 12 (p. ex. *arriver/arrivée, conduire/conduite, venir/venue*) et 13 (p. ex. *corrélér/corrélat, instituer/institut, substituer/substitut*) sont conservées si l'existence d'un patron constructionnel autonome peut être garanti. À titre d'exemple, les noms dénotant des entités animées patientives (p. ex. *blesé, mutilé*) ne sont pas retenus, car ils peuvent être analysés comme des noms dérivés par conversion d'adjectifs, eux-mêmes issus de verbes par conversion (Tribout, 2010, p. 85-86).

Au final, l'échantillon complet comptabilise 5'274 paires verbe-nom. Le nombre total de paires par procédé morphologique est indiqué dans le tableau 2.2 ($M = 115$, $ET = 178$).

Procédé	Nombre de paires	Procédé	Nombre de paires
-ade	110	-eur (fém.)	7
-age	688	-eur (masc.)	780
-ail	3	-eure	50
-aille	15	-euse	485
-ain	7	-in	14
-aire (fém.)	40	-ine	4
-aire (masc.)	63	-ing	27
-aison	66	-ion	522
-ance/-ence	124	-is	71
-ant/-ent	117	-ise	3
-ante/-ente	88	-ment	421
-ard	58	-oir/-oire	139
-arde	31	-oire	24
-asse	5	-on	73
-eau	16	-onne	10
-elle	10	-ot	11
-er	16	-ote	9
-er/-ier	86	-rice	97
-ère/-ière	76	-ure	165
-eresse	4	Conversion - Thème 0	198
-erie	226	Conversion - Thème 12	104
-et	23	Conversion - Thème 13	64
-ette	89	Conversion - Infinitif	35

TABLEAU 2.2 – Nombre de paires verbe-nom par procédé morphologique dans la base de données finale

2.2 Principes d'analyse

Il s'agissait, suite à l'échantillonnage, d'analyser finement et de manière systématique la sémantique des verbes et des noms ainsi récoltés. Nous présentons dans ce qui suit les trois groupes de propriétés sémantiques décrites (cf. chapitre 1), à savoir :

- (i) le type sémantique du nom, selon une double typologie ontologique et relationnelle (section 2.2.1);
- (ii) l'aspect lexical du verbe et du nom (section 2.2.2), découpé en plusieurs traits (dynamacité, durativité, télicité et post-phase);
- (iii) les rôles sémantiques pouvant être assignés aux arguments du verbe et du nom (section 2.2.3).

De plus amples informations sur les différentes propriétés examinées et sur la procédure d'analyse sémantique sont disponibles dans un guide d'annotation spécifique (Salvadori et al., 2023), également fourni dans la documentation complémentaire de la thèse ²⁹.

Fondamental pour cette étude, l'examen de l'ambiguïté des noms et des verbes apparentés a aussi fait l'objet d'une attention particulière (section 2.2.4). L'analyse sémantique proposée repose sur l'hypothèse que les procédés de formation des mots s'appliquent à des lexèmes, entendus comme des éléments spécifiés sémantiquement (cf. Mel'čuk, 1993; Fradin & Kerleroux, 2003). Dans le cadre de ce travail, un nom ou un verbe ambigu correspond de fait à différents lexèmes, que cette ambiguïté relève de la polysémie ou de l'homonymie.

2.2.1 Type sémantique des noms

Les classifications proposées dans la littérature pour décrire la sémantique des noms déverbaux (Ježek, 2008; Fradin, 2012; Melloni, 2011; Lieber, 2016; *inter alia*) n'établissent pas toujours une distinction claire entre les informations relevant de la nature du référent (p. ex. événement, état, entité animée) et celles se rapportant à la relation avec l'éventualité dénotée par le verbe de base (p. ex. patient, résultat, instrument). Ces composantes sémantiques, respectivement ontologiques et relationnelles, sont pourtant de nature différente, compatibles plutôt que mutuellement exclusives, et en partie au moins indépendantes (Huyghe, 2015, 2021). Comme en témoignent les paires en (44)-(45), les nominalisations renvoyant à un artéfact peuvent par exemple exprimer différentes relations avec l'éventualité dénotée par leur verbe de base (RÉSULTAT, INSTRUMENT, LIEU). *A contrario*, les noms exprimant le résultat d'une éventualité peuvent dénoter différents types de référents (Artéfact, État, Entité animée) ³⁰.

29. Pour rappel, cette documentation est disponible à l'adresse suivante : https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2.

30. Nous conserverons la mise en forme suivante pour le reste de ce travail : les types relationnels sont indiqués en petites majuscules (p. ex. INSTRUMENT, RÉSULTAT, THÈME, TRANSPOSITION), et les types ontologiques débutent par une majuscule standard (p. ex. Artéfact, État, Événement, Objet cognitif).

- (44) a. *croquer* → *croquis* [Artéfact-RÉSULTAT]
 b. *raser* → *rasoir* [Artéfact-INSTRUMENT]
 c. *garer* → *garage* [Artéfact-LIEU]
- (45) a. *croquer* → *croquis* [Artéfact-RÉSULTAT]
 b. *énervé* → *énervement* [État-RÉSULTAT]
 c. *créer* → *créature* [Entité animée-RÉSULTAT]

Nous présentons brièvement les typologies sémantiques dédiées aux composantes ontologiques et relationnelles ci-dessous.

Types ontologiques

15 types ontologiques simples ont été distingués sur la base de tests distributionnels proposés dans la littérature (Godard & Jayez, 1996; Flaux & Van de Velde, 2000; Huyghe, 2015; Haas et al., 2023; *inter alia*). Un nom peut par exemple se voir attribuer le type ontologique Événement s'il admet d'être employé comme objet des verbes *effectuer* (46a), *procéder à* (46b) et *accomplir* (46c), ou comme sujet des verbes *se produire* (47a) et *avoir lieu* (47b) :

- (46) a. *La mécanicienne a effectué une **réparation** difficile.*
 b. *La start-up a procédé à un **licenciement** économique.*
 c. *Mon frère a accompli une **formation** spécifique.*
- (47) a. *Un **éboulement** s'est produit au sud de la ville.*
 b. *Une **fusillade** a eu lieu hier matin.*

Les noms d'Artéfact sont identifiés sur la base de leur capacité à être utilisés comme objet des verbes *fabriquer* (48a) et *déchirer* (48b), ou des verbes *construire* (48c) et *confectionner* (48d) suivis de divers compléments les caractérisant d'un point de vue physique :

- (48) a. *Elle a fabriqué une **génératrice** dans son garage.*
 b. *J'ai déchiré mon **attestation** ce matin.*
 c. *Ils ont construit des **logements** gris en béton.*
 d. *Son frère a confectionné des **roulades** au fromage.*

Plus marginalement, les noms d'Objet financier doivent pouvoir être intégrés dans les structures *verser le N en euros* (49a) ou *obtenir un N modique* (49b) :

- (49) a. *Nous avons versé la **redevance** en euros.*
 b. *Vous avez obtenu un **rendement** modique.*

D'un point de vue méthodologique, les différents tests distributionnels ont été appliqués à un sens nominal à la fois, suivant un arbre de décision proposé par Haas et ses collègues

2.2. Principes d'analyse

(2023). Les noms n'étant pas spécifiés ontologiquement, comme *échappatoire*, *composant* ou *mélange*, ont reçu l'étiquette N/A.

Outre les 15 types simples, 7 catégories complexes composées de deux types ontologiques simples ont également été incluses dans la classification pour rendre compte des noms présentant une structure sémantique hybride (cf. Pustejovsky, 1995; Cruse, 1995; Godard & Jayez, 1996; Kleiber, 1999; Asher, 2011; *inter alia*). Pour rappel (cf. section 1.2.1), ces derniers se distinguent des polysèmes standards par le fait qu'ils admettent la coprédication, i.e. qu'ils peuvent être insérés dans des constructions grammaticales où des prédicats de types sémantiques distincts s'appliquent conjointement à un argument sans pour autant générer d'effet de zeugme. Comme le montre la phrase en (50), le nom *discussion* se combine par exemple aisément avec des prédicats caractéristiques à la fois des noms d'Événement (*avoir lieu*) et des noms d'Objet cognitif (*porter sur*). Les deux interprétations ne s'excluant pas mutuellement, ce dérivé peut par conséquent se voir attribuer le type complexe Événement*Objet cognitif.

(50) *La **discussion** qui a eu lieu hier à huis clos portait également sur les modalités de suivi, dans la perspective d'une action commune future.* (web)

Les 15 types simples et 7 types complexes composant la classification ontologique sont présentés en intégralité dans le tableau 2.3.

Type ontologique simple	Abréviation	Exemple
Artéfact	Art	<i>bouilloire</i>
Domaine	Dom	<i>jardinage</i>
Entité animée	Anm	<i>collaboratrice</i>
État	Sta	<i>agacement</i>
Événement	Evt	<i>accouchement</i>
Institution	Ins	<i>association</i>
Maladie	Dis	<i>pelade</i>
N/A	N/A	<i>échappatoire</i>
Objet cognitif	Cog	<i>corrélat</i>
Objet financier	Fin	<i>redevance</i>
Objet naturel	Nat	<i>nageoire</i>
Phénomène	Phn	<i>senteur</i>
Propriété	Ppt	<i>persévérance</i>
Quantité	Qua	<i>lichette</i>
Temps	Tim	<i>échéance</i>
Type ontologique complexe	Abréviation	Exemple
Artéfact*Cognitif	Art*Cog	<i>circulaire</i>
Artéfact*Institution	Art*Ist	<i>restaurant</i>
Événement*Cognitif	Cog*Evt	<i>témoignage</i>

Événement*État	Evt*Sta	<i>disparition</i>
Événement*Financier	Evt*Fin	<i>paiement</i>
Événement*Naturel	Evt*Nat	<i>inflammation</i>
Événement*Phénomène	Evt*Phn	<i>crissement</i>

TABLEAU 2.3 – Types ontologiques

La typologie ontologique présente une autre caractéristique notable : la possibilité pour presque toutes les catégories mentionnées³¹ d'être associées à une étiquette additionnelle Collectif (abrégée *Coll*), utilisée pour indiquer que le dérivé concerné possède une référence plurielle lorsqu'il est employé au singulier (cf. Flaux, 1999; Lammert & Lecolle, 2014; de Vries, 2021; *inter alia*). Le nom *assistance* (51a), par exemple, dénote un groupe de personnes et non pas un individu isolé; *naissain* (51b), un ensemble de mollusques à l'état larvaire et non pas une seule larve; et *braderie* (51c), une foire où plusieurs personnes bradent des objets et non pas une transaction unique.

- (51) a. *assister* → *assistance* [Entité animée-Collectif]
 b. *naître* → *naissain* [Objet naturel-Collectif]
 c. *brader* → *braderie* [Événement-Collectif]

La catégorie Événement-Collectif (p. ex. *braderie*, *manifestation*, *ruée*, *tuerie*) a en particulier nécessité la création et la mise au point de tests idoines, ce type sémantique restant très peu documenté dans la littérature (mais voir Thornton, 2008 sur les reduplications en italien et Salvadori & Huyghe, 2023 sur les noms déverbaux suffixés en français). L'évaluation porte notamment sur la nature des arguments, qui renvoient volontiers à une pluralité d'individus :

- (52) a. *La manifestation des grévistes de la centrale n'avait pas d'incidence sur les opérations de déchargement-rechargement du combustible.* (web)
 b. *?la manifestation du gréviste*

Types relationnels

La typologie relationnelle comprend 18 types représentant des participants potentiels à des éventualités. Adaptés de travaux existants sur les rôles sémantiques, en particulier *VerbNet* (Kipper-Schuler, 2005) et *LIRICS* (Petukhova & Bunt, 2019)³², ils sont identifiés au moyen de définitions conceptuelles (en l'absence de tests distributionnels applicables). Par exemple, le nom *cambricoleur* peut se voir attribuer le type relationnel AGENT, car il désigne une personne effectuant intentionnellement l'action dénotée par le verbe *cambrioter*. Le nom *centrifugeuse* se voit quant à lui assigner le type INSTRUMENT, puisqu'il renvoie à une

31. La seule exception est le type QUANTITÉ, qui peut déjà impliquer une pluralité d'éléments.

32. Voir également le numéro spécial de *Computational Linguistics* édité par Márquez et ses collègues (2008) et l'ouvrage collectif de Palmer et ses collègues (2010) au sujet de l'identification de rôles sémantiques.

2.2. Principes d'analyse

entité manipulée dans le but d'effectuer l'action dénotée par *centrifuger*. La liste complète des types relationnels et des différentes définitions est présentée dans le tableau 2.4.

En complément des rôles sémantiques majeurs comme AGENT, INSTRUMENT ou encore THÈME, la classification inclut également une catégorie TRANSPOSITION servant à rendre compte des dérivés exprimant plus ou moins le même type d'éventualité que leur base (cf. Beard, 1995; Spencer, 2013; Ten Hacken, 2021; Lieber, 2015; *inter alia*). Le nom *licenciement*, par exemple, renvoie à une action peu ou prou équivalente à celle de sa base *licencier*, tout comme *atterrissage* par rapport à *atterrir*, et *falsification* à l'aune de *falsifier*. La transposition est abordée sous un angle relativement large dans le cadre de ce travail. Elle correspond schématiquement à la préservation du trait aspectuel de dynamicité entre la base et le nom (cf. section 2.2.2), que ceux-ci soient dynamiques (53) ou statifs (54).

- (53) a. **Construire** ce chalet a pris du temps.
 b. La **construction** de ce chalet a pris du temps.
- (54) a. Il n'apprécie pas que Jean **se méfie** de lui.
 b. Il n'apprécie pas la **méfiance** de Jean envers lui.

Type relationnel	Abréviation	Définition	Exemple
AGENT	AGT	Entité effectuant intentionnellement une action	<i>forger/forgeron</i>
BÉNÉFICIAIRE	BEN	Entité qui reçoit ou est dépossédée de quelque chose, ou qui est avanta-gée ou désavantagée par un événe-ment ou un état	<i>hériter/héritier</i>
CAUSE	CAU	Entité déclenchant ou à l'origine d'une éventualité (pas nécessaire-ment de manière intentionnelle)	<i>agglutiner/agglutinine</i>
CHEMIN	PTH	Trajectoire suivie lors d'un change-ment de localisation	<i>dévier/déviation</i>
DESTINATION	DES	Point d'arrivée lors d'un change-ment de localisation	<i>buter/butoir</i>
ÉTENDUE	EXT	Valeur extensive liée à un événe-ment, ou ampleur mesurable d'un état	<i>conteneur/contenance</i>
EXPÉRIENCEUR	EXP	Entité se trouvant ou entrant dans un état particulier en raison d'une stimulation psychologique, percep-tive ou physiologique	<i>adorer/adorateur</i>
INSTRUMENT	INS	Entité manipulée afin d'effectuer une action	<i>arroser/arrosoir</i>

Type relationnel	Abréviation	Définition	Exemple
LIEU	LOC	Entité servant de point de repère pour localiser une autre entité ou un événement	<i>fumer/fumoir</i>
MANIÈRE	MAN	Manière dont une action est effectuée ou intensité d'un état	<i>prononcer/prononciation</i>
PATIENT	PAT	Entité subissant un changement (potentiel) d'état ou de structure	<i>mourir/mourant</i>
PIVOT	PVT	Entité à laquelle est attribuée une propriété ou qui se trouve dans un état non stimulé	<i>composer/composant</i>
RÉSULTAT	RES	Entité créée lors d'un événement	<i>égratigner/égratignure</i>
SOURCE	SRC	Point de départ d'un changement de localisation	<i>plonger/plongeoir</i>
STIMULUS	STI	Entité qui provoque un état psychologique, perceptif ou physiologique	<i>épouvanter/épouvantail</i>
THÈME	THM	Entité qui se trouve à un certain endroit ou qui change de localisation	<i>traîner/traîneau</i>
TOPIQUE	TPC	Entité faisant l'objet d'une réflexion, d'une discussion ou d'une activité cognitive	<i>supposer/supposition</i>
TRANSPOSITION	TSP	Éventualité dénotée par le lexème-base	<i>atterrir/atterrissage</i>

TABLEAU 2.4 – Types relationnels

Autre particularité de cette typologie : la possibilité de capturer les acceptions métaphoriques, comme *lacet*₂ en (55), et métonymiques, comme *planteur*₂ en (56).

- (55) a. *lacer* → *lacet*₁ 'cordon étroit' [INSTRUMENT]
b. *lacet*₁ → *lacet*₂ 'zigzag' [(INSTRUMENT)·FIGURATIF]
- (56) a. *planter* → *planteur*₁ 'agriculteur' [AGENT]
b. *planteur*₁ → *planteur*₂ 'cocktail' [(AGENT)·FIGURATIF]

Ces lectures, qui constituent des extensions d'autres interprétations nominales, n'entretiennent pas de lien sémantique direct avec les lexèmes-bases. Par convention, le choix est fait d'annoter relationnellement les acceptions de ce type par une double étiquette composée (i) du type relationnel du sens dont elles dérivent, qui sert ainsi à identifier leur source, et (ii) d'une étiquette FIGURATIF (abrégée FIG), qui permet de rendre compte de leur absence de relation au verbe.

2.2.2 Aspect lexical

L'aspect lexical des verbes et des noms a été décomposé en quatre traits élémentaires, examinés indépendamment les uns des autres³³ : la dynamicité, la durativité, la télélicité et la post-phase. Nous discutons brièvement ces différents traits ci-dessous.

Dynamicité Dans le cadre de ce travail, un verbe ou un nom est considéré comme dynamique si l'éventualité qu'il dénote implique un quelconque changement, et comme statif si l'éventualité à laquelle il renvoie n'en implique pas (Pustejovsky & Batiukova, 2019, p. 159). Par exemple, le verbe *marcher* dans son sens 'se déplacer à pied' est dynamique, tandis que le verbe *posséder* dans son sens 'avoir en sa possession' est statif.

Durativité Un verbe ou un nom est considéré comme duratif si l'éventualité qu'il dénote « occupe du temps » ; il est ponctuel si l'éventualité à laquelle il renvoie est instantanée et peut être idéalisée « comme un point dans le temps³⁴ » (Kearns, 2011, p. 157). Le verbe *chanter* est par exemple duratif, par opposition au verbe *naître*, qui est ponctuel.

Télélicité Un verbe ou un nom est considéré comme télélique si l'éventualité qu'il dénote implique un *telos*, i.e. un point culminant qui actualise l'action décrite (cf. Garey, 1957; Vendler, 1967; Rothstein, 2004; Demonte & McNally, 2012; Filip, 2012; *inter alia*). Le verbe *fabriquer* – télélique – se distingue par exemple du verbe *jardiner* – atélélique – en raison du fait qu'il renvoie à un procès comportant une fin inhérente.

Certains verbes et noms se distinguent par ailleurs par leur télélicité intrinsèquement variable, décrivant le changement progressif d'une propriété fondamentale d'un participant dans le cadre d'une situation donnée (Dowty, 1979; Hay et al., 1999; Kearns, 2007; Kennedy & Levin, 2008; Rothstein, 2008; Martínez Vera, 2021; *inter alia*). Dans la phrase suivante, cette propriété fondamentale est par exemple la température du mets pour le verbe *refroidir* :

- (57) a. *Quand le contenu du plat **a refroidi**, égouttez-le dans une passoire placée au-dessus d'une casserole.* (web)

Les lexèmes à télélicité variable ont cela de particulier qu'ils sont compatibles aussi bien avec des compléments en *en* (58a) qu'en *pendant* (58b), associés respectivement à la télélicité et à l'atélélicité :

- (58) a. *L'appareil **a refroidi** en 20 minutes.*
b. *L'appareil **a refroidi** pendant 20 minutes.*

33. Il existe malgré tout des interdépendances logiques entre les différentes valeurs aspectuelles (cf. Salvadori et al., 2023). Par exemple, un verbe statif (p. ex. *connaître*, *posséder*) ne peut pas être télélique.

34. [*En anglais dans le texte*] « A durative event occupies time, in contrast with a non-durative event which is idealized to a point in time ».

De plus, ils acceptent volontiers d'apparaître aux côtés d'adverbes intensificateurs comme *beaucoup* et *considérablement* (59) :

(59) *L'appareil a {beaucoup / considérablement} refroidi.*

Post-phase La post-phase peut être définie comme le fait, pour une éventualité dynamique, de comporter un état résultant duratif (cf. Apothéloz, 2008; Piñón, 1997, 1999; Fradin, 2011; Haas & Jugnet, 2013; *inter alia*). Le complément introduit par *pendant* en (60), par exemple, ne porte non pas sur l'action de disparaître, mais sur la situation qui en résulte. Aussi une post-phase peut-elle être attribuée au verbe *disparaître* dans sa lecture 'ne plus être vu'.

(60) *Un petit Américain de 3 ans a disparu pendant deux jours avant d'être retrouvé sain et sauf dans un bois.* (web)

À l'inverse, le complément introduit par *pendant* en (61) porte sur l'action de patiner. On peut donc estimer que le verbe *patiner* dans son sens 'glisser sur la glace' ne comporte pas de post-phase.

(61) *Elle a patiné pendant des heures sans se fatiguer.* (web)

Pour résumer cet aperçu, les valeurs admises pour les différents traits aspectuels sont les suivantes : verbe/nom dynamique ou statif pour la dynamicité; verbe/nom duratif ou ponctuel pour la durativité; verbe/nom téléique, atéléique ou à téléicité variable pour la téléicité; et verbe/nom avec ou sans post-phase pour la post-phase.

Les quatre traits aspectuels considérés ont été analysés au moyen de tests linguistiques inspirés de la littérature (Vendler, 1967; Dowty, 1979; Rothstein, 2004; Haas et al., 2008; Filip, 2012; *inter alia*). Même s'ils partagent parfois quelques similitudes, ces derniers restent fondamentalement distincts pour les verbes et les noms. Un verbe peut par exemple être considéré comme duratif (*vs* ponctuel) s'il est compatible avec l'une des deux expressions présentées en (62).

- (62) a. X a {commencé à / continué de / arrêté de} V (Y)
Maurice a continué de chanter aux premières lueurs de l'aube et aucune conciliation n'a pu avoir lieu entre les parties. (web)
- b. X a Vé (Y) {en / pendant} x temps
Kjell Elvis a chanté pendant plus de 50 heures, pulvérisant de plus de sept heures le précédent record du monde. (web)

Dans le domaine nominal en revanche, un dérivé peut être considéré comme duratif s'il admet d'être inséré dans au moins une des structures en (63) :

2.2. Principes d'analyse

- (63) a. Le N a duré x temps
*La **conférence** a duré une semaine, entre le 4 et le 11 février 1945.* (web)
- b. un N de x temps
*Dans le cadre de l'Année internationale des langues autochtones 2019, une **conférence** internationale de trois jours sera organisée.* (web)
- c. Le N s'est déroulé à tel endroit
*Cette **conférence** s'est déroulée, samedi, à l'Agora.* (web)
- d. x temps de N (au singulier)
*Dijon : deux jours de **conférence** sur les mineurs non accompagnés* (web)

Concernant la télicité, les verbes transitifs ont été systématiquement analysés avec un argument interne délimité, certains verbes téliques comme *imprimer* (64a) pouvant en effet recevoir une lecture atélique d'activité lorsqu'ils sont employés avec un argument non circonscrit (64b) (cf. Verkuyl, 1993; Rothstein, 2004).

- (64) a. *Matthieu a **imprimé** le livre vert.*
- b. *Matthieu a **imprimé** des livres verts.*

Comme le montre le tableau 2.5, les quatre traits aspectuels examinés peuvent au besoin servir à reconstituer les classes vendlériennes traditionnelles³⁵ d'accomplissement, d'achèvement, d'activité et d'état (Vendler, 1967), ainsi que quatre catégories supplémentaires (accomplissement gauche, achèvement gauche, achèvement graduel et achèvement graduel gauche). Ces dernières classes découlent à la fois (i) de l'inclusion de la post-phase comme trait aspectuel distinctif et (ii) de la possibilité d'indiquer qu'un verbe donné possède une télicité intrinsèquement variable.

2.2.3 Rôles sémantiques

La structure argumentale des verbes et des noms dérivés a été décrite en termes de rôles sémantiques assignés aux différents arguments, les extensions maximales étant systématiquement privilégiées dans l'analyse. Par exemple, la structure encodée pour le verbe *lire*

35. La littérature est divisée quant à savoir si les sémelfactifs (p. ex. *cligner, frapper, tousser*) constituent (Comrie, 1976; Smith, 1991) ou non (Dowty, 1979; Rothstein, 2004) une cinquième classe aspectuelle majeure aux côtés des activités, des achèvements, des accomplissements et des états. Les verbes et noms concernés se distinguent par leur capacité à dénoter des occurrences uniques d'événements – se rapprochant ainsi des achèvements – tout en étant habituellement considérés comme des actions atéliques, pouvant avoir une interprétation d'activité (Rothstein, 2004, p. 184-185). Dans le cadre de ce travail, ces lexèmes ont été traités comme des achèvements standards, en vertu du fait qu'ils correspondent à des événements dynamiques et ponctuels (ia). Les lectures atéliques qu'ils peuvent parfois recevoir (ib) ont été considérées comme itératives et n'ont pas été analysées en tant que telles.

- (i) a. *À quelques centimètres du visage de Maï, un homme assis en tailleur s'inclina en avant, **toussa** une fois dans ses mains, puis se renversa en arrière, en appui sur ses coudes.* (web)
- b. *Ben, je l'ai rencontré en 1966 à Varsovie, j'avais reçu un prix, j'étais dans un hôtel et quelqu'un a **toussé** toute la nuit dans la chambre d'à côté.* (web)

Classe aspectuelle	Dynamacité	Durativité	Télicité	Post-phase	Exemple verbal	Exemple nominal
Accomplissement	Dynamique	Duratif	Télique	Sans post-phase	<i>imprimer</i>	<i>rasage</i>
Accomplissement gauche	Dynamique	Duratif	Télique	Avec post-phase	<i>boutonner</i>	<i>défixation</i>
Achèvement	Dynamique	Ponctuel	Télique	Sans post-phase	<i>gober</i>	<i>craquement</i>
Achèvement gauche	Dynamique	Ponctuel	Télique	Avec post-phase	<i>kidnapper</i>	<i>emprisonnement</i>
Achèvement graduel	Dynamique	Duratif	Variable	Sans post-phase	<i>mûrir</i>	<i>calcination</i>
Achèvement graduel gauche	Dynamique	Duratif	Variable	Avec post-phase	<i>refroidir</i>	<i>russification</i>
Activité	Dynamique	Duratif	Atélique	Sans post-phase	<i>covoiturer</i>	<i>clapotis</i>
État	Statif	Duratif	Atélique	Sans post-phase	<i>rutiller</i>	<i>discordance</i>

TABLEAU 2.5 – Classes aspectuelles et combinaison de valeurs correspondant à chacune d'elles

2.2. Principes d'analyse

comprend à la fois un sujet et un objet (65a), bien que dans certains contextes, ce dernier puisse être omis (65b). Il en va de même pour son dérivé *lecture* dans son sens événementiel (65c).

- (65) a. *Pierre **lit** le tract.*
b. *Pierre **lit**.*
c. *la **lecture** du tract par Pierre*

Seuls les arguments à la fois syntaxiques et sémantiques ont été pris en compte. Ainsi, les arguments sémantiques mais non syntaxiques n'ont pas été inclus dans l'annotation, et aucun rôle instrumental n'a par exemple été retenu pour le verbe *gribouiller*. S'il peut effectivement paraître difficile de gribouiller quelque chose sans recourir à un instrument, le fait qu'un syntagme instrumental comme *avec un stylo rouge* en (66b) puisse être associé à *le faire* en reprise du verbe indique qu'il ne s'agit pas d'un argument syntaxique (cf. Bonami, 1999).

- (66) a. *Marie **a gribouillé** quelques notes avec un stylo rouge.*
b. *Marie **a gribouillé** quelques notes, et elle l'a fait avec un stylo rouge.*

Les arguments syntaxiques mais non sémantiques n'ont pas non plus été pris en considération dans l'analyse. Aucun rôle bénéfactif n'a par exemple été attribué au verbe *dire* : bien qu'une reprise en *le faire* avec le destinataire soit impossible (67b), ce qui indique qu'il s'agit bien d'un argument syntaxique, le verbe *dire* n'inclut pas nécessairement un argument bénéfactif dans son sémantisme.

- (67) a. *Jean **a dit** une bêtise à son fils.*
b. *?Jean **a dit** une bêtise, et il l'a fait à son fils.*

S'agissant de l'évaluation des arguments nominaux, seuls ceux de dérivés dénotant une éventualité ont été évalués d'office. Les arguments de nominalisations renvoyant à une entité ont été annotés si et seulement si leur rôle est du type de ceux que peut assigner un verbe. Les arguments strictement nominaux, dépendant par exemple de relations de partie-tout (*un composant de cet appareil*), de membre-collection (*un bataillon de soldats*) ou de relations interpersonnelles (*le chiard de Pierre*), n'ont pas été pris en compte.

Les rôles sémantiques assignés aux arguments verbaux et nominaux sont sensiblement les mêmes que ceux employés pour les types relationnels (cf. section 2.2.1 et tableau 2.3), à l'exception du type TRANSPOSITION, qui ne correspond pas à un rôle sémantique. Fondée sur la lexicographie et des recherches en corpus, l'attribution de ces rôles a été effectuée pour des entrées lexicales, impliquant la prise en compte de toutes les variantes possibles pour chaque type d'argument. Par exemple, le rôle de cause a été attribué par défaut au sujet du verbe *casser* (68a), bien que dans certains contextes (68b), celui-ci puisse être volitionnel.

- (68) a. *Dans le quartier de Villeneuve-les-Salines, le vent **a cassé** de nombreuses branches.*
(web)
b. *Alors qu'il venait d'obtenir un aménagement de peine, un détenu **a cassé** son bracelet électronique, saccagé son appartement et menacé de mort des policiers.*
(web)

Cela signifie en substance que pour déterminer le rôle d'un argument donné, les annotateur.ices ont examiné un large éventail de scénarios et ne se sont pas cantonné-es à des situations prototypes impliquant exclusivement des entités animées.

2.2.4 Ambiguïté

Omniprésente dans la nominalisation (cf. Rainer, 2014; Lieber, 2018; *inter alia*), l'ambiguïté lexicale a fait l'objet d'une attention toute particulière lors de l'analyse sémantique. Les différents sens des noms et des verbes apparentés – toujours illustrés par un exemple de corpus dans la base de données – ont été identifiés par la variation d'au moins une des valeurs attribuées aux propriétés annotées (base verbale, type sémantique, aspect lexical, rôles des arguments). Deux interprétations ont ainsi été assignées au nom *séchoir* en vertu de sa compatibilité avec différents types relationnels : INSTRUMENT (69a) et LIEU (69b).

- (69) a. *Au cours du processus de séchage, la température à l'intérieur du local où est installé le **séchoir** augmente.* (web)
b. *Le gîte a une histoire originale car c'était autrefois un **séchoir** à chanvre.* (web)

Deux sens ont également été attribués à *ébaucheur*, ambigu entre deux types ontologiques : Entité animée (70a) et Artéfact (70b).

- (70) a. *Seul l'appareilleur avait la parfaite connaissance de ces données et c'est sous les ordres de ce spécialiste hautement qualifié que travaillaient les **ébaucheurs**.* (web)
b. *Le forêt en spirale pour bois est équipé d'une pointe de centrage et d'un **ébaucheur**.*
(web)

Une analyse complète est présentée dans le tableau 2.6 pour le nom *vol* et le verbe apparenté *voler*. Chaque colonne correspond à un sens nominal et représente une entrée dans la base de données. Au total, trois sens nominaux et deux sens verbaux se différenciant notamment par leurs types sémantiques, leur aspect et les rôles sémantiques qu'ils convoquent ont été distingués et appariés selon le principe de la plus grande proximité sémantique, étant entendu que les processus de dérivation s'appliquent à des éléments spécifiés sémantiquement (cf. Mel'čuk, 1993; Fradin & Kerleroux, 2003). Le sens *voler*₁ 'dérober' (71a) a ainsi été associé à *vol*₁ 'action de dérober' (71b), tandis que *voler*₂ 'se déplacer' (72a) a été mis en relation avec *vol*₂ 'action de se déplacer' (72b) et *vol*₃ 'ensemble d'animaux et d'insectes se déplaçant' (72c) :

2.2. Principes d'analyse

- (71) a. *J'**ai volé** un morceau de tuyau de cuivre à mon chef de chantier il y a quelques semaines...* (web)
b. *Le **vol** du téléphone a eu lieu dans les transports en commun, alors que la jeune fille aidait une maman avec une poussette.* (web)
- (72) a. *La balle **a volé** pendant 13 secondes.* (web)
b. *Ce jeudi, la Suisse a organisé un **vol** spécial à destination du Nigéria.* (web)
c. *C'est toujours mieux qu'un **vol** de criquets ou qu'un essaim de guêpes [...].* (web)

Quelques remarques peuvent encore être formulées au sujet des alternances de construction dans le domaine verbal, dans la mesure où elles influencent l'analyse de l'appariement verbe-nom, et par la force des choses, celle de l'ambiguïté nominale. Par défaut, les verbes permettant des alternances syntaxiques systématiques ont été encodés comme des entrées lexicales uniques. Cela concerne notamment les formes réflexives comme (*se*) *regarder*, mais aussi les verbes que la littérature anglo-saxonne désigne sous le nom de *load verbs* (Levin, 1993). Par exemple, bien que deux structures argumentales soient possibles pour le verbe *charger* 'faire porter une charge' – *X charge Y de Z* (73a) et *X charge Y dans Z* (73b) –, seule l'une des deux constructions (en l'occurrence, la seconde) a été enregistrée dans la base de données.

- (73) a. *Les uns **chargent** les camions de nourriture, les autres grimpent dans les cabines.* (web)
b. *Les pêcheurs **chargent** les 200 kg de civelles vivantes dans le camion.* (web)

Certaines formes verbales réfléchies se distinguent aussi par leur autonomie lexicale (Barque et al., 2019), identifiable notamment par un changement d'aspect et des rôles sémantiques en jeu. Elles ont par conséquent été analysées comme des entrées distinctes, si tant est qu'au moins un sens nominal leur corresponde. Les phrases suivantes illustrent par exemple deux lectures ayant été relevées pour le verbe *frotter* : 'exercer une pression sur quelque chose' (74a) et 'avoir des relations avec quelque chose ou quelqu'un' (74b).

- (74) a. *Ensuite, on **frottait** le linge dans des cuvettes remplies d'eau froide, on le battait, on le rinçait et on l'essorait à la main.* (web)
b. *On **s'est frotté** à l'autonomie, à l'éducation et à la formation en travail social.* (web)

Un cas plus délicat est celui des alternances causatives-anticausatives, qui se caractérisent par une certaine systématisme (voir p. ex. Heidinger, 2010 et Labelle & Doron, 2010 pour une analyse des propriétés des anticausatifs en français). L'option choisie a été de séparer ces alternances en deux entrées, en raison notamment de l'existence de doublons morphologiques à lecture agentive/non agentive sélectionnant l'une ou l'autre interprétation (Kelling, 2001; F. Martin, 2010; Fradin, 2016; *inter alia*). En français, les noms *émiettage* (75a) et *émiet-*

Dérivé	<i>vol</i> ₁ 'action de dérober'	<i>vol</i> ₂ 'action de se déplacer'	<i>vol</i> ₃ 'ensemble d'animaux ou d'insectes se déplaçant'
Type ontologique	Événement	Événement	Entité animée-Collectif
Type relationnel	Transposition	Transposition	Agent
Dynamicité	Dynamique	Dynamique	–
Durativité	Ponctuel	Duratif	–
Télicité	Télique	Atélique	–
Post-phase	Sans post-phase	Sans post-phase	–
Rôle du 1 ^{er} argument	Thème	Thème	–
Rôle du 2 ^e argument	Bénéficiaire	–	–
Rôle du 3 ^e argument	Agent	–	–
Base verbale	<i>voler</i> ₁ 'dérober'	<i>voler</i> ₂ 'se déplacer'	<i>voler</i> ₂ 'se déplacer'
Dynamicité	Dynamique	Dynamique	Dynamique
Durativité	Ponctuel	Duratif	Duratif
Télicité	Télique	Atélique	Atélique
Post-phase	Sans post-phase	Sans post-phase	Sans post-phase
Rôle du sujet	Agent	Thème	Thème
Rôle de l'objet	Thème	–	–
Rôle de l'oblique	Bénéficiaire	–	–

TABLEAU 2.6 – Analyse sémantique complète du nom ambigu *vol* et du verbe apparenté *voler*. Trois sens nominaux et deux sens verbaux ont été distingués et appariés (*vol*₁ / *voler*₁, *vol*₂ / *voler*₂, *vol*₃ / *voler*₂)

2.3. Procédure d'analyse

tement (75b) semblent par exemple dériver respectivement des formes *émietter* (76a) et *s'émietter* (76b).

- (75) a. *Ensuite, un **émiettage** du sol s'avère indispensable. Cette opération peut se faire à l'aide d'un râteau ou d'une griffe.* (web)
b. *Nous assistons, en fait, à un **émiettement** du service public d'éducation nationale.* (web)
- (76) a. *Puis elle sortait le plat du four, **émiettait** le contenu avec deux fourchettes et servait le Schmarren accompagné de laitue.* (web)
b. *Nos déchets, enfin, ont atteint la taille d'un continent dans l'océan, qui **s'émiette** peu à peu pour s'infiltrer dans nos assiettes.* (web)

Il s'ensuit que le verbe *humidifier* (77a) et son équivalent anticausatif *s'humidifier* (77b), par exemple, ont été chacun associés à un sens du dérivé *humidification* : respectivement *humidification*₁ 'action d'humidifier' (78a) et *humidification*₂ 'action de s'humidifier' (78b).

- (77) a. *Dimitri **a humidifié** la terre avec le vaporisateur.* (web)
b. *Selon l'Institut Fédéral pour l'Etude de la Neige et des Avalanches, le manteau neigeux **s'est humidifié** et fortement tassé jusqu'à environ 2000 m d'altitude [...].* (web)
- (78) a. *Pour ce faire, elle procède à l'**humidification**, puis au balayage ou à l'aspiration des déchets.* (web)
b. *S'ensuit une période où la tendance générale est à la baisse des températures et une **humidification** croissante du climat [...].* (web)

2.3 Procédure d'analyse

L'analyse sémantique a été réalisée par six personnes membres du projet général sur la sémantique des noms déverbaux, dont l'autrice du présent travail (cf. introduction du chapitre)³⁶. Le but étant de décrire les sens lexicaux des noms et des verbes examinés, elle s'est fondée sur l'examen de définitions lexicographiques disponibles dans *Le Petit Robert* (Éditions Le Robert, s. d.-a) et le *Trésor de la Langue Française informatisé* (ATILF - CNRS & Université de Lorraine, s. d.), ainsi que sur l'observation d'occurrences dans le corpus FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015) et sur la Toile, notamment pour les mots non documentés dans la lexicographie. Comme mentionné précédemment, un guide d'annotation détaillé (Salvadori et al., 2023) a par ailleurs été élaboré pour les besoins de la tâche. Nous présentons les deux phases de cette analyse dans les sous-sections suivantes.

36. Les personnes suivantes ont été impliquées (par ordre alphabétique) : Lucie Barque, Pauline Haas, Richard Huyghe, Alizée Lombard, Matthieu Monney et Justine Salvadori.

Catégorie grammaticale	Propriété	Accord brut	κ de Cohen	PABAK
Verbe	Dynamicité	0.99	0.90	0.98
	Durativité	0.86	0.53	0.74
	Télicité	0.73	0.55	0.61
	Post-phase	0.83	0.44	0.67
	Rôle du sujet	0.75	0.60	0.72
	Rôle de l'objet	0.73	0.66	0.71
	Rôle de l'oblique	0.84	0.44	0.82
Nom	Type ontologique	0.73	0.66	0.71
	Type relationnel	0.80	0.72	0.78
	Dynamicité	1	1	1
	Durativité	0.91	0.83	0.87
	Télicité	0.87	0.79	0.83
	Post-phase	0.91	0.83	0.87
	Rôle du 1 ^{er} argument	0.67	0.58	0.64
	Rôle du 2 ^e argument	0.81	0.64	0.79
	Rôle du 3 ^e argument	0.97	0.39	0.95
Moyenne		0.84	0.67	0.79

TABLEAU 2.7 – Taux d'accord entre annotateur·ices par propriété

κ	Niveau d'accord	κ	Niveau d'accord
<0.00	Pauvre	0.00-0.20	Inexistant
0.00-0.20	Minime	0.21-0.39	Minimal
0.21-0.40	Bon	0.40-0.59	Faible
0.41-0.60	Modéré	0.60-0.79	Modéré
0.61-0.80	Considérable	0.80-0.90	Fort
0.81-1.00	Presque parfait	0.90	Presque parfait

(a) Adapté de Landis et Koch (1977) (b) Adapté de McHugh (2012)

TABLEAU 2.8 – Deux échelles proposées pour l'interprétation du κ de Cohen

2.3.1 Évaluation du modèle

La première étape de l'analyse a consisté à contrôler la qualité du modèle d'annotation proposé en examinant l'accord entre annotateur·ices pour les 16 propriétés sémantiques présentées dans les sous-sections précédentes (type sémantique, aspect lexical, rôles sémantiques). À cet effet, 10 échantillons de 50 paires verbe-nom extraits de la base de données ont été annotés en aveugle par des équipes de deux personnes en variant la composition de ces binômes. L'accord pour chaque échantillon a ensuite été calculé à l'aide de deux indicateurs : (i) le κ de Cohen (Cohen, 1960), un coefficient fréquemment employé pour

2.3. Procédure d'analyse

mesurer l'accord entre deux personnes lors d'une évaluation qualitative et qui tient compte des accords fortuits; (ii) et le PABAK (*prevalence-adjusted bias-adjusted kappa*) de Byrt et ses collègues (1993), qui ne souffre pas de certains biais reprochés au κ de Cohen. Il est en effet bien connu que ce dernier dépend fortement de la prévalence de certaines catégories d'annotation, et peut de ce fait renvoyer un score excessivement faible même si l'accord est en réalité élevé (Feinstein & Cicchetti, 1990; Di Eugenio & Glass, 2004). Les résultats complets sont présentés dans le tableau 2.7. Selon les échelles proposées par Landis et Koch (1977) et McHugh (2012) et reproduites dans le tableau 2.8, il ressort de l'analyse que l'accord est globalement satisfaisant pour toutes les propriétés annotées.

La qualité de l'évaluation de l'ambiguïté a également été vérifiée, en suivant toutefois une procédure différente. En effet, la tâche ne consistait pas à sélectionner une étiquette parmi d'autres pour une acception d'un nom ou d'un verbe (p. ex. un type relationnel), mais à distinguer un certain nombre de sens pour une forme nominale donnée. Pour ce faire, des équipes composées de deux personnes ont également analysé en aveugle 10 échantillons tirés de la base de données (cf. section 2.1), comprenant chacun 50 paires verbe-nom. Un score de fiabilité a ensuite été calculé au moyen du coefficient de corrélation intra-classe (abrégé ICC en anglais)³⁷, qui permet de déterminer dans quelle mesure plusieurs personnes identifient un nombre de sens similaire pour un mot donné. Pour les dix échantillons, le taux de fiabilité brut se monte au final à 0.63, et l'ICC, à 0.54. Ce second score est modéré selon l'échelle proposée par Koo et Li (2016) : les valeurs d'ICC inférieures à 0.5 indiquent en effet une fiabilité médiocre; celles comprises entre 0.5 et 0.75, une fiabilité moyenne; celles comprises entre 0.75 et 0.9, une bonne fiabilité; et celles supérieures à 0.90, une excellente fiabilité. Ce résultat doit toutefois être considéré à l'aune de la nature de la tâche, réputée complexe (McCrae et al., 2022).

Soulignons pour conclure cette partie que les données ont été adjudiquées collégalement dans tous les cas de désaccord observés (que ce soit pour les propriétés sémantiques annotées ou pour le nombre de sens distingués).

2.3.2 Annotation

Il s'agissait, suite au contrôle de la qualité du schéma d'annotation, de procéder à l'analyse des données à proprement parler. Pour ce faire, un certain nombre de paires verbe-nom ont été attribuées aux six annotateur-ices. Ils et elles avaient pour consigne de procéder de manière individuelle à leur examen complet, ce qui consistait d'une part à identifier et à apparier les différents sens verbaux et nominaux, et d'autre part à sélectionner des valeurs pour toutes les propriétés examinées. Compte tenu de la complexité de la tâche, des mises en commun régulières ont été organisées avec les autres membres du projet de manière à pouvoir statuer collégalement sur les cas difficiles. Une fois l'analyse achevée,

³⁷ Il existe différents formes d'ICC (Koo & Li, 2016). Nous avons ici choisi l'ICC (2,1) dans la typologie de Shrout et Fleiss (1979).

trois personnes³⁸ – dont l’auteurice du présent travail – ont procédé à l’homogénéisation et au contrôle final des données afin de constituer une base de données cohérente, lisible et simple d’utilisation.

2.4 Bilan

Ce chapitre avait pour objectif de présenter la méthodologie que nous avons adoptée pour la collecte et l’analyse sémantique d’un vaste ensemble de nominalisations et de verbes apparentés (i.e. plus de 4’000 paires verbe-nom). Résumons ici plusieurs points clés :

Échantillonnage Plus de 50’000 paires verbe-nom ont été extraites automatiquement du corpus web FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015). Ces paires, qui représentent à la fois des formes lexicalisées et des néologismes, couvrent un large éventail de procédés morphologiques permettant la nominalisation des verbes en français, avec 46 suffixes et formes de conversion recensés. Sur cette base, une sélection manuelle a été effectuée pour aboutir à un échantillon final de 5’274 paires verbe-nom.

Grille d’analyse sémantique Une grille d’analyse sémantique détaillée a été mise au point pour rendre compte des types sémantiques que les nominalisationsinstancient, aussi bien sur un plan ontologique que sur un plan relationnel. Cette grille prend également en considération les propriétés aspectuelles et les rôles sémantiques des verbes et des noms dérivés, ainsi que l’ambiguïté des formes verbales et nominales. Au total, 16 propriétés sémantiques ont été retenues.

Procédure d’annotation La grille d’analyse sémantique a ensuite été appliquée à l’échantillon précédemment constitué. La fiabilité du modèle d’annotation proposé a été soigneusement vérifiée, et six annotateur·ices ont identifié et apparié les différents sens verbaux et nominaux, tout en sélectionnant les valeurs appropriées pour chacune des propriétés examinées. Dans un dernier temps, trois personnes ont procédé à l’homogénéisation de la base de données afin de garantir une certaine cohérence dans l’ensemble des annotations.

Originalité et contributions de la ressource La ressource ainsi constituée représente une nouveauté pour les travaux situés à l’interface de la morphologie et de la sémantique lexicale. Elle associe un volume important de données empiriques à une analyse qualitative approfondie des propriétés sémantiques des nominalisations et des verbes correspondants. Plus spécifiquement, la ressource présente plusieurs atouts pour l’étude de l’ambiguïté dans la nominalisation en français. Elle rend premièrement possibles l’identification des sens multiples en jeu et l’appariement des acceptions verbales et nominales. La description précise du sémantisme des lexèmes-bases permet par ailleurs d’évaluer comment les valeurs aspectuelles et les rôles sémantiques

38. Les personnes suivantes ont été impliquées (par ordre alphabétique) : Richard Huyghe, Justine Salvadori et Rossella Varvara.

2.4. Bilan

des verbes peuvent favoriser l'émergence de certains schémas d'ambiguïté dans le domaine nominal. Enfin, l'inclusion d'un maximum de procédés morphologiques utilisés pour la nominalisation de verbes nous offre la possibilité d'adopter une perspective systémique au regard du phénomène de l'ambiguïté dans le lexique déverbal.

Chapitre 3

Profil sémantique des noms déverbaux et des verbes de base

Ce chapitre vise à établir le profil sémantique des nominalisations et des verbes apparentés que nous avons analysés. Notre approche, essentiellement descriptive, est motivée par plusieurs questions clés relatives à la sémantique des noms déverbaux. Quels types sémantiques les nominalisations investissent-elles? Quelles valeurs aspectuelles et quels rôles sémantiques les dérivés et les verbes correspondants mettent-ils en jeu? Ces différentes propriétés sémantiques sont-elles interdépendantes? Enfin, en ce qui concerne l'ambiguïté des nominalisations, comment les sens verbaux et nominaux entrent-ils en relation?

La section 3.1 présente le comportement des lexèmes nominaux (3.1.1) et verbaux (3.1.2) à l'aune des différentes propriétés évoquées dans le chapitre 2, telles que l'aspect lexical ou les types ontologiques et relationnels. Cela nous conduira dans la section 3.2 à l'examen des interdépendances entre les variables collectées (p. ex. les traits d'aspect lexical). Enfin, nous déterminerons dans la section 3.3 de quelle façon les lexèmes nominaux et verbaux présentés précédemment s'associent les uns aux autres au sein de situations d'ambiguïté. Une fois ces quelques bases posées, nous pourrons alors explorer plus avant les sources potentielles de l'ambiguïté des noms déverbaux dans les chapitres 4 et 5.

3.1 Informations générales

Nous décrivons dans les deux sections suivantes les caractéristiques des noms (3.1.1) et des verbes (3.1.2) analysés en considérant plusieurs propriétés, telles que le nombre de sens qu'ils comportent, leurs traits d'aspect lexical, ou encore les rôles sémantiques qu'ils convoquent.

3.1. Informations générales

3.1.1 Noms

Nombre de sens

Parmi les 5'274 noms que compte la base de données, 3'252 (62 %) sont monosémiques et 2'022 (38 %) sont ambigus, pour un total de 8'206 sens. En moyenne, un nom a 1.56 sens ($ET = 0.89$). Si seuls les noms ambigus sont considérés, cette moyenne monte à 2.45 sens ($ET = 0.87$). La figure 3.1 présente la distribution des noms collectés en fonction du nombre de sens auxquels ils sont associés. Seuls 77 noms (1.5 %) sont liés à 5 sens ou plus.

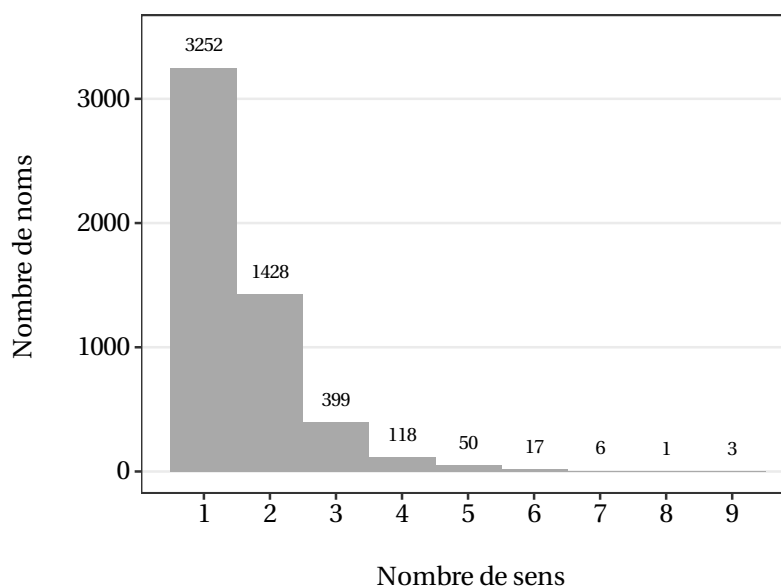


FIGURE 3.1 – Distribution des noms annotés en fonction du nombre de sens qu'ils comportent

La répartition des noms et des sens ainsi que le taux d'ambiguïté par procédé morphologique sont présentés dans le tableau 3.1. Les procédés morphologiques associés aux plus hauts taux d'ambiguïté, calculés en divisant le nombre de sens identifiés par le nombre de noms, sont la conversion sur thème 0 (2.1 sens par nom), la conversion sur thème 12 (2.05), *-ise* (2), *-ure* (1.92) et *-ment* (1.87). On peut se demander à propos de ces résultats si les taux d'ambiguïté élevés sont liés à la grande diversité sémantique des procédés, généralement admise pour la conversion notamment (voir p. ex. Plag, 1999, p. 221). Le rapport entre poly-fonctionnalité sémantique des procédés morphologiques et taux d'ambiguïté des dérivés sera étudié dans le chapitre 5.

Les 5 procédés morphologiques présentant les taux d'ambiguïté les plus bas sont *-eresse* (1), *-onne* (1), *-aire* (fém.) (1.02), *-er* (1.06) et *-eure* (1.06). À l'exception de *-er* qui forme surtout des noms d'agent et d'instrument monosémiques, il s'agit ici principalement de suffixes permettant la création de noms d'agent féminins (p. ex. *chasseresse*, *matonne*, *protestataire*, *manageure*).

Chapitre 3. Profil sémantique des noms déverbaux et des verbes de base

Procédé morphologique	Nombre de noms	Nombre de sens	Taux d'ambiguïté
<i>-ade</i>	110	161	1.46
<i>-age</i>	688	1230	1.79
<i>-ail</i>	3	5	1.67
<i>-aille</i>	15	21	1.4
<i>-ain</i>	7	9	1.29
<i>-aire</i> (fém.)	40	41	1.02
<i>-aire</i> (masc.)	63	69	1.10
<i>-aison</i>	66	113	1.71
<i>-ance/-ence</i>	124	196	1.58
<i>-ant/-ent</i>	117	147	1.26
<i>-ante/-ente</i>	88	98	1.11
<i>-ard</i>	58	67	1.16
<i>-arde</i>	31	34	1.10
<i>-asse</i>	5	6	1.2
<i>-eau</i>	16	25	1.56
<i>-elle</i>	10	16	1.6
<i>-er</i>	16	17	1.06
<i>-er/-ier</i>	86	97	1.13
<i>-ère/-ière</i>	76	92	1.21
<i>-eresse</i>	4	4	1
<i>-erie</i>	226	370	1.64
<i>-et</i>	23	35	1.52
<i>-ette</i>	89	128	1.44
<i>-eur</i> (fém.)	7	12	1.71
<i>-eur</i> (masc.)	780	1071	1.37
<i>-eure</i>	50	53	1.06
<i>-euse</i>	485	695	1.43
<i>-in</i>	14	16	1.14
<i>-ine</i>	4	6	1.5
<i>-ing</i>	27	41	1.52
<i>-ion</i>	522	918	1.76
<i>-is</i>	71	107	1.51
<i>-ise</i>	3	6	2
<i>-ment</i>	421	786	1.87
<i>-oir/-oire</i>	139	170	1.22
<i>-oire</i>	24	29	1.21
<i>-on</i>	73	93	1.27
<i>-onne</i>	10	10	1
<i>-ot</i>	11	12	1.09
<i>-ote/-otte</i>	9	13	1.44
<i>-rice</i>	97	108	1.11
<i>-ure</i>	165	316	1.92
Conversion - Thème 0	198	415	2.10

3.1. Informations générales

Procédé morphologique	Nombre de noms	Nombre de sens	Taux d'ambiguïté
Conversion - Thème 12	104	213	2.05
Conversion - Thème 13	64	77	1.20
Conversion - Infinitif	35	58	1.66

TABLEAU 3.1 – Répartition des noms et sens en fonction du procédé morphologique. Le taux d'ambiguïté, calculé en divisant le nombre de sens identifiés par le nombre de noms, est également précisé

Types sémantiques

Pour rappel, la sémantique des noms a été décrite au moyen d'une double classification ontologique et relationnelle. D'un point de vue ontologique, les 4 types les plus fréquents (79) sont Événement (27 % des lexèmes), Entité animée (26 %), Artéfact (14 %) et Événement*État (11 %), le reste des types (p. ex. Temps, Phénomène, Maladie) représentant chacun moins de 4 % des lexèmes nominaux.

- (79) a. *chamaillerie, flagellation, manœuvre* [Événement]
 b. *broutard, pondeuse, squatteur* [Entité animée]
 c. *décrottoir, imprimante, sifflet* [Artéfact]
 d. *accoutumance, fâcherie, patrimonialisation* [Événement*État]

S'agissant du plan d'analyse relationnel, les 4 types les plus fréquents (80) sont TRANSPOSITION (48 % des lexèmes), AGENT (25 %), INSTRUMENT (9 %) et RÉSULTAT (7 %). À l'inverse, 10 types relationnels (p. ex. ÉTENDUE, CHEMIN, SOURCE) représentent chacun moins de 1 % des données.

- (80) a. *coexistence, confrontation, étripade* [TRANSPOSITION]
 b. *commanditaire, contributeur, papotier* [AGENT]
 c. *décapant, étuveuse, questionnaire* [INSTRUMENT]
 d. *distillat, polycopie, sculpture* [RÉSULTAT]

Seule une minorité de lexèmes (1 %), tels que *lavette*₂ 'personne sans énergie' en (81), n'entretiennent pas de lien sémantique direct avec un verbe. Comme indiqué dans le chapitre précédent, ces noms ont été annotés comme figuratifs (cf. section 2.2.1). Nous les excluons pour l'instant de la présentation des résultats généraux du fait de leur statut particulier.

- (81) a. *laver*₁ → *lavette*₁ [Artéfact-INSTRUMENT]
 b. *lavette*₁ → *lavette*₂ [Entité animée-(INSTRUMENT)·FIGURATIF]

Nous avons vu dans le chapitre 2 que les typologies ontologique et relationnelle peuvent aussi être combinées de manière à obtenir un aperçu plus complet de la sémantique des noms.

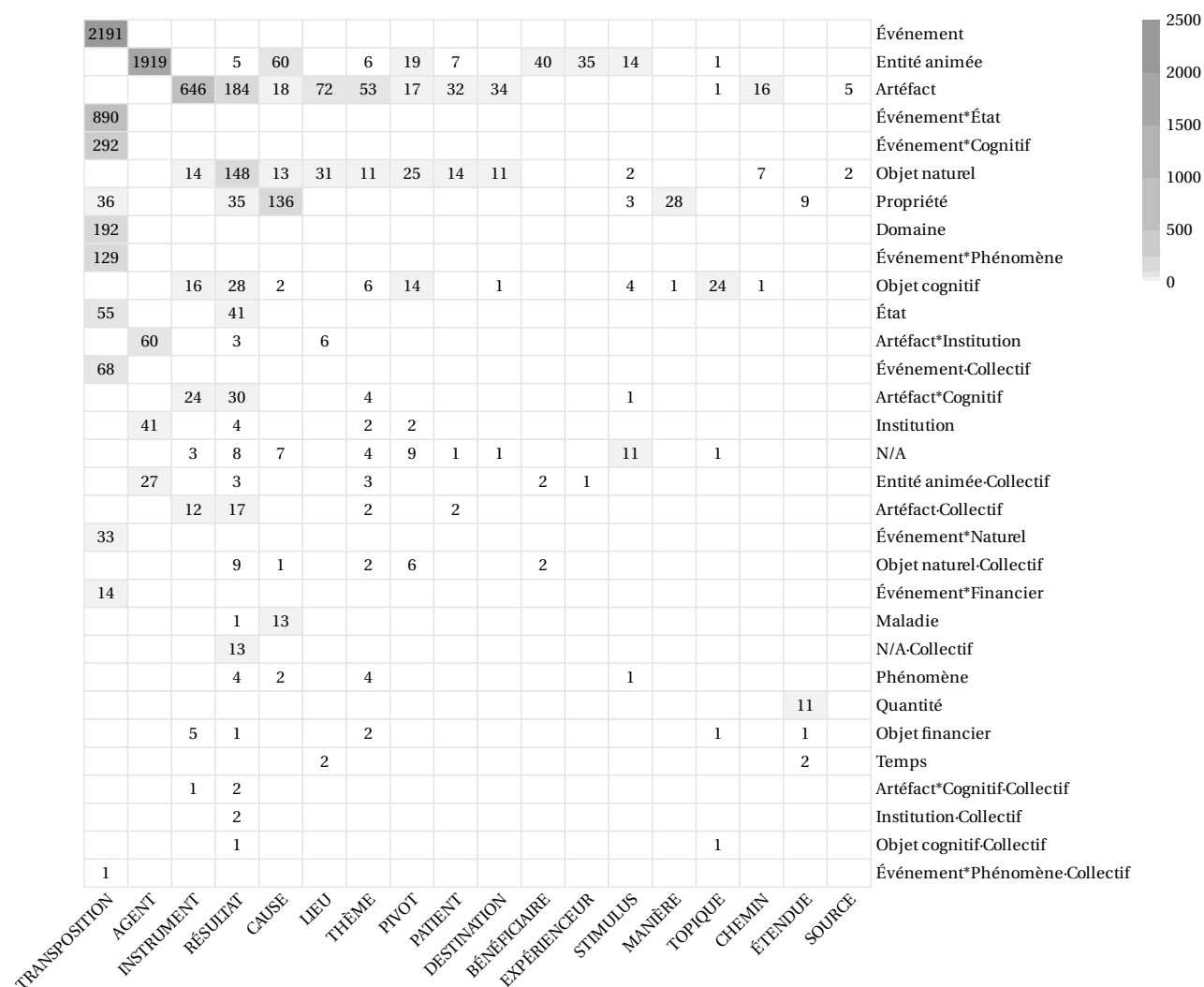


FIGURE 3.2 – Fréquence d’association entre types ontologiques (lignes) et relationnels (colonnes) en nombre de lexèmes. Les types sémantiques sont ordonnés de gauche à droite et de haut en bas par ordre de fréquence décroissant

3.1. Informations générales

Si l'on exclut les items comportant un type relationnel figuratif, un total de 113 types combinés peuvent être identifiés. Les 4 types les plus fréquents (82) sont Événement-TRANSPOSITION (27 % des lexèmes), Entité animée-AGENT (23 %), Événement*État-TRANSPOSITION (11 %) et Artéfact-INSTRUMENT (8 %).

- | | | | |
|------|----|---|--------------------------------|
| (82) | a. | <i>baignade, pondaison, repos</i> | [Événement-TRANSPOSITION] |
| | b. | <i>chasseur, pillarde, régente</i> | [Entité animée-AGENT] |
| | c. | <i>accoutumance, pourrissement, relâche</i> | [Événement*État-TRANSPOSITION] |
| | d. | <i>autotest, cueilloir, perforatrice</i> | [Artéfact-INSTRUMENT] |

À l'autre bout du continuum, 20 types combinés ne sont représentés qu'une seule fois dans la base. Sont notamment concernés les types Phénomène-STIMULUS (*senteur*) et Entité animée-TOPIQUE (*connaissance* 'personne que l'on connaît'). Le nombre de lexèmes instanciant chaque association entre types ontologiques et relationnels est présenté dans la figure 3.2.

Propriétés aspectuelles

Sur l'ensemble des lexèmes nominaux étudiés, 4'162 (51 %) expriment une éventualité, qu'il s'agisse d'un événement, d'un état ou d'une propriété, et possèdent de ce fait des propriétés aspectuelles. Les événements et les états sont évalués en fonction de quatre traits d'*Aktionsart* : la dynamicité, la durativité, la télicité et la post-phase (cf. section 2.2.2). Seul le trait de dynamicité est en revanche pertinent pour les propriétés. En effet, étant atemporelles, celles-ci ne se prêtent pas aux traits de durée, de télicité et de post-phase.

Comme le montre la figure 3.3, il ressort de l'analyse que les lexèmes dotés de propriétés aspectuelles sont principalement dynamiques (83a), duratifs (83b) et téliques (83c), et qu'ils ne possèdent généralement pas de post-phase (83d).

- | | | | |
|------|----|---|----------------|
| (83) | a. | <i>customisation, équarrissage, gesticulade</i> | [+ Dynamicité] |
| | b. | <i>gratouillage, pesée, plumaison</i> | [+ Durativité] |
| | c. | <i>allumage, forgerment, oblitération</i> | [+ Télicité] |
| | d. | <i>exécution, flibuste, labourage</i> | [– Post-phase] |

Nous l'avons vu dans le chapitre 2, les différents traits aspectuels peuvent aussi être considérés conjointement de manière à obtenir des classes d'inspiration vendliérienne (activités, accomplissements, etc.). Leur fréquence relative est présentée dans la figure 3.4. Si la plupart des dérivés correspondent à des activités (84a), des quantités relativement importantes d'accomplissements (84b) et d'achèvements (84c) peuvent aussi être relevées.

- | | | | |
|------|----|--|-------------------|
| (84) | a. | <i>gazouillis, jonglerie, roupillon</i> | [Activité] |
| | b. | <i>cuirassement, profération, raccommode</i> | [Accomplissement] |
| | c. | <i>rejet, trouvaille, trahison</i> | [Achèvement] |

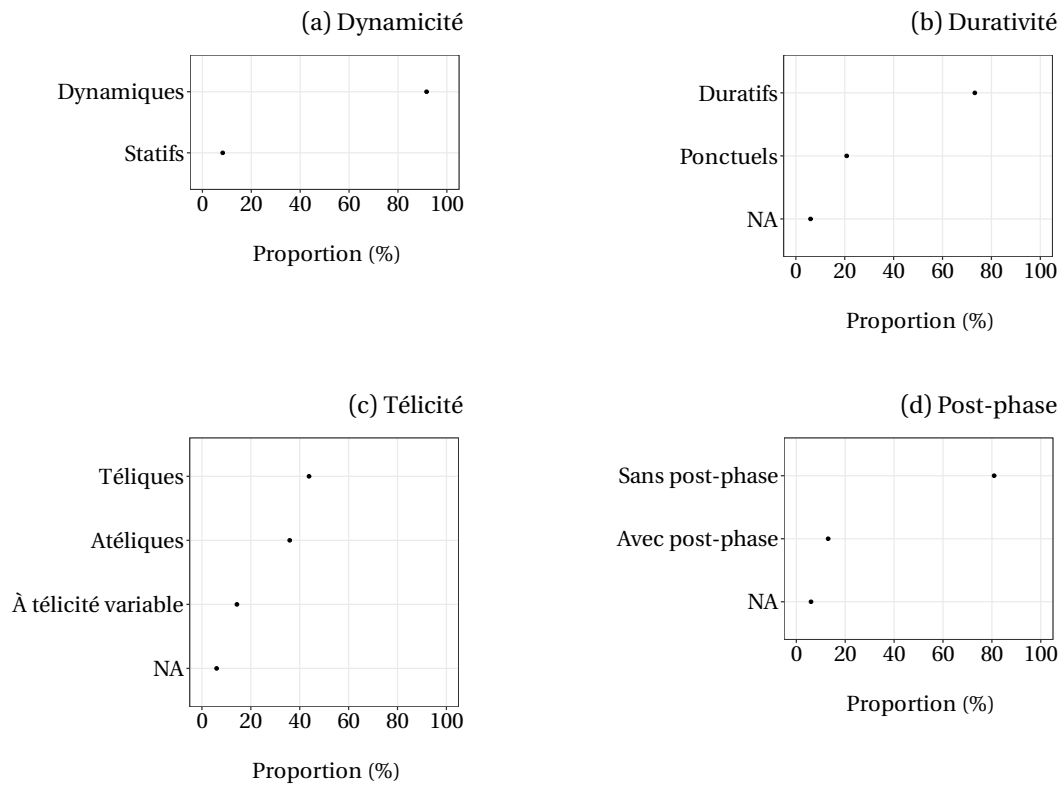


FIGURE 3.3 – Traits aspectuels des noms d'éventualité annotés ($n = 4'162$)

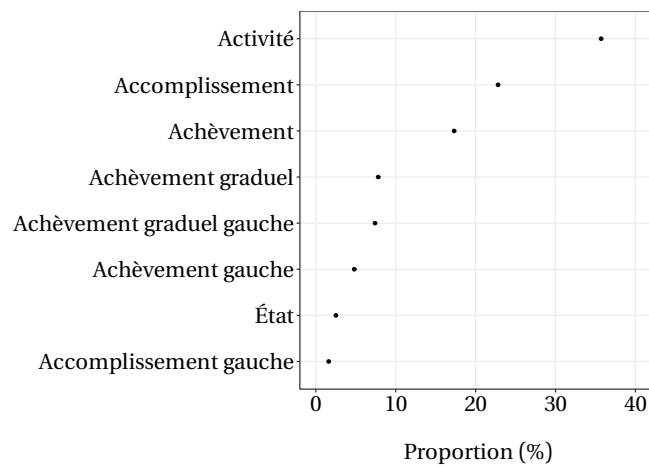


FIGURE 3.4 – Classe aspectuelle des noms d'éventualité annotés ($n = 4'162$)

3.1. Informations générales

Rôles sémantiques

Un total de 4'452 (54 %) lexèmes nominaux ont été encodés comme possédant une structure argumentale dans notre base de données. Ce résultat appelle deux observations. D'une part, nous relevons que tous les noms d'éventualité n'ont pas nécessairement une structure argumentale. Sont notamment concernés les noms de domaine (*camping*), les noms d'événement météorologique (*saucée*), les noms d'événement avec une dimension collective (*braderie*), ainsi que les noms d'événement hybride impliquant un phénomène (*tapage*) ou un objet naturel (*éboulis*). D'autre part, il convient de souligner que les noms dénotant des entités animées peuvent avoir une structure argumentale lorsque celle-ci se fonde sur une éventualité sous-jacente (cf. Roy & Soare, 2014) (p. ex. *l'inventeur de cette machine*, *le destinataire de ce courrier*, *les tenants de cette idéologie*).

D'un point de vue ontologique, les lexèmes pourvus d'une structure argumentale sont majoritairement des dérivés dénotant des éventualités (88 %) et des entités animées (12 %). Sur le plan relationnel, la plupart sont transpositionnels (82 %) et agentifs (11 %). Le nombre d'arguments qu'ils convoquent et les rôles qui y sont associés sont variables. Environ 44 % d'entre eux possèdent un seul argument (85a), 50 % ont deux arguments (85b), et 5 % comportent une structure argumentale à trois arguments (85c). La figure 3.5 présente la fréquence relative des rôles liés à ces différents arguments.

- (85) a. *le discours de X, le roucoulement de X, l'exécuteur de Y*
b. *le financement de Y par X, la géocodification de Y par X, la levée de Y par X*
c. *l'assignement de Y à Z par X, la déviation de Y dans Z par X, la livraison de Y à Z par X*

3.1.2 Verbes

Nombre de sens

Passons maintenant à la description des lexèmes verbaux. Parmi les 1'710 verbes que compte notre base de données, 1'008 (59 %) sont monosémiques et 702 (41 %) sont ambigus, pour un total de 2'810 sens. En moyenne, un verbe a 1.64 sens ($ET = 1.04$). Si seuls les verbes ambigus sont pris en compte, la moyenne atteint 2.57 sens ($ET = 1.08$). La figure 3.6 présente la distribution des verbes décrits en fonction du nombre de sens qu'ils comportent.

Un test t de Welch ($t(2567.5) = -3.13$, $p < 0.001$) indique par ailleurs que les verbes examinés sont dans l'ensemble plus ambigus que les noms de notre ressource. Ce résultat va dans le sens d'un constat déjà établi dans la littérature (cf. chapitre 1), selon lequel les verbes sont globalement plus ambigus que les noms dans le lexique général (Gentner, 1981; Miller & Fellbaum, 1991; Levickij et al., 1999). Nos observations attestent d'autant plus de ce déséquilibre d'ambiguïté que seuls les sens verbaux liés à des nominalisations ont été pris en compte dans ce travail.

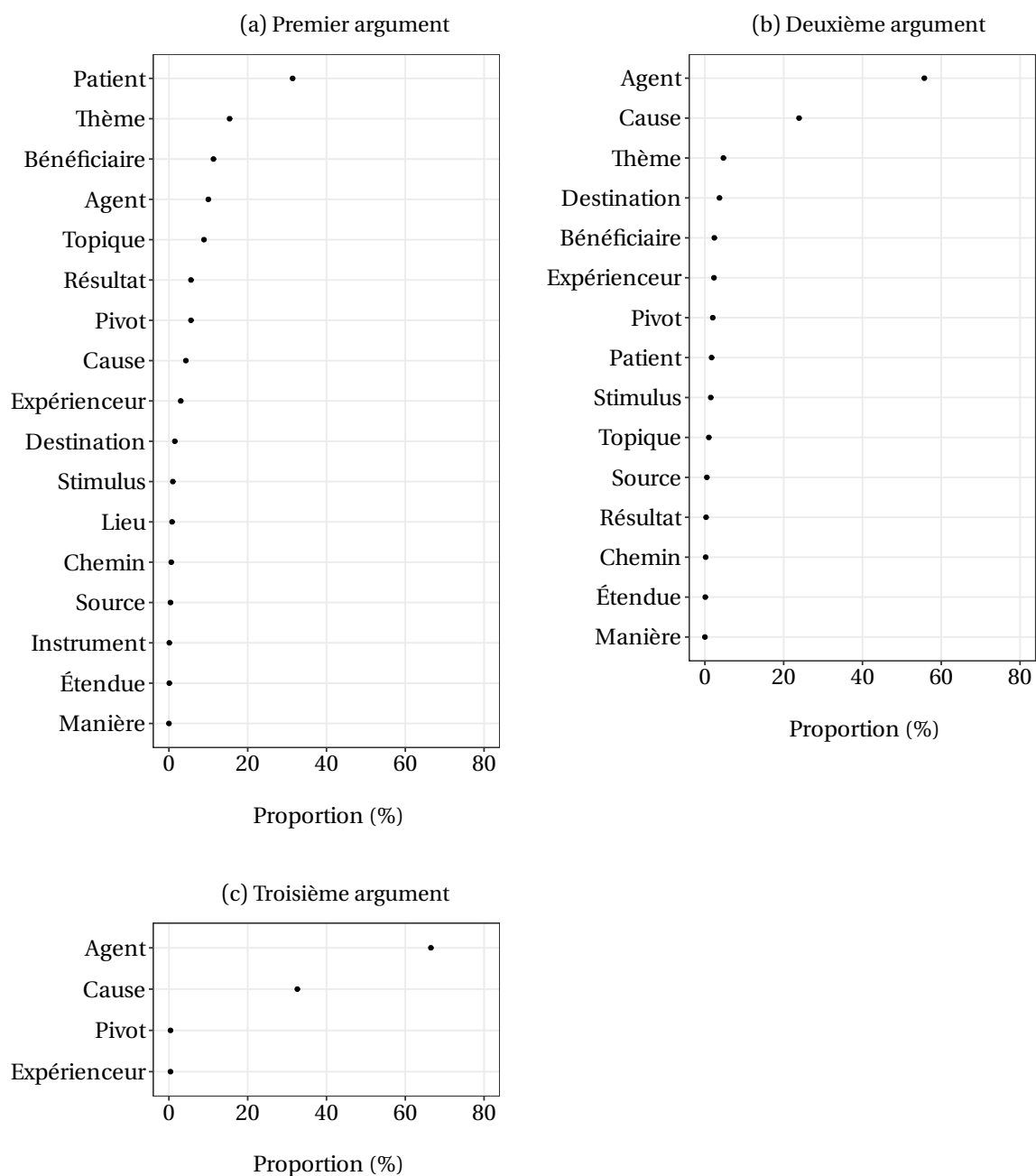


FIGURE 3.5 – Rôles sémantiques assignés par les lexèmes dotés d'une structure argumentale ($n = 4'452$)

3.1. Informations générales

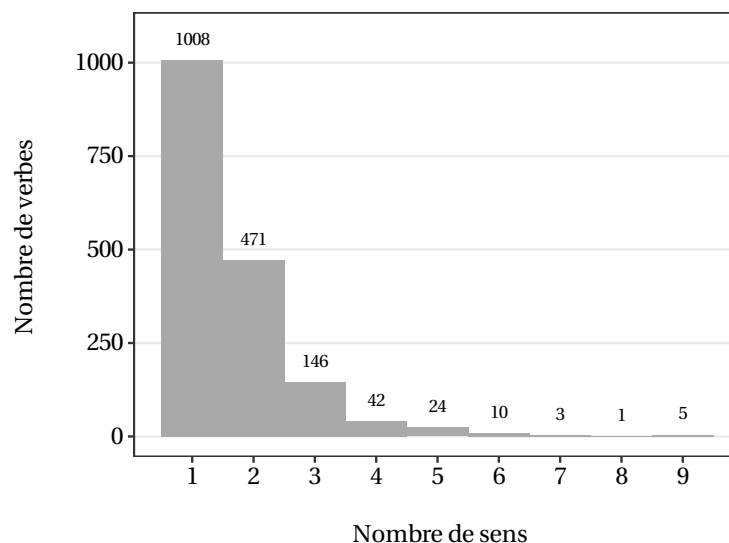


FIGURE 3.6 – Distribution des verbes annotés en fonction du nombre de sens qu'ils comportent

Propriétés aspectuelles

Comme pour les noms, l'aspect lexical des verbes a été analysé selon quatre traits distincts : la dynamicité, la durativité, la télicité et la post-phase. La figure 3.7 présente la fréquence relative des valeurs associées à chacun de ces traits³⁹. Il apparaît que les lexèmes verbaux de notre base de données sont principalement dynamiques (86a), duratifs (86b), téliques (86c) et sans post-phase (86d).

- | | | | |
|------|----|----------------------------------|----------------|
| (86) | a. | <i>broder, caqueter, saler</i> | [+ Dynamicité] |
| | b. | <i>dormir, forger, surfer</i> | [+ Durativité] |
| | c. | <i>plonger, cercler, germer</i> | [+ Télicité] |
| | d. | <i>chuchoter, dévaler, gober</i> | [– Post-phase] |

Les traits de dynamicité, de durativité, de télicité et de post-phase peuvent aussi être considérés conjointement pour obtenir des classes aspectuelles d'inspiration vendlérienne. Comme le montre la figure 3.8, la plupart des verbes de notre base de données se révèlent être des activités (87a), des accomplissements (87b) et des achèvements (87c).

- | | | | |
|------|----|--|-------------------|
| (87) | a. | <i>jardiner, dorloter, pleurnicher</i> | [Activité] |
| | b. | <i>déshabiller, guérir, souder</i> | [Accomplissement] |
| | c. | <i>cohériter, parier, réceptionner</i> | [Achèvement] |

39. Dans ce qui suit, les lexèmes verbaux associés à plusieurs sens nominaux ne sont comptés qu'une seule fois. Tel est le cas, par exemple, du lexème *importer*₁ 'introduire dans un pays', lié à *importation*₁ 'action d'introduire dans un pays' et *importation*₂ 'ce qui est introduit dans un pays'.

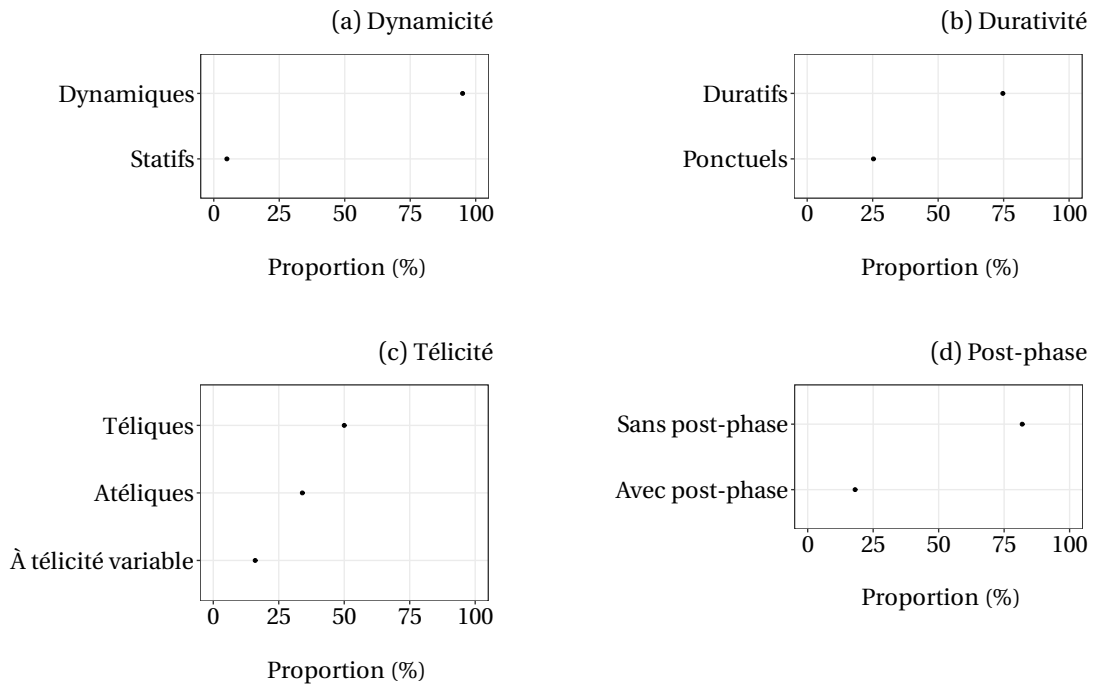


FIGURE 3.7 – Traits aspectuels des verbes annotés

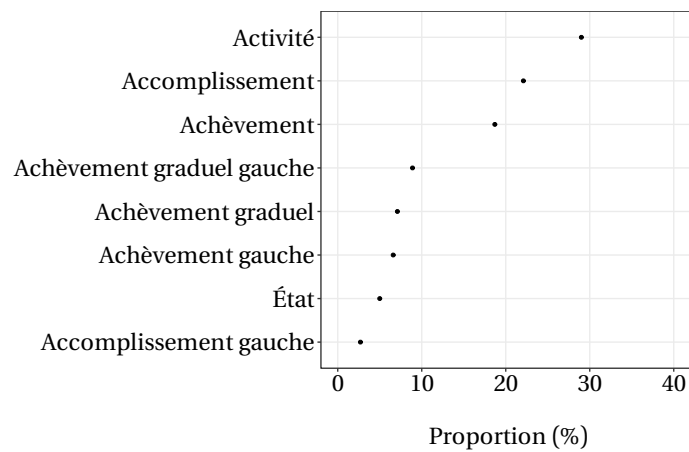


FIGURE 3.8 – Classe aspectuelle des verbes annotés

3.1. Informations générales

Ce travail ne vise pas à analyser en détail la relation aspectuelle entre verbes et noms d'éventualité dérivés. Cependant, dans la mesure où plusieurs études ont abordé cette problématique par le passé (Balvet et al., 2011; Fábregas et al., 2012; *inter alia*), il nous semble intéressant de relever que l'aspect lexical du verbe est souvent (mais pas systématiquement) préservé à travers la nominalisation dans nos données. Le taux global de préservation des valeurs aspectuelles atteint en effet 84.5 %, 3'515 des 4'162 noms dénotant des éventualités s'étant vu attribuer exactement les mêmes valeurs aspectuelles que leurs bases verbales. Plus spécifiquement, ce taux de préservation s'élève à 94.2 % pour la dynamicité, 90.6 % pour la durativité, 87.7 % pour la télicité, et 91.3 % pour la post-phase. Ces résultats confirment que la tendance est à la préservation, mais qu'il subsiste néanmoins une part non négligeable de décalages entre verbes et noms.

Rôles sémantiques

Les lexèmes verbaux examinés varient en fonction (i) du nombre d'arguments, (ii) des fonctions syntaxiques et (iii) des rôles sémantiques qu'ils admettent. S'agissant du nombre d'arguments, 27 % des lexèmes de notre base de données n'en convoquent qu'un seul, 66 % en ont deux, et 7 % en comportent trois. Un seul verbe – *grêler* – a été annoté comme exempt d'arguments. Plusieurs combinaisons de fonctions syntaxiques peuvent par ailleurs être observées. 27 % des lexèmes ne présentent qu'un sujet (88a), 56 % ont un sujet et un objet mais pas d'oblique (88b), 10 % ont un sujet et un oblique mais pas d'objet (88c), et 7 % peuvent convoquer un sujet, un objet et un oblique (88d).

- (88) a. *X clopine, X grimace, X triche*
b. *X ratisse Y, X écorne Y, X garde Y*
c. *X adhère à Z, X cohabite avec Z, X dépend de Z*
d. *X abreuve Y de Z, X puise Y dans Z, X vole Y à Z*

Nous relevons en outre d'importantes disparités en matière de rôles sémantiques. Comme le montre la figure 3.9, les sujets se voient principalement attribuer le rôle d'agent, les objets celui de patient, et les obliques celui de destination. Une vue d'ensemble de la structure argumentale des verbes peut également être obtenue en combinant les rôles des sujets, des objets et des obliques. Dans ce contexte, les verbes transitifs à objet patient et à sujet agent (12 %) (89a) ou cause (12 %) (89b) sont majoritaires, suivis de près par les verbes intransitifs à sujet patient (10 %) (89c). Ces résultats vont dans le sens de l'hypothèse selon laquelle, lorsqu'un verbe demande un argument agentif, celui-ci correspond en principe à son sujet plutôt qu'à un potentiel objet ou oblique (Bowerman, 1990; Wechsler, 2015; Wilson et al., 2022).

- (89) a. *X bêche Y, X égorge Y, X peint Y*
b. *X chamboule Y, X déshumanise Y, X souille Y*
c. *X décroît, X meurt, X sèche*

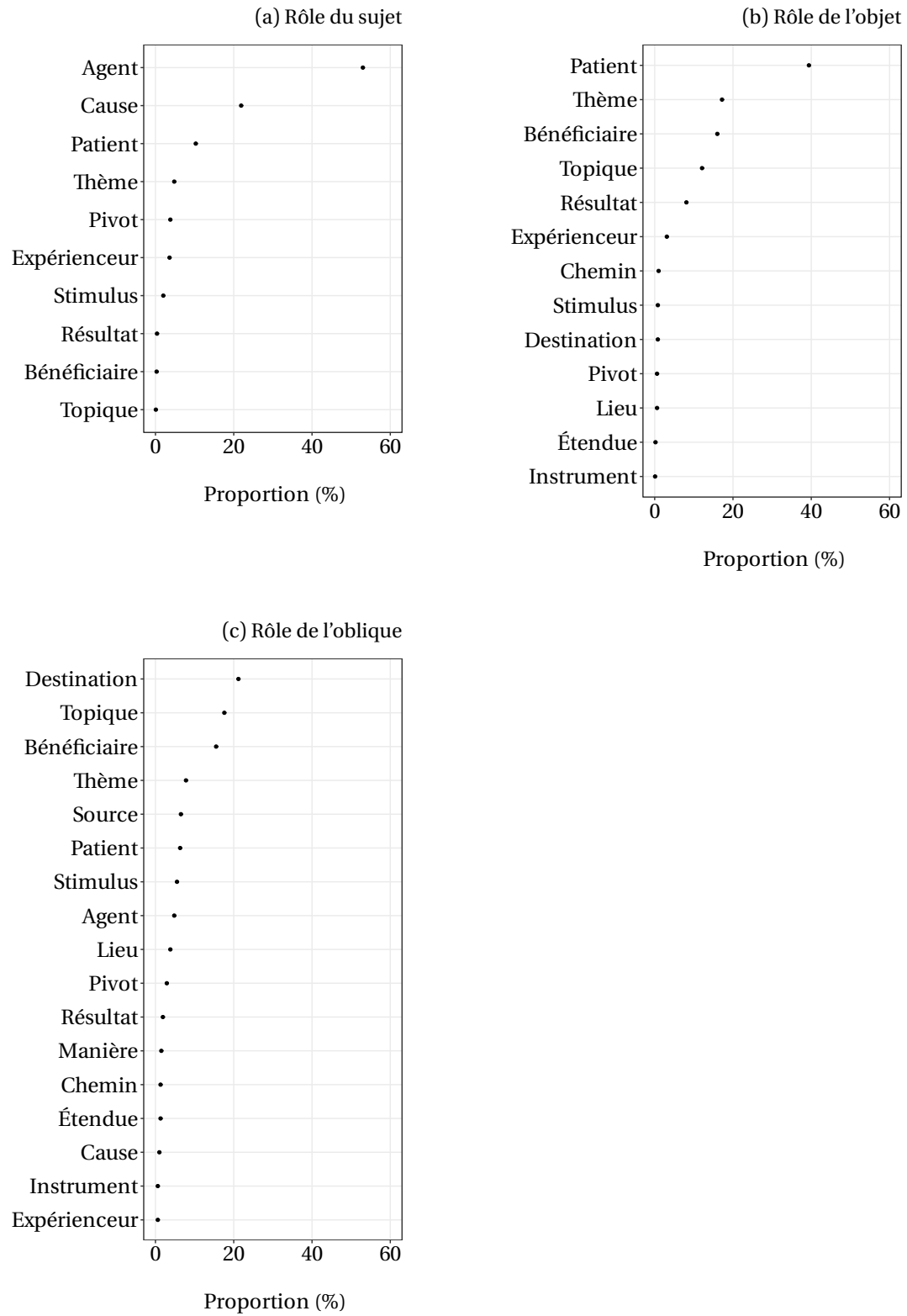


FIGURE 3.9 – Rôles sémantiques assignés par les verbes annotés

3.2. Associations entre propriétés

Enfin, quelques observations peuvent être faites concernant la relation entre les rôles des verbes et ceux des noms dérivés. Nous constatons que 4'202 des 4'452 noms (94 %) ayant été encodés comme possédant une structure argumentale dans notre base de données reprennent au moins un des rôles de leur base verbale. Plus précisément, 17.5 % des dérivés préservent partiellement les rôles de leur base, tandis que 76.9 % les conservent intégralement. Seuls 5.6 % des noms ne reprennent aucun rôle du lexème-base.

3.2 Associations entre propriétés

Maintenant que les principales caractéristiques sémantiques des lexèmes ont été établies, et avant de passer à l'analyse des différentes configurations d'ambiguïté entre verbes et noms dérivés, il peut être intéressant de chercher à déterminer si les propriétés que nous avons analysées sont interdépendantes. Par exemple, les typologies ontologique et relationnelle sont-elles associées de manière significative? Et qu'en est-il des relations entre les différents traits aspectuels ou entre les rôles sémantiques assignés aux arguments des verbes et des noms?

Dans ce qui suit, nous revenons premièrement (3.2.1) sur les relations entre les différentes propriétés sémantiques des noms, puis des verbes. La seconde partie de cette section (3.2.2) est quant à elle dédiée à l'examen du comportement des noms déverbaux du français au regard de la loi sens-fréquence de Zipf (1945), selon laquelle plus un mot est fréquent, plus il a de significations.

3.2.1 Variables catégorielles

Propriétés nominales

Nous cherchons premièrement à déterminer s'il existe des relations significatives entre les différentes propriétés nominales, à savoir les types sémantiques (ontologiques et relationnels), les traits aspectuels (dynamicité, durativité, télicité et post-phase), les rôles des arguments (premier, deuxième et troisième) et le procédé morphologique. Malgré la taille importante de notre échantillon, nous recourons pour ce faire au test exact de Fisher avec valeur p basée sur 2'000 répétitions plutôt qu'au test du χ^2 ou au test G d'indépendance, les valeurs attendues étant en effet très basses pour plusieurs paires de propriétés (cf. McDonald, 2014, p. 89). Il ressort de l'analyse que toutes les variables sont associées de manière significative ($p < 0.001$).

Nous examinons dans un second temps la magnitude des associations entre variables en faisant ici appel au τ de Goodman et Kruskal (1954). Contrairement au V de Cramér, qui jouit pourtant d'une plus grande popularité, ce test d'association présente l'avantage d'être directionnel : il permet en effet d'évaluer, pour deux variables x et y , la dépendance unilatérale de x à y et de y à x . La mesure d'association $\tau(x, y)$ quantifie plus précisément la diminution

de la proportion d'erreurs que nous commettrions en choisissant y tout en connaissant x ⁴⁰. En pratique, un $\tau(x, y)$ de 0.75 signifie ainsi que connaître x réduit de 75 % les chances de faire une erreur lors du choix de y par rapport à une situation où aucune information préalable n'est disponible. L'échelle s'étend de $\tau(x, y) = 0$, lorsque y est totalement indépendant de x , à $\tau(x, y) = 1$, lorsque y est totalement dépendant de x .

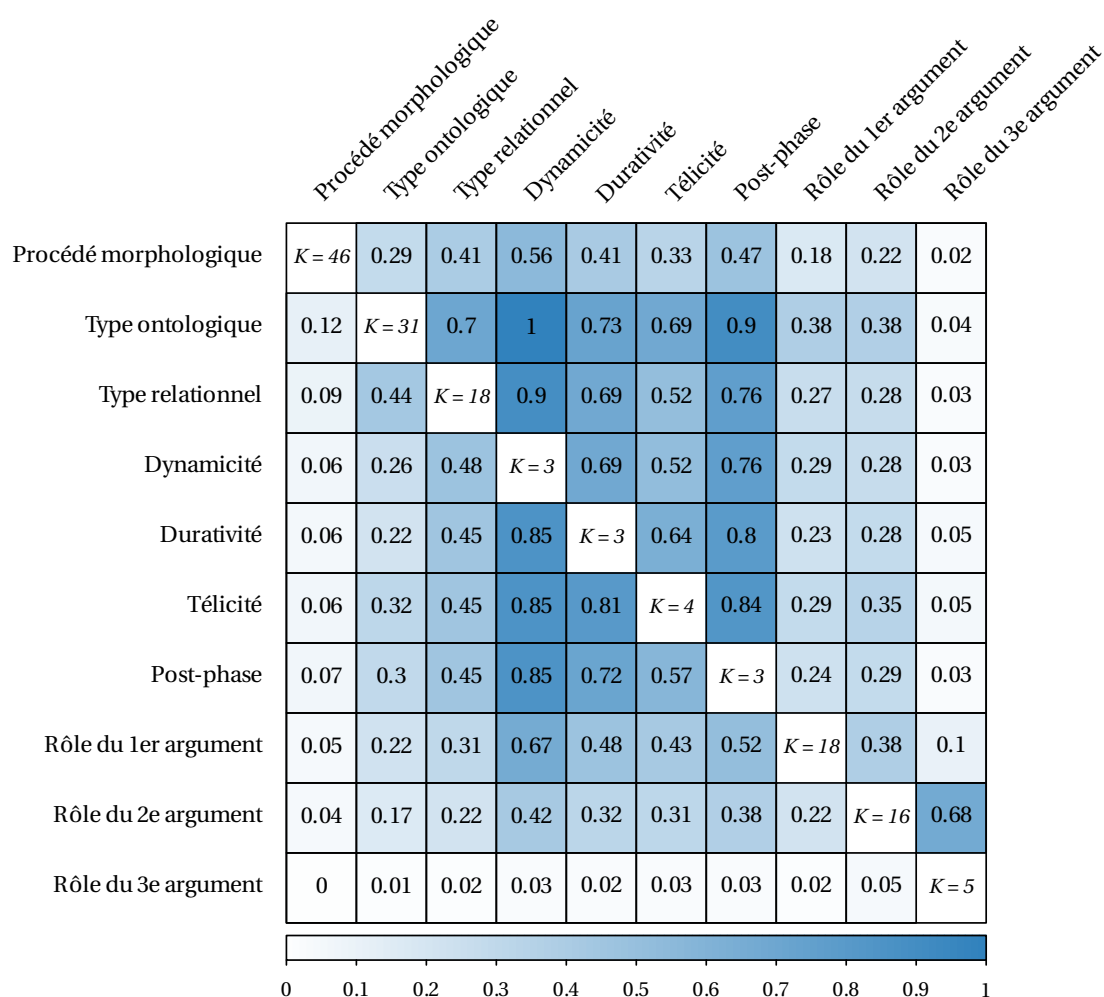


FIGURE 3.10 – Associations entre propriétés nominales selon le τ de Goodman et Kruskal. Les valeurs correspondent à la mesure d'association directionnelle $\tau(x, y)$, où x constitue la condition (ligne) et y le dépendant (colonne). La diagonale indique le nombre de niveaux que comporte x

La figure 3.10 synthétise les résultats du τ de Goodman et Kruskal obtenus pour nos variables nominales. De nombreuses associations présentent un τ modéré voire élevé, ce qui suggère une bonne interprédictibilité des différentes propriétés examinées. En effet,

40. Il s'agit ainsi d'un test de type *RPE*, pour « Réduction Proportionnelle de l'Erreur » (*PRE* en anglais, pour *Proportional Reduction in Error*).

3.2. Associations entre propriétés

47 (52 %) des 90 paires de propriétés examinées obtiennent un score égal ou supérieur à 0.3. L'hypothèse peut toutefois être formulée que ces résultats reposent sur la fréquence très élevée d'un petit nombre de combinaisons de valeurs plutôt que sur un nombre élevé de combinaisons récurrentes au sein de chaque paire de variables. Étant donné qu'ils ne possèdent pas de propriétés d'aspect lexical, tous les dérivés dénotant des entités dans notre base de données ($n = 3'940$, 49 %) se sont par exemple vus attribuer la valeur « NA » pour la dynamicité, la durativité, la télélicité et la post-phase.

Le caractère asymétrique de la plupart des relations entre variables nominales peut également être relevé. Nous constatons d'une part que certaines propriétés semblent exercer une influence plus prégnante que d'autres. Par exemple, les types ontologiques et relationnels, et dans une moindre mesure la télélicité et la durativité, sont influents pour pratiquement toutes les autres variables, alors que le rôle du troisième argument ne fournit que peu d'informations sur le reste des propriétés. Inversement, certaines variables sont plus fortement conditionnées que d'autres. De nombreuses propriétés peuvent notamment servir à déterminer les différentes valeurs associées aux traits d'aspect lexical (dynamicité, durativité, télélicité et post-phase), tandis qu'aucune propriété n'est très informative lorsqu'il s'agit de déterminer le procédé morphologique.

Il convient aussi d'examiner plus avant la relation entre types ontologiques et relationnels. Les études sur la sémantique des noms déverbaux tendent en effet à privilégier l'un ou l'autre plan d'analyse (cf. chapitre 1), ce qui amène à s'interroger sur la pertinence d'une double typologie. La figure 3.10 montre qu'une association modérée et asymétrique existe entre les deux classifications, avec un τ de 0.7 pour la prédiction du type relationnel à partir du type ontologique, contre 0.44 pour l'inverse. Bien que certaines associations entre types ontologiques et relationnels soient fréquentes dans nos données (la plupart des noms dénotant des entités animées expriment par exemple un agent par rapport à leur lexème-base), ces résultats montrent qu'elles ne sont pas systématiques : connaître le type ontologique d'un lexème permet de déduire des informations substantielles sur son type relationnel, mais l'inverse est nettement moins probant. Le type ontologique Maladie, par exemple, est exclusivement associé aux types relationnels CAUSE et RÉSULTAT, tandis que ces derniers sont liés respectivement à 9 et 20 types ontologiques différents. Ces observations font écho à l'une de nos études antérieures reposant sur la même grille d'analyse bipartite pour un échantillon de données différent (Salvadori & Huyghe, 2023), où un test de χ^2 avec valeur p basée sur 2'000 répétitions avait révélé une relation significative entre les deux typologies ($p < 0.001$), et un test V de Cramér, une force d'association modérée ($V = 0.41$)⁴¹. Il semble ainsi que l'analyse relationnelle, souvent privilégiée pour décrire le sens des dérivés, ne soit pas suffisante pour rendre compte de la complexité sémantique du lexique construit. Étant donné qu'ils contribuent à caractériser des aspects distincts de la sémantique des

41. Pour information, le V de Cramér entre types ontologiques et types relationnels se monte à 0.45 dans nos données.

noms déverbaux, nous continuerons donc, dans ce qui suit, à distinguer les plans d'analyse relationnel et ontologique.

Enfin, la relation entre procédés morphologiques et types sémantiques appelle également quelques commentaires, puisqu'elle touche directement au rapport sens-forme dans le lexique. Les associations sont ici toutes asymétriques : $\tau(\text{procédé morphologique}, \text{type ontologique}) = 0.29$, $\tau(\text{type ontologique}, \text{procédé morphologique}) = 0.12$, $\tau(\text{procédé morphologique}, \text{type relationnel}) = 0.41$ et $\tau(\text{type relationnel}, \text{procédé morphologique}) = 0.09$. Ces résultats suggèrent un déséquilibre dans la relation multiple (*many-to-many*) entre les différents types sémantiques et les procédés morphologiques. En effet, si le procédé morphologique influence dans une certaine mesure le type ontologique et relationnel des noms dérivés, connaître le type ontologique ou relationnel d'un lexème fournit en revanche très peu d'informations sur le procédé morphologique avec lequel il a été formé. Par exemple, le suffixe *-rice* construit principalement des noms d'agent animé ($n = 92$, soit 85 % des dérivés en *-rice*), alors que le type combiné Entité animée-AGENT est mobilisé par pas moins de 28 procédés morphologiques différents. Nous tenterons de déterminer dans le chapitre 5 si les préférences que montrent les procédés morphologiques en matière de types sémantiques s'étendent également à l'ambiguïté des dérivés (cf. section 1.2.2).

Propriétés verbales

Examinons maintenant les associations entre les différentes propriétés verbales, à savoir les traits aspectuels (dynamicité, durativité, télicité et post-phase) et les rôles sémantiques des sujets, des objets et des obliques. Comme pour les noms, nous commençons l'analyse par un test exact de Fisher afin de déterminer si les différentes variables prises deux à deux sont interdépendantes. Plus précisément, nous procédons à un test exact de Fisher avec valeur p simulée sur la base de 2'000 répétitions pour toutes les paires sauf 3, pour lesquelles le test standard peut être appliqué⁴². Les résultats indiquent que toutes les variables sont associées de manière significative, les valeurs p obtenues étant toujours inférieures à 0.01.

Nous avons ensuite appliqué le test τ de Goodman et Kruskal, utilisé précédemment pour les variables nominales, afin d'évaluer le degré d'association entre les différentes propriétés verbales. Les scores sont synthétisés dans la figure 3.11. Un premier résultat notable est la proportion élevée de paires symétriques, où $\tau(x, y)$ est proche de $\tau(y, x)$, notamment pour les traits de dynamicité, de durativité et de post-phase. Par ailleurs, nous observons que les valeurs sont globalement inférieures à celles obtenues pour les noms. Seules 2 paires de propriétés verbales (5 %) affichent en effet un score égal ou supérieur à 0.3. Cette différence suggère que l'interdépendance entre les variables est plus faible dans le domaine verbal que dans le domaine nominal, une disparité qui pourrait s'expliquer par fait que les traits aspectuels s'appliquent uniquement à certains noms, alors qu'ils concernent tous les

42. L'espace de travail disponible dans R était suffisant pour des tests exacts. Les 3 relations en jeu sont celles entre durativité et post-phase, durativité et dynamicité, et dynamicité et post-phase.

3.2. Associations entre propriétés

verbes. Les catégories non pertinentes dans le domaine nominal (« NA ») montrent une forte interdépendance, ce qui est moins le cas pour les verbes où les relations entre propriétés aspectuelles sont moins systématiques.

Il reste que plusieurs associations dans les données verbales sont asymétriques et présentent des scores modérés, voire élevés. La relation entre le rôle du sujet et la dynamicité se révèle particulièrement intéressante à cet égard : si le rôle du sujet fournit beaucoup d'informations sur la dynamicité du verbe ($\tau = 0.78$), l'inverse n'est pas vérifié ($\tau = 0.07$). Cette asymétrie semble résulter du fait que tous les verbes ayant un sujet agent ($n = 1'488$) sont dynamiques dans notre base de données, tandis que les sujets des verbes dynamiques couvrent 10 rôles sémantiques différents. L'idée communément admise selon laquelle l'agentivité implique la dynamicité, mais la dynamicité n'implique pas nécessairement l'agentivité (Filip, 2016, p. 19) se vérifierait donc. Bien que cette découverte souligne une relation forte et asymétrique entre aspect lexical et programmes de complémentation dans le domaine verbal, elle se limite toutefois à un seul trait d'aspect lexical et une seule fonction syntaxique.

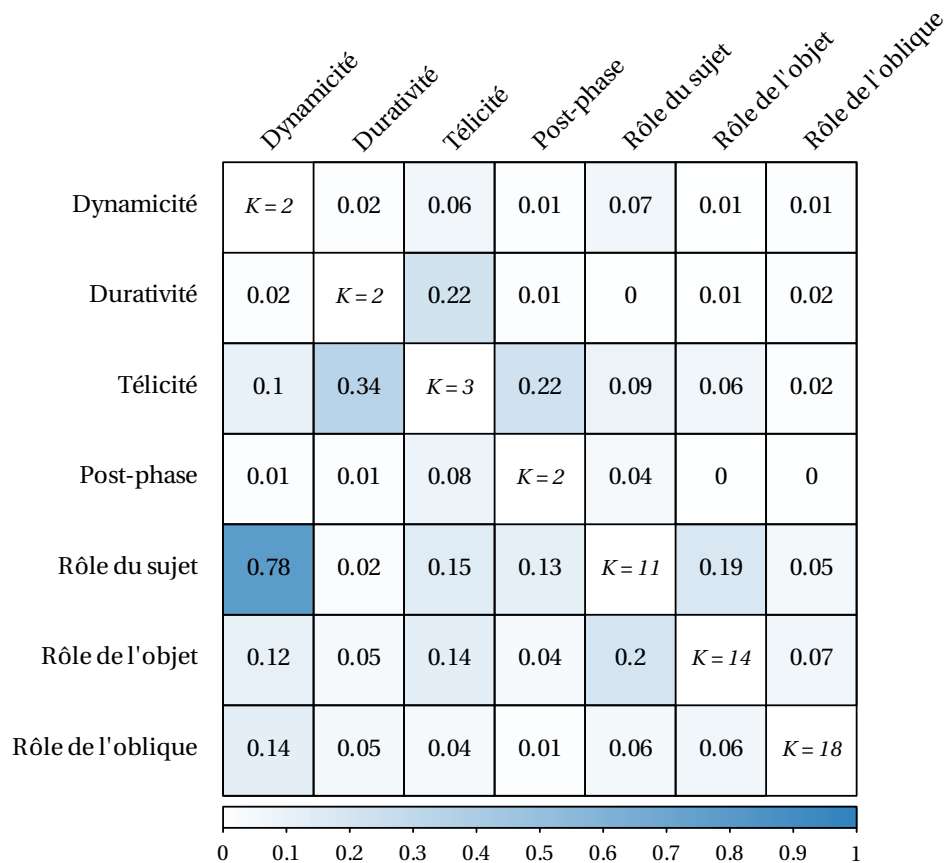


FIGURE 3.11 – Associations entre propriétés verbales selon le τ de Goodman et Kruskal. Les valeurs correspondent à la mesure d'association directionnelle $\tau(x, y)$, où x constitue la condition (ligne) et y le dépendant (colonne). La diagonale indique le nombre de niveaux que comporte x

3.2.2 Loi sens-fréquence

Selon la loi sens-fréquence (*meaning-frequency law* en anglais) formulée par Zipf (1945), plus un mot est fréquent, plus il a de significations. Cette loi a été vérifiée dans plusieurs familles de langues (indo-européennes, japoniques, sino-tibétaines et austronésiennes), avec des méthodes variées, et sur la base de différents corpus (Ilgen & Karaoglan, 2007; Hernández-Fernández et al., 2016; Bond et al., 2019; Casas et al., 2019; *inter alia*). Elle est de fait souvent considérée comme relevant des universaux linguistiques.

Nous cherchons dans cette section à déterminer si cette loi sens-fréquence structure également le rapport entre ambiguïté et fréquence s’agissant des noms déverbaux et des verbes apparentés en français, en nous inspirant pour ce faire de la méthodologie employée par Casas et ses collègues (2019) dans leur étude sur l’anglais, l’espagnol et le néerlandais. Définissant la loi sens-fréquence comme une association significative et positive entre la fréquence des mots et le nombre de sens qu’ils comportent (Casas et al., 2019, p. 21), les auteurs évaluent pour chaque langue la relation entre le nombre de synsets renseignés pour différents mots dans WordNet⁴³ et leur fréquence dans deux corpus à l’aide de tests de corrélation. Ils n’opèrent pas de distinction entre polysémie et homonymie, et traitent conjointement les différentes catégories grammaticales qu’ils étudient (noms, verbes, adjectifs et adverbes). Leur approche présente *a minima* trois avantages (Casas et al., 2019, p. 21-23) : (i) elle ne présuppose pas de loi de puissance ou de généralisation d’une loi de puissance, comme il est souvent le cas dans les études traditionnelles sur les lois statistiques du langage ; (ii) elle est agnostique quant à la dépendance fonctionnelle entre variables ; (iii) enfin, elle évite de devoir recourir à un regroupement des données par classe (méthode du *binning*).

Jeu de données	Moyenne	Écart type
Noms	24’501	116’925
Verbes	157’560	2’013’686
Noms et verbes	63’030	1’089’515

TABLEAU 3.2 – Fréquence des noms et verbes analysés dans le corpus FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015)

Au vu de la méthode d’analyse sémantique adoptée dans ce travail, nous confrontons pour notre part le nombre de sens ayant été attribués aux nominalisations et verbes apparentés dans notre base de données et leur fréquence en tant que noms et verbes (voir tableau 3.2) dans notre corpus de référence, FRCOW16A (Schäfer & Bildhauer, 2012; Schäfer, 2015). Nous procédons concrètement à trois tests τ de Kendall afin d’évaluer le rapport sens-fréquence pour les noms, les verbes, mais aussi pour les noms et verbes pris conjointement. Comme le

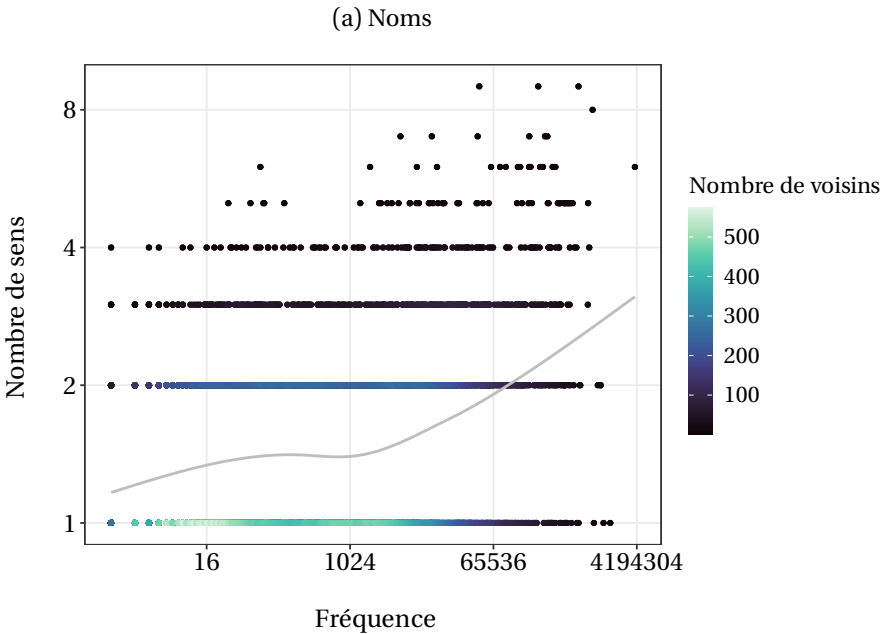
43. Les auteurs utilisent plus précisément les WordNet suivants : le *Princeton WordNet* (Fellbaum, 1998) pour l’anglais, le *Multilingual Central Repository* (Gonzalez-Agirre et al., 2012) pour l’espagnol, et l’*Open Dutch WordNet* (Postma et al., 2016) pour le néerlandais.

3.2. Associations entre propriétés

montre le tableau 3.3, il ressort de l'analyse que toutes les associations sont significatives et positives (avec une corrélation généralement modérée et un score légèrement plus élevé pour les verbes), une tendance que semblent soutenir les représentations visuelles des données dans la figure 3.12. Nos résultats confirment ainsi la loi sens-fréquence telle que définie par Casas et ses collègues (2019), et ce, dans des proportions similaires (Casas et al., 2019, p. 44-46).

Jeu de données	τ	Valeur p
Noms	0.21	< 0.001
Verbes	0.28	< 0.001
Noms et verbes	0.22	< 0.001

TABLEAU 3.3 – Résultats des analyses de corrélation (test τ de Kendall) entre le nombre de sens et la fréquence des noms et verbes apparentés



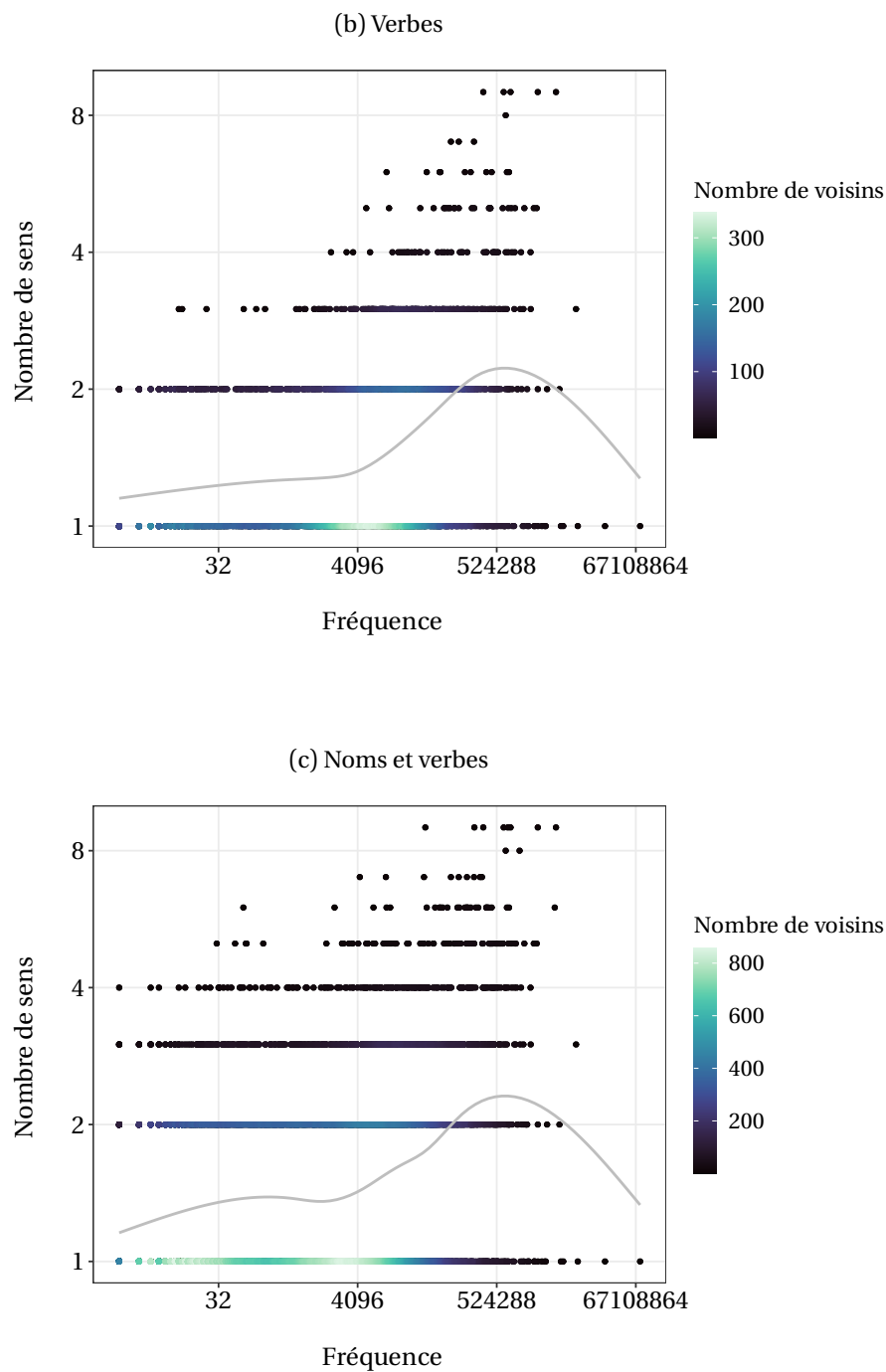


FIGURE 3.12 – Distribution des mots en fonction de leur fréquence dans le corpus FRCOW16A et du nombre de sens qu’ils comportent dans la base de données. Pour plus de visibilité, les résultats sont transformés suivant une double échelle logarithmique binaire. Les observations sont colorées en fonction du nombre de voisins qu’elles possèdent pour représenter la densité des données. Une courbe lissée non paramétrique est ajoutée afin de signaler la tendance générale

3.3 Configurations d'ambiguïté

Nous avons jusqu'à présent décrit le comportement sémantique des noms et des verbes suivant différentes propriétés (section 3.1), puis établi l'existence de relations significatives entre ces dernières (section 3.2). Il reste à examiner de quelle façon les lexèmes présentés précédemment s'associent les uns aux autres en cas d'ambiguïté verbale et nominale, un travail déjà amorcé dans Salvadori et Huyghe (2022).

En guise d'illustration, considérons deux noms ayant été annotés comme bisémiques dans notre base de données : *gâchage* ('action de délayer avec de l'eau' et 'action de compromettre') et *pliage* ('action de plier' et 'objet plié'). Comme le montrent les exemples en (90) et (91), le rapport qu'entretient *gâchage* avec sa base *gâcher* est sensiblement différent de celui qu'a *pliage* avec *plier*, eu égard au nombre de lexèmes concernés dans chaque cas et la façon dont ceux-ci se combinent. Respectivement deux lexèmes-bases et un lexème-base sont en effet impliqués dans la formation des différentes lectures de *gâchage* et de *pliage* :

- (90) a. *gâcher*₁ 'délayer avec de l'eau' → *gâchage*₁ 'action de délayer avec de l'eau'
b. *gâcher*₂ 'compromettre' → *gâchage*₂ 'action de compromettre'
- (91) a. *plier*₁ 'rabattre une matière sur elle-même' → *pliage*₁ 'action de plier'
b. *plier*₁ 'rabattre une matière sur elle-même' → *pliage*₂ 'objet plié'

Nous désignons ces deux scénarios sous le nom de configurations d'ambiguïté « à bases multiples » *versus* « à base unique » et les schématisons dans les figures 3.13a et 3.13b.

$$V_1/V_2 \longrightarrow N_1/N_2$$

(a) Bases multiples

$$V_1 \longrightarrow N_1/N_2$$

(b) Base unique

FIGURE 3.13 – Configurations générales d'ambiguïté

Qu'en est-il de nos données? Le tableau 3.4 synthétise la distribution des noms identifiés comme ambigus ($n = 2'022$) en fonction des deux configurations d'ambiguïté que nous venons de distinguer. Certains noms comme *éplucheur* en (92) correspondent aux deux cas et sont par conséquent comptés deux fois.

- (92) a. *éplucher*₁ 'peler (qqch)' → *éplucheur*₁ 'personne qui pèle (qqch)'
b. *éplucher*₁ 'peler (qqch)' → *éplucheur*₂ 'instrument servant à peler (qqch)'
c. *éplucher*₂ 'passer (qqch) au crible' → *éplucheur*₃ 'personne qui passe (qqch) au crible'

Dans ce qui suit, nous développons la distinction entre configurations d'ambiguïté à bases multiples (3.3.1) et unique (3.3.2) en décrivant plusieurs scénarios secondaires pour les relations entre lexèmes pris deux à deux (N_1 et N_2 pour les noms, V_1 et V_2 pour les verbes). Si

certains de ces scénarios peuvent être quantifiés sur la base de nos données actuelles, d'autres nécessiteraient des annotations et analyses supplémentaires dépassant malheureusement le cadre du présent travail. Nous nous satisferons donc, dans leur cas, d'une description qualitative.

Configuration d'ambiguïté	Nombre de noms	Nombre de sens
Bases multiples	1'196	3'153
Base unique	1'110	2'884

TABLEAU 3.4 – Distribution des noms ambigus et sens apparentés en fonction des configurations d'ambiguïté

3.3.1 Bases multiples

Nous l'avons vu dans le chapitre 1, les noms construits ambigus associés à différents lexèmes-bases peuvent se présenter sous deux formes selon que les sens de la base entretiennent ou non un lien sémantique entre eux. Un examen qualitatif des données révèle que ces deux cas de figure ont également cours dans le lexique déverbal en français. Tel qu'il est décrit dans l'exemple (93), le verbe *mater* dont est tiré le nom ambigu *matage* est de toute évidence homonymique, tandis que les sens du verbe *harponner* dont dérive *harponnage* sont liés par une métaphore en (94) :

- (93) a. *mater*₁ 'rendre mat' → *matage*₁ 'action de rendre mat'
b. *mater*₂ 'réprimer' → *matage*₂ 'action de réprimer'
- (94) a. *harponner*₁ 'frapper avec un harpon' → *harponnage*₁ 'action de frapper avec un harpon'
b. *harponner*₂ 'saisir au passage' → *harponnage*₂ 'action de saisir au passage'



FIGURE 3.14 – Configurations d'ambiguïté à bases multiples

Nous schématisons ces deux scénarios respectivement en 3.14a et 3.14b, en figurant également la double motivation parfois possible pour le nom dans les situations où la même figure sémantique (métaphore ou métonymie) s'observe entre les sens du verbe et entre ceux du nom. Dans le cas du dérivé ambigu *harponnage* en (94), par exemple, un lien

3.3. Configurations d'ambiguïté

sémantique transparaît non seulement entre V_2 et N_2 par l'intermédiaire de la construction morphologique, mais aussi entre N_1 et N_2 à travers une métaphore semblable à celle que l'on observe entre *harponner*₁ et *harponner*₂.

Le cas des alternances causatives/anticausatives⁴⁴ (cf. section 2.2.4) comme en (95) est plus délicat :

- (95) a. *européaniser*₁ 'donner un caractère européen' → *européanisation*₁ 'fait de donner un caractère européen'
b. *s'eupéaniser*₂ 'prendre un caractère européen' → *eupéanisation*₂ 'fait de prendre un caractère européen'

Ce genre d'alternance pousse en effet à s'interroger à la fois sur l'orientation de la relation entre les deux formes du verbe et sur la nature du lien sémantique qu'elles entretiennent. Partant du principe que l'alternant anticausatif dérive de l'alternant causatif, certaines études défendent par exemple l'hypothèse que le premier (96a) constitue une extension métonymique du second (96b) (Ruiz de Mendoza Ibáñez & Otal Campo, 2002; Paducheva, 2003; *inter alia*). Ces travaux ont en commun de postuler que le passage de la forme causative à la forme anticausative s'accompagne d'une rétrogradation de la cause ou de l'agent responsable de l'événement, qui reste toutefois récupérable sémantiquement. Cette position est néanmoins contestée. Panther (2003, p. 284) fait par exemple remarquer qu'il existe des événements pour lesquels aucune cause n'est identifiable (96c), mettant de fait en doute l'existence d'un chemin conceptuel entre les deux alternants. Par défaut, nous ne considérons donc pas les alternants causatifs et anticausatifs comme liés par un lien métonymique et les traitons *a priori* comme des cas d'ambiguïté avec bases indépendantes.

- (96) a. *La fenêtre s'est fermée.*
b. *Pierre a fermé la fenêtre.*
c. *La fenêtre s'est fermée toute seule.*

3.3.2 Bases uniques

Examinons maintenant les configurations d'ambiguïté à base unique, en commençant par quelques statistiques descriptives. Le tableau 3.5 présente la distribution des 159 paires verbe-nom dans lesquelles un seul lexème-base est lié à plusieurs lexèmes nominaux en fonction du nombre de sens qu'elles impliquent. Un nom peut être compté plus d'une fois s'il est lié à deux bases verbales différentes et si chacune d'entre elles est en outre associée à plus d'un sens nominal. Le nom *patinage*, par exemple, possède deux sens liés au lexème-base *patiner*₁ 'couvrir de patine', ainsi que deux lectures pouvant être mises en relation avec le lexème-base *patiner*₂ 'glisser' :

44. La littérature parle aussi d'alternances causatives/inchoatives.

- (97) a. *patiner*₁ ‘couvrir de patine’ → *patinage*₁ ‘action de couvrir de patine’
 b. *patiner*₁ ‘couvrir de patine’ → *patinage*₂ ‘patine obtenue’
 c. *patiner*₂ ‘glisser’ → *patinage*₃ ‘action de glisser’
 d. *patiner*₂ ‘glisser’ → *patinage*₄ ‘art de glisser sur la glace’

Nombre de sens	Nombre de paires verbe-nom	Nombre de relations
2	986	986
3	148	418
4	18	97
5	6	49
6	1	15

TABLEAU 3.5 – Distribution des paires verbe-nom où un seul lexème-base est lié à plusieurs lexèmes nominaux en fonction du nombre de sens qu’elles comportent

Ces 1’159 paires fournissent 1’565 relations sémantiques bipartites entre lexèmes nominaux pouvant faire l’objet d’une analyse : par exemple, celle entre les deux sens du nom *annonce* en (98) (*annonce*₁ et *annonce*₂), ou encore celles entre les différents sens du nom trisémiq*ue* *teinture* en (99) (*teinture*₁ et *teinture*₂, *teinture*₁ et *teinture*₃, *teinture*₂ et *teinture*₃).

- (98) a. *annoncer*₁ ‘faire savoir’ → *annonce*₁ ‘action de faire savoir’
 b. *annoncer*₁ ‘faire savoir’ → *annonce*₂ ‘écrit servant à faire savoir’
- (99) a. *teindre*₁ ‘colorer’ → *teinture*₁ ‘action de colorer’
 b. *teindre*₁ ‘colorer’ → *teinture*₂ ‘substance servant à colorer’
 c. *teindre*₁ ‘colorer’ → *teinture*₃ ‘coloration obtenue’

Nous ajoutons au tableau 3.5 la répartition des relations bipartites en fonction du nombre de sens concernés, en écartant certaines relations illogiques impliquant des sens figuratifs n’entretenant pas de relation avec la base verbale. S’agissant du nom trisémiq*ue* *sauterelle* sous (100), par exemple, seules les relations entre *sauterelle*₁ et *sauterelle*₂, et entre *sauterelle*₁ et *sauterelle*₃ sont conservées, celle entre les sens figuratifs *sauterelle*₂ et *sauterelle*₃ étant non pertinente.

- (100) a. *sauter*₁ ‘bondir’ → *sauterelle*₁ ‘insecte’
 b. *sauterelle*₁ ‘insecte’ → *sauterelle*₂ ‘personne maigre’
 c. *sauterelle*₁ ‘insecte’ → *sauterelle*₃ ‘fausse équerre’

Ces relations N_1 - N_2 se répartissent en deux configurations d’ambiguïté principales, esquissées en 3.15a et 3.15b, qui se distinguent par le rapport que le sens N_2 entretient avec la base verbale V_1 . Nous identifions d’une part les extensions sémantiques pures, comme *pilon* en (101), où N_2 est sémantiquement indépendant de V_1 (figure 3.15a). L’ambiguïté se construit

3.3. Configurations d'ambiguïté

ici par étapes, avec l'obtention d'un premier sens par dérivation ($V_1 \rightarrow N_1$), puis la création d'un sens second par une extension sémantique de type métaphorique ou métonymique ($N_1 \rightarrow N_2$). En tout, 113 relations N_1 - N_2 (7 % des 1'565 relations possibles) s'apparentent à ce cas de figure dans nos données. Parmi celles-ci, 72 se rencontrent dans des noms bisémiques, 31 dans des noms trisémiques, 6 dans des noms à 4 sens, et 4 dans des noms à 5 sens.

- (101) a. *piler*₁ 'écraser' → *pilon*₁ 'instrument servant à écraser'
 b. *pilon*₁ 'instrument servant à écraser' → *pilon*₂ 'partie d'une cuisse de volaille'

La seconde configuration que nous repérons concerne les codérivations comme *promenade* en (102), où les deux sens nominaux N_1 et N_2 entretiennent un lien sémantique avec la base verbale V_1 (figure 3.15b). La grande majorité ($n = 1'452$, 93 %) des relations N_1 - N_2 s'apparentent à ce cas de figure dans nos données. 914 d'entre elles concernent un nom bisémique, 387 un nom trisémique, 91 un nom à quatre sens, 45 un nom à 4 sens, et 15 un nom à six sens.

- (102) a. *se promener*₁ 'déambuler' → *promenade*₁ 'action de déambuler'
 b. *se promener*₁ 'déambuler' → *promenade*₂ 'lieu où l'on déambule'

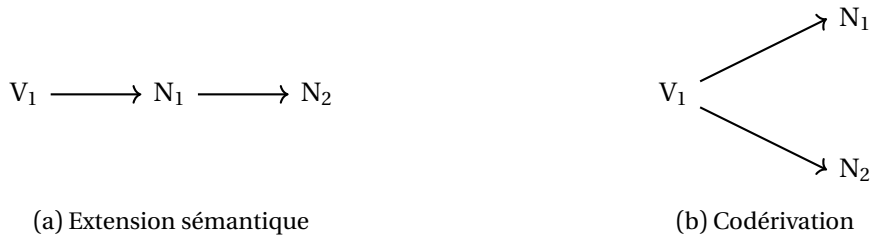


FIGURE 3.15 – Configurations d'ambiguïté à base unique

Un examen qualitatif de la ressource révèle par ailleurs que la codérivation peut prendre plusieurs formes en fonction de la force du lien conceptuel entre V_1 et N_1 d'une part, et entre N_1 et N_2 d'autre part. Nous distinguons et schématisons ci-dessous quatre sous-scénarios prototypiques au sein desquels la construction morphologique et l'extension sémantique, en tant que mécanismes de création de sens, interviennent de manière variable. Notre annotation dans son état actuel ne permet pas de distinguer précisément chacun de ces sous-cas supplémentaires, mais il va sans dire que des analyses quantitatives seraient à l'avenir les bienvenues pour chiffrer précisément ces différents phénomènes.

La première configuration que nous identifions (schématisée dans la figure 3.16) se rapporte aux noms ambigus pour lesquels V_1 , N_1 et N_2 entretiennent tous un lien sémantique relativement fort. Dans le cas de *gribouillage* en (103), par exemple, la relation entre *gribouiller*₁ et les deux sens nominaux *gribouillage*₁ et *gribouillage*₂ est claire. Il en va de même pour celle entre *gribouillage*₁ et *gribouillage*₂, que l'on peut lier par une métonymie. Une

double motivation de N_2 , à la fois morphologique ($V_1 \rightarrow N_1$) et sémantique ($N_1 \rightarrow N_2$), semble ainsi se dessiner.

- (103) a. *gribouiller*₁ ‘griffonner’ → *gribouillage*₁ ‘action de griffonner’
 b. *gribouiller*₁ ‘griffonner’ → *gribouillage*₂ ‘résultat de l’action de griffonner’

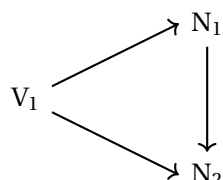


FIGURE 3.16 – Première forme de codérivation. V_1 , N_1 et N_2 entretiennent tous un lien conceptuel fort

Le second scénario que nous distinguons (figure 3.17) concerne les noms ambigus pour lesquels les liens sémantiques entre V_1 et N_1 et entre N_1 et N_2 paraissent intuitivement plus forts que celui entre V_1 et N_2 . Cette configuration semble s’expliquer par des effets de fréquence associés aux procédés morphologiques. Dans nos données, le sens *parloir*₂ en (104) est par exemple le seul lexème nominal construit avec *-oir* s’étant vu attribuer une étiquette sémantique événementielle. Rien de surprenant ici, le suffixe *-oir* étant connu pour former des noms d’entité plutôt que d’éventualité. Le *Robert méthodique* (Rey-Debove, 1985, p. 960-961) fournit par exemple la définition suivante pour ce suffixe : « Élément d’adjectifs et de noms [...] qui signifie “qui fait, sert à, ou contribue à faire” l’action exprimée par la base verbale ».

- (104) a. *parler*₁ ‘discuter’ → *parloir*₁ ‘local où l’on discute’
 b. *parler*₁ ‘discuter’ → *parloir*₂ ‘discussion dans un local’

Il s’ensuit que l’hypothèse de l’extension sémantique ($N_1 \rightarrow N_2$) semble plus probable que celle de la construction morphologique ($V_1 \rightarrow N_2$) dans le cas de *parloir*₂, bien que le lien sémantique entre V_1 et N_2 reste récupérable.

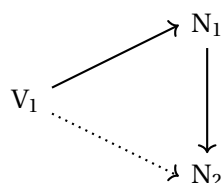


FIGURE 3.17 – Deuxième forme de codérivation. N_2 entretient un lien conceptuel clair avec N_1 mais incertain avec V_1

3.3. Configurations d'ambiguïté

Le troisième scénario pouvant être mis en avant (figure 3.18) se rapporte aux noms ambigus pour lesquels le lien sémantique entre N_1 et N_2 est nébuleux. Les deux sens du nom *causeuse* en (105) se voient par exemple consacrer des articles distincts dans le *Petit Robert* (Éditions Le Robert, s. d.-a) et le *TLFi* (ATILF - CNRS & Université de Lorraine, s. d.). Nous en concluons que ces deux lexèmes nominaux sont vraisemblablement dérivés de manière indépendante, la construction morphologique ($V_1 \rightarrow N_1$ et $V_1 \rightarrow N_2$) prenant ainsi le pas sur l'extension sémantique ($N_1 \rightarrow N_2$).

- (105) a. *causer*₁ 'converser' \rightarrow *causeuse*₁ 'personne qui converse'
b. *causer*₁ 'converser' \rightarrow *causeuse*₂ 'petit canapé'

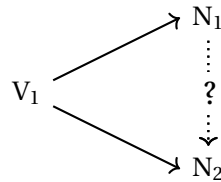


FIGURE 3.18 – Troisième forme de codérivation. N_2 entretient un lien conceptuel plus fort avec V_1 qu'avec N_1

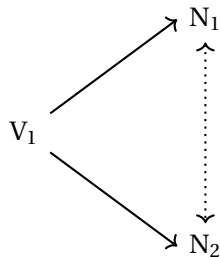


FIGURE 3.19 – Quatrième forme de codérivation. Le lien conceptuel entre N_1 et N_2 est indirect

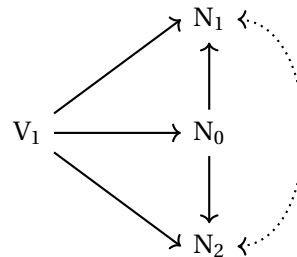


FIGURE 3.20 – Sens événementiel N_0 mettant en relation N_1 et N_2

Enfin, le quatrième et dernier scénario que nous identifions (figure 3.19) se rapporte aux noms pour lesquels les liens sémantiques entre V_1 et N_1 et entre V_1 et N_2 paraissent plus robustes que celui entre N_1 et N_2 . Cette configuration touche principalement des lexèmes correspondant à des noms à trois sens ou plus, comme les lectures instrumentale *teinture*₂ et résultative *teinture*₃ déjà rencontrées dans l'exemple (99) mentionné plus haut.

- (99) a. *teindre*₁ 'colorer' \rightarrow *teinture*₁ 'action de colorer'
b. *teindre*₁ 'colorer' \rightarrow *teinture*₂ 'substance servant à colorer'
c. *teindre*₁ 'colorer' \rightarrow *teinture*₃ 'résultat de l'action de colorer'

Nous faisons ici l'hypothèse, illustrée par la figure 3.20, que les sens *teinture*₂ et *teinture*₃ se retrouvent associés par le truchement de la lecture événementielle *teinture*₁ plutôt que par un lien conceptuel direct. À notre connaissance, en effet, aucune mention n'est faite dans la littérature sur les noms déverbaux d'une forme de polysémie INSTRUMENT/RÉSULTAT sans qu'un troisième sens événementiel ne soit évoqué (voir p. ex. Lieber, 2018), alors même que les alternances régulières ÉVÉNEMENT/INSTRUMENT et ÉVÉNEMENT/RÉSULTAT ont fait l'objet de plusieurs travaux (cf. chapitre 1). Cette forme d'association indirecte a pour conséquence qu'il peut être difficile de postuler une relation orientée entre les deux lexèmes.

3.4 Bilan

Ce chapitre avait pour objectif d'établir le profil sémantique des noms déverbaux et des verbes apparentés figurant dans notre base de données. Nous résumons ci-dessous les principaux résultats de cet aperçu.

Nombre de sens Les noms déverbaux du français et verbes apparentés sont plus souvent monosémiques qu'ambigus. Les verbes sont globalement plus ambigus que les noms.

Types sémantiques Les noms déverbaux dénotent en majorité des formes d'événements transpositionnels, des entités animées agentives et des artéfacts instrumentaux.

Aspect lexical Les lexèmes nominaux événementiels et les verbes de base sont majoritairement dynamiques, duratifs, téliques et sans post-phase. La classe aspectuelle vendliérienne la plus fréquente est celle des activités toutes catégories grammaticales confondues.

Rôles sémantiques De nombreux noms sont dotés d'une structure argumentale. Si programme de complémentation il y a, celui-ci est généralement composé de deux arguments. Le rôle sémantique le plus fréquent pour les premiers arguments nominaux est celui de patient; pour les deuxièmes et troisièmes arguments, celui de d'agent. S'agissant des verbes de base, la plupart d'entre eux comportent deux arguments. La combinaison de fonctions syntaxiques la plus fréquente est sujet-objet. D'un point de vue sémantique, les sujets renvoient généralement à des agents, les objets à des patients, et les obliques à des destinations.

Relations entre propriétés Les différentes propriétés mobilisées dans la description sémantique des noms et verbes apparentés sont toutes associées de manière significative, mais à des degrés divers et souvent de manière asymétrique. Les noms déverbaux du français vérifient la loi sens-fréquence de Zipf (1945) telle qu'opérationnalisée par Casas et ses collègues (2019).

Configurations d'ambiguïté Les sens des noms déverbaux ambigus s'inscrivent dans différents modèles morphosémantiques selon les rapports qu'ils entretiennent les uns avec les autres, de même qu'avec les bases verbales. La construction morphologique

3.4. *Bilan*

et l'extension sémantique, en tant qu'outils de création de sens, ont une influence variable au sein de ces modèles.

Les deux prochains chapitres seront consacrés respectivement aux pans sémantique et morphologique de l'analyse des configurations d'ambiguïté à base unique, où un seul lexème-base est associé à plusieurs lexèmes nominaux. Ces configurations restent en effet les plus intéressantes pour qui souhaite examiner les apports sémantiques et morphologiques dans la construction de l'ambiguïté des noms déverbaux.

Chapitre 4

Associations sémantiques régulières et réseau d'ambiguïté

Nous l'avons vu dans le chapitre 1, les noms déverbaux sont connus pour la fréquence supposément élevée à laquelle ils instancient certaines associations de types sémantiques, renvoyant par exemple volontiers à un événement et un résultat (106b), à un événement et un agent (106b), ou encore à un agent et un instrument (106c).

- (106) a. *collage* 'action de coller' / 'composition artistique', *contrefaçon* 'action de contre-faire' / 'reproduction frauduleuse', *réalisation* 'action de réaliser' / 'entité qui a été réalisée'
- b. *gouvernement* 'action de gouverner' / 'pouvoir qui gouverne', *rébellion* 'action de se rebeller' / 'groupe de personnes qui se rebellent', *immigration* 'action d'immigrer' / 'groupe de personnes qui immigrer'
- c. *balayeuse* 'employée qui balaie les rues' / 'véhicule destiné au balayage des rues', *couveuse* 'poule qui couve' / 'étuve servant à l'incubation', *pétrisseur* 'ouvrier qui pétrit' / 'pétrin mécanique'

Le but de ce chapitre, à l'aune de ce constat, est double. Il s'agira d'une part de quantifier et de décrire ces différents schémas d'association. Nous tenterons notamment de déterminer avec quelle régularité les déverbaux les réalisent au-delà de simples suppositions théoriques, et s'il existe par ailleurs des motifs d'ambiguïté passés sous le radar pour l'instant. Le second objectif que nous nous fixons vise la mise au jour de potentiels mécanismes d'extension sémantique sous-jacents à ces patrons. Un certain nombre d'études quantitatives (Xu et al., 2020; Karjus et al., 2021; Brochhagen & Boleda, 2022; *inter alia*) se sont penchées ces dernières années sur les principes cognitifs généraux façonnant le lexique ambigu dans les langues du monde, tels que l'associativité conceptuelle ou la parenté sémantique. Il reste à établir de

quelle manière ceux-ci se présentent dans le cas des nominalisations déverbaux au-delà des idiosyncrasies morphosémantiques observables.

Une telle entreprise nécessite évidemment des outils adaptés à la nature particulière des données examinées. La méthodologie que nous sélectionnons pour satisfaire ces attentes repose sur deux cadres d'analyse complémentaires, exploratoires plutôt qu'inférentiels. Nous effectuerons premièrement une analyse d'association (Agrawal et al., 1993) afin de détecter des règles intéressantes et orientées de forme $\{A\} \Rightarrow \{B\}$ entre types sémantiques, pouvant être interprétées de la manière suivante : si un nom instancie un type sémantique A (p. ex. Artéfact-RÉSULTAT), alors il est fort probable qu'il réalise aussi un type sémantique B (p. ex. Événement-TRANSPOSITION)⁴⁵. Il s'agit en d'autres termes de « corrélations entre attributs exprimées sous forme de règles » (Barra et al., 2021, p. 215). Sera menée dans un second temps une analyse de réseau (cf. Wasserman & Faust, 1994; Newman, 2018) basée sur les relations précédemment identifiées, dans le but cette fois-ci de pouvoir apprécier les types sémantiques en tant qu'entités interagissant au sein d'une structure globale. En guise d'exemple, imaginons que 5 règles d'association orientées entre types sémantiques puissent être détectées pour les noms déverbaux : $\{A\} \Rightarrow \{B\}$, $\{C\} \Rightarrow \{B\}$, $\{D\} \Rightarrow \{B\}$, $\{D\} \Rightarrow \{C\}$ et $\{D\} \Rightarrow \{E\}$. La représentation de ces dernières sous la forme d'un réseau (figure 4.1) permet – entre autres choses – de visualiser rapidement les types sémantiques (D , et dans une moindre mesure, A et C) dont la présence est conditionnée à celle d'autres types, et à l'inverse, les types (B , et dans une moindre mesure, C et E) dont d'autres dépendent.

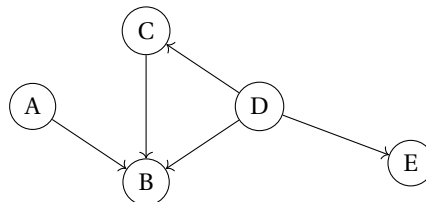


FIGURE 4.1 – Exemple de réseau de règles d'association orientées entre types sémantiques

Le chapitre est organisé comme suit. Nous proposerons dans la section 4.1 un aperçu fréquentiel des associations entre types ontologiques (p. ex. Objet cognitif, Maladie), relationnels (p. ex. CAUSE, MANIÈRE) et combinés (p. ex. Artefact-INSTRUMENT, Événement-TRANSPOSITION) pouvant être observées dans notre base de données. Cet examen initial nous amènera, dans la section 4.2, à l'extraction de règles d'association pour les types combinés. La section 4.3 sera quant à elle dévolue à la construction et à l'analyse descriptive d'un réseau général de relations sémantiques.

45. Les règles d'association peuvent aussi naturellement accommoder plus de deux types sémantiques, puisqu'il existe des règles de forme $\{A, B\} \Rightarrow \{C\}$.

4.1 Relations sémantiques observées

Les sections suivantes présentent les associations les plus fréquentes entre types ontologiques (4.1.1), relationnels (4.1.2) et combinés (4.1.3) dans notre base de données. L'objectif étant d'analyser les cas d'ambiguïté dans les configurations de « base unique » (*vs* de « bases multiples »), nos objets d'étude sont ici les noms distingués d'après l'appariement à un même lexème-base. Par commodité, nous appellerons désormais ces noms « vocables ». Deux vocables seront par exemple considérés pour le nom *vol* : le vocable monosémique *vol_A*, que l'on peut lier au lexème-base *voler₁* 'dérober' (107a) ; et le vocable bisémique *vol_B*, associé au lexème-base *voler₂* 'se déplacer' (107b).

- (107) a. *voler₁* 'dérober' → *vol_A* 'action de dérober'
 b. *voler₂* 'se déplacer' → *vol_B* 'action de se déplacer' / 'groupe d'animaux ou d'insectes qui se déplacent'

En tout, la base de données contient 6'841 vocables nominaux comme *vol_A* et *vol_B*.

4.1.1 Types ontologiques

Nous examinons dans ce qui suit deux formes d'association entre types ontologiques : les associations complètes et les associations bipartites. Les premières correspondent aux ensembles complets de types ontologiques instanciés par les 6'841 vocables. Par exemple, les alternances complètes considérées sont Événement/Artéfact pour le vocable bisémique *gribouillage* (108), Événement/Domaine/Artéfact pour le vocable triséémique *sculpture* (109), et Événement/Artéfact/Artéfact pour le vocable triséémique *teinture* (110) :

- (108) a. *gribouillage₁* 'action de gribouiller' [Événement]
 b. *gribouillage₂* 'dessin gribouillé' [Artéfact]
- (109) a. *sculpture₁* 'action de sculpter' [Événement]
 b. *sculpture₂* 'art de sculpter' [Domaine]
 c. *sculpture₃* 'œuvre sculptée' [Artéfact]
- (110) a. *teinture₁* 'action de teindre' [Événement]
 b. *teinture₂* 'substance servant à teindre' [Artéfact]
 c. *teinture₃* 'coloration obtenue' [Artéfact]

La seconde forme d'association que nous considérons correspond aux ensembles de types ontologiques pris deux à deux. Au total, les 6'841 vocables examinés mettent en jeu 1'565 paires de types sémantiques pouvant faire l'objet d'une analyse. Parmi ces dernières figurent par exemple :

4.1. Relations sémantiques observées

- la paire Événement/Artéfact (*gribouillage*₁ / *gribouillage*₂) pour *gribouillage* en (108);
- les paires Événement/Domaine (*sculpture*₁ / *sculpture*₂), Événement/Artéfact (*sculpture*₁ / *sculpture*₃) et Domaine/Artéfact (*sculpture*₂ / *sculpture*₃) pour *sculpture* en (109);
- les paires Événement/Artéfact (*teinture*₁ / *teinture*₂), Événement/Artéfact (*teinture*₁ / *teinture*₃) et Artéfact/Artéfact (*teinture*₂ / *teinture*₃) pour *teinture* en (110).

Associations complètes

Au total, 230 associations complètes distinctes peuvent être relevées pour les types ontologiques. La majorité d'entre elles ($n = 140$, 61 %) ne sont instanciées que par un vocable, à l'instar de l'alternance Objet naturel/Quantité réalisée par *picotin* 'ration d'avoine' / 'mesure valant environ 2 litres et demi'. Le tableau 4.1 présente les 10 associations complètes les plus fréquentes.

Association	<i>n</i>	%	Exemple
Anm/Art	213	3.1	<i>balayeuse</i> 'personne qui balaie' / 'véhicule servant à balayer'
Evt/Art	141	2.1	<i>gribouillage</i> 'action de gribouiller' / 'dessin gribouillé'
Evt/Nat	64	0.9	<i>brûlis</i> 'action de brûler' / 'essart'
Evt/Ppt	50	0.7	<i>goinfrerie</i> 'repas' / 'voracité'
Evt/Dom	41	0.6	<i>voltige</i> 'action de voltiger' / 'art de voltiger'
Art/Art	35	0.5	<i>urinoir</i> 'édifice public' / 'cuvette'
Cog*Evt/Ppt	27	0.4	<i>flatterie</i> 'fausse louange' / 'caractère flatteur'
Art*Ist/Dom	18	0.3	<i>distillerie</i> 'usine où l'on distille' / 'industrie de la distillation'
Evt/Evt-Coll	18	0.3	<i>bousculade</i> 'heur rapide' / 'remous de foule'
Nat/Art	17	0.2	<i>pelure</i> 'peau de fruit ou de légume' / 'vêtement'

TABEAU 4.1 – Associations complètes les plus fréquentes entre types ontologiques. *n* = Nombre total de vocables réalisant l'association. % = Pourcentage de vocables réalisant l'association

Impliquant toutes deux sens, les associations les plus courantes témoignent d'une claire préférence pour les domaines des éventualités (Événement, Événement*Cognitif, Événement-Collectif, Domaine, Propriété) et des objets concrets, qu'ils soient construits (Artéfact) ou naturels (Objet naturel). Un autre résultat notable concerne la présence marquée de patrons ne faisant intervenir que des éventualités (Événement/Propriété, Événement/Domaine, Événement*Cognitif/Domaine, Événement/Événement-Collectif), pourtant peu relevés dans la littérature sur la sémantique des nominalisations⁴⁶. Le motif Événement/Événement-Collectif semble à cet égard particulièrement intrigant. En effet, les travaux sur les noms collectifs (Flaux, 1999; Lammert & Lecolle, 2014; de Vries, 2021; *inter alia*) ne font pas mention

46. Mais voir Montermini (2015) au sujet des associations entre sens d'événement et de propriété.

de l'existence d'éventualités plurielles – à notre connaissance du moins. Enfin, la présence de types complexes (Événement*Cognitif, Artéfact*Institution) parmi les associations les plus fréquentes révèle que la régularité entre sens peut aussi se doubler d'une régularité entre facettes sémantiques.

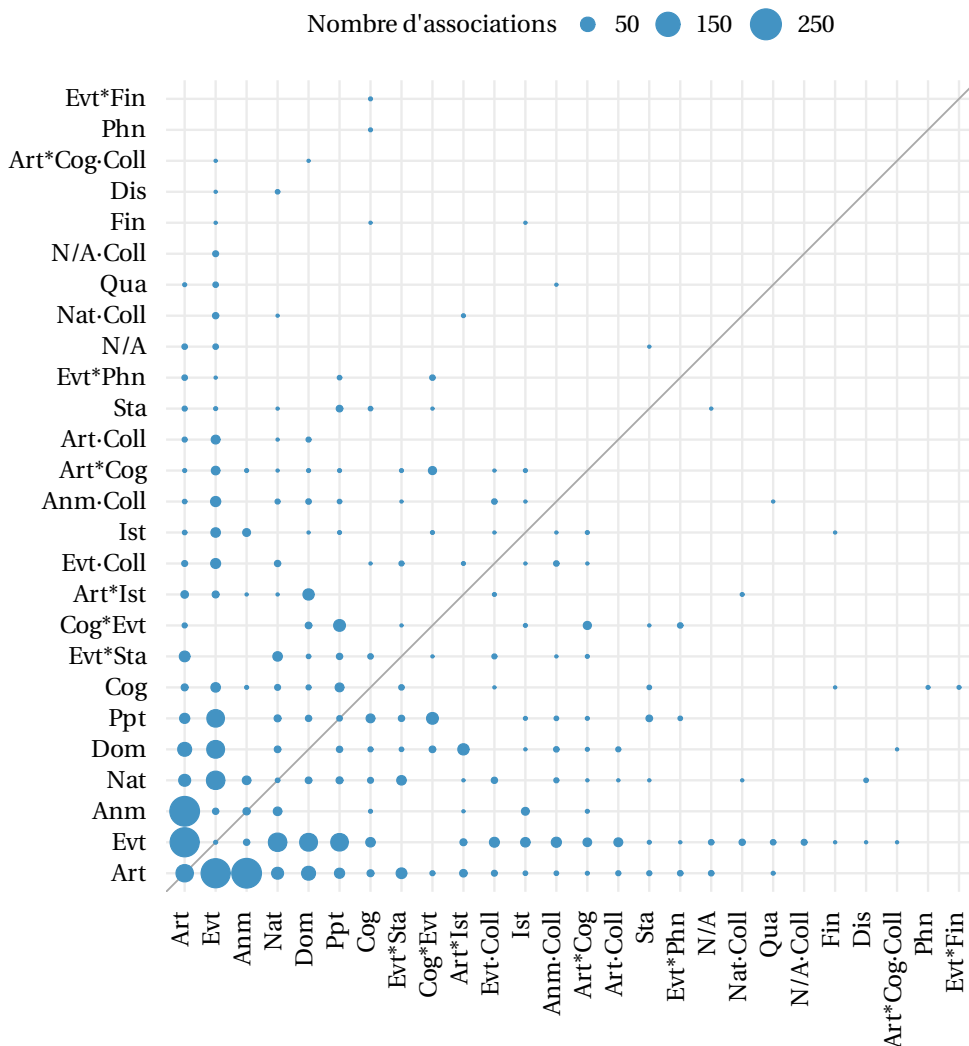


FIGURE 4.2 – Associations bipartites entre types ontologiques. Seules les associations observées au moins deux fois dans l'ensemble des données sont représentées. Les types sémantiques sont ordonnés de gauche à droite et de bas en haut par ordre de fréquence décroissant

Associations bipartites

Au total, 161 associations bipartites distinctes (p. ex. Événement/Artéfact, Artéfact/Artéfact) peuvent être relevées pour les types ontologiques. La figure 4.2 présente les 102 associations bipartites (63 % des 161 associations de départ) instanciées au moins 2 fois. La

4.1. Relations sémantiques observées

fréquence d'apparition de ces associations varie considérablement. Alors que 74 d'entre elles sont observées moins de 10 fois, une poignée d'associations se distinguent néanmoins par leur prépondérance, à savoir Entité animée/Artéfact ($n = 230$ paires sur 1'565), Événement/Artéfact ($n = 221$), Événement/Objet naturel ($n = 84$), Événement/Domaine ($n = 78$), Événement/Propriété ($n = 76$) et Artéfact/Artéfact ($n = 74$). Ces associations bipartites correspondent à certaines associations complètes évoquées précédemment (se référer au tableau 4.1 pour des exemples spécifiques).

Nonobstant la fréquence d'instanciation des associations, les résultats révèlent par ailleurs que certains types sont plus « généralistes » que d'autres, au sens où ils admettent davantage d'associations différentes. De ce point de vue, les types les plus importants sont Événement (qui s'associe au moins deux fois à 21 types autres que lui-même), Artéfact (18), Objet naturel (15), Domaine (13), Propriété (12) et Objet cognitif (12). À l'autre extrémité du continuum, 3 types ontologiques (N/A-Collectif, Événement*Financier, Phénomène) ne s'associent qu'à une seule autre catégorie.

4.1.2 Types relationnels

Passons maintenant aux types relationnels. Comme pour les types ontologiques, nous considérons ci-dessous les associations complètes et les associations bipartites. Les 113 lexèmes sans lien sémantique direct avec une base verbale (cf. chapitre 2, section 2.2.1), comme *pilon*₂ en (111b), ont été rassemblés au préalable sous une bannière unique FIGURATIF (111c).

- | | | | |
|-------|----|---|--------------------------|
| (111) | a. | <i>pilon</i> ₁ 'instrument servant à écraser' | [INSTRUMENT] |
| | b. | <i>pilon</i> ₂ 'partie d'une cuisse de volaille' | [(INSTRUMENT)·FIGURATIF] |
| | c. | <i>pilon</i> ₂ 'partie d'une cuisse de volaille' | [FIGURATIF] |

Associations complètes

Au total, 161 associations complètes différentes ont été identifiées pour les types relationnels. La majorité d'entre elles ($n = 97$, 60 %) ne sont observées qu'une seule fois. Le tableau 4.2 présente les 10 associations complètes les plus fréquentes. Il est intéressant de constater que la plupart de ces associations comprennent à la fois un sens transpositionnel et un sens de participant à une éventualité (p. ex. résultat, instrument, agent, lieu), soulignant ainsi l'importance des relations de forme thématique pour les déverbaux. La prépondérance de l'association AGENT/INSTRUMENT, bien documentée dans la littérature (voir p. ex. Booij, 1986; Luschützky & Rainer, 2011), se trouve par ailleurs confirmée. Plus surprenante, l'association INSTRUMENT/FIGURATIF montre que l'ambiguïté des déverbaux ne se limite pas nécessairement à des associations entre lexèmes entretenant un lien sémantique avec les verbes de base, et que les extensions sémantiques se construisent volontiers à partir du sens instrumental. Enfin, la présence d'une association à trois composantes

Association	<i>n</i>	%	Exemple
TSP/RES	234	3.4	<i>pliage</i> 'action de plier'/'objet plié'
AGT/INS	194	2.8	<i>testeur</i> 'personne qui teste'/'objet servant à tester'
TSP/TSP	86	1.3	<i>dressage</i> 'action de dresser un animal'/'art équestre'
TSP/CAU	77	1.1	<i>reniflette</i> 'action de renifler'/'inflammation de la muqueuse nasale'
TSP/INS	67	1.0	<i>rafraîchissement</i> 'action de rafraîchir'/'boisson'
TSP/AGT	47	0.7	<i>relève</i> 'action de remplacer'/'personnes qui remplacent'
TSP/THM	25	0.4	<i>livraison</i> 'action de livrer'/'marchandise livrée'
TSP/LOC	20	0.3	<i>baignade</i> 'action de se baigner'/'lieu où l'on se baigne'
INS/FIG	19	0.3	<i>éventail</i> 'accessoire pour (s')éventer'/'ensemble de choses'
TSP/INS/RES	15	0.2	<i>scan</i> 'action de scanner'/'numériseur'/'document scanné'

TABLEAU 4.2 – Associations complètes les plus fréquentes entre types relationnels. *n* = Nombre total de vocables réalisant l'association. % = Pourcentage de vocables réalisant l'association

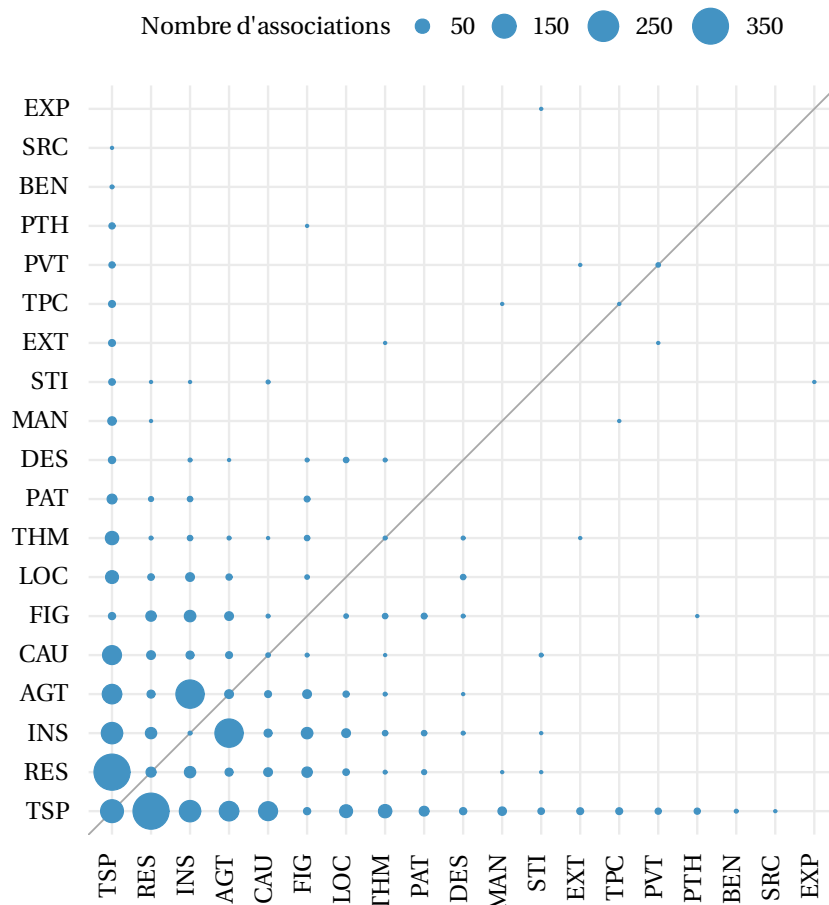


FIGURE 4.3 – Associations bipartites entre types relationnels. Seules les associations observées au moins deux fois dans l'ensemble des données sont représentées. Les types sémantiques sont ordonnés de gauche à droite et de bas en haut par ordre de fréquence décroissant

4.1. Relations sémantiques observées

(TRANSPOSITION/INSTRUMENT/RÉSULTAT) parmi les schémas les plus fréquents vient confirmer de manière quantitative une hypothèse déjà formulée dans la littérature, selon laquelle les noms déverbaux peuvent se soumettre à des patrons de polysémie régulière impliquant plus de deux sens (cf. Ježek, 2008; Huyghe, 2014; Lieber, 2018).

Associations bipartites

En tout, 90 associations bipartites différentes peuvent être observées parmi les 1'565 paires de types sémantiques relationnels. Environ un tiers de ces associations ($n = 29$, 32 %) ne se manifestent qu'une seule fois. La figure 4.3 présente les associations instanciées au moins deux fois. Les plus fréquentes correspondent pour l'essentiel aux associations complètes citées ci-dessus, à savoir TRANSPOSITION/RÉSULTAT ($n = 354$), AGENT/INSTRUMENT ($n = 214$), TRANSPOSITION/TRANSPOSITION ($n = 137$), TRANSPOSITION/INSTRUMENT ($n = 119$) et TRANSPOSITION/AGENT ($n = 98$).

Les résultats montrent également que certains types relationnels admettent un plus grand nombre d'associations différentes. Sans surprise, le type TRANSPOSITION se distingue comme étant le plus généraliste, s'associant à 17 autres types au moins deux fois. Il est suivi des types RÉSULTAT (10), INSTRUMENT (10), FIGURATIF (10) et AGENT (8).

4.1.3 Types combinés

Nous concluons cet aperçu fréquentiel par l'examen des associations complètes et bipartites entre types combinés, en rassemblant pour ce faire les typologies ontologique et relationnelle. Dans un souci de cohérence avec l'analyse proposée pour les types relationnels, nous simplifions les étiquettes contenant un type figuratif. Ainsi, l'étiquette retenue pour *pilon*₂ est celle présentée en (112c) plutôt que celle en (112b).

- | | | | |
|-------|----|---|-----------------|
| (112) | a. | <i>pilon</i> ₁ 'objet servant à écraser' | [Art-INS] |
| | b. | <i>pilon</i> ₂ 'partie d'une cuisse de volaille' | [Nat-(INS-)FIG] |
| | c. | <i>pilon</i> ₂ 'partie d'une cuisse de volaille' | [Nat-FIG] |

Associations complètes

Un total de 384 associations complètes différentes peuvent être identifiées pour les types combinés à partir des 6'841 vocables nominaux. Parmi elles, 274 (71 %) ne sont réalisées qu'une seule fois. Le tableau 4.3 présente les 10 associations les plus fréquentes. Dans l'ensemble, les résultats obtenus ici reflètent ceux se rapportant aux classifications ontologique et relationnelle. Tous les schémas les plus courants sont composés de 2 sens, dont l'un peut éventuellement présenter des facettes sémantiques (Artéfact*Institution, Événement*État, Événement*Cognitif). Le patron « événement/résultat », sujet privilégié dans la littérature sur la sémantique des noms déverbaux (cf. Grimshaw, 1990; Jacquy, 2006; Bisetto & Melloni,

2007; Lieber, 2018; *inter alia*), est également bien représenté. Toutefois, nous relevons qu'il peut se décliner de différentes manières d'un point de vue ontologique, tant sur le plan événementiel (Événement, Événement*État) que résultatif (Objet naturel, Artéfact).

Association	<i>n</i>	%	Exemple
Anm-AGT/Art-INS	191	2.8	<i>arroreuse</i> 'personne qui arrose'/'véhicule qui sert à arroser'
Evt-TSP/Art-RES	83	1.2	<i>photocopie</i> 'action de photocopier'/'document photocopie'
Evt-TSP/Nat-RES	50	0.7	<i>écorchement</i> 'action d'écorcher'/'déchirure légère'
Evt-TSP/Dom-TSP	41	0.6	<i>jonglage</i> 'action de jongler'/'art de jongler'
Evt-TSP/Ppt-CAU	36	0.5	<i>abstinence</i> 'action de s'abstenir'/'tempérance'
Evt-TSP/Art-INS	27	0.4	<i>autotest</i> 'action de s'autotester'/'outil de dépistage'
Cog*Evt-TSP/Ppt-CAU	25	0.4	<i>radoterie</i> 'discours radoté'/'propension à radoter'
Evt-TSP/Evt-Coll-TSP	18	0.3	<i>exposition</i> 'action d'exposer'/'présentation publique d'œuvres d'art'
Art*Ist-AGT/Dom-TSP	17	0.2	<i>imprimerie</i> 'établissement qui imprime'/'art d'imprimer'
Evt*Sta-TSP/Nat-RES	15	0.2	<i>gonflement</i> 'action de (se) gonfler'/'boursouflure'

TABEAU 4.3 – Associations complètes les plus fréquentes entre types combinés. *n* = Nombre total de vocables réalisant l'association. % = Pourcentage de vocables réalisant l'association

Associations bipartites

Au total, 400 associations bipartites différentes peuvent être observées parmi les 1'565 paires de types sémantiques combinés. 240 (60 %) d'entre elles ne sont réalisées qu'une seule fois. Les plus fréquentes sont présentées dans la figure 4.4. On y retrouve notamment les associations Entité animée-AGENT/Artéfact-INSTRUMENT (*n* = 202), Événement-TRANSPOSITION/Artéfact-RÉSULTAT (*n* = 120), Événement-TRANSPOSITION/Domaine-TRANSPOSITION (*n* = 78), Événement-TRANSPOSITION/Objet naturel-RÉSULTAT (*n* = 61) et Événement-TRANSPOSITION/Artéfact-INSTRUMENT (*n* = 51). Toutes ces associations bipartites ont déjà été identifiées parmi les associations complètes les plus fréquentes, répertoriées dans le tableau 4.3.

Pour ce qui est de la diversité des associations admises, les types les plus généralistes sont principalement événementiels et transpositionnels. En tête de liste, le type Événement-TRANSPOSITION se distingue en s'associant au moins deux fois avec 44 autres types combinés. Viennent ensuite les types Artéfact-INSTRUMENT (24), Domaine-TRANSPOSITION (20), Entité animée-AGENT (12) et Événement*État-TRANSPOSITION (12).

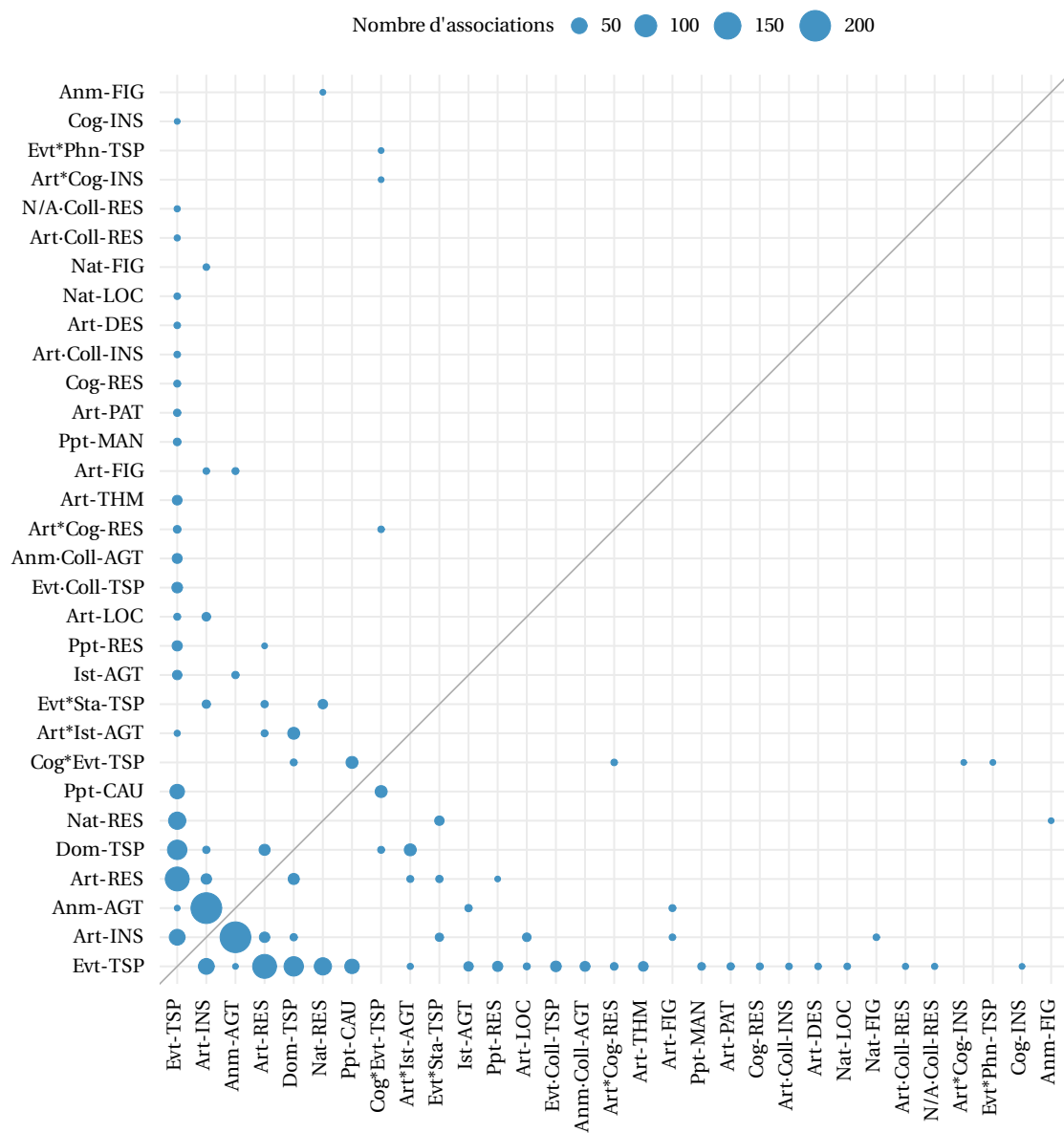


FIGURE 4.4 – Associations bipartites entre types combinés. Seules les associations observées au moins six fois dans l'ensemble des données sont représentées. Les types sémantiques sont ordonnés de gauche à droite et de bas en haut par ordre de fréquence décroissant

4.2 Règles d'association

Les tendances d'associations que nous venons de décrire paraissent somme toute instructives, mais leur examen peut encore être développé. En effet, une simple description fréquentielle des associations ne suffit pas à déterminer si celles-ci sont le fruit du hasard, ni à évaluer la place qu'occupent les noms monosémiques dans le système sémantique des nominalisations. Rien ne dit, par exemple, qu'une relation entre deux types *A* et *B* relativement fréquente au regard des autres associations ne soit pas marginale proportionnellement au nombre total de noms – monosémiques ou non – qui réalisent chaque type.

Pour approfondir notre enquête, nous recourons donc à une méthode d'analyse de données non supervisée largement éprouvée en *data mining* : l'analyse de règles d'association (Agrawal et al., 1993). Comme son nom l'indique, ce type d'analyse vise à la détection d'associations, de relations ou de dépendances « intéressantes⁴⁷ » entre entités au sein de grands ensembles d'éléments de données (Cios et al., 2007, p. 289). Il peut s'agir de produits souvent achetés ensemble dans un magasin, de comorbidités récurrentes chez plusieurs patient-es, ou encore de médicaments fréquemment prescrits dans une même ordonnance (voir p. ex. Belyi et al., 2016 ; Hernández et al., 2019).

À notre connaissance, l'extraction de règles d'association reste peu exploitée en linguistique fondamentale (voir tout de même Gries, 2009 ; Gries et al., 2020). Elle semble pourtant particulièrement adaptée à notre étude, dans la mesure où elle s'accommode très bien d'ensembles de données éparpillées comportant de nombreuses observations (comme des noms), mais peu d'informations par observation (comme les types sémantiques instanciés par des noms) (Chapman & McDonnell Feit, 2019, p. 360). En pratique, nous cherchons ainsi à extraire des règles orientées comme celle en (113), qui indique que les types Artéfact-RÉSULTAT et Événement-TRANSPOSITION sont non seulement corrélés dans notre ressource, mais qu'un nom réalisant le premier type a de fortes chances de réaliser également le second.

- (113) {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION}
moulage 'objet moulé' / 'action de mouler'

La section 4.2.1 détaille la méthode utilisée pour l'extraction initiale, ainsi que la sélection des règles d'association entre types sémantiques combinés à partir de notre base de données⁴⁸. Les résultats finaux sont présentés et analysés dans la section 4.2.2.

47. Les indices servant à évaluer « l'intérêt » des règles sont nombreux. Nous en détaillons certains dans les sous-sections suivantes.

48. L'infrastructure pour extraire, manipuler et analyser des règles d'association est fournie par le package *arules* (Hahsler et al., 2005) dans R. Des visualisations interactives peuvent être obtenues avec son extension *arulesViz* (Hahsler, 2017).

4.2. Règles d'association

4.2.1 Extraction

Préparation des données

La première étape de l'extraction a consisté à reformater nos données. En effet, l'algorithme employé par la suite pour la détection des règles a été conçu à l'origine pour l'analyse de paniers d'achats (*market basket analysis* ou *MBA* en anglais). La procédure suivie pour arriver au format requis est résumée à travers les tableaux 4.4, 4.5 et 4.6. À partir de la liste des vocables nominaux, lexèmes verbaux apparentés et types sémantiques (tableau 4.4), il s'agissait dans un premier temps de générer une matrice booléenne exprimant la (non-)réalisation de types sémantiques par les vocables nominaux (tableau 4.5), puis de créer à partir de celle-ci un fichier de « transactions », où les différentes paires verbe-nom constituent des « paniers », et les types sémantiques, les « items » présents dans ces paniers (tableau 4.6).

La transformation de nos données en « transactions » a malheureusement conduit à l'exclusion de deux formes d'ambiguïté, qui restent toutefois très marginales au regard de l'ensemble des lexèmes (1.4 % des 8'206 lexèmes nominaux présents dans la ressource). D'une part, 113 sens figuratifs sans lien sémantique direct avec une base verbale ont été supprimés, comme *allumette*₂ (114b), dérivé de *allumette*₁ (114a). Dans le cas des noms ayant trois sens ou plus, l'algorithme aurait en effet pu associer les sens figuratifs à des sens nominaux avec lesquels ils n'entretiennent pas de relation sémantique réelle, ce qui aurait biaisé l'analyse.

- (114) a. *allumette*₁ 'bâtonnet de bois inflammable' [Artéfact-INSTRUMENT]
 b. *allumette*₂ 'pâtisserie à base de pâte feuilletée' [Artéfact-FIGURATIF]

La seconde forme d'ambiguïté qu'il a fallu écarter concerne 5 cas d'association impliquant deux fois le même type combiné (p. ex. Événement-TRANSPPOSITION/Événement-TRANSPPOSITION pour *monte* 'accouplement'/'trot à cheval'). Seule la (non-)réalisation de types pouvant être prise en compte par l'algorithme (cf. tableau 4.5), les types sémantiques réalisés à double par des noms n'ont été comptés qu'une seule fois.

Vocable nominal	Lexème-base	Type combiné
<i>aiguïsoir</i>	<i>aiguïser</i> ₁	Art-INS
...		
<i>sculpture</i>	<i>sculpter</i> ₁	Evt-TSP
<i>sculpture</i>	<i>sculpter</i> ₁	Dom-TSP
<i>sculpture</i>	<i>sculpter</i> ₁	Art-RES
...		
<i>vol</i>	<i>voler</i> ₁	Evt-TSP
<i>vol</i>	<i>voler</i> ₂	Evt-TSP
<i>vol</i>	<i>voler</i> ₂	Anm-Coll-TSP

TABLEAU 4.4 – Extrait de la base de données d'origine

	Anm-Coll-AGT	Art-INS	Art-RES	Dom-TSP	Evt-TSP
<i>aiguisoir ~ aiguiser</i> ₁	FAUX	VRAI	FAUX	FAUX	FAUX
...					
<i>sculpture ~ sculpter</i> ₁	FAUX	FAUX	VRAI	VRAI	VRAI
...					
<i>vol ~ voler</i> ₁	FAUX	FAUX	FAUX	FAUX	VRAI
<i>vol ~ voler</i> ₂	VRAI	FAUX	FAUX	FAUX	VRAI

TABEAU 4.5 – Matrice booléenne exprimant la (non-)réalisation de types sémantiques par certains vocables

Panier	Contenu
<i>aiguisoir~aiguiser</i> ₁	Art-INS
...	
<i>sculpture~sculpter</i> ₁	Art-RES, Dom-TSP, Evt-TSP
...	
<i>vol~voler</i> ₁	Evt-TSP
<i>vol~voler</i> ₂	Anm-Coll-AGT, Evt-TSP

TABEAU 4.6 – Exemples de « transactions » pouvant être extraites de la base de données

La base de données adaptée contient finalement 6'841 vocables nominaux – également appelés « paniers » – pour 113 types sémantiques – ou « items » – totalisant 8'088 occurrences sur l'ensemble des paniers. Nous comptons par exemple dans le tableau 4.6 une occurrence du type Artéfact-INSTRUMENT, une du type Artéfact-RÉSULTAT, une du type Domaine-TRANSPOSITION, une du type Entité animée-Collectif-AGENT, ainsi que trois occurrences du type Événement-TRANSPOSITION. Conformément aux observations du chapitre précédent, 84 % des paniers ne contiennent qu'un seul item, 14 % en comportent deux, et 2 % en ont trois ou plus. Les 5 items les plus fréquents sont Événement-TRANSPOSITION (27 % des 8'088 occurrences), Entité animée-AGENT (24 %), Événement*État-TRANSPOSITION (11 %), Artéfact-INSTRUMENT (8 %) et Événement*Cognitif-TRANSPOSITION (4 %). Le reste des types représentent collectivement environ 27 % des occurrences.

Application de l'algorithme

La deuxième étape de l'extraction a consisté à appliquer l'algorithme *Apriori* (Agrawal et al., 1993; Agrawal et al., 1996) aux données transformées afin de déterminer quels types sémantiques sont le plus souvent associés. Les cooccurrences ainsi obtenues sont formulées sous la forme de règles d'association orientées de type $\{A\} \Rightarrow \{B\}$, où l'antécédent $\{A\}$ correspond à un ou plusieurs types sémantiques, le conséquent $\{B\}$ à un type sémantique unique,

4.2. Règles d'association

et où $\{A\}$ et $\{B\}$ sont nécessairement différents⁴⁹. L'intérêt des règles est généralement évalué de prime abord sur la base de deux indices appelés « support » et « confiance ». Le support représente dans le cadre de cette étude la proportion de vocables examinés réalisant tous les types sémantiques inclus dans l'antécédent $\{A\}$ et le conséquent $\{B\}$ d'une règle d'association $\{A\} \Rightarrow \{B\}$:

$$\begin{aligned} support(A \Rightarrow B) &= P(A \cap B) \\ &= \frac{\text{Nombre de vocables réalisant } A \text{ et } B}{\text{Nombre total de vocables}} \end{aligned}$$

La confiance permet quant à elle d'évaluer la précision d'une règle. Pour une règle d'association $\{A\} \Rightarrow \{B\}$, elle représente en effet la proportion de vocables réalisant l'antécédent $\{A\}$ qui instantient également le conséquent $\{B\}$:

$$\begin{aligned} confiance(A \Rightarrow B) &= P(B|A) = \frac{P(A \cap B)}{P(A)} \\ &= \frac{\text{Nombre de vocables réalisant } A \text{ et } B}{\text{Nombre de vocables réalisant } A} \end{aligned}$$

L'exemple (115) ci-dessous présente une règle d'association montrant que les noms qui réalisent les types Artéfact-INSTRUMENT et Artéfact-RÉSULTAT instantient également le type Événement-TRANSPPOSITION. Dans nos données, 15 vocables sur 6'841 réalisent les trois types sémantiques considérés, ce qui équivaut à un support de 0.2 %. On observe par ailleurs que 21 vocables investissent à la fois les types Artéfact-INSTRUMENT et Artéfact-RÉSULTAT (avec ou sans le type Événement-TRANSPPOSITION). Il s'ensuit que la confiance de la règle posée s'élève à 15/21, soit 71.4 %.

- (115) $\{\text{Art-RES}, \text{Art-INS}\} \Rightarrow \{\text{Evt-TSP}\}$
 Support de 0.2 % et confiance de 71.4 %
teinture 'substance servant à teindre' / 'coloration obtenue' / 'action de teindre'

L'un des problèmes majeurs liés à la recherche de cooccurrences « d'intérêt » tient au fait que le nombre de règles d'association possibles croît de manière exponentielle en fonction du nombre d'attributs examinés (Larose & Larose, 2019, p. 199). De manière à optimiser ce processus, les indices de support et de confiance font également office de paramètres d'extraction devant être spécifiés par l'analyste lorsqu'elle ou il génère des règles d'association. L'algorithme *Apriori*, certainement le plus utilisé à l'heure actuelle, met en jeu deux étapes principales (Tufféry, 2012, p. 323; Cios et al., 2007, p. 298) : (i) la recherche de sous-groupes d'items présentant un support supérieur à un seuil choisi par l'analyste dans la liste de transactions, puis (ii) la décomposition de ces sous-groupes en règles dont l'indice de confiance

49. Cette dernière propriété évite de générer des règles insignifiantes comme $\{\text{Artéfact-INSTRUMENT}, \text{Entité animée-AGENT}\} \Rightarrow \{\text{Artéfact-INSTRUMENT}\}$ (Larose & Larose, 2019, p. 200).

dépasse également un seuil prédéfini. En substance, il s'agit ainsi d'extraire des cooccurrences suffisamment fréquentes, puis d'identifier parmi celles-ci des règles « intéressantes ». Il va sans dire que le nombre de règles obtenues dépend fortement des valeurs sélectionnées pour les paramètres. Comme le précisent Chapman et McDonnell Feit (2019, p. 345), ces valeurs sont généralement déterminées par tâtonnement. L'objectif reste dans tous les cas de parvenir à un ensemble raisonnable de règles comprenant un nombre adéquat d'items.

Les noms ambigus étant minoritaires parmi les nominalisations étudiées, des seuils de support et de confiance relativement faibles ont été choisis pour cette étude : 0.1 % pour le support, et 10 % pour la confiance. Suivant ces seuils, le sous-ensemble en (116) a par exemple été retenu à l'étape (i), car son support est de 0.2 %.

- (116) {Art-RES, Art-INS, Evt-TSP}
Support de 0.2 %
teinture 'substance servant à teindre' / 'coloration obtenue' / 'action de teindre'

Toutes les règles pouvant être dérivées de ce sous-ensemble à l'étape (ii) ont également été conservées puisque leurs scores de confiance sont supérieurs à 10 % :

- (117) a. {Art-RES, Art-INS} \Rightarrow {Evt-TSP}
Support de 0.2 % et confiance de 71.4 %
b. {Art-RES, Evt-TSP} \Rightarrow {Art-INS}
Support de 0.2 % et confiance de 12.5 %
c. {Art-INS, Evt-TSP} \Rightarrow {Art-RES}
Support de 0.2 % et confiance de 29.4 %

L'extraction effectuée selon les seuils précités de support et de confiance a abouti à la génération de 42 règles englobant 59 types sémantiques distincts (52 % des 113 types examinés initialement). Parmi ces règles, 33 contiennent un type sémantique unique en antécédent, tandis que les 9 restantes en présentent deux. Leurs propriétés sont résumées dans le tableau 4.7. Quoiqu'instructifs, ces résultats requièrent néanmoins des analyses supplémentaires. En effet, toutes les associations respectant les seuils minimaux de support et de confiance ne sont pas nécessairement pertinentes. L'évaluation finale des règles extraites est présentée dans la section suivante.

	Minimum	Maximum	Moyenne	Écart type
Support	0.1	2.9	0.4	0.6
Confiance	10.5	100	33.6	20.5
Nombre de vocables concernés	7	201	29	44.3

TABLEAU 4.7 – Propriétés des 42 règles extraites

4.2. Règles d'association

Évaluation finale

Il est généralement admis que les indices de support et de confiance ne suffisent pas à eux seuls à évaluer l'intérêt de règles. Ces deux mesures présentent en effet deux désavantages (Barra et al., 2021, p. 224). D'une part, elles tendent à produire un nombre important de règles, ce qui peut compliquer une éventuelle analyse qualitative. D'autre part, elles ne garantissent pas que pour une règle $\{A\} \Rightarrow \{B\}$, la probabilité d'apparition du conséquent $\{B\}$ ne soit pas supérieure ou égale à la probabilité du conséquent $\{B\}$ étant donné l'antécédent $\{A\}$ (Tufféry, 2012, p. 322). Pour illustrer brièvement ce second point, nous pouvons examiner la règle présentée en (118), qui figure parmi les 42 règles détectées à l'étape précédente.

(118) $\{\text{Ist-AGT}\} \Rightarrow \{\text{Anm-AGT}\}$
Support de 0.1 % et confiance de 24.4 %

S'appliquant à environ 0.1 % des données, cette règle indique que 24 % des noms qui instantient le type Institution-AGENT réalisent également le type Entité animée-AGENT. Le résultat obtenu semble *a priori* instructif, un certain nombre de noms déverbaux du français réalisant l'alternance en question : *distributeur* 'personne qui distribue' / 'société qui distribue', *éditeur* 'personne qui édite' / 'maison d'édition', *promoteur* 'personne qui promeut' / 'société qui promeut', etc. Il convient toutefois de considérer cette règle à la lumière de la fréquence élevée du conséquent Entité animée-AGENT. Instancié par 1'916 des 6'841 vocables considérés dans ce chapitre, ce type sémantique présente une probabilité d'apparition (28 %) qui dépasse l'indice de confiance de la règle en (118). Il s'ensuit que la règle $\{\text{Institution-AGENT}\} \Rightarrow \{\text{Entité animée-AGENT}\}$ a peu d'intérêt si l'on souhaite prédire le type Entité animée-AGENT.

Un nombre élevé de mesures d'intérêt complémentaires ont été proposées au fil du temps pour pallier ces différents problèmes. Geng et Hamilton (2006, p. 9) en répertorient pas moins de 38 dans leur vue d'ensemble consacrée à ce sujet, alors que plus de 50 indices sont implémentés dans le package *arules* dédié à l'extraction et l'analyse de règles d'association dans R (Hahsler, 2015). La littérature s'accorde sur le fait que le choix d'une mesure plutôt qu'une autre dépend à la fois des objectifs de l'étude et de la manière dont on définit l'intérêt des règles. Nous faisons ici le choix de recourir à une mesure additionnelle appelée « lift » (Brin et al., 1997) en complément des indices de support et de confiance. Cet indice, fréquemment employé, sert à comparer la confiance que nous pouvons avoir dans l'utilisation d'une règle $\{A\} \Rightarrow \{B\}$ à la probabilité d'obtenir son conséquent $\{B\}$ au hasard sans avoir recours à une règle d'association (Larose & Larose, 2019, p. 201) :

$$\begin{aligned} \text{lift}(A \Rightarrow B) &= \frac{P(B|A)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \\ &= \frac{\text{Confiance de la règle}}{\text{Probabilité d'observer le conséquent seul}} \end{aligned}$$

Le lift s'interprète de la manière suivante (Cios et al., 2007, p. 301) : s'il est supérieur à 1, alors $\{A\}$ et $\{B\}$ sont positivement corrélés, au sens où $\{A\}$ et $\{B\}$ ont tendance à apparaître ensemble plus souvent que ne le prévoirait le hasard; s'il est inférieur à 1, alors l'antécédent $\{A\}$ est négativement corrélé au conséquent $\{B\}$, la présence de l'un décourageant la présence de l'autre; enfin, s'il est égal à 1, alors $\{A\}$ et $\{B\}$ sont indépendants. S'élevant à 0.9, le lift de la règle en (118) indique ainsi qu'il est peu probable que le type sémantique Entité animée-AGENT soit réalisé conjointement avec le type Institution-AGENT. À l'inverse, celui de la règle en (115) est égal à 2.2, ce qui signifie que la combinaison des catégories {Artéfact-RÉSULTAT, Artéfact-INSTRUMENT, Événement-TRANSPOSITION} se produit environ 2.2 fois plus souvent que ce à quoi nous pourrions nous attendre si les deux sous-ensembles {Artéfact-RÉSULTAT, Artéfact-INSTRUMENT} et {Événement-TRANSPOSITION} étaient indépendants (Chapman & McDonnell Feit, 2019, p. 343).

Parmi les 42 règles de départ, 8 ont été supprimées en raison d'un lift inférieur ou égal à 1, leurs antécédent et conséquent n'étant pas positivement corrélés. Deux autres règles ont également été retirées de l'extraction initiale, car elles étaient redondantes en comparaison d'autres associations. En principe, une règle X peut être considérée comme redondante par rapport à une règle Y si (i) X comporte le même conséquent que Y , (ii) X inclut au moins un item de plus dans son antécédent en comparaison de Y , et si (iii) X présente une confiance inférieure ou égale à Y (Hahsler et al., 2005). Par exemple, la règle {Artéfact-RÉSULTAT, Domaine-TRANSPOSITION} \Rightarrow {Événement-TRANSPOSITION} a été retirée de la liste puisqu'il existe la règle plus générale {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION} présentant une confiance supérieure (58 % *vs* 65 %). En fin de compte, 32 règles ont été conservées.

4.2.2 Résultats et discussion

Les 32 règles finalement sélectionnées sont reproduites en intégralité dans le tableau 4.8⁵⁰. Représentant 21 types sémantiques différents (19 % des 113 types combinés d'origine), elles comportent en moyenne un support de 0.5 % ($ET = 0.7$), une confiance de 36.6 % ($ET = 21.3$) et un lift de 6.1 ($ET = 9.7$)⁵¹. Nous revenons dans ce qui suit sur certaines de leurs propriétés.

Probabilité d'orientation sémantique

Un premier résultat notable concerne le degré de réversibilité des associations observées. Sur l'ensemble des règles étudiées, seules 3 règles bipartites (n°1 et 2, n°9 et 15, n°12 et 17) et 2 règles tripartites (n°5, 6 et 19, n°8, 11 et 16) sont réciproques. Nous définissons la réciprocité comme suit. Pour une règle bipartite $\{A\} \Rightarrow \{B\}$, la réciproque attendue est $\{B\} \Rightarrow \{A\}$. S'agissant

50. Une version interactive de ce tableau est également disponible à l'adresse https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2. Elle permet notamment de trier les règles en fonction de chaque indice et de sélectionner celles contenant des types sémantiques précis.

51. Selon un test τ de Kendall, les indices de support, de confiance et de lift ne sont pas corrélés ($p > 0.05$).

4.2. Règles d'association

des règles tripartites, la réciproque de $\{A, B\} \Rightarrow \{C\}$ devrait logiquement être $\{C\} \Rightarrow \{A, B\}$. Cependant, étant donné que les conséquents des règles ne peuvent contenir qu'un seul type sémantique, nous considérons une forme particulière de « réciprocity tripartite », où la réciproque de $\{A, B\} \Rightarrow \{C\}$ est représentée par les règles $\{A, C\} \Rightarrow \{B\}$ et $\{B, C\} \Rightarrow \{A\}$.

Il y a lieu de réfléchir à la raison pour laquelle certaines règles ne sont pas accompagnées de leur(s) réciproque(s) dans la sélection finale. D'un point de vue technique, cette situation s'explique par le fait que la confiance de la réciproque peut être inférieure au seuil de 10 % spécifié lors de l'extraction, même si l'antécédent et le conséquent de la règle sont corrélés positivement selon l'indice de lift. Par exemple, la règle bipartite {Entité animée-Collectif-AGENT} \Rightarrow {Événement-TRANSPOSITION} (*relève* 'personnes qui remplacent' / 'action de remplacer') figure dans la liste finale avec une confiance de 70.4 %, alors que sa réciproque a été exclue automatiquement lors de l'extraction, son niveau de confiance n'atteignant que 0.9 %. On peut également s'interroger sur la manière d'interpréter ces différences de réciprocity au plan théorique. Rappelons que les règles que nous avons extraites constituent des associations positives entre types sémantiques qui ne sont pas le fruit du hasard. Ces associations non aléatoires suggèrent que la cooccurrence de types sémantiques est potentiellement motivée, une motivation qui pourrait à son tour être favorisée par des relations sémantiques de type métonymique ou métaphorique entre types. À supposer que ces relations sémantiques existent effectivement, nous pouvons faire l'hypothèse que le degré de réversibilité des règles reflète la probabilité d'orientation de l'extension sémantique, mettant potentiellement en lumière l'existence de schémas de transfert conceptuel plus ou moins symétriques en synchronie.

Si cette analyse se vérifie, les associations sémantiques seraient à orienter dans le sens contraire de celui des règles, l'existence des antécédents étant conditionnée à celle des conséquents. Les indices de confiance associés au couple de types {Entité animée-Collectif-AGENT, Événement-TRANSPOSITION}, par exemple, indiquent que le premier est très souvent accompagné du second (confiance de 70.4 %), mais que l'inverse est beaucoup moins concluant (confiance de 0.9 %). Il paraît ainsi plus probable, s'agissant des noms *vol*, *remonte* et *relève*⁵², que les sens Entité animée-Collectif-AGENT constituent des extensions métonymiques des sens Événement-TRANSPOSITION que l'inverse :

- (119) a. *vol*₁ 'action de se déplacer' \rightarrow *vol*₂ 'groupe d'animaux ou d'insectes qui se déplacent'
b. *remonte*₁ 'action de remonter une rivière' \rightarrow *remonte*₂ 'poissons qui remontent une rivière'
c. *relève*₁ 'action de remplacer' \rightarrow *relève*₂ 'personnes qui remplacent'

52. Le fait que ces trois exemples constituent des conversions plutôt que des formes ouvertement suffixées est intrigant. Nous aborderons dans le chapitre 5 les restrictions formelles pesant sur les schémas d'association sémantique dans le lexique déverbal.

ID	Antécédent	Conséquent	Structure	ID réciproque(s)	Support		Confiance		Lift	
					Valeur (%)	Rang	Valeur (%)	Rang	Valeur	Rang
1	{Art-INS}	⇒ {Anm-AGT}	Bipartite	2	2.9	1	31.1	19	1.1	26
2	{Anm-AGT}	⇒ {Art-INS}	Bipartite	1	2.9	1	10.5	30	1.1	26
3	{Art-LOC}	⇒ {Art-INS}	Bipartite	–	0.2	14	19.4	23	2.1	14
4	{Art-RES}	⇒ {Art-INS}	Bipartite	–	0.3	8	11.4	29	1.2	23
5	{Art-RES, Evt-TSP}	⇒ {Art-INS}	Tripartite	6 et 19	0.2	13	12.5	28	1.3	18
6	{Art-INS, Evt-TSP}	⇒ {Art-RES}	Tripartite	5 et 19	0.2	13	29.4	20	10.9	5
7	{Art*Ist-AGT}	⇒ {Art-RES}	Bipartite	–	0.1	17	15.0	25	5.6	8
8	{Art*Ist-AGT, Dom-TSP}	⇒ {Art-RES}	Tripartite	11 et 16	0.1	17	32.1	18	12.0	4
9	{Dom-TSP}	⇒ {Art-RES}	Bipartite	15	0.4	6	12.5	28	4.6	10
10	{Dom-TSP, Evt-TSP}	⇒ {Art-RES}	Tripartite	–	0.2	14	17.9	24	6.7	6
11	{Art-RES, Dom-TSP}	⇒ {Art*Ist-AGT}	Tripartite	8 et 16	0.1	17	37.5	13	42.8	1
12	{Dom-TSP}	⇒ {Art*Ist-AGT}	Bipartite	17	0.4	5	14.6	26	16.6	3
13	{Art*Cog-RES}	⇒ {Cog*Evt-TSP}	Bipartite	–	0.1	18	26.7	21	6.2	7
14	{Ppt-CAU}	⇒ {Cog*Evt-TSP}	Bipartite	–	0.4	5	20.6	22	4.8	9
15	{Art-RES}	⇒ {Dom-TSP}	Bipartite	9	0.4	6	13.0	27	4.6	10
16	{Art-RES, Art*Ist-AGT}	⇒ {Dom-TSP}	Tripartite	8 et 11	0.1	17	100	1	35.6	2
17	{Art*Ist-AGT}	⇒ {Dom-TSP}	Bipartite	12	0.4	5	46.7	8	16.6	3
18	{Anm-Coll-AGT}	⇒ {Evt-TSP}	Bipartite	–	0.3	10	70.4	3	2.2	12
19	{Art-INS, Art-RES}	⇒ {Evt-TSP}	Tripartite	5 et 6	0.2	13	71.4	2	2.2	11
20	{Art-RES}	⇒ {Evt-TSP}	Bipartite	–	1.8	2	65.2	5	2.0	15
21	{Art-THM}	⇒ {Evt-TSP}	Bipartite	–	0.3	11	34.0	16	1.1	27
22	{Art-Coll-INS}	⇒ {Evt-TSP}	Bipartite	–	0.1	18	66.7	4	2.1	13
23	{Art-Coll-RES}	⇒ {Evt-TSP}	Bipartite	–	0.1	19	41.2	11	1.3	21

ID	Antécédent	Conséquent	Structure	ID réciproque(s)	Support		Confiance		Lift	
					Valeur (%)	Rang	Valeur (%)	Rang	Valeur	Rang
24	{Art*Cog-RES}	⇒ {Evt-TSP}	Bipartite	–	0.2	15	36.7	14	1.1	24
25	{Cog-RES}	⇒ {Evt-TSP}	Bipartite	–	0.1	17	32.1	18	1.0	29
26	{Dom-TSP}	⇒ {Evt-TSP}	Bipartite	–	1.1	3	40.6	12	1.3	22
27	{Evt-Coll-TSP}	⇒ {Evt-TSP}	Bipartite	–	0.3	7	32.4	17	1.0	28
28	{Ist-AGT}	⇒ {Evt-TSP}	Bipartite	–	0.2	12	41.5	9	1.3	19
29	{N/A·Coll-RES}	⇒ {Evt-TSP}	Bipartite	–	0.1	19	53.8	7	1.7	17
30	{Nat-RES}	⇒ {Evt-TSP}	Bipartite	–	0.9	4	41.2	10	1.3	20
31	{Ppt-MAN}	⇒ {Evt-TSP}	Bipartite	–	0.1	16	35.7	15	1.1	25
32	{Ppt-RES}	⇒ {Evt-TSP}	Bipartite	–	0.3	9	57.1	6	1.8	16

TABLEAU 4.8 – Règles d’association finales, triées par ordre alphabétique en fonction des conséquents, puis des antécédents. Les cooccurrences ayant obtenu les scores les plus élevés pour les indices de support, de confiance et de lift sont colorées en vert

Il reste que la cooccurrence de types N_1 et N_2 pourrait aussi s'expliquer par la grande productivité de patrons $V \rightarrow N_1$ et $V \rightarrow N_2$, sans que N_2 ne constitue nécessairement une extension métaphorique ou métonymique de N_1 . On peut imaginer, par exemple, que les noms formés avec *-eur* qui sont ambigus entre une lecture agentive et une lecture instrumentale, comme *accordeur* 'personne qui accorde' / 'instrument servant à accorder', viennent au jour en raison de la forte prévalence de ces types parmi les dérivés en *-eur*, plutôt que par le biais d'une extension sémantique de type agent \rightarrow instrument ou instrument \rightarrow agent. De même, une différence de réciprocité pourrait résulter d'une productivité inégale entre deux patrons. Le type Événement-TRANSPOSITION est de toute évidence beaucoup plus fréquent que le type Entité animée-Collectif-AGENT s'agissant de la règle n°18. Néanmoins, la présomption d'extension sémantique lorsqu'une règle d'association est observée reste forte, notamment en cas d'absence de réciprocité (la métonymie étant fondamentalement orientée).

Un élément de confirmation important pour l'hypothèse d'une extension sémantique serait qu'un type sémantique constituant l'antécédent d'une règle ne soit réalisé que par des noms ambigus instanciant aussi le conséquent (cf. Salvadori & Huyghe, 2022). Plus généralement, la « figurativité » pourrait aussi être envisagée comme une propriété scalaire, susceptible d'être approximée par un indice de réalisation autonome. Nous proposons de calculer cet indice en évaluant la probabilité que l'antécédent $\{A\}$ d'une règle $\{A\} \Rightarrow \{B\}$ soit instancié par des noms monosémiques, i.e. en divisant simplement le nombre de lexèmes correspondant à $\{A\}$ et réalisés en monosémie par le nombre total de lexèmes instanciant $\{A\}$. Dans ce contexte, une probabilité très faible, voire nulle, constituerait un indice du fait que l'antécédent $\{A\}$ est un type sémantique secondaire, c'est-à-dire dérivé sémantiquement. Dans le cas de la règle n°18, par exemple, le type sémantique Entité animée-Collectif-AGENT est instancié par 3 lexèmes en monosémie et 27 lexèmes au total. Il s'ensuit que l'indice de réalisation autonome de ce type sémantique est relativement bas ($3/27 = 0.11$), ce qui semble indiquer qu'il peut être obtenu par extension sémantique.

Si l'on reprend l'argument dans son ensemble, l'hypothèse figurative (i.e. impliquant des extensions sémantiques métonymiques ou métaphoriques) pourrait être vérifiée en fonction des éléments suivants – pour les règles bipartites du moins ⁵³ :

- (i) La première étape consisterait à vérifier, sur la base de l'indice de réalisation autonome, si l'antécédent $\{A\}$ d'une règle d'association a peu d'autonomie, c'est-à-dire s'il est conditionné à la présence d'un autre type sémantique. Si tel est le cas, on peut supposer qu'en cas de règle $\{A\} \Rightarrow \{B\}$, $\{A\}$ résulte d'une extension sémantique à partir de $\{B\}$ (plutôt que d'une stricte codérivation).
- (ii) La non-réciprocité de $\{A\} \Rightarrow \{B\}$ pourrait ensuite confirmer l'orientation de l'extension sémantique.
- (iii) Enfin, une confiance élevée pour $\{A\} \Rightarrow \{B\}$ pourrait indiquer que $\{B\}$ est le principal ou l'un des principaux motivateurs sémantiques de $\{A\}$.

53. Les règles tripartites étant particulières, nous les discutons dans la prochaine sous-section.

4.2. Règles d'association

Le tableau 4.9 présente le score de réalisation autonome obtenu par chaque type sémantique constituant l'antécédent ou le conséquent d'une règle. De l'examen comparatif de ce tableau avec les règles exposées dans le tableau 4.8 émergent plusieurs résultats marquants. Si l'on considère l'argument exposé ci-dessus, 4 règles (n°18, 20, 22 et 32) se distinguent premièrement par les caractéristiques suivantes : le type constituant leur antécédent obtient un score de réalisation autonome très faible (inférieur à 0.20); leur réciproque n'est pas présente dans la sélection finale; leur niveau de confiance est relativement élevé (de 57.1 à 70.4 %); enfin, leur conséquent correspond systématiquement au type Événement-TRANSPPOSITION. Tout porte ainsi à croire que le type présent en antécédent de la règle – Entité animée-Collectif-AGENT (120a), Artéfact-RÉSULTAT (120b), Artéfact-Collectif-INSTRUMENT (120c), Propriété-RÉSULTAT (120d) – dérive sémantiquement du type Événement-TRANSPPOSITION, *a priori* par métonymie. Cette hypothèse semble corroborer l'idée largement admise selon laquelle les sens événementiels constituent une source majeure d'extensions sémantiques dans le lexique déverbal, d'autant que le score de réalisation autonome du type Événement-TRANSPPOSITION est l'un des plus élevés (0.77).

- (120) a. *relève*₁ 'action de remplacer' → *relève*₂ 'personnes qui remplacent'
b. *moulage*₁ 'action de mouler' → *moulage*₂ 'objet moulé'
c. *boutonnage*₁ 'action de boutonner (un vêtement)' → *boutonnage*₂ 'ensemble de boutons'
d. *agencement*₁ 'action d'agencer' → *agencement*₂ 'manière dont quelque chose est agencé'

Il convient néanmoins de souligner que les sens potentiellement obtenus par figure sémantique dans les exemples ci-dessus (*relève*₂, *moulage*₂, *boutonnage*₂ et *agencement*₂) entretiennent malgré tout des liens sémantiques évidents avec les lexèmes-bases (respectivement *relever*, *mouler*, *boutonner* et *agencer*). De fait, il paraît difficile d'évacuer complètement l'implication des verbes apparentés dans la construction des sens nominaux *N*₂. Une solution potentielle à ce problème pourrait être de considérer que la construction morphologique et l'extension sémantique collaborent dans l'élaboration du sens *N*₂. Il serait alors envisageable de parler de contributions convergentes lorsque, simultanément, un procédé morphologique (*V*₁ → *N*₂) et une extension sémantique (*N*₁ → *N*₂) interviennent pour former un même lexème, voire de contributions enchâssées, lorsque l'on construit deux lexèmes liés sémantiquement dont les sens sont conditionnés par la forme (*V*₁ → *N*₁ + *N*₂). Nous schématisons ces deux hypothèses en 4.5 et 4.6.

Un deuxième cas notable émergeant de la comparaison des tableaux 4.8 et 4.9 concerne les règles réciproques n°1 et 2 mettant en jeu les types Artéfact-INSTRUMENT et Entité animée-AGENT (p. ex. *plumeuse* 'personne qui plume des volailles'/'machine servant à plumer des volailles', *verseur* 'personne qui verse'/'objet servant à verser', *bobineuse* 'personne qui bobine'/'machine servant à bobiner'). Les scores de réalisation autonome obtenus par les deux types sont parmi les plus élevés, atteignant respectivement 0.52 et 0.88, tandis que les niveaux

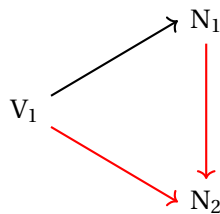


FIGURE 4.5 – Contributions convergentes. Une extension sémantique ($N_1 \rightarrow N_2$) et un procédé morphologique ($V_1 \rightarrow N_2$) sont appliqués simultanément pour former un même lexème

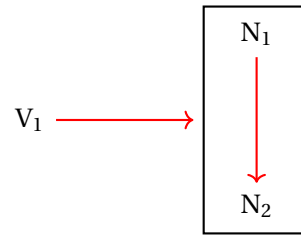


FIGURE 4.6 – Contributions enchâssées. La construction morphologique et l'extension sémantique collaborent pour former deux lexèmes liés sémantiquement dont les sens sont conditionnés par la forme ($V_1 \rightarrow N_1 + N_2$)

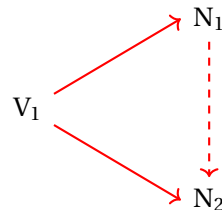


FIGURE 4.7 – Contributions cloisonnées. Deux procédés morphologiques sont appliqués de manière indépendante pour former deux lexèmes ($V_1 \rightarrow N_1$, $V_1 \rightarrow N_2$). Un lien sémantique entre les lexèmes ainsi formés peut ensuite se dessiner ($N_1 \rightarrow N_2$)

de confiance, qui sont de 31.1 % et de 10.5 %, restent inférieurs à la moyenne de l'ensemble des règles (36.6 %). En prenant en compte la réciprocité des deux associations, on en déduit qu'une codérivation stricte ($V_1 \rightarrow N_1$, $V_1 \rightarrow N_2$) semble plus plausible qu'une extension sémantique. Cela étant, la littérature tend à reconnaître un lien métonymique ou métaphorique entre les catégories agentive et instrumentale. Panther et Thornburg (2002, p. 309) soutiennent par exemple que le sens instrumental 'somnifère' du nom polysémique *sleeper* constitue une extension métonymique de son sens agentif principal 'dormeur' dans leur analyse du suffixe nominalisateur *-er* en anglais. Booij (2005, p. 221) privilégie quant à lui une approche métaphorique, arguant que la notion d'agent peut être transférée au domaine des choses matérielles inanimées, conçues comme des agents exécutant une tâche particulière. Dans les deux perspectives, l'orientation prédominante entre les types s'établit toujours de la catégorie agentive à la catégorie instrumentale, ce que tendent à confirmer les niveaux de confiance obtenus. La confiance de la règle Entité animée-AGENT \Rightarrow Artéfact-INSTRUMENT est en effet supérieure d'environ 20 % à celle de sa réciproque. Dès lors, comment arbitrer entre l'hypothèse de la codérivation stricte et celle de l'extension sémantique ? Une approche combinée apparaît encore une fois pertinente ici : l'alternance observée pourrait bel et bien résulter d'une codérivation stricte ($V_1 \rightarrow N_1$, $V_1 \rightarrow N_2$), comme le laissent penser les taux élevés de réalisation autonome, la réciprocité des règles et les niveaux de confiance relativement bas, puis s'accompagner de « l'élaboration » éventuelle d'un lien sémantique allant de

4.2. Règles d'association

N_1 à N_2 . La construction morphologique et l'extension sémantiques contribueraient ainsi de manière « cloisonnée » à la formation de dérivés ambigus. Nous illustrons cette hypothèse en 4.7.

Au-delà de ces premiers cas, il convient enfin de relever que pour un certain nombre de règles non réciproques, le score de réalisation autonome de l'antécédent et le niveau de confiance ne sont ni spécialement élevés, ni spécialement bas. Tel est le cas, par exemple, des règles n°23 et 30 :

- (121) {Art·Coll-RES} \Rightarrow {Evt-TSP}
Confiance de 41.2 %, taux de réalisation autonome de 0.41 pour Art·Coll-RES
empilement 'ensemble de choses empilées' / 'action d'empiler'
- (122) {Nat-RES} \Rightarrow {Evt-TSP}
Confiance de 41.2%, taux de réalisation autonome de 0.41 pour Nat-RES
brûlis 'essart' / 'action de défricher par le feu'

L'hypothèse peut être formulée que ces résultats reposent – au moins en partie – sur les préférences sémantiques des procédés morphologiques avec lesquels les dérivés sont formés. Dans le cas de la règle n°30 en (122), par exemple, un examen manuel des dérivés concernés indique que le suffixe *-is* semble bel et bien favoriser la coréalisation du type Objet naturel-RÉSULTAT avec la catégorie Événement-TRANSPOSITION (p. ex. *brûlis* 'action de défricher par le feu' / 'essart', *dégueulis* 'action de dégueuler' / 'matière vomie', *semis* 'action de semis' / 'terre ensemencée', etc.). Le suffixe *-ure*, en revanche, forme de nombreux noms instanciant le type Objet naturel-RÉSULTAT en monosémie (p. ex. *chiure* 'excrément', *épluchure* 'déchet', *râpure* 'matière râpée', etc.). Si seuls ces derniers étaient considérés dans l'analyse, l'indice de réalisation autonome du type Objet naturel-RÉSULTAT devrait théoriquement être élevé, ce qui n'est pas le cas ici. L'implication des procédés morphologiques dans la formation de noms déverbaux ambigus sera examinée plus en détail dans le chapitre 5.

Structure bipartite vs tripartite

Les études sur la colexification, la polysémie régulière et la sémantique des nominalisations se concentrent majoritairement sur des schémas d'association bipartites. Xu et ses collègues (2020, p. 3) considèrent par exemple qu'il y a colexification lorsque deux concepts distincts partagent le même mot, Lombard et ses collègues (2024) examinent la régularité de patrons de polysémie à deux sens uniquement (p. ex. plante \rightarrow partie du corps, nourriture \rightarrow quantité), et Ferret et Villoing (2015) explorent les conditions de coréalisation de sens instrumentaux par les noms déverbaux événementiels en *-age* du français (p. ex. *allumage*, *bandage*, *chauffage*). Nos résultats indiquent pourtant que les associations intéressantes instanciées par les noms déverbaux peuvent mobiliser deux ($n = 25$), mais aussi trois ($n = 7$) sens distincts. Les observations formulées à ce sujet dans l'aperçu fréquentiel présenté en première partie de ce chapitre (section 4.1.2) se voient ainsi consolidées. Il en va naturelle-

Type sémantique	ID antécédent	ID conséquent	Nombre de lexèmes en monosémie	Nombre total de lexèmes	Réalisation autonome
Anm-AGT	2	1	1'684	1'916	0.88
Anm-Coll-AGT	18	–	3	27	0.11
Art-INS	1, 6 et 19	2 à 5	337	646	0.52
Art-LOC	3	–	31	72	0.43
Art-RES	4, 5, 11, 15, 16 et 20	6 à 10	35	184	0.19
Art-THM	21	–	26	53	0.49
Art-Coll-INS	22	–	2	12	0.17
Art-Coll-RES	23	–	7	17	0.41
Art*Cog-RES	13 et 24	–	4	30	0.13
Art*Ist-AGT	7, 8, 16 et 17	11 et 12	25	60	0.42
Cog-RES	25	–	9	28	0.32
Cog*Evt-TSP	–	13 et 14	222	292	0.76
Dom-TSP	8, 9, 10, 11, 12 et 26	15, 16 et 17	45	192	0.23
Evt-TSP	5, 6 et 10	18 à 32	1'677	2'189	0.77
Evt-Coll-TSP	27	–	24	68	0.35
Ist-AGT	28	–	6	41	0.15
N/A-Coll-RES	29	–	5	13	0.38
Nat-RES	30	–	60	148	0.41
Ppt-CAU	14	–	38	136	0.28
Ppt-MAN	31	–	13	28	0.46
Ppt-RES	32	–	4	35	0.11

TABLEAU 4.9 – Taux de réalisation autonome des types sémantiques constituant l'antécédent ou le conséquent de règles d'association. Le taux est calculé en divisant le nombre de lexèmes réalisant un type sémantique donné en monosémie par le nombre total de lexèmes instanciant le même type. Les taux les plus élevés sont colorés en vert clair. L'ID des règles contenant chaque type en antécédent ou en conséquent est également fourni

4.2. Règles d'association

ment de même pour les réflexions de Ježek (2008), Huyghe (2014) et Lieber (2018), bien que la forme conjointe *vs* disjointe (cf. chapitre 1) de certains des patrons d'association mentionnés dans ces travaux puisse parfois être questionnée.

L'existence de schémas à trois sens est intrigante, d'autant que plusieurs règles à antécédent multiple que nous avons identifiées présentent des similitudes. La quasi-totalité des associations tripartites semble en effet se répartir en deux groupes de trois règles : les n°5, 6 et 19 (voir tableau 4.8), composées des types Événement-TRANSPOSITION, Artéfact-INSTRUMENT et Artéfact-RÉSULTAT (123); et les n°8, 11 et 16, qui combinent les types Domaine-TRANSPOSITION, Artéfact-RÉSULTAT et Artéfact*Institution-AGENT (124).

(123) Événement-TRANSPOSITION/Artéfact-INSTRUMENT/Artéfact-RÉSULTAT
teinture 'action de teindre'/'substance servant à teindre'/'coloration obtenue'

(124) Domaine-TRANSPOSITION/Artéfact-RÉSULTAT/Artéfact*Institution-AGENT
menuiserie 'art de menuiser'/'ouvrage menuisé en bois'/'atelier qui menuise'

Un premier point commun aux deux ensembles réside dans le degré élevé de réciprocité (telle que définie dans la sous-section précédente pour les règles tripartites) des différentes règles. Toutes les configurations possibles en termes d'antécédent et de conséquent se définissent en effet par une confiance supérieure à 10 % et un lift supérieur à 1 : $\{A, B\} \Rightarrow \{C\}$, $\{B, C\} \Rightarrow \{A\}$ et $\{A, C\} \Rightarrow \{B\}$. Un autre résultat notable tient au fait que des règles bipartites sous-tendent certaines des règles tripartites au sein de chaque ensemble, comme $\{\text{Artéfact-RÉSULTAT}\} \Rightarrow \{\text{Événement-TRANSPOSITION}\}$, ou encore $\{\text{Artéfact*Institution-AGENT}\} \Rightarrow \{\text{Domaine-TRANSPOSITION}\}$. Plus générales et plus fréquentes, elles se caractérisent toutefois par une confiance et un lift moindres. Enfin, la règle présentant le score de confiance le plus élevé au sein de chaque groupe comporte toujours un type événementiel (i.e. Événement ou Domaine) en conséquent. Le niveau de confiance de la règle $\{\text{Artéfact-INSTRUMENT, Artéfact-RÉSULTAT}\} \Rightarrow \{\text{Événement-TRANSPOSITION}\}$ atteint en effet 71.4 % (contre 12.5 et 29.4 % pour les règles réciproques n°5 et 6), et même 100 % pour la règle $\{\text{Artéfact-RÉSULTAT, Artéfact*Institution-AGENT}\} \Rightarrow \{\text{Domaine-TRANSPOSITION}\}$ (contre 32.1 % et 37.5 % pour les règles réciproques n°8 et 11).

Ces observations soulèvent plusieurs hypothèses. La réciprocité « tripartite » entre les différentes règles de chaque groupe suggère d'abord une forte dépendance conceptuelle au sein des triplets de types sémantiques concernés. Il s'ensuit que certaines associations bipartites signalées comme plus ou moins fréquentes dans la littérature (p. ex. Événement-TRANSPOSITION/Artéfact-INSTRUMENT) sont peut-être à replacer dans un espace conceptuel plus large et potentiellement parsemé de microparadigmes sémantiques, comme le suggère par ailleurs la présence de règles bipartites moins « intéressantes » que les règles tripartites apparentées. Pour autant, tous les types sémantiques concernés par ces schémas à trois têtes ne se trouvent pas nécessairement sur un pied d'égalité. D'une part, la présence du type Artéfact-RÉSULTAT dans chacun des deux groupes suggère que celui-ci tient un rôle

central dans la construction d'associations sémantiques tripartites. D'autre part, la présence privilégiée des types événementiels Événement et Domaine en conséquent des règles laisse supposer que ceux-ci sont peut-être à l'origine de l'émergence des schémas tripartites étudiés. Ce résultat pourrait ainsi confirmer le fort potentiel métonymique de la description événementielle, qui repose sur des relations thématiques entre actions et participants (cf. Salvadori & Huyghe, 2023).

Types sémantiques associés

Nous l'avons vu dans le chapitre 1, les noms déverbaux se distinguent par leur tendance à favoriser certaines associations sémantiques. Nos résultats confirment effectivement l'importance de plusieurs alternances discutées dans la littérature. Le patron général « événement/résultat » (Grimshaw, 1990; Jacquy, 2006; Bisetto & Melloni, 2007; Lieber, 2018; *inter alia*), notamment, est particulièrement bien représenté dans nos données, avec pas moins de huit règles bipartites (n°13, 20, 23, 24, 25, 29, 30 et 32). Il convient toutefois de souligner deux points enrichissant les discussions sur ce patron. D'une part, la règle n°13 révèle que le type événementiel peut être complexe :

- (125) Événement*Cognitif-TRANSPOSITION/Artéfact*Cognitif-RÉSULTAT
dictée 'action de dicter un texte'/'texte transcrit par écrit'

D'autre part, il apparaît que les « résultats » peuvent être de différents types ontologiques. La règle n°13 présentée ci-dessus et la n°24 ci-dessous impliquent par exemple le même résultat à ontologie complexe avec des facettes concrète et cognitive :

- (126) Événement-TRANSPOSITION/Artéfact*Cognitif-RÉSULTAT
relevé 'action de prendre en note'/'document écrit'

Trois autres règles (n°20, 23 et 30) comprennent par ailleurs des objets concrets, construits (127)-(128) ou naturels (129) :

- (127) Événement-TRANSPOSITION/Artéfact-RÉSULTAT
moulage 'action de mouler'/'objet moulé'
- (128) Événement-TRANSPOSITION/Artéfact-Collectif-RÉSULTAT
lattage 'action de lacter'/'ouvrage de lattes'
- (129) Événement-TRANSPOSITION/Objet naturel-RÉSULTAT
souillure 'action de souiller'/'tache'

4.2. Règles d'association

Enfin, trois règles moins attendues (n°25, 29 et 32) font intervenir des résultats se rapportant respectivement à des entités abstraites (130), à des groupes d'entités sous-déterminés⁵⁴ (131) et à des propriétés (132) :

- (130) Événement-TRANSPOSITION/Objet cognitif-RÉSULTAT
préprogrammation 'action de préprogrammer un appareil' / 'programme qui a été prédéfini'
- (131) Événement-TRANSPOSITION/N/A-Collectif-RÉSULTAT
assemblage 'action d'assembler' / 'ensemble de choses assemblées'
- (132) Événement-TRANSPOSITION/Propriété-RÉSULTAT
agencement 'action d'agencer' / 'manière dont quelque chose est agencé'

Relevons par ailleurs que certains types résultatifs au sein de ces ensembles affichent des scores de réalisation autonome particulièrement bas en tant qu'antécédent, notamment Propriété-RÉSULTAT (0.11), Artéfact*Cognitif-RÉSULTAT (0.13) et Artéfact-RÉSULTAT (0.19) (voir tableau 4.9). De plus, aucune de ces règles n'est accompagnée de sa réciproque dans la sélection finale, et les niveaux de confiance excèdent systématiquement 25 %. Ces observations semblent souligner l'importance des mécanismes d'extension sémantique de type métonymique dans la formation d'alternances « événement/résultat », où le sens résultatif dériverait du sens événementiel. Toutefois, comme nous l'avons évoqué précédemment, l'implication de la construction morphologique ne saurait être négligée (cf. schémas 4.5 et 4.6).

D'autres patrons d'ambiguïté déjà abordés dans la littérature peuvent également être relevés. Outre l'alternance agent/instrument mentionnée ci-dessus, nous retrouvons le schéma instrument/lieu (Luschützky & Rainer, 2013) sous la forme Artéfact-INSTRUMENT/Artéfact-LIEU (133), le patron événement/agent et ses voisins événement/humain et événement/groupe humain (Scott, 2010; Ježek, 2008) au travers des formes Événement-TRANSPOSITION/Entité animée-Collectif-AGENT (119) et Événement-TRANSPOSITION/Institution-AGENT (134), ainsi que l'association événement/moyen et sa voisine événement/instrument (Fradin, 2012; Ferret & Villoing, 2015). Ce dernier cas reste toutefois particulier. En effet, l'alternance ne se vérifie que si l'instrument est collectif d'un point de vue ontologique (135), ou si le sens instrumental s'accompagne du type Artéfact-RÉSULTAT dans un schéma tripartite (123).

- (133) Artéfact-INSTRUMENT/Artéfact-LIEU
équarri 'objet servant à équarrir' / 'lieu où l'on équarrir'

54. Le type ontologique N/A-Collectif n'est pas sans rappeler la catégorie sémantique des « métacollectifs » décrite par Lammert (2008), i.e. des noms collectifs décrivant des ensembles d'éléments non spécifiés (p. ex. *ensemble*, *groupe*, *collection* versus *famille*, *forêt*, *troupeau*).

- (134) Événement-TRANSPOSITION/Institution-AGENT
gouvernement 'action de gouverner'/'pouvoir politique qui gouverne'
- (135) Événement-TRANSPOSITION/Artéfact-Collectif-INSTRUMENT
dînette 'petit repas'/'service de table miniature'

4.3 Réseau général

L'analyse d'association, bien qu'efficace pour identifier des règles individuelles, présente une limitation majeure : comme le relève Tufféry (2012, p. 333), elle ne permet pas de dégager « une vision globale du phénomène, avec mise en évidence de facteurs prépondérants et d'interactions entre les facteurs ». Pour pallier cette difficulté, nous proposons dans cette section une modélisation des règles d'association prenant la forme d'un réseau dont les sommets représentent les types sémantiques, et les liens, les implications observées. Nous inscrivons ce faisant cette étude dans le champ général de « l'analyse de réseau » (cf. Wasserman & Faust, 1994; Newman, 2018), un cadre d'analyse polyvalent et transdisciplinaire devant beaucoup à la sociologie, la théorie des graphes et la physique statistique (cf. Zweig, 2016). L'avantage d'une telle approche réside essentiellement dans son caractère systémique. En replaçant les schémas associatifs individuels au sein d'un espace sémantique plus large, nous pouvons en effet espérer atteindre une compréhension plus holistique du rapport forme-sens dans le lexique déverbal.

Cette démarche en deux temps, combinant analyses d'association et de réseau, a déjà été adoptée dans différents domaines par le passé (voir p. ex. Belyi et al., 2016 en pharmacosurveillance et Hernández et al., 2019 en médecine). À notre connaissance, elle reste toutefois inédite en linguistique, bien que l'analyse de réseau soit elle-même bien implantée. Des exemples notables en morphologie incluent les études de Sims (2020) sur la structure organisationnelle de systèmes flexionnels et de Dattner et ses collègues (2023) sur le développement morphologique de la catégorie verbale en acquisition du langage. Dans une autre veine, Schulte (2015a, 2015b) propose des cartes sémantiques des suffixes *-age* et *-ery* en anglais inspirées des *semantic maps* de Haspelmath (2003), avec toutefois une dimension plus quantitative. Les études sur la colexification recourent elles aussi à des outils d'analyse de réseau : la base de données en ligne CLICS³ (Rzymiski et al., 2020), par exemple, permet d'examiner les colexifications interlinguistiques à travers des réseaux sémantiques globaux et régionaux, tandis que Jackson et ses collègues (2019) et Fukuya et ses collègues (2023) ont étudié des réseaux d'émotions.

Dans cette section, nous tenterons pour notre part de proposer des réponses aux questions suivantes :

- (i) Quels sont les types sémantiques les plus importants au sein du réseau sémantique formé par les associations précédemment identifiées? De quelle manière leur importance se profile-t-elle? Et à quel point le réseau dépend-il de ces types importants?

4.3. Réseau général

- (ii) En faisant abstraction des types eux-mêmes, l'ambiguïté des nominalisations se caractérise-t-elle par des motifs d'association récurrents d'un point de vue structurel? Par exemple, si un type *A* amène à un type *B*, *B* amène-t-il aussi systématiquement à *A*? De même, si un type *A* est en relation avec un type *B* lui-même lié à un type *C*, le type *A* a-t-il également tendance à s'associer à *C*?

Ces deux volets de questions seront tour à tour abordés dans les parties 4.3.2 et 4.3.3. La sous-section 4.3.1 présente en préambule la constitution du réseau.

4.3.1 Constitution du réseau

Nous avons jusqu'à présent extrait 42 règles d'association de notre base de données, puis nettoyé cette première liste en supprimant les règles redondantes et les ensembles de types qui n'étaient pas corrélés positivement. Les 32 règles finales, analysées individuellement dans la section précédente, comprennent au total 21 types sémantiques différents et 39 implications si les associations tripartites sont comptées à double. Deux implications peuvent par exemple être relevées pour la règle {Artéfact-RÉSULTAT, Artéfact-INSTRUMENT} \Rightarrow {Événement-TRANSPOSITION} : celle reliant les types Artéfact-RÉSULTAT et Événement-TRANSPOSITION ; et celle rassemblant les types Artéfact-INSTRUMENT et Événement-TRANSPOSITION.

La figure 4.8 présente le réseau⁵⁵ constitué à partir des données définitives, composé donc de 21 sommets (représentant ici les types sémantiques) et de 39 liens orientés (reflétant les différentes implications entre types)⁵⁶. Ces derniers sont à la fois proportionnés selon le degré de confiance qui caractérise la règle dans laquelle ils s'inscrivent et colorés de manière à refléter leur implication dans des schémas bipartites *vs* tripartites. Plus précisément, les flèches bleues indiquent des relations bipartites (pour deux nœuds *A* et *B*, une relation $\{A\} \Rightarrow \{B\}$), et les flèches rouges, des relations faisant partie de schémas tripartites (i.e. pour deux nœuds *A* et *B*, une relation $\{A, X\} \Rightarrow \{B\}$). La taille des nœuds au sein du réseau traduit quant à elle la fréquence plus ou moins élevée des types dans la liste de règles. Soulignons au passage que la distance visuelle entre types n'est pas représentative d'un quelconque indice d'intérêt (support, confiance ou lift). En effet, la disposition générale du réseau a été déterminée automatiquement à l'aide de l'algorithme Davidson-Harel (Davidson & Harel, 1996) de façon à améliorer la visibilité du graphe.

Plusieurs aspects fondamentaux du réseau peuvent déjà être relevés avant un examen plus approfondi dans les sections suivantes. Il apparaît en premier lieu que le système étudié ne se compose pas de règles isolées, mais forme une seule composante connectée. La densité du réseau, qui correspond au rapport entre le nombre de liens présents ($n = 30$ en réduisant

55. Différentes infrastructures ont été développées pour construire, visualiser et analyser des réseaux dans R. Nous avons utilisé les packages *igraph* (Csardi & Nepusz, 2006), *tidygraph* (Pedersen, 2024b), *manynet* (Hollway, 2024), *miagraph* (Hollway, 2021) et *sna* (Butts, 2023) pour les manipulations générales, *ggraph* (Pedersen, 2024a) pour les visualisations, et *intergraph* (Bojanowski, 2023) pour le passage d'un environnement à un autre.

56. Nous utiliserons de manière interchangeable les termes « nœud », « sommet » et « acteur » pour parler des points composant un graphe, et les termes « lien » et « relation » pour décrire les lignes entre ces points.

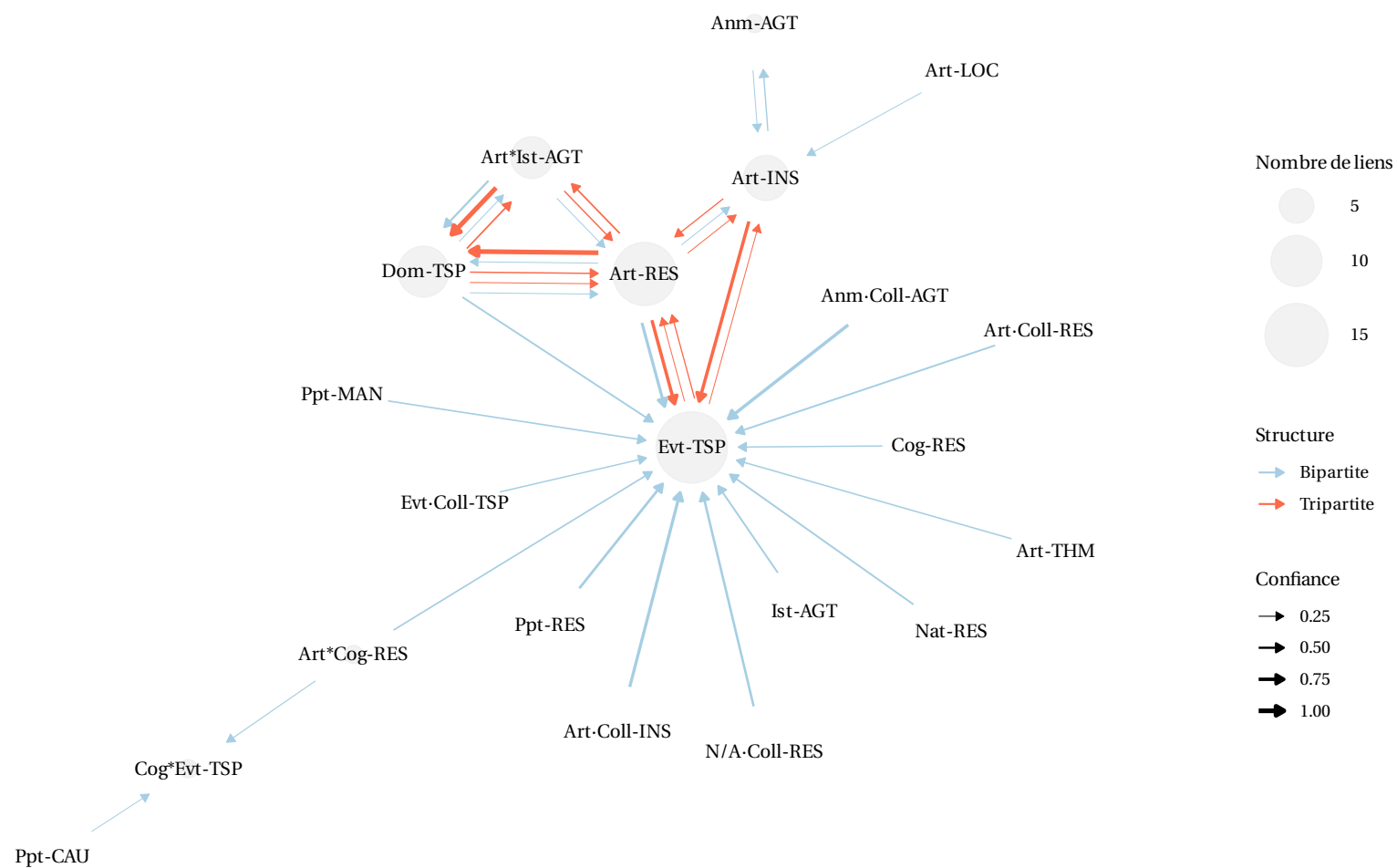


FIGURE 4.8 – Réseau général d'associations de types sémantiques pour les noms déverbaux ambigus

4.3. Réseau général

à une relation unique les liens multiples) et le plus grand nombre possible de liens dans le graphe ($n = 420$), reste toutefois relativement faible à 0.071. Concernant les types sémantiques observés, les catégories « marginales » (p. ex. Institution-AGENT) semblent intuitivement plus nombreuses que les types « importants » comme Événement-TRANSPOSITION et Artéfact-INSTRUMENT, ces derniers semblant par ailleurs s'attirer mutuellement. Structurellement, le réseau révèle un motif en étoile, extrêmement rare selon Zweig (2016, p. 244). Toutes ces impressions demandent toutefois à être précisées.

4.3.2 Importance des types

Cette section vise à évaluer l'importance respective des différents types sémantiques composant le réseau d'associations (analyse de centralité), puis à quantifier le degré de dépendance du réseau vis-à-vis de ses nœuds les plus importants (analyse de centralisation). Pour cela, nous fusionnons au préalable les liens multiples du réseau présenté dans la figure 4.8, certaines des mesures employées amenant en effet à des résultats erronés en leur présence (Csardi & Nepusz, 2006). Par exemple, les 5 relations observées entre les types Domaine-TRANSPOSITION et Artéfact-RÉSULTAT peuvent être réduites à un lien bidirectionnel unique. Le réseau simplifié compte au total 16 liens unidirectionnels (\rightarrow) et 7 liens bidirectionnels tous composés de deux relations orientées (\rightarrow et \leftarrow).

Centralité

L'importance des sommets composant un graphe est généralement évaluée sur la base d'indices de « centralité », qui permettent d'établir un classement reflétant leur influence ou leur rôle au sein du réseau (Koschützki et al., 2005). À l'instar des indices d'intérêt en analyse d'association (support, confiance, lift, etc.), de très nombreuses mesures de centralité ont été proposées au fil du temps, chacune offrant une perspective différente sur ce qui caractérise un nœud important. Comme le résumant Brughmans et Peeples (2023, p. 133), le nœud le plus important peut par exemple être celui (i) qui a de nombreuses relations (centralité de degré), (ii) qui est connecté à beaucoup d'autres sommets eux-mêmes importants (centralité de vecteur propre), (iii) qui fait le pont entre plusieurs groupes de sommets (centralité d'intermédiarité), ou encore (iv) qui est proche de tous les autres sommets du réseau (centralité de proximité).

La mise en œuvre des indices associés aux scénarios (iii) et (iv) s'accompagne d'un prérequis délicat dans notre cas, puisque les liens du réseau devraient théoriquement pouvoir s'interpréter comme des intermédiaires pour le flux de ressources (p. ex., les chemins que pourrait emprunter un virus pour se propager d'une personne *A* à une personne *Z* à l'autre bout du réseau). Nous choisissons par conséquent de nous limiter à des mesures reflétant les deux premières situations.

Centralité de degré Certainement l'une des mesures les plus simples et les plus intuitives, la centralité de degré (*degree centrality* en anglais) évalue l'importance d'un sommet donné en fonction du nombre de liens qui le rattachent à d'autres nœuds du réseau. Reflétant ainsi essentiellement une « activité relationnelle directe » (Lemieux & Ouimet, 2004, p. 23), elle se décline en trois variantes dans le cas des graphes orientés : (i) la centralité de degré sortant, qui mesure le nombre de liens dirigés *depuis* un nœud donné ; (ii) la centralité de degré entrant, qui évalue le nombre de liens dirigés *vers* un sommet donné ; (iii) et la centralité de degré total, qui comptabilise le nombre de liens entrants et sortants associés à un sommet donné. Selon ces différents indices de centralité, un type sémantique sera important dans notre cas s'il est souvent présent respectivement (i) en antécédent de règles, (ii) en conséquent de règles, (iii) en antécédent et/ou en conséquent de règles.

Centralité de vecteur propre La centralité de vecteur propre (*eigenvector centrality* en anglais) mesure l'importance d'un type proportionnellement à l'importance des nœuds avec lesquels il est en relation. L'intérêt de cet indice réside dans le fait qu'il tient compte non seulement de la quantité de liens que peuvent entretenir les différents types sémantiques du réseau, mais aussi de la qualité de ces connexions. Il peut toutefois produire des résultats incohérents si l'orientation des liens est prise en compte (Newman, 2018, p. 161). Pour cette raison, la direction des relations a été ignorée lors du calcul de cet indice.

Nous présentons dans le tableau 4.10 les scores de centralité atteints par chaque type sémantique. La figure 4.9 résume quant à elle les résultats de tests de corrélation (test τ de Kendall, comme recommandé par Shao et al., 2018) entre les différentes mesures de centralité sous la forme d'une carte de chaleur.

Type sémantique	Degré sortant		Degré entrant		Degré total		Vecteur propre	
	Valeur	Rang	Valeur	Rang	Valeur	Rang	Valeur	Rang
Anm-AGT	1	4	1	4	2	6	0.25	6
Anm-Coll-AGT	1	4	0	5	1	7	0.16	8
Art-INS	3	2	4	2	7	3	0.75	3
Art-LOC	1	4	0	5	1	7	0.12	9
Art-RES	4	1	4	2	8	2	0.97	2
Art-THM	1	4	0	5	1	7	0.16	8
Art-Coll-INS	1	4	0	5	1	7	0.16	8
Art-Coll-RES	1	4	0	5	1	7	0.16	8
Art*Cog-RES	2	3	0	5	2	6	0.17	7
Art*Ist-AGT	2	3	2	3	4	5	0.54	5
Cog-RES	1	4	0	5	1	7	0.16	8
Cog*Evt-TSP	0	5	2	3	2	6	0.03	10
Dom-TSP	3	2	2	3	5	4	0.66	4

4.3. Réseau général

Type sémantique	Degré sortant		Degré entrant		Degré total		Vecteur propre	
	Valeur	Rang	Valeur	Rang	Valeur	Rang	Valeur	Rang
Evt-TSP	2	3	15	1	17	1	1.00	1
Evt-Coll-TSP	1	4	0	5	1	7	0.16	8
Ist-AGT	1	4	0	5	1	7	0.16	8
N/A-Coll-RES	1	4	0	5	1	7	0.16	8
Nat-RES	1	4	0	5	1	7	0.16	8
Ppt-CAU	1	4	0	5	1	7	0.00	11
Ppt-MAN	1	4	0	5	1	7	0.16	8
Ppt-RES	1	4	0	5	1	7	0.16	8

TABLEAU 4.10 – Scores de centralité des types sémantiques examinés. Les scores les plus élevés sont colorés en vert. Les types sémantiques sont classés par ordre alphabétique

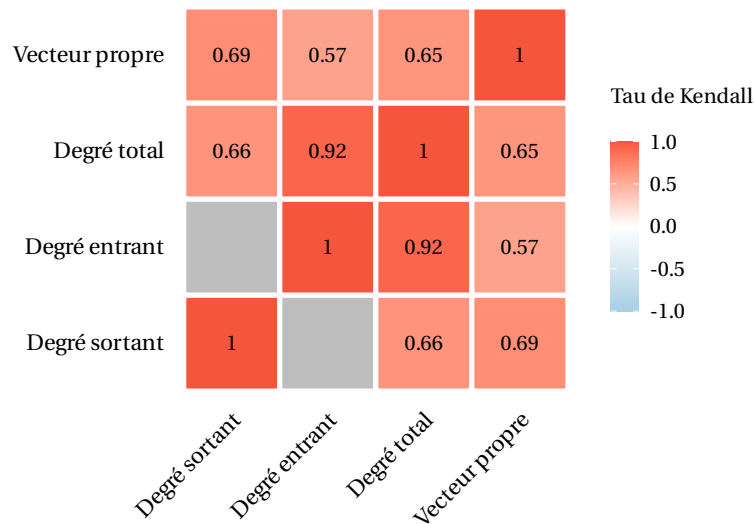


FIGURE 4.9 – Corrélations de rang (test τ de Kendall) entre les différentes mesures de centralité. Les résultats non significatifs n'apparaissent pas

Sans grande surprise, seule une poignée de types sémantiques se révèlent importants au sein du réseau, notamment Événement-TRANSPPOSITION, Artéfact-RÉSULTAT et Artéfact-INSTRUMENT. Le type Événement-TRANSPPOSITION se distingue en étant à la fois le type le plus présent en conséquent de règles et le type le plus susceptible de s'associer à d'autres types importants, tandis que le type Artéfact-RÉSULTAT apparaît comme le plus fréquent en antécédent de règles. Si l'hypothèse sur la direction des extensions sémantiques présentée dans la section 4.2.2 se vérifie, ces résultats pourraient indiquer que le type Événement-TRANSPPOSITION est le plus enclin à déclencher des phénomènes de polysémie, en servant d'acception de base dans des schémas de dérivation sémantique (en l'occurrence de métonymies). Ils pourraient également signaler que le type Artéfact-RÉSULTAT constitue la catégorie

sémantique la plus à même de se profiler comme un point de convergence conceptuel, suivant sa capacité à figurer comme sens cible dans des schémas de dérivation sémantique (pour la métonymie également). Pour rappel, le score de réalisation autonome obtenu pour ce type (0.19) dans la section 4.2.2 est très faible, ce qui constitue un indice supplémentaire du fait que cette catégorie peut être dérivée sémantiquement. De manière plus générale, l'absence de corrélation significative entre les scores de centralité de degré entrant et de centralité de degré sortant (figure 4.9) indique que le fait d'être une source privilégiée ne présage en rien du fait d'être ou non une cible privilégiée.

Centralisation

Si l'importance des types sémantiques a jusqu'à présent été examinée à un niveau local, elle peut également être mesurée à l'échelle du réseau entier à l'aide d'indices globaux appelés mesures de « centralisation ». Représentant la distribution des scores de centralité, ces dernières servent essentiellement à quantifier le degré de dépendance d'un graphe par rapport à ses nœuds les plus centraux (Zweig, 2016, p. 244). Nous retenons pour cette analyse les mesures de centralisation de degré total et de vecteur propre, chacune permettant de mettre en lumière une structure hiérarchisée particulière (Csardi & Nepusz, 2006) : un graphe en étoile pour la centralisation de degré, et un réseau composé d'une seule relation (et potentiellement de nombreux types isolés) pour la centralisation de vecteur propre. Dans les deux cas, un score de 0 signifie qu'aucun nœud ne domine, tandis qu'un score de 1 indique qu'un nœud unique prévaut sur tous les autres.

Les valeurs obtenues sont de 0.64 pour la centralisation de degré total et de 0.81 pour la centralisation de vecteur propre, ce qui suggère qu'un petit nombre de types concentrent la plupart des liens et que ces mêmes types tendent à se connecter les uns aux autres. Il reste que ces scores doivent être interprétés avec prudence. En l'absence d'une référence comparative, il demeure difficile de déterminer si ceux-ci constituent le fruit du hasard étant donné la taille et la densité du graphe, ou s'ils reflètent au contraire une propriété distinctive du réseau étudié.

Une solution à ce problème consiste en l'application d'un test CUG (*Conditional Uniform Graph* en anglais) (Butts, 2008), qui permet d'évaluer si une caractéristique spécifique d'un graphe est présente à un niveau statistiquement significatif (Cranmer et al., 2020, p. 20). L'analyse met en jeu trois étapes pour une mesure d'intérêt x telle que l'indice de centralisation de degré et de vecteur propre :

- (i) la génération d'un grand nombre de réseaux aléatoires ($n = 2'000$ dans notre cas) partageant certaines caractéristiques du réseau étudié (ici, le nombre de sommets et la densité) ;
- (ii) le calcul de x pour chacun de ces réseaux simulés ;
- (iii) et enfin, la comparaison du score x du réseau étudié avec les valeurs x des réseaux simulés.

4.3. Réseau général

L'output du test se résume au final à deux proportions (Stoltz & Taylor, 2024, p. 265-266) : $\Pr(X \geq \text{Obs})$, soit la proportion de x simulés supérieurs ou égaux au x observé dans notre réseau (i.e. la p-valeur empirique du test unilatéral droit); et $\Pr(X \leq \text{Obs})$, soit la proportion de x simulés inférieurs ou égaux au x observé dans notre réseau (i.e. la p-valeur empirique du test unilatéral gauche). À compter que le seuil de signification soit fixé à 0.05, nous pouvons conclure qu'il est peu probable que le hasard explique la valeur observée x si celle-ci est supérieure ou inférieure à 95 % des valeurs x simulées.

Comme le montrent les figures 4.10 et 4.11, les résultats des tests CUG appliqués à nos deux mesures de centralisation s'avèrent concluants, puisque dans les deux cas, on observe $\Pr(X \geq \text{Obs}) = 0$ et $\Pr(X \leq \text{Obs}) = 1$. Les valeurs observées pour les réseaux simulés sont systématiquement inférieures aux scores de centralisation obtenus pour notre graphe. Ces résultats confirment que les niveaux de centralisation observés sont statistiquement significatifs, soulignant en particulier l'importance du caractère radial du réseau (figure 4.10).

Il convient de réfléchir aux origines d'une structure non aléatoire aussi hiérarchisée. Une explication souvent avancée dans le cas de distributions de scores de centralité très déséquilibrées comme c'est le cas ici est le processus d'attachement préférentiel, qui veut que les acteurs « populaires » tendent à le devenir encore plus au fil du temps, car ils bénéficient dès le départ d'une grande visibilité (Robins, 2015, p. 28). Nous remarquons que le cœur de notre réseau est le type Événement-TRANSPPOSITION, ce qui n'est pas anodin compte tenu de la relation étroite que les nominalisations déverbiales entretiennent avec les verbes et le domaine de l'action. Ces résultats pourraient ainsi concourir à montrer que l'héritage morphologique du verbe façonne l'activité ambiguë des noms déverbaux (cf. Melloni, 2011; Scott, 2010), les différenciant fondamentalement des autres catégories de noms (p. ex. simples ou désadjectivaux).

4.3.3 Structures fréquentes

Nous laissons de côté les types sémantiques pour la fin de ce chapitre afin de nous concentrer sur la structure même du graphe. Cette section vise concrètement à mettre au jour des motifs d'associations récurrents au sein du réseau étudié, pouvant contribuer à la compréhension des dynamiques et normes relationnelles qui le sous-tendent. En analyse de réseau, deux types de microstructures sont généralement considérés pour ce faire (Wasserman & Faust, 1994, p. 18-19) : la dyade, un ensemble de deux nœuds et le lien qui les rattache potentiellement; et la triade, un triplet de sommets pouvant ou non être reliés entre eux.

Dyades

Constituants les plus élémentaires d'un réseau, les dyades peuvent prendre trois formes dans le contexte des graphes orientés : les dyades nulles, caractérisées par l'absence de lien entre les deux nœuds ($\bullet \not\rightarrow \bullet$); les dyades asymétriques, où un lien unidirectionnel relie deux sommets ($\bullet \rightarrow \bullet$); et les dyades mutuelles ($\bullet \longleftrightarrow \bullet$), marquées par la présence d'un

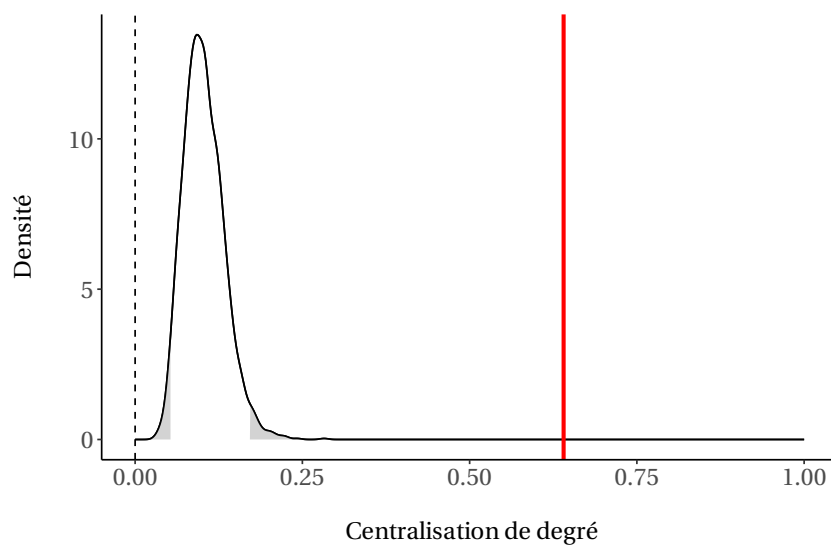


FIGURE 4.10 – Résultats d'un test CUG pour la centralisation de degré. Le diagramme de densité illustre les scores obtenus pour 2'000 réseaux simulés. Les queues grises de la distribution représentent les 5 % de scores les plus extrêmes. La ligne verticale rouge représente le score obtenu pour le réseau étudié

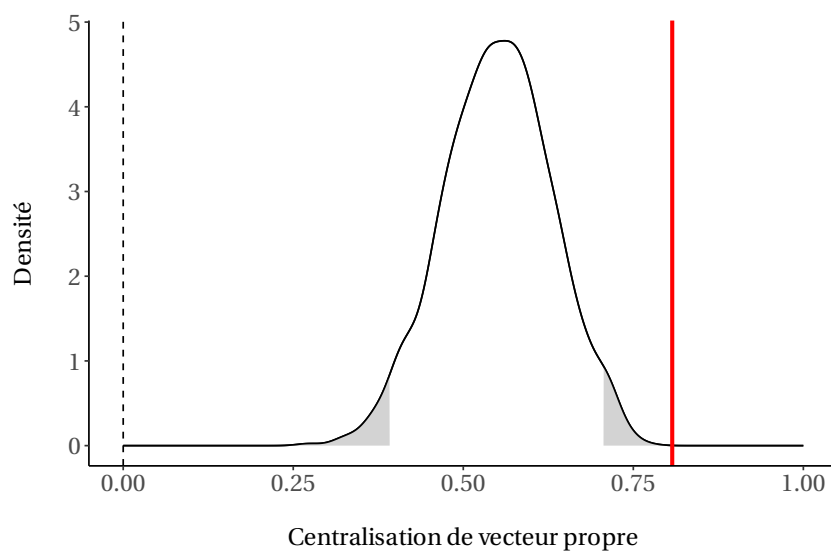


FIGURE 4.11 – Résultats d'un test CUG pour la centralisation de vecteur propre. Le diagramme de densité illustre les scores obtenus pour 2'000 réseaux simulés. Les queues grises de la distribution représentent les 5 % de scores les plus extrêmes. La ligne verticale rouge représente le score obtenu pour le réseau étudié

4.3. Réseau général

lien bidirectionnel. Malgré leur apparente simplicité, les dyades sont d'une grande utilité en analyse systémique, notamment pour l'étude des dynamiques de réciprocité (Rawlings et al., 2023).

Nous avons observé dans la section 4.2.2 des variations notables de réversibilité au sein des règles d'association, en conjecturant que celles-ci pouvaient refléter le caractère plus ou moins asymétrique de schémas d'extension conceptuelle. Il reste toutefois à proposer une analyse plus formelle de cette supposition. Nous avons ainsi entrepris un recensement des différentes formes de dyades présentes dans le réseau d'associations. Après amalgamation des liens multiples, les résultats de ce comptage révèlent que le réseau étudié comporte 7 liens mutuels, 16 liens asymétriques et 187 liens nuls. En d'autres termes, 30 % des liens observés sont mutuels, tandis que 70 % sont asymétriques.

Comme pour les mesures de centralisation abordées précédemment, évaluer la qualité de ces valeurs reste difficile sans un cadre de référence adéquat. Nous avons donc à nouveau réalisé un test CUG basé sur 2'000 simulations. Les résultats montrent que le taux de réciprocité observé dans notre réseau (0.304, soit 30.4 %) est systématiquement supérieur à ceux de réseaux simulés de densité comparable. Les taux de réciprocité maximum et moyen atteignent en effet respectivement 0.2 (soit 20 %) et 0.038 (ou 3.8 %) pour les graphes aléatoires. Il ressort donc de l'analyse que le réseau observé présente un taux de réciprocité allant bien au-delà des attentes du hasard.

Il convient d'examiner les raisons pour lesquelles le réseau étudié affiche un tel niveau de réciprocité. L'analyse du graphe général présenté en 4.8 met premièrement en lumière un recoupement considérable entre la quasi-totalité des liens réciproques relevés ($n = 6/7$) et les liens correspondant à des règles tripartites, représentés en rouge dans le réseau. Cette configuration pointe vers l'existence de deux « microparadigmes » sémantiques, déjà évoqués dans la section précédente : le premier regroupe les types Événement-TRANSPOSITION, Artéfact-INSTRUMENT et Artéfact-RÉSULTAT (123) ; le second, les types Domaine-TRANSPOSITION, Artéfact-RÉSULTAT et Artéfact*Institution-AGENT (124).

(123) Événement-TRANSPOSITION/Artéfact-INSTRUMENT/Artéfact-RÉSULTAT
teinture 'action de teindre' / 'substance servant à teindre' / 'coloration obtenue'

(124) Domaine-TRANSPOSITION/Artéfact-RÉSULTAT/Artéfact*Institution-AGENT
menuiserie 'art de menuiser' / 'ouvrage menuisé en bois' / 'atelier qui menuise'

Reste un dernier lien mutuel, celui entre les types Entité animée-AGENT et Artéfact-INSTRUMENT, qui semble échapper à cette structuration paradigmatique. Comme mentionné précédemment, il est plausible que l'origine des noms ambigus entre ces deux types soit davantage morphologique que purement sémantique (par exemple, en raison de la forte productivité des patrons de construction agentif et instrumental pour les dérivés en *-eur*), ce qui pourrait expliquer la réciprocité du lien dans leur cas.

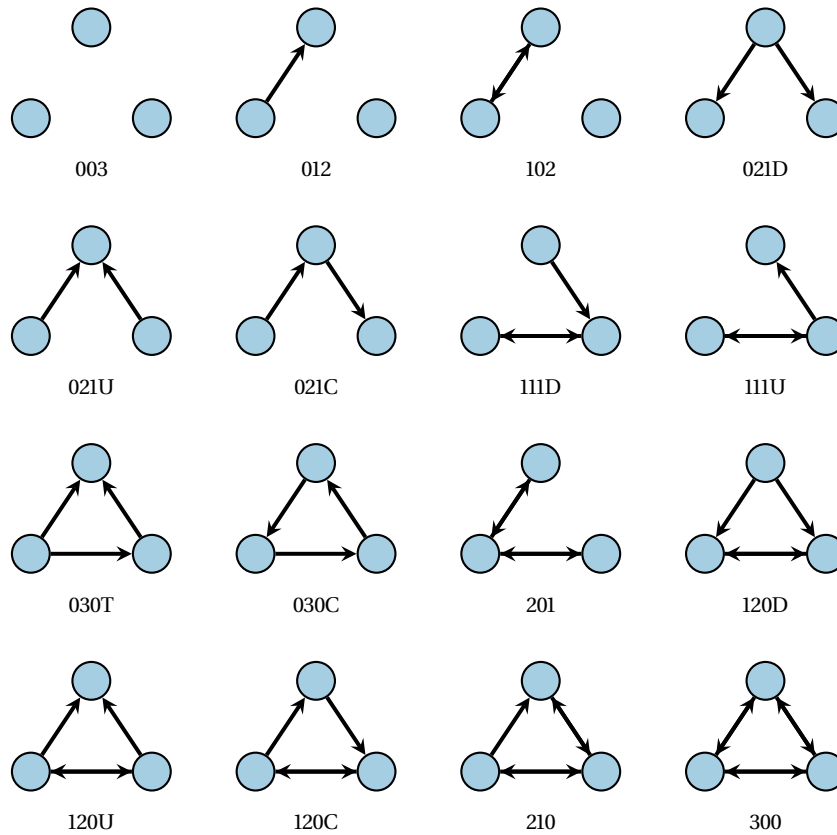


FIGURE 4.12 – Triades pouvant être observées au sein d'un réseau orienté (adapté de Rawlings et al., 2023, p. 150)

Triades

Bien qu'intéressantes, les dyades n'offrent au final qu'une image limitée de la structure du graphe. Les triades, surnommées les « molécules des réseaux » par Kadushin (2012, p. 204), en représentent une extension bien connue, permettant d'appréhender des phénomènes plus complexes tels que la transitivité (si un nœud *A* est lié à *B*, et que *B* est lié à *C*, *A* et *C* sont-ils également en relation?). Les réseaux orientés comme le nôtre peuvent en théorie inclure 16 triades différentes, illustrées par la figure 4.12. Chacune d'entre elle est généralement associée à un code « MAN » (Holland & Leinhardt, 1970), qui fournit des informations sur le nombre de dyades mutuelles (M), asymétriques (A) ou nulles (N) qu'elle contient. Par exemple, la triade 012 ne comporte aucun lien mutuel, mais comprend un lien asymétrique et deux liens nuls, alors que la triade 210 se caractérise par deux liens mutuels, un lien asymétrique, mais aucun lien nul. Certaines triades s'accompagnent par ailleurs de lettres renseignant sur la direction des relations : U (*up*) pour les relations orientées vers un nœud supérieur ; D (*down*) pour les relations descendant vers un nœud inférieur ; C (*cyclical*) pour les relations circulaires entre trois nœuds ; et T (*transitive*) pour signaler que si un nœud *A* est lié à un nœud *B* et

4.3. Réseau général

que ce nœud *B* est lié à un nœud *C*, alors *A* est également lié à *C*. La triade 021U se définit ainsi par 0 lien mutuel, 2 liens asymétriques et 1 lien nul, le tout pointant vers le haut.

Nous avons, comme pour les dyades, effectué un recensement des différentes formes de triades au sein du réseau. Les résultats obtenus, synthétisés dans la deuxième colonne du tableau 4.11, révèlent que si certaines d'entre elles sont très présentes (003, 012, 102, 021U, 111D), d'autres sont complètement absentes (021C, 030T, 030C, 120D, 120U, 120C). Il reste que ces différents comptages sont difficiles à interpréter sans point de comparaison adéquat. Comme le recommandent Rawlings et ses collègues (2023), nous avons donc à nouveau eu recours à des tests CUG dans le but de déterminer quelles formes de triades, parmi les 16 possibles, sont observées à des niveaux supérieurs ou inférieurs à ceux qu'on pourrait attendre par chance. Le conditionnement des réseaux simulés repose ici sur le nombre de dyades mutuelles, asymétriques et nulles que comporte notre réseau d'intérêt.

Triade	Val. obs.	Min.	Max.	<i>M</i>	Pr($X \geq \text{Obs}$)	Pr($X \leq \text{Obs}$)
003	1'013	925	959	937.5	0	1
012	114	207	263	243.2	1	0
102	86	83	119	106.4	0.9995	0.0005
021D	1	0	15	4.923	0.999	0.0255
021U	79	0	14	4.926	0	1
021C	0	1	21	9.794	1	0
111D	28	1	18	9.148	0	1
111U	1	2	20	9.169	1	0
030T	0	0	5	0.353	1	0.704
030C	0	0	2	0.1265	1	0.8815
201	5	0	11	3.43	0.2375	0.898
120D	0	0	2	0.1870	1	0.8255
120U	0	0	3	0.167	1	0.8465
120C	0	0	4	0.38	1	0.679
210	1	0	4	0.2995	0.2615	0.967
300	2	0	1	0.0275	0	1

TABLEAU 4.11 – Recensement de triades et résultats de tests CUG basés sur 2'000 réplifications. Val. obs. = Nombre de triades dans le réseau étudié. Min., Max., *M* = Valeurs minimum, maximum et moyennes obtenues pour les 2'000 réseaux aléatoires. Pr($X \geq \text{Obs}$) et Pr($X \leq \text{Obs}$) = Proportion de valeurs simulées respectivement supérieures ou égales, et inférieures ou égales à la valeur du réseau observé. Les résultats significatifs à $p < 0.05$ sont surlignés en vert clair

Les résultats des tests CUG sont ajoutés au tableau 4.11 sous la forme de cinq colonnes : la troisième, la quatrième et la cinquième indiquent combien de triades d'une forme donnée sont observées dans les réseaux aléatoires (nombre minimum, nombre maximum et moyenne) ; la sixième, la proportion de valeurs simulées supérieures ou égales à la valeur

du réseau observé; et la septième, la proportion de valeurs simulées inférieures ou égales à la valeur du réseau observé. La comparaison de ces deux dernières colonnes révèle des disparités statistiques importantes. Les triades 003, 021U, 111D et 300 apparaissent en effet plus fréquemment que prévu, tandis que les triades 012, 102, 021D, 021C et 111U sont moins courantes qu'attendu.

On peut s'interroger sur la manière d'interpréter ces particularités structurelles. La fréquence élevée des triades 021U et 111D, comparées à la rareté de la triade 021C, suggère que l'ambiguïté régulière des noms déverbaux tend à se structurer de manière radiale plutôt que par ressemblance de famille⁵⁷. En effet, les chaînes de relations unidirectionnelles de type $A \rightarrow B \rightarrow C$ sont statistiquement très rares. Il apparaît par ailleurs que cette structure radiale s'oriente très distinctement vers le cœur du réseau, comme en témoignent la triade 021U et sa variante avec un lien bidirectionnel 111D, largement plus fréquentes que la triade 021D et sa voisine 111U. Ces résultats, cohérents avec ce qui a été relevé au sujet de la centralisation du réseau, poussent à s'intéresser au type de structure privilégié selon la partie du lexique concerné. L'hypothèse peut être formulée que les déverbaux se singularisent par une prédilection pour des structures d'ambiguïté radiale, par opposition au reste du lexique nominal. Cette hypothèse serait bien sûr à vérifier avec un examen des structures d'ambiguïté du lexique non construit, ou d'autres types de dérivés (p. ex. désadjectivaux).

Un autre résultat notable réside dans la sous-représentation des triades 012 et 102, qui ne mettent en jeu qu'une seule relation entre trois sommets. Ce constat suggère une préférence des déverbaux pour les relations multiples impliquant plus de deux types. L'importance des structures tripartites dans l'espace sémantique des nominalisations semble en outre confirmée par la fréquence marquée de la triade 300, qui reste toutefois le seul ensemble à trois relations à être statistiquement intéressant (*vs* 030C, 030T, 120C, 120D, 120U et 210). On peut avancer l'idée que les occurrences de triade 300 sont principalement dues aux règles tripartites « réciproques », et ainsi peut-être aux microparadigmes relevés précédemment (cf. figure 4.8), bien que ces occurrences puissent théoriquement résulter de combinaisons de règles bipartites.

4.4 Bilan

Ce chapitre visait d'une part à quantifier et à décrire les différentes associations sémantiques instanciées par les noms déverbaux, et d'autre part à mettre au jour de potentiels mécanismes d'extension sémantique à l'œuvre derrière ces schémas d'ambiguïté. Nous résumons ci-dessous nos principaux résultats.

Non-arbitrarité et motivation L'ambiguïté des noms déverbaux est jalonnée par un nombre important d'associations sémantiques, la plupart du temps très peu fréquentes. Une

57. Darmesteter (1887, p. 73-83) distinguait déjà il y a plus de cent ans les « transformations » de sens par « rayonnement » de celles par « enchaînement ».

4.4. Bilan

analyse d'association a révélé qu'une partie des schémas observés ne peuvent être imputés au hasard, ce qui peut amener à penser que la cooccurrence de types sémantiques est potentiellement motivée. Cette motivation pourrait s'expliquer en grande partie par des relations sémantiques de type métonymique entre types.

Contenu des associations Notre analyse a confirmé de manière quantitative l'existence de plusieurs associations ayant déjà fait l'objet de travaux par le passé (p. ex. événement/résultat, agent/instrument), en apportant toutefois un éclairage plus fin sur leur contenu, tant sur le plan ontologique que relationnel. Elle a aussi dévoilé un certain nombre d'associations encore largement inconnues (p. ex. événement/événement collectif, événement/objet naturel).

Structure des relations Il ressort des analyses que les associations régulières peuvent être composées de deux, mais aussi de trois types sémantiques, signalant potentiellement l'existence de microparadigmes sémantiques dans le système sémantique des nominalisations. Il apparaît par ailleurs que la régularité entre types peut se doubler d'une régularité intratypes, comme en témoigne la présence marquée de facettes sémantiques parmi les schémas examinés. Enfin, il a aussi été montré que les associations entre types sont tendanciellement dirigées (si *A* implique *B*, *B* n'implique pas nécessairement *A*). Ces préférences pourraient refléter le caractère plus ou moins asymétrique de schémas d'extension conceptuelle à l'œuvre dans la constitution de formes déverbales ambiguës.

Organisation systémique Une analyse de réseau fondée sur les règles d'association précédemment extraites a montré que l'ambiguïté des noms déverbaux présente une organisation hiérarchique très centralisée, prenant essentiellement la forme d'un réseau en étoile dont le cœur est le type sémantique Événement-TRANSPOSITION. Ces résultats ont au moins deux implications potentielles : que l'ambiguïté des nominalisations se structure radialement plutôt que par ressemblance de famille, mais aussi que la sémantique des noms déverbaux se différencie probablement par certains aspects de celle du lexique simple en raison de la présence du verbe en ascendant morphologique.

Nos résultats soulignent dans l'ensemble l'importance que revêt la construction morphologique pour l'ambiguïté des déverbaux. Le prochain et dernier chapitre nous servira à explorer plus avant ce phénomène, en considérant à la fois les procédés morphologiques et les propriétés des verbes de base des nominalisations.

Chapitre 5

L'ambiguïté nominale au prisme de la construction morphologique

La littérature s'accorde généralement sur le fait que plusieurs contraintes d'ordre morphologique pèsent sur les catégories de sens instanciées par les noms déverbaux. En français, certains procédés morphologiques nominalisateurs sont connus pour leurs préférences sémantiques : par exemple, *-aire* forme régulièrement des noms d'humain comme *abonnataire* et *signataire* (Schnedecker & Aleksandrova, 2016); *-age* et *-ment*, des noms d'événement comme *nettoyage* et *déblaiement* (Fradin, 2016); et *-ance*, des noms de qualité comme *rutilance* et des noms de procès comme *accoutumance* (Dal & Namer, 2010). Les propriétés sémantiques des verbes de base des nominalisations, qu'il s'agisse de l'aspect lexical ou du type de procès décrit, sont aussi fréquemment évoquées dans ce contexte. Nous l'avons vu dans le chapitre 1, l'*Aspect Preservation Hypothesis* (Fábregas et al., 2012) postule qu'un nom construit peut répliquer l'aspect lexical du verbe dont il dérive s'il dénote une éventualité. Aussi le verbe d'activité *patiner* amène-t-il naturellement au nom d'activité *patinage*, le verbe d'état *espérer* au nom d'état *espérance*, et le verbe d'accomplissement *façonner* au nom d'accomplissement *façonnement*. S'agissant des rôles sémantiques, on reconnaît volontiers que les dérivés peuvent lexicaliser des arguments syntaxiques ou sémantiques de leurs bases. Par exemple, les noms dénotant des instruments sont souvent liés à des verbes requérant un participant instrumental (Plag et al., 2023).

Ces contraintes peuvent, selon toute vraisemblance, également s'étendre à l'ambiguïté des nominalisations. Rappelons que selon Scott (2010, p. 713), le suffixe *-ung* en allemand serait le seul procédé morphologique permettant de former régulièrement des noms déverbaux actionnels avec une seconde lecture concrète (p. ex. *Tätowierung* 'action de tatouer'/'tatouage'). Pour le français, Huyghe et Lombard (2022) ont examiné comment la télicité du verbe peut conduire à la polysémie événement/résultat dans le cas des noms déverbaux en *-age*. Enfin, concernant les types de procès décrits, Melloni (2011, p. 239) suggère

5.1. Procédés morphologiques

que les verbes de création sont intimement liés à la forme de polysémie événement/produit en italien (p. ex. *costruzione* ‘action de construire’/‘résultat de l’action de construire’).

Il reste qu’une étude systématique et à grande échelle de ces phénomènes manque encore. Dans ce qui suit, nous cherchons ainsi à développer l’analyse holistique des associations sémantiques proposée dans le chapitre précédent en évaluant dans quelle mesure et de quelle façon la construction morphologique peut conditionner l’ambiguïté des noms déverbaux. Nous nous attardons en particulier sur l’influence que peuvent avoir les procédés morphologiques (section 5.1), ainsi que les propriétés sémantiques des verbes de base (section 5.2) sur la construction de nominalisations ambiguës. La dernière partie de ce chapitre examine enfin de quelle manière ces deux catégories d’attributs peuvent collaborer (section 5.3), de possibles affinités entre verbes et procédés morphologiques étant en effet possibles.

Quelques mots sur notre méthodologie avant de commencer. Si dans la section 5.1 nous reprenons essentiellement certains outils employés dans le chapitre précédent (analyse d’association standard et analyse de réseau), les résultats présentés dans les sections 5.2 et 5.3 reposent en revanche sur une version dite « séquentielle » de l’analyse d’association (Agrawal & Srikant, 1995), qui rend possible l’identification de motifs d’association récurrents dans des listes ordonnées d’éléments. En pratique, nous cherchons ainsi à déchiffrer (i) quelles propriétés sémantiques des bases verbales et/ou quels procédés morphologiques sont les plus à même de déclencher (ii) la réalisation de certaines catégories sémantiques ou formes d’ambiguïté par les dérivés. Le tableau 5.1 illustre un extrait de l’input et de l’output d’une extraction de règles séquentielles effectuée sur nos données.

Propriété(s) morphosémantique(s)	Type(s) sémantique(s) combiné(s)
{V à télicité variable, V à objet patientif, Suffixe <i>-ion</i> }	⇒ {Evt*Sta-TSP}
{Suffixe <i>-ing</i> }	⇒ {Dom-TSP}
{V duratif, V télique, V à sujet agentif, Suffixe <i>-age</i> }	⇒ {Evt-TSP, Art-RES}
{V télique, V à objet patientif, Suffixe <i>-ure</i> }	⇒ {Evt-TSP, Nat-RES}
{V duratif, V à objet patientif, Suffixe <i>-oir</i> }	⇒ {Art-INS, Art-LOC}

TABLEAU 5.1 – Extrait de l’input (propriétés morphosémantiques) et de l’output (types sémantiques uniques ou ensembles de types sémantiques) d’une extraction de règles séquentielles

5.1 Procédés morphologiques

Cette section teste l’hypothèse selon laquelle les procédés morphologiques peuvent influencer les cooccurrences de sens possibles dans le domaine nominal. Vu sous un autre angle, ce postulat revient à affirmer que la capacité à former des mots régulièrement ambigus entre plusieurs types sémantiques pourrait constituer un trait distinctif des procédés

morphologiques. Nous rappelons à ce titre la distinction établie dans le chapitre 1 entre deux notions proches mais distinctes, à savoir l'ambiguïté lexicale et la « polysémie » affixale, que nous nommerons désormais « polyfonctionnalité affixale » pour plus de clarté. Comme indiqué précédemment, ces deux phénomènes ne sont pas équivalents. Le fait qu'un affixe soit associé à plusieurs fonctions sémantiques n'implique pas nécessairement que tous ses dérivés les réalisent conjointement. Inversement, le fait qu'une nominalisation instancie plusieurs types sémantiques n'implique pas que le procédé morphologique avec lequel elle est formée ait ces sens pour fonctions sémantiques – notamment en cas d'extension figurative idiosyncrasique (Salvadori & Huyghe, 2023).

Notre étude du rapport entre ambiguïté lexicale et polyfonctionnalité affixale se divise en deux étapes. Seront premièrement examinées les potentialités sémantiques des procédés morphologiques au niveau du lexème (p. ex. *accordeur*₁ 'personne qui accorde', *accordeur*₂ 'instrument pour accorder') et du vocable (p. ex. *accordeur* 'personne qui accorde' / 'instrument pour accorder') (section 5.1.1). Nous chercherons en particulier à déterminer quelles catégories sémantiques sont réalisées par quel procédé morphologique et à quelle fréquence, mais surtout à vérifier si l'ambiguïté des dérivés déverbaux varie en fonction du suffixe ou de la forme de conversion qu'ils représentent. Pour approfondir notre examen d'un point de vue statistique, nous proposerons dans un second temps (section 5.1.2) une étude de cas portant sur les procédés morphologiques les plus polyfonctionnels de notre base de données (-ion, -age, -ment, -ure, -erie, conversion sur thèmes 0 et 12). Cette étude prendra la forme d'une analyse d'association couplée à une analyse de réseau, suivant la méthodologie en deux parties employée dans le chapitre précédent. L'objectif sera ici de proposer une analyse plus formelle des particularités sémantiques des dérivés ambigus construits avec ces différents procédés, tant sur le plan de l'interdépendance des catégories sémantiques (analyse d'association) que sur celui de leurs relations structurelles (analyse de réseau).

5.1.1 Premières considérations

Nous avons jusqu'à présent éludé la place que peuvent prendre les procédés morphologiques dans le sémantisme des nominalisations. Dans ce qui suit, nous revenons ainsi sur ce que permettent « sémantiquement » les suffixes et les formes de conversion du point de vue lexématique, puis présentons quelques considérations relatives au rapport entre polyfonctionnalité affixale et ambiguïté nominale.

Richesse sémantique et fréquence de réalisation

La figure 5.1 illustre les différents types sémantiques réalisés au moins deux fois par les 46 procédés morphologiques recensés dans la base de données. Ce seuil arbitraire de deux occurrences vise à éliminer autant que possible les types sémantiques qui pourraient être

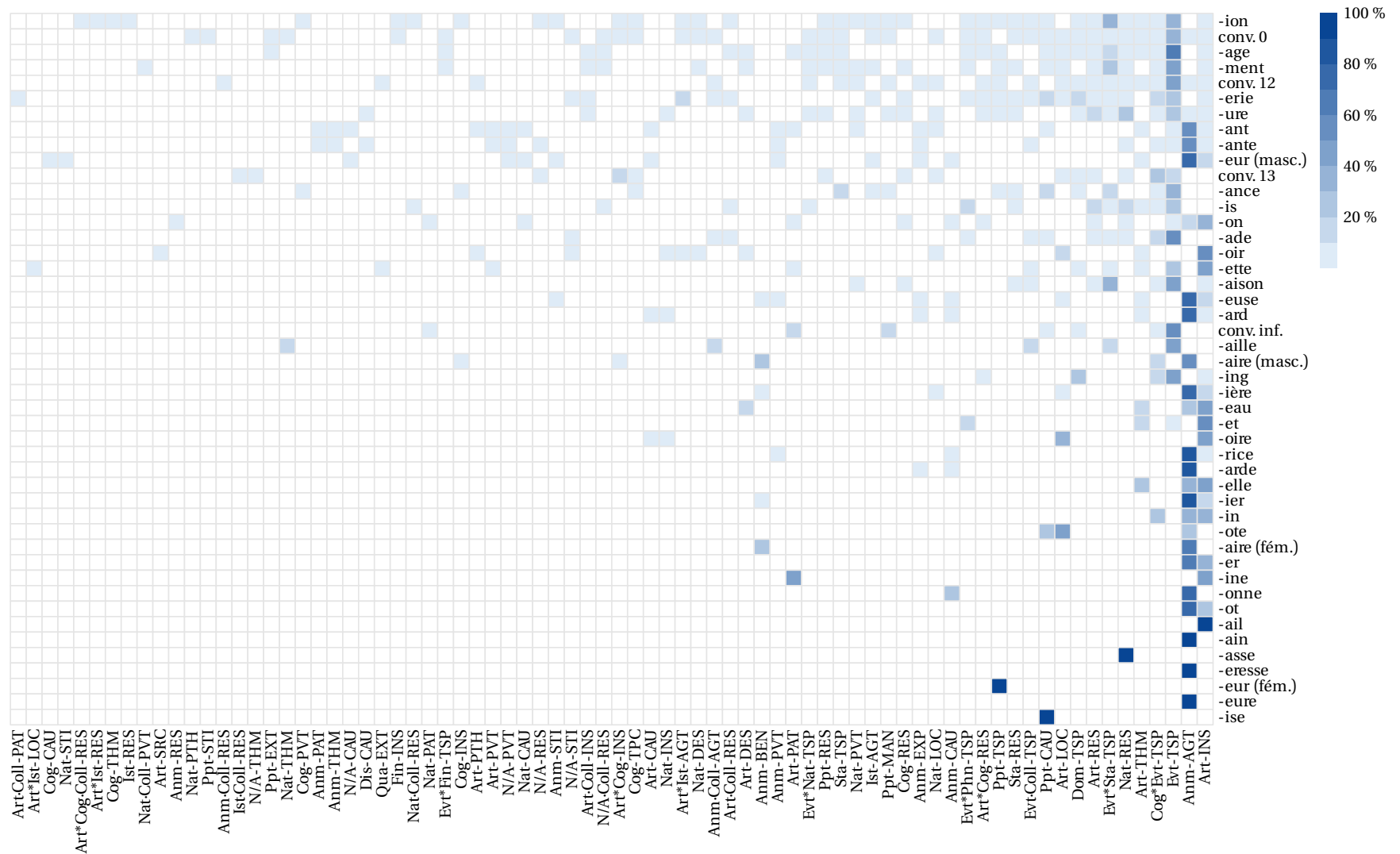


FIGURE 5.1 – Types sémantiques combinés par procédé morphologique. Seuls les types instanciés au moins deux fois par un procédé morphologique sont représentés. Les couleurs renseignent sur le pourcentage de dérivés formés avec un procédé morphologique donné. Les procédés morphologiques sont classés de bas en haut, du moins polyfonctionnel au plus polyfonctionnel. Les types sont classés de gauche à droite, du moins mobilisé au plus mobilisé

idiosyncrasiques et résulter d'effets de lexicalisation, ne dépendant ainsi pas de la dérivation elle-même (Salvadori & Huyghe, 2023)⁵⁸.

Un premier résultat intéressant concerne ici le nombre de types sémantiques combinés instanciés par ces différents suffixes et formes de conversion. Les plus riches sémantiquement sont dans l'ordre : le suffixe *-ion* ($n = 35$ types), la conversion sur thème 0 ($n = 34$), *-age* ($n = 23$), *-ment* ($n = 23$), la conversion sur thème 12 ($n = 23$), *-erie* ($n = 22$) et *-ure* ($n = 21$). À l'autre bout du spectre, quelques suffixes ne réalisent qu'un seul type sémantique : *-ail*, *-ain*, *-asse*, *-eresse*, *-eur* (fém.), *-eure* et *-ise*. Un second résultat notable touche à la fréquence de réalisation (calculée en nombre de lexèmes) des différents types. Les procédés morphologiques les plus polyfonctionnels manifestent une préférence marquée pour un (conversion sur thèmes 0 et 12) ou deux (*-ion*, *-age*, *-ment*) types particuliers, soit Événement-TRANSPOSITION et Événement*État-TRANSPOSITION, tout en réalisant de nombreuses autres catégories de sens à des fréquences relativement basses. À l'inverse, d'autres suffixes sont plus équilibrés sémantiquement (p. ex. *-in*, *-elle*).

Capacités de réalisation ambiguë

Nous pouvons chercher à déterminer si les différents procédés morphologiques ont une influence sur l'ambiguïté de leurs dérivés, en commençant par examiner le taux d'ambiguïté nominale qui leur est associé. L'ambiguïté est évaluée ici sans tenir compte des sens figurés ($n = 113$) et uniquement pour les types réalisés au moins deux fois par un procédé, afin de maintenir la cohérence avec le nombre de types mentionnés précédemment. Comme le montre le tableau 5.2, nos résultats révèlent d'importantes disparités entre les différents procédés morphologiques. 14 d'entre eux (p. ex. *-aire* (fém.), *-eresse*, *-eure*, *-onne*, *-ot*) – souvent agentifs et féminins – ne forment que des vocables monosémiques, alors que 5 autres sont au contraire associés à des taux d'ambiguïté nominale supérieurs à 1.3 (*-erie*, conversion sur thème 0, *-ure*, *-ing*, *-ine*). Si le taux identifié pour *-ine* n'est de toute évidence pas très informatif du fait du peu de vocables impliqués ($n = 3$), on peut en revanche estimer que ceux obtenus pour *-erie*, la conversion sur thème 0, *-ure* et *-ing* traduisent des comportements sémantiques particuliers qu'il serait intéressant d'examiner plus avant.

Nous ajoutons également au tableau 5.2 (dernière colonne) une mesure représentant la combinaison entre ambiguïté nominale et polyfonctionnalité des procédés morphologiques, de manière à documenter l'envergure sémantique de ces derniers. Il s'agit essentiellement d'un calcul de pondération (i.e. l'ambiguïté moyenne divisée par le nombre de types), où un score de 1 indique que chaque mot en *-x* instancie toutes les fonctions possibles de *-x*. Cette information répond (notamment pour les procédés polyfonctionnels) à la question de savoir à quel point un procédé est « génèreux » dans l'attribution de ses types à ses dérivés. Si l'on exclut les procédés morphologiques monofonctionnels (p. ex. *-ail*, *-ain*, *-asse*), les résultats

58. Comme dans le chapitre précédent, nous avons aussi exclu les 113 sens figuratifs pour les analyses présentées ci-dessous.

5.1. Procédés morphologiques

indiquent dans l'ensemble que plus un procédé morphologique est polyfonctionnel, moins il tend à répartir largement ses types entre ses dérivés.

Procédé	Vocables	Sens	Ambig. moyenne	ET	Types sémantiques	Pondération
-ade	131	150	1.15	0.39	11	0.10
-age	1'047	1'209	1.15	0.40	23	0.05
-ail	3	3	1	0	1	1
-aille	13	16	1.23	0.44	5	0.25
-ain	5	5	1	0	1	1
-aire (fém.)	40	40	1	0	2	0.50
-aire (masc.)	59	63	1.07	0.25	5	0.21
-aison	94	106	1.13	0.37	9	0.13
-ance/-ence	145	179	1.23	0.47	13	0.09
-ant/-ent	127	134	1.06	0.23	18	0.06
-ante/-ente	87	89	1.02	0.15	14	0.07
-ard	53	55	1.04	0.19	7	0.15
-arde	30	30	1	0	3	0.33
-asse	2	2	1	0	1	1
-eau	14	15	1.07	0.27	4	0.27
-elle	8	9	1.12	0.35	3	0.38
-er	16	16	1	0	2	0.50
-er/-ier	84	87	1.04	0.19	3	0.35
-ère/-ière	75	80	1.07	0.25	5	0.21
-eresse	3	3	1	0	1	1
-erie	243	363	1.49	0.66	22	0.07
-et	17	20	1.18	0.39	4	0.29
-ette	86	100	1.16	0.46	10	0.12
-eur (fém.)	2	2	1	0	1	1
-eur (masc.)	936	1'054	1.13	0.35	14	0.08
-eure	52	52	1	0	1	1
-euse	592	690	1.17	0.38	9	0.13
-in	8	8	1	0	3	0.33
-ine	3	4	1.33	0.58	2	0.67
-ing	27	37	1.37	0.56	5	0.27
-ion	796	901	1.13	0.38	35	0.03
-is	80	94	1.18	0.38	12	0.10
-ise	2	2	1	0	1	1
-ment	705	762	1.08	0.28	23	0.05
-oir/-oire	142	156	1.10	0.30	11	0.10
-oire	19	21	1.11	0.32	4	0.28
-on	65	69	1.06	0.24	12	0.09
-onne	8	8	1	0	2	0.50
-ot	7	7	1	0	2	0.50

Procédé	Vocables	Sens	Ambig. moyenne	ET	Types sémantiques	Pondération
-ote/-otte	7	7	1	0	3	0.33
-rice	102	105	1.03	0.17	4	0.26
-ure	202	291	1.44	0.75	21	0.07
Conv. - Thème 0	264	393	1.49	0.78	34	0.04
Conv. - Thème 12	151	181	1.20	0.43	23	0.05
Conv. - Thème 13	58	63	1.09	0.28	14	0.08
Conv. - Infinitif	35	43	1.23	0.43	7	0.18

TABLEAU 5.2 – Analyse des différents procédés morphologiques (classés par ordre alphabétique) en termes de nombre de vocables, nombre de sens, ambiguïté moyenne au niveau nominal (avec écart-type) et nombre de fonctions sémantiques. L'ambiguïté est calculée sans les sens figuratifs ($n = 113$) et uniquement pour les types réalisés au moins deux fois par un procédé. Les 5 valeurs les plus élevées sont à chaque fois surlignées en vert

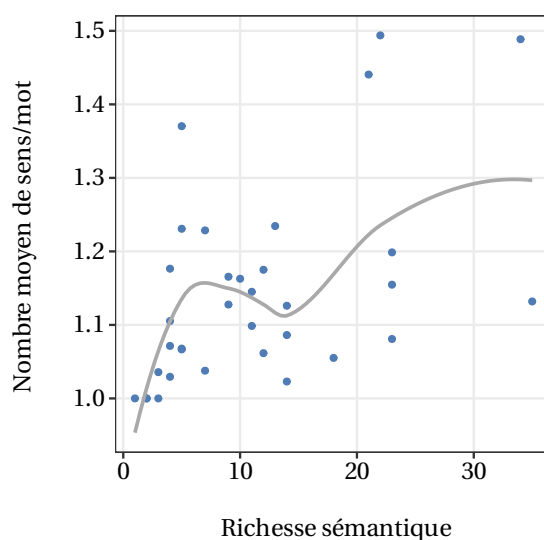


FIGURE 5.2 – Rapport entre richesse sémantique et nombre moyen de sens par vocable pour 34 procédés morphologiques. Une courbe de lissage obtenue par régression locale (LOESS) est ajoutée pour montrer la tendance générale

Ces premiers résultats nous conduisent à vouloir estimer la relation entre la richesse sémantique des différents procédés morphologiques (appréhendée ici à travers le nombre de types sémantiques combinés auxquels ils sont associés) et le taux d'ambiguïté moyen de leurs dérivés. En d'autres termes, nous cherchons à déterminer si plus un procédé morphologique est polyfonctionnel, plus les mots qu'il produit ont de sens. Nous effectuons pour ce faire un test de corrélation de Kendall en excluant les procédés morphologiques ($n = 12$) pour lesquels moins de 10 vocables sont identifiés et dont les taux d'ambiguïté sont de ce fait peu informatifs : -asse, -eur (fém.), -ise, -ine, -ail, -eresse, -ain, -ot, -ote, -elle, -in et -onne. Les résultats de ce test indiquent qu'il existe une corrélation statistiquement significative entre

5.1. Procédés morphologiques

richesse sémantique et taux d'ambiguïté nominale ($r_t = 0.36$, $p < 0.01$), suggérant que lorsque la première augmente, le second tend à augmenter également (voir figure 5.2). Certaines exceptions peuvent néanmoins être relevées. Le taux d'ambiguïté moyen des vocables formés avec *-ion* n'atteint par exemple que 1.13 sens par nom, alors même que ce suffixe réalise au moins deux fois 35 types sémantiques différents.

Sur un plan plus qualitatif, il paraît enfin pertinent de se demander comment les différents types sémantiques répertoriés dans la figure 5.1 se combinent dans des configurations d'ambiguïté suivant le suffixe ou la forme de conversion. Un examen manuel des données confirme que les procédés morphologiques peuvent contraindre, à des degrés divers, les associations sémantiques qu'instancient leurs dérivés ; aucun d'entre eux ne donne en effet lieu aux mêmes associations de catégories sémantiques dans les mêmes proportions. Puisqu'il serait de toute évidence fastidieux et peu informatif de vouloir recenser exhaustivement toutes les associations observées, nous proposons dans le tableau 5.3 une sélection d'associations se distinguant par leur fréquence élevée pour 15 suffixes et 3 formes de conversion.

Plusieurs éléments intéressants émergent de cet aperçu. Nos résultats corroborent d'une part certaines affinités déjà établies dans la littérature entre procédés morphologiques et associations de catégories sémantiques. Par exemple, plusieurs suffixes « événementiels » (*-age*, *-ion*, *-ment*, *-ure*) montrent une préférence marquée pour l'alternance événement/résultat à travers les variantes Événement-TRANSPOSITION/Artéfact-RÉSULTAT et Événement-TRANSPOSITION/Objet naturel-RÉSULTAT. D'autres associations peuvent aussi être relevées à un niveau plus spécifique : Événement-TRANSPOSITION/Artéfact-INSTRUMENT pour le suffixe *-age* (Ferret & Villoing, 2015) ; Entité animée-AGENT/Artéfact-INSTRUMENT pour le suffixe *-eur* (Huyghe & Tribout, 2015) ; ou encore Artéfact-INSTRUMENT/Artéfact-LIEU pour le suffixe *-oir* (Namer & Villoing, 2008). Les résultats révèlent d'autre part une série de préférences encore inconnues de la littérature – à notre connaissance du moins. Les suffixes *-ade* et *-aison* semblent par exemple avantager l'association Événement-TRANSPOSITION/Événement-Collectif-TRANSPOSITION, avec la nuance supplémentaire que le type Événement-Collectif-TRANSPOSITION correspond plus spécifiquement à des saisons ou des périodes de l'année dans le cas des dérivés en *-aison* :

- (136) a. *cueillaison* 'action de cueillir' / 'période de l'année durant laquelle on cueille'
b. *montaison* 'migration d'un saumon' / 'saison où les saumons remontent des cours d'eau'
c. *fauchaison* 'action de faucher' / 'saison où l'on fauche'

Il ressort par ailleurs de l'analyse que plusieurs suffixes associent volontiers des catégories de sens événementielles : Événement-TRANSPOSITION/Domaine-TRANSPOSITION dans le cas de la conversion sur thème 0, *-age*, *-ing* ; et Événement-TRANSPOSITION/Propriété-CAUSE dans le cas de *-ance* et *-erie*. Nous notons enfin deux formes de préférence dans le cadre spécifique de la conversion : le patron Événement-TRANSPOSITION/Quantité-ÉTENDUE pour la conversion sur thème 12, qui semble procéder de l'appétence plus générale de ce procédé morpholo-

Procédé morphologique	Association	Exemple	Nombre de vocables
-ade	Evt-TSP/Evt-Coll-TSP	<i>bousculade</i> 'action de bousculer' / 'remous de foule'	6
-age	Evt-TSP/Art-RES	<i>collage</i> 'action de coller' / 'composition artistique'	49
	Evt-TSP/Dom-TSP	<i>jonglage</i> 'action de jongler' / 'art de jongler'	19
	Evt-TSP/Art-INS	<i>drapage</i> 'action de draper' / 'drap'	12
-aison	Evt-TSP/Evt-Coll-TSP	<i>olivaïson</i> 'action de cueillir des olives' / 'saison où l'on cueille des olives'	6
-ance	Evt-TSP/Ppt-CAU	<i>abstinence</i> 'action de s'abstenir' / 'capacité à s'abstenir'	6
-ant	Art-CAU/Ppt-CAU	<i>croquant</i> 'biscuit qui croque sous la dent' / 'capacité à croquer sous la dent'	3
-erie	Cog*Evt-TSP/Ppt-CAU	<i>radoterie</i> 'discours radoté' / 'propension à radoter'	22
	Evt-TSP/Ppt-CAU	<i>lambinerie</i> 'fait de lambiner' / 'propension à lambiner'	19
	Dom-TSP/Art*Ist-AGT	<i>mégisserie</i> 'art de préparer les cuirs' / 'entreprise qui prépare des cuirs'	15
-eur (masc.)	Anm-AGT/Art-INS	<i>éplucheur</i> 'personne qui épluche' / 'objet servant à éplucher'	91
	Anm-AGT/Ist-AGT	<i>promoteur</i> 'personne qui promeut' / 'société qui promeut'	7
-euse	Anm-AGT/Art-INS	<i>brodeuse</i> 'personne qui brode' / 'machine à broder'	89
-ière	Anm-AGT/Art-INS	<i>facturière</i> 'personne qui facture' / 'machine pour la facturation'	4
-ing	Evt-TSP/Dom-TSP	<i>coaching</i> 'action de coacher' / 'activité de mentorat'	4
-ion	Evt-TSP/Art-RES	<i>macération</i> 'action de macérer' / 'préparation macérée'	8
-is	Evt-TSP/Nat-RES	<i>brûlis</i> 'action de défricher par le feu' / 'terrain brûlé'	5
-ment	Evt-TSP/Art-RES	<i>cuirassement</i> 'action de cuirasser' / 'cuirasse'	8
	Evt-TSP/Nat-RES	<i>vomissement</i> 'action de vomir' / 'matière vomie'	8
	Evt*Sta-TSP/Nat-RES	<i>enflement</i> 'action d'enfler' / 'enflure'	5
-oir	Art-LOC/Art-INS	<i>équarri-soir</i> 'endroit où l'on équarrit' / 'instrument pour équarrire'	9
-ure	Evt-TSP/Art-RES	<i>jaspure</i> 'action de jasper' / 'bigarrure'	11
	Evt-TSP/Nat-RES	<i>tonture</i> 'action de tondre' / 'déchet de matière tondue'	7
Conversion - Thème 0	Evt-TSP/Dom-TSP	<i>voltige</i> 'action de voltiger' / 'art de voltiger'	7
	Evt-TSP/Nat-RES	<i>labour</i> 'action de labourer' / 'terre labourée'	6
Conversion - Thème 12	Evt-TSP/Nat-RES	<i>percée</i> 'action de percer' / 'trou percé'	3
	Evt-TSP/Qua-EXT	<i>pesée</i> 'action de peser' / 'quantité pesée'	3
Conversion - Infinitif	Evt-TSP/Art-PAT	<i>goûter</i> 'action de faire une collation' / 'nourriture consommée en goûtant'	4

TABLEAU 5.3 – Associations entre types sémantiques les plus fréquentes pour une sélection de procédés morphologiques

5.1. Procédés morphologiques

gique pour les noms de quantité, construits sur base verbale ou nominale (p. ex. *bouchée*, *brassée*, *chiée*, *flopée*, *pincée*, *trifouillée*) ; et l'association Événement-TRANSPOSITION/Artéfact-PATIENT pour la conversion de forme infinitive (137), qui s'inscrit dans le champ sémantique très restreint des repas (cf. Barque et al., 2014).

- (137) a. *petit-déjeuner* 'action de petit-déjeuner' / 'nourriture consommée en petit-déjeunant'
b. *goûter* 'action de faire une collation' / 'nourriture consommée en goûtant'
c. *dîner* 'action de dîner' / 'nourriture consommée en dinant'

5.1.2 Étude de cas : les réseaux d'associations des procédés les plus polyfonctionnels

Les différents schémas d'association que nous venons de mettre en évidence sont certes intéressants, mais n'ont pour l'instant reçu aucune validation statistique. Dans le prolongement de la précédente partie, cette section est dévolue à une analyse plus formelle du rôle joué par les procédés morphologiques dans la construction de l'ambiguïté des noms déverbaux. Nous nous soumettons pour ce faire au même exercice exploratoire que dans le chapitre 4. Il s'agira premièrement de mettre au jour des règles d'association orientées et non aléatoires entre types sémantiques pour différents procédés morphologiques, puis de construire, à partir de ces ensembles de règles, des réseaux d'associations pouvant être contrastés et comparés.

Nous choisissons de concentrer notre analyse sur les procédés instanciant au moins deux fois plus de 20 types sémantiques différents, à savoir *-ion* ($n = 35$ types), la conversion sur thème 0 ($n = 34$), *-age* ($n = 23$), *-ment* ($n = 23$), la conversion sur thème 12 ($n = 23$), *-erie* ($n = 22$) et *-ure* ($n = 21$). Au-delà du fait qu'elle garantit une certaine richesse sémantique *a priori* fructueuse pour l'analyse, cette sélection se justifie à nos yeux pour plusieurs raisons.

Conversion sur thèmes 0 et 12 L'inclusion de formes de conversion permet de comparer directement sémantique de la suffixation et sémantique de la conversion, un sujet de recherche où persistent de nombreuses zones d'ombre malgré un nombre croissant d'études comparatives (Missud & Villoing, 2021 ; Barbu Mititelu et al., 2023 ; Valera, 2023 ; *inter alia*).

-age, -ion et -ment Ces trois suffixes sont régulièrement traités comme des « rivaux » dans la littérature en vertu du fait que leurs dérivés investissent des types sémantiques similaires (Dal et al., 2018 ; Wauquier, 2020 ; Huyghe et al., 2023 ; *inter alia*). Déceler des différences, même minimes, dans la réalisation ambiguë de leurs dérivés pourrait souligner la nécessité de prendre en compte l'ambiguïté lexicale dans les analyses portant sur la rivalité ou la similarité affixale.

-erie et -ure Ces deux suffixes ont fait l'objet d'un nombre de travaux limité en comparaison d'autres procédés nominalisateurs en français (mais voir Drapeau & Boulanger, 1982 ;

Temple, 1991, 1996; Lecomte, 1997; *inter alia*). Comme nous l'avons vu plus haut, ils sont pourtant associés à des niveaux élevés d'ambiguïté nominale laissant présager des découvertes intéressantes.

Ensembles de règles d'association

La première étape de notre étude de cas a consisté à détecter, pour chaque procédé morphologique sélectionné, des règles de forme $\{A\} \Rightarrow \{B\}$ s'entendant comme des associations orientées et non aléatoires entre types sémantiques. Les étapes suivies pour arriver à ce résultat sont sensiblement les mêmes que celles décrites pour la base de données générale dans le chapitre précédent.

Formatage des données Un fichier de « transactions » représentant les différents types sémantiques instanciés par des vocables a premièrement été créé pour chaque procédé morphologique. Les sens figuratifs ont été exclus d'office (p. ex. *pâtisserie*₂ 'moulage en stuc' obtenu par métaphore à partir de *pâtisserie*₁ 'préparation sucrée').

Extraction des règles d'association Des règles d'association ont ensuite été extraites de chaque fichier de transactions à l'aide de l'algorithme *Apriori* (Agrawal et al., 1996). Comme dans le chapitre 4, le seuil de confiance a pour ce faire été fixé à 10 %. Les seuils de support ont en revanche été modulés de façon à ce que les associations identifiées s'appliquent à 2 vocables au minimum pour chaque procédé morphologique, assurant ainsi un socle commun de régularité.

Nettoyage des règles Enfin, les règles redondantes par rapport à d'autres (i.e. celles comportant plus d'éléments et étant de qualité moindre) ou dont l'antécédent et le conséquent n'étaient pas corrélés positivement (i.e. celles ayant obtenu un lift inférieur ou égal à 1) ont été éliminées.

Nous proposons une vue d'ensemble des résultats de l'analyse d'association dans le tableau 5.4⁵⁹. Plusieurs résultats intéressants se dégagent de cette synthèse. Nous notons d'emblée une grande variabilité dans le nombre de règles retenues, allant de 4 pour la conversion sur thème 12 à 52 pour la conversion sur thème 0, *a priori* sans corrélation directe avec le nombre de vocables examinés. Presque trois fois plus de noms ont par exemple été analysés dans le cas du suffixe *-ment* ($n = 710$) que dans le cas du suffixe *-erie* ($n = 245$), mais ce dernier donne lieu à un ensemble de règles plus important ($n = 23$ vs $n = 10$). Un autre point d'intérêt concerne le nombre d'items inclus dans les règles des différents ensembles. Si la majorité des règles extraites, tous procédés morphologiques confondus, sont constituées d'un type sémantique en antécédent et d'un autre en conséquent, des schémas d'ambiguïté à 3 (conversion sur thème 0, *-age*, *-erie*, *-ure*), voire à 4 (*-ure*) éléments affluent également. Enfin, d'importantes variations peuvent être observées quant aux taux

59. Des listes interactives des règles retenues pour chaque procédé sont accessibles à l'adresse https://osf.io/dpr s6/?view_only=0cbb83b7036a4b608ec772386c2100f2.

	Vocables	Support initial (%)	Règles	Structure			Support (%)		Confiance (%)		Lift	
				Bipartite	Tripartite	Quadripartite	<i>M</i>	<i>ET</i>	<i>M</i>	<i>ET</i>	<i>M</i>	<i>ET</i>
<i>-age</i>	1'052	0.2	19	14	5	0	0.7	1.2	72.2	34.6	1.9	1.3
<i>-erie</i>	245	0.8	23	16	7	0	4.0	3.5	60.4	28.4	4.0	2.0
<i>-ion</i>	798	0.3	21	21	0	0	0.4	0.3	63.9	31.7	4.3	4.1
<i>-ment</i>	710	0.3	10	10	0	0	0.6	0.4	58.7	33.0	2.0	1.2
<i>-ure</i>	208	1.0	34	20	13	1	3.3	2.9	55.9	31.7	5.4	7.7
Conv. - Thème 0	268	0.7	52	39	13	0	1.1	0.8	45.7	33.8	5.3	6.0
Conv. - Thème 12	160	1.3	4	4	0	0	1.4	0.3	56.2	18.5	7.3	6.9

TABLEAU 5.4 – Synthèse de l'extraction des règles d'association pour une sélection de procédés morphologiques. Vocables = Nombre de vocables analysés. Support initial = Seuil de support appliqué lors de l'extraction initiale. Règles = Nombre de règles d'association retenues après nettoyage. Structure = Nombre de règles retenues présentant une structure bi-, tri- et quadripartite

de support, de confiance et de lift. Les règles identifiées pour le suffixe *-age* se distinguent par exemple par leur fiabilité (plus de 70 % en moyenne), tandis que celles extraites pour *-ure* et la conversion sur thèmes 0 et 12 se démarquent par leur lift élevé, malgré une grande variabilité.

Avant de passer à l'analyse de réseau, nous revenons ci-dessous sur quelques éléments particulièrement notables au sein de chaque ensemble de règles.

Conversion sur thème 0 La conversion sur thème 0 se distingue par la taille importante de son ensemble de règles. Souvent réciproques et tripartites, ces dernières ont ceci de remarquable qu'elles ne s'appliquent la plupart du temps qu'à 2 vocables ($n = 34/52$ règles). Trois associations intéressantes et relativement fréquentes peuvent toutefois être relevées :

- (138) {Objet naturel-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION}
labour 'terre labourée' / 'action de labourer'
 Support de 4.1 % ($n = 11$ vocables), confiance de 84.6 % et lift de 1.5
- (139) {Entité animée-Collectif-AGENT} \Rightarrow {Événement-TRANSPOSITION}
relève 'personnes qui remplacent' / 'action de remplacer'
 Support de 3.4 % ($n = 9$ vocables), confiance de 90 %, lift de 1.6
- (140) {Objet naturel-LIEU} \Rightarrow {Événement-TRANSPOSITION}
estive 'lieu où l'on estive' / 'action d'estiver'
 Support de 2.2 % ($n = 6$ vocables), confiance de 100 %, lift de 1.8

Conversion sur thème 12 La conversion sur thème 12 est le procédé morphologique donnant lieu au plus petit ensemble de règles ($n = 4$). Comme pour la conversion sur thème 0, celles-ci ne s'appliquent qu'à un nombre extrêmement limité de vocables (min. = 2, max. = 3). Nous relevons en particulier une association déjà mentionnée dans la section précédente :

- (141) {Objet naturel-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION}
percée 'trou percé' / 'action de percer'
 Support de 1.9 % ($n = 3$ vocables), confiance de 75 %, lift de 1.5

-age Les règles d'association extraites pour le suffixe *-age* favorisent largement les types Événement-TRANSPOSITION et Événement*État-TRANSPOSITION en conséquent ($n = 15/19$). Comme en témoigne l'association en (142), les antécédents de ces règles renvoient quant à eux volontiers à des objets concrets artéfactuels d'un point de vue ontologique ($n = 12/19$) et résultatifs sur un plan relationnel ($n = 10/19$).

- (142) {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION}
gribouillage 'dessin gribouillé' / 'action de gribouiller'
 Support de 5.7 % ($n = 60$ vocables), confiance de 93.8 %, lift de 1.7

5.1. Procédés morphologiques

Nous relevons également l'existence de 5 règles tripartites impliquant systématiquement les types Événement-TRANSPOSITION et Artéfact-RÉSULTAT, et mobilisant alternativement les types Propriété-RÉSULTAT, Artéfact-INSTRUMENT et Domaine-TRANSPOSITION comme troisième acteur. Parmi elles figure par exemple l'association suivante :

- (143) {Artéfact-INSTRUMENT, Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPOSITION}
bandage 'bande de tissu servant à bander' / 'pansement posé' / 'action de bander'
 Support de 0.5 % ($n = 5$ vocables), confiance de 100 %, lift de 1.4

-erie Les règles sélectionnées pour le suffixe *-erie* ont cela d'intéressant qu'elles se trouvent fréquemment accompagnées de leurs réciproques, avec des indices de confiance très proches voire égaux. Tel est le cas, par exemple, des associations présentées en (144) et (145).

- (144) {Propriété-CAUSE} \Rightarrow {Événement*Cognitif-TRANSPOSITION}
radoterie 'propension à radoter' / 'discours radoté'
 Support de 9.4 % ($n = 23$ vocables), confiance de 45.1 %, lift de 2
- (145) {Événement*Cognitif-TRANSPOSITION} \Rightarrow {Propriété-CAUSE}
flatterie 'fausse louange' / 'propension à flatter'
 Support de 9.4 % ($n = 23$ vocables), confiance de 41.1 %, lift de 2

Nous relevons également dans ce contexte une forte affinité entre les types Domaine-TRANSPOSITION, Artéfact*Institution-AGENT et Artéfact-RÉSULTAT, mobilisés à la fois dans des règles bipartites et tripartites, comme celles en (146) et (147).

- (146) {Domaine-TRANSPOSITION} \Rightarrow {Artéfact*Institution-AGENT}
fonderie 'commerce de la fonte' / 'usine qui fond du minerai'
 Support de 9.8 % ($n = 24$ vocables), confiance de 53.3 %, lift de 2.9
- (147) {Artéfact*Institution-AGENT, Artéfact-RÉSULTAT} \Rightarrow {Domaine-TRANSPOSITION}
menuiserie 'atelier qui travaille le bois' / 'ouvrage menuisé en bois' / 'travail du bois'
 Support de 3.3 % ($n = 8$ vocables), confiance de 100 %, lift de 5.4

-ion Toutes bipartites, les règles sélectionnées pour le suffixe *-ion* se caractérisent par une grande diversité ontologique en antécédent (p. ex. Objet cognitif, Artéfact, Institution, Objet financier) et le fait qu'elles sollicitent majoritairement le type Événement-TRANSPOSITION en conséquent. Nous notons également le nombre restreint de vocables concernés par chaque règle en comparaison du nombre élevé de noms retenus pour l'analyse ($n = 798$). Les deux règles les plus largement applicables ne concernent en effet que 8 items :

- (148) {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPPOSITION}
oblitération 'marque apposée sur un document' / 'action d'oblitérer un document'
 Support de 1 % ($n = 8$ vocables), confiance de 88.9 % et lift de 2.1
- (149) {Domaine-TRANSPPOSITION} \Rightarrow {Événement-TRANSPPOSITION}
gestion 'science de l'administration' / 'action de gérer'
 Support de 1 % ($n = 8$ vocables), confiance de 57.1 %, lift de 1.4

-ment Le suffixe *-ment* se caractérise par un nombre relativement restreint de règles ($n = 10$) en comparaison de ses « rivaux » *-age* ($n = 19$) et *-ion* ($n = 21$). Nous constatons que tous les conséquents sont constitués de types événementiels transpositionnels (Événement-TRANSPPOSITION, Événement*État-TRANSPPOSITION, Événement*Cognitif-TRANSPPOSITION et Événement*Phénomène-TRANSPPOSITION). Autre point intéressant, les deux règles s'appliquant au plus grand nombre de lexèmes impliquent en antécédent un type concret sur un plan ontologique et résultatif d'un point de vue relationnel :

- (150) {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPPOSITION}
griffonnement 'dessin' / 'action de griffonner'
 Support de 1.4 % ($n = 10$ vocables), confiance de 90.9 % et lift de 1.8
- (151) {Objet naturel-RÉSULTAT} \Rightarrow {Événement-TRANSPPOSITION}
écorchement 'partie écorchée' / 'action d'écorcher'
 Support de 1.1 % ($n = 8$ vocables), confiance de 57.1 % et lift de 1.1

-ure Enfin, le suffixe *-ure* est le seul procédé morphologique analysé pour lequel une règle quadripartite a été détectée (152). D'autres règles – tripartites et bipartites – incluant les mêmes types ont également été retenues, telles que celle présentée sous (153).

- (152) {Événement-TRANSPPOSITION, Domaine-TRANSPPOSITION, Artéfact-RÉSULTAT} \Rightarrow {Artéfact-INSTRUMENT}
argenteur 'action d'argenter' / 'art de l'argenteur-euse' / 'couche argentée' / 'produit servant à argenter'
 Support de 1.4 % ($n = 3$ vocables), confiance de 37.5 %, lift de 5.6
- (153) {Artéfact-RÉSULTAT} \Rightarrow {Événement-TRANSPPOSITION}
moulure 'objet moulé' / 'action de mouler'
 Support de 12.5 % ($n = 26$ vocables), confiance de 70.3 %, lift de 1.7

Réseaux individuels

Comme dans le chapitre précédent, nous poursuivons notre analyse en générant un réseau d'associations orienté pour chaque ensemble de règles. Le but est ici d'explorer plus avant, à travers l'examen de « cartographies » de relations, comment se structure l'ambiguïté

5.1. Procédés morphologiques

des nominalisations suivant le procédé morphologique avec lequel elles sont formées. Nous présentons les 7 graphes ainsi créés dans la figure 5.3 et synthétisons dans le tableau 5.5 leurs principales caractéristiques, à savoir : (i) le nombre de nœuds qu'ils comportent, correspondant aux types sémantiques mobilisés en antécédent ou en conséquent de règles ; (ii) le nombre total de liens entre nœuds, qui coïncident avec les implications entre types sémantiques au sein des règles⁶⁰ ; (iii) le nombre de liens uniques entre nœuds, obtenu en réduisant à un seul lien les liens multiples entre deux mêmes nœuds ; (iv) leur densité, c'est-à-dire le ratio entre le nombre de liens uniques observés et le nombre total de liens possibles entre les nœuds présents dans chaque graphe ; et (v) le nombre de composantes, soit le nombre de sous-graphes disjoints pouvant être observés dans chaque graphe.

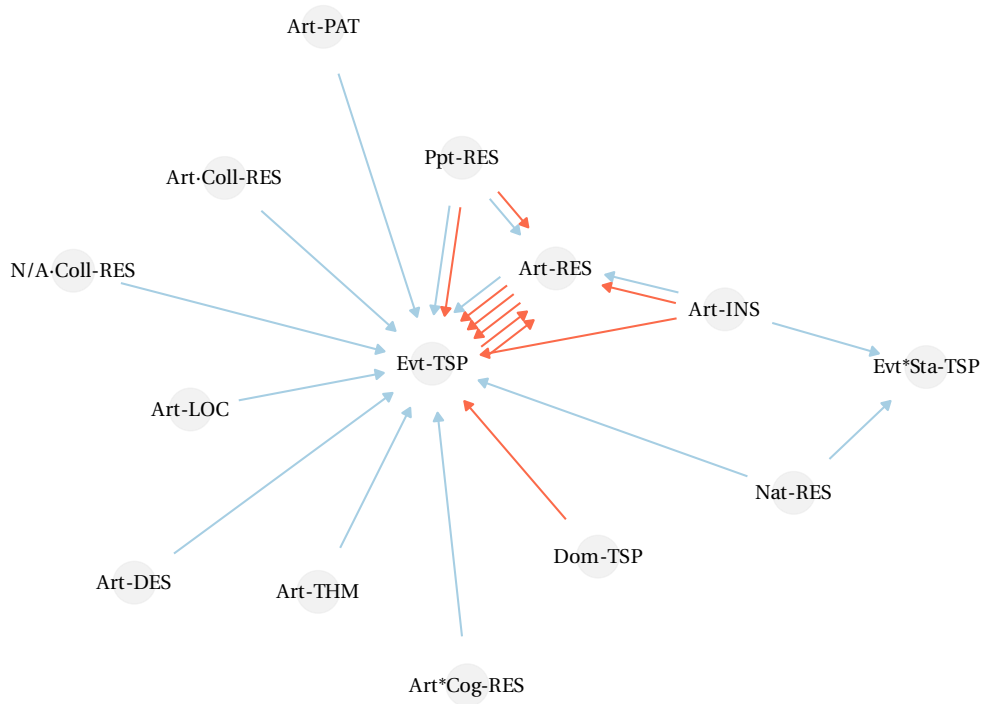
	Nœuds	Liens	Liens uniques	Densité	Composantes
<i>-age</i>	14	24	17	0.09	1
<i>-erie</i>	11	30	20	0.18	1
<i>-ion</i>	20	21	21	0.06	2
<i>-ment</i>	11	10	10	0.09	2
<i>-ure</i>	12	49	22	0.17	2
Conversion - Thème 0	20	65	49	0.13	1
Conversion - Thème 12	5	4	4	0.20	2

TABEAU 5.5 – Principales propriétés des réseaux constitués à partir des règles d'association extraites pour chaque procédé morphologique

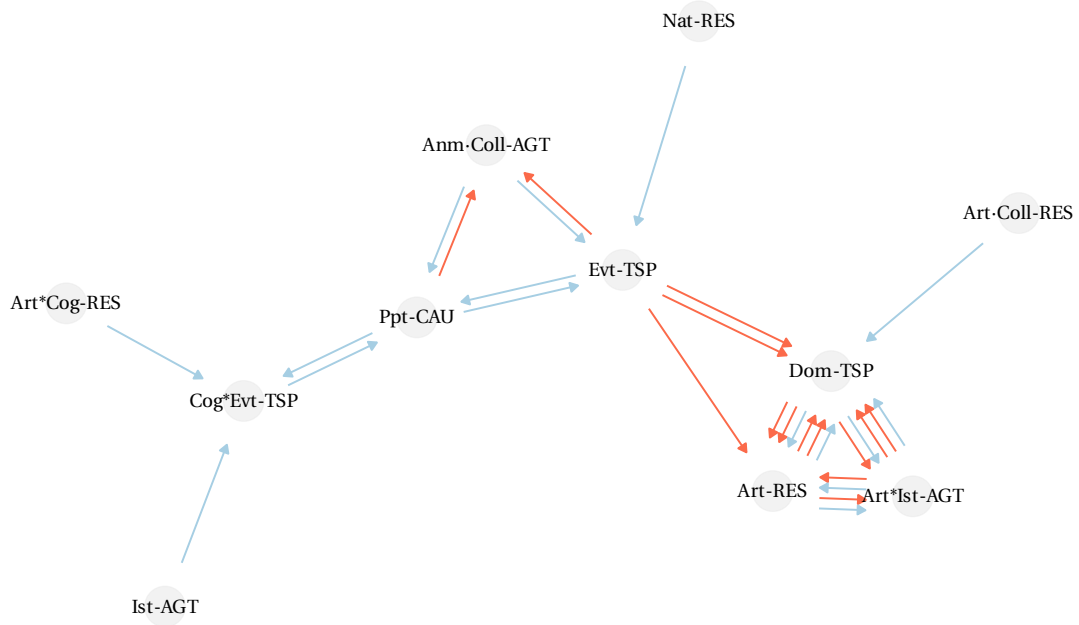
Plusieurs enseignements peuvent être tirés de cette synthèse initiale. Nous relevons tout d'abord d'importantes variations relativement au nombre de types sémantiques mobilisés, la taille des différents réseaux allant en effet de 5 à 20 nœuds. Les vocables ambigus obtenus par conversion sur thème 0 ou formés avec le suffixe *-ion* semblent à ce titre occuper des espaces sémantiques particulièrement riches. Un autre résultat intéressant concerne la densité relativement élevée des réseaux obtenus pour la conversion sur thèmes 0 et 12 ainsi que les suffixes *-erie* et *-ure*, qui contraste avec la plus faible connectivité des graphes générés pour les « rivaux » *-age*, *-ion* et *-ment*. Enfin, la présence de différentes composantes disjointes dans le cas de la conversion sur thème 12 et des suffixes *-ion*, *-ment* et *-ure* est potentiellement révélatrice de sous-groupes sémantiques distincts.

Certaines propriétés de ces réseaux peuvent être examinées plus en détail, à commencer par l'importance que revêtent les différents types sémantiques au sein de chacun d'eux. Le chapitre précédent a révélé que seules quelques catégories sémantiques jouaient un rôle crucial dans le réseau général d'ambiguïté, en particulier Événement-TRANSPOSITION,

60. Comme dans le chapitre 4, les implications au sein de règles incluant plus d'un élément en antécédent sont comptées à double (règles tripartites) ou à triple (règle quadripartite). Deux liens sont par exemple générés dans le cas de la règle {Artéfact-RÉSULTAT, Artéfact-INSTRUMENT} \Rightarrow {Événement-TRANSPOSITION} identifiée pour le suffixe *-ure* : l'un entre les types Artéfact-RÉSULTAT et Événement-TRANSPOSITION ; l'autre entre les types Artéfact-INSTRUMENT et Événement-TRANSPOSITION.



(a) Suffixe -age



(b) Suffixe -erie

FIGURE 5.3 – Réseaux d'associations obtenus pour les différents procédés morphologiques. Les liens relevant de règles bipartites, tripartites et quadripartites sont colorés respectivement en bleu, rouge et vert

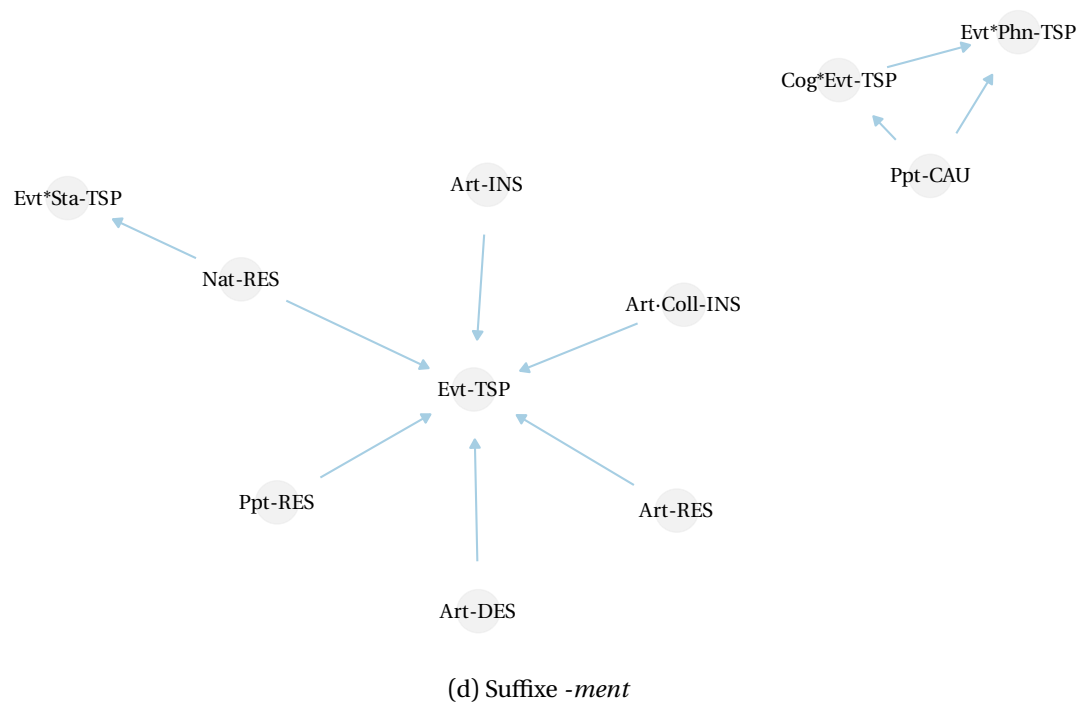
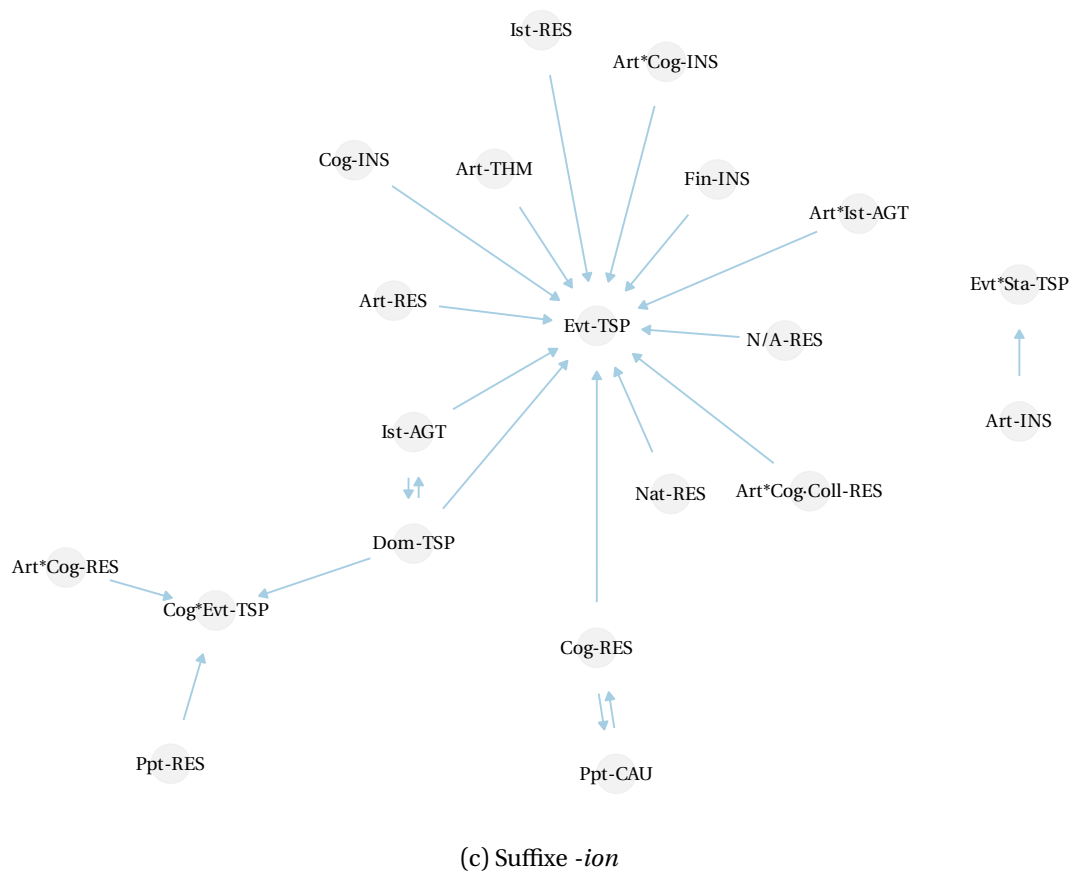
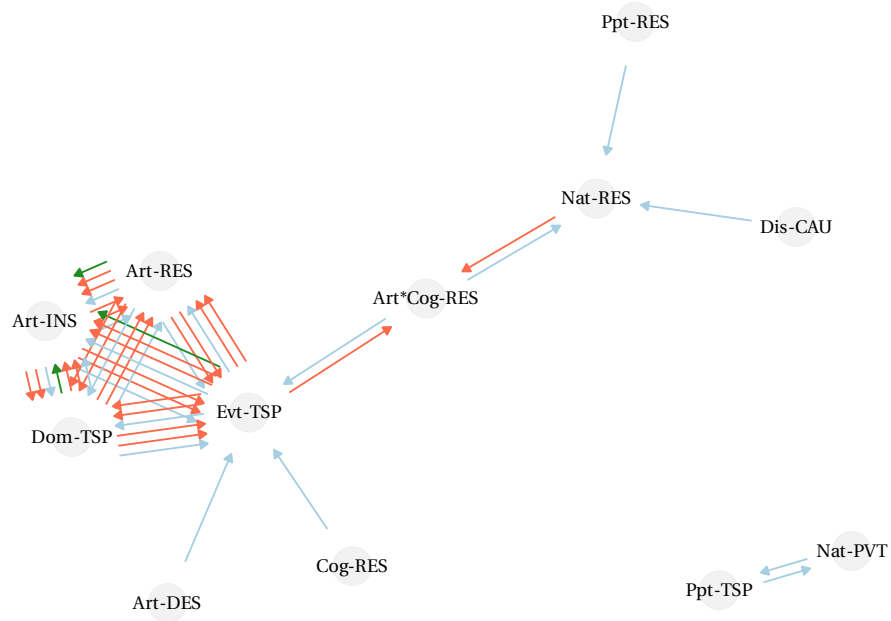
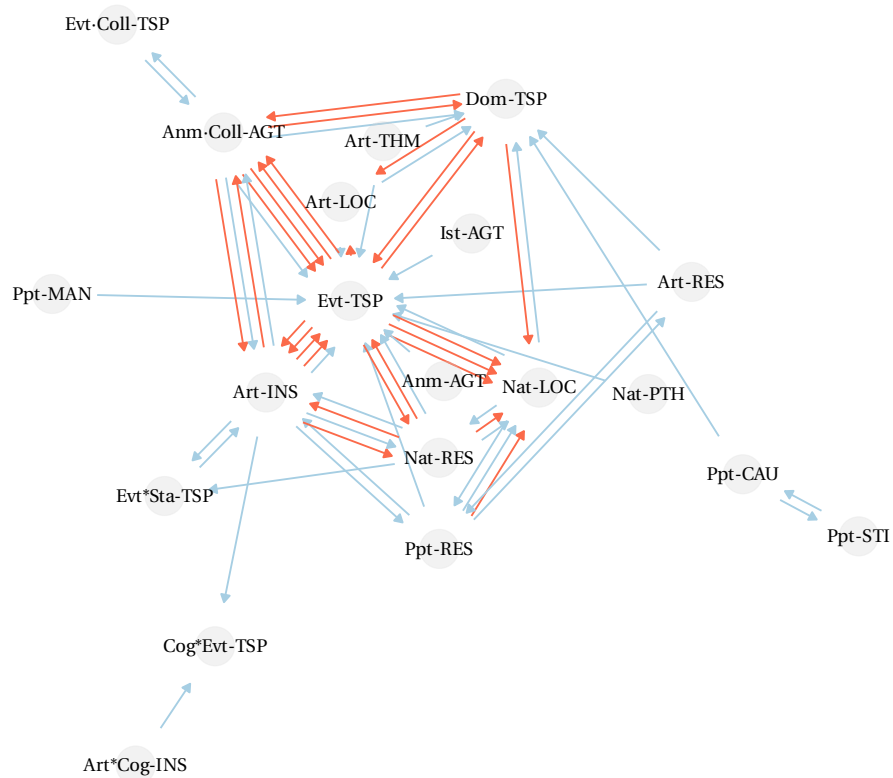


FIGURE 5.3 – Réseaux d’associations des différents procédés morphologiques (suite). Les liens relevant de règles bipartites, tripartites et quadripartites sont colorés respectivement en bleu, rouge et vert



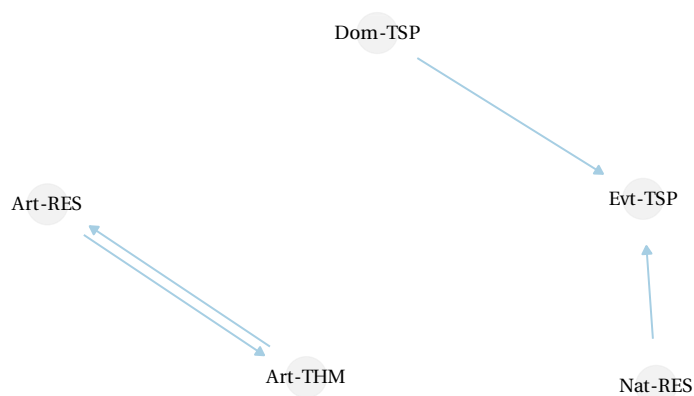
(e) Suffixe *-ure*



(f) Conversion sur thème 0

FIGURE 5.3 – Réseaux d'associations des différents procédés morphologiques (suite). Les liens relevant de règles bipartites, tripartites et quadripartites sont colorés respectivement en bleu, rouge et vert

5.1. Procédés morphologiques



(g) Conversion sur thème 12

FIGURE 5.3 – Réseaux d'associations des différents procédés morphologiques (suite). Les liens relevant de règles bipartites, tripartites et quadripartites sont colorés respectivement en bleu, rouge et vert

Artéfact-RÉSULTAT et Artéfact-INSTRUMENT. Il paraît à ce titre pertinent de se demander si cette observation varie en fonction du procédé morphologique. Pour répondre à cette question, nous avons calculé quatre scores de centralité reflétant différentes formes « d'importance » pour tous les nœuds de chaque graphe : (i) la centralité de degré total, pour identifier les types sémantiques recevant et/ou initiant le plus de liens ; (ii) la centralité de degré entrant, pour repérer les types recevant le plus de liens ; (iii) la centralité de degré sortant, pour distinguer les types initiant le plus de liens ; et (iv) la centralité de vecteur propre, pour identifier les types dont l'influence découle de leur connexion à d'autres nœuds importants.

Nous présentons dans le tableau 5.6 les types sémantiques ayant obtenu les scores de centralité les plus élevés pour chaque procédé morphologique⁶¹. Nos résultats soulignent dans l'ensemble le rôle capital joué par le type Événement-TRANSPOSITION dans l'ambiguïté des noms déverbaux, transcendant les différences formelles incarnées par les procédés morphologiques. Ce type événementiel obtient en effet le plus haut score de centralité de degré total et de centralité de vecteur propre dans tous les réseaux à l'exception de celui de la conversion sur thème 12, plus particulier en raison de sa taille réduite. Ces nouveaux résultats semblent donc confirmer la propension de cette catégorie de sens à former des connexions avec de nombreux autres types sémantiques (centralité de degré total), et notamment avec d'autres catégories de sens importantes (centralité de vecteur propre).

Pour le reste, les différents réseaux varient considérablement. Les trois rivaux *-age*, *-ion* et *-ment* ne s'accordent pas sur le type sémantique le plus susceptible d'initier des liens (Artéfact-INSTRUMENT pour *-age*, Objet naturel-RÉSULTAT et Propriété-CAUSE pour *-ment*,

61. La liste complète des scores est disponible dans le matériel supplémentaire de la thèse (https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2).

Domaine-TRANSPOSITION pour *-ion*), alors même qu'ils privilégient tous les trois le type Événement-TRANSPOSITION comme « récepteur » de liens. De plus, nous pouvons relever que les rapports entre catégories de sens semblent plus équilibrés dans le cas du suffixe *-ure* que dans celui de *-age*, *-ion* et *-ment*. S'agissant de *-ure*, ce sont en effet systématiquement les 4 mêmes types qui sont importants selon les 4 indices de centralité, et toujours dans le même ordre : Événement-TRANSPOSITION, suivi par Artéfact-INSTRUMENT, Artéfact-RÉSULTAT et Domaine-TRANSPOSITION au second rang.

Procédé	Degré total	Degré entrant	Degré sortant	Vecteur propre
<i>-age</i>	Evt-TSP	Evt-TSP	Art-INS	Evt-TSP
<i>-erie</i>	Evt-TSP	Dom-TSP	Evt-TSP	Evt-TSP
<i>-ion</i>	Evt-TSP	Evt-TSP	Dom-TSP	Evt-TSP
<i>-ment</i>	Evt-TSP	Evt-TSP	Nat-RES, Ppt-CAU	Evt-TSP
<i>-ure</i>	Evt-TSP	Evt-TSP	Evt-TSP	Evt-TSP
Conv. - Thème 0	Evt-TSP	Evt-TSP	Art-INS	Evt-TSP
Conv. - Thème 12	Evt-TSP, Art-RES, Art-THM	Evt-TSP	Art-RES, Art-THM, Dom-TSP, Nat-RES	Art-THM, Art-RES

TABLEAU 5.6 – Types sémantiques présentant les scores de centralité les plus élevés pour chaque procédé morphologique examiné (avec indication des ex aequo le cas échéant)

Une autre propriété qu'il peut être intéressant d'examiner est l'organisation structurelle des différents réseaux suivant l'importance accordée aux catégories sémantiques. Dans le chapitre précédent, nous avons en effet découvert que le réseau général d'associations se caractérisait par une structure hiérarchique extrêmement centralisée : en d'autres termes, qu'un petit nombre de types sémantiques accaparaient la plupart des liens (centralisation de degré) et que ces mêmes types tendaient à se connecter les uns aux autres (centralisation de vecteur propre). Comme pour l'importance des différents types sémantiques, on peut se demander si ces résultats varient selon les réseaux individuels.

Nous présentons dans les tableaux 5.7 et 5.8 les scores de centralisation de degré et de vecteur propre calculés pour chaque procédé morphologique, ainsi que les résultats de tests CUG basés sur 2'000 simulations et conditionnés au nombre de sommets et à la densité des réseaux. Pour rappel, ces tests permettent de déterminer si une caractéristique spécifique d'un graphe (ici, le degré de centralisation) est statistiquement distinctive (Butts, 2008; Cranmer et al., 2020). Il apparaît dans l'ensemble que la plupart des réseaux (conversion sur thème 0, *-age*, *-ion*, *-ment*, *-ure*) présentent des scores de centralisation de degré et de vecteur propre significativement plus élevés que dans le cas d'associations au hasard, bien que des variations puissent être relevées. Nous notons d'une part que les procédés morphologiques tendent davantage vers un graphe en étoile (centralisation de degré) que vers un graphe composé d'une seule relation et potentiellement de nombreux acteurs isolés (centralisation de vecteur propre), si l'on compare les valeurs observées et les scores de centralisation moyens

5.1. Procédés morphologiques

obtenus pour les graphes aléatoires. Un autre point intéressant concerne les différences entre procédés morphologiques. À en croire les scores obtenus pour les simulations, les suffixes *-age* et *-ion* sont associés à une structure radiale plus marquée que ce qui peut être observé pour les suffixes *-ure* et *-erie*.

	Val. obs.	Min.	Max.	<i>M</i>	Pr($X \geq \text{Obs}$)	Pr($X \leq \text{Obs}$)
<i>-age</i>	0.75	0.04	0.36	0.14	0	1
<i>-erie</i>	0.26	0.04	0.45	0.16	0.0535	0.9465
<i>-ion</i>	0.58	0.03	0.24	0.10	0	1
<i>-ment</i>	0.42	0.03	0.43	0.15	0.0005	0.9995
<i>-ure</i>	0.35	0.02	0.44	0.15	0.001	0.999
Conversion - Thème 0	0.52	0.04	0.30	0.12	0	1
Conversion - Thème 12	0.20	0	0.59	0.20	0.47	0.53

TABLEAU 5.7 – Centralisation de degré des différents procédés morphologiques et résultats de tests CUG basés sur 2'000 répétitions. Val. obs. = Score de centralisation observé. Min., Max., *M* = Valeurs minimum, maximum et moyenne obtenues pour les 2'000 réseaux aléatoires. Pr($X \geq \text{Obs}$) et Pr($X \leq \text{Obs}$) = Proportion de valeurs simulées respectivement supérieures ou égales, et inférieures ou égales à la valeur du réseau observé. Les résultats significatifs à $p < 0.05$ sont surlignés en vert clair

	Val. obs.	Min.	Max.	<i>M</i>	Pr($X \geq \text{Obs}$)	Pr($X \leq \text{Obs}$)
<i>-age</i>	0.75	0.22	0.78	0.49	0.0005	0.9995
<i>-erie</i>	0.64	0.17	0.78	0.45	0.015	0.985
<i>-ion</i>	0.83	0.32	0.83	0.58	0	1
<i>-ment</i>	0.81	0.22	0.90	0.52	0.002	0.998
<i>-ure</i>	0.75	0.17	0.78	0.49	0.002	0.998
Conversion - Thème 0	0.75	0.23	0.70	0.47	0	1
Conversion - Thème 12	0.86	0	1	0.51	0.053	0.9665

TABLEAU 5.8 – Centralisation de vecteur propre des différents procédés morphologiques et résultats de tests CUG basés sur 2'000 répétitions. Val. obs. = Score de centralisation observé. Min., Max., *M* = Valeurs minimum, maximum et moyenne obtenues pour les 2'000 réseaux aléatoires. Pr($X \geq \text{Obs}$) et Pr($X \leq \text{Obs}$) = Proportion de valeurs simulées respectivement supérieures ou égales, et inférieures ou égales à la valeur du réseau observé. Les résultats significatifs à $p < 0.05$ sont surlignés en vert clair

Nous terminons notre étude de cas par un examen des différentes formes de dyades – nulles ($\bullet \not\leftrightarrow \bullet$), asymétriques ($\bullet \rightarrow \bullet$) et mutuelles ($\bullet \longleftrightarrow \bullet$) – pouvant être observées au sein des 7 réseaux. Dans le chapitre précédent, nous avons en effet découvert que le réseau général d'associations sémantiques se caractérisait par un taux de réciprocité allant au-delà des associations au hasard. Mais cette observation se vérifie-t-elle pour tous les procédés morphologiques examinés? Pour répondre à cette question, nous présentons

dans le tableau 5.9 les taux de réciprocité observés pour les différents suffixes et formes de conversion, ainsi que les résultats de tests CUG basés sur 2'000 simulations et conditionnés au nombre de dyades nulles, asymétriques et mutuelles.

	Val. obs.	Min.	Max.	M	Pr(X ≥ Obs)	Pr(X ≤ Obs)
-age	0.06	0	0.31	0.05	0.5455	0.8365
-erie	0.54	0	0.43	0.10	0	1
-ion	0.11	0	0.24	0.03	0.101	0.987
-ment	0	0	0.43	0.05	1	0.644
-ure	0.69	0	0.38	0.09	0	1
Conversion - Thème 0	0.58	0	0.20	0.07	0	1
Conversion - Thème 12	0.33	0	1	0.11	0.2995	0.9915

TABLEAU 5.9 – Taux de réciprocité des différents procédés morphologiques et résultats de tests CUG basés sur 2'000 réplifications. Val. obs. = Score de centralisation observé. Min., Max., M = Valeurs minimum, maximum et moyenne obtenues pour les 2'000 réseaux aléatoires. Pr(X ≥ Obs) et Pr(X ≤ Obs) = Proportion de valeurs simulées respectivement supérieures ou égales, et inférieures ou égales à la valeur du réseau observé. Les résultats significatifs à $p < 0.05$ sont surlignés en vert clair

Nos résultats révèlent que seuls les suffixes *-erie* et *-ure*, ainsi que la conversion sur thème 0 présentent des taux de réciprocité supérieurs à ce qu'on pourrait attendre de la chance. Une explication possible aux résultats obtenus pour *-erie* et *-ure* pourrait être la présence de microparadigmes sémantiques structurés par des liens systématiquement bidirectionnels (cf. chapitre 4), rendant difficile l'identification d'une source d'extension sémantique unique. Les catégories sémantiques concernées seraient vraisemblablement Artéfact*Institution-AGENT, Artéfact-RÉSULTAT et Domaine-TRANSPOSITION pour *-erie* (154), et les types Événement-TRANSPOSITION, Domaine-TRANSPOSITION, Artéfact-RÉSULTAT et Artéfact-INSTRUMENT dans le cas du suffixe *-ure* (155) :

- (154) Artéfact*Institution-AGENT/Artéfact-RÉSULTAT/Domaine-TRANSPOSITION
menuiserie 'atelier qui travaille le bois'/'ouvrage menuisé en bois'/'travail du bois'
- (155) Événement-TRANSPOSITION/Domaine-TRANSPOSITION/Artéfact-RÉSULTAT/Artéfact-INSTRUMENT
argenture 'action d'argenter'/'art de l'argenteur-euse'/'couche argentée'/'produit servant à argenter'

Il paraît à cet égard intéressant de souligner que les autres suffixes, i.e. les « rivaux » *-age*, *-ment* et *-ion*, se caractérisent davantage par des relations sémantiques uniorientées, où le type Événement-TRANSPOSITION occupe une place prépondérante. Quant à la conversion sur thème 0, nous faisons l'hypothèse que le taux de réciprocité élevé observé pour ses relations

5.2. Propriétés sémantiques des verbes de base

témoigne de la diversité sémantique et de l'équilibre dans la polyvalence du procédé, aucun ensemble de règles ne semblant particulièrement sortir du lot.

Tous ces éléments nous amènent aux conclusions suivantes pour cette étude de cas :

-age, -ion, -ment Les divergences mises en évidence dans le cas des suffixes « rivaux » *-age*, *-ion* et *-ment* soulignent la nécessité de considérer l'ambiguïté des dérivés dans l'analyse des situations de compétition affixale. Il ressort en particulier que seuls les dérivés en *-age* peuvent s'inscrire dans des schémas d'association tripartites. Toutefois, force est de constater que *-age*, *-ion* et *-ment* génèrent tous les trois des réseaux d'associations avec une organisation hiérarchique similaire, formant une étoile dont le centre est le type Événement-TRANSPOSITION. L'absence significative de réciprocité au sein des trois réseaux suggère par ailleurs que ce type sémantique constitue probablement la principale source d'ambiguïté pour les noms formés avec ces suffixes.

-erie et -ure Nos résultats tendent à corroborer l'idée que les dérivés ambigus formés avec les suffixes *-erie* et *-ure* montrent un comportement sémantique particulier. Leurs réseaux d'associations se caractérisent par des taux de réciprocité élevés et significatifs, ce qui suggère des rapports plus équilibrés entre types sémantiques que ceux qui sont observés dans le cas des suffixes *-age*, *-ion* et *-ment*. Ce résultat étaye par ailleurs l'hypothèse selon laquelle les deux réseaux sont ponctués de microparadigmes ambigus incluant notamment les types Événement-TRANSPOSITION, Artéfact-RÉSULTAT et Domaine-TRANSPOSITION.

Conversion sur thèmes 0 et 12 Deux enseignements généraux relatifs à la sémantique de la conversion *vs* de la suffixation peuvent être tirés de cette étude. Nous relevons d'une part que la conversion sur thème 0 et la conversion sur thème 12 présentent un caractère très idiosyncrasique par rapport aux suffixes étudiés. D'autre part, les résultats montrent qu'il convient de distinguer différents types de conversion, la conversion sur thème 0 et la conversion sur thème 12 donnant en effet lieu à des réseaux très différents. La conversion sur thème 0, notamment, se démarque des autres procédés par la diversité et l'équilibre des types dans les dérivés ambigus. Son profil particulier semble aller dans le sens d'un constat fréquent dans la littérature sur la sémantique de la conversion, selon lequel les conversions peuvent être considérées comme sémantiquement plus diversifiées que les mots affixés (Plag, 1999, p. 221).

5.2 Propriétés sémantiques des verbes de base

Nous laissons momentanément de côté les procédés morphologiques pour mieux les retrouver dans la section 5.3, dédiée à la collaboration entre attributs. Dans l'intervalle, la présente section se concentre sur la seule implication des propriétés sémantiques des verbes de base dans la construction de l'ambiguïté des nominalisations. L'analyse portera d'abord sur différents traits d'aspect lexical, à savoir la dynamicité, la durativité, la télicité et la post-

phase (section 5.2.2), puis sur les rôles sémantiques des arguments, qu'il s'agisse du sujet, de l'objet ou de l'oblique (section 5.2.3). En préambule, la section 5.2.1 expose le fonctionnement de l'extraction de règles séquentielles⁶², le reste de ce travail reposant entièrement sur cette méthode d'analyse.

5.2.1 L'extraction de règles séquentielles

Notre objectif, dans ce qui suit, est d'identifier les propriétés sémantiques verbales et/ou les procédés morphologiques susceptibles de conduire certains dérivés à instancier des types ou des groupes de types sémantiques particuliers. Comme mentionné précédemment, nous nous intéressons aux propriétés aspectuelles et aux rôles sémantiques des arguments dans les sous-sections 5.2.2 et 5.2.3, ainsi qu'à l'interaction entre propriétés aspectuelles, rôles sémantiques et procédés morphologiques dans la section 5.3. Pour mener à bien cette recherche, une analyse d'association standard pourrait être à nouveau envisagée. Cependant, cette approche ne semble pas pleinement adaptée à nos besoins actuels. Les propriétés sémantiques des verbes et les types sémantiques des noms, par exemple, seraient traités de manière équivalente, ce qui nous amènerait potentiellement à générer un nombre important de règles non pertinentes, telles que {Verbe avec post-phase} \Rightarrow {Verbe télique}, n'incluant pas de types sémantiques nominaux. C'est pourquoi nous optons ici pour une approche différente mais relativement proche, qui permet de distinguer entre propriétés morphosémantiques et types sémantiques : l'analyse de motifs séquentiels.

Introduite par Agrawal et Srikant (1995), l'analyse de motifs séquentiels peut être *grosso modo* définie comme une extension de l'analyse de règles d'association intégrant une succession de conditions. Contrairement à l'analyse d'association « classique », qui vise à détecter des ensembles d'items présentant des affinités plus ou moins fortes, cette méthode permet en effet de modéliser des « enchaînements d'ensembles d'items » (Masseglia et al., 2004). Considérons à titre d'exemple une application au domaine commercial. Si l'analyse d'association standard sert à identifier quels produits tendent à apparaître conjointement dans des paniers d'achats (p. ex. couches-culottes, bières et cigarettes), l'analyse de motifs séquentiels permet quant à elle de découvrir l'ordre dans lequel des client-es effectuent des transactions (p. ex. couches-culottes le lundi, bières et cigarettes le mardi).

Les domaines d'application de l'analyse de motifs séquentiels sont nombreux, et s'étendent à tous les cas dans lesquels une relation de succession (*happens-after relationship*) peut être postulée (Zaki, 2001, p. 34). En sciences biomédicales, Wright et ses collègues (2015) ont par exemple identifié des relations temporelles entre les différents médicaments pris par des patient-es pour prédire de prochaines prescriptions. Nous modélisons pour notre part la construction des noms déverbaux selon une succession d'ordre logique mettant en jeu (i) les propriétés d'une base verbale et/ou un procédé morphologique pour déterminer (ii)

62. Toutes les extractions de règles séquentielles ont été réalisées dans R avec le package `arulesSequences` (Buchta et al., 2023) sur la base du tutoriel en ligne proposé par la Prof. Allison Koenecke (Koenecke, 2019).

5.2. Propriétés sémantiques des verbes de base

le sens d'un nom dérivé. En pratique, l'analyse ne consiste ainsi plus à détecter des associations non aléatoires entre types sémantiques (156) comme dans le chapitre 4 et le début du présent chapitre, mais à identifier des associations non aléatoires entre propriétés morphosémantiques et types sémantiques (157)-(158). Les éléments inclus dans les antécédents (propriétés morphosémantiques) et les conséquents (types sémantiques) des règles sont donc désormais contraints.

- (156) Type(s) sémantique(s) $X \Rightarrow$ Type sémantique Y
Support de [...] %, confiance de [...] %, lift de [...]
- (157) Propriété(s) morphosémantique(s) $X \Rightarrow$ Type(s) sémantique(s) Y
Support de [...] %, confiance de [...] %, lift de [...]
- (158) {Suffixe *-age*, Verbe télique, Verbe à sujet agent} \Rightarrow {Evt-TSP, Art-RES}
Support de 0.7 %, confiance de 12.9 %, lift de 7.4

Comme le montrent les schémas généraux en (156) et (157), ainsi que la règle en (158) obtenue lors de l'examen de l'ensemble de propriétés morphosémantiques (section 5.3), l'analyse d'association standard (156) et l'analyse séquentielle (157)-(158) donnent lieu à des outputs similaires : des règles orientées caractérisées par des niveaux de support, de confiance et de lift variables. À la différence de l'analyse d'association « classique », l'analyse de motifs séquentiels autorise toutefois la présence de plusieurs items (i.e. plus d'un type sémantique dans notre cas) dans les conséquents des règles, ce qui paraît ici particulièrement profitable. De plus, l'interprétation des niveaux de support, de confiance et de lift est légèrement différente, puisque l'analyse ne porte plus uniquement sur des dérivés nominaux, mais sur des paires verbe-nom :

Support Le support d'une règle séquentielle $\{A\} \Rightarrow \{B\}$ représente la proportion de paires verbe-nom instanciant l'antécédent $\{A\}$ (qui correspond à une ou plusieurs propriétés morphosémantiques) et le conséquent $\{B\}$ (qui équivaut à un ou plusieurs types sémantiques) :

$$\begin{aligned} support(A \Rightarrow B) &= P(A \cap B) \\ &= \frac{\text{Nombre de paires V-N réalisant A et B}}{\text{Nombre total de paires V-N}} \end{aligned}$$

Par exemple, 50 des 6'841 paires verbe-nom étudiées contiennent un verbe télique avec un sujet agentif et un dérivé construit avec le suffixe *-age* réalisant conjointement les types Événement-TRANSPOSITION et Artéfact-RÉSULTAT. Il s'ensuit que le support de la règle posée en (158) est de 0.7 % (50/6'841).

Confiance La confiance d'une règle séquentielle $\{A\} \Rightarrow \{B\}$ correspond à la proportion de paires verbe-nom réalisant l'antécédent $\{A\}$ qui instancient également le conséquent $\{B\}$:

$$\begin{aligned} \text{confiance}(A \Rightarrow B) &= P(B|A) = \frac{P(A \cap B)}{P(A)} \\ &= \frac{\text{Nombre de paires V-N réalisant A et B}}{\text{Nombre de paires V-N réalisant A}} \end{aligned}$$

À titre d'illustration, nous observons que 387 paires verbe-nom de la base de données comportent les propriétés morphosémantiques apparaissant dans l'antécédent de la règle sous (158). Par ailleurs, 50 de ces 387 paires incluent aussi les types mentionnés dans le conséquent de la règle. La confiance de cette dernière est donc de 12.9 % (50/387).

Lift Enfin, le lift compare la confiance d'une règle séquentielle $\{A\} \Rightarrow \{B\}$ à la probabilité d'obtenir son conséquent $\{B\}$ au hasard (i.e. sans considérer le procédé morphologique avec lequel les dérivés sont construits ou les propriétés sémantiques des bases verbales) :

$$\begin{aligned} \text{lift}(A \Rightarrow B) &= \frac{P(B|A)}{P(B)} = \frac{P(A \cap B)}{P(A)P(B)} \\ &= \frac{\text{Confiance de la règle}}{\text{Probabilité d'observer le conséquent}} \end{aligned}$$

Le lift de la règle sous (158) indique par exemple que son antécédent et son conséquent sont corrélés positivement : nous avons en effet 7.4 fois plus de chance de voir apparaître ensemble les éléments {Suffixe *-age*, Verbe téléique, Verbe avec sujet agentif} et {Événement-TRANSPPOSITION, Artéfact-RÉSULTAT} que s'ils étaient indépendants l'un de l'autre. Ce résultat s'obtient en divisant la confiance de la règle (12.9199 % exactement) par la proportion de paires verbe-nom comprenant un nom qui instancie conjointement les types Événement-TRANSPPOSITION et Artéfact-RÉSULTAT (1.754129 % des 6'841 paires verbe-nom examinées).

D'un point de vue technique, l'analyse de motifs fréquents fait intervenir des étapes similaires à celles de l'analyse d'association standard, avec en premier lieu la préparation des données pour l'extraction. À titre d'illustration, nous présentons dans le tableau 5.10 un extrait de notre base de données formaté pour une extraction de règles séquentielles basée sur l'ensemble des propriétés morphosémantiques : le procédé morphologique, les différents traits d'aspect lexical (dynamicité, durativité, téléicité et post-phase) et les rôles sémantiques du sujet, de l'objet et de l'oblique⁶³. Dans la terminologie de l'analyse de motifs séquentiels, les propriétés morphosémantiques et les types sémantiques sont considérés comme deux « événements » distincts. Les paires verbe-nom examinées sont quant à elles appelées « séquences ». Toutes les analyses présentées dans le reste de ce travail reposent sur

63. Comme pour toutes les analyses d'association dans ce travail, les types figuratifs (p. ex. Entité animée-FIGURATIF) ont été systématiquement exclus. Il en va de même pour les doublons de types sémantiques (p. ex. Événement-TRANSPPOSITION pour *monte*₁ 'accouplement' et *monte*₂ 'trot à cheval'), qui ont été réduits à une seule occurrence.

5.2. Propriétés sémantiques des verbes de base

6'841 paires verbe-nom, distinguées d'après l'appariement de sens nominaux à un même lexème-base (cf. chapitre 4). Deux séquences sont par exemple distinguées pour le nom *vol* :

- (159) a. *voler*₁ 'dérober' ⇒ *vol* 'action de dérober'
 b. *voler*₂ 'se déplacer' ⇒ *vol* 'action de se déplacer' / 'groupe d'animaux ou d'insectes qui se déplacent'

Séquence	Événement	Items	Nombre d'items
<i>gribouiller</i> ₁ ⇒ <i>gribouillage</i>	1	{Suffixe <i>-age</i> , V à sujet agent, V à objet résultat, V sans oblique, V dynamique, V duratif, V téléique, V sans post-phase}	8
	2	{Evt-TSP, Art-RES}	2
<i>kidnapper</i> ₁ ⇒ <i>kidnapping</i>	1	{Suffixe <i>-ing</i> , V à sujet agent, V à objet bénéficiaire, V sans oblique, V dynamique, V ponctuel, V téléique, V avec post-phase}	8
	2	{Evt*Sta-TSP}	1
<i>voler</i> ₁ ⇒ <i>vol</i>	1	{Conversion sur thème 0, V à sujet agent, V à objet thème, V sans oblique, V dynamique, V ponctuel, V téléique, V sans post-phase}	8
	2	{Evt-TSP}	1
<i>voler</i> ₂ ⇒ <i>vol</i>	1	{Conversion sur thème 0, V à sujet thème, V sans objet, V sans oblique, V dynamique, V duratif, V atélique, V sans post-phase}	8
	2	{Evt-TSP, Anm-Coll-AGT}	2

TABLEAU 5.10 – Extrait d'un tableau de données préparé pour l'extraction de règles séquentielles. Les Événements 1 correspondent aux ensembles de propriétés morphosémantiques examinées (procédés morphologiques, propriétés sémantiques des verbes de base); les Événements 2, aux types sémantiques observés dans le domaine nominal

Une fois les données converties au format requis, il s'agit essentiellement (i) d'extraire des motifs dont le support dépasse un certain seuil, qu'il s'agisse de propriétés morphosémantiques et/ou de types sémantiques, (ii) de détecter parmi ces motifs des règles dont la confiance dépasse un certain seuil, puis enfin (iii) de sélectionner les règles dont l'antécédent et le conséquent sont positivement corrélés (i.e. dont le lift est supérieur à 1) et qui ne sont pas redondantes. L'extraction de motifs fréquents à l'étape (i) a été réalisée à l'aide de l'algorithme cSPADE (Zaki, 2001) avec un support correspondant à 7 paires verbe-nom pour

les trois analyses présentées dans les sections 5.2.2, 5.2.3 et 5.3. Ce seuil a comme d'habitude été fixé par tâtonnement, de façon à limiter le nombre de règles générées tout en assurant une certaine variété sémantique. Le tableau 5.11 présente une sélection de motifs retenus pour l'analyse de l'ensemble des procédés morphosémantiques (section 5.3).

Motif	Proportion de paires V-N (%)	Paires V-N
{Anm-AGT}	28.0	1'916
{Suffixe <i>-ante</i> , V télique}	0.5	32
{Suffixe <i>-age</i> , V télique, V sans post-phase, V à sujet agent, V à objet patient}, {Evt-TSP, Art-RES}	0.5	35
{Suffixe <i>-ion</i> , V à télicité variable, V sans oblique}, {Evt*Sta-TSP}	3.5	239

TABLEAU 5.11 – Sélection de motifs fréquents extraits dans le cadre de l'analyse de l'ensemble des propriétés morphosémantiques (procédés morphologiques, aspect lexical et rôles sémantiques assignés par les verbes de base)

S'agissant de l'étape (ii), nous avons arbitrairement fixé le seuil de confiance à 1 % pour l'analyse des propriétés aspectuelles (section 5.2.2) et des rôles sémantiques (section 5.2.3), et à 10 % pour l'analyse conjointe des propriétés sémantiques des bases et des procédés morphologiques (section 5.3). Comme nous le verrons dans les deux sections suivantes, les associations entre types sémantiques sont en effet plus difficiles à prévoir lorsque seuls les traits d'aspect lexical ou les rôles des arguments des verbes sont considérés. Le tableau 5.12 synthétise un extrait des résultats obtenus pour l'analyse de l'ensemble des propriétés morphosémantiques (section 5.3).

Antécédent		Conséquent	Support (%)	Confiance (%)	Lift	Paires V-N
{Suffixe <i>-age</i> , V télique, V à objet patient}	⇒	{Evt-TSP, Art-RES}	0.6	18.6	10.6	42
{Suffixe <i>-ette</i> }	⇒	{Art-INS}	0.7	50.0	5.3	49
{V à télicité variable, V avec post-phase}	⇒	{Evt*Sta-TSP}	4.5	81.6	6.3	306
{Suffixe <i>-ment</i> , V duratif, V à sujet agent}	⇒	{Evt-TSP}	2.1	71.9	2.2	146

TABLEAU 5.12 – Extrait des règles séquentielles obtenues dans le cadre de l'analyse de l'ensemble des propriétés morphosémantiques (procédés morphologiques, aspect lexical et rôles sémantiques assignés par les verbes de base)

5.2.2 Aspect lexical

Nous cherchons dans cette section à évaluer l'impact que peut avoir l'aspect lexical des verbes de base sur le sémantisme de leurs dérivés, en particulier sur leur ambiguïté. La télicité, c'est-à-dire la capacité d'un verbe à décrire une éventualité présentant un point culminant, a déjà fait l'objet de commentaires à ce sujet par le passé (voir par ex. Osswald, 2005; Huyghe & Lombard, 2022). Il paraît toutefois pertinent d'examiner aussi le rôle joué par d'autres traits d'aspect lexical tels que la dynamicité, la durativité et la post-phase, ainsi que leur potentielle coopération. Le fait qu'un verbe télique présente une post-phase (p. ex. *(se) gonfler*) favorise-t-il par exemple l'émergence de certaines combinaisons de types sémantiques parmi ses dérivés (p. ex. Événement*État-TRANSPOSITION et Objet naturel-RÉSULTAT pour *gonflement* 'action de gonfler'/'boursofflure')? Nous proposons ci-dessous une analyse de règles séquentielles pour explorer cette problématique.

Extraction des règles séquentielles

La première étape de l'analyse de règles séquentielles a consisté à extraire des motifs composés de valeurs aspectuelles et/ou de types sémantiques instanciés par au moins 7 des 6'841 paires verbe-nom que compte la base de données (p. ex. *(se) gonfler*₁ / *gonflement*). Au total, 1'814 motifs « fréquents » ont été identifiés, avec un support moyen de 1.9 % ($ET = 6.8$), ce qui représente 132 paires verbe-nom. Nous avons ensuite identifié 749 règles d'association séquentielles « intéressantes »⁶⁴ parmi ces motifs fréquents. Le seuil d'extraction a arbitrairement été fixé à un niveau très bas, soit 1 %, afin de garantir la présence de plusieurs conséquents composés d'au moins deux types sémantiques. Avec un seuil de confiance à 10 % comme dans le chapitre 4 et dans la section 5.1.2 du présent chapitre, aucune règle comportant plus d'un type sémantique en conséquent n'était en effet générée et seul un nombre très limité de types sémantiques étaient représentés. Enfin, comme pour les précédentes analyses d'association, la première extraction a été nettoyée : 200 règles ont été immédiatement éliminées, car leur lift était inférieur ou égal à 1, et 334 règles supplémentaires ont ensuite été retirées parce qu'elles étaient redondantes.

Suite à ce tri, 215 règles séquentielles ont été conservées pour l'analyse des propriétés aspectuelles des verbes. Elles se caractérisent en moyenne par un support de 3.5 % ($ET = 6.1$, $n = 239$ paires verbe-nom), une confiance de 7.9 % ($ET = 13.2$), et un lift de 2.7 ($ET = 5.1$). La sélection finale est consultable en intégralité et dans un format interactif à l'adresse https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2.

64. Toutes les valeurs ne doivent pas nécessairement être renseignées pour identifier un motif. Certaines propriétés aspectuelles peuvent par conséquent être sous-déterminées dans les règles décrites.

Résultats et discussion

Nos résultats confirment dans l'ensemble que des valeurs ou des combinaisons de valeurs aspectuelles dans le domaine verbal sont positivement corrélées à des associations de types sémantiques dans le domaine nominal. Parmi les 215 règles retenues, 58 (environ 27 %) comportent en effet deux types sémantiques en conséquent, représentant 12 associations sémantiques différentes. Le tableau 5.13 reprend la règle la plus « intéressante » pour chacune de ces associations, sélectionnée en fonction des niveaux de lift et de confiance. Nous constatons que ces derniers restent au demeurant très bas. Le taux le plus élevé pour une règle à conséquent bipartite n'atteint en effet que 7.6 % (voir tableau 5.13). En comparaison, le taux de confiance le plus élevé dans le cas des règles à conséquent unique se monte à 81.6 % pour l'ensemble {Verbe à télicité variable, Verbe avec post-phase} \Rightarrow {Événement*État-TRANSPOSITION} ($n = 306$ vocables), c'est-à-dire lorsque le verbe correspond à un achèvement graduel gauche (cf. Dowty, 1979; Hay et al., 1999; Rothstein, 2008; *inter alia*).

Nous relevons dans tous les cas 3 configurations aspectuelles dans le domaine verbal présentant des affinités plus ou moins fortes avec des formes d'ambiguïté dans le domaine nominal.

Verbes à télicité variable Les verbes à télicité variable (i.e. les verbes d'achèvement graduel) tendent à favoriser deux formes d'ambiguïté : l'association entre les types Entité animée-AGENT et Artéfact-INSTRUMENT d'une part; et deux associations incluant le type complexe Événement*État-TRANSPOSITION d'autre part. La première configuration montre que les dérivés ont la possibilité de nominaliser conjointement l'agent et l'instrument œuvrant au changement d'état graduel décrit par le verbe. La seconde configuration indique quant à elle que les dérivés peuvent nominaliser non seulement le changement d'état gradable décrit par leur base (via le type Événement*État-TRANSPOSITION), mais aussi un instrument y contribuant (Artéfact-INSTRUMENT) ou sa conséquence concrète (Objet naturel-RÉSULTAT).

Verbes duratifs, téliques et sans post-phase Les verbes duratifs, téliques et sans post-phase (i.e. les verbes d'accomplissement selon la classification vendlérienne) sont associés au plus grand nombre de règles « intéressantes » ($n = 7$ dans le tableau 5.13), dont 4 comportent un type résultatif (Artéfact-RÉSULTAT ou Objet naturel-RÉSULTAT). L'hypothèse selon laquelle la télicité favorise la résultativité semble ainsi se vérifier, avec la nuance qu'elle doit être couplée à une notion de durée et à l'absence de post-phase pour être réellement opérationnelle. Nous notons en outre que la catégorie aspectuelle d'accomplissement semble favoriser l'émergence de patrons ne comportant pas nécessairement une composante résultative (p. ex. instrument, agent).

Verbes dynamiques et atéliques Les verbes dynamiques et atéliques (i.e. des verbes d'activité selon la classification vendlérienne) paraissent encourager l'association du type Propriété-CAUSE avec les catégories Événement-TRANSPOSITION et Événement*Cognitif-TRANSPOSITION. Cette tendance, qui n'a jamais été documentée auparavant à notre

Dynamicité	Durativité	Télicité	Post-phase		Types sémantiques	Support (%)	Confiance (%)	Lift	Paires V-N	Exemple
—	—	Variable	—	⇒	Evt*Sta-TSP, Art-INS	0.1	1.0	5.4	9	<i>lisser</i> ₁ ‘rendre lisse’ ⇒ <i>lissage</i> ‘action de lisser’ / ‘produit pour lisser’
—	—	Variable	Sans	⇒	Evt*Sta-TSP, Nat-RES	0.2	2.2	8.7	11	<i>mutiler</i> ₁ ‘altérer’ ⇒ <i> mutilation</i> ‘action d’altérer’ / ‘blessure’
—	—	Variable	Sans	⇒	Anm-AGT, Art-INS	0.6	7.6	2.6	39	<i>sécher</i> ₁ ‘rendre sec’ ⇒ <i>sécheuse</i> ‘personne qui sèche’ / ‘machine pour sécher’
—	Duratif	Télique	Sans	⇒	Evt-TSP, Art-RES	1.6	6.2	3.6	110	<i>ébaucher</i> ₁ ‘concevoir’ ⇒ <i>ébauche</i> ‘action d’ébaucher’ / ‘première version d’une œuvre’
—	Duratif	Télique	Sans	⇒	Evt-TSP, Nat-RES	0.4	1.6	1.8	29	<i>dégueuler</i> ₁ ‘vomir’ ⇒ <i>dégueulis</i> ‘action de dégueuler’ / ‘matière vomie’
—	Duratif	Télique	Sans	⇒	Evt-TSP, Art-INS	0.5	2.0	2.7	35	<i>s'autotester</i> ₁ ‘s’évaluer soi-même’ ⇒ <i>autotest</i> ‘action de s’autotester’ / ‘outil pour s’autotester’
—	Duratif	Télique	Sans	⇒	Evt-TSP, Dom-TSP	0.4	1.6	1.4	29	<i>arpenter</i> ₂ ‘mesurer la superficie’ ⇒ <i>arpentage</i> ‘action d’arpenter’ / ‘science de l’arpenteur-euse’
—	Duratif	Télique	Sans	⇒	Dom-TSP, Art*Ist-AGT	0.3	1.1	2.6	19	<i>distiller</i> ₁ ‘séparer par la chaleur’ ⇒ <i>distillerie</i> ‘art de distiller’ / ‘entreprise qui distille’
—	Duratif	Télique	Sans	⇒	Dom-TSP, Art-RES	0.3	1.2	3.6	22	<i>damasser</i> ₁ ‘fabriquer une étoffe’ ⇒ <i>damassure</i> ‘art de damasser’ / ‘tissu damassé’
—	Duratif	Télique	Sans	⇒	Art-INS, Art-RES	0.3	1.1	3.5	19	<i>mouler</i> ₁ ‘fabriquer’ ⇒ <i>mouleau</i> ‘moule’ / ‘objet moulé’
Dynamique	—	Atélique	—	⇒	Ppt-CAU, Cog*Evt-TSP	0.3	1.0	2.5	23	<i>flatter</i> ₁ ‘louer’ ⇒ <i>flatterie</i> ‘propension à flatter’ / ‘louange’
Dynamique	—	Atélique	—	⇒	Ppt-CAU, Evt-TSP	0.4	1.3	2.2	30	<i>minauder</i> ₁ ‘faire certaines mines’ ⇒ <i>minauderie</i> ‘propension à minauder’ / ‘action de minauder’

TABLEAU 5.13 – Sélection de règles extraites pour l’analyse des propriétés aspectuelles des bases verbales. Nous présentons la règle la plus intéressante (selon les niveaux de confiance et de lift) pour chacune des 12 combinaisons de types sémantiques de la liste finale

connaissance, constitue une exception notable à l'*Aspect Preservation Hypothesis* (Fábregas et al., 2012; voir aussi Haas et al., 2008; Balvet et al., 2011). En effet, des événements dynamiques dans le domaine verbal semblent ici pouvoir donner lieu à des associations ambiguës comportant un type statif et un type dynamique dans le domaine nominal.

5.2.3 Rôles sémantiques des arguments

Dans la continuité de nos premières observations sur l'aspect lexical, cette section examine l'influence des structures de rôles sémantiques associées aux verbes de base sur l'ambiguïté des noms dérivés, qu'il s'agisse de rôles fréquents (agent, patient, instrument) ou plus rares (étendue, bénéficiaire). Nous réalisons pour ce faire une deuxième analyse de règles séquentielles.

Extraction des règles séquentielles

Comme pour l'aspect lexical, la première étape de l'analyse a consisté à extraire des motifs fréquents (composés de rôles sémantiques attribués par les verbes de bases et/ou de types sémantiques nominaux) à partir des 6'841 paires verbe-nom contenues dans la base de données. Pour assurer une certaine cohérence avec ce qui a été présenté plus tôt, nous avons maintenu le même seuil de support que dans l'étude de l'aspect lexical, à savoir 7 paires verbe-nom. Au total, 1'279 motifs ont été extraits, avec un support moyen de 1.3 % ($SD = 4.4$, $n = 87$ paires verbe-nom). La deuxième étape de l'extraction a consisté à détecter des règles « intéressantes » parmi ces motifs fréquents. Toujours dans un souci de cohérence, le même seuil que dans l'analyse de l'aspect lexical a été appliqué. Ainsi, 741 règles se caractérisant par un support minimum de 1 % ont été générées. Nous avons enfin procédé au nettoyage du premier ensemble de règles, en éliminant d'abord celles dont l'antécédent et le conséquent ne présentaient pas de corrélation positive ($n = 166$), puis en supprimant les règles redondantes ($n = 152$).

La sélection finale comprend 423 règles, accessibles en intégralité et dans un format interactif à l'adresse suivante : https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2. Elles se caractérisent en moyenne par un support de 1 % ($ET = 2.4$), une confiance de 14.8 % ($ET = 17.5$) et un lift de 6.8 ($ET = 15.9$).

Résultats et discussion

Nos résultats confirment l'existence d'associations positives et non aléatoires entre les rôles sémantiques assignés par les verbes de base à leurs arguments et des combinaisons de types sémantiques dans le domaine nominal. 92 (22 %) des 423 règles de la sélection finale comportent en effet deux ou trois types sémantiques en conséquent, représentant au total 16 associations différentes. Le tableau 5.14 présente pour chacune de ces associations la règle la plus intéressante selon les niveaux de lift et de confiance. À l'instar de ce qui a été

5.2. Propriétés sémantiques des verbes de base

découvert pour l'aspect lexical, ces derniers restent relativement bas comparativement à ce qui peut être observé dans le cas des règles à conséquent unique. Le taux de confiance le plus élevé parmi les règles à conséquent multiple n'atteint en effet que 7.5 %. En comparaison, il y a 63.5 % de chance que le dérivé d'un verbe à sujet patientif, sans objet ni oblique, investisse le type sémantique complexe Événement*État-TRANSPOSITION (p. ex. *s'autodégrader*₁ / *autodégradation*).

Nous identifions malgré tout plusieurs « pôles » sémantiques verbaux (cf. Levin, 1993) pouvant potentiellement favoriser l'émergence de formes spécifiques d'ambiguïté nominale.

Verbes de changement d'état La moitié des 16 règles « intéressantes » présentées dans le tableau 5.14 incluent un participant patientif en antécédent, que celui-ci soit sujet ou objet. L'association {Entité animée-AGENT, Artéfact-INSTRUMENT} montre dans ce contexte que les noms déverbaux peuvent nominaliser à la fois l'entité animée autrice d'un changement d'état⁶⁵ et l'instrument avec lequel l'objet du verbe est affecté. On observe par ailleurs que l'affectation de l'objet peut conduire à l'interprétation résultative : plusieurs des règles les plus « intéressantes » comportent en effet un objet patientif en antécédent et un type résultatif (Propriété-RÉSULTAT, Artéfact-RÉSULTAT) en conséquent. L'association {Événement*État-TRANSPOSITION, Objet naturel-RÉSULTAT} indique en outre qu'à l'instar des objets patientifs, les sujets patientifs – caractéristiques des formes verbales anticausatives – peuvent aussi donner lieu à la nominalisation d'un type résultatif, en l'occurrence non imputable à une action humaine.

Verbes de création Parmi les 16 règles présentées dans le tableau 5.14, 3 comportent à la fois un objet résultatif en antécédent et un type sémantique résultatif (Artéfact-RÉSULTAT) accompagné d'un type événementiel (Événement-TRANSPOSITION, Domaine-TRANSPOSITION) en conséquent. Ce résultat confirme que les noms déverbaux ont la possibilité de nominaliser à la fois un résultat et l'éventualité qui en est à l'origine.

Verbes de mouvement 2 des 16 règles les plus « intéressantes » comprennent un thème (i.e. un objet localisé ou en mouvement) en antécédent. Si les associations correspondantes incluent toutes deux la catégorie sémantique Événement-TRANSPOSITION, la nature du second type impliqué dans l'association est sujette à variation : il peut s'agir soit de l'objet mis en mouvement dans le contexte de l'événement décrit par le verbe (Artéfact-THÈME), soit du résultat généré par un déplacement de matière (Objet naturel-RÉSULTAT).

Verbes de comportement et/ou de parole 2 des 16 règles les plus « intéressantes » relient des verbes de comportement ou de parole agentifs et intransitifs en antécédent au type Propriété-CAUSE en conséquent. Dans ce contexte, les dérivés ont la possibilité de nominaliser à la fois l'action dénotée par le verbe et la propension à l'accomplir.

65. Il est intéressant de remarquer que l'agentivité des noms déverbaux concernés par cette association contraste avec le sujet du verbe de base, qui renvoie à un agent causal n'étant pas forcément intentionnel. Autrement dit, le sujet du lexème-base peut simplement provoquer un changement d'état sans être l'acteur intentionnel de cette action, alors que l'agentivité du nom déverbal indique une participation active et intentionnelle à l'événement décrit.

Sujet	Objet	Oblique	Types sémantiques	Support %	Confiance %	Lift	Paires V-N	Exemple ⁶⁶
Patient	NA	NA	⇒ Evt*Sta-TSP, Nat-RES	0.1	2.3	9.1	9	<i>repousser</i> ₂ ‘pousser de nouveau’ ⇒ <i>repousse</i> ‘action de repousser’ / ‘nouvelle pousse d’un végétal’
Cause	Patient	NA	⇒ Anm-AGT, Art-INS	0.9	7.5	2.6	62	<i>blanchir</i> ₁ ‘rendre blanc’ ⇒ <i>blanchisseur</i> ‘personne qui blanchit’ / ‘produit pour blanchir’
Agent	Patient	—	⇒ Evt-TSP, Art-RES	1.0	6.7	3.8	66	<i>cuirasser</i> ₁ ‘revêtir d’une cuirasse’ ⇒ <i>cuirassement</i> ‘action de cuirasser’ / ‘cuirasse’
Agent	Patient	NA	⇒ Evt-TSP, Ppt-RES	0.2	1.1	3.9	11	<i>raffiner</i> ₁ ‘épurer’ ⇒ <i>raffinement</i> ‘action de raffiner’ / ‘caractère raffiné’
Agent	Patient	NA	⇒ Evt-TSP, Art-INS	0.3	1.9	2.5	18	<i>rembourrer</i> ₁ ‘garnir de bourre’ ⇒ <i>rembourrage</i> ‘action de rembourrer’ / ‘matière servant à rembourrer’
Agent	Patient	NA	⇒ Art-INS, Art-RES	0.2	1.5	5	15	<i>glacer</i> ₂ ‘revêtir d’un vernis’ ⇒ <i>glacis</i> ‘produit pour vernir’ / ‘couche de vernis’
Agent	Patient	NA	⇒ Evt-TSP, Art-INS, Art-RES	0.2	1.1	5.2	11	<i>frire</i> ₁ ‘cuire dans l’huile’ ⇒ <i>friture</i> ‘action de frire’ / ‘huile pour frire’ / ‘aliment frit’
Agent	Patient	NA	⇒ Dom-TSP, Art*Ist-AGT	0.2	1.2	3	12	<i>mégir</i> ₁ ‘tanner’ ⇒ <i>mégisserie</i> ‘action de mégir’ / ‘entreprise qui mégite’
Agent	Résultat	NA	⇒ Evt-TSP, Dom-TSP	0.2	3.2	2.8	13	<i>charpenter</i> ₁ ‘construire (une structure en bois)’ ⇒ <i>charpentage</i> ‘action de charpenter’ / ‘art de charpenter’
—	Résultat	NA	⇒ Evt-TSP, Dom-TSP, Art-RES	0.1	1.7	8.5	8	<i>graver</i> ₁ ‘tracer en creux’ ⇒ <i>gravure</i> ‘action de graver’ / ‘art de graver’ / ‘objet gravé’
—	Résultat	NA	⇒ Dom-TSP, Art-RES	0.1	2	5.6	9	<i>maçonner</i> ₁ ‘construire (un mur)’ ⇒ <i>maçonnerie</i> ‘art de maçonner’ / ‘objet maçonné’
Agent	NA	NA	⇒ Ppt-CAU, Evt-TSP	0.2	2.3	3.7	14	<i>lambiner</i> ₁ ‘agir avec lenteur’ ⇒ <i>lambinerie</i> ‘propension à lambiner’ / ‘action de lambiner’

66. Les noms cités comme exemples peuvent instancier d’autres types sémantiques que ceux mentionnés.

Sujet	Objet	Oblique		Types sémantiques	Support %	Confiance %	Lift	Paires V-N	Exemple ⁶⁶
Agent	NA	NA	⇒	Ppt-CAU, Cog*Evt-TSP	0.2	2.6	6.4	16	<i>bavasser</i> ₁ ‘bavarder’ ⇒ <i>bavasserie</i> ‘propension à bavarder’ / ‘bavardage’
Cause	Thème	NA	⇒	Evt-TSP, Nat-RES	0.2	5.9	6.7	12	<i>vomir</i> ₁ ‘dégueuler’ ⇒ <i>vomissement</i> ‘action de vomir’ / ‘matière vomie’
Agent	Thème	—	⇒	Evt-TSP, Art-THM	0.2	2.3	8.6	12	<i>importer</i> ₁ ‘introduire dans un pays’ ⇒ <i>importation</i> ‘action d’importer’ / ‘produit importé’
—	Bénéficiaire	NA	⇒	Evt-TSP, Ist-AGENT	0.2	1.9	7.6	14	<i>gouverner</i> ₂ ‘administrer’ ⇒ <i>gouvernement</i> ‘action de gouverner’ / ‘institution qui gouverne’

TABLEAU 5.14 – Sélection de règles extraites pour l’analyse des rôles des bases verbales. Nous présentons la règle la plus intéressante (selon les niveaux de confiance et de lift) pour chacune des 16 combinaisons de types sémantiques de la liste finale

66. Les noms cités comme exemples peuvent instancier d’autres types sémantiques que ceux mentionnés.

Verbes bénéfactifs Enfin, une dernière règle intéressante suggère que les noms déverbaux peuvent nominaliser à la fois l'éventualité dénotée par le verbe et une institution dont on bénéficie de l'autorité.

5.3 Collaboration entre attributs

On peut remarquer que plusieurs des associations identifiées dans les sections précédentes se recoupent. La combinaison {Événement-TRANSPOSITION, Artéfact-RÉSULTAT} présente par exemple des affinités avec la conversion sur thème 0 et les suffixes *-age*, *-ion*, *-ment* et *-ure* (section 5.1.2), mais aussi avec les verbes d'accomplissement (section 5.2.2) et les verbes à sujet agentif et objet patientif (section 5.2.3). Cette situation invite à réfléchir aux éventuelles interactions entre les différentes propriétés morphosémantiques examinées jusqu'à présent. Comme nous l'avons mentionné dans le chapitre 3, le fait qu'un verbe requière un sujet agentif implique par exemple qu'il soit dynamique, bien que la réciproque ne soit pas vraie (Filip, 2016, p. 19). De même, Missud (2019) a mis en évidence le fait que le suffixe *-age* marque une préférence pour les verbes d'accomplissement contrairement à la conversion de verbe à nom. Toutes ces observations nous conduisent à vouloir examiner pour cette fin de chapitre la contribution conjointe des traits d'aspect lexical, des rôles argumentaux des bases, ainsi que des procédés morphologiques dans la construction de l'ambiguïté des dérivés déverbaux. Nous en venons ainsi à une troisième et dernière étude de règles séquentielles.

5.3.1 Extraction des règles

Comme pour les analyses précédentes, la première étape de l'extraction a consisté à identifier des motifs instanciés par au moins 7 paires verbe-nom sur 6'841. Au total, 67'184 motifs fréquents ont été identifiés, avec un support moyen de 0.7 % ($ET = 2.2$), soit 47 noms. Nous avons dans un second temps détecté 29'013 règles « intéressantes » se dégageant de ces motifs fréquents. Pour ce faire, le seuil de confiance a été fixé à un niveau plus élevé que pour les deux analyses de règles séquentielles précédentes (i.e. 10 % *vs* 1 %), les règles étant en effet de meilleure qualité que celles extraites pour l'aspect lexical et les rôles sémantiques. Comme dernière étape, nous avons supprimé les règles dont l'antécédent et le conséquent n'étaient pas corrélés ($n = 1'836$), puis celles qui étaient redondantes ($n = 23'070$).

L'échantillon final contient 4'107 règles, accessibles en intégralité à l'adresse https://osf.io/dprs6/?view_only=0cbb83b7036a4b608ec772386c2100f2. Leur support est en moyenne de 1 % ($ET = 2.4$, $n = 67$ noms), leur confiance de 47.1 % ($ET = 28.6$ %) et leur lift de 8.5 ($ET = 13.9$).

5.3.2 Résultats et discussion

En accord avec ce que nous avons découvert jusqu'à présent, nos résultats montrent que des motifs d'ambiguïté dans le domaine nominal sont associés positivement à des procédés morphologiques et/ou à des propriétés sémantiques des bases. Parmi les 4'107 règles analysées, 435 (11 %) incluent en effet deux ou trois types sémantiques en conséquent. Les sous-sections suivantes examinent successivement deux groupes de questions soulevées par cette situation :

- (i) Quelles sont ces associations de catégories sémantiques? De quelle qualité sont les règles au sein desquelles elles apparaissent?
- (ii) Quelles sont les propriétés morphosémantiques qui conduisent à ces associations? Comment collaborent-elles? Des restrictions existent-elles quant aux procédés morphologiques, aux valeurs aspectuelles et aux rôles sémantiques officiant comme potentiels facteurs d'ambiguïté?

Associations sémantiques

Dans l'ensemble, seules 12 associations combinant 11 types sémantiques différents émergent des 435 règles à conséquent multiple. Le tableau 5.15 présente une synthèse des règles extraites pour chacune de ces associations et renseigne également sur leurs niveaux moyens de support, de confiance et de lift. Un premier élément intéressant se dégageant de ces résultats est que pratiquement toutes les associations entre types sémantiques exposées ici ont été identifiées dans le chapitre 4, consacré notamment aux mécanismes d'extension sémantique sous-tendant les schémas d'ambiguïté investis par les noms déverbaux. Pour rappel, il a été montré dans ce chapitre que la coprésence de certains types sémantiques parmi les nominalisations ambiguës était potentiellement motivée (en tout cas en partie) par des relations sémantiques de type métonymique. Ces nouveaux résultats semblent ainsi confirmer qu'extension sémantique et construction morphologique peuvent bel et bien œuvrer conjointement à la formation de noms ambigus dans le lexique déverbal.

Cela étant, il reste que certaines associations identifiées précédemment, comme {Événement-TRANSPPOSITION, Événement-Collectif-TRANSPPOSITION} (*bousculade* 'action de bousculer'/'remous de foule') ou encore {Événement-TRANSPPOSITION, Artéfact-Collectif-RÉSULTAT} (*empilage* 'action d'empiler'/'ensemble d'objets empilés'), ne figurent pas dans la sélection finale de cette section. Deux pistes explicatives peuvent être formulées pour justifier ces absences. Au-delà du fait que nous avons employé deux méthodes d'analyse d'association différentes (standard *vs* séquentielle), il paraît d'une part envisageable que des effets de fréquence affectent les résultats présentés dans le tableau 5.15 ci-dessous. En effet, les associations manquantes par rapport au chapitre 4 incluent souvent un type sémantique très peu fréquent (p. ex. Artéfact-Collectif-RÉSULTAT), ce qui peut entraîner des niveaux de confiance inférieurs à 10 %, soit le seuil minimum fixé pour l'extraction de cette partie. Une autre explication, plus intéressante d'un point de vue théorique, pourrait être que ces

associations manquantes se forment hors du giron des procédés morphologiques et des propriétés sémantiques des bases. Par exemple, il se peut que l'association {Événement*Cognitif-TRANSPOSITION, Artéfact*Cognitif-RÉSULTAT} (*dictée* 'action de dicter'/'transcription écrite') identifiée dans le chapitre précédent découle d'une forme de proximité conceptuelle basée sur le contenu informationnel véhiculé par les types sémantiques complexes plutôt que des procédés morphologiques ou des propriétés sémantiques des bases.

Association	Règles	Support (%)		Confiance (%)		Lift		Paires V-N	
		M	ET	M	ET	M	ET	M	ET
Anm-AGT, Art-INS	155	0.40	0.35	23.0	10.7	7.8	3.6	27	24
Art-INS, Art-LOC	4	0.12	0.00	12.6	0.5	61.4	2.6	8	0
Art-INS, Art-RES	4	0.13	0.00	11.6	1.6	37.7	5.1	9	0
Dom-TSP, Art*Ist-AGT	41	0.28	0.06	19.6	8.8	47.8	21.6	19	4
Dom-TSP, Art-RES	22	0.12	0.01	12.1	2.2	34.5	6.2	8	0
Dom-TSP, Evt-TSP	8	0.13	0.01	11.2	1.0	9.8	0.9	9	1
Evt-TSP, Art-INS	19	0.13	0.02	12.7	2.0	17.0	2.6	9	1
Evt-TSP, Art-INS, Art-RES	4	0.13	0.00	11.6	1.6	52.8	7.2	9	0
Evt-TSP, Art-RES	93	0.46	0.23	19.7	5.2	11.2	3.0	31	15
Evt-TSP, Nat-RES	9	0.15	0.01	10.7	0.5	12.0	0.5	10	0
Ppt-CAU, Evt-TSP	36	0.23	0.05	14.7	2.8	23.9	4.5	16	4
Ppt-CAU, Cog*Evt-TSP	40	0.27	0.04	17.1	3.5	41.7	8.6	18	3

TABLEAU 5.15 – Synthèse des règles extraites pour chaque association sémantique. Les valeurs les plus élevées sont surlignées en vert

Nous relevons par ailleurs que toutes les associations ne présentent pas les mêmes niveaux de support, de confiance et de lift. Autrement dit, il semble que certaines associations soient plus faciles à anticiper que d'autres à partir des propriétés sémantiques des bases et des procédés morphologiques. Les règles incluant les associations {Événement-TRANSPOSITION/Artéfact-RÉSULTAT} et {Entité animée-AGENT, Artéfact-INSTRUMENT} en conséquent se caractérisent par exemple par les taux moyens de support et de confiance les plus élevés, ce qui indique qu'elles s'appliquent non seulement à un nombre relativement important de paires verbe-nom, mais qu'elles sont aussi généralement fiables. Les règles dont le conséquent est l'association {Artéfact-INSTRUMENT, Artéfact-LIEU} constituent un autre cas notable, puisqu'elles se définissent par le lift moyen le plus élevé. La fréquence d'association entre ce motif sémantique et certaines propriétés morphosémantiques dépasse ainsi largement ce que l'on pourrait attendre du hasard.

Propriétés morphosémantiques

Ces premières observations nous conduisent à vouloir explorer de manière plus approfondie les propriétés morphosémantiques (procédés morphologiques, valeurs aspectuelles,

5.3. Collaboration entre attributs

rôles sémantiques) sous-jacentes aux 12 associations identifiées. À cette fin, nous présentons dans le tableau 5.16 la règle la plus « intéressante » en termes de confiance et de lift pour chaque association, ainsi qu’une synthèse qualitative des différentes propriétés morphosémantiques susceptibles de contribuer à chaque motif d’ambiguïté dans le tableau 5.17⁶⁷. Un premier résultat notable est que toutes les règles les plus « intéressantes » combinent plusieurs propriétés morphosémantiques en antécédent, ce qui suggère que la contribution morphologique à la construction de l’ambiguïté des déverbaux est essentiellement multifactorielle. L’association {Artéfact-INSTRUMENT, Artéfact-LIEU} illustre parfaitement ce point : si elle figure bien dans la sélection finale de cette section, elle est en revanche absente des analyses antérieures axées uniquement sur l’aspect lexical (section 5.2.2) et les rôles sémantiques (section 5.2.3). Cette association semble donc résulter d’une convergence de facteurs (suffixe *-oir*, caractère duratif du verbe, absence de post-phase, objet patientif) plutôt que de propriétés opérant de manière isolée.

Nous pouvons encore nuancer ces premiers constats, dans la mesure où tous les éléments présents en antécédent de règles ne contribuent pas de manière équivalente à la construction des formes déverbales ambiguës. Certaines propriétés (p. ex. procédé morphologique, rôle de l’objet) et certaines valeurs (p. ex. suffixe *-erie*, objet patientif) semblent exercer une influence plus marquée que d’autres. En particulier, nous observons que toutes les règles du tableau 5.16 incluent un procédé morphologique en antécédent, que toutes sauf une mentionnent une valeur pour le trait aspectuel de télicité, et que toutes sauf deux spécifient une valeur pour l’objet. Ces observations nous conduisent à formuler l’hypothèse que ces trois propriétés sont les plus déterminantes dans la construction de l’ambiguïté des déverbaux, en accord avec ce qui a déjà été mis en évidence dans la littérature (p. ex. Osswald, 2005 ; Melloni, 2011 ; Huyghe & Lombard, 2022). La dynamicité, la post-phase et le rôle de l’oblique semblent *a contrario* plus négligeables. Quant aux valeurs spécifiques, nos résultats révèlent que seule une fraction des procédés morphologiques, des valeurs aspectuelles et des rôles sémantiques contribue de manière significative à la construction de l’ambiguïté des déverbaux. Le comportement particulier du suffixe *-erie* et dans une moindre mesure du suffixe *-ure*, qui éclipsent des procédés morphologiques souvent plus étudiés, mérite encore une fois d’être souligné (cf. section 5.1.2 de ce chapitre). Par exemple, bien que l’association {Événement-TRANSPOSITION, Artéfact-INSTRUMENT} soit associée positivement au suffixe *-age* dans la sélection finale de règles (voir le tableau 5.17), le niveau de confiance le plus élevé pour une règle incluant *-age* en antécédent atteint 13 % pour cette association, contre 17 % dans le cas de *-erie*. Dans le domaine verbal, seules certaines propriétés sémantiques semblent aussi jouer un rôle décisif : la durativité, la télicité et l’absence de post-phase pour l’aspect lexical ; et les sujets agentifs et les objets patientifs pour les rôles sémantiques.

67. Pour rappel, toutes les valeurs n’ont pas à être renseignées pour identifier un motif.

Procédé	Dyn.	Dur.	Tél.	P.-p.	Suj.	Obj.	Obl.	Types sémantiques	Supp. (%)	Conf. (%)	Lift	Paires	Exemple ⁶⁸
<i>-euse</i>	—	—	Var.	Sans	CAU	PAT	NA	⇒ Anm-AGT, Art-INS	0.2	54.2	18.4	13	<i>polir</i> ₁ ‘rendre lisse’ ⇒ <i>polisseuse</i> ‘ouvrière qui polit’ / ‘machine pour polir’
<i>-oir</i>	—	Dur.	—	Sans	—	PAT	—	⇒ Art-INS, Art-LOC	0.1	13.1	64.1	8	<i>sécher</i> ₁ ‘rendre sec’ ⇒ <i>séchoir</i> ‘appareil pour sécher’ / ‘lieu où l’on sèche’
<i>-ure</i>	—	Dur.	Tél.	—	—	PAT	—	⇒ Art-INS, Art-RES	0.1	13.8	45.1	9	<i>doubler</i> ₂ ‘garnir l’intérieur’ ⇒ <i>doublure</i> ‘étoffe servant à doubler’ / ‘surface qui a été doublée’
<i>-erie</i>	—	Dur.	Tél.	—	—	PAT	—	⇒ Dom-TSP, Art*Ist-AGT	0.2	36.4	88.8	12	<i>distiller</i> ₁ ‘séparer par la chaleur’ ⇒ <i>distillerie</i> ‘art de distiller’ / ‘entreprise’
<i>-erie</i>	—	Dur.	Tél.	—	—	—	NA	⇒ Dom-TSP, Art-RES	0.1	16.4	46.6	9	<i>maçonner</i> ₁ ‘construire (un mur)’ ⇒ <i>maçonnerie</i> ‘art de maçonner’ / ‘objet maçonné’
<i>-ure</i>	—	Dur.	Tél.	—	AGT	—	—	⇒ Evt-TSP, Dom-TSP	0.1	13.4	11.8	9	<i>peindre</i> ₁ ‘figurer avec des couleurs’ ⇒ <i>peinture</i> ‘action de peindre’ / ‘art de peindre’
<i>-ure</i>	—	Dur.	Tél.	—	—	PAT	—	⇒ Evt-TSP, Art-INS, Art-RES	0.1	13.8	63.1	9	<i>teindre</i> ₁ ‘colorer’ ⇒ <i>teinture</i> ‘action de teindre’ / ‘produit pour teindre’ / ‘coloration obtenue’
<i>-ure</i>	—	Dur.	Tél.	—	AGT	PAT	—	⇒ Evt-TSP, Art-INS	0.1	17.0	22.8	8	<i>frir</i> ₁ ‘cuire dans l’huile’ ⇒ <i>friture</i> ‘action de frir’ / ‘huile pour frir’
<i>-ure</i>	—	Dur.	Tél.	—	AGT	PAT	—	⇒ Evt-TSP, Art-RES	0.2	31.9	18.2	15	<i>relier</i> ₁ ‘couvrir un ouvrage d’une matière rigide’ ⇒ <i>reliure</i> ‘action de couvrir un ouvrage’ / ‘couverture d’un livre relié’
<i>-ure</i>	—	—	Tél.	Sans	—	PAT	NA	⇒ Evt-TSP, Nat-RES	0.1	11.5	12.9	10	<i>tondre</i> ₁ ‘couper à ras’ ⇒ <i>tonture</i> ‘action de tondre’ / ‘déchet de matière tondue’
<i>-erie</i>	—	—	Atél.	—	AGT	NA	NA	⇒ Ppt-CAU, Evt-TSP	0.2	20.3	33.1	13	<i>minauder</i> ₁ ‘faire certaines mines’ ⇒ <i>minauderie</i> ‘propension à minauder’ / ‘action de minauder’
<i>-erie</i>	—	—	Atél.	—	AGT	NA	NA	⇒ Ppt-CAU, Cog*Evt-TSP	0.2	23.4	57.3	15	<i>bavasser</i> ₁ ‘bavarder’ ⇒ <i>bavasserie</i> ‘propension à bavarder’ / ‘bavardage’

TABLEAU 5.16 – Sélection de règles extraites pour l’analyse des procédés morphologiques, des traits d’aspect lexical et des rôles des arguments. Nous présentons la règle la plus intéressante (selon les niveaux de confiance et de lift) pour chacune des 12 combinaisons de types sémantiques de la liste finale

68. Les noms cités comme exemples peuvent instancier d’autres types sémantiques que ceux mentionnés.

Association	Procédé	Dynamicité	Durativité	Télicité	Post-phase.	Sujet	Objet	Oblique
Anm-AGT, Art-INS	<i>-eur</i> (masc.), <i>-euse</i>	Dynamique	Duratif	Télique, Variable, Atélique	Sans, Avec	Cause, Agent	Patient, Résultat, Thème, Topique	NA, Destination
Art-INS, Art-LOC	<i>-oir</i>	—	Duratif	—	Sans	—	Patient	—
Art-INS, Art-RES	<i>-ure</i>	—	Duratif	Télique	Sans	—	Patient	—
Dom-TSP, Art*Ist-AGT	<i>-erie</i>	Dynamique	Duratif	Télique	Sans	Agent	Patient	NA
Dom-TSP, Art-RES	<i>-erie, -ure</i>	—	Duratif	Télique	Sans	Agent	Patient	NA
Evt-TSP, Dom-TSP	<i>-ure</i> , conv. th. 0	—	Duratif	Télique	Sans	Agent	NA	NA
Evt-TSP, Art-INS	<i>-ure, -age</i>	—	Duratif	Télique	Sans	Agent	Patient, Bénéficiaire	NA
Evt-TSP, Art-INS, Art-RES	<i>-ure</i>	—	Duratif	Télique	Sans	—	Patient	—
Evt-TSP, Art-RES	<i>-ure, -age</i>	Dynamique	Duratif	Télique	Sans	Agent	Patient, Résultat	NA
Evt-TSP, Nat-RES	<i>-ure</i>	Dynamique	Duratif	Télique	Sans	Cause	Patient, Thème	NA
Ppt-CAU, Evt-TSP	<i>-erie</i>	Dynamique	Duratif	Atélique	Sans	Agent	NA	NA
Ppt-CAU, Cog*Evt-TSP	<i>-erie</i>	Dynamique	Duratif	Atélique	Sans	Agent	NA	NA

TABLEAU 5.17 – Propriétés morphosémantiques favorisant chaque association de types sémantiques

5.4 Bilan

Ce chapitre avait pour objectif d'évaluer dans quelle mesure et de quelle façon la construction morphologique conditionne l'ambiguïté des noms déverbaux. Nous retenons les éléments suivants au terme de notre analyse.

Procédés morphologiques La polyfonctionnalité des procédés et l'ambiguïté des mots construits constituent des phénomènes distincts qui n'en entretiennent pas moins des rapports étroits : tendanciellement, plus un procédé de nominalisation est polyfonctionnel, plus les dérivés formés avec celui-ci sont ambigus.

Les procédés morphologiques les plus polyfonctionnels peuvent donner lieu à différentes formes d'organisation systémique dans le lexique ambigu déverbal : structure radiale marquée et composée de liens majoritairement unidirectionnels pointant vers le type Événement-TRANSPOSITION (-age, -ion, -ment); présence de microparadigmes sémantiques composés de 3 ou 4 catégories de sens entretenant des relations réciproques fortes (-erie, -ure); structure éclatée (conversion sur thème 0). Ces différentes tendances ont des implications pour l'étude de la rivalité ou de la similarité affixale, mais aussi pour celle du rapport entre sémantique de l'affixation et sémantique de la conversion.

Propriétés sémantiques des bases Deux analyses de règles d'association séquentielles ont montré que des schémas d'ambiguïté nominale (p. ex. Événement-TRANSPOSITION/Artéfact-RÉSULTAT) sont positivement corrélés avec des valeurs aspectuelles et des rôles sémantiques dans le domaine verbal. La dynamacité, la durativité, la télélicité et l'absence de post-phase semblent particulièrement influentes dans le cas de l'aspect lexical. L'importance de l'agentivité des sujets et de la patientivité des objets peut être mise en évidence s'agissant des rôles sémantiques.

Collaboration entre attributs Une troisième analyse de règles d'association séquentielles portant sur la collaboration entre procédés morphologiques, valeurs aspectuelles et rôles sémantiques a révélé que ces différentes propriétés morphosémantiques œuvrent plus efficacement ensemble que seules à la construction de noms déverbaux ambigus. Néanmoins, toutes ne sont pas de la même importance : la contribution des procédés morphologiques semble plus déterminante que celle des rôles sémantiques et de l'aspect lexical.

Enfin, nous avons constaté que la quasi-totalité des associations sémantiques corrélées positivement avec des procédés morphologiques, des valeurs aspectuelles et des rôles sémantiques avait déjà été identifiée dans le chapitre précédent. Pour rappel, celui-ci a montré que la coprésence de certains types sémantiques parmi les nominalisations ambiguës était potentiellement motivée (en tout cas en partie) par des relations sémantiques de type métonymique. Nos résultats tendent ainsi à confirmer que l'extension sémantique et la construction morphologique peuvent contribuer conjointement à la formation de noms ambigus dans le lexique déverbal.

Conclusion

Deux objectifs principaux ont guidé cette étude :

- (i) proposer une description étendue et systématique de l'ambiguïté des noms déverbaux du français en synchronie ;
- (ii) examiner la contribution respective et la combinaison des procédés d'extension sémantique (métonymie, métaphore) et de construction morphologique (procédés morphologiques, contraintes sémantiques sur les bases verbales) dans la formation des noms déverbaux ambigus.

Dans ce qui suit, nous synthétisons nos principaux résultats, fournissons un commentaire général, puis revenons enfin sur les limites et les perspectives de ce travail.

Résumé des résultats

Le **chapitre 1** a posé des jalons théoriques utiles au reste de notre étude en s'appuyant sur des recherches en morphologie dérivationnelle et en sémantique lexicale. Trois constats faisant consensus dans la littérature ont été mis en lumière. Il s'avère premièrement que certaines catégories (p. ex. événement, résultat, agent) et associations sémantiques (p. ex. événement/résultat, agent/instrument) se manifestent de manière récurrente parmi les noms déverbaux. Nous avons également relevé le fait que l'ambiguïté des nominalisations peut être de forme conjointe (*certificat, attestation*) ou disjointe (*construction, accordeur*), selon que les interprétations sont compatibles ou non en contexte. Enfin, il a été rappelé que les procédés morphologiques et les propriétés sémantiques des lexèmes-bases jouent un rôle déterminant dans l'ambiguïté des dérivés.

Le **chapitre 2** a présenté la méthodologie que nous avons adoptée pour la collecte et l'analyse sémantique manuelle d'un vaste corpus de noms déverbaux et de lexèmes-bases. Il s'agissait en effet de rassembler une grande quantité de données finement analysées et de garantir l'identification des sens multiples, ce que les outils de sémantique distributionnelle et le typage sémantique (semi-)automatique ne permettent *a priori* pas. Nous avons premièrement détaillé dans ce chapitre l'échantillonnage des dérivés et des bases, en soulignant l'intérêt d'inclure des procédés morphologiques rares ou peu documentés. La grille d'analyse élaborée pour l'étude des types sémantiques nominaux, des rôles sémantiques, de l'aspect

Conclusion

lexical et de l'ambiguïté des verbes et des noms a par la suite été introduite. Enfin, nous avons décrit la procédure d'analyse, qui a notamment impliqué une phase d'annotation en double aveugle et le calcul de scores d'accord (Kappa de Cohen et PABAK). Au total, 46 procédés morphologiques (42 suffixes et 4 formes de conversion), 5'274 paires verbe-nom et 8'206 sens nominaux ont été examinés.

Le **chapitre 3** a établi le profil sémantique des nominalisations étudiées ainsi que de leurs verbes de base. Trois étapes ont été suivies pour arriver à ce résultat : nous avons décrit le comportement des lexèmes nominaux et verbaux en fonction des propriétés exposées dans le chapitre 2, examiné les éventuelles interdépendances entre les attributs collectés, puis déterminé comment les lexèmes nominaux et verbaux s'associent en cas d'ambiguïté. Plusieurs résultats notables ont émergé de ce cheminement, à commencer par le fait que les dérivés et les verbes qui leur sont apparentés sont plus souvent monosémiques qu'ambigus en français, les noms étant par ailleurs moins ambigus que les verbes. Il s'avère également que les noms déverbaux dénotent principalement des formes d'événements transpositionnels (*atterrissage*, *transformation*), des entités animées agentives (*pondeuse*, *écrivain*) et des artéfacts instrumentaux (*arrosoir*, *pétrin*). S'agissant des interdépendances entre propriétés, il a été démontré que les différentes caractéristiques mobilisées pour la description sémantique des noms et des verbes apparentés étaient toutes associées de manière significative, bien que de façon variée et souvent asymétrique. Enfin, un examen détaillé des configurations d'appariement verbe-nom a révélé que les sens des noms déverbaux ambigus s'inscrivent dans divers modèles morphosémantiques, selon les relations qu'ils entretiennent entre eux et avec les bases verbales. La construction morphologique et l'extension sémantique exercent une influence variable au sein de ces modèles en tant qu'outils de création de sens.

Le **chapitre 4** visait à quantifier et à décrire les différentes associations sémantiques investies par les nominalisations, mais aussi à identifier les potentiels mécanismes d'extension sémantique à l'origine de ces schémas d'ambiguïté. Une analyse de règles d'association a premièrement révélé une série de combinaisons de types sémantiques dont l'existence ne peut être attribuée au hasard, suggérant que celles-ci – ou en tout cas une partie d'entre elles – mettent potentiellement en jeu des relations sémantiques de type métonymique. Ces associations, qu'elles soient familières (p. ex. événement/résultat) ou moins connues (p. ex. événement/événement-collectif), présentent plusieurs caractéristiques notables. Elles sont tendanciellement dirigées, reflétant de fait la potentielle asymétrie des schémas d'extension conceptuelle. Elles sont bipartites (conformément aux attentes de nombreuses études sur la polysémie régulière), mais aussi tripartites, ce qui pourrait indiquer qu'il existe des microparadigmes sémantiques. Enfin, elles peuvent illustrer deux formes de régularité – intra- et inter-types –, incluant en effet parfois des facettes de sens (p. ex. artéfact et objet cognitif dans le cas du nom *attestation* 'témoignage écrit'). Ces différentes associations ont ensuite été étudiées sous la forme d'un réseau dans le but d'atteindre une vision globale du phénomène de l'ambiguïté dans la nominalisation. Les résultats montrent que l'ambiguïté des noms déverbaux est fortement centralisée d'un point de vue systémique, se présentant essentiellement sous la forme d'un réseau en étoile dont le centre de gravité est le type

Événement-TRANSPOSITION. L'hypothèse a été formulée que l'ambiguïté des déverbaux se structure de manière radiale plutôt que par ressemblance de famille, et que l'ascendant morphologique verbal joue un rôle crucial dans la constitution de cette ambiguïté.

Dernière étape de ce travail, le **chapitre 5** a évalué l'influence de la construction morphologique sur l'ambiguïté des noms déverbaux en examinant la contribution des procédés morphologiques et des propriétés sémantiques des lexèmes-bases. Il a été démontré que la polyfonctionnalité – ou « polysémie » – des affixes et l'ambiguïté des noms dérivés constituent des phénomènes distincts, mais aussi étroitement liés. Une analyse de règles d'association couplée à une analyse de réseau a révélé que les procédés morphologiques les plus polyfonctionnels engendrent différentes formes d'organisation systémique dans le lexique des déverbaux ambigus, avec des implications notables pour l'étude de la rivalité ou de la similarité affixale, ainsi que pour celle de la relation entre sémantique de la suffixation et sémantique de la conversion. Concernant les propriétés des verbes de base, deux analyses de règles d'association séquentielles ont montré que des schémas d'ambiguïté nominale sont positivement corrélés dans le domaine verbal avec certaines valeurs aspectuelles (dynamacité, durativité, télicité et absence de post-phase) et des rôles sémantiques (sujets agentifs, objets patientifs). Enfin, une troisième analyse de règles d'association séquentielles a mis en évidence que les procédés morphologiques, les valeurs aspectuelles et les rôles sémantiques assignés aux arguments œuvrent plus efficacement ensemble que séparément à la construction de noms déverbaux ambigus. Ces propriétés morphosémantiques ne revêtent cependant pas toutes la même importance : les procédés morphologiques jouent un rôle plus crucial que les rôles sémantiques, lesquels sont eux-mêmes plus déterminants que l'aspect lexical. Un dernier constat notable est que l'extension sémantique et la construction morphologique peuvent probablement contribuer conjointement à la formation de noms ambigus dans le lexique déverbal. En effet, la quasi-totalité des associations sémantiques positivement corrélées avec des procédés morphologiques, des valeurs aspectuelles et des rôles sémantiques met en jeu des relations sémantiques de type métonymique.

Commentaire général

On peut s'interroger sur la mise en perspective du travail et des résultats dans un contexte de recherche plus large. Au plan méthodologique, notre étude se distingue par la combinaison inédite d'analyses manuelles expertes menées sur de vastes échantillons représentatifs de l'ensemble du système étudié (les noms dérivés de verbes en français) et de généralisations statistiques s'appuyant sur des méthodes quantitatives pas ou peu exploitées en linguistique (recherche de règles d'association classiques et séquentielles). D'un point de vue théorique, l'intérêt réside dans notre positionnement à l'intersection de la sémantique lexicale – alors que celle-ci ne prend pas toujours en compte la structure morphologique – et de la morphologie dérivationnelle, laquelle n'examine pas toujours en détail les propriétés sémantiques des mots construits. Notre travail plaide ainsi en faveur d'une collaboration accrue entre ces deux champs.

Conclusion

Concernant les contributions spécifiques à notre sujet d'étude, notre recherche a démontré que l'extension sémantique et la morphologie dérivationnelle jouent un rôle actif dans la création de noms déverbaux ambigus. Nous l'avons vu dans les chapitres 4 et 5, les nominalisations donnent lieu à des associations orientées et non aléatoires entre types sémantiques évoquant des patrons de polysémie régulière, ainsi qu'à des corrélations positives et non aléatoires entre associations sémantiques et propriétés morphosémantiques. Ces résultats corroborent l'observation de Brdar (2017, p. 1) selon laquelle « tant la métonymie (et la métaphore) que la formation des mots ont une importance immense dans l'enrichissement de notre stock lexical, et donc dans l'extension, le raffinement et la systématisation de la capacité de conceptualisation humaine ⁶⁹ ».

De manière plus intéressante, notre travail a également montré qu'extension sémantique et construction morphologique peuvent opérer séparément, mais aussi et surtout conjointement (cf. chapitre 3). Les associations que nous avons identifiées dans les chapitres consacrés à l'extension sémantique (chapitre 4) et à la construction morphologique (chapitre 5) se chevauchent en effet en grande partie, révélant plusieurs formes de contributions conjointes :

- (i) des contributions cloisonnées, lorsque l'on applique deux procédés à la suite de manière indépendante pour former deux lexèmes ($V_1 \rightarrow N_1$, $V_1 \rightarrow N_2$), et qu'un lien sémantique (typiquement métonymique) peut se dessiner après coup ($N_1 \rightarrow N_2$);
- (ii) des contributions convergentes, lorsque l'on applique simultanément une extension ($N_1 \rightarrow N_2$) et un procédé ($V_1 \rightarrow N_2$) pour former un même lexème;
- (iii) voire des contributions enchâssées, lorsque l'on construit deux lexèmes dont les lectures sont conditionnées par la forme ($V_1 \rightarrow N_1 + N_2$).

Certains apports peuvent aussi être dégagés relativement aux travaux existants dans des domaines connexes. Au-delà de l'analyse de la contribution sémantique des procédés dérivationnels et de la comparaison structurelle de l'ambiguïté des mots et des affixes comme deux manifestations de la relation *one form-many meanings*, notre étude ouvre de nouvelles perspectives sur la compétition morphologique. Les résultats présentés au chapitre 5 révèlent en effet que les affixes et les formes de conversions examinés n'engendrent pas les mêmes alternances sémantiques, même lorsque la littérature les considère comme concurrents. Il s'ensuit que les capacités de réalisation ambiguë des procédés morphologiques devraient être intégrées dans l'analyse des situations de similarité ou de compétition. La similarité entre procédés pourrait s'évaluer non seulement en termes de types sémantiques instanciés et de fréquence à laquelle ces catégories sont investies par des dérivés, mais aussi en fonction de la manière dont les types sémantiques s'associent entre eux. Les outils d'analyse de réseau pourraient offrir ici une implémentation technique intéressante, permettant de modéliser ces trois dimensions.

69. [En anglais dans le texte] « both metonymy (together with metaphor) and word-formation are of immense importance in enriching our vocabulary stock, and therefore in extending, refining and systematizing human conceptualizing capacity ».

Dans un tout autre domaine, notre travail met aussi en lumière la possibilité d'élargir les analyses de la colexification, habituellement centrées sur des référents individuels, à des catégories de référents, comme l'a initié Tjuka (2024) dans ses recherches sur l'ambiguïté partie du corps/objet. Nos résultats plaident également pour une prise en compte accrue de la construction morphologique dans les études sur la colexification (cf. List, 2023; Tjuka & List, 2024). Sur un plan systémique, l'ascendant morphologique verbal paraît en effet jouer un rôle déterminant dans la structuration de l'ambiguïté des noms déverbaux, qui présente des affinités fortes avec le domaine de l'action et les différents participants que les procès peuvent impliquer. Les événements transpositionnels endossent dans ce contexte le rôle d'intermédiaires sémantiques privilégiés, dont la présence paraît cruciale pour que d'autres catégories de sens puissent émerger ou interagir (cf. chapitres 3 et 4).

Perspectives

La recherche menée dans ce travail se heurte à plusieurs limites. En se concentrant exclusivement sur le français et les nominalisations déverbaux, elle n'a évidemment pas épuisé la question du rapport forme-sens dans le domaine nominal. En outre, seules les configurations d'ambiguïté de base unique (i.e. émergeant d'un seul lexème-base) ont été étudiées en détail. Cette situation laisse entrevoir différentes perspectives de recherche, également nourries par les apports et les découvertes de ce travail. Nous discutons certaines d'entre elles ci-dessous.

Configurations d'ambiguïté à bases multiples Les configurations d'ambiguïté à bases multiples n'ont été que partiellement explorées dans ce travail et pourraient faire l'objet d'une nouvelle campagne d'annotation. En effet, des résultats différents de ceux obtenus pour les configurations d'ambiguïté à base unique pourraient émerger. Concernant les relations sémantiques dont elles témoignent, on peut émettre l'hypothèse que si la métonymie est prédominante dans les configurations d'ambiguïté à base unique, elle cède le pas à la métaphore dans les configurations à bases multiples. Comme évoqué précédemment, la métaphore est en effet plus courante que la métonymie dans le domaine verbal (cf. King & Gentner, 2022), et les configurations d'ambiguïté à bases multiples se fondent justement sur le principe que les noms héritent de l'ambiguïté du verbe (p. ex. *harponner*₁ 'frapper avec un harpon' → *harponnage*₁ 'action de frapper avec un harpon', *harponner*₂ 'saisir au passage' → *harponnage*₂ 'action de saisir au passage'). Il paraît également possible que de nouveaux microparadigmes morphosémantiques puissent être identifiés, avec d'autres formes de motivation combinée pour la construction de sens comme *harponnage*₂ 'action de saisir au passage'.

Lexique simple et autres noms construits Les structures d'ambiguïté du lexique nominal mériteraient d'être examinées au-delà des seules nominalisations déverbaux. Dans une perspective systémique, le postulat de Croft (1991) selon lequel les noms simples dénotent prototypiquement des objets, les noms construits déverbaux des actions, et

Conclusion

les noms construits désadjectivaux des propriétés, pourrait être évalué à l'échelle de l'ambiguïté des unités lexicales. L'hypothèse peut être formulée que certaines alternances sémantiques sont plus fréquentes parmi les noms déverbaux que dans le reste du lexique nominal, et qu'inversement, certaines associations sont beaucoup plus rares parmi les noms déverbaux que dans le reste du lexique nominal. Notre travail a par exemple confirmé que le schéma événement/résultat (*gribouillage, construction*) est prédominant dans la nominalisation. Il n'est pas certain que dans le lexique simple, les noms d'événement (*baptême, grève, safari*) présentent aussi volontiers un sens additionnel résultatif. De même, il semble envisageable que les alternances impliquant exclusivement des entités concrètes comme animal/viande (*agneau, veau*) ou matière/objet (*feutre, fer*) restent principalement la chasse gardée du lexique simple plutôt que du lexique déverbal ou désadjectival. De manière générale, on peut aussi s'interroger sur les possibles conséquences de l'existence d'alternances sémantiques distinctives des noms construits. Parmi elles, la présence dans le lexique de patrons morphologiques complexes enchâssant une extension sémantique régulière, et permettant ainsi de former des noms ambigus dont les types sémantiques sont à la fois liés et spécifiés, pourrait être examinée plus avant (cf. Salvadori & Huyghe, 2022).

Diversité linguistique Il y a lieu de s'interroger sur l'existence dans d'autres langues de liens conceptuels et de réseaux d'ambiguïté comparables à ceux mis au jour pour le français, et plus largement sur le rôle de la morphologie dans la colexification de types observée dans les langues du monde. Y a-t-il par exemple des invariants conceptuels et cognitifs qui sous-tendent les relations sémantiques et qui en même temps s'appuient sur des relations morphologiques partagées? L'orientation des extensions sémantiques pourrait elle aussi être examinée, étant potentiellement sujette à variation. L'emploi de noms de parties du corps pour désigner des objets, s'il est par exemple fréquent dans de nombreuses langues dont le français (*pied, coude, dos*, etc.), demeure rare et idiosyncratique dans d'autres (Tjuka & List, 2024).

Sens produits par la morphologie vs la sémantique Un autre élément pouvant être approfondi tient à la comparaison des sens produits par les opérations morphologiques et les figures sémantiques. On peut en effet se demander dans quelle mesure morphologie et extension, en tant que mécanismes permettant de générer du sens à partir de nouvelles formes (morphologie) ou de formes existantes (sémantique), ont des capacités sémantiques partagées vs des capacités distinctives, que ce soit en termes d'opérations sémantiques réalisées ou de types sémantiques créés.

Organisation des dérivés ambigus dans le lexique mental Enfin, une réflexion pourrait être menée sur l'organisation des dérivés ambigus dans le lexique mental. Il serait pertinent d'examiner si le caractère construit ou non des mots ambigus affecte leur représentation et leur traitement. En outre, nous pourrions chercher à déterminer si les différents types de figures (métaphore, métonymie) ou les types complexes (p. ex. Événement*Etat, Artéfact*Institution) sont affectés par la construction morphologique et la possible médiation morphosémantique par un verbe de base. Une autre question

Conclusion

méritant d'être creusée concerne l'existence d'une possible contrepartie psychologique aux différents rapports repérés entre dérivation et figure : par exemple, une instance de l'alternance agent/instrument a-t-elle la même représentation mentale et induit-elle le même rapport entre acceptions qu'une métonymie action/résultat ?

Dans l'ensemble, nous espérons avoir montré dans ce travail l'intérêt de se pencher sur l'ambiguïté des mots complexes, ainsi que la fécondité des études à l'interface de la sémantique lexicale et de la morphologie. Beaucoup reste à faire.

Bibliographie

- Agrawal, R., & Srikant, R. (1995). Mining sequential patterns. In P. S. Yu & A. L. P. Chen (Éd.), *ICDE '95: Proceedings of the Eleventh International Conference on Data Engineering* (p. 3-14). IEEE Computer Society. <https://doi.org/10.1109/ICDE.1995.380415>
- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *SIGMOD Rec.*, 22(2), 207-216. <https://doi.org/10.1145/170036.170072>
- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy (Éd.), *Advances in knowledge discovery and data mining* (p. 307-328). American Association for Artificial Intelligence.
- Agrell, S. (1908). *Aspektänderung und Aktionsartbildung beim polnischen Zeitworte : Ein Beitrag zum Studium der indogermanischen Präverbia und ihrer Bedeutungsfunktionen*. Håkan Ohlssons Buchdruckerei.
- Alexiadou, A. (2001). *Functional structure in nominals. Nominalization and ergativity*. John Benjamins.
- Alexiadou, A. (2019). Event/Result in morphology. In M. Aronoff (Éd.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.555>
- Aliquot-Suengas, S. (2003). La productivité actuelle de la forme constructionnelle *-ade*. *Langue française*, 140, 38-55.
- Anderson, S. R. (1994). The form and content of derivational relations. *Proceedings of the Mid-America Linguistics Conference*, 3-22.
- Anscombre, J.-C. (1986). L'article zéro en français : un imparfait du substantif ? *Langue française*, 72, 4-39. <https://doi.org/10.3406/lfr.1986.4656>
- Apothéloz, D. (2008). *Entrer quelques instants vs arriver quelques instants : le problème de la spécification de la durée de l'état résultant*. *Verbum*, 30(2-3), 199-219.
- Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, 12(142), 5-32. <https://doi.org/10.1515/ling.1974.12.142.5>
- Arapinis, A., & Vieu, L. (2015). A plea for complex categories in ontologies. *Applied Ontology*, 10(3-4), 285-296. <https://doi.org/10.3233/ao-150156>
- Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.

Bibliographie

- Asher, N. (2011). *Lexical meaning in context. A web of words*. Cambridge University Press.
- ATILF - CNRS & Université de Lorraine. (s. d.). *TLFi : Trésor de la langue Française informatisé*. <http://www.atilf.fr/tlfi>
- Bach, E. (1981). On time, tense, and aspect: An essay in English metaphysics. In P. Cole (Éd.), *Radical pragmatics* (p. 63-81). Academic Press.
- Balvet, A., Barque, L., Condette, M.-H., Haas, P., Huyghe, R., Marín, R., & Merlo, A. (2011). Nomage: An electronic lexicon of French deverbal nouns based on a semantically annotated corpus. In B. Sagot (Éd.), *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011* (p. 9-16).
- Barbu Mititelu, V., Leseva, S., & Stoyanova, I. (2023). Semantic analysis of verb – noun zero derivation in Princeton WordNet. *Zeitschrift für Sprachwissenschaft*, 42(1), 181-207. <https://doi.org/10.1515/zfs-2022-2017>
- Barque, L. (2008). *Description et formalisation de la polysémie régulière du français* (thèse de doct.). Université Paris 7 Denis Diderot, UFR L laboratoire LaTTiCe.
- Barque, L., Candito, M., & Huyghe, R. (2019). La classification des verbes réfléchis à l'épreuve d'une annotation en corpus. *Langages*, 216(4), 121-137. <https://doi.org/10.3917/lang.216.0121>
- Barque, L., Fábregas, A., & Marín, R. (2012). Les noms d'état psychologique et leurs 'objets' : étude d'une alternance sémantique. *Lexique*, 20, 21-42.
- Barque, L., & Haas, P. (2022). La polysémie régulière : approches empiriques. Présentation. *Lexique*, 31, 5-25. <https://doi.org/10.54563/lexique.809>
- Barque, L., Haas, P., & Huyghe, R. (2014). La polysémie nominale *événement/objet* : quels objets pour quels événements ? *Neophilologica*, 26, 170-187.
- Barque, L., Haas, P., & Huyghe, R. (2018). Polysémie régulière et néologie sémantique. Constitution d'une ressource pour l'étude des sens nouveaux. *Neologica*, 12, 91-108. <https://doi.org/10.15122/isbn.978-2-406-08196-8.p.0091>
- Barra, V., Cornuéjols, A., & Miclet, L. (2021). *Apprentissage artificiel : concepts et algorithmes - de Bayes et Hume au Deep Learning*. Eyrolles.
- Barz, I. (1982). Motivation und Wortbildungsbedeutung: Eine Diskussion sowjetischer Forschungsergebnisse. *Beiträge zur Erforschung der deutschen Sprache*, 2, 5-21.
- Bauer, L. (1997). Derivational paradigms. In G. Booij & J. van Marle (Éd.), *Yearbook of morphology 1996* (p. 17-59). Springer. https://doi.org/10.1007/978-94-017-3718-0_13
- Bauer, L. (2017). Metonymy and the semantics of word-formation. In N. Koutsoukos, J. Audring & F. Masin (Éd.), *Morphological variation: Synchrony and diachrony. Online Proceedings of the Eleventh Mediterranean Morphology Meeting (MMM11)* (p. 1-13). <https://doi.org/10.26220/mmm.2868>
- Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. OUP.
- Beard, R. (1990). The nature and origins of derivational polysemy. *Lingua*, 81(2-3), 101-140.
- Beard, R. (1995). *Lexeme-morpheme base morphology: A general theory of inflection and word formation*. SUNY Press.
- Beard, R. (s. d.). *Transposition*. <https://rbeard.org/linguistics/transpos.html>

- Becker, T. (1993). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In G. Booij & J. van Marle (Éd.), *Yearbook of morphology 1993* (p. 1-25). Kluwer.
- Belyi, E., Giabbanelli, P. J., Patel, I., Balabhadrapathruni, N. H., Abdallah, A. B., Hameed, W., & Mago, V. K. (2016). Combining association rule mining and network analysis for pharmacosurveillance. *The Journal of Supercomputing*, 72, 2014-2034. <https://doi.org/10.1007/s11227-016-1714-y>
- Bierwisch, M. (1990-1991). Event nominalizations: Proposals and problems. *Acta Linguistica Hungarica*, 40(1-2), 19-84. <http://www.jstor.org/stable/44306682>
- Bisetto, A., & Melloni, C. (2007). Result nominals: A lexical-semantic investigation. In G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli & S. Scalise (Éd.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* (p. 393-412).
- Bojanowski, M. (2023). *intergraph: Coercion routines for network data objects* [R package version 2.0-3]. <https://mbojan.github.io/intergraph/>
- Bonami, O. (1999). *Les constructions du verbe : le cas des groupes prépositionnels argumentaux. Analyse syntaxique, sémantique et lexicale* (thèse de doct.). Université Paris VII – Denis Diderot.
- Bonami, O. (2014). *La structure fine des paradigmes de flexion : études de morphologie descriptive, théorique et formelle* (Mémoire d'habilitation à diriger des recherches). 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., & Kerleroux, F. (2009). L'allomorphie radicale et la relation flexion-construction. In B. Fradin, F. Kerleroux & M. Plénat (Éd.), *Aperçus de morphologie du français* (p. 103-125). Presses de l'Université de Vincennes.
- Bonami, O., & Guzmán Naranjo, M. (2023). Distributional evidence for derivational paradigms. In S. Kotowski & I. Plag (Éd.), *The semantics of derivational morphology: Theory, methods, evidence* (p. 219-258). De Gruyter. <https://doi.org/10.1515/9783111074917-008>
- Bonami, O., & Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, 29, 167-197. <https://doi.org/10.1007/s11525-018-9322-6>
- Bond, F., Janz, A., Maziarz, M., & Rudnicka, E. (2019). Testing Zipf's meaning-frequency law with wordnets as sense inventories. In P. Vossen & C. Fellbaum (Éd.), *Proceedings of the 10th Global Wordnet Conference* (p. 342-352). Global Wordnet Association. <https://aclanthology.org/2019.gwc-1.44>
- Booij, G. (1986). Form and meaning in morphology: The case of Dutch agent nouns. *Linguistics*, 24(3), 503-517. <https://doi.org/10.1515/ling.1986.24.3.503>
- Booij, G. (2005). *The grammar of words: An introduction to linguistic morphology*. Oxford University Press.
- Booij, G. (2019). The role of schemas in Construction Morphology. *Word Structure*, 12(3), 385-395. <https://doi.org/10.3366/word.2019.0154>

Bibliographie

- Booij, G., & Lieber, R. (2004). On the paradigmatic nature of affixal semantics in English and Dutch. *Linguistics*, 42(2), 327-357. <https://doi.org/10.1515/ling.2004.011>
- Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations. In J. Moore & M. Polinsky (Éd.), *The nature of explanations in linguistic theory* (p. 31-67). CSLI; University of Chicago.
- Bowerman, M. (1990). Mapping thematic roles onto syntactic functions: Are children helped by innate linking rules? *Linguistics*, 28(6), 1253-1290. <https://doi.org/10.1515/ling.1990.28.6.1253>
- Brdar, M. (2017). *Metonymy and word-formation: Their interactions and complementation*. Cambridge Scholars Publishing.
- Brdar, M., & Brdar-Szabó, R. (2014). Where does metonymy begin? Some comments on Janda (2011). *Cognitive Linguistics*, 25(2), 313-340. <https://doi.org/10.1515/cog-2014-0013>
- Bréal, M. (1897). *Essai de sémantique*. Librairie Hachette.
- Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *SIGMOD Rec.*, 26(2), 255-264. <https://doi.org/10.1145/253262.253325>
- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.
- Brochhagen, T., & Boleda, G. (2022). When do languages use the same word for different meanings? The Goldilocks principle in colexification. *Cognition*, 226, 105179. <https://doi.org/10.1016/j.cognition.2022.105179>
- Brughmans, T., & Peeples, M. A. (2023). *Network science in archaeology*. Cambridge University Press.
- Buchta, C., Hahsler, M., & Diaz, D. (2023). *arulesSequences: Mining frequent sequences* [R package version 0.2-30]. <https://CRAN.R-project.org/package=arulesSequences>
- Buljan, G. (2018). The Croatian suffix *-stv(o)*: A study of meaning and polysemy in word formation. *Journal of Slavic Linguistics*, 26(2), 185-244. <https://doi.org/10.1353/jsl.2018.0008>
- Burdy, P. (2013). *Die mittels -aison und Varianten gebildeten Nomina des Französischen von den Anfängen bis zur Gegenwart: Eine Studie zur diachronen Wortbildung*. Vittorio Klostermann.
- Butts, C. T. (2008). Social networks: A methodological introduction. *Asian Journal of Social Psychology*, 11(1), 13-41. <https://doi.org/10.1111/j.1467-839X.2007.00241.x>
- Butts, C. T. (2023). *sna: Tools for social network analysis* [R package version 2.7-2]. <https://CRAN.R-project.org/package=sna>
- Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), 423-429. [https://doi.org/10.1016/0895-4356\(93\)90018-v](https://doi.org/10.1016/0895-4356(93)90018-v)
- Casas, B., Hernández-Fernández, A., Català, N., Ferrer-i-Cancho, R., & Baixeries, J. (2019). Polysemy and brevity versus frequency in language. *Computer Speech & Language*, 58, 19-50. <https://doi.org/10.1016/j.csl.2019.03.007>
- Chapman, C., & McDonnell Feit, E. (2019). *R for marketing research and analytics*. Springer.

- Chomsky, N. (1970). Remarks on nominalization. In R. A. Jacobs & P. S. Rosenbaum (Éd.), *Readings in English Transformational Grammar* (p. 184-221). Ginn; Company.
- Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L. A. (2007). *Data mining: A knowledge discovery approach*. Springer.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.
- Condette, M.-H., Marín, R., & Merlo, A. (2012). La structure argumentale des noms déverbaux : du corpus au lexique et du lexique au corpus. In F. Neveu, V. M. Toke, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost & S. Teston-Bonnard (Éd.), *Congrès Mondial de Linguistique Française – CMLF 2012* (p. 845-858). <https://doi.org/10.1051/shsconf/20120100271>
- Copot, M., & Bonami, O. (2024). Baseless derivation: The behavioural reality of derivational paradigms. *Cognitive Linguistics*, 35(2), 221-250. <https://doi.org/10.1515/cog-2023-0018>
- Corbin, D. (1976). Peut-on faire l'hypothèse d'une dérivation en morphologie ? In J.-C. Chevalier (Éd.), *Grammaire transformationnelle : syntaxe et lexique* (p. 47-91). Presses Universitaires de Lille.
- Corbin, D. (1991). *Morphologie dérivationnelle et structuration du lexique*. Presses universitaires du Septentrion.
- Cranmer, S. J., Desmarais, B. A., & Morgan, J. W. (2020). *Inferential network analysis*. Cambridge University Press.
- Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. University Press of Chicago.
- Croft, W., & Cruse, A. D. (2004). *Cognitive linguistics*. Cambridge University Press.
- Cruse, A. D. (1986). *Lexical semantics*. Cambridge University Press.
- Cruse, A. D. (1995). Polysemy and related phenomena from a cognitive linguistic viewpoint. In P. Saint-Dizier & E. Viegas (Éd.), *Computational lexical semantics* (p. 33-49). Cambridge University Press.
- Cruse, A. D. (2000). Lexical 'facets': Between monosemy and polysemy. In S. Beckmann, P.-P. König & G. Wolf (Éd.), *Sprachspiel und Bedeutung: Festschrift für Franz Hundsnurscher zum 65. Geburtstag* (p. 25-36). Max Niemeyer Verlag. <https://doi.org/10.1515/9783110913705.25>
- Cruse, A. D. (2004). Lexical facets and metonymy. *Ilha do Desterro*, 47, 73-96.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. <https://igraph.org>
- 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 (Éd.), *La morphologie des dérivés évaluatifs. Forum de morphologie* (p. 37-47).

Bibliographie

- Dal, G., Hathout, N., Lignon, S., Namer, F., & Tanguy, L. (2018). Toile *versus* dictionnaires : les nominalisations du français en *-age* et en *-ment*. In F. Neveu, B. Harmegnies, L. Hriba & S. Prévost (Éd.), *Congrès Mondial de Linguistique Française - CMLF 2018* (p. 08003). Institut de Linguistique Française. <https://doi.org/10.1051/shsconf/20184608003>
- Dal, G., & Namer, F. (2010). Les noms en *-ance/-ence* du français : quel(s) patron(s) constructionnel(s)? In F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada & S. Prévost (Éd.), *CMLF 2010 - 2ème Congrès Mondial de Linguistique Française* (p. 893-907). Institut de Linguistique Française. <https://doi.org/10.1051/cmlf/2010154>
- Darmesteter, A. (1887). *La vie des mots étudiée dans leurs significations*. Librairie Ch. Delagrave.
- Dattner, E., Ashkenazi, O., Ravid, D., & Levie, R. (2023). Explaining dynamic morphological patterns in acquisition using Network Analysis. *Morphology*, 33, 511-556. <https://doi.org/10.1007/s11525-022-09394-0>
- Davidson, R., & Harel, D. (1996). Drawing graphs nicely using simulated annealing. *ACM Transactions on Graphics*, 15(4), 301-331. <https://doi.org/10.1145/234535.234538>
- Demonte, V., & McNally, L. (Éd.). (2012). *Telicity, change, and state: A cross-categorical view of event structure*. Oxford University Press.
- Denistia, K., Shafaei-Bajestan, E., & Baayen, R. H. (2022). Exploring semantic differences between the Indonesian prefixes *PE-* and *PEN-* using a vector space model. *Corpus Linguistics and Linguistic Theory*, 18(3), 573-598. <https://doi.org/10.1515/cllt-2020-0023>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- de Vries, H. (2021). Collective nouns. In P. Cabredo Hofherr & J. Doetjes (Éd.), *The Oxford handbook of grammatical number* (p. 257-274). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198795858.013.12>
- Di Eugenio, B., & Glass, M. (2004). The Kappa statistic: A second look. *Computational Linguistics*, 30(1), 95-101. <https://doi.org/10.1162/089120104773633402>
- Di Natale, A., Pellert, M., & Garcia, D. (2021). Colexification networks encode affective meaning. *Affective Science*, 2, 99-111. <https://doi.org/10.1007/s42761-021-00033-1>
- Dixon, R. M. W. (2005). *A semantic approach to English grammar*. OUP.
- Dölling, J. (2020). Systematic polysemy. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullmann & T. E. Zimmermann (Éd.), *The Wiley Blackwell companion to semantics*. John Wiley & Sons. <https://doi.org/10.1002/9781118788516.sem099>
- Dowty, D. R. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. D. Reidel Publishing Company.

- Drapeau, L., & Boulanger, A. (1982). Les suffixes *-erie* en français populaire : implications pour la théorie morphologique. *Revue québécoise de linguistique*, 11(2), 73-91. <https://doi.org/10.7202/602488ar>
- Dubois, J. (1962). *Étude sur la dérivation suffixale en français moderne et contemporain : essais d'interprétation des mouvements observés dans le domaine de la morphologie des mots construits*. Larousse.
- Dubois, J., & Dubois-Charlier, F. (1999). *La dérivation suffixale en français*. Nathan.
- Éditions Le Robert. (s. d.-a). *Le Petit Robert de la langue française*. petitrobert.lerobert.com
- Éditions Le Robert. (s. d.-b). *Petit dictionnaire des suffixes du français*. <https://petitrobert.lerobert.com/AidePR/Pages/Suffixes.HTML>
- Fábregas, A., & Marín, R. (2012). The role of Aktionsart in deverbal nouns: State nominalizations across languages. *Journal of Linguistics*, 48(1), 35-70. <https://doi.org/10.1017/S002226711000351>
- Fábregas, A., Marín, R., & McNally, L. (2012). From psych verbs to nouns. In V. Demonte & L. McNally (Éd.), *Telicity, change, and state: A cross-categorical view of event structure* (p. 162-184). Oxford University Press.
- Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, 157, 1-16. <https://doi.org/10.1016/j.lingua.2015.02.002>
- Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa : The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43, 543-549. [https://doi.org/10.1016/0895-4356\(90\)90158-L](https://doi.org/10.1016/0895-4356(90)90158-L)
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press.
- Ferret, K., & Villoing, F. (2015). French N-age instrumentals: Semantic properties of the base verb. *Morphology*, 25, 473-496. <https://doi.org/10.1007/s11525-015-9273-0>
- Filip, H. (2012). Lexical aspect. In R. I. Binnick (Éd.), *The Oxford handbook of tense and aspect* (p. 721-751). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195381979.013.0025>
- Filip, H. (2016). *Aspect, eventuality types and nominal reference*. Routledge.
- Flaux, N. (1999). À propos des noms collectifs. *Revue de linguistique romane*, 63(251), 471-502.
- 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. (2011). Remarks on state denoting nominalizations. *Recherches linguistiques de Vincennes*, 40, 73-99. <https://doi.org/10.4000/rlv.1946>
- Fradin, B. (2012). Les nominalisations et la lecture "moyen". *Lexique*, 20, 129-156.
- Fradin, B. (2016). L'interprétation des nominalisations en N-age et N-ment en français. In E. Büchi, J.-P. Chauveau & J.-M. Pierrel (Éd.), *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013)* (p. 53-66). ÉLiPhi.
- Fradin, B. (2023). La polysémie et l'argument dérivationnel. In S. Palma & A. Theissen (Éd.), *Res per nomen VIII. Polysémie et référence* (p. 53-66). ÉPURE - Éditions et Presses universitaires de Reims.

Bibliographie

- Fradin, B., & Kerleroux, F. (2003). Quelle base pour les procédés de la morphologie constructionnelle ? *Silexicales*, 3, 76-84.
- François, A. (2008). Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. In M. Vanhove (Éd.), *From polysemy to semantic change: Towards a typology of lexical semantic associations* (p. 163-215). John Benjamins.
- Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(6), 1366-1383. <https://doi.org/10.1037/0278-7393.25.6.1366>
- Fuchs, C. (1996). *Les ambiguïtés du français*. Ophrys.
- Fukuya, M., Matsumoto, T., Shimada, Y., & Ikeguchi, T. (2023). Central emotions and hubs in a colexification network. *Scientific Reports*, 13, 21823. <https://doi.org/10.1038/s41598-023-48922-8>
- Gaeta, L. (2000). On the interaction between morphology and semantics: The Italian suffix *-ata*. *Acta Linguistica Hungarica*, 47(1), 205-229.
- Gamillscheg, E. (1921). Grundzüge der galloromanischen Wortbildung. In E. Gamillscheg & L. Spitzer (Éd.), *Beiträge zur romanischen Wortbildungslehre* (p. 1-80). Léo S. Olschki.
- Gardani, F., Rainer, F., & Luschützky, H. C. (2019). Competition in morphology: A historical outline. In F. Rainer, Gardani, W. U. Dressler & H. C. Luschützky (Éd.), *Competition in inflection and word-formation* (p. 3-36). Springer. https://doi.org/10.1007/978-3-030-02550-2_1
- Garey, H. B. (1957). Verbal aspect in French. *Language*, 33(2), 91-110. <https://doi.org/10.2307/410722>
- Geeraerts, D. (1993). Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics*, 4(3), 223-272. <https://doi.org/10.1515/cogl.1993.4.3.223>
- Geeraerts, D. (2017). Lexical semantics. In M. Aronoff (Éd.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.29>
- Geng, L., & Hamilton, H. J. (2006). Interestingness measures for data mining: A survey. *ACM Computing Surveys*, 38(3), 1-32. <https://doi.org/10.1145/1132960.1132963>
- Gentner, D. (1981). Some interesting differences between nouns and verbs. *Cognition and Brain Theory*, 4, 161-178.
- Glaser, K. (1910). Le sens péjoratif du suffixe *-ard* en français. *Romanische Forschungen*, 27(3), 932-983. <http://www.jstor.org/stable/27935761>
- Godard, D., & Jayez, J. (1996). Types nominaux et anaphores : le cas des objets et des événements. In W. De Mulder, L. Tasmowski-De Ryck & C. Veters (Éd.), *Anaphores temporelles et (in-)cohérence, Cahiers Chronos I* (p. 41-58). Rodopi.
- Gonzalez-Agirre, A., Laparra, E., & Rigau, G. (2012). Multilingual Central Repository version 3.0. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (Éd.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (p. 2525-2529). European Language

- Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/293_Paper.pdf
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268), 732-764. <https://doi.org/10.2307/2281536>
- Gries, S. T. (2009). *Quantitative corpus linguistics with R: A practical introduction* (1^{re} éd.). Routledge.
- Gries, S. T., Jansegers, M., & Miglio, V. G. (2020). Quantitative methods for corpus-based contrastive linguistics. In R. Enghels, B. Defrancq & M. Jansegers (Éd.), *New approaches to contrastive linguistics: Empirical and methodological challenges* (p. 53-84). De Gruyter Mouton. <https://doi.org/10.1515/9783110682588-003>
- Grimshaw, J. (1990). *Argument structure*. MIT Press.
- Guzmán Naranjo, M., & Bonami, O. (2023). A distributional assessment of rivalry in word formation. *Word structure*, 16(1), 87-114. <https://doi.org/10.3366/word.2023.0222>
- Haas, P. (2009). *Comment l'aspect vient aux noms. Les propriétés aspectuelles des noms à l'épreuve des restrictions de sélection imposées par certaines prépositions* (thèse de doct.). Université de Lille III – Charles de Gaulle.
- Haas, P., Barque, L., Huyghe, R., & Tribout, D. (2023). Pour une classification sémantique des noms en français appuyée sur des tests linguistiques. *Journal of French Language Studies*, 33(1), 52-81. <https://doi.org/10.1017/S0959269522000187>
- Haas, P., Huyghe, R., & Marín, R. (2008). Du verbe au nom : calques et décalages aspectuels. In D. J., H. B. & L. B. (Éd.), *Congrès Mondial de Linguistique Française - CMLF'08* (p. 2051-2065). Institut de Linguistique Française. <https://doi.org/10.1051/cmlf08108>
- Haas, P., & Jugnet, A. (2013). De l'existence des prédicats d'achèvements. *Linguisticæ Investigationes*, 36(1), 56-89. <https://doi.org/10.1075/li.36.1.03haa>
- Hahsler, M. (2015). *A probabilistic comparison of commonly used interest measures for association rules*. <https://mhahsler.github.io/arules/docs/measures>
- Hahsler, M. (2017). arulesViz: Interactive visualization of association rules with R. *The R Journal*, 9(2), 163-175. <https://doi.org/10.32614/RJ-2017-047>
- Hahsler, M., Grün, B., & Hornik, K. (2005). arules - A computational environment for mining association rules and frequent item sets. *Journal of Statistical Software*, 14(15), 1-25. <https://doi.org/10.18637/jss.v014.il5>
- Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello (Éd.), *The new psychology of language* (p. 211-242). Erlbaum.
- Haspelmath, M. (2023). Coexpression and synexpression patterns across languages: Comparative concepts and possible explanations. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1236853>
- Haspelmath, M., & Sims, A. D. (2013). *Understanding morphology*. Routledge.

Bibliographie

- Hathout, N., & Namer, F. (2014a). Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, 11. <https://aclanthology.org/2014.lilt-11.6>
- Hathout, N., & Namer, F. (2014b). Discrepancy between form and meaning in word formation: The case of over- and under-marking in French. In F. Rainer, W. U. Dressler, F. Gardani & H. C. Luschützky (Éd.), *Morphology and meaning (Selected papers from the 15th International Morphology Meeting, Vienna, February 2012)* (p. 177-190). John Benjamins.
- Hay, J., Kennedy, C., & Levin, B. (1999). Scalar structure underlies telicity in "degree achievements". In T. Matthews & D. Strolovitch (Éd.), *Proceedings of SALT 9* (p. 127-144). CLC Publications. <https://doi.org/10.3765/salt.v9i0.2833>
- Heidinger, S. (2010). *French anticausatives. A diachronic perspective*. Mouton De Gruyter.
- Hernández, B., Reilly, R. B., & Kenny, R. A. (2019). Investigation of multimorbidity and prevalent disease combinations in older Irish adults using network analysis and association rules. *Scientific Reports*, 9(1), 14567. <https://doi.org/10.1038/s41598-019-51135-7>
- Hernández-Fernández, A., Casas, B., Ferrer-i-Cancho, R., & Baixeries, J. (2016). Testing the robustness of laws of polysemy and brevity versus frequency. In P. Král & C. Martín-Vide (Éd.), *Statistical language and speech processing* (p. 19-29). Springer International Publishing.
- Holland, P. W., & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 76(3), 492-513. <https://doi.org/10.1086/224954>
- Hollway, J. (2021). *migraph: Multimodal network analysis and more*. <https://github.com/stocnet/migraph>
- Hollway, J. (2024). *manynet: Many ways to make, modify, mark, and map myriad networks* [R package version 0.4.4]. <https://stocnet.github.io/manynet/>
- Huguin, M., Barque, L., Haas, P., & Tribout, D. (2023). Typage sémantique des noms dans la ressource morphologique Démonette. *Lexique*, 33, 41-56. <https://doi.org/10.54563/lexique.1086>
- Huyghe, R. (2011). (A)telicity and the mass-count distinction. *Recherches linguistiques de Vincennes*, 40, 101-126. <https://doi.org/10.4000/rlv.1952>
- Huyghe, R. (2014). La sémantique des noms d'action : quelques repères. *Cahiers de lexicologie*, 105(2), 181-201.
- Huyghe, R. (2015). Les typologies nominales : présentation. *Langue française*, 185(1), 5-27.
- Huyghe, R. (2021). Building a lexical database to investigate the semantics of French deverbal nouns. In F. Namer, N. Hathout, S. Lignon, M. Ševčíková & Z. Žabokrtský (Éd.), *Proceedings of the Third International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2021)* (p. 21-29). ATILF (CNRS & Université de Lorraine).
- Huyghe, R., Barque, L., Delafontaine, F., & Salvadori, J. (2024). The ambiguous nature of complex semantic types: An experimental investigation. *Language and Cognition*, 1-26. <https://doi.org/10.1017/langcog.2023.73>

- 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. <https://doi.org/10.1017/S0959269520000320>
- Huyghe, R., Lombard, A., Salvadori, J., & Schwab, S. (2023). Semantic rivalry between French deverbal neologisms in *-age*, *-ion* and *-ment*. In S. Kotowski & I. Plag (Éd.), *The semantics of derivational morphology: Theory, methods, evidence* (p. 143-176). De Gruyter. <https://doi.org/10.1515/9783111074917-006>
- Huyghe, R., & Tribout, D. (2015). Noms d'agents et noms d'instruments : le cas des déverbaux en *-eur*. *Langue française*, 185(1), 99-112. <https://doi.org/10.3917/lf.185.0099>
- Iacobini, C. (2000). Base and direction of derivation. In G. Booij, C. Lehmann & J. Mugdan (Éd.), *Morphology. An international handbook on inflection and word-formation* (p. 865-876). Mouton de Gruyter.
- Ilgen, B., & Karaoglan, B. (2007). Investigation of Zipf's 'law-of-meaning' on Turkish corpora. *2007 22nd international symposium on computer and information sciences*, 1-6. <https://doi.org/10.1109/ISCIS.2007.4456846>
- Jackendoff, R., & Audring, J. (2020). *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford University Press. <https://doi.org/10.1093/oso/9780198827900.001.0001>
- Jackson, J. C., Watts, J., Henry, T. R., List, J.-M., Forkel, R., Mucha, P. J., Greenhill, S. J., Gray, R. D., & Lindquist, K. A. (2019). Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472), 1517-1522. <https://doi.org/10.1126/science.aaw8160>
- Jacquey, E. (2006). Un cas de « polysémie logique » : modélisation de noms d'action en français ambigus entre processus et artefact. *TAL*, 47(1), 137-166. <https://aclanthology.org/2006.tal-1.6>
- Jacquey, E. (2013). Déverbaux en français, lexicographie et corpus. *Cahiers de lexicologie*, (103), 63-83. <https://doi.org/10.15122/isbn.978-2-8124-2079-5.p.0063>
- Janda, L. A. (2011). Metonymy in word-formation. *Cognitive Linguistics*, 22(2), 359-392. <https://doi.org/10.1515/cogl.2011.014>
- Janda, L. A. (2014). Metonymy and word-formation revisited. *Cognitive Linguistics*, 25(2), 341-349. <https://doi.org/10.1515/cog-2014-0008>
- Ježek, E. (2008). Polysemy of Italian event nominals. *Faits de langue*, 30, 251-264.
- Ježek, E., & Melloni, C. (2009). Complex types in the (morphologically) complex lexicon. *Proceedings of the GL 2009, 5th International Conference on Generative Approaches to the Lexicon*, 59-67.
- Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language*, 72(3), 533-578.
- Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. Oxford University Press.

Bibliographie

- Karjus, A., Blythe, R. A., Kirby, S., Wang, T., & Smith, K. (2021). Conceptual similarity and communicative need shape colexification: An experimental study. *Cognitive Science*, 45(9), e13035. <https://doi.org/https://doi.org/10.1111/cogs.13035>
- Kawaletz, L. (2023). *The semantics of English -ment nominalizations*. Language Science Press.
- Kawaletz, L., & Plag, I. (2015). Predicting the semantics of English nominalizations: A frame-based analysis of *-ment* suffixation. In L. Bauer, P. Štekauer & L. Körtvélyessy (Éd.), *Semantics of complex words* (p. 289-319). Springer.
- Kearns, K. (2007). Telic senses of deadjectival verbs. *Lingua*, 117(1), 26-66. <https://doi.org/10.1016/j.lingua.2005.09.002>
- Kearns, K. (2011). *Semantics*. Palgrave Macmillan.
- Kelling, C. (2001). Agentivity and suffix selection. In M. Butt & T. Holloway King (Éd.), *Proceedings of the LFG01 Conference* (p. 147-162). CSLI Publications. <https://web.stanford.edu/group/cslipublications/cslipublications/LFG/6/pdfs/lfg01kelling.pdf>
- Kennedy, C., & Levin, B. (2008). Measure of change: The adjectival core of degree achievements. In L. McNally & C. Kennedy (Éd.), *Adjectives and adverbs. Syntax, semantics and discourse* (p. 156-182). Oxford University Press.
- Kerleroux, F. (1996). *La coupure invisible : études de syntaxe et de morphologie*. Presses Universitaires du Septentrion.
- Kerleroux, F. (2007). On a subclass of non-affixed deverbal nouns in French. In G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli & S. Scalise (Éd.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* (p. 93-104). <https://doi.org/10.26220/mmm.2384>
- King, D., & Gentner, D. (2022). Verb metaphoric extension under semantic strain. *Cognitive Science*, 46(5), e13141. <https://doi.org/10.1111/cogs.13141>
- Kiparsky, P. (1982). Word-formation and the lexicon. In F. Ingemann (Éd.), *Proceedings of the 1982 Mid-America Linguistics Conference* (p. 3-32). University of Kansas.
- Kipper-Schuler, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon* (thèse de doct.). University of Pennsylvania.
- Kleiber, G. (1997). Massif / comptable et partie / tout. *Verbum*, 9, 321-337.
- Kleiber, G. (1999). *Problèmes de sémantique. La polysémie en questions*. Presses universitaires du Septentrion.
- Klepousniotou, E., Pike, G. B., Steinhauer, K., & Gracco, V. (2012). Not all ambiguous words are created equal: An EEG investigation of homonymy and polysemy. *Brain and Language*, 123(1), 11-21. <https://doi.org/10.1016/j.bandl.2012.06.007>
- Koenecke, A. (2019). *Tutorial: Sequential pattern mining in R for business recommendations*. <https://blog.revolutionanalytics.com/2019/02/sequential-pattern-mining-in-r.html>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>

- Koontz-Garboden, A. (2007). *States, changes of state, and the Monotonicity Hypothesis* (thèse de doct.). Stanford University.
- Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., & Zlotowski, O. (2005). Centrality indices. In U. Brandes & T. Erlebach (Éd.), *Network analysis: Methodological foundations* (p. 16-61). Springer. https://doi.org/10.1007/978-3-540-31955-9_3
- Kotowski, S., & Plag, I. (2023). The semantics of derivational morphology: Introduction. In S. Kotowski & I. Plag (Éd.), *The semantics of derivational morphology: Theory, methods, evidence* (p. 1-14). De Gruyter. <https://doi.org/10.1515/978311074917-001>
- Labelle, M., & Doron, E. (2010). Anticausative derivations (and other valency alternations) in French. *Probus*, 22(2), 303-316. <https://doi.org/10.1515/prbs.2010.011>
- Laca, B. (2001). Derivation. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Éd.), *Language typology and language universals / Sprachtypologie und sprachliche Universalien / La typologie des langues et les universaux linguistiques* (p. 1214-1227). De Gruyter Mouton. <https://doi.org/10.1515/9783110194265-026>
- Lammert, M. (2008). *Ensemble, groupe et collection : des noms collectifs bien particuliers. Scolia : Sciences Cognitives, Linguistiques et Intelligence Artificielle*, 23, 85-107. https://www.persee.fr/doc/scoli_1253-9708_2008_act_23_1_1117
- Lammert, M., & Lecolle, M. (2014). Les noms collectifs en français : une vue d'ensemble. *Cahiers de lexicologie*, 105(2), 203-222. <https://doi.org/10.15122/isbn.978-2-8124-3689-5.p.0203>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Langacker, R. W. (1991). *Concept, image, and symbol: The cognitive basis of grammar*. Mouton de Gruyter.
- Larose, C. D., & Larose, D. T. (2019). *Data science using Python and R*. Wiley.
- Lecomte, E. (1997). Tous les mots possibles en *-ure* existent-ils ? In D. Corbin, B. Fradin, B. Habert, F. Kerleroux & M. Plénat (Éd.), *Silexicales 1 : Mots possibles et mots existants* (p. 191-200). Presses de l'Université de Lille.
- Lees, R. B. (1960). *The grammar of English nominalizations*. Indiana University Research Center in Anthropology, Folklore and Linguistics.
- Lemieux, V., & Ouimet, M. (2004). *L'analyse structurale des réseaux sociaux*. De Boeck Supérieur. <https://doi.org/10.3917/dbu.lemie.2004.01>
- Levickij, V., Drebet, V., & Kiiko, S. (1999). Some quantitative characteristics of polysemy of verbs, nouns and adjectives in the German language. *Journal of Quantitative Linguistics*, 6(2), 172-187. <https://doi.org/10.1076/jqul.6.2.172.4131>
- Levin, B. (1993). *English verb classes and alternations: A preliminary survey*. The University of Chicago Press.
- Lieber, R. (2004). *Morphology and lexical semantics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486296>

Bibliographie

- Lieber, R. (2015). The semantics of transposition. *Morphology*, 25, 353-369. <https://doi.org/10.1007/s11525-015-9261-4>
- Lieber, R. (2016). *English nouns: The ecology of nominalization*. Cambridge University Press.
- Lieber, R. (2018). Nominalization: General overview and theoretical issues. In M. Aronoff (Éd.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.501>
- Lieber, R., & Plag, I. (2022). The semantics of conversion nouns and -ing nominalizations: A quantitative and theoretical perspective. *Journal of Linguistics*, 58(2), 307-343. <https://doi.org/10.1017/S0022226721000311>
- Liebesman, D., & Magidor, O. (2017). Copredication and property inheritance. *Philosophical Issues*, 27(1), 131-166. <https://doi.org/10.1111/ph1512104>
- Lipka, L. (1990). *An outline of English lexicology: Lexical structure, word semantics, and word-formation*. Max Niemeyer.
- List, J.-M. (2023). Inference of partial colexifications from multilingual wordlists. *Frontiers in Psychology*, 14, 1156540. <https://doi.org/10.3389/fpsyg.2023.1156540>
- Lombard, A., Huyghe, R., Barque, L., & Gras, D. (2023). Regular polysemy and novel word-sense identification. *The Mental Lexicon*, 18(1). <https://doi.org/10.1075/ml.21002.lom>
- Lombard, A., Ulicheva, A., Korochkina, M., & Rastle, K. (2024). The regularity of polysemy patterns in the mind: Computational and experimental data. *Glossa Psycholinguistics*, 3(1), 1-24. <https://doi.org/10.5070/G60111327>
- Lopukhina, A., Laurinavichyute, A., Lopukhin, K., & Dragoy, O. (2018). The mental representation of polysemy across word classes. *Frontiers in Psychology*, 9, 192. <https://doi.org/10.3389/fpsyg.2018.00192>
- Luschützky, H. C., & Rainer, F. (2011). Agent-noun polysemy in a cross-linguistic perspective. *STUF - Language Typology and Universals*, 64(4), 287-338. <https://doi.org/10.1524/stuf.2011.0023>
- Luschützky, H. C., & Rainer, F. (2013). Instrument and place nouns: A typological and diachronic perspective. *Linguistics*, 51(6), 1301-1359. <https://doi.org/10.1515/ling-2013-0051>
- Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, 122(3), 485-515. <https://doi.org/10.1037/a0039267>
- Màrquez, L., Carreras, X., Litkowski, K. C., & Stevenson, S. (2008). Semantic role labeling: An introduction to the special issue. *Computational Linguistics*, 34(2), 145-159. <https://doi.org/10.1162/coli.2008.34.2.145>
- Martin, F. (2010). The semantics of eventive suffixes in French. In M. Rathert & A. Alexiadou (Éd.), *The semantics of nominalizations across languages and frameworks* (p. 109-140). De Gruyter Mouton. <https://doi.org/10.1515/9783110226546.109>
- Martin, R. (1983). *Pour une logique du sens*. Presses universitaires de France.
- Martínez Vera, G. (2021). Degree achievements and maximalization: A crosslinguistic perspective. *Glossa: A journal of general linguistics*, 6(1), 1-28. <https://doi.org/10.16995/glossa.5883>

- Marzo, D. (2015). Motivation, compositionality, idiomatization. In P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (Éd.), *Word-formation: An international handbook of the languages of Europe* (p. 984-1001). De Gruyter Mouton. <https://doi.org/10.1515/9783110246278-012>
- Masseglia, F., Teisseire, M., & Poncelet, P. (2004). Extraction de motifs séquentiels. Problèmes et méthodes. *Ingénierie des Systèmes d'Information (ISI)*, 9(3-4), 183-210. <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00108563>
- McCrae, J. P., Fransen, T., Ahmadi, S., Buitelaar, P., & Goswami, K. (2022). Towards an integrative approach for sense distinctions. *Frontiers in Artificial Intelligence*, 5. <https://doi.org/10.3389/frai.2022.745626>
- McDonald, J. (2014). *Handbook of biological statistics* (3^e éd.). Sparky House Publishing. <http://www.biostathandbook.com/index.html>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276-282. <https://doi.org/10.11613/BM.2012.031>
- Mel'čuk, I. (1993). *Cours de morphologie générale : Théorique et descriptive*. Presses de l'Université de Montréal.
- Melloni, C. (2011). *Event and result nominals. A morpho-semantic approach*. Peter Lang Verlag. <https://doi.org/10.3726/978-3-0351-0335-9>
- Melloni, C. (2015). Result nouns. In P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (Éd.), *Word-formation: An international handbook of the languages of Europe* (p. 1253-1268). De Gruyter Mouton. <https://doi.org/10.1515/9783110246278-027>
- Meyer-Lübke, W. (1890). *Italienische Grammatik*. O.R. Reisland.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1301.3781>
- Milicevic, J., & Polguère, A. (2010). Ambivalence sémantique des noms de communication langagière du français. In F. Neveu, V. Muni Toke, T. Klingler, J. Durand, L. Mondada & S. Prévost (Éd.), *CMLF 2010 - 2ème Congrès Mondial de Linguistique Française* (p. 1029-1050). Institut de Linguistique Française. <https://doi.org/10.1051/cmlf/2010102>
- Miller, G. A., & Fellbaum, C. (1991). Semantic networks of English. *Cognition*, 41(1), 197-229. [https://doi.org/10.1016/0010-0277\(91\)90036-4](https://doi.org/10.1016/0010-0277(91)90036-4)
- Missud, A. (2019). *Modélisation quantitative de la rivalité entre la suffixation en -age et la conversion de verbe à nom* (mémoire de recherche). 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, 41-68. https://www.atilf.fr/wp-content/uploads/publications/verbum/XLIII/atilf_Verbum_XLIII_02_Missud_Villoing.pdf
- Montermini, F. (2015). Using distributional data to explore derivational under-markedness: A study of the event/property polysemy in nominalization. In V. Pirrelli, C. Marzi &

Bibliographie

- M. Ferro (Éd.), *Word structure and word usage. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015*. <https://ceur-ws.org/Vol-1347/paper16.pdf>
- Müller, P. O. (2011). The polysemy of the German suffix *-er*: Aspects of its origin and development. *STUF – Language Typology and Universals*, 64(1), 33-40. <https://doi.org/10.1524/stuf.2011.0003>
- Murphy, E. (2021). *Linguistic representation and processing of copredication* (thèse de doct.). University College London. <https://discovery.ucl.ac.uk/id/eprint/10119848/>
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511780684>
- 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 (Éd.), *Competition in word-formation* (p. 1-31). John Benjamins. <https://doi.org/10.1075/la.284.01nag>
- Namer, F. (2012). Nominalisation et composition en français : d'où viennent les verbes composés ? *Lexique*, 20, 173-205.
- Namer, F., & Hathout, N. (2020). ParaDis and Démonette. From theory to resources for derivational paradigms. *The Prague Bulletin of Mathematical Linguistics*, 114(1), 5-34. <https://doi.org/10.14712/00326585.001>
- Namer, F., Hathout, N., Amiot, D., Barque, L., Bonami, O., Boyé, G., Calderone, B., Cattini, J., Dal, G., Delaporte, A., Duboisindien, G., Falaise, A., Grabar, N., Haas, P., Henry, F., Huguin, M., Juniarta, N., Liégeois, L., Lignon, S., ... Tribout, D. (2023). Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions. *Lexique*, 33, 6-40. <https://doi.org/10.54563/lexique.1242>
- 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)*. In J. Durand, B. Habert & B. Laks (Éd.), *Congrès Mondial de Linguistique Française – CMLF* (p. 1539-1557). Institut de Linguistique Française. <https://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08226.pdf>
- Nerlich, B., & Clarke, D. D. (2003). Polysemy and flexibility: Introduction and overview. In B. Nerlich, Z. Todd, V. Herman & D. D. Clarke (Éd.), *Polysemy: Flexible patterns of meaning in mind and language* (p. 3-30). De Gruyter Mouton. <https://doi.org/10.1515/9783110895698.3>
- Newman, M. (2018). *Networks* (2^e éd.). Oxford University Press. <https://doi.org/10.1093/oso/9780198805090.001.0001>
- Nicolas, D. (2002). *La distinction entre noms massifs et noms comptables*. Peeters.
- Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 3(2), 143-184. <https://www.jstor.org/stable/25001016>
- Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics*, 12(2), 109-132. <https://doi.org/10.1093/jos/12.2.109>
- Ortega-Andrés, M., & Vicente, A. (2019). Polysemy and co-predication. *Glossa: A journal of general linguistics*, 4(1). <https://doi.org/10.5334/gjgl.564>

- Osswald, R. (2005). On result nominalization in German. *Proceedings of Sinn Und Bedeutung*, 9, 256-270. <https://doi.org/10.18148/sub/2005.v9i0.732>
- Padó, S., Herbelot, A., Kisselew, M., & Šnajder, J. (2016). Predictability of distributional semantics in derivational word formation. In Y. Matsumoto & R. Prasad (Éd.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (p. 1285-1296). The COLING 2016 Organizing Committee. <https://aclanthology.org/C16-1122>
- Paducheva, E. V. (2003). Is there an "ANTICAUSATIVE" component in the semantics of de-causatives? *Journal of Slavic Linguistics*, 11(1), 173-198. <http://www.jstor.org/stable/24599713>
- Palmer, M., Gildea, D., & Xue, N. (Éd.). (2010). *Semantic role labeling*. Springer. <https://doi.org/10.1007/978-3-031-02135-0>
- Panther, K.-U. (2003). Review of "Metonymy, Grammar, and Communication". Francisco José Ruiz de Mendoza Ibáñez and José Luis Otaño Campo (eds.) *Annual Review of Cognitive Linguistics*, 1, 276-288. <https://doi.org/10.1075/arcl.1.16pan>
- Panther, K.-U., & Thornburg, L. L. (2002). The roles of metaphor and metonymy in English -er nominals. In R. Dirven & R. Pörings (Éd.), *Metaphor and metonymy in comparison and contrast* (p. 279-322). De Gruyter Mouton. <https://doi.org/10.1515/9783110219197.2.279>
- Pedersen, T. L. (2024a). *ggraph: An implementation of grammar of graphics for graphs and networks* [R package version 2.2.1.9000, <https://github.com/thomasp85/ggraph>]. <https://ggraph.data-imaginist.com>
- Pedersen, T. L. (2024b). *tidygraph: A tidy API for graph manipulation* [R package version 1.3.1.9000, <https://github.com/thomasp85/tidygraph>]. <https://tidygraph.data-imaginist.com>
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. <https://doi.org/10.3115/v1/D14-1162>
- Petukhova, V., & Bunt, H. (2019). LIRICS semantic role annotation: Design and evaluation of a set of data categories. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis & D. Tapias (Éd.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (p. 39-45). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/pdf/17_paper.pdf
- Pinker, S. (2015). *The sense of style: The thinking person's guide to writing in the 21st century*. Penguin.
- Piñón, C. (1997). Achievements in an event semantics. In A. Lawson (Éd.), *Proceedings of SALT 7* (p. 276-293). CLC Publications. <https://doi.org/10.3765/salt.v7i0.2781>
- Piñón, C. (1999). Durative adverbials for result states. In S. Bird, A. Carnie, J. D. Haugen & P. Norquest (Éd.), *WCCFL 18 Proceedings* (p. 420-433). Cascadia Press.

Bibliographie

- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter. <https://doi.org/10.1515/9783110802863>
- Plag, I., Andreou, M., & Kawaletz, L. (2018). A frame-semantic approach to polysemy in affixation. In O. Bonami, G. Boyé, G. Dal, H. Giraudo & F. Namer (Éd.), *The lexeme in descriptive and theoretical morphology* (p. 467-486). Language Science Press. <https://doi.org/10.5281/zenodo.1402520>
- Plag, I., Kawaletz, L., Arndt-Lappe, S., & Lieber, R. (2023). Analogical modeling of derivational semantics: Two case studies. In S. Kotowski & I. Plag (Éd.), *The semantics of derivational morphology: Theory, methods, evidence* (p. 103-142). De Gruyter. <https://doi.org/10.1515/9783111074917-005>
- Plénat, M. (1999). *Poissonnaille, poiscaïl (et poiscaïlle) : forme et sens des dérivés en -aille*. In M. Plénat, M. Aurnague, A. Condamines, J.-P. Maurel, C. Molinier & C. Muller (Éd.), *L'emprise du sens. Structures linguistiques et interprétation. Mélanges de syntaxe et de sémantique offerts à Andrée Borillo par un groupe d'amis, de collègues et de disciples* (p. 249-269). Brill.
- Polguère, A. (2018). A lexicographic approach to the study of copolysemy relations. *Russian Journal of Linguistics*, 22(4), 788-820. <https://doi.org/10.22363/2312-9182-2018-22-4-788-820>
- Postma, M., van Miltenburg, E., Segers, R., Schoen, A., & Vossen, P. (2016). Open Dutch WordNet. In C. Fellbaum, P. Vossen, V. B. Mititelu & C. Forascu (Éd.), *Proceedings of the 8th Global WordNet Conference (GWC)* (p. 302-310). Global Wordnet Association. <https://aclanthology.org/2016.gwc-1.43>
- Pustejovsky, J. (1995). *The generative lexicon*. MIT Press.
- Pustejovsky, J., & Batiukova, O. (2019). *The lexicon*. Cambridge University Press. <https://doi.org/10.1017/9780511982378>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienne, Autriche. <https://www.R-project.org/>
- Rainer, F. (1996). La polysémie des noms abstraits : historique et état de la question. In N. Flaux, M. Glatigny & D. Samain (Éd.), *Les noms abstraits. Histoire et théories* (p. 117-126). Presses universitaires du Septentrion.
- Rainer, F. (2014). Polysemy in derivation. In R. Lieber & P. Štekauer (Éd.), *The Oxford handbook of derivational morphology* (p. 338-353). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641642.013.0019>
- Ravin, Y., & Leacock, C. (2000). Polysemy: An overview. In Y. Ravin & C. Leacock (Éd.), *Polysemy: Theoretical and computational approaches* (p. 1-29). Oxford University Press. <https://doi.org/10.1093/oso/9780198238423.003.0001>
- Rawlings, C. M., Smith, J. A., Moody, J., & McFarland, D. A. (2023). *Network analysis: Integrating social network theory, method, and application with R*. Cambridge University Press. <https://doi.org/10.1017/9781139794985>
- Rey-Debove, J. (1985). *Le Robert méthodique : dictionnaire méthodique du français actuel*. Le Robert.

- Robins, G. (2015). *Doing social network research: Network-based research design for social scientists*. SAGE Publications Ltd.
- Roché, M. (1998). *Deux études sur la dérivation en -ier(e)*. Université de Toulouse - Le Mirail.
- Roché, M. (2009). Pour une morphologie lexicale. *La morphologie lexicale : un domaine autonome de la grammaire ?*, 65-87.
- Rothstein, S. (2004). *Structuring events: A study in the semantics of lexical aspect*. Blackwell Publishing. <https://doi.org/10.1002/9780470759127>
- Rothstein, S. (2008). Two puzzles for a theory of lexical aspect: Semelfactives and degree achievements. In J. Dölling, T. Heyde-Zybatow & M. Schäfer (Éd.), *Event structures in linguistic form and interpretation* (p. 175-198). De Gruyter. <https://doi.org/10.1515/9783110925449.175>
- Roy, I., & Soare, E. (2014). On the internal eventive properties of -er nominals. *Lingua*, 141, 139-156. <https://doi.org/10.1016/j.lingua.2013.11.006>
- Ruiz de Mendoza Ibáñez, F. J., & Otal Campo, J. L. (2002). *Metonymy, grammar, and communication*. Editorial Comares.
- Ryder, M. E. (1999). Bankers and blue-chippers: An account of -er formations in Present-day English. *English Language and Linguistics*, 3(2), 269-297. <https://doi.org/10.1017/S1360674399000246>
- Rzysmski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., Chang, S., Lai, Y., Morozova, N., Arjava, H., Hübler, N., Koile, E., Pepper, S., Proos, M., Van Epps, B., ... List, J.-M. (2020). The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific Data*, 7, 1-12. <https://doi.org/10.1038/s41597-019-0341-x>
- Salvadori, J., Barque, L., Haas, P., Huyghe, R., Lombard, A., Monney, M., Schwab, S., Tribout, D., Varvara, R., & Wauquier, M. (2023). *The semantics of deverbal nouns in French. Annotation guide*. <https://github.com/semantics-deverbal-nouns/annotation-guide>
- Salvadori, J., & Huyghe, R. (2022). When morphology meets regular polysemy. *Lexique*, 31, 85-113. <https://doi.org/10.54563/lexique.857>
- Salvadori, J., & Huyghe, R. (2023). Affix polyfunctionality in French deverbal nominalizations. *Morphology*, 33, 1-39. <https://doi.org/10.1007/s11525-022-09401-4>
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In P. Bański, H. Biber, E. Breiteneder, M. Kupietz, H. Lungen & A. Witt (Éd.), *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)* (p. 28-34). Institut für Deutsche Sprache.
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the Web using a new efficient tool chain. In N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odiijk & S. Piperidis (Éd.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (p. 486-493). European Language Resources Association.

Bibliographie

- Schneedecker, C., & Aleksandrova, A. (2016). Les noms d'humains en *-aire* : essai de classification. *Congrès Mondial de Linguistique Française - CMLF 2016*. <https://doi.org/10.1051/shsconf/20162712001>
- Schulte, M. (2015a). Polysemy and synonymy in derivational affixation—a case study of the English suffixes *-age* and *-ery*. *Morphology*, 25, 371-390. <https://doi.org/10.1007/s11525-015-9271-2>
- Schulte, M. (2015b). *The semantics of derivational morphology. A synchronic and diachronic investigation of the suffixes -age and -ery in English*. Narr.
- Scott, A. K. (2010). Accounting for the semantic extension of derived action nouns. *Journal of Linguistics*, 46(3), 711-734. <https://doi.org/10.1017/S0022226710000034>
- Shao, C., Cui, P., Xun, P., Peng, Y., & Jiang, X. (2018). Rank correlation between centrality metrics in complex networks: An empirical study. *Open Physics*, 16(1), 1009-1023. <https://doi.org/10.1515/phys-2018-0122>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Sims, A. D. (2020). Inflectional networks: Graph-theoretic tools for inflectional typology. *Society for Computation in Linguistics*, 3, 88-98. <https://doi.org/10.7275/c1f4-pg94>
- Smith, C. S. (1991). *The parameter of aspect*. Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-011-5606-6>
- Spencer, A. (2013). *Lexical relatedness: A paradigm-based model*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199679928.001.0001>
- Štekauer, P. (2014). Derivational paradigms. In R. Lieber & P. Štekauer (Éd.), *Oxford handbook of derivational morphology* (p. 354-369). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641642.013.0020>
- Štekauer, P. (2015). Backformation. In P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (Éd.), *Word-formation: An international handbook of the languages of Europe* (p. 340-352). De Gruyter Mouton. <https://doi.org/10.1515/9783110246254-020>
- Stoltz, D. S., & Taylor, M. A. (2024). *Mapping texts: Computational text analysis for the social sciences*. Oxford University Press. <https://doi.org/10.1093/oso/9780197756874.001.0001>
- Strnadová, J. (2014). *Les réseaux adjectivaux. Sur la grammaire des adjectifs dénominaux en français* (thèse de doct.). Université Paris Diderot (Paris 7) Sorbonne Paris Cité et Univerzita Karlova. <https://theses.hal.science/tel-01536100>
- Stump, G. (2019). Some sources of apparent gaps in derivational paradigms. *Morphology*, 29, 271-292. <https://doi.org/10.1007/s11525-018-9329-z>
- Stupak, I. V., & Baayen, R. H. (2022). An inquiry into the semantic transparency and productivity of German particle verbs and derivational affixation. *The Mental Lexicon*, 17(3), 422-457. <https://doi.org/10.1075/ml.22012.stu>
- Temple, M. (1991). Définir le sens d'un mot construit : application aux noms d'activité en *-erie*. *Lexique*, 10, 163-209.

- Temple, M. (1996). *Pour une sémantique des mots construits*. Presses universitaires du Septentrion. <https://doi.org/10.4000/books.septentrion.115578>
- Ten Hacken, P. (2021). Transposition and the limits of word formation. In L. Bauer, L. Körtvélyessy & P. Štekauer (Éd.), *Semantics of complex words* (p. 187-216). Springer. https://doi.org/10.1007/978-3-319-14102-2_10
- Thiele, J. (1987). *La formation des mots en français moderne*. Presses de l'Université de Montréal.
- Thornton, A. M. (2008). Italian verb-verb reduplicative action nouns. *Lingue e linguaggio*, 7(2), 209-232. <https://doi.org/10.1418/28096>
- Tjuka, A. (2024). Objects as human bodies: Cross-linguistic colexifications between words for body parts and objects. *Linguistic Typology*, version imprimée à paraître. <https://doi.org/10.1515/lingty-2023-0032>
- Tjuka, A., & List, J.-M. (2024). Partial colexifications reveal directional tendencies in object naming. *PsyArXiv*. <https://doi.org/10.31234/osf.io/hc3j5>
- Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français* (thèse de doct.). Université Paris Diderot (Paris 7).
- Tribout, D. (2012). Verbal stem space and verb to noun conversion in French. *Word Structure*, 5(1), 109-128. <https://doi.org/10.3366/word.2012.0022>
- Tufféry, S. (2012). *Data mining et statistique décisionnelle : l'intelligence des données*. Technip.
- Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, 4(3), 273-290. <https://doi.org/10.1515/cogl.1993.4.3.273>
- Uth, M. (2008). The division of the causative eventive chain by means of *-ment* and *-age*. In F. Schäfer (Éd.), *Working Papers of the SFB 732 Incremental Specification in Context 01* (p. 209-234). University of Stuttgart. <http://nbn-resolving.de/urn:nbn:de:bsz:93-opus-35566>
- Valera, S. (2020). Polysemy versus homonymy. In M. Aronoff (Éd.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.617>
- Valera, S. (2023). The semantics of noun-to-verb zero-derivation in English and Spanish. *Zeitschrift für Sprachwissenschaft*, 42(1), 153-180. <https://doi.org/10.1515/zfs-2022-2016>
- Valera, S., & Ruz, A. E. (2021). Conversion in English: Homonymy, polysemy and paronymy. *English Language and Linguistics*, 25(1), 181-204. <https://doi.org/10.1017/S1360674319000546>
- Van de Velde, D. (Éd.). (1995). *Le spectre nominal. Des noms de matières aux noms d'abstractions*. Peeters.
- Varvara, R., Lapesa, G., & Padó, S. (2021). Grounding semantic transparency in context. *Morphology*, 31, 409-446. <https://doi.org/10.1007/s11525-021-09382-w>
- Vendler, Z. (1967). *Linguistics in philosophy*. Cornell University Press. <https://doi.org/10.7591/9781501743726>

Bibliographie

- Vendryes, J. (1946). Sur le suffixe *-is* du français. In Société de publications romanes et françaises (Éd.), *Études romanes dédiées à Mario Roques par ses amis, collègues et élèves de France* (p. 103-110). E. Droz.
- Verkuyl, H. J. (1993). *A theory of aspectuality: The interaction between temporal and atemporal structure*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511597848>
- Vicente, A. (2018). Polysemy and word meaning: An account of lexical meaning for different kinds of content words. *Philosophical Studies*, 175, 947-968. <https://doi.org/10.1007/s11098-017-0900-y>
- Vicente, A. (2021). Approaches to co-predication: Inherent polysemy and metaphysical relations. *Journal of Pragmatics*, 182, 348-357. <https://doi.org/10.1016/j.pragma.2021.02.007>
- Vicente, A., & Falkum, I. L. (2017). Polysemy. In M. Aronoff (Éd.), *Oxford research encyclopedia of linguistics*. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.325>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>
- Wauquier, M. (2020). *Confrontation des procédés dérivationnels et des catégories sémantiques dans les modèles distributionnels* (thèse de doct.). Université Toulouse le Mirail - Toulouse II.
- Wauquier, M. (2022). Apports de la sémantique distributionnelle pour la morphologie dérivationnelle. *Corpus*, 23, 1-14. <https://doi.org/10.4000/corpus.6303>
- Wechsler, S. (2015). *Word meaning and syntax: Approaches to the interface*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199279883.001.0001>
- Wikimedia Foundation. (s. d.). *Wiktionnaire - Le dictionnaire libre*. fr.wiktionary.org
- Wilson, V. A. D., Zuberbühler, K., & Bickel, B. (2022). The evolutionary origins of syntax: Event cognition in nonhuman primates. *Science Advances*, 8(25), eabn8464. <https://doi.org/10.1126/sciadv.abn8464>
- Wright, A. P., Wright, A. T., McCoy, A. B., & Sittig, D. F. (2015). The use of sequential pattern mining to predict next prescribed medications. *Journal of Biomedical Informatics*, 53, 73-80. <https://doi.org/10.1016/j.jbi.2014.09.003>
- Xu, Y., Duong, K., Malt, B. C., Jiang, S., & Srinivasan, M. (2020). Conceptual relations predict colexification across languages. *Cognition*, 201, 104280. <https://doi.org/10.1016/j.cognition.2020.104280>
- Yurchenko, A., Lopukhina, A., & Dragoy, O. (2020). Metaphor is between metonymy and homonymy: Evidence from event-related potentials. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.02113>
- Zaki, M. J. (2001). SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*, 42, 31-60. <https://doi.org/10.1023/A:1007652502315>
- Zellmer, E. O. (1935). Die Wörter auf *-ure* in den französischen Mundarten des 19. Jahrhunderts. Ein Beitrag zur Wortbildungs- und Bedeutungslehre. *Zeitschrift für französische Sprache und Literatur*, 59(5/6), 271-318. <http://www.jstor.org/stable/40615546>

- Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, 33(2), 251-256. <https://doi.org/10.1080/00221309.1945.10544509>
- Zweig, K. A. (2016). *Network analysis literacy: A practical approach to the analysis of networks*. Springer. <https://doi.org/10.1007/978-3-7091-0741-6>