

Université Paris 7 Paris-Diderot
U.F.R. Linguistique
Mémoire pour l'obtention du Master 2 « Sciences du Langage –
Linguistique Théorique et Expérimentale »

Mémoire dirigé par :
Madame Heather Burnett et Monsieur Olivier Bonami

LA TRONCATION EN FRANÇAIS

RÉMI ANSELME



Soutenu à Paris, le 21 juin 2019

"GIVE ME DEMOCRACY OR GIVE ME DEATH"

Remerciements

Pour Lew Ruenn Shen.

J'apprends beaucoup des interactions que j'ai avec mon entourage. Je ne pourrais m'imaginer futur chercheur si je ne pouvais pas échanger mon point de vue d'autres personnes. Les gens qui m'entourent sont divers autant dans leur personnalité que dans leurs opinions, et c'est cette complexité dans la mixité qui m'a tant plu pendant mes années de master.

Sans les conseils et la supervision de mes co-directeurs, ce mémoire n'aurait jamais vu le jour. Je leur suis reconnaissant de la confiance qu'ils m'ont accordée. Je souhaite tout d'abord remercier Heather Burnett sans qui ce mémoire ne serait jamais né. Je la remercie d'avoir accepté de m'encadrer cette année. Je remercie aussi Olivier Bonami pour son encadrement. Le petit étudiant en L1 que j'étais n'aurait sûrement jamais cru que je collaborerais avec lui un jour. Je suis reconnaissant envers mes directeurs pour leur soutien, leurs conseils, pour avoir toujours relu ce que je leur envoyais et pour m'avoir toujours permis d'améliorer mon travail. Merci pour ces nombreux rendez-vous, ces échanges constructifs que nous avons eus et qui ont contribué à ma formation.

Je souhaite aussi remercier Ewan Dunbar et Doriane Gras qui m'ont aidé avec mes analyses statistiques. Mes remerciements vont aussi à Ioana Chitoran et les autres professeurs et chercheurs rattachés à l'UFR de Linguistique de l'université Paris-Diderot pour leurs enseignements et leur aide.

Plus généralement, je tiens à remercier toutes les personnes qui m'ont motivé cette année, qui m'ont inspiré et fait comprendre l'importance d'être toujours dans une dynamique de donner le meilleur de soi lorsqu'on fait de la recherche.

Je tiens à remercier Nika Jurov qui aura été présente à m'encourager et m'aider à me concentrer pendant ces deux années dernières années. Merci d'être une bonne collègue qui ne pas me laisse pas manger seul comme un phoque aux buffets. J'espère que par la suite nous aurons à nouveau la chance de faire des statistiques dans la joie et la ~~bonne humeur~~ le chocolat.

Je remercie aussi mes amis d'ici et d'ailleurs, Amatoullah, Angélique, Jiaqi, Piña, Titish, Yiming (et tous ceux que je ne cite pas explicitement), pour leur soutien et de toujours venir me voir s'il s'agit de continuer à parler de passé surcomposé ou de troncation. Merci pour vos nombreux exemples bizarres de mots tronqués.

Je me suis sous-estimé dans ma capacité à faire des fautes d'orthographe et de grammaire. Heureusement que mes deux locutrices natives du français préférées Laëtitia Raison-Aubry et Anna Gwenn Marie Ridez ont été là pour tout corriger. Je tiens particulièrement à les remercier pour le temps qu'elles auront passé à me relire et à s'intéresser à la troncation. S'il reste des fautes, c'est entièrement de ma faute car elles ont déjà fait beaucoup pour rendre ce mémoire propre.

Finalement, je veux remercier ma famille et leur soutien sans lequel je ne serais pas là où je suis. Merci à eux.

TABLE DES MATIÈRES

1	Introduction générale	11
2	État de l'art	14
2.1	Définition de la troncation	14
2.1.1	Présuppositions sur la troncation	15
2.2	Pourquoi la troncation ?	17
2.2.1	Les motivations pour tronquer	17
2.2.1.1	Principe du moindre effort	18
2.2.1.2	Influence de la situation de communication	20
2.2.1.3	Démarcation sociale	22
2.2.2	Comment tronquons-nous ?	22
2.2.2.1	Observations sur la structure syllabique . .	23
2.2.2.2	Les finales des mots tronqués	24
2.2.2.3	Observations sur les catégories de mots tron- cables	25
2.3	Parce que tous les noms ne sont pas tronqués	27
2.3.1	Les différentes classes de noms	28
2.3.1.1	Les noms d'événements complexes (ASN) .	28
2.3.1.2	Les noms résultatifs (RN)	29
2.3.1.3	Les noms d'événements simples (SEN) . . .	31
2.3.2	Comportement des troncats déverbaux	32

2.3.2.1	Ce que nous observons	32
2.3.2.2	Discussion	38
3	Étude quantitative des contraintes sur la troncation	39
3.1	Les contraintes sur la troncation	39
3.1.1	Ce qui est tronnable et ce qui ne l'est pas	40
3.1.1.1	Ce qui n'est pas tronnable mais l'est quand même	41
3.1.2	Étude des contraintes sur la troncation	44
3.1.2.1	Notre corpus	45
3.1.2.2	Élaboration d'une liste de troncat	46
3.1.2.3	Élaboration d'une liste de déterminants	47
3.1.2.4	Recherche sur corpus	47
3.1.2.5	Statistiques	48
3.1.3	Conclusion intermédiaire	52
4	Quel est le sens de la troncation ?	53
4.1	État de l'art	54
4.2	Étude sur les acceptions des troncats	56
4.2.1	Méthodologie de notre travail sur corpus	59
4.2.1.1	Corpus	59
4.2.1.2	Listes de sens	59
4.2.1.3	Annotation	60
4.2.2	Distribution des sens	61
4.2.2.1	Observation des distributions	61
4.2.2.2	Résultats et analyses	63
4.2.2.3	Troncats, occurrences et forums	67
4.2.3	Entropie des distributions de sens	68
4.2.4	Modèle linéaire mixte de la spécialisation des sens	73
4.2.4.1	Les données	74
4.2.5	Conclusion intermédiaire	77
4.3	Le sens de la troncation	78
4.3.1	Comparaison avec les diminutifs	79
4.3.1.1	Quand nous ajoutons au lieu d'enlever...	80
4.3.1.2	Résumé des résultats	85

5 Conclusion	86
5.1 Les règles de construction de lexèmes	86
5.2 Discussion	91
Bibliographie	93
Annexes	97
A Listes de sens	98
A.1 COMPILATION/COMPIL	98
A.2 DÉCORATION/DÉCO	99
A.3 DÉMONSTRATION/DÉMO	100
A.4 INFORMATION/INFO	101
A.5 INTOXICATION/INTOX	103
A.6 MANIFESTATION/MANIF	104
A.7 PROMOTION/PROMO	105
A.8 RÉCRÉATION/RÉCRÉ	107
A.9 RÉDACTION/RÉDAC	108
A.10 SONORISATION/SONO	109
B Listes de sens pour les diminutifs	111
B.1 GLANDER/GLANDOUILLER	111
B.2 GRATTER/GRAT(T)OULLER	112
B.3 MÂCHER/MÂCHOULLER	114
B.4 MORDRE/MORDILLER	115
B.5 SAUTER/SAUTILLER	117

TABLE DES FIGURES

2.1	Obtention de la forme réus par morphologie dérivationnelle	16
2.2	Obtention de la forme réus par morphologie flexionnelle . . .	16
2.3	Structure du nom d'événement complexe DÉCORATION	30
2.4	Structure du nom résultatif DÉCORATION	31
3.1	Proportions relatives d'utilisation des déterminants par les truncats et les non-truncats	49
3.2	Proportions relatives d'utilisation de déterminants au singu- lier et au pluriel par les bases et les truncats	50
3.3	Proportions relatives d'utilisation de déterminants défini et indéfini par les bases et les truncats	51
4.1	Distribution des sens de promotion/promo	62
4.2	Répartition du nombre de cooccurrences des truncats et bases en fonction de leur présence en forum ou dans hors forum . .	68
4.3	Modèle Linéaire Mixte pour prédire la forme d'un lexème en fonction du sens et de la formalité du contexte	76
4.4	Distribution des sens de <i>mâcher/mâchouiller</i>	83
5.1	RCL et non-spécialisation des sens	87
5.2	RCL et spécialisation des sens	88
5.3	Obtention de n lexèmes-dérivés à partir d'un lexème polysé- mique	89

5.4	RCL et distribution de probabilités des sens	91
A.1	Distribution des sens de COMPILATION et de COMPIL	99
A.2	Distribution des sens de DÉCORATION et de DÉCO	100
A.3	Distribution des sens de DÉMONSTRATION et de DÉMO	101
A.4	Distribution des sens de INFORMATION et de INFO	102
A.5	Distribution des sens de INTOXICATION et de INTOX	103
A.6	Distribution des sens de MANIFESTATION et de MANIF	105
A.7	Distribution des sens de PROMOTION et de PROMO	106
A.8	Distribution des sens de RÉCRÉATION et de RÉCRÉ	108
A.9	Distribution des sens de RÉDACTION et de RÉDAC	109
A.10	Distribution des sens de SONORISATION et de SONO	110
B.1	Distribution des sens de GLANDER et de GLANDOUILLER	112
B.2	Distribution des sens de GRATTER et de GRATOUILLER	114
B.3	Distribution des sens de MÂCHER et de MÂCHOUILLER	116
B.4	Distribution des sens de MORDRE et de MORDILLER	118
B.5	Distribution des sens de SAUTER et de SAUTILLER	122

LISTE DES TABLEAUX

2.1	Nombre de syllabes des mots-sources en français, repris de Antoine [2000]	23
2.2	Nombre de syllabes des troncats en français, repris de Antoine [2000]	24
2.3	Nombre de syllabes de ce qui est tronnable et leur proportion, repris à partir de Fridrichová [2012]	24
2.4	Nombre de troncats en fonction de leur structure syllabique, et ce qu'ils représentent, repris à partir de Fridrichová [2012]	25
2.5	Répartition des parties du discours Fridrichová [2012]	26
3.1	Liste des paires de bases-troncats utilisée	46
3.2	Liste des déterminants utilisée	47
3.3	Nombre d'occurrences et de cooccurrences des bases et troncats	48
4.1	Résultats des test d'indépendance du Chi2	66
4.2	Mesure d'entropie pour les distributions de sens	72
4.3	Compte des sens événementiels et des sens non-événementiels	74
4.4	Résultats des test d'indépendance du Chi2	84
4.5	Mesure d'entropie pour les distributions de sens	85

CHAPITRE 1

INTRODUCTION GÉNÉRALE

POURQUOI est-il important d'étudier la troncation ? Cela pourrait sembler répétitif d'entendre parler de la troncation pour celui qui souhaite lire quelque chose de nouveau sur la morphologie française. En effet, beaucoup d'auteurs se sont déjà intéressés à ce phénomène (Dister [1998], Kerleroux [1999], Antoine [2000], Fridrichová [2013], Raccah [2015], J. K. Kanwal [2018]). Tous ceux qui ont eu un intérêt pour troncation ont cherché à répondre à des questions particulières. Certains auteurs se sont demandés quels étaient les mots tronqués et leurs caractéristiques. D'autres auteurs se sont penchés sur les motivations de la troncation, ou encore comment le phénomène était perçu dans la société (Fridrichová [2012]).

De ce fait, ce qui, dans un premier temps, nous a poussé à faire des recherches sur le phénomène c'est notre curiosité pour le profil des locuteurs qui rendent ce phénomène productif. Dans un second temps, nos recherches ont été poussées par l'envie de caractériser l'apport de la troncation pour la linguistique à travers une approche de morphologie quantitative.

Par troncation d'un mot, il est compris que nous référons au processus qui permet de passer d'un mot à forme longue à un mot à forme courte. Il est possible de se placer à plusieurs niveaux d'analyse. Les linguistes se sont particulièrement intéressés à la partie perdue du mot. D'autres caractéris-

tiques, intrinsèques au mot, comme le sens, sont également intéressantes à observer.

Comme les mots ne sont pas isolés dans le discours de celui qui parle, il est aussi légitime de regarder le contexte dans lequel ces mots sont produits. Ce concept étant souvent mis de côté, nous avons trouvé intéressant l'étudier.

Ce qui nous a paru curieux avec la troncation et qui continuera sûrement à nous fasciner, c'est de voir à quel point le phénomène est productif.

▷ « On fera ça ens. » [ɔ̃fɔʁasãsə]

Quand une amie rapporta que sa soeur avait employé le mot « ens » à l'oral, quelle ne fut pas pour nous la difficulté à retrouver le mot d'origine sous-jacent. « ens » était utilisé pour l'adverbe « ensemble ». Une des questions qui peut émerger lorsque nous entendons un mot en français pour la première fois est quel est le sens que nous lui rattachons.

▷ « J'ai une réu à 14h. »

Une seconde question émerge, concernant le type de personne utilisant ce genre de mot raccourci. Qui dit « réu » pour « réunion » ?

Dans les deux exemples que nous venons de donner, les formes nous étaient inconnues. Pourtant la forme « réu » une fois découverte nous a directement évoqué un type de personne en particulier. En demandant l'opinion de notre entourage, ce type de locuteurs a été subjectivement associé aux personnes que nous pouvons retrouver travaillant en start-up ou en communication/marketing. Ainsi, nous voyons ces formes comme informatives sur notre identité et nous détaillerons ce point de vue dans le mémoire.

Ce qui est tronçable est donc variable, et cette variation, nous pensons, peut être caractérisée au niveau de l'individu et/ou du groupe puisque toutes les personnes n'ont pas accès à la même information qui permet de tronquer. Pourtant, comme le français est productif vis-à-vis des mots tronqués, cela met en évidence qu'il doit exister une caractéristique de la langue qui permet de tronquer. Cette évidence doit se différencier des motivations de la troncation. Nous allons développer cette idée au cours de ce mémoire.

Ce mémoire a donc pour objectif de revenir sur ce qui a été dit sur la troncation car il est toujours intéressant de comprendre ce qui a pu être observé par d'autres linguistes ainsi que leur méthodologie. Nous avons également pour objectif d'apporter de nouveaux éléments de compréhension de la troncation à travers une quantification du phénomène.

Après cette brève introduction générale, notre premier chapitre sera dédié à l'état de l'art concernant la troncation. Nous verrons comment les locuteurs utilisent le processus morphologique et quelles sont leurs motivations pour tronquer. Nous aborderons la question des différentes catégories de noms déverbaux qui ont été proposés par la littérature. Une catégorie particulière retiendra notre attention, celle des noms d'événements complexes. Il s'agit d'un type de noms où l'action est vue dans son déroulement. A cause de cette caractéristique, ces noms sont dits non-troncables. Nous observerons leur comportement à partir d'un exemple.

Nous poursuivrons notre état de l'art par une première étude de corpus visant à mettre en avant les contraintes sur la troncation. Ces restrictions ont des conséquences syntaxiques quantifiables. Ainsi nous allons porter notre intérêt sur la définitude et le nombre des déterminants précédant les noms ainsi que les prépositions en *de* suivant ces noms, trois caractéristiques impliquées dans la distribution des noms d'événements complexes. Nous montrerons que des contraintes syntaxiques existent et qu'elles jouent un rôle dans la distribution des formes tronquées.

Nous accompagnerons nos observations de ces contraintes par une étude de corpus sur les acceptions des noms déverbaux. Nous reviendrons sur les contraintes sémantiques proposées par la littérature et plus particulièrement sur la réduction affirmée des sens des mots lors de la troncation. Nous mettrons nos résultats en parallèle avec les résultats d'une étude sur les modificateurs verbaux permettant la formation de diminutifs. Les résultats étant similaires nous conclurons sur les contraintes sémantiques des processus impliquant une dérivation, en spécifiant les caractéristiques propres à chaque processus pour former de nouveaux mots.

CHAPITRE 2

ÉTAT DE L'ART

2.1 Définition de la troncation

AVANT DE SE LANCER dans une revue de littérature sur ce qui a été dit sur la troncation, il est déjà important de savoir définir ce terme et ce qu'il englobe.

Dans le présent mémoire, nous allons nous référer à la troncation comme décrite dans le Trésor de la Langue Française Informatisé (TLFI), c'est-à-dire comme un « procédé d'abrègement des mots polysyllabiques qui consiste à supprimer une ou plusieurs syllabes à l'initiale ou, plus souvent, à la finale ». Cette définition n'est qu'un point de départ et n'engage pas encore sur l'intérêt du processus en soi qui, nous le verrons, n'est pas juste un « abrègement ». Supprimer une ou plusieurs syllabes en début de mot porte le nom d'aphérèse comme en (1) où la forme *bleme* est utilisée pour la forme *problème* et en fin de mot le nom d'apocope comme en (2) où la forme *réus* est employée pour la forme *réunions*. Lorsque la troncation a lieu au milieu du mot, on parle alors de syncope. Dans ce mémoire nous référerons uniquement au processus d'apocope quand nous emploierons le terme *troncation*, et laisserons de côté l'aphérèse et la syncope.

— Aphérèse :

- (1) « slt voila j'ai un *bleme* je fait du karate . quand j'ai demenagé je suis allé dans un autre club et je suis la seule fille jusque la il y avait pas de probleme [...] »¹

— Apocope :

- (2) « Salut vous tous, je vous souhaite une belle journée et d'être bien en vous même. Avec les *réus* je peux apprendre à être mieux et plus souvent en moi pour exprimer ce qu'il y au fond de mon coeur. »²

2.1.1 Présuppositions sur la troncation

Pour comprendre la démarche de ce mémoire, il est important de préciser qu'à ce stade nous utiliserons le terme de « lexème » de manière générique pour se référer, d'un côté, aux formes sources et, de l'autre côté, aux formes résultantes de la troncation. Dans le domaine de la morphologie, le terme *lexème* est employé pour se référer à une collection abstraite de mots qui partagent un sens lexical, c'est ce sens qui est décrit par les dictionnaires. Le problème qui se pose est de savoir s'il faut considérer la troncation comme un processus flexionnel, comme défini en (3), ou comme un processus dérivationnel, comme défini en (4). Nous montrerons que la dénomination de « lexème » est justifiée, et qu'il ne faut pas lui préférer le terme de « mot » dénotant la flexion. En effet, la forme raccourcie doit être considérée comme un lexème distinct de la forme non-raccourcie, car elle dispose d'une autonomie qui est propre aux lexèmes.

- (3) ***Inflection*** : the relationship between word-forms of a lexeme ; a part of morphology that is characterized by relatively abstract morphological meanings, semantic regularity, almost unlimited applicability, etc. (Haspelmath & Sims [2013])

1. « Blème » : [consulté le 24 mars 2019]. http://forum.ados.fr/love/Amour/bleme-sujet_24513_1.htm

2. « Réu » : [consulté le 24 mars 2019]. <https://www.forum-na.fr/viewtopic.php?pid=7841>

- (4) **Derivation** : the relationship between lexemes of a word family ; a part of morphology that is characterized by relatively concrete morphological meanings, potential semantic irregularity, restrictions on applicability, etc. (Haspelmath & Sims [2013])

Reprenons l'exemple en (2) de **réus**. D'un côté, nous avons l'approche dérivationnelle qui a permis d'aboutir à ce mot que nous représentons en FIGURE 2.1 où les lexèmes RÉUNION et RÉU sont reliés directement. D'un autre côté, nous avons l'approche flexionnelle permettant d'aboutir par un moyen flexionnel, au travers du mot *réunions*, à **réus**, que nous représentons en FIGURE 2.2.

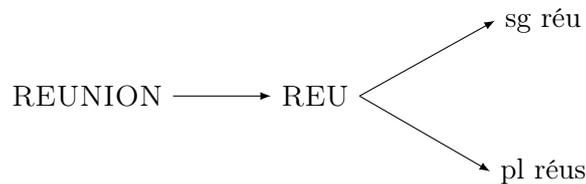


FIGURE 2.1 – Obtention de la forme **réus** par morphologie dérivationnelle

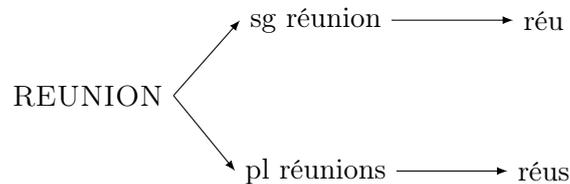


FIGURE 2.2 – Obtention de la forme **réus** par morphologie flexionnelle

Comme nous nous intéressons à un niveau qui est de l'ordre du niveau lexémique (Fradin & Kerleroux [2009]), il nous est important de comprendre si la troncation crée un nouveau lexème. Nous allons développer dans ce mémoire des arguments allant dans ce sens.

Avant de continuer, il est aussi important de porter l'attention du lecteur sur un terme que nous utiliserons très fréquemment et qui est celui de « **troncat** » que nous empruntons à Antoine [2000]. Contrairement à ce que nous pourrions penser, le terme troncat n'est pas une forme tronquée de

troncation mais bien un nom déverbal de TRONQUER³. Le troncat peut donc se définir comme le résultant de la troncation. Les « formes tronquées d'un mot » et « troncats » font référence au même concept. En opposition au terme troncat, nous utiliserons plusieurs expressions dont « base », « forme non-tronquée », « non-troncat », « forme tronnable » ou encore « forme source » pour parler du mot qui peut subir la troncation mais n'est pas encore tronqué, sans pour autant faire d'hypothèse sur la nature du processus morphophonologique qu'est la troncation.

2.2 Pourquoi la troncation ?

2.2.1 Les motivations pour tronquer

La littérature sur la troncation a essayé de trouver des motivations pour justifier le fait qu'un tel processus morphophonologique puisse exister. En effet, il pourrait sembler facile de ne décrire le phénomène que d'un point de vue formel en ne tenant pas compte du point de vue des locuteurs, mais pour autant il y aurait là une perte informationnelle. Il est donc pertinent et important de regarder du côté du comportement des utilisateurs par rapport au raccourcissement des unités lexicales avant de décrire les implications qu'a la troncation pour la linguistique.

Nous avons trouvé dans la littérature trois principales hypothèses qui pousseraient les locuteurs à tronquer. Ces hypothèses se basent sur l'idée que la troncation agit comme une fonction identité et donc qu'elle ne change pas le contenu descriptif du lexème. La troncation est alors vue comme une opération changeant la forme mais pas le sens.

La première hypothèse dont nous allons parler est celle qui se rattache au principe du moindre effort (Zipf [1949]). Le moindre effort se réalise dans la troncation par le fait d'utiliser des formes plus courtes là où il existe une variante plus longue, le mot d'origine, le mot-base. Il a été supposé que plus un mot est fréquent, plus ce mot a de chances d'être tronqué (J. K. Kanwal [2018]). Cette première hypothèse est centrale dans la mesure où elle

3. De manière analogique, nous avons *agrégat* qui dérive de *agréger*. Voir Bonami et al. [2009] pour une discussion plus approfondie du processus dérivationnel sous-jacent.

touche directement le comportement interne des locuteurs et peut être mise en évidence dans des phénomènes qui ne sont pas directement liés au langage. Cela démontre que la paresse est centrale dans notre comportement dès lors que nous essayons d’optimiser notre effort langagier. La deuxième hypothèse est celle de l’accélération de la communication, autrement dit faire vite donc court. Une personne qui tronque aura tendance à tronquer certains mots dans des situations où il faut parler vite, par exemple pour transmettre de l’information en un minimum de temps, et où le sens des formes tronquées est partagé par l’ensemble des interlocuteurs. Dans ce cas, ces formes tronquées appartiennent souvent à des jargons. La troisième hypothèse est celle de la démarcation sociale. Les troncats sont souvent connotés comme appartenant à des groupes sociaux spécifiques car appartenant à des argots spécifiques. Ces formes servent à véhiculer, en plus du sens lexical du mot, des propriétés sociolinguistiques qui sont celles du groupe de personnes qui utilise les formes tronquées. Cette troisième raison se distingue de la deuxième car un nouveau sens est ajouté avec la troncation, le sens social⁴.

Dans les sous-sections qui suivent nous allons développer plus explicitement ces trois principales hypothèses fondatrices pour comprendre le comportement des locuteurs vis-à-vis de la troncation.

2.2.1.1 Principe du moindre effort

Plus un mot est fréquent plus ce dernier a tendance à être court. Quand il est possible d’utiliser une forme courte, le locuteur préférera toujours utiliser cette forme par rapport à une forme analogue plus longue. Ces deux simples phrases sont celles qui nous permettent de résumer au mieux les conclusions des travaux de Zipf du deuxième quart du XXème siècle et permettent d’introduire le principe du moindre effort (Zipf [1949]).

Avant d’évoquer le principe du moindre effort, nous devons revenir sur la loi de l’abréviation. Il s’agit de l’une des deux lois de Zipf, mais c’est

4. Nous renvoyons le lecteur à la lecture de Ochs [1993] sur le lien entre la construction de l’identité sociale et le sens social.

surement la plus connue, la deuxième étant celle du principe du moindre effort.

La loi de l'abréviation stipule qu'il existe une relation inverse entre la fréquence d'un mot et sa longueur. Ainsi les mots les plus longs sont les moins fréquents, et les mots les plus courts sont les plus fréquents. Cette relation inverse a été démontrée dans nombre de langues bien qu'il a été montré que le contenu informationnel pouvait être un meilleur indicateur pour la longueur d'un mot (Piantadosi et al. [2011]). Ainsi, par exemple, un locuteur du français s'exprimant à l'oral dans un discours normal utilisera beaucoup de déterminants comme *le* ou *la*, de prépositions comme *à* ou *de* mais très peu de noms comme *triskaïdékaphobie*, la peur du nombre 13. D'un côté nous avons des prépositions et des déterminants courts qui sont fréquents et de l'autre des noms qui sont longs et moins fréquents. Mais même pour des mots courts, il est possible de retrouver cette loi de l'abréviation à travers la forme phonétique des mots. Les mots monosyllabiques auront tendance à avoir une réalisation phonétique plus brève, de l'ordre de quelques millisecondes lorsqu'ils sont fréquents, par rapport aux mots monosyllabiques non fréquents (Wright [1979]).

Cette loi de l'abréviation est une caractéristique générale des langues et va au delà du langage humain. En effet, chez certaines espèces animales, comme les primates ou les chauves-souris par exemple (Semple et al. [2010], Luo et al. [2013]), les signaux des cris des individus se trouvent raccourcis lorsqu'ils sont fréquents.

Pourtant si cette loi de l'abréviation existe, il ne faut pas tomber dans l'ambiguïté qui serait causée par l'utilisation de formes courtes qui se ressemblent trop. La paresse, vue comme une force façonnant le langage, veut que nous produisions des mots moins longs mais aussi moins de mots pour communiquer. Ainsi émerge le principe du moindre effort. Pour éviter les ambiguïtés et rester optimal, le locuteur a tendance à associer pour deux formes, une forme courte et une forme longue, le concept le plus saillant à la forme courte et le concept le moins saillant à la forme longue. Les mots les plus visibles dans le discours sont les plus prévisibles et donc peuvent être raccourcis en opposition aux mots les moins visibles qui doivent être pro-

noncés avec effort (J. Kanwal et al. [2017]). Ainsi il a été postulé que l'information devait rester constante pendant le discours (Frank & Jaeger [2008]), il est moins coûteux de réduire la forme d'un mot long qui ne transmet que peu d'information par rapport à un mot long qui en transmet beaucoup. Tous les mots n'apportent donc pas la même quantité d'information et c'est pour cela que les mots les moins prévisibles, les plus surprenants doivent être articulés dans des contextes où leur information sera préservée. Il est donc possible de calculer la probabilité en contexte des mots. Les mots qui auront une fois de plus été mentionnés beaucoup de fois pendant le discours, seront plus prévisibles ainsi que les mots qui leur sont sémantiquement reliés, et donc il leur sera préférée une forme moins articulée (Aylett & Turk [2004]). Ainsi, les locuteurs, en tronquant, gardent l'information distribuée uniformément dans le discours et celle associée au mot, mais économisent l'énergie associée à l'effort articulatoire de ce qui a été tronqué.

Le principe du moindre effort est par conséquent un outil pour comprendre pourquoi le phénomène de la troncation existe, et qu'il n'est pas limité à la langue française. Il est par exemple retrouvé en anglais où les troncats sont aussi présents (Jamet [2009]). Nous devons prendre en compte que les locuteurs ont à leur disposition deux formes pour désigner un même concept : une forme pleine étant le mot de base long, et le mot tronqué étant le troncat court. Il en ressort donc naturellement que les locuteurs vont privilégier l'utilisation de troncats lorsque le concept rattaché au mot est prévisible dans le contexte, le mot « simple » est préféré, et l'utilisation de la forme longue lorsque que le concept vient d'être introduit dans le discours, puisque l'information qu'il apporte est nouvelle. Dans ce second cas, la quantité d'information transmise par le mot est trop importante pour qu'il puisse être tronqué.

2.2.1.2 Influence de la situation de communication

Antoine [2000] souligne le fait que les locuteurs peuvent avoir besoin de communiquer rapidement dans certains contextes. Il faut prendre en compte que les locuteurs ne communiquent pas avec n'importe qui. Ils sont limités, lors de la communication, à un cercle de personnes de leur entourage. Cet

entourage peut être, par exemple, professionnel. A chaque milieu, il est possible d'associer un vocabulaire spécifique que nous appelons « jargon » dont les termes sont moins fréquents en dehors du milieu dans lequel ils évoluent. Dans le monde médical, par exemple, une chirurgienne peut utiliser les mots *perf* pour parler de *perfusion*, *réa* pour *réanimation* ou encore *staphylo* pour parler de la bactérie staphylocoque. Elle utilisera ces termes avec ses collègues, son entourage médical, et les gens travaillant avec elle comprendront sans problème ce à quoi elle fait référence et pourront lui répondre par une forme courte aussi. La communication sera réussie, efficace et rapide.

Pourtant, cette accélération de la communication à travers des formes courtes ne trouvera pas de motivation lorsque la chirurgienne parlera à un patient pour annoncer qu'il a une infection due à un staphylocoque. En effet, le patient n'est pas familiarisé avec le jargon du monde médical et donc chaque mot de ce jargon est très informatif. La chirurgienne devra donc éviter les troncats pour se faire comprendre.

Il n'est pas rare de trouver dans de nombreux domaines techniques beaucoup de troncats qui sont compris seulement par les personnes du métier. Ainsi, dans le domaine linguistique, il n'est pas rare d'utiliser des troncats. *Phono* peut être utilisé pour parler de *phonologie*. Mais dès lors qu'il s'agira d'expliquer la linguistique à une personne novice, la forme complète *phonologie* sera préférée car pour bien expliquer il faut prendre son temps.

La troncation, étant un type particulier d'abréviation, est majoritairement associée à des contextes où la communication a un coût. Par exemple, lors de la prise de notes, écrire un mot long est plus coûteux en temps. Le temps investi pour écrire est du temps qui n'est pas investi dans l'attention portée à l'interlocuteur. Sur le réseau social Twitter, il faut être également capable d'écrire en un minimum de caractères. Donc, la tergiversation n'a pas sa place et les abréviations peuvent être de mise.

Les formes tronquées peuvent ensuite être exportées dans le langage courant et une des motivations à leur utilisation sera le principe du moindre effort.

2.2.1.3 Démarcation sociale

En parcourant le dictionnaire d'Antoine sur les troncats, nous réalisons que beaucoup de troncats n'appartiennent, non pas à des jargons mais bien à des argots. Antoine argumente pour une nouvelle motivation pour la troncation. Par la troncation, le locuteur peut se démarquer socialement. De cette façon les groupes sociaux utilisent certaines formes qui ne seront pas comprises par des personnes extérieures au groupe. Le but est double. D'un côté il faut marquer son appartenance, et d'un autre il faut exclure celui qui ne ressemble pas. Il y a un cryptage de la communication qui s'opère à travers une opacification des formes. C'est ainsi que des mots passés dans le langage courant sont verlanisés puis tronqués pour en donner de nouvelles formes comme *pédé* qui donne *dépé* qui donnera *dèp*⁵, et donc la dérivation n'est pas directe par rapport à *pédéastre*. Ainsi pour Antoine « l'utilisateur de l'argot s'efforce [...] toujours de garder une avance sur la langue familière. ». L'emploi de certains suffixes, parfois arbitraire, après la troncation permet d'obtenir le même effet en cachant la forme d'origine du mot comme par exemple *dirlingue* qui remplace *directeur*.

Il est aussi possible pour les locuteurs, une fois qu'il ont compris comment utiliser les troncats qui appartiennent aux jargons ou aux argots, de les utiliser et de se les attribuer pour s'intégrer, ainsi le troncat sert à exprimer un « lien social ». Ce sens social est important car il signifie que les locuteurs qui partagent ces formes ont quelque chose en commun, et ce commun se reflète dans la manière dont ils parlent tout en évoluant avec le groupe social. Malheureusement, cet aspect ne pas étudié dans la suite de ce mémoire.

2.2.2 Comment tronquons-nous ?

Nous avons vu qu'il était possible de catégoriser les motivations de la troncation en fonction du comportement des locuteurs. Mais nous n'avons

5. a. La forme *dèp* peut aussi être analysée comme provenant de *pédale* par troncation puis vernalisation. *pédale* dérive de *pédéastre* par troncation puis suffixation.
b. *dep* peut aussi vu comme l'apocope de *dépanneur*. Au Québec cette forme est courante.
c. FUN FACT : "dèp" signifiant *beau* en vietnamien, voyez cela comme un compliment si on vous traite de *dèp*.

pas abordé, si ce n'est rapidement survolé, les stratégies utilisées par les locuteurs pour tronquer.

Nous retrouvons très tôt dans la littérature que chaque langue utilise de préférence une forme minimale pour former ses mots, et cette forme minimale correspond à un mot prosodique minimal (McCarthy & Prince [1996], Plénat [1994] cité par Kerleroux [1999]). Ce mot peut être vu comme une forme optimale. En français, il s'agit d'un mot dissyllabique, autrement dit les mots à deux syllabes sont privilégiés par la langue française. Or beaucoup de mots en français font plus de deux syllabes. Comme le mot dissyllabique est privilégié par la langue, il est avec la troncation un moyen de passer d'un mot de plusieurs syllabes à un mot à deux syllabes.

2.2.2.1 Observations sur la structure syllabique

Antoine puis Fridrichová [2012], dans sa thèse, ont essayé de quantifier le nombre de syllabes résultant après la troncation du troncat. D'un côté Antoine avec sa liste composée de 911 troncats résultant des différents phénomènes de troncation, a mis en évidence que plus de la moitié de ses mots-sources possédait une structure syllabique à plus de deux syllabes (TABLEAU 2.1) et après troncation, les troncats dissyllabiques représentaient à eux seuls 61,6% de son corpus (TABLEAU 2.2).

Nombre de syllabes	1	2	3	4	5	6	7	> 7
Nombre d'items	4	224	329	254	76	19	3	2
%	0,4	24,6	36,1	27,9	8,3	2,1	0,3	0,2

TABLE 2.1 – Nombre de syllabes des mots-sources en français, repris de Antoine [2000]

Ces résultats sont confirmés par les résultats de Fridrichová qui tire ses troncats d'une analyse de corpus issue de la revue de journaux francophones variés. L'étude de la linguiste repose sur huit journaux à public différent dont 20 minutes, Cosmopolitan, Inrockuptibles ou Le Nouvel Observateur, ce qui lui permet d'avoir un échantillon et un lexique plus diversifiés, plus complexes.

Syllabes	1	2	3	4 et +
Nombre d'items	200	561	140	10
%	21,9	61,6	15,4	1,1

TABLE 2.2 – Nombre de syllabes des troncats en français, repris de Antoine [2000]

Nombre de syllabes	1	2	3	4	5	6	7
Nombre d'occurrences	12	84	191	203	76	16	4
%	2	14	32	35	13	3	1

TABLE 2.3 – Nombre de syllabes de ce qui est tronnable et leur proportion, repris à partir de Fridrichová [2012]

Ces différents résultats nous permettent d'observer qu'à plus de douze ans d'intervalle, les deux linguistes obtiennent des résultats similaires qui vont dans le sens d'une stabilité des formes à deux syllabes. En effet, il est possible de supposer que le vocabulaire couvert par Fridrichová a été couvert par le dictionnaire d'Antoine bien qu'il soit antérieur. Le vocabulaire d'une langue n'évolue que très peu au cours d'une période de temps si courte, ce dernier étant soumis à des forces dont les manifestations ne sont visibles qu'à travers différentes générations (Smith [2004]).

2.2.2.2 Les finales des mots tronqués

Il est aussi intéressant de se pencher sur les finales des troncats. Premièrement, les finales vocaliques, toutes catégories confondues, sont les plus nombreuses avec 65% des troncats. Parmi les finales consonantiques (représentant donc 35%), le phonème /k/ est le plus représenté, comme dans /kolok/ pour *coloc* de *colocation*. En ce qui concerne les finales vocaliques, c'est le phonème /o/ qui est le plus représenté comme dans /eko/ pour *éco* de *écologie*.

Ce qui fait la caractéristique des troncats c'est le fait que la plupart du temps le mot tronqué peut être suffixé ou redupliqué directement après la

Syllabes	1	2	3	4
Nombre d'occurrences	143	348	86	9
%	24	59	15	2

TABLE 2.4 – Nombre de troncats en fonction de leur structure syllabique, et ce qu'ils représentent, repris à partir de Fridrichová [2012]

troncation. Ainsi, il est possible d'obtenir *cinoche* à partir de *cin-* issu de *cinéma* ensuite suffixé par *-oche*, ou *joujou* obtenu par redoublement hypocoristique à partir de *jou* forme résultante de la troncation de *jouet*. Les rédupliques peuvent être connotés comme appartenant au langage enfantin (Fridrichová [2012], Plénat [1999]) ou à des argots particuliers. Il est possible de se demander si les noms hypocoristiques de personnes doivent être vus comme présentant un comportement similaire. L'obtention de *Nini* à partir de *Nika*⁶ doit-elle être vue comme issue d'une apocope puis d'un redoublement hypocoristique? Plénat [1999] argumente que la formation des hypocoristiques, résulte juste dans ce cas présent, de la copie du matériel segmental du nom, souvent la première syllabe, ce qui ne devrait pas être vu comme un troncat à proprement parlé.

Raccah [2015] met en évidence en travaillant sur la troncation que nombre de troncats se terminent par la voyelle *-o* de manière naturelle ou bien par suffixation. Il en postule un effet sémantique associé à cette voyelle pour la troncation.

2.2.2.3 Observations sur les catégories de mots troncables

Finalement, nous allons regarder quelles parties du discours sont tronquées. Celui qui ne connaît pas l'état de l'art de ce qui se dit sur la troncation pensera sûrement que le troncat ne peut être qu'un nom ou un adjectif. Et cette personne n'aura pas tort. Les noms (ou substantifs) ainsi que les ad-

6. Nous avons eu d'ailleurs une conversation intéressante avec une locutrice native du Slovène dont le nom *Nika* est issu. Il se trouve que la formation d'un nom hypocoristique n'y est pas la même puisque *Nini* ne fera jamais référence à *Nika* mais bien à *Nina*, là où en français cela est possible.

jectifs sont les plus représentés parmi les troncats. En effet, Fridrichová a étudié la répartition des parties du discours. Nous reprenons ses résultats dans le TABLEAU 2.5.

Partie du discours	Nombre	%
Substantifs	351	59,49
Adjectifs	134	22,71
Noms propres	86	14,58
Adverbes	14	2,37
Interjections	2	0,34
Verbes	2	0,34
Préposition	1	0,17

TABLE 2.5 – Répartition des parties du discours Fridrichová [2012]

De ce tableau, nous remarquons que de sa liste, 75% des troncats sont des noms incluant substantifs et noms propres. 23% sont des adjectifs. Nous attirons l'attention du lecteur sur les pourcentages restant répartis entre les adverbes, les interjections, les verbes et les prépositions. En effet, il semble possible de tronquer tout mot dont la longueur serait un obstacle à sa prononciation et/ou son écriture. Une étude qualitative sur les troncats verbaux que nous avons rapidement réalisée à partir de l'outil recherche de Google et de Twitter, nous a révélé que les verbes sont troncables en fréquence plus importante que celle spécifiée par Fridrichová lorsque leur longueur pose problème. Ainsi, il est possible de retrouver *déj* pour le verbe *déjeuner*. Nous proposons néanmoins que la troncation d'une forme verbale n'est possible que si un nom déverbal associé à ce verbe existe et est lui-même tronnable. En effet, il est à considérer la possibilité de la troncation de DÉJEUNER.VERBE en *déj* car la forme *déj* existe déjà pour DÉJEUNER.NOM. Ceci n'est qu'une hypothèse mais elle serait intéressante à développer.

Dans la suite de ce mémoire, nous allons nous intéresser principalement aux substantifs car c'est à propos d'eux que la littérature scientifique a le plus parlé.

2.3 Parce que tous les noms ne sont pas tronqués

La troncation est un phénomène productif en français. C'est sûrement pour cela que tout paraît tronnable parmi les noms. Mais pourtant à l'intérieur de la catégorie des noms, il a été dit que tout n'est pas tronnable (Kerleroux [1999]). Avant de réfléchir davantage sur les caractéristiques de ce qui n'est pas tronnable et afin de mieux comprendre ce qui peut être tronqué, nous devons comprendre quelles classes de noms ont été proposées, et quelle est la justification sous-jacente derrière cette classification. Ainsi nombre de linguistes s'étant intéressés aux noms déverbaux ont essayé de les regrouper dans différentes catégories pour justifier les contrastes qui pouvaient s'observer avec leur usage. Nous allons particulièrement examiner les noms déverbaux en *-tion* car c'est un genre de noms qui est largement repris par la littérature sur la troncation et qui être vu comme possédant plusieurs interprétations.

Les trois classes qui ont été proposées par Grimshaw [1990] sont celles des « noms d'événements complexes » (Borer [2003] s'y réfère comme les "Argument-Structure Nominals" abrégés en ASN), des « noms d'événements simples » (SEN) et celle des « noms résultatifs » (RN). Les deux dernières classes peuvent être mises en opposition face à la classe des noms d'événements complexes. La structure argumentale de ces noms est un facteur important qui a poussé Grimshaw à opérer avec cette distinction⁷. Après avoir caractérisé ces trois classes de noms à travers leurs différentes caractéristiques et leurs différentes sensibilités vis-à-vis des tests linguistiques qui ont été proposés nous allons directement les mettre en relation avec la troncation avec comme attente que les noms d'événements complexes ne sont pas tronlables.

7. Il existe dans la littérature un autre point de vue développé à propos des différentes classes de noms. Nous renvoyons le lecteur à l'article de Grimm & McNally [2013] pour une approche basée sur le corpus. Les résultats suggèrent que les arguments obligatoires ne sont pas une condition suffisante pour catégoriser les noms. Cela a pour effet de placer les noms sur un continuum et d'abandonner la distinction de catégories proposée par Grimshaw.

2.3.1 Les différentes classes de noms

2.3.1.1 Les noms d'événements complexes (ASN)

Les noms d'événements complexes correspondent à une première classe de noms. Nous donnons un exemple en (5) avec le nom *destruction* qui est un nom d'événement complexe.

- (5) La *destruction* de l'immeuble par les ingénieurs

La première caractéristique de ces noms est qu'ils sont associés à une structure argumentale obligatoire, autrement dit ce type de noms doit prendre un complément prépositionnel en *de* obligatoire et de manière facultative un complément en *par*. L'obligation de la structure argumentale est vue comme une manifestation du caractère événementiel du nom. Nous énumérons ci-dessous les caractéristiques qui ont été proposées.

- (6) Les noms d'événements complexes
- a. Ils sont assignateurs de rôle θ . Ces rôles thématiques sont vus comme des étiquettes abstraites caractérisant la relation sémantique entre le nom et ses arguments (Saint-Dizier [2006]). Les arguments du nom d'événement complexe sont obligatoires.
 - b. Ils dénotent des événements.
 - c. Ils sont compatibles avec un modifieur orienté-agent (par exemple, ils sont compatibles avec *intentionnel* ou *délibéré*).
 - d. Ils peuvent avoir un complément en *par* qui est un argument facultatif et non un adjectif. La présence du *par* implique la présence du *de*.
 - e. Ils sont compatibles au singulier et seulement au singulier avec modifieurs comme *fréquent* ou *constant*.
 - f. Ils sont compatibles avec des modifieurs aspectuels comme *en une heure*.
 - g. Ce sont des noms massiques.

Par exemple, si nous reprenons l'exemple en (5), nous aurions :

- (7) La *destruction* de l'immeuble par les ingénieurs :
- a. La destruction *(de l'immeuble) *(par les ingénieurs)

- b. La destruction délibérée de l'immeuble par les ingénieurs
- c. La constante destruction de l'immeuble par les ingénieurs
- d. * Les constantes destructions de l'immeuble par les ingénieurs
- e. La destruction en une heure de l'immeuble par les ingénieurs

Borer [2003] a cherché à mettre en avant que la structure interne des noms déverbaux ASN dérivait du verbe à partir duquel les noms sont formés. Il s'agit d'une approche dirigée par la syntaxe. De même que les verbes doivent sélectionner des arguments, les ASN sont vus comme des noms prédictifs tenus d'en sélectionner. Le point commun est qu'il est supposé que le verbe et le nom ont une même entrée lexicale. Ainsi pour Borer, il est possible de retrouver une fonction verbalisante qui est associée avec la structure argumentale de l'ASN et lui permet d'assigner des rôles thématiques à ses arguments. L'interprétation du nom déverbal en tant que nom d'événement complexe n'est pas due à son affixe obtenu par dérivation, mais est bien acquis dans sa projection structurale interne à travers la tête fonctionnelle Asp_{EV} (Aspect of Event) qui est remplie par le suffixe comme représenté dans l'arbre en FIGURE 2.3 pour *décoration* adapté de Roy & Soare [2013]. Asp_Q (Aspect of Quantity) est la tête utilisée par l'argument obligatoire du nom.

2.3.1.2 Les noms résultatifs (RN)

Il a été tout d'abord opposé aux noms d'événements complexes, les noms résultatifs qui n'impliquent pas de participants dans leur interprétation. Pour cela nous en discuterons avant de parler des noms d'événements simples. Ces noms résultatifs ont un comportement opposé à celui des ASN ce qui a justifié dans la littérature un traitement séparé des ASN et donc un début de classification. En effet, contrairement aux ASN, ils n'ont pas de structure argumentale obligatoire, autrement dit, s'ils comportent des compléments, il s'agit de compléments facultatifs. Ainsi, s'ils ont un complément en *par* c'est un adjectif et non un argument. Les noms RN ne sont pas assignateurs de rôles thématiques contrairement aux noms ASN. Un des critères les plus importants est sûrement le fait que les RN n'ont pas de lecture événementielle.

Par exemple, en (8), nous avons un nom résultatif.

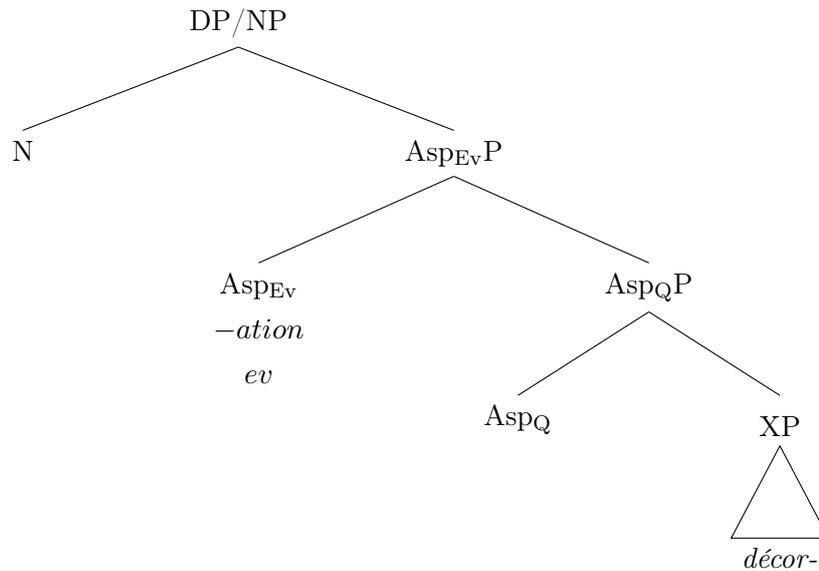


FIGURE 2.3 – Structure du nom d’événement complexe DÉCORATION

- (8) « Quand on corrige les *dissertations* du concours, on essaye de voir si le candidat structure bien son écrit [...]. »⁸

Un des tests pour vérifier qu’il ne s’agit pas d’un ASN est l’incompatibilité au singulier et uniquement au singulier avec un modifieur comme *fréquent* ou *constant*. Pour l’exemple en (8), nous obtenons donc les résultats que nous présentons en (9).

- (9) On corrige les *dissertations* du concours. :
- a. On corrige les dissertations (du concours).
 - b. On corrige les fréquentes dissertations du concours.
 - c. * On corrige la constante dissertation du concours.

A la différence des ASN, les RN ne disposent pas d’une interprétation événementielle héritée du verbe dont ils dérivent. Ainsi leur structure fonctionnelle est plate et n’implique pas autant de têtes fonctionnelles que les ASN comme représenté dans la FIGURE 2.4 pour *décoration*.

8. Exemple tiré de FRCOW.

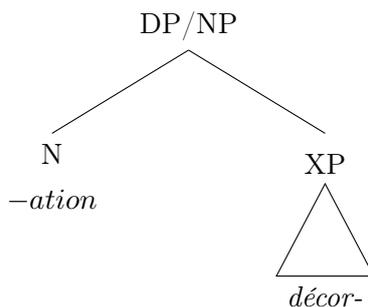


FIGURE 2.4 – Structure du nom résultatif DÉCORATION

L’interprétation que nous pouvons avoir de DÉCORATION est ambiguë. En effet, le nom déverbal est compatible avec la structure d’un événement complexe mais aussi avec celle d’une structure de nom résultatif. Certaines ambiguïtés résultent de la structure syntaxique et non du lexique. Des tests existent donc pour contrer cette ambiguïté. Pour autant, ces deux classes de noms ne sont pas uniques et il en existe une troisième : les noms d’événements simples. DÉCORATION peut aussi être un nom d’événement simple. Alors comment faire la distinction entre ces trois catégories ?

2.3.1.3 Les noms d’événements simples (SEN)

Les noms d’événements simples ont la particularité d’avoir une interprétation événementielle comme les ASN, mais de se comporter structurellement comme des RN (FIGURE 2.4). Ainsi s’il fallait seulement désambiguïser entre les ASN et les RN, la tâche serait simple, mais ajouter les SEN demande de repenser différemment les frontières de catégories. Roy & Soare [2013] rappellent que les propriétés des SEN oscillent entre celles des ASN comme en (10a) et en (10b) et des RN en (10c), raison pour laquelle ils ont souvent été laissés de côté par les théories s’attaquant aux noms événementiels. Ces dernières argumentent aussi pour considérer l’interprétation événementielle comme une propriété lexicale du SEN et non comme étant dérivée dans sa structure syntaxique. La contrepartie de « l’événementialité grammaticale » des ASN est vue ici comme une événementialité lexicale.

- (10) a. La manifestation a duré longtemps.

- b. Durant la manifestation...
- c. * La manifestation en deux heures

2.3.2 Comportement des troncats déverbaux

Kerleroux [1999] dit que les noms d'événements complexes ne sont pas troncables et donne le contraste suivant :

- (11) a. La manifestation de la vérité aura pris cinquante ans
 b. * La manif de la vérité aura pris cinquante ans

Kerleroux introduit donc deux caractéristiques [\pm ev.] et [\pm s.a.] représentant respectivement, d'un côté, le caractère événementiel ou non du nom, et de l'autre, la présence obligatoire d'une structure argumentale attachée à ce nom. Les noms ayant un + spécifié pour leur structure argumentale [+ s.a.] ne peuvent pas être tronqués. Les événements complexes sont spécifiés [+ s.a.] en plus de [+ ev.]. Ainsi, cela permet de distinguer les SEN, les ASN et les RN. Les noms dénotant un procès vu dans son déroulement ne sont pas troncables contrairement aux noms dénotant un événement vu dans sa globalité.

2.3.2.1 Ce que nous observons

Pour comprendre le comportement des troncats déverbaux et à quelle(s) classe(s) ils appartiennent, nous avons décidé de partir de l'exemple du comportement de la paire de lexèmes DÉCORATION/DÉCO dans le but de voir s'il était possible d'en dégager une tendance.

Nous avons ainsi, à partir de corpus FRCOW⁹ (Schäfer [2015], Schäfer & Bildhauer [2012]), cherché des occurrences où *décoration* ou *déco* étaient dans des environnements satisfaisants des contraintes des événements complexes dans le but de vérifier leur nature nominale. Pour cela nous avons fait une requête CQL (Corpus Query Language) qui nous permet de chercher des patrons sur le corpus. Nous avons spécifié que nous souhaitions un complément en *par*, car nous avons vu dans la manière de caractériser les ASN que bien que la présence du complément en *par* n'était pas obligatoire,

9. Nous décrivons le corpus en détail dans la section 3.1.2.1.

lorsque ce dernier était présent, il fallait avoir une interprétation événementielle du nom correspondant à celle d'un nom d'événement complexe. En (12) nous spécifions donc ce complément mais aussi un complément prépositionnel en *de* qui doit être le complément à minima obligatoire marquant le thème de l'action dénoté par le nom là où avec *par* l'agent est introduit.

(12) [*word* = "*déco*"] [*word* = "*de*"] []{0,4} [*word* = "*par*"]

Nous avons aussi pris un exemple construit d'événement complexe pour nous servir de référence de ce à quoi doit ressembler DÉCORATION et DÉCO lorsqu'ils valident les propriétés des ASN :

— La décoration|(?déco) du sapin par les enfants

Savoir si DÉCO peut être un ASN, autrement dit savoir si DÉCO permet de voir l'événement de décorer dans son déroulement, est toute la question ici. Cependant, selon Kerleroux, les troncats ne peuvent être des ASN. Ainsi, avant de procéder à une recherche sur corpus, nous ne nous attendions pas à trouver DÉCO dénotant un événement complexe.

Nous avons repris quelques concordances entre notre patron et les résultats, relevés manuellement car bruités, de notre corpus. Nous nous sommes seulement cantonnés à certains tests donnés par la littérature pour vérifier si nous faisons face à des ASN autant pour *déco* que pour *décoration*. Le premier test que nous avons cherché à vérifier était celui de l'obligation d'avoir au minimum un argument.

(13) Possession d'argument(s) non facultatifs

- a. La décoration *(de la chapelle de la Vierge)
- b. La décoration *(d'un mur de la cour de récréation) à l'aide d'une mosaïque
- c. La décoration *(du bureau de l'Élysée) *(par P. Stark)
- d. La déco *(de la salle) *(par Cocteau)
- e. La déco *(de chez Mustapha) *(par Mustapha)
- f. La décoration *(du sapin) *(par les enfants)
- g. La déco *(du sapin) *(par les enfants)

L'interprétation de ces sept exemples (13), dont deux construits, comme des noms d'événements complexes, nous permet de mettre en évidence que

les arguments sont effectivement obligatoires. Il est tout de même important de souligner que la lecture d'événement complexe doit être forcée, car il est peu aisé d'interpréter certains de ces noms aussi comme des résultatifs. Grimshaw parlait d'ambiguïté à cet égard.

Contrairement à nos prédications, le troncet DÉCO peut également être un nom d'événement complexe. De plus, l'expression en (13g) nous montre qu'il est totalement possible d'interpréter comme grammaticale une phrase où DÉCO serait un événement complexe pouvant être paraphrasé par *l'action de décorer le sapin*.

Pour autant, nous ne nous sommes pas limités à un seul test, nous en avons utilisé d'autres pour consolider l'argument soutenant qu'il peut aussi arriver aux troncets d'être des ASN. Un deuxième test que nous avons appliqué est l'ajout d'un modifieur aspectuel comme *en une heure* pour observer s'il était toujours possible de considérer nos troncets comme des noms d'événements complexes. Là, aussi il était pertinent de supposer que les troncets ne devraient pas accepter de modifieurs.

(14) Possession de modifieur aspectuel

- a. La décoration de la chapelle de la Vierge en une heure
- b. La décoration d'un mur de la cour de récréation en une heure à l'aide d'une mosaïque
- c. La décoration du bureau de l'Elysée en une heure par P. Stark
- d. La déco de la salle en une heure par Cocteau
- e. La déco de chez Mustapha en une heure par Mustapha
- f. La décoration du sapin en heure par les enfants
- g. La déco du sapin en une heure par les enfants

Pourtant ce que nous observons en (14g) met en évidence, une fois de plus, que le troncet donne un résultat positif à un test permettant d'identifier un nom d'événement complexe. En (14g), il est possible de paraphraser sous la forme *La déco du sapin aura pris une heure*. En supprimant le complément en *par*, la phrase semble naturelle. Ici donc, il n'est pas possible de contester que nous sommes face à un événement complexe et non à un nom d'événement simple ou à un nom résultatif, là où parfois le doute reste.

Nous devons aussi relever que ce test est positif pour les autres exemples où ces derniers sont interprétés comme des nominalisations d'actions avec une éventualité persistante.

Nous proposons ensuite un autre test qui est celui de la possession d'un modifieur orienté-agent comme *intentionnel* ou *délibéré*. Le modifieur orienté-agent permet de mettre en évidence implicitement la structure argumentale du nom d'événement complexe. En effet, par ce modifieur l'argument agent permet de montrer que certains noms ont besoin de référer à un agent. Les noms qui ne sont pas des événements complexes doivent bannir ce genre de modifieurs car l'agent n'est pas inclus dans leur structure.

(15) Possession d'un modifieur orienté-agent

- a. La décoration délibérée de la chapelle de la Vierge
- b. La décoration intentionnelle d'un mur de la cour de récréation à l'aide d'une mosaïque
- c. La décoration intentionnelle du bureau de l'Elysée en une heure par P. Stark
- d. La déco intentionnelle de la salle par Cocteau
- e. La déco délibérée de chez Mustapha par Mustapha
- f. La décoration intentionnelle du sapin par les enfants
- g. La déco intentionnelle du sapin par les enfants

ori

Dans ce nouveau test (15), nous obtenons des résultats similaires aux deux tests précédents. En (15g), il est toujours possible d'interpréter le troncat comme un événement complexe. Le modifieur orienté-agent *intentionnelle* sous-entend que c'est la *déco*, *l'action de décorer* qui a été intentionnelle. Les enfants ont donc décoré intentionnellement le sapin.

Finalement, nous proposons un dernier test en (16), celui de l'ajout de modifieurs adjectivaux au pluriel. Cela devrait rendre les phrases agrammaticales pour des événements complexes uniquement.

(16) Possession d'un modifieur adjectival au pluriel

- a. * Les décorations constantes de la chapelle de la Vierge

- b. * Les décorations constantes d'un mur de la cour de récréation à l'aide d'une mosaïque
- c. * Les décorations constantes du bureau de l'Elysée par P. Stark
- d. * Les déco constantes de la salle par Cocteau
- e. * Les déco constantes de chez Mustapha par Mustapha
- f. * Les décorations constantes du sapin par les enfants
- g. * Les décors constants du sapin par les enfants

Les modifieurs adjectivaux sont ici post-nominaux. Toutes les expressions sont jugées agrammaticales avec une mauvaise interprétation incluant des expressions avec troncat comme en (16g). Le processus cognitif de penser à des actions constantes de décorer le sapin par les enfants n'est pas possible. Cependant, le modifieur adjectival aurait fonctionné au singulier donnant « *la déco constante du sapin par les enfants* ».

Le problème des jugements d'acceptabilité, lorsqu'ils n'ont pas été obtenus pendant une expérience, est qu'ils sont principalement binaires, grammaticaux ou agrammaticaux. Or certaines phrases semblent plus grammaticales que d'autres. Il serait intéressant de mener une expérience de jugement d'acceptabilité pour voir si les locuteurs acceptent comme grammatical ce qui ne doit pas être grammatical, autrement dit, si les locuteurs considèrent que les troncats dont la structure fonctionnelle est celle d'un ASN, constituent des énoncés grammaticaux.

Pour mettre en contraste ces noms d'événements complexes, nous avons aussi relevé dans notre corpus FRCOW des noms non-ASN troncats avec une structure similaire dont nous présentons les expressions en (17).

- (17) a. « La nouvelle déco du Chat Noir réalisée par Stéphane Lauper et Jerrycan [...] »
- b. « La déco du salon a été pensé par Yan [...] »
- c. « [...] ma déco de salle de bain par Gérard [...] »
- d. « La magnifique déco de cette chambre est signée par la marque italienne Alta Moda [...] »

- e. « La déco de cet endroit, réalisée par Philippe le tenancier de la péniche est remarquable [...] »

De même que pour les noms d'événements complexes, nous avons fait passer nos quatre tests pour vérifier nos dires en (18).

(18) Noms non-ASN

- a. La nouvelle déco (du Chat Noir) (réalisée par Stéphane Lauper et Jerrycan)
- b. La déco du salon (*en une heure) a été pensée par Yan
- c. Ma déco (*intentionnelle) de salle de bain par Gérard
- d. Les magnifiques décors constants de cette chambre sont signées par la marque italienne Alta Moda
- e. La déco (de cet endroit), (réalisée par Philippe le tenancier de la péniche) est remarquable

Dans les expressions (18a) et (18e), nous pouvons observer que les arguments sont facultatifs, dans l'expression (18b), l'ajout d'un modifieur aspectuel a pour conséquence de rendre l'expression agrammaticale. Dans l'expression (18c), l'ajout d'un modifieur orienté-agent rend l'expression agrammaticale mettant en évidence qu'il ne s'agit pas d'un nom d'événement complexe. Finalement, en (18d) le fait de pouvoir d'accepter le modifieur adjectival au pluriel montre bien qu'il ne s'agit pas d'un nom d'événement complexe.

Nous venons de montrer que les tests d'identification des noms d'événements complexes fonctionnent. Nous avons ainsi pu certifier que DÉCORATION peut être un nom d'événement complexe, mais aussi plus surprenant que DÉCO pouvait être interprété de la sorte.

Bien que nous n'ayons analysé que le lexème tronqué DÉCO, nous devons nous demander à quel point nos remarques peuvent s'étendre à d'autres noms déverbaux tronqués. Si ces remarques s'appliquent à d'autres noms alors nous devons nous interroger sur les mécanismes sous-jacents rendant cela possible. Nous tenterons d'apporter des éléments de réponse dans ce mémoire.

Comme nous l'avons postulé, il est raisonnable de penser que nous avons deux structures associées aux noms. D'un côté la structure « complexe » des noms d'événements complexes, et de l'autre côté la structure « simple » des noms d'événements simples et des noms résultatifs. Nous supposons que la structure simple est celle qui est favorisée par la troncation.

2.3.2.2 Discussion

Nous avons vu d'un côté les motivations de la troncation, et d'un autre côté les contraintes associées à ce phénomène. La motivation principale de la troncation est celle associée au principe du moindre effort. La contrainte principale de la troncation est celle associée au caractère non-apocopable des noms d'événements complexes. Nous avons pourtant réussi à interpréter le troncat DÉCO comme un nom d'événement complexe, c'est-à-dire qu'il existe une interprétation où l'action de décorer est vue dans son déroulement et non comme ayant déjà eu lieu ce qui contredit les propos de Kerleroux. Pour autant, est-il possible de dire qu'elle a tort ? C'est ce que nous allons nous demander dans le troisième chapitre de ce mémoire où nous allons essayer de quantifier les conséquences de l'impossibilité pour les noms d'événements complexes d'être des troncats comme supposée par Kerleroux.

CHAPITRE 3

ÉTUDE QUANTITATIVE DES CONTRAINTES SUR LA TRONCATION

3.1 Les contraintes sur la troncation

TOUTS LES MOTS ne semblent pas être égaux vis-à-vis de la troncation. Après avoir vu pourquoi nous tronquons, nous avons observé que les verbes étaient défavorisés par la troncation au profit des adjectifs et des noms qui étaient majoritairement représentés. De même, nous avons vu qu'à l'intérieur de la catégorie des noms déverbaux, tous les noms n'étaient pas égaux dans leur comportement vis-à-vis de la troncation. Par la suite, nous nous intéresserons uniquement aux noms déverbaux en *-tion*. Kerleroux [1999] réceptive à ce manque de sensibilité pour la troncation de certains noms postule que celle-ci ne s'applique pas à tous les noms, et plus particulièrement aux noms d'événements complexes qui ne peuvent pas donner de forme tronquée.

3.1.1 Ce qui est tronnable et ce qui ne l'est pas

Kerleroux explique que la troncation s'applique aux noms qui correspondent à des dénominations d'entités, c'est-à-dire à des objets concrets ou abstraits. Ces noms là peuvent subir l'apocope. Pour illustrer, en (19) nous avons un exemple de troncation issu de Yahoo! Answers. La personne s'exprimant utilise la forme *doc* pour *documentaire* qui est un nom substantif. Ce nom est une dénomination d'entité et peut donc être tronqué. Cet exemple est aussi intéressant dans la mesure où il reflète le comportement des locuteurs qui, généralement, lorsqu'ils se permettent d'utiliser un troncat en utilisent plusieurs. Ainsi nous retrouvons l'adjectif *antisarko* pour *antisarkoziste* et le nom propre *Sarko* pour *Sarkozy*.

- (19) « Vous ne pensez pas que le **doc antisarko** de Arte n'était pas très convainquant ?

La mauvaise foi des journaliste anglais (qui pratique souvent du french bashing était flagrante), ils puait l'antisarkozysme et les critiques était tellement stupide que ca peut se retourner contre eux et jouer en faveur de **Sarko** »

Ces noms non-déverbaux de dénominations d'entités sont troncables. Ils s'opposent aux noms déverbaux qui comprennent les noms référentiels, la classe des noms d'événements complexes et la classe des noms d'événements simples. Nous avons vu que la littérature est catégorique sur le fait que les ASN ne sont pas troncables, et donc que seulement les SEN et les RN sont troncables.

Exemples :

- (20) a. *L'intro du lynx dans le Vercors provoque des tollés¹
b. L'intro de ta dissertation nous a bien fait rire.¹
- (21) a. *L'huma va à sa perte.¹
b. Acheter l'Huma le dimanche matin.

1. Exemples empruntés à Kerleroux [1999]

Observons les contrastes en (20) et en (21). En (20a), nous remarquons que INTRODUCTION ne peut pas être tronqué alors que en 20b INTRO est tronqué. Ce qui a été proposé est que les deux noms ne sont pas du même type. En effet, en (20a) il s’agit d’un nom d’événement complexe, alors que en (20b), il s’agit d’un nom d’événement simple. Nous avons argumenté dans le deuxième chapitre de ce mémoire que l’interprétation entre un nom d’événement complexe et un nom d’événement simple pouvait être ambiguë pour un nom déverbal.

De même, en prenant le contraste existant en (21) pour *humanité*, nous avons en (21a) une phrase jugée agrammaticale par la présence de la forme tronquée *huma* alors que la même forme en (21b) ne rend pas la phrase grammaticale. *humanité* lorsqu’il correspond à un nom désadjectival, représentant une « sorte de Nom collectif » (Kerleroux [1999]), ne peut pas être tronqué et il est possible d’en faire le parallèle avec les noms déverbaux non-troncables. Alors que lorsqu’il correspond à un nom propre celui du journal *Humanité*, il est tronnable, représentant donc un nom propre. Les noms propres ne sont pas contraints par la troncation. En suivant cette logique pour *humanité*, nous avons d’un côté un type de nom tronnable (22b) et d’un autre un type non tronnable (22a).

(22) HUMANITÉ → ?HUMA

- a. « La Fête de l’Huma est une fête organisée tous les ans par le journal L’Humanité. ² »
- b. *Les crimes contre l’huma commis contre la population civile durant la guerre.

Kerleroux argumente donc contre le troncat *huma* s’il ne réfère pas au nom propre du journal. Kerleroux voit à travers les noms désadjectivaux un moyen de confirmer les observations et les contrastes qui pouvaient être faits avec les noms déverbaux.

3.1.1.1 Ce qui n’est pas tronnable mais l’est quand même

Si Kerleroux disait vrai, alors nous n’aurions pas dû nous retrouver avec des occurrences de *déco* dont l’interprétation est celle d’un nom d’événement

² Google Search : « Fête de l’Huma » : [consulté le 24 mars 2019]. <http://paris13.pcf.fr/5547>

complexe. Il est donc là une force qui doit être mise en avant et permettre la troncation. Cette force est sûrement à mettre en relation avec le principe du moindre effort. Pour vérifier ce qu'il était possible de tronquer, nous avons recherché sur trois moteurs de recherche, Google Search, Yahoo! Answers et Twitter, l'expression exacte *crime contre l'huma*. Nous avons trouvé des résultats mettant en évidence que le mot *huma* qui ne réfère pas au journal existe.

(23) « La france est elle fiere de ces 200.000 ivg par an, est-ce cela la modernité, ou est ce un crime contre l'huma ? »³

(24) « @afpfr @Guilhem_Klein fallait y penser avan d'y semer le chaos non ? Faut y envoyer sarko pour qu'il y soit jugé pour crime contre l'huma. »⁴

Il est possible de protester qu'il s'agit d'exemples qui ne sont pas représentatifs de par leur provenance. En effet, Twitter est connu pour proposer d'écrire dans un format limité en caractères. Ainsi jusqu'à novembre 2017, seulement 140 caractères pouvaient être inclus dans une publication, contre 280 après novembre 2017. La firme de Twitter affirme cependant que 7% des tweets écrits à l'époque atteignaient 140 caractères⁵. Cela peut jouer en la faveur de l'idée que *crime contre l'huma* est possible dans la grammaire de ceux qui l'utilisent et qu'il ne s'agit pas de la conséquence d'une simple limite de caractères. Dans cette continuité, nous pouvons expliquer l'expression *crime contre l'huma* issue de Yahoo! Answers en (23) comme une manifestation de cette possibilité incluse dans la grammaire des locuteurs.

La littérature est donc catégorique entre ce qui peut être tronqué et ce qui ne le peut pas mais ce n'est pas la bonne vision à adopter. Il devient

3. Yahoo! Answers : « Crime contre l'huma » : [consulté le 29 octobre 2018]. https://fr.answers.yahoo.com/question/index;_ylt=AwrC1C20JMtbiMAIxfP2IIQ;_ylu=X3oDMTBydDI5cXVubGNvbG8DYmYxBHBvcwM2BHZ0aWQDBHNIYwNzcg--?qid=20100204115901AABcFG9

4. Twitter : « Crime contre l'huma » : [consulté le 29 octobre 2018]. <https://twitter.com/afpfr/status/685518849467219968>

5. Le Monde : [consulté le 10 mai 2019]. https://www.lemonde.fr/entreprises/article/2017/11/07/twitter-generalise-les-messages-en-280-caracteres_5211616_1656994.html

ainsi légitime de se demander ce qui est réellement produit et ce qui peut être réellement produit.

Dans le but de répondre à cette question, nous avons fait de nouvelles recherches sur internet pour nous faire un avis. Nous avons trouvé des attestations pour tous les exemples qui selon la littérature ne devraient pas exister pour différentes raisons, comme pour leur structure argumentale, ou pour la « manière lacunaire » dont s'exerce la troncation sur les différents sens d'un mot. Certains de ces exemples sont donc dérivés de bases qui ne sont pas forcément des noms déverbaux en *-tion*. Ils sont présentés comme ayant les mêmes caractéristiques. Les lexèmes doivent par conséquent présenter les mêmes contrastes dans l'utilisation de leurs formes. Il est donc possible de retrouver certains emplois restreints par la troncation.

(25) MANIFESTATION → MANIF

- a. « En d'autres termes, cette **manif d'humeur** interpelle les autorités à démettre "Cissé Technologies" du marché [...] »⁶

(26) PUBLICITÉ → PUB

- a. « Rappelle qu'un arrêt d'assemblée plénière du 24/11/1989 dit que les plaidoiries participent des droits de la défense et de la **pub des débats** »⁷

(27) EXERCICE → EXO

- a. « @eridann "86, Mitterrand savait qu'il perdrait, ms il était possible que le 1er Ministre issu de l'opposition soit usé par l'**exo du pouvoir**. »⁸
- b. « Good news pour #Leo il est libre de s'engager auprès d'un nouveau club sans craindre d'être limité dans l'**exo de ses fonctions**. #TeamPSG »⁹

6. « Manif d'humeur » : [consulté le 24 mai 2019]. <https://maliactu.net/mali-reparation-du-scanner-de-gabriel-toure-la-societe-cisse-technologies-a-du-mal-a-voiler-ses-carences/>

7. « Pub des débats » : [consulté le 24 mai 2019]. <https://twitter.com/fatadecaro/status/840135341902258177>

8. « Exo du pouvoir » : [consulté le 24 mai 2019]. <https://twitter.com/MatthieuGB/status/73487370619596800>

9. « Exo de ses fonctions » : [consulté le 24 mai 2019]. <https://twitter.com/MatthieuGB/status/73487370619596800>

MANIF ne peut être interprété que comme référant à un cortège de personnes dont le but est de communiquer une opinion. Son occurrence en (25a) ne devrait pas être possible mais pourtant existe. PUB est possible que s’il ne comporte pas d’arguments, donc il n’est pas possible de tronquer PUBLICITÉ dans l’expression *publicité des débats*. Son occurrence en (26a) ne devrait pas être possible mais pourtant existe. Kerleroux argumente que EXERCICE peut disposer de vrais arguments bien qu’il s’agisse d’un mot lexical simple. S’il en possède alors il n’est pas tronnable. Pour autant en (27a) ou en (27b), il est possible de dire que EXO dispose d’une structure argumentale alors qu’il n’aurait pas dû être possible d’obtenir ces occurrences.

Par ces contre-exemples nous ne souhaitons pas avancer des résultats absolus mais simplement mettre en évidence qu’une étude qualitative permet de prendre du recul par rapport aux propos de Kerleroux. Il ne s’agit que d’une étude préliminaire pour tester la solidité des hypothèses mises en avant par la littérature puisque les contrastes qui doivent s’observer selon Kerleroux ont des répercussions sur la syntaxe des mots et leur sémantique. Il y aurait donc une différence de distribution entre les mots non-tronqués et les mots tronqués au sein d’un groupe nominal. C’est sur cette première question de recherche que nous développerons notre étude visant à vérifier ou infirmer cette différence de distribution qui se veut absolue comme exprimé par Kerleroux.

3.1.2 Étude des contraintes sur la troncation

Nous sommes partis d’une simple observation de Kerleroux sur les noms de procès, autrement dit les noms qui dénotent un événement complexe et ne sont pas tronnable, et sont caractérisés dans leur syntaxe par la réalisation d’un argument sous la forme d’un complément. De plus, ces noms doivent être au singulier avec un déterminant défini, et non au pluriel ou avec un déterminant indéfini. Ainsi à cause de l’argument obligatoire sous forme de complément, ces noms apparaissent avec un groupe prépositionnel en *de*.

Pour vérifier cette affirmation, nous avons mené une étude quantitative sur corpus. En effet, il est possible de quantifier à partir d’un corpus du

français des proportions d'utilisation du singulier, des déterminants définis ou des groupes prépositionnels en *de*. Pour cela, si Kerleroux avait vu juste nous devrions nous attendre à plusieurs proportions dans l'analyse du corpus pour les différents paramètres analysés. Ces proportions sont liées au fait que les noms sont ambigus. Premièrement, nous nous attendons à une proportion plus importante de formes au singulier pour les non-troncats que pour les troncats. Deuxièmement, nous nous attendons à une proportion plus importante de formes en cooccurrence avec un déterminant défini et finalement à une proportion plus importante de formes en cooccurrence avec groupe prépositionnel en *de*. Cela met directement en opposition les noms d'événements complexes non-troncables aux autres classes de noms troncables.

3.1.2.1 Notre corpus

Nous avons travaillé avec le corpus FRCOW, il s'agit d'un corpus web du français de 2016 (Schäfer [2015], Schäfer & Bildhauer [2012]). Le corpus comportait 8 789 873 699 mots lors de sa dernière compilation en décembre 2017 (10/12/2017) pour 10 790 863 069 occurrences au total. 403 981 lexèmes sont représentés¹⁰, le corpus, de plus, comporte 27 642 351 mots-formes/formes fléchies¹¹. Il est aussi possible d'avoir une information sur les parties du discours. L'annotation des parties du discours présente dans le corpus a été faite de manière automatique. Les nombres fournis ci-dessous sont aussi ceux donnés par la documentation du corpus. N'étant pas un corpus validé manuellement, il faut garder en tête que la lemmatisation et l'annotation des parties du discours ne sont pas parfaites.

Le corpus FRCOW présente aussi l'avantage lors de la recherche d'avoir accès aux métadonnées sur le pays d'où le site web est indexé. Plus de 90 pays sont répertoriés. Près de 80% du corpus provient de France, du Canada, de Suisse, de Belgique et du Luxembourg. Il est aussi possible de catégoriser le corpus en deux sous-parties, en fonction de si le site-web est ou non un site de forum. Il s'agit là aussi d'une catégorisation qui a été faite

10. Sketch Engine : Lemma is the basic form of a word, typically the form found in dictionaries.

11. Sketch Engine : Word form - Refers to one of the word forms that a lemma can take.

automatiquement.

3.1.2.2 Élaboration d'une liste de troncat

Ne pouvant travailler de manière exhaustive sur tous les troncats, nous avons dû nous restreindre à l'étude d'un sous-ensemble de troncats pour faire nos recherches. Nous avons fait le choix de ne travailler que sur des noms déverbaux en *-tion* car il s'agissait des noms les plus étudiés dans la littérature et dont nous avons vu l'ambiguïté dans l'interprétation. J. K. Kanwal [2018] avait travaillé sur une liste de 30 troncats de ces noms déverbaux. Ayant de plus accès aux troncats du dictionnaire de Antoine Antoine [2000] et à la liste de Raccah [2015], nous avons élaboré une liste de 10 paires de bases/troncats présents dans les ouvrages des trois auteurs que nous présentons en TABLE 3.1. Seule la finalité d'avoir des troncats suffisamment utilisés pour être trouvés lors d'une analyse de corpus nous à motiver à procéder à l'intersection des listes pour élaborer la notre. Nous utiliserons cette même liste dans la deuxième partie du présent mémoire car nous avons fait un travail d'annotation manuelle, ce qui justifie la taille de la liste.

Lexème-base	→	Lexème-troncat
COMPILATION	→	COMPIL
DÉCORATION	→	DÉCO
DÉMONSTRATION	→	DÉMO
INFORMATION	→	INFO
INTOXICATION	→	INTOX
MANIFESTATION	→	MANIF
PROMOTION	→	PROMO
RÉCRÉATION	→	RÉCRÉ
RÉDACTION	→	RÉDAC
SONORISATION	→	SONO

TABLE 3.1 – Liste des paires de bases-troncats utilisée

3.1.2.3 Élaboration d'une liste de déterminants

Avant d'étudier le corpus, nous avons aussi dressé une liste de déterminants qui nous permettraient d'observer concrètement les contrastes attendus en fonction de notre question de recherche qui était de caractériser les emplois des bases et des troncats. Cette liste est présentée dans la TABLE 3.2. Nous avons choisi onze déterminants. Ces déterminants sont définis et indéfinis. Ils peuvent être singulier ou pluriel. Ce choix se justifie par le fait que les recherches que nous avons faites ont été des recherches manuelles et donc qu'il était important de se limiter dans la quantité de données à chercher. Nous n'avons, de plus, pas pris en compte les déterminants contractés comme par exemple *aux* pour *à + les*.

	Défini	Indéfini
Singulier	la cette	une chaque aucune toute
Pluriel	les ces	plusieurs quelques toutes

TABLE 3.2 – Liste des déterminants utilisée

3.1.2.4 Recherche sur corpus

Nous avons recherché manuellement le nombre de cooccurrences des bases et des troncats avec les différents déterminants dans un premier temps. Les nombres d'occurrences ont été annotés dans la TABLE 3.3. Certaines cooccurrences ne sont pas attestées dans notre corpus. De plus, nous avons dû supprimer la paire *sonorisation/sono* car le corpus était bruité lors de la recherche de *sono*. En effet, l'indexation des pages web étant faite automatiquement, des sites webs français dont la langue majoritaire est l'italien ont été indexés et a bruité nos résultats. Là où nous nous attendions à peu d'occurrences de *sono*, forme tronquée de *sonorisation*, nous nous sommes

retrouvés avec un nombre d’occurrences trop élevé. La forme *sono* est une forme fléchie de ESSERE en italien justifiant son emploi massif. Les résultats que nous présentons en TABLE 3.3 peuvent aussi être bruités. A cause du nombre important d’occurrences et cooccurrences, nous avons procédé au survol de quelques occurrences pour une vérification manuelle, et cette dernière n’a pas motivé la mise à l’écart d’autres couples.

Mot \ Déterminant	Nombre de cooccurrences avec un déterminant (en %)											Nombre total de cooccurrences avec un déterminants (en %)	Nombre total d’occurrences
	la	une	les	cette	ces	chaque	quelques	aucune	toute	toutes	plusieurs		
compilation	51.4950	35.1520	3.6956	8.1842	0.4947	0.2255	0.2037	0.1128	0.1564	0.0073	0.2728	42.1753	65185
compil	75.1019	15.3312	3.4254	4.9796	0.4980	0.1056	0.2113	0.0151	0.0453	0.0000	0.2867	64.6789	10246
décoration	74.3764	13.5824	7.6176	2.3615	0.7286	0.0871	0.3162	0.3956	0.2979	0.0474	0.1894	24.8839	263090
déco	78.3034	15.9165	3.7446	1.2865	0.2262	0.0667	0.2002	0.1594	0.0185	0.0037	0.0742	14.5236	185711
démonstration	53.0224	32.0857	6.1537	5.4877	1.1243	0.1021	0.4402	0.7216	0.5015	0.0170	0.3438	37.3540	235964
démo	56.0486	27.8293	7.6247	6.3843	0.5339	0.1662	0.8568	0.1726	0.0032	0.0000	0.3804	27.3080	114545
information	49.9224	8.4147	28.2447	3.6476	5.2853	0.0477	0.9875	1.0906	1.7570	0.5114	0.0913	30.5902	5730896
info	46.3116	5.5028	36.8258	3.9796	3.3707	0.0272	2.3913	0.9701	0.3468	0.2197	0.0544	16.2864	2347651
intoxication	44.9685	31.9289	17.2455	3.1322	1.7031	0.0518	0.0740	0.2444	0.4961	0.0444	0.1111	39.9734	33785
intex	84.6474	11.1248	1.7898	1.8516	0.2006	0.0154	0.2469	0.0926	0.0154	0.0154	0.0000	32.4927	19946
manifestation	31.9742	21.9523	21.7488	15.8345	5.2654	0.3474	0.2954	0.4567	1.1076	0.1853	0.8325	42.8191	691825
manif	56.0821	19.2867	18.0858	3.8047	1.5399	0.4219	0.2957	0.2020	0.1443	0.0180	0.1190	51.1898	54169
promotion	76.5809	4.8358	4.5578	13.2334	0.1894	0.2013	0.0363	0.1514	0.1331	0.0299	0.0506	46.6996	630781
promo	59.9025	14.9069	20.5895	2.6640	0.6693	0.3147	0.2305	0.3989	0.1241	0.0887	0.1108	16.4018	137546
récréation	67.6508	11.1746	17.8201	1.1217	0.5291	1.2381	0.2011	0.0635	0.1376	0.0000	0.0635	29.2606	32296
récré	92.7578	1.4585	4.9455	0.1341	0.0168	0.6203	0.0335	0.0168	0.0000	0.0000	0.0168	39.4224	15131
rédaction	96.0335	1.6225	1.6730	0.5073	0.0432	0.0278	0.0109	0.0191	0.0236	0.0014	0.0379	60.3201	590899
rédiac	99.2938	0.3262	0.2725	0.0806	0.0000	0.0154	0.0038	0.0000	0.0000	0.0000	0.0077	80.9860	32171

TABLE 3.3 – Nombre d’occurrences et de cooccurrences des bases et troncats

3.1.2.5 Statistiques

Nous avons cherché à étudier les proportions relatives des différentes cooccurrences en fonction de différents paramètres.

En nous intéressant aux différentes proportions globales d’utilisation des déterminants par les bases et les troncats, il est possible de représenter ces proportions pour mettre en évidence la dépendance de la distribution des déterminants en fonction de la forme du mot, ou de l’absence de dépendance. La FIGURE 3.1 met en évidence ces différentes proportions. Nous

remarquons que sur la totalité des déterminants, *la* représente la majorité de ceux-ci, tant pour les bases que pour les troncats. De plus, il est possible de dire, graphiquement, que cette distribution semble dépendante de la forme du mot. La troncation semble donc influencer la distribution des déterminants.

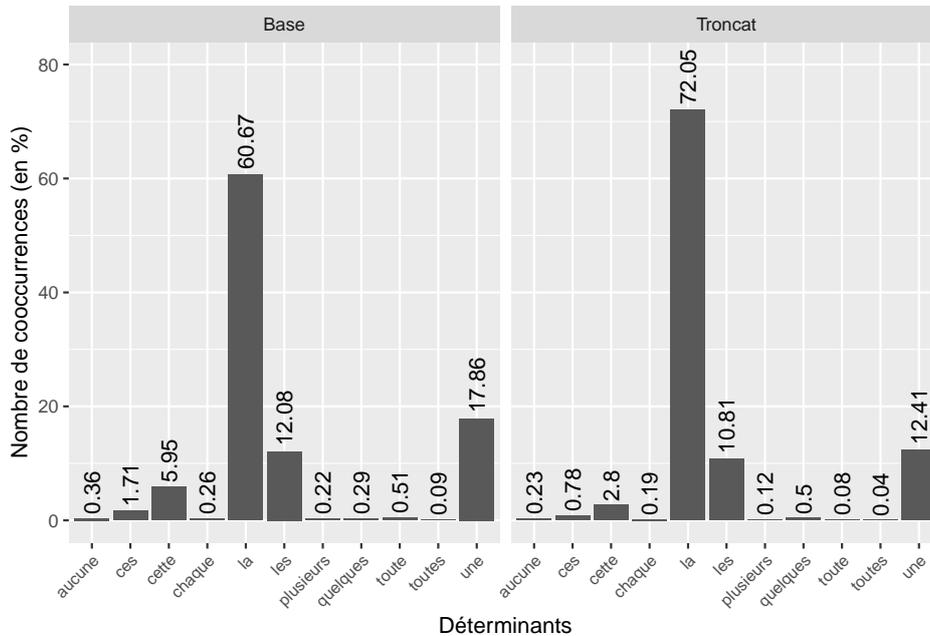


FIGURE 3.1 – Proportions relatives d’utilisation des déterminants par les troncats et les non-troncats

Pour vérifier cette observation, nous avons effectué un test de χ^2 sur nos données ($\chi^2(10) = 33257, p < .001$). Cela nous a permis de considérer qu’il peut exister un lien entre les deux variables troncation et distribution des déterminants. Autrement dit, le fait d’être ou non un mot tronqué a donc une influence sur le nombre de cooccurrence du déterminant avec le nom. De plus, nous soulignons le fait que les résidus standardisés de Pearson ont mis en évidence que seulement *les* et *quelques* sont sur-représentés avec les troncats. Seul le déterminant *aucune* à son nombre d’occurrences qui est indépendant du fait qu’il soit apparaisse en cooccurrence avec un troncat ou une base. Les résultats nous permettent de dire, lorsqu’ils sont informatifs, que les autres déterminants sont sous-représentés avec les troncats.

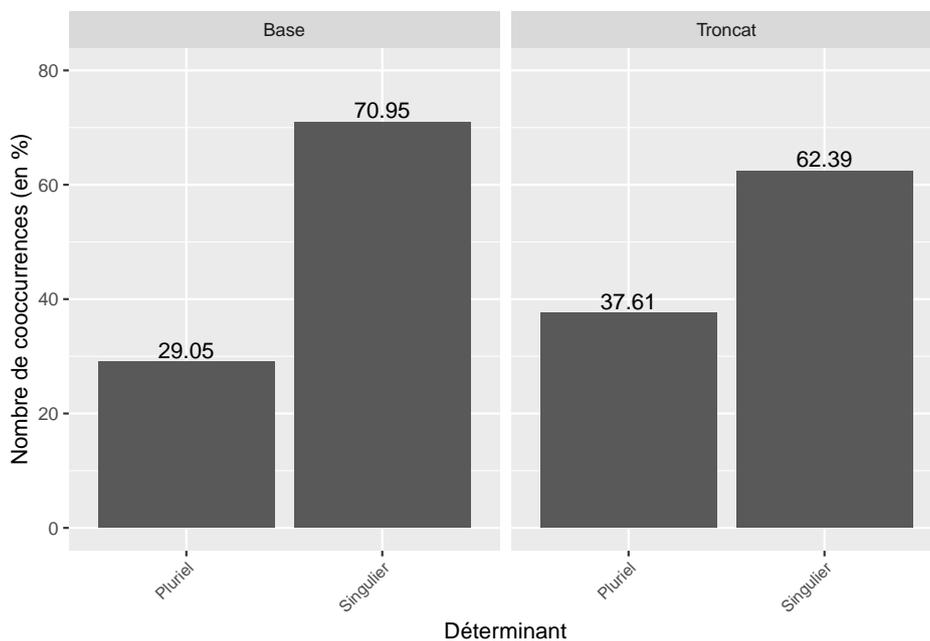


FIGURE 3.2 – Proportions relatives d’utilisation de déterminants au singulier et au pluriel par les bases et les troncats

En regardant cette fois-ci les proportions d’utilisation du pluriel et du singulier au niveau des déterminants (FIGURE 3.2), l’utilisation au singulier est moins prépondérante pour les troncats par rapport aux formes non-tronquées. Le fait pour le mot d’être ou non tronqué a une influence sur le nombre de cooccurrences qu’il aura avec un déterminant singulier ou un déterminant pluriel ($\chi^2(1) = 14805, p < .001$). Le pluriel est sur-représenté avec les troncats tandis que le singulier est sous-représenté. L’asymétrie que nous mettons en évidence va dans le sens d’une interprétation où le troncat au pluriel est possible et que ce qui est défavorisé au moment de la troncation est sûrement un type particulier de troncats correspondant aux noms d’événements complexes qui se doivent d’être au singulier pour être interprétés comme tel.

Les proportions d’utilisation de déterminants définis et indéfinis varient aussi en fonction de la forme du mot (FIGURE 3.3). En gardant à l’esprit

que nous avons trouvé un effet de la forme des mots sur le nombre des déterminants, nous trouvons aussi un effet des mots sur la définitude des déterminants ($\chi^2(1) = 1018.7, p < .001$). La proportion de déterminants définis pour les troncats est plus importante que pour les bases. Ces résultats vont donc dans le sens de nos résultats précédents. Si la troncation contraint, c'est qu'elle défavorise les noms avec un déterminant indéfini et donc favorise ceux définis. Pourtant, une fois de plus, il ne s'agit pas d'une contrainte absolue.

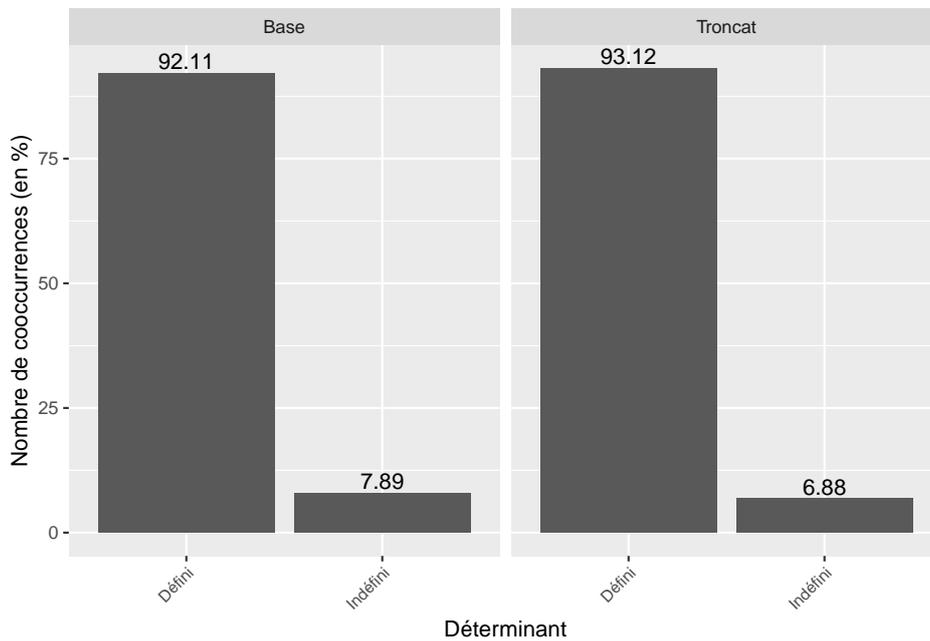


FIGURE 3.3 – Proportions relatives d'utilisation de déterminants défini et indéfini par les bases et les troncats

Finalement, nous avons voulu mettre en évidence qu'il existe une différence dans les proportions d'utilisations de la préposition post-nominal *de* en fonction des noms bases et des noms tronqués. Nos résultats nous indiquent que la proportion de préposition *de* est différente ($\chi^2(1) = 472010, p < .001$) en fonction de la forme du lexème. La proportion de *de* pour les non-troncats est plus importante que pour les troncats ($t(8) = -2.9174, p = 0.0096$). Les troncats, en comparaison avec les bases, ne font appel qu'à très

peu de groupes prépositionnels post-nominaux, montrant bien une fois de plus qu'ils sont contraints.

3.1.3 Conclusion intermédiaire

Kerleroux propose des contraintes catégoriques. Nous avons montré que celles-ci ne tiennent pas en tant que telles, mais que la tendance statistique va dans la direction suggérée par Kerleroux. Tout est tronnable, mais tout n'est pas tronnable dans les mêmes proportions. De ce fait, les distributions des troncats et des non-troncats sont différentes au sein de la phrase comme nous l'avons montré. Nous avons mis en avant que la troncation a une influence sur le nombre du troncat. Cette influence existe aussi sur la définitude du déterminant qui précède le nom tronqué.

Notre travail quantitatif n'a porté que sur certains noms déverbaux, mais il serait intéressant d'observer ce qui se passe avec les noms désadjectivaux qui eux aussi sont contraints par la troncation. Dans une perspective différente, il serait aussi intéressant de se demander quelles sont les contraintes de la troncation sur les verbes et les adjectifs.

Ainsi, les résultats présentés ici devraient aussi être mis en relation avec des résultats qui seraient potentiellement similaires avec une étude des troncats déverbaux en anglais, puisque la littérature sur les différentes classes de noms porte principalement sur la langue anglaise. Il ne serait pas surprenant que les troncats en français et en anglais partagent des propriétés morphologiques communes vis-à-vis de la troncation.

CHAPITRE 4

QUEL EST LE SENS DE LA TRONCATION ?

DANS CETTE NOUVELLE PARTIE, nous nous sommes intéressés aux sens que peuvent revêtir les mots apocopables. Cet intérêt provient du décalage que nous avons pu avoir entre le sens donné par la littérature (Racah [2015], Kerleroux [1999]) et nos propres intuitions de locuteurs natifs.

Si la troncation existe c'est qu'elle doit apporter quelque chose. Cet apport est caractérisé d'un côté par une réduction de forme, des mots plus courts sont obtenus par troncation. D'un autre côté, nous verrons que la troncation se caractérise aussi par les différentes acceptions, les différents sens, des mots qu'elle choisit. Kerleroux parle d'une « troncation lacunaire ».

Comprendre l'importance de l'apport de la troncation, c'est comprendre que la troncation n'agit pas comme une fonction identité au niveau sémantique. Il y a donc une information morphologique et sémantique apportée par le phénomène. Nous allons explorer l'information sémantique dans ce chapitre en nous plaçant du côté de la sémantique lexicale.

4.1 État de l'art

La troncation modifie la signification d'un mot sans pour autant en modifier la référence. C'est derrière cette remarque reprise à Mahfoudh-Hubert [2009] que Raccah [2015] construit son argumentation visant à proposer une description sémantique de la troncation, et plus précisément de l'apocope. Son argumentation pourtant ne se base ni sur une approche morphologique, ni phonologique, bien qu'il puisse y avoir des liens entre ces différentes disciplines linguistiques et dont l'auteur est d'ailleurs conscient. Raccah aborde la sémantique de la troncation avec les substantifs laissant de côté les adjectifs et les verbes, comme d'autres auteurs l'ont déjà fait. En effet, il a besoin dans son argumentaire de mettre en évidence ce qui diffère entre l'utilisation d'un troncat et l'utilisation du mot-source dont est issu le troncat, et, de plus, des effets que ces utilisations peuvent produire dans le discours. Cela permet, selon lui, de construire l'orientation générale du discours souhaitée par les locuteurs. Pour ce faire, il est important pour lui de trouver des « instructions ou des contraintes sémantiques stables », autrement dit ce qui va permettre le changement entre le mot-source et le troncat. Kerleroux [1999] perçoit ce changement comme une « opération de singularisation des dénominations ».

Premièrement, lorsque nous parlons de troncation, nous pouvons voir le phénomène avec une approche pragmatique consistant à dire qu'un énoncé avec des troncats n'a pas les mêmes conditions d'emploi qu'un énoncé avec les mots-sources correspondant aux troncats. Les conditions d'utilisation du troncat sont plus restreintes lorsque les mots tronqués sont présents, autrement dit, comme nous l'avons vu, tout n'est pas tronnable, et cela se manifeste à travers les contraintes syntaxiques que nous avons mises en avant. Mais ici, le fait de dire que tous les contextes ne permettent pas l'utilisation de troncats exprime le besoin d'avoir des contraintes sémantiques pour la production de troncats puisqu'il faut une certaine cohérence par rapport au discours alors qu'avec les mot-sources de telles contraintes ne semblent pas présentes.

Mais pourtant la troncation n'affecte pas que les conditions d'emplois

et donc les effets produits ne sont pas que pragmatiques puisque la situation d'énonciation n'est pas la seule à rendre irrecevable un énoncé. Des « instructions sémantiques » sont utiles pour mettre en évidence que ce n'est pas seulement l'énonciation qui pose problème. Raccach définit les conditions d'emploi comme « des contraintes qui mettent en relation des segments linguistiques avec des situations d'énonciation », autrement dit les conditions agissent comme une interface entre la morphologie et les formes utilisables, et la pragmatique avec le contexte d'élocution. Les instructions sémantiques, quant à elles, permettent d'entrevoir les intentions du locuteur sur son discours. Il s'agit donc d'une restriction ou d'une modification du domaine de référence par rapport à celui du mot-source.

Pour mettre en avant cette restriction, Raccach analyse des occurrences de formes tronquées en les comparant avec les formes non-tronquées. En guise d'exemple pour son argumentaire, il utilise le contraste qui existe entre *examen* et *exam* lorsque la forme est précédée du déterminant indéfini *un* et suivie de l'adjectif *médical*. Le contraste d'interprétation en (28) est justifié par le fait que le domaine de référence de *exam* est limité au domaine scolaire.

- (28) a. un **examen** médical
 b. ? un **exam** médical

Ainsi, il n'est pas possible de parler d'un examen médical avec sa forme tronquée. De même, pour la paire *intoxication/intox*, le domaine de restriction du domaine n'inclut là non plus pas le domaine médical, l'auteur argumente pour ce qu'il appelle une « translation métaphorique du domaine ». Il n'est donc pas possible de parler d'*intox alimentaire* pour parler d'une *intoxication alimentaire*.

Raccach liste de cette façon des restrictions de domaines pour 142 apocopes substantivales. Il ne peut donner des indications de restriction pour une trentaine de troncats car certaines formes n'imposent pas de restrictions, ou tout simplement car il n'est pas évident de caractériser la restriction opérée par certaines formes. Bien que ce ne soit pas aisé, il reste que les conditions d'emplois, pouvant être en relation avec la familiarité du discours, jouent un rôle important. La subjectivité du lecteur entre donc en

compte lors de la modification du domaine de référence, c'est sa manière de voir le monde qui façonne sa manière de parler et donc de tronquer ou non. Il s'agit donc d'un choix de la part du locuteur qui n'est pas basé sur des critères objectifs et qui ne seront donc pas commun à l'ensemble des locuteurs d'une langue, mais bien personnels, et peuvent néanmoins être partagés par un groupe d'individus. Il s'agit d'un point de vue sur une situation. Raccach reconnaît l'utilité d'une étude de corpus pour vérifier ses hypothèses.

4.2 Étude sur les acceptions des troncats

En tenant compte de ce qui avait été dit dans la littérature (Raccach [2015], Kerleroux [1999]), plusieurs questions surgissent. Tout l'argumentaire de Raccach s'est fondé sur une réduction du domaine de référence sans pour autant définir clairement le domaine de référence et sa quantification. De combien de domaines de référence dispose un mot polysémique ? Un domaine de référence est-il égal à un sens ? Qu'est-ce qu'un sens ? Pour simplifier, nous considérerons ici qu'un sens est une acception d'un mot, autrement dit, il s'agit d'une des définitions données par le dictionnaire. Cela correspond donc à des concepts, des idées, et qui peuvent être exprimées sous forme de mots.

Ainsi, un mot polysémique est un mot qui correspond à plusieurs concepts généralement reliés par métaphore, ou métonymie par exemple. Pour autant, ces concepts peuvent être difficile à catégoriser. Nous interprétons les catégories comme des domaines de références. Un mot peut posséder des domaines, chaque domaine se référant à un concept. Ce serait là ce dont Raccach parle, des mots dont les acceptions se perdent avec la troncation. Autrement dit, un troncat a moins d'acceptions possibles qu'un mot-source car d'une certaine manière le troncat ne fait pas référence au même concept. Cette hypothèse est forte, pour cela nous nous sommes demandés comment tester cette affirmation empiriquement.

La méthodologie que nous proposons repose sur une analyse quantitative. Nombre d'auteurs, en parlant des acceptions des mots, ne font qu'une analyse qualitative en sélectionnant les acceptions que les mots doivent avoir, et en argumentant comment les processus morphologiques opèrent

avec ces acceptions. Ainsi, les acceptions ne sont pas quantifiées, et il est donc possible d'inférer que les locuteurs disposent de toutes ces acceptions lorsqu'ils utilisent une occurrence d'un mot, ce qui est discutable. Notre méthodologie, quant à elle, ne fait pas de présupposition sur les acceptions que doivent posséder les troncats.

Nous avons eu recours à nouveau à une étude de corpus et avons choisi de quantifier les acceptions qu'un mot peut avoir. Nous avons présupposé que le nombre d'acceptions des mots, les lexèmes-bases, pouvait être obtenu à travers un dictionnaire, et qu'il ne s'agirait là que de faire un compte. Le travail de quantification, plus chronophage donc pas envisagé par la littérature, consistait à regarder ce qui se passe quand le mot est utilisé en corpus. Il faut se demander comment le mot est utilisé et avec quelle acception il est employé. Nous avons donc procédé à une annotation manuelle de notre corpus pour obtenir des résultats analysables.

Notre travail s'est basé sur plusieurs questions de recherches que nous avons développées. D'un côté, nous cherchions à mettre en avant l'existence de différences entre le sens des troncats et le sens des bases, et d'un autre d'établir un rôle de la situation sociale.

Pour cela, comme nous avons vu que la troncation était accompagnée d'instructions et de contraintes, nous nous sommes demandés, dans un premier temps, s'il existait un lien de dépendance entre la distribution des sens d'un mot, c'est-à-dire le nombre de sens disponibles pour un mot et les proportions d'utilisation de chaque sens, et le fait que ce mot soit sous forme tronquée ou non. Autrement dit, il s'agit de savoir si la troncation a un effet sur la manière pour les mots d'utiliser les sens, les concepts qui lui sont associés et donc d'avoir la possibilité de mettre en évidence des instructions et des contraintes sémantiques.

Dans le même cadre, nous avons réfléchis sur l'existence possible d'un lien de dépendance entre la distribution des sens et le niveau de formalité du contexte encadrant l'occurrence d'intérêt. Il est possible que le niveau de formalité du discours ait une influence sur les sens véhiculés d'un mot, et que donc certains sens soient uniquement présents dans des contextes infor-

mels et d'autres uniquement dans des contextes formels, en tenant compte, en plus, du fait que ces sens sont directement relié avec un troncat ou une base. Le locuteur choisit l'orientation de son discours en utilisant des formes différentes pour exprimer des sens différents.

Dans un deuxième temps, nous avons voulu savoir s'il est possible de mettre en évidence que certaines formes de mots ont un contenu informatif plus important que d'autres formes. Il faut comprendre que ce qui est informatif est peu prédictible, alors, s'il existe une simplification possible de distribution des sens possibles d'un mot par rapport à sa forme, une singularisation des sens, c'est que certaines formes doivent bien avoir des sens qui deviennent plus prédictibles, donc plus fréquents et moins informatifs. Cela revient à se demander si passer d'une forme non-tronquée à une forme tronquée rend les sens du mots plus prédictibles. Pour cela il faut quantifier l'information relative à chaque mot étudié.

Finalement, une dernière question a émergé. Il s'agissait de savoir s'il était possible de mettre en évidence que certains sens étaient privilégiés par certaines formes d'un mot. Pour cela, il fallait se demander si faire la dichotomie entre sens référant à une action, donc un événement, et les autres sens est importante pour comprendre un des mécanismes sous-jacent de la troncation. Kerleroux a proposé que les événements complexes ne permettaient pas l'apocope à cause de plusieurs propriétés qui leur étaient intrinsèques. Dans la mesure où nous avons vu que nos résultats allaient dans la direction suggérée par Kerleroux, dire que les sens actionnels sont désavantagés par la troncation permet d'appuyer l'argument des contraintes syntaxiques que nous avons vu dans la première partie de ce mémoire. Nous voulons donc mettre en évidence une répercussion de ces contraintes au niveau sémantique.

Le travail sur corpus nous permettra d'esquisser des réponses à ces questions à travers une annotation manuelle des sens des mots-bases et des troncats.

4.2.1 Méthodologie de notre travail sur corpus

4.2.1.1 Corpus

Pour cette deuxième analyse sur corpus, nous avons utilisé le même corpus FRCOW (Schäfer & Bildhauer [2012], Schäfer [2015]) que dans la première partie du mémoire. Néanmoins, nous avons opéré différemment au moment des recherches. Sachant que nous voulions limiter la variation dans les emplois, nous avons établi un sous-corpus issu du corpus général francophone en spécifiant que nous désirions des textes issus de sites web de France, dont la date, indiquée dans les métadonnées des documents du corpus, était postérieure à 1995. De ce corpus, il a ensuite été aisé de sortir à nouveau deux sous-corpus. Nous avons supposé que le niveau de formalité d'un contexte était dépendant de la présence d'une occurrence dans un forum. Ainsi, nous sommes partis de l'idée qu'une occurrence présente dans un forum était associée à un niveau de *formalité informel*, et qu'une occurrence absente d'un forum était associée à un niveau de *formalité formel*. Le premier sous-corpus obtenu correspondait aux documents issus de forums et le deuxième sous-corpus correspondait aux documents issus de non-forums.

Nous avons repris la liste des paires de troncats/bases que nous avons utilisée dans le premier chapitre de ce mémoire pour mettre en évidence les différentes contraintes qui étaient imposées par la troncation. Comme le travail de cette partie aura été de faire une annotation manuelle des sens, nous avons pu réintégrer à notre liste la paire SONO/SONORISATION après avoir trier les occurrences qui n'étaient pas du français.

4.2.1.2 Listes de sens

Nous avons ensuite, pour préparer l'annotation, collecté et élaboré des listes de sens. Dans un premier temps, à partir du Trésor de la langue informatisé (TLFI), nous avons récupéré les listes de définitions complètes pour les dix formes non-tronquées que nous étudions, et, lorsqu'elle était disponible dans le TLFI, une liste de définitions des mots tronquées. A partir des acceptions récupérées, nous avons fait un premier travail d'élaboration de listes de sens en regroupant les acceptions suffisamment proches sémantiquement et en séparant les acceptions suffisamment éloignées sémantiquement

et méritant d’avoir un sens à part entière. Dans un deuxième temps, nous avons fait le même travail avec le Wiktionnaire.

Les définitions dans le TLFi sont organisées selon un plan suivant les différentes constructions possibles du mot ainsi que ses différents emplois. Ainsi, les acceptions sont toujours subsumées. Dans le Wiktionnaire, les définitions ne sont pas catégorisées de la sorte. Nous y retrouvons fréquemment en premier les sens les plus utilisés. De plus, les définitions données par le Wiktionnaire rendent compte de plus de sens émergents des mots, là où « la rédaction du TLF [Trésor de la Langue Française] est terminée depuis 1994 »¹.

Les définitions pour les troncats étaient plus riches dans le Wiktionnaire que dans le TLFi. De plus, nous avons de manière anecdotique ajouté certaines définitions provenant de pages Wikipédia détaillées pour les troncats, nous permettant d’enrichir nos listes de sens que nous avons seulement élaborées avec le TLFi.

En ANNEXE, A page 98, nous présentons les listes de définitions que nous avons retenu pour nos dix paires de bases/troncats. Nous avons ajouté des exemples issus de FRCOW pour illustrer les définitions. Nous illustrerons nos propos par la suite avec la paire PROMOTION/PROMO.

4.2.1.3 Annotation

Pour annoter, nous avons généré à partir de nos deux nouveaux sous-corpus des tirages aléatoires d’occurrences en contexte de mots tronqués ainsi que de mots non-tronqués. Nous reviendrons sur le nombre de tirages aléatoires après avoir expliqué les caractéristiques de ces tirages.

Premièrement, nous avons souhaité travailler avec un nombre assez important d’occurrences, mais comme l’annotation était manuelle nous nous sommes limités à l’annotation de 100 occurrences par forme. Ayant 20 lexèmes, 10 tronqués et 10 non-tronqués, cela représente 2000 occurrences dont nous avons annoté manuellement les sens. Parmi les 100 occurrences par lexème, nous retrouvons 50 occurrences qui proviennent du sous-corpus que nous avons élaboré pour la condition informelle (forum), et 50 autres

1. <http://atilf.atilf.fr/>

occurrences du sous-corpus pour la condition formelle (hors forum). Malheureusement nous n'avons pas pu ne tirer que 50 occurrences aléatoires pour chaque mot dans chaque condition. En effet, nous avons rapidement réalisé que le corpus était bruité et qu'à maintes reprises, les occurrences dans un même contexte se répétaient bien qu'indexées dans des pages web différentes. Pour remédier à ces difficultés, nous avons tirés au moins 100 occurrences à chaque fois pour n'en annoter que 50.

Nous avons récupéré les occurrences dans un document CSV, puis nous avons élaboré des tableaux dans un tableur contenant les informations relatives à la provenance des occurrences, à leur présence ou non dans un forum, à leur caractère tronqué ou non, et des métadonnées en général pour pouvoir répliquer en cas de besoin les tirages aléatoires.

Finalement, nous avons annoté nos 2000 occurrences grâce à nos listes de sens. L'annotation a été réalisé par une seule personne. Nous avons ensuite importé les différents tableaux en format CSV sur Rstudio (RStudio Team [2015]) pour l'analyse des résultats.

4.2.2 Distribution des sens

Nous nous étions posés trois questions principales. Les données que nous avons collectées nous ont permis de répondre à ces questions. Dans un premier temps, nous allons nous intéresser à la distribution des sens, dans un deuxième temps, nous regarderons l'entropie des distributions pour l'information associée aux formes, et finalement, nous analyserons si un type de sens, le sens actionnel, est défavorisé ou bien donné comme absent par la troncation.

4.2.2.1 Observation des distributions

A l'aide de RStudio, nous avons pu visualiser et analyser nos données de distribution de sens. Une distribution des sens correspond à la manière dont les acceptions se répartissent pour un même mot. Autrement dit, ce qui nous intéresse avec les distributions de sens c'est de savoir quelles sont les acceptions sur-représentées pour un mot, et celles qui sont sous-représentées pour le même mot. Les sous-représentations au niveau des troncats peuvent

être vues comme la manifestation de contraintes ou d'instructions comme le supposait Raccach.

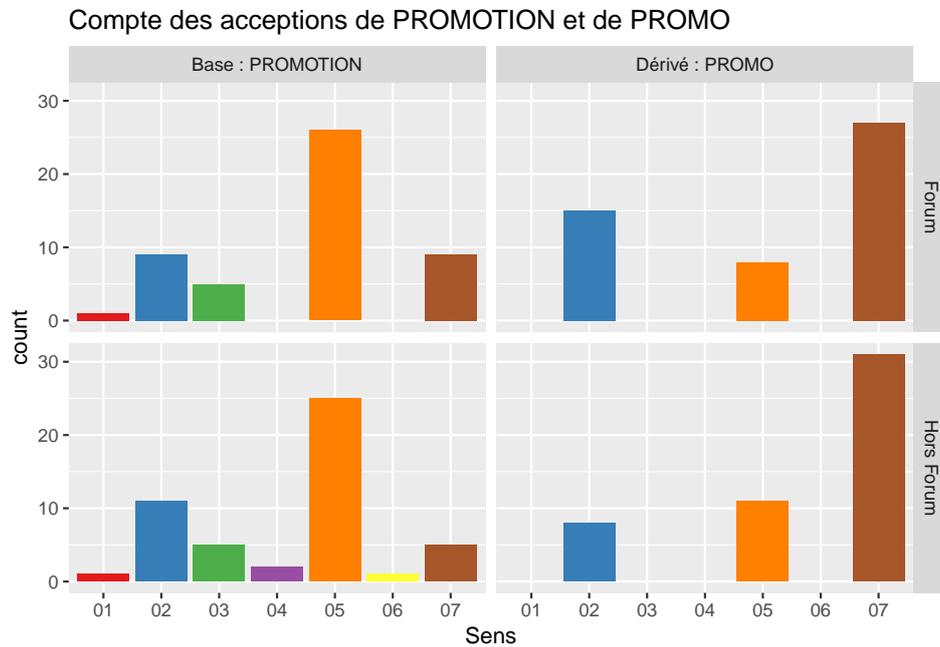


FIGURE 4.1 – Distribution des sens de promotion/promo

Dans le graphique en FIGURE 4.1, nous observons quatre diagrammes à bâtons. Nous nous référerons à chaque diagramme comme représentant une condition particulière. Nous avons donc les sous-histogrammes suivants :

1. En haut à gauche : la distribution des sens du lexème PROMOTION lorsqu'il est dans un forum
2. En haut à droite : la distribution des sens du lexème PROMO lorsqu'il est dans un forum
3. En bas à gauche : la distribution des sens du lexème PROMOTION lorsqu'il est hors forum
4. En bas à droite : la distribution des sens du lexème PROMO lorsqu'il est hors forum

En axe des abscisses, nous retrouvons les différentes acceptions que peut avoir un lexème. Pour le couple PROMOTION/PROMO, sept acceptions sont disponibles. En axe des ordonnées, nous avons le compte du nombre d'oc-

currences associé à chaque acception. Donc, chaque barre dans la figure correspond à un nombre d'occurrences d'un lexème tronqué ou non-tronqué avec une acception spécifique.

Pour notre paire de lexème PROMOTION/PROMO, nous observons que certaines acceptions², les acceptions 1, 3, 4 et 6 ne sont pas représentées pour la forme tronquée *promo*, autrement dit lors de l'annotation de nos 100 occurrences, nous n'avons pas identifié et relié ces sens à la forme tronquée. Néanmoins, cela ne signifie pas qu'avec un plus grand nombre d'occurrences nous ne pourrions rencontrer une forme avec une de ces acceptions. Les acceptions de *promo* se distribuent donc entre les acceptions 2, 5 et 7, l'acception 7 étant graphiquement la plus représentée autant en condition forum qu'en condition non-forum. Les sens de *promotion* sont tous représentés dans la condition non-forum, mais dans la condition forum les sens 4 et 6 ne sont pas représentés.

Dans cet exemple de *promotion*, tous les sens ont été représentés au moins une fois dans une condition, mais ce ne fut pas le cas pour certains autres lexèmes dont nous avons annoté les acceptions. Cela montre une fois de plus que ce que nous observons sont des tendances, et que ce qui est absent ici peut devenir présent mais rare dans un plus grand corpus annoté.

4.2.2.2 Résultats et analyses

Maintenant que nous avons annoté nos données, nous pouvons les analyser. Pour cela, nous avons réalisé des tests du χ^2 de Pearson à l'aide de RStudio. Nous avons simulés les p-valeurs grâce à la simulation de Monte-Carlo avec 1000 répliquions pour diminuer le risque d'erreur. Nous présentons dans un premier temps nos résultats pour le couple de lexème PROMOTION/PROMO.

Pour tester si la présence ou l'absence d'une forme dans un forum avait un effet sur la distribution des sens, nous nous sommes intéressés au compte des sens d'occurrence pour PROMOTION en forum et en non-forum. Les ré-

2. Pour la correspondance des acceptions au numéro que nous attribuerons, se reporter à la liste des définitions en ANNEXE A.7 page 105.

sultats statistiques ($\chi^2(6) = 4.3625$, $p = 0.7223$) ne sont pas significatifs ne nous permettant pas de rejeter l'hypothèse nulle d'indépendance. Nos résultats sont donc compatibles avec l'interprétation possible de qu'il n'y a pas d'effet sur la distribution des sens pour le lexème PROMOTION en fonction de son emploi dans un forum ou dans un non-forum.

De même, en cherchant à savoir si la présence du troncet PROMO dans un forum à un effet sur la distribution de ses acceptions, nous obtenons des résultats similaires ($\chi^2(2) = 2.88$, $p = 0.2517$), autrement dit nous ne pouvions donc pas rejeter l'hypothèse nulle d'indépendance. Nos résultats étaient donc compatibles avec l'interprétation possible qu'il n'y a pas d'effet sur la distribution des sens pour le mot *promo* en fonction de s'il est dans un forum ou dans un autre type de contexte. Donc, il ne semble pas que la présence dans un forum, condition que nous avons étudiée pour vérifier un sens social possible à travers la formalité supposée des énoncés, ait un effet sur la distribution des sens possibles de PROMOTION/PROMO.

Après avoir vu que la formalité n'avait pas d'influence sur la distribution des sens, nous nous sommes demandé si la forme tronquée ou non avait un effet sur la distribution de sens. Lorsque nous ne sommes pas dans des forums, nous obtenons des résultats significatifs ($\chi^2(6) = 33.696$, $p < .001$) nous permettant de rejeter l'hypothèse d'indépendance, autrement dit, la distribution des sens des lexèmes PROMOTION et PROMO est dépendante du lexème tronquée ou non et ce lorsque le lexème est dans un contexte hors d'un forum. De plus, avec les résidus standardisés de Pearson, il nous est possible de regarder quelles sont les acceptions sous-représentées et sur-représentées par rapport à l'hypothèse d'indépendance. Lorsque les lexèmes ne sont pas dans un forum, nous pouvons dire que l'acception 7 du lexème tronquée PROMO est la seule sur-représentée, et que ses acceptions 3 et 5 sont sous-représentées. Les résidus ne sont pas informatifs pour les autres acceptions, les sens 1, 2, 4 et 6.

En s'intéressant cette fois-ci à ce qui se passe dans un forum, nous retrouvons aussi des résultats significatifs ($\chi^2(2) = 26.029$, $p < .001$) nous permettant de rejeter à nouveau l'hypothèse d'indépendance entre la forme

d'un mot et sa distribution de sens. De même que lorsque nous n'étions pas en forum, nous retrouvons que la distribution des sens des lexèmes PROMOTION et PROMO est dépendante du fait que la forme est tronquée ou non dans un forum. Par rapport à l'hypothèse d'indépendance, les résidus standardisés de Pearson nous indiquent que seule l'acceptation 7 est sur-représentée pour le troncat PROMO et que les acceptations 3 et 5 sont sous-représentées lorsque le lexème PROMO est dans un forum. Les résidus ne sont pas informatifs pour les acceptations 1 et 2. Nous rappelons que les acceptations 4 et 6 étaient absentes des annotations pour cette condition forum autant pour PROMOTION que pour PROMO.

À partir du couple de lexème base/troncat qu'est PROMOTION/PROMO, nous avons déjà pu avoir un premier aperçu de l'effet de la forme sur ses différentes acceptations. Néanmoins, nous n'avons pas observé le fait d'être dans un forum pour une forme avait un effet sur la distribution des acceptations.

Nous avons répliqué nos analyses statistiques à l'ensemble de nos mots. Nous présentons ces résultats de test de χ^2 de Pearson dans le TABLEAU 4.1. Nous avons coloré en rouge pâle les cases du tableau correspondant à un test du χ^2 dont le résultat était significatif. De même que pour la paire PROMOTION/PROMO, nous observons à partir du tableau que nos résultats significatifs mettent en avant un effet de la forme d'un mot sur sa distribution de sens. Pour 8 couples base/troncat sur 10, leur distribution de sens est complètement indépendante d'être dans un forum ou non. Pour les deux derniers couples, uniquement la forme présente ou absente dans un forum a un effet sur la distribution des sens.

Il est aussi intéressant de voir que même en analysant un nombre de paires plus important, nous avons à nouveau un manque de résultats significatifs pour dire que la présence d'une forme dans un forum ou non peut avoir un effet sur la distribution de ses sens. En effet, pour six cas sur dix, nous avons une absence totale de résultats significatifs au test de χ^2 mettant en avant une fois de plus que nos résultats sont compatibles avec l'interprétation possible de l'absence d'effet de la présence d'une forme dans un forum ou non sur la distribution de ses sens ou qu'en tout cas si différence il y a elle est trop subtile pour être observable avec un jeu de données de cette taille.

	Effet de Forum/Hors Forum						Effet de Troncat/Base					
	Troncat			Base			Forum			Hors Forum		
	χ^2	DL	Pval	χ^2	DL	Pval	χ^2	DL	Pval	χ^2	DL	Pval
Compilation	3.8655	2	0.1778	1.569	2	0.5485	11.663	2	0.0049	6.348	2	0.0229
Décoration	5.3752	1	0.0039	5.3752	2	0.0699	15.316	2	0.0009	5.0571	2	0.0599
Démonstration	5.6579	6	0.4575	6.3039	5	0.2957	60.202	7	0.0009	49.334	6	0.0009
Information	5.7222	4	0.2158	9.8772	5	0.0669	1.1151	4	0.9061	20.036	5	0.0009
Intoxication	2.0833	2	0.4186	3.4572	3	0.3397	74.785	2	0.0009	60.6	3	0.0009
Manifestation	2.0408	1	0.4955	20.057	5	0.0019	29.6	5	0.0009	47.059	3	0.0009
Promotion	2.88	2	0.2517	4.3625	6	0.7223	26.029	4	0.0009	33.696	6	0.0009
Récréation	2.2104	1	0.2318	8.2101	3	0.0339	6.5473	3	0.0359	15.082	3	0.0019
Rédaction	3.4667	3	0.3626	14.466	4	0.0029	18.108	4	0.0009	36.641	4	0.0009
Sonorisation	3.5202	2	0.2118	4.2605	2	0.1159	34.41	2	0.0009	29.963	2	0.0009

TABLE 4.1 – Résultats des test d’indépendance du Chi2

Nos données ne nous permettent pas de dire qu’il existe un sens social lié directement à la distribution des sens d’un mot. Nos données vont en effet dans le sens de notre première hypothèse qui était de dire qu’il existe un lien entre la forme d’un lexème, apocopé ou non, et les sens disponibles auxquels ce lexème peut être associé. La distribution des sens d’un mot dépend de sa forme.

Néanmoins, nos résultats mettent en avant que la distribution des sens d’un mot ne dépend pas majoritairement du fait que ce mot soit présent dans un contexte formel ou non. Ce manque de lien entre cette distribution et la formalité pourrait s’expliquer du fait que le corpus est bruité et donc que les tirages aléatoires sont eux-mêmes bruités, entraînant un manque de représentativité des acceptions. Il est aussi possible d’expliquer ce manque de lien par la taille insuffisante des données, qui rend difficile la détection d’un effet subtil, dès lors que nous supposons qu’il y a un effet. Si le travail d’annotation n’était pas manuel, il serait intéressant de travailler avec un

nombre beaucoup plus important de couples de bases/troncats. Finalement, une dernière explication importante possible est de dire tout simplement que la distinction que nous avons opér e entre forum et hors forum n'est pas suffisamment corr el ee avec la distinction formel/informel. Ainsi, il faudrait pouvoir trouver d'autres moyens de mesurer le degr e de formalit e en corpus.

4.2.2.3 Troncats, occurrences et forums

L'hypoth ese de l'importance du contexte est justifi ee puisque cela correspond  a une des motivations de la troncation. Cependant, nous n'avons pas trouv e d'effet significatif lorsqu'il s'agissait de mettre en lien les sens et la formalit e du contexte. Souhaitant v erifier cette hypoth ese de l'importance du contexte pour la troncation, nous avons repris nos occurrences en forum et hors forum. Nous les avons quantifi es. Le contexte devrait avoir un r ole d'inhibiteur ou d'activateur de la troncation. Ce r ole est coupl e avec d'autres effets et n'est pas seul responsable de ce qui va  tre tronqu e ou non. Nous avons choisi de repr esenter ce contexte   travers son niveau de formalit e.

Nous avons pour cela compt e les troncats et les bases en fonction de leur pr esence dans un forum ou non en nous limitant aux paires de notre liste. Nous pr esentons le graphique en FIGURE 4.2 correspondant aux diff erentes proportions obtenues pour tous les lex emes en fonction de leur pr esence dans un forum ou non et sous leur forme tronqu ee ou non. Le tout est normalis e.

Nous pouvons dire qu'il y a une diff erence de nombres d'occurrences relatives pour les troncats selon que nous avons trouv e en forum et hors forum. Si nous nous int eressons dans un premier temps aux bases, nous pouvons dire qu'en proportions relatives, elles repr esentent plus de 80% des occurrences dans les non-forums contre environ 40% des occurrences dans les forums. A l'inverse, les lex emes tronqu es ne repr esentent qu'environ 20% des occurrences dans les non-forums, et environ 60% des occurrences dans les forums. Les troncats sont significativement sur-repr esent es dans les forums et sous-repr esent es dans les non-forums ($\chi^2(12) = 257290$, $p < .001$), autrement dit le nombre d'occurrences de troncats et de non-troncats est donc d ependant de sa pr esence du lex eme dans un forum ou non.

Donc bien que le niveau de formalit e ne joue pas un r ole sur le sens

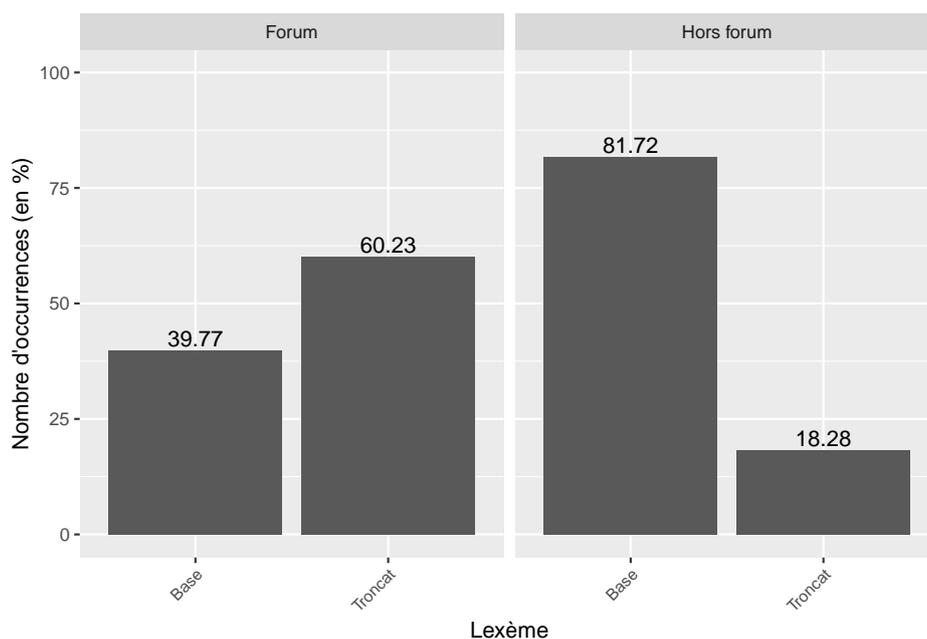


FIGURE 4.2 – Répartition du nombre de cooccurrences des troncats et bases en fonction de leur présence en forum ou dans hors forum

véhiculé par une forme, il semble néanmoins qu'il joue un rôle sur le nombre relatif de formes utilisées.

4.2.3 Entropie des distributions de sens

Notre deuxième question de recherche de cette partie consistait à vérifier si lorsqu'un mot est tronqué, cela induit une spécialisation dans son domaine de référence. En effet, nous avons vu, par exemple, que EXAMEN une fois tronqué donne EXAM, et que le troncat résultant avait un domaine de référence limité au domaine scolaire. Ainsi, il était possible de parler d'*exam de chimie* mais pas d'un *exam médical*. Cette spécialisation du domaine de référence peut être vu comme une réduction dans les sens possibles d'un mot, et donc l'information de chaque distribution de sens. Mais qu'est-ce que l'information d'une distribution de sens ?

Pour comprendre la notion de quantité d'information, nous devons retourner à la théorie de l'information et de la communication (Shannon

[1948]) et à la notion de surprise. Certains éléments dans le discours sont plus prévisible que d'autres. La notion de surprise se base sur ce qui est prévisible. Ce qui est surprenant ce sont les éléments auxquels nous ne nous attendons pas. Par exemple, prenons le début de phrase suivante : *Je bois de la soupe pour...*, il est peu surprenant de continuer la phrase avec le verbe *grandir* alors qu'il le serait plus avec *manifeste*. L'information apportée est plus importante avec *manifeste* qu'avec *grandir*. Ce qui est probable transmet peu d'information alors que ce qui est peu probable transmet plus d'information. Cette quantité d'information transmise est appelée la *surprise*. La surprise est noté S et vaut $S = -\log_2 p$, avec p une probabilité et comme unité le bit. Lorsque p tend vers 0, S tend vers $+\infty$, lorsque $p = 1$, $S = 0$. La surprise est infinie quand un événement impossible se produit, en revanche elle est nulle lorsqu'un événement certain se produit.

L'entropie H d'une distribution est une mesure d'information qui reprend cette idée de surprise et correspond à une valeur moyenne de la surprise de la distribution. Le nombre de valeurs que contient cette distribution correspond au nombre de valeur que peut prendre une variable aléatoire, l'entropie étant celle d'une variable aléatoire. L'entropie rend compte de la quantité d'information moyenne produite et de l'incertitude moyenne associée à cette quantité d'information. Cette entropie H vaut :

$$H = - \sum_{i=1}^n (p_i \times \log_2 p_i)$$

Avec p_1, \dots, p_n appartenant à un ensemble de probabilités qui correspondent aux différentes valeurs à l'intérieur de notre distribution. De la sorte, il est possible d'en déduire que si tous les p_i sont égaux à $\frac{1}{n}$, alors H croit en fonction de n , plus ce dernier est élevé, plus la valeur de H sera grande. Une grande valeur de H correspond à l'entropie d'une distribution statistique équilibrée, pouvant être associée à beaucoup de valeurs. Pour revenir à nos distributions de sens, nous pouvons imaginer le cas d'un mot possédant 5 acceptions. Lors de l'annotation, nous aurions révélé qu'en forum chaque acception est présente 10 fois, et qu'en dehors des forums, chaque acception est aussi présente 10 fois. Par exemple, nous aurions pour l'entropie associée à la distribution des acceptions dans les forums :

$$\begin{aligned}
H &= -\left(\left(\frac{1}{5} \times \log_2 \frac{1}{5}\right) + \left(\frac{1}{5} \times \log_2 \frac{1}{5}\right)\right) \\
&\approx 2.3219
\end{aligned}$$

De même, nous aurions eu une valeur égale en non-forum. Pour nous donner un ordre de comparaison, si nous avions eu une distribution avec exactement deux sens équiprobables, alors nous aurions eu une entropie $H = 1$.

Et d'un autre côté, il est aussi possible d'avoir une distribution statistique déséquilibrée et/ou avec peu de valeurs. Dans ce cas là, l'entropie serait basse. Reprenons notre exemple mais cette fois-ci avec un autre mot possédant toujours 5 acceptions mais dont le nombre d'occurrences annotées est disproportionné entre les acceptions. Supposons qu'une acception représente 92% des occurrences annotées (pour cette exemple nous ne faisons pas de distinction entre forum et hors forum), elle est donc très probable, et que les 4 autres acceptions se partagent équitablement les 8% restant, alors nous aurions eu l'entropie suivante :

$$\begin{aligned}
H &= -\left(\left(\frac{92}{100} \times \log_2 \frac{92}{100}\right) + \left(\frac{2}{100} \times \log_2 \frac{2}{100}\right) + \left(\frac{2}{100} \times \log_2 \frac{2}{100}\right) + \left(\frac{2}{100} \times \log_2 \frac{2}{100}\right) + \left(\frac{2}{100} \times \log_2 \frac{2}{100}\right)\right) \\
&\approx 0.5622
\end{aligned}$$

Cette valeur d'entropie basse reflète une situation sans trop d'incertitude. Nous avons donc un exemple où la surprise moyenne est moins importante que celle de l'exemple précédent.

Finalement, notre dernier exemple correspond à un mot polysémique, toujours à 5 acceptions, mais dont nous n'avons annoté qu'une acception, représentant donc 100% des acceptions annotées. Nous obtenons l'entropie suivante :

$$\begin{aligned}
H &= -\left(\left(1 \times \log_2(1)\right)\right) \\
&= 0
\end{aligned}$$

Une valeur d'entropie nulle est donc associée à une distribution sans incertitude.

Maintenant que nous avons vu ce qu'était l'entropie, revenons à notre question de recherche. En effet, s'il y a une spécialisation des sens au moment de la troncation, alors nous pouvons nous attendre à une diminution de l'entropie. La surprise moyenne des bases doit être plus élevée que celle des troncats. Les bases non spécialisées doivent regrouper plus de sens, que nous supposons répartis équitablement. Cela équivaut à une entropie élevée. Pour les troncats, nous nous attendons à ce que l'entropie soit relativement plus basse. Le processus de troncation a pour conséquence une réduction du domaine de référence qui provoque indirectement un renforcement de certains sens, en excluant donc les autres, ce qui rend les acceptations du mot plus prévisibles. La distribution comportant moins de valeurs doit avoir une entropie plus basse.

Nous avons calculé les différentes entropies de nos distributions en fonction de nos quatre conditions et les résultats sont présentés dans le TABLEAU 4.2. Nous avons coloré en bleu les cases qui correspondent à une diminution d'entropie, et en rouge les cases qui correspondent à une augmentation de l'entropie.

A partir du tableau, nous pouvons observer que pour une paire base/troncat dans un contexte autre qu'un forum, dans 9 cas sur 10, nous avons une diminution de l'entropie. Lorsque la paire est dans un forum, dans 7 cas sur 10 nous avons une diminution de l'entropie.

Nous avons vérifié nos résultats en procédant à deux tests de Student appariés sur les moyennes des entropies. Dans un premier temps, nous avons considéré la condition forum et comparé la moyenne des entropies des troncats avec celle des bases. Le résultat du test de Student est significatif ($t(9) = -2.1741, p = 0.028$), ce qui nous permet de rejeter l'hypothèse nulle que les deux moyennes sont égales. Ayant comme alternative que la différence entre les moyennes ($\mu_{entropie-troncat} - \mu_{entropie-base}$) est inférieure à 0, nous pouvons dire que la moyenne des entropies pour les troncats est significativement inférieure à celle des bases. Dans un forum, avoir un lexème

	Hors Forum		Forum	
	Base	Troncat	Base	Troncat
compilation	1.0512	0.5411	1.2299	0.9161
décoration	1.2597	0.8994	0.9284	0.9849
démonstration	1.9582	1.6319	2.2579	1.4325
information	2.3412	2.0024	2.0256	2.0995
intoxication	1.1273	1.1914	0.6395	1.1069
manifestation	1.6146	0.0000	2.1484	0.2422
promotion	2.0564	1.3311	1.8262	1.4241
récréation	1.6494	0.6800	1.0844	0.4021
rédaction	1.3925	1.1024	1.8111	0.8841
sonorisation	1.1974	0.3825	0.9593	0.8176

TABLE 4.2 – Mesure d’entropie pour les distributions de sens

tronqué présente moins d’incertitude et est donc plus informatif sur les sens possibles véhiculés par rapport à un lexème non-tronqué. Nous retrouvons bien la réduction de sens recherchée.

Dans un second temps, nous avons considéré la condition hors forum et nous avons de nouveau comparé les moyennes des entropies. Le résultat du test de Student était à nouveau significatif ($t(9) = -3.976, p < .001$), mettant en évidence qu’il y a une réduction de l’entropie moyenne entre la distribution des sens des bases et la distribution des sens des troncats.

Nous avons cherché à vérifier s’il existait une réduction de sens entre deux bases lorsque nous passons d’un contexte hors forum à forum ou inversement, néanmoins, nos résultats n’étaient pas significatifs ($p > 0.2$) ce qui est compatible avec l’interprétation possible que la formalité n’a pas de rôle dans la réduction de la distribution des sens pour une même forme de lexème.

A travers cette deuxième question de recherche, nous avons mis en avant que, comme prédit par la littérature, il y a bien une réduction des sens lorsque nous passons de la distribution des sens d'un mot non-tronqué à un mot tronqué. Cette diminution de l'entropie a été vue comme une restriction du domaine de référence. N'ayant analysé que des mots polysémiques, nous aurions dû nous attendre à une spécialisation et une réduction des sens complète correspondant au passage d'une entropie haute à une entropie nulle, mais ce ne fut pas le cas.

4.2.4 Modèle linéaire mixte de la spécialisation des sens

Finalement nous analyserons si la troncation favorise ou défavorise certains types de sens. L'intérêt de répondre à cette question nous vient de l'interprétation des arguments de Kerleroux et de la littérature sur les noms d'événements. En effet, selon la linguiste les noms d'événements complexes ne sont pas apocopables. Néanmoins, nous avons argumenté que son point de vue ne devrait pas être considéré comme absolu mais plutôt comme une tendance. Les noms déverbaux d'événements simples et les noms résultatifs, quant à eux, sont troncables. Kerleroux argumentait pour une troncation s'exerçant de manière lacunaire sur ces catégories.

Pour ce nouveau chapitre, nous sommes partis de l'idée que les noms d'événements, indépendamment de si l'action est vue dans son déroulement ou non, sont défavorisés par la troncation. Par noms d'actions nous faisons référence aux noms d'événements complexes ainsi qu'aux noms d'événements simples. Si les noms d'actions sont défavorisés par la troncation, tout ce qui n'est pas un nom d'action a plus de place pour faire un troncat. Autrement dit, les noms résultatifs doivent être représentatifs de la troncation. Pour cela, nous avons cherché à savoir s'il était possible de prédire la forme d'un lexème en fonction de son sens et de son contexte d'apparition. Autrement dit, nous souhaitons mettre en évidence qu'il existe un lien entre si le lexème est tronqué ou non, son sens et la formalité du contexte, ce qui correspond aux informations dont nous disposions.

4.2.4.1 Les données

Nous avons récupéré les données issues de nos annotations où nous avons des occurrences en contexte avec un sens qui leur était associé. Nous avons repris nos listes de sens, et nous avons regroupé d'un côté tous les sens événementiels, c'est-à-dire tous les sens qui étaient en relation avec une action, et d'un autre côté les autres sens, non-événementiels. Par exemple, si nous reprenons la paire *promotion/promo* nous avons d'un côté les sens du type « *Action d'élever simultanément plusieurs personnes à un même grade* » et d'un autre côté les sens du type « *Opération commerciale afin d'attirer l'attention sur une offre ou un produit* ». En annexe, nous présentons avec nos listes de sens, les regroupement de sens que nous avons effectué. A chaque définition est associé une étiquette, [*Ev*] lorsqu'il s'agit d'un sens événementiel, et [*NoEv*] lorsqu'il ne s'agit pas d'un sens événementiel.

	Hors forum		Forum	
	Base	Troncat	Base	Troncat
Sens non-événementiels	269 (59,78%)	363 (80,85%)	279 (62%)	360 (80,72%)
Sens événementiels	181 (40,22%)	86 (19,15%)	171 (38%)	86 (19,28%)

TABLE 4.3 – Compte des sens événementiels et des sens non-événementiels

A partir du TABLEAU 4.3, nous pouvons observer la présence de deux phénomènes, et l'absence d'un phénomène. Premièrement, nous observons que pour les sens événementiels, il y a une diminution de moitié entre la base et le troncat, autant en forum que hors des forums.

Deuxièmement, nous observons une augmentation uniquement pour les sens non-événementiels de leur nombre entre la base et le troncat, autant en forum que hors des forums. Il est important de souligner que proportionnellement, le nombre d'occurrences avec un sens non-événementiel est toujours supérieur au nombre d'occurrences avec un sens événementiel. Finalement, nous n'observons pas de différence entre les proportions d'utilisation des sens en condition forum et hors forum.

Nous avons ensuite élaboré notre modèle linéaire mixte grâce au pa-

ckage Lme4 et la fonction Glmer (Bates et al. [2015]) sous R qui cherchait à prédire la forme d'un mot (pouvant avoir deux valeurs, tronqué ou non tronqué) en fonction de la condition forum correspondant à la formalité (si le mot est dans un forum ou non) et de son sens (s'il s'agit d'un sens événementiel ou non-événementiel). Nous avons ajouté un effet aléatoire pour l'influence que les différents lexèmes pouvaient avoir, correspondant aux couples base/troncat, puisqu'il est possible de trouver une variabilité dans la distribution des sens en fonction des couples. Nous avons centré nos effets fixes et aléatoires. Nous présentons nos résultats³ en FIGURE 4.3 page 76.

Les résultats de notre modèle nous indiquent que pour prédire la forme d'un mot, il y a un effet très significatif ($p < .001$) du sens. Nous ne retrouvons pas d'effet de la présence du mot en forum. Le modèle ne rend pas compte d'effet de la formalité sur la forme d'un lexème. Le contexte ne semble donc pas une source révélatrice d'information pour savoir si la forme qui sera utilisée le sera pour véhiculer un sens précis d'un mot. Nous n'avons pas de corrélation entre le sens et le forum sur la forme que peut avoir un mot, le modèle ne rend pas compte d'interaction entre les deux effets fixes.

Cela est cohérent avec nos attentes. En effet, comme les sens actionnels sont désavantagés il est possible d'inférer qu'une forme avec un sens actionnel aura plus de chance d'être une forme longue, alors qu'une forme avec un sens non-actionnel aura proportionnellement plus de chance d'être une forme tronquée. Pourtant, il faut tenir compte du fait que les deux types de sens sont présents pour des mots qui ont une forme tronquée et non-tronquée, et qu'il ne s'agit donc pas ici d'un résultat absolu mais bien d'une tendance. Nous pouvons toujours nous demander quels sont les facteurs qui conditionnent l'émergence des troncats que nous pourrions rajouter dans notre modèle. En effet, nous avons vu que le nombre d'occurrences des troncats et des bases change en fonction de si le lexème est en ou hors forum, mais rajouter cette information pourrait poser problème si nous souhaitions refaire un modèle qui en tient compte. Il faudrait imaginer la pos-

3. Mes remerciements vont à Doriane Gras qui m'a aidé à faire ce modèle.

```

## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: Troncat ~ Sens * Formalite + (Sens * Forum | Lexeme)
## Data: table_4regression
## Control: lme4::glmerControl(optimizer = c("bobyqa"))
##
##      AIC      BIC   logLik deviance df.resid
## 2398.7  2475.6 -1185.3  2370.7    1781
##
## Scaled residuals:
##      Min      1Q  Median      3Q      Max
## -2.7068 -0.9136  0.3533  1.0383  1.8508
##
## Random effects:
## Groups Name          Variance Std.Dev. Corr
## Lexeme (Intercept) 0.12746  0.3570
##      Sens           0.47490  0.6891  -0.26
##      Forum          0.01104  0.1050   0.25  0.83
##      Sens:Forum     0.84230  0.9178  -0.14  0.32  0.49
## Number of obs: 1795, groups: Lexeme, 9
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.05402    0.14055   0.384   0.701
## Sens        -1.45833    0.30797  -4.735 2.19e-06 ***
## Forum         0.06875    0.13449   0.511   0.609
## Sens:Forum  -0.29786    0.46132  -0.646   0.518
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Sens   Forum
## Sens          -0.344
## Forum          0.046  0.163
## Sens:Forum    -0.089  0.168 -0.203

```

FIGURE 4.3 – Modèle Linéaire Mixte pour prédire la forme d'un lexème en fonction du sens et de la formalité du contexte

sibilité d'annoter le sens de toutes les occurrences des troncats et des bases en contexte forum et hors-forum. Il serait aussi possible d'échantillonner le nombre d'occurrences à annoter en fonction des proportions d'occurrences que nous retrouvons dans le corpus. Néanmoins, pour cette analyse nous en resterons ici, car le résultat, bien qu'issu d'un échantillon, reflète déjà la tendance qu'a la troncation de défavoriser les noms d'action, et de favoriser les noms qui ne sont pas d'actions.

4.2.5 Conclusion intermédiaire

Dans cette deuxième partie du mémoire, nous avons vu que la distributions des sens d'un mot est dépendante de sa forme. Ainsi, la manière pour un mot d'être associé à des concepts change en fonction de sa forme. Cela peut être dû à des instructions ou contraintes sémantiques qui spécifient que certaines acceptions doivent être associées différemment avec les troncats.

Les résultats de nos tests d'entropie vont dans ce sens. Nous avons vu que l'entropie des distributions de sens diminuait, mettant en évidence la diminution de la représentation des sens opérée par la troncation. La troncation est donc un phénomène qui simplifie la forme des mots d'un côté, mais également leur sens qui deviennent donc moins surprenants car plus prédictibles.

Pour voir quels sens étaient réduits par la troncation et étaient directement touchés par les contraintes sémantiques, nous avons élaboré un modèle linéaire mixte qui nous a permis de faire ressortir la possibilité de prédire la forme d'un mot en fonction de son sens. Le modèle a mis en évidence que les sens événementiels sont désavantagés par la troncation ce qui confirme l'hypothèse que la troncation préfère les noms de dénomination d'entités.

Nous n'avons pas trouvé d'effet significatif nous permettant de dire que la formalité avait un effet sur les distributions de sens, néanmoins, nous devons garder en tête que le nombre d'occurrences des troncats est lui dépendant de la formalité du contexte. Nous avons retrouvé proportionnellement plus de troncats en forum qu'en dehors des forums.

Finalement, nos résultats vont dans le sens d'une interprétation tendancielle de la troncation et non absolue comme cela avait été défendu dans la littérature.

4.3 Le sens de la troncation

Nous n'avions pas voulu faire d'hypothèse sur le statut morphologique de la troncation en employant le terme *base*, terme connoté linguistiquement. Sans aucune intention de tergiverser, nous devons enfin nous fixer sur son statut. Nous avons argumenté dans la première partie de ce mémoire qu'il était possible de voir la troncation comme une réalisation du moindre effort. Les locuteurs pour minimiser l'effort articulatoire tronquent certains mots. En effet, la forme de certains mots peut être prévisible en fonction du contexte et du début du mot, l'information contenue dans la fin du mot n'est plus importante, et le locuteur peut se permettre de tronquer. La forme tronquée devient simplement ce que nous pourrions considérer comme une abréviation qui peut être reprise ensuite par les locuteurs pour en faire des formes d'argots ou de jargons.

Néanmoins, bien que minimiser l'effort soit important pour l'émergence des troncats, nous pensons que ce n'est pas la seule raison. En effet, nous devons caractériser l'information morphologique apportée par le phénomène pour voir ce qui est spécifique de la troncation et ce qui relève de la morphologie. Nous allons argumenter que la troncation est un phénomène morphologique à part entière qui permet la formation d'un nouveau lexème à partir d'un lexème. Un lexème représente une collection abstraite de mots qui partagent un sens lexical. Nous avons appelé *base* et *troncat* deux lexèmes déverbaux polysémiques qui sont liés par une règle de construction de lexème (Fradin & Kerleroux [2009]). Nous reviendrons sur cette règle de construction de lexèmes par la suite. De la même manière qu'en (29a) nous avons une règle pour passer de *manifestester* à *manifestation*, nous avons une règle en (29b) pour passer de *manifestation* à *manif*.

- (29) a. *manifestester* → *manifestation*
b. *manifestation* → *manif*
c. *mathématiques* → *maths* → *matheux* ↔ *matheuse*
d. *cinématographe* → *cinéma* → *ciné*

Un lexème est donc dérivable par une opération. Il est donc normal de s'attendre à ce que les troncats soient eux-mêmes dérivables, ce qui est le

cas en (29c) où nous retrouvons une règle qui permet d’exprimer avec le suffixe X-eux celui qui fait des mathématiques, des maths. Nous pouvons supposer que l’évolution de *matheux* \leftrightarrow *matheuse* est directement corrélée à celle du lexème MATHS.

Un autre argument pour considérer la troncation comme un moyen de formation des lexèmes nous vient du fait qu’il est possible de tronquer à nouveau un lexème qui a déjà été tronqué, quand il reste encore de la matière à tronquer. Par exemple, en (29d) le lexème CINÉMA, qui est déjà une forme tronquée, peut lui-même être tronqué sous la forme *ciné*, considérée comme un nouveau lexème.

Nous avons donc montré que la troncation permettait la dérivation d’un nouveau lexème à travers les règles de construction de lexèmes. Pourtant, le rôle de ces règles n’est pas encore clair. Pour cela, il nous faut nous demander si les changements que nous avons observés sont spécifiques à la troncation ou s’ils sont présents dans d’autres phénomènes de dérivation. Dans le deuxième cas, cela aurait pour implication de les rendre caractéristiques des règles de dérivation des lexèmes. Nous avons décidé de refaire le travail d’annotation des sens avec un autre phénomène dérivationnel pour trancher. Nous avons réutilisé la même méthodologie pour rendre nos résultats comparables à ceux de la troncation. Nous avons donc travaillé avec les diminutifs verbaux. Si nous partons du postulat que raccourcir la forme d’un lexème réduit son nombre de sens, alors allonger la forme d’un lexème pourrait avoir pour conséquence d’augmenter son nombre de sens, ou à minima de laisser inchangé son nombre de sens.

4.3.1 Comparaison avec les diminutifs

Avant d’attaquer directement notre comparaison avec les diminutifs, nous devons comprendre ce que sont les diminutifs verbaux. Ce que nous appelons diminutifs correspond à des modificateurs verbaux (Dubois et al. [1999]). Ce type de modificateurs se présente sous forme de suffixe. Du point de vue morphologique ce suffixe peut-être vu soit comme un morphème soit comme un pseudo-morphème étant donné son manque d’autonomie (To-

vena & Kihm [2008]). Les diminutifs comme catégorie sémantique servent à exprimer plusieurs concepts à travers les langues comme la petite taille, l'affection, l'approximation, l'intensification, l'imitation ou encore le genre féminin (Jurafsky [1996]). Les modificateurs verbaux quant à eux peuvent servir à atténuer ou ajouter un aspect de fréquence répétitif, ou péjoratif. Chaque suffixe peut transmettre un sens supplémentaire au verbe qu'il dérive. Nous retrouvons ainsi par exemple les verbes en *-iller*, les verbes en *-ailler*, *-asser*, *-oyer*, *-ouiller*, etc. Il est même possible dans certains cas d'ajouter plusieurs suffixes comme avec *mordre* + *ill* + *onn* donnant MOR-DILLONNER.

L'intérêt d'étudier les diminutifs est d'établir une comparaison avec la troncation pour en dégager des mécanismes communs. En effet, comme pour la troncation, le processus de formation des diminutifs n'entraîne pas un changement des parties du discours. Nous obtenons un verbe déverbal, là où avec la troncation nous obtenions un nom dénominal issu d'un nom déverbal. Comme pour la troncation, la base et le dérivé peuvent désigner la même chose. Cela permet de parler du même concept de différentes manières. Le troncat, comme le diminutif, peut apporter en plus, comme nous l'avons supposé, un sens social et une connotation qui lui est associée, des éléments de compréhension du discours qui ne sont pas d'ordre lexical. Finalement, étudier les verbes déverbaux nous permet de travailler directement avec des lexèmes qui dénotent des événements. Cela constitue un atout dans la mesure où nous avons vu que la troncation défavorise les noms d'événements.

Les diminutifs allongent donc la forme des mots, là où les troncats la raccourcissent, rendant la comparaison entre les deux intéressante.

4.3.1.1 Quand nous ajoutons au lieu d'enlever...

Pour étudier les modificateurs verbaux avec la même méthodologie que celle que nous avons utilisé avec les troncats, nous avons constitué une liste de cinq paires de lexèmes. Chaque lexème-base que nous avons sélectionné donnait un lexème diminutif. Nous avons dû sélectionner des paires suffi-

samment représentées dans notre corpus, c'est-à-dire avec au minimum 100 occurrences, 50 présentes dans un forum et 50 présentes hors forum. Le premier problème que nous avons rencontré est que beaucoup des verbes diminutifs donnés dans la littérature étaient peu attestés dans notre sous-corpus en forum. Après avoir analysé les troncats, nous n'avions aucune raison de penser que la présence ou l'absence d'un diminutif dans un forum pourrait avoir un effet sur ses sens ou sa distribution de sens, néanmoins, nous avons préféré chercher des verbes conforme à la méthodologie que nous avons proposée. Nous n'avons pas motivé la sélection de nos verbes sur d'autres critères qui auraient pu être aspectuels. C'est ainsi que nous avons choisi de travailler avec une liste de verbes diminutifs avec *X-ouiller* et *X-iller* comme désinence suffixale pour exprimer le diminutif que nous proposons en (30).

- (30) a. Glander → Glandouiller
 b. Gratter → Gratouiller
 c. Mâcher → Mâchouiller
 d. Mordre → Mordiller
 e. Sauter → Sautiller

Dans un premier temps, après avoir sélectionné nos paires, nous avons constitué des listes de sens toujours à partir du TLFi et de Wiktionnaire. Il est intéressant de remarquer qu'en travaillant avec des verbes, nous avons des lexèmes hautement polysémiques. En moyenne nous avons $\mu = 13,8$ acceptions, la médiane de notre nombre d'acceptions est à 10. Pour illustrer nos propos, nous allons nous intéresser au cas de MÂCHER/MÂCHOUILLER comme nous l'avons fait avec PROMOTION/PROMO en guise d'exemple.

Nous avons pour cela élaboré une liste de 10 acceptions pour MÂCHER/MÂCHOUILLER. Ci-dessous, nous énumérons ces sens en les exemplifiant avec notre corpus. Les exemples sont donc le reflet de la manière dont nous avons interprété les sens et avons annoté. L'ensemble des listes de définitions est présenté en ANNEXE B, page 111.

1. Action de broyer un aliment à l'aide des dents et de la mâchoire.
 — *Les oiseaux ne **mâchent** pas ce qu' ils mangent [...]*

2. Réduire en mordillant avec les dents et/ou en humectant de salive.
 - [...] *les feuilles de coca par exemple , **mâchées** dans certaines régions pauvres [...]*
3. Mordiller, mâchonner.
 - [...] *on ne me suce plus , on me **mâche** [...]*
4. Préparer pour faciliter la tâche à quelqu'un.
 - [...] *cherche tu trouveras je ne vais pas te **mâcher** le travail [...]*
5. Prononcer indistinctement.
 - [...] *et je **mâche** mes mots [...]*
6. S'exprimer de façon claire et sans ménagement [si forme négative].
 - [...] *Jean-Louis Triaud n' a pas **mâché** ses mots au micro [...]*
7. Penser sans cesse à quelque chose.
 - *Enfermé dans ses ratiocinations **mâchouillant** de mornes pensées [...]*
8. Couper quelque chose sans faire une section nette, déchirer.
 - [...] *on **mâche** dans la mouillure de la feuille de salade [...]*
9. User ou marquer de façon irrégulière.
 - [...] *il s' agit peut-être d' un docu déjà **mâché** [...]*
10. Meurtrir.
 - *Les combos lourds pour **mâcher** et broyer ses adversaires et les combos légers [...]*

Nous avons ensuite procédé à l'annotation suivant nos quatre conditions comme définies pour l'annotation des troncats. Nous présentons les résultats de cette annotation pour *mâcher/mâchouiller* dans la FIGURE B.3 où nous pouvons observer les distributions de sens selon nos conditions. Chaque condition correspond à un panel de la figure incluant un diagramme en barre.

La première chose que nous observons est que nous n'avons pas d'occurrences associées à toutes les acceptions pour le verbe-base *mâcher*. En effet l'acception 7 issue du TLFi pouvant être paraphrasée par REMÂCHER,

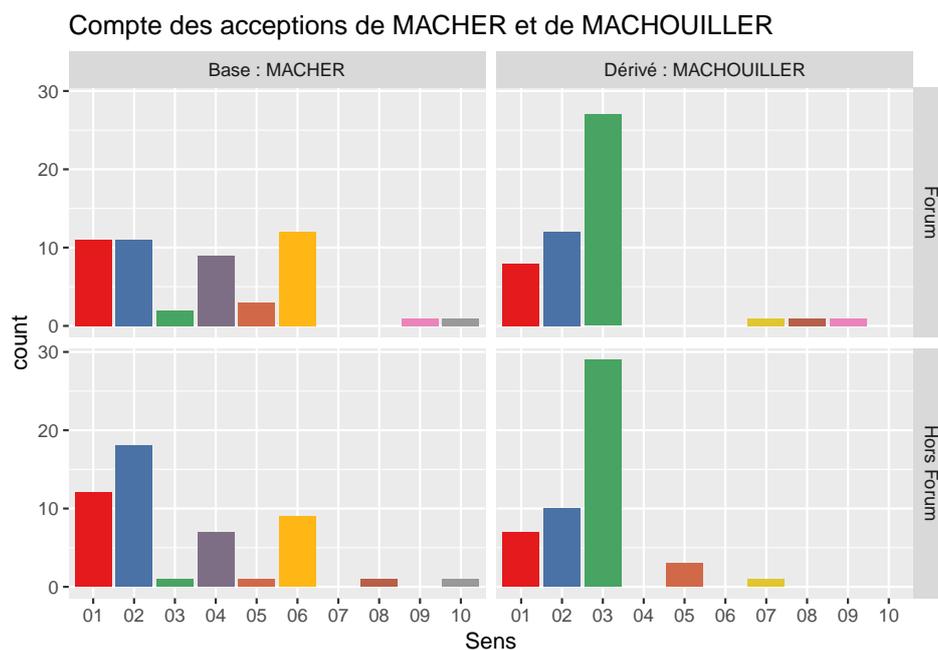


FIGURE 4.4 – Distribution des sens de *mâcher/mâchouiller*

RUMINER⁴ n'était présente que pour le diminutif MÂCHOUILLER. De plus, nous pouvons observer que l'acceptation 3 semble être privilégiée par la forme diminutive là où elle est marginalisée, mais quand même présente, pour le verbe base. La distribution des sens en forum change du verbe base au verbe diminutif ($\chi^2(9) = 49.069$, $p < .001$), tout comme la distribution des sens hors forum ($\chi^2(8) = 49.735$, $p < .001$), confirmant que la formation du diminutif entraîne bien un changement dans la distribution des sens. En effet, les acceptations privilégiées par le diminutif sont différentes des acceptations privilégiées par le verbe-base. Nous n'avons retrouvé d'effet significatif pour l'effet de la formalité à travers forum/non-forum ni pour les bases ($\chi^2(6) = 5.3199$, $p = 0.5165$) ni pour les diminutifs ($\chi^2(8) = 5.745$, $p = 0.7592$). Nous présentons dans le TABLEAU 4.4, afin de les analyser, les résultats des tests d'indépendance du χ^2 de Pearson pour les autres couples de verbes-bases/diminutifs.

Les résultats obtenus pour tous les verbes nous indiquent qu'il y a une tendance pour les verbes et leurs diminutifs à avoir une distribution diffé-

4. Ces synonymes nous sont donnés directement par le TLFi

	Effet de Forum/Hors Forum						Effet de Dérivé/Non-Dérivé					
	Dérivé			Non-Dérivé			Forum			Hors Forum		
	Chi2	DL	Pval	Chi2	DL	Pval	Chi2	DL	Pval	Chi2	DL	Pval
Glander	1.1111	2	1	1.6123	3	0.7582	4.2941	3	0.2977	3.277	3	0.4126
Gratter	8.1208	9	0.5584	15.247	13	0.2787	30.292	13	0.0009	41.658	12	0.0009
Mâcher	5.3199	6	0.5165	5.745	8	0.7592	49.069	9	0.0009	49.735	8	0.0009
Mordre	6.3443	7	1	5.7808	11	0.9131	87.952	11	0.0009	70.328	11	0.0009
Sauter	20.978	9	0.0029	29.929	22	0.0649	73.938	19	0.0009	77.752	21	0.0009

TABLE 4.4 – Résultats des test d'indépendance du Chi2

rente de leur sens. Il est possible de voir que la présence d'une forme dans un forum ou hors d'un forum n'a aucun effet sur la distribution de ses sens en général.

Nos résultats sont donc cohérents avec ce que nous avons obtenu pour la troncation et peut laisser penser que le changement dans la distribution des sens des lexèmes polysémiques est interne à la dérivation en général et non aux processus spécifiques comme la troncation ou le processus relatif aux diminutifs.

Finalement, nous nous sommes intéressés à voir si ce changement dans les distributions était lui-aussi accompagné d'une réduction dans les sens, avec certains sens sur-représentés et d'autres sous-représentés ou absents. Nous avons pour cela calculé l'entropie des distributions, c'est à dire la surprise moyenne. Nous présentons nos résultats dans le TABLEAU 4.5 avec une coloration en bleu associée aux cases où nous observons une diminution de l'entropie.

Nos résultats montrent que les sens des diminutifs sont plus prédictibles, il y a donc moins d'incertitude associée à un diminutif qu'à un verbe-base. La réduction des sens entre la base et le dérivé est significative dans un forum ($t(4) = -3.1509, p = 0.01724$) et dans un non-forum

	Hors Forum		Forum	
	Base	Dérivé	Base	Dérivé
glander	0.9004	0.4689	1.0361	0.5411
gratter	3.0130	2.7911	3.2338	2.6688
mâcher	2.3186	1.6737	2.5556	1.7358
mordre	3.0859	0.7011	2.9665	0.5221
sauter	3.9235	2.3815	3.9431	1.9209

TABLE 4.5 – Mesure d’entropie pour les distributions de sens

($t(4) = -2.59, p = 0.03034$).

Nos résultats sont une fois de plus similaires à ceux que nous avons obtenus avec la troncation. Il semblerait que les processus dérivationnels aient pour caractéristique de sur-représenter certains sens et d’en sur-représenter d’autres ayant pour conséquence de rendre les sens plus prévisibles pour les dérivés. Nous allons argumenter ce point de vue dans la sous-section suivante.

4.3.1.2 Résumé des résultats

Autant pour les distributions de sens que pour les mesures d’entropie, nous obtenons des résultats similaires à ceux obtenus avec la troncation. Une fois de plus, la distribution des sens d’un mot est dépendante de sa forme qui peut être dérivée à l’aide d’un modifieur verbal ou à l’aide de la troncation par exemple. Les acceptions dont la fréquence a été modifiée deviennent plus prédictibles avec la forme dérivée. Ces résultats laissent penser que pour les mots polysémiques il peut exister un phénomène sous-jacent associé à la dérivation qui permet en surface d’obtenir ces résultats.

CHAPITRE 5

CONCLUSION

LA MÉTHODOLOGIE que nous avons élaborée pour étudier les différentes acceptions des bases et des dérivés doit être intégrée dans une perspective plus grande que celle de ce mémoire. En effet, les langues, et nous supposons que c'est valable pour toutes les langues où la dérivation est possible, disposent d'outils pour créer de nouveaux mots à partir de mots préexistants. Ces outils, nous les avons définis comme des règles de construction de lexèmes en argumentant que les dérivés étaient eux-mêmes aptes à resservir à des règles morphologiques. Nous souhaitons montrer dans cette conclusion que Fradin & Kerleroux [2009] ne rendent compte que d'une partie de la complexité de ces règles.

5.1 Les règles de construction de lexèmes

Nous avons abordé la question des règles de construction de lexèmes pour comprendre ce qui était de leur ressort et ce qui était spécifique à chaque phénomène dérivationnel. Nous avons mis en relation un changement dans la distribution des sens communs à la troncation et à la formation de dérivés. La question qui surgit est donc de savoir comment les règles de construction de lexèmes opèrent lorsqu'il s'agit de construire de nouveaux lexèmes avec des sens particuliers, des acceptions en relation.

Premièrement, nous sommes partis de l'idée que les sens du dérivé ne sont pas spécifiés par la dérivation. Nous parlerons d'une non-spécialisation des sens. En effet, il serait tentant d'affirmer que les dérivés ont accès aux mêmes acceptions que les bases dont ils sont issus. Dans le cas particulier de la troncation, cela permettrait de justifier le manque de contraintes apparentes pour annoter les différentes acceptions des lexèmes tronqués puisque toutes les significations étaient à priori possibles car données pour la base. Par exemple, en FIGURE 5.1 nous avons deux lexèmes reliés par une règle de construction de lexèmes. Le phénomène de dérivation ici est sous-spécifié. Le lexème base à gauche dans le graphique possède n acceptions. Ce sont ces n acceptions qui sont reprises par la règle de construction de lexèmes (RCL) et qui sont directement associées au nouveau lexème. Aucune contrainte sur la manière d'associer ces sens n'est spécifiée. Ainsi, il est possible d'obtenir un lexème dérivé possédant l'intégralité des sens de la base. Pourtant, cette vision n'est pas idéale, car instinctivement un locuteur n'a pas accès à toutes les acceptions d'une base, et encore moins à toutes les acceptions du dérivé. Si les locuteurs avaient accès à toutes les acceptions, alors, lors des annotations, toutes les acceptions devraient être représentés de manière égale ce qui n'a été le cas ni pour la troncation ni pour les diminutifs.

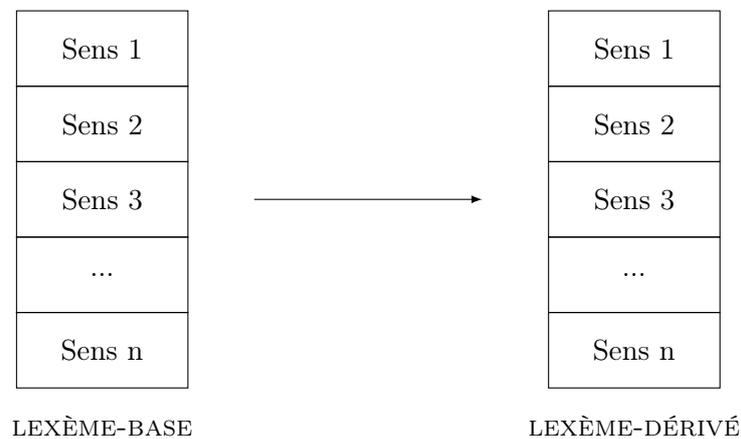


FIGURE 5.1 – RCL et non-spécialisation des sens

Alors pourquoi cette vision n'est pas la bonne ? N'est-ce pas ce qui a été

avancé dans la littérature ? Non. La littérature semble diverger de ce point de vue. En effet, il a souvent été argumenté que les RCL construisent des lexèmes sur la base d'un sens (Fradin & Kerleroux [2009]). Fradin et Kerleroux argumentent que pour les lexèmes polysémiques « tout dérivé se forme sur une des significations associées à [ce lexème] ». Cela leur permet de justifier que tous les dérivés morphologiques avec des significations spécifiques ne sont pas possibles. L'opération de dérivation est vue comme une opération discriminante, et les RCL en sont les protagonistes puisqu'elles favorisent une acception particulière aux dépens des autres à partir de spécifications sémantiques.

Cette acception privilégiée sera donc associée au dérivé, le dérivé pouvant lui-même acquérir de nouvelles acceptions par la suite. Dans notre exemple FIGURE 5.2, nous illustrons ce point de vue où la dérivation est accompagnée d'une spécialisation des sens. Parmi nos n sens qui sont inhérents au lexème de base, seule l'acception 3 est associée au dérivé (le choix de cette acception particulière sert simplement à illustrer notre propos mais n'a aucune implication par la suite).

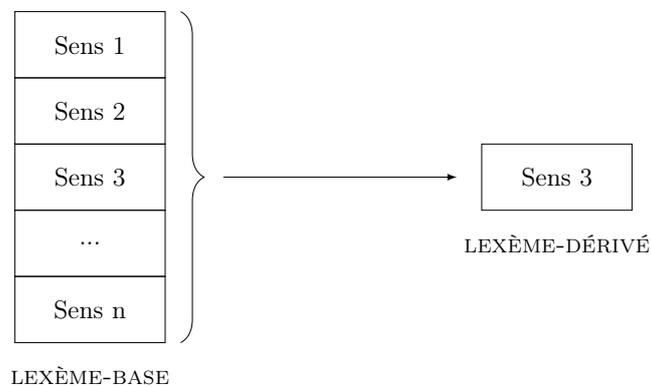


FIGURE 5.2 – RCL et spécialisation des sens

Dans cette logique, il est normal de dire, en supposant toutes les acceptions que le lexème *examen* a, que la forme tronquée *exam* ne se réfère qu'au domaine scolaire et est compatible avec cette vision de l'opération de règles de construction de lexèmes. Néanmoins à aucun moment cette analyse ne prend en compte la possibilité d'avoir un *exam* qui se réfère à

autre chose qu'un examen du milieu scolaire. Or, nous avons trouvé des occurrences d'*exam* en faisant des recherches basiques sur Google, sur Yahoo Answers et sur Twitter se référant au domaine médical. Pour représenter que certains dérivés peuvent être associés à plusieurs acceptions, Fradin et Kerleroux parlent de pluralité des lexèmes distincts. Cela leur permet d'éviter de parler de lexème-dérivé polysémique. Nous en faisons la représentation graphique dans la FIGURE 5.3 où un nombre n de lexèmes est obtenu à partir d'un lexème polysémique. Cette vision a la principale implication de dire que pour que les n lexèmes soient attestés, il y a besoin d'autant de spécifications sémantiques associées à autant de RCL.

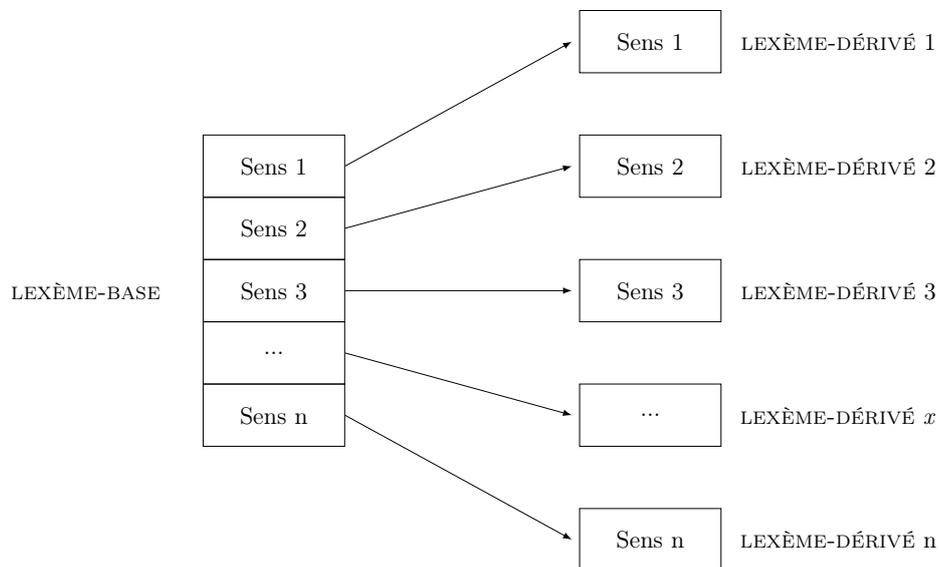


FIGURE 5.3 – Obtention de n lexèmes-dérivés à partir d'un lexème polysémique

Il est donc peu concevable de garder cette vision de la dérivation car elle ne semble pas parcimonieuse dans son utilisation des spécifications sémantiques. Pour cela, nous proposons une alternative qui conserve l'attractivité d'une approche non-spécialisée du sens mais qui inclut des contraintes rendant compte de la spécialisation des sens observée (surreprésentation ou sous-représentation de certains sens). Ainsi, nous supposons que les spécifications sémantiques doivent être vues comme un outil de pondération sur les acceptions au moment de la dérivation.

Notre argumentaire consiste à établir que les règles de construction de lexèmes ne se fondent pas sur des acceptions particulières mais des distributions de probabilités. Nous avons vu que tous les sens n'étaient pas représentés de la même manière. A chaque lexème-base est associée une distribution de probabilités correspondant à la connaissance globale du mot et de ses acceptions par les locuteurs de la langue. En surface, au moment de l'annotation, cette distribution est visible à travers les comptes de sens pour chaque occurrence. Le compte total pour certains sens est plus important que pour d'autres, certains sens étant absents, cela traduit de leur rareté mais non de leur impossibilité. Comme représenté dans la FIGURE 5.4, supposons que nous avons un lexème-base, à gauche, possédant n sens. Dans ce cas de figure, nous pouvons présumer que le lexème-dérivé, à droite, aura autant de sens que le lexème de base, ce qui correspond à un point de vue où les acceptions choisies par la dérivation ne sont pas spécifiées. Pourtant la règle de construction de lexèmes permettant le passage d'une forme à l'autre tient compte de l'information statistique contenue dans le lexème de base, intègre l'information spécifique au phénomène de dérivation, et à partir de ces différentes informations élabore une nouvelle distribution statistique pour le lexème dérivé donnant un lexème-dérivé polysémique et non plusieurs lexèmes-dérivés monosémiques. Par exemple, une des informations apportées par la troncation est que les sens événementiels sont désavantagés. Cette information aura pour conséquence de désavantager la fréquence des lexèmes avec une acception qui correspond à une interprétation événementielle.

En outre, dans notre exemple, l'acception 3 n'est pas désavantagée par la RCL aux dépens des autres sens. La fréquence d'utilisation des autres acceptions diminue. Cela peut être vu comme une information qui a été apportée par le processus morphologique dérivationnel et qui défavorise toutes les acceptions sauf la troisième. Les autres acceptions sont toujours présentes mais sont moins représentées. Dans cette sous-représentation, il existe encore des inégalités dues à d'autres informations apportées à la RCL. Dans le langage courant, cela se traduit par peu de personne utilisant un lexème dérivé avec une acception marginalisée. Ainsi notre représentation de la dé-

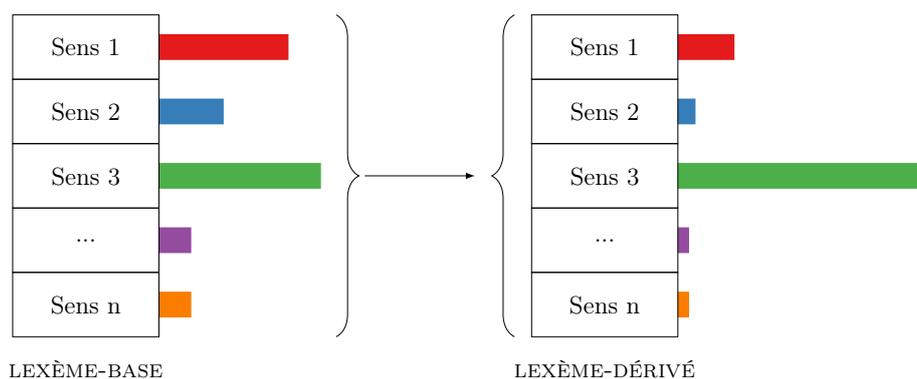


FIGURE 5.4 – RCL et distribution de probabilités des sens

rivation a le mérite de faire exister ces sens et de potentiellement les rendre accessibles aux locuteurs. Cette analyse rend aussi compte du fait qu’une première acception peut être sur-représentée, et qu’une autre peut être fortement représentée, cela dépendra de l’information de la RCL.

5.2 Discussion

Dans ce mémoire, nous avons cherché à mettre en avant ce qui avait été dit par la littérature à propos de la troncation. L’état de l’art sur la troncation n’est pas pauvre, pourtant nous avons cherché à garder une position détachée de ce qui avait été affirmé. A travers deux études de corpus, et un parallèle avec les modificateurs verbaux sous forme de diminutifs, nous avons essayé de déterminer quel était le but de la troncation. Notre intérêt s’est porté sur ce qui était spécifique à la troncation et ce qui était commun à différents processus morphologiques. Nous avons mis en avant l’économie syllabique obtenue par la troncation. Les mots tronqués sont plus courts, et proportionnellement plus utilisés en contexte informel.

Tous les mots ne sont pas facilement tronçables. Ainsi parmi les différentes classes de noms existant, celles qui font références à des événements, c’est-à-dire les noms représentant une action plutôt qu’une entité, sont désavantagées. La structure sous-jacente syntaxique interne au mot est plus compliquée pour les événements lorsqu’ils sont vus dans leur déroulement.

Il y a là une complexité qui défavorise ces formes. De plus, c'est justement cette structure qui les rend proches des verbes desquels elles dérivent dont la structure est similaire. En omettant ce qui caractérise les noms déverbaux en *-tion*, c'est-à-dire le suffixe *-tion*, il est possible de supposer que les locuteurs ont moins de repères pour identifier le sens du nom comme événementiel ce qui rend difficile la troncation.

Pour autant, la troncation est un processus morphologique productif. Il y a d'un autre côté des forces qui permettent de tronquer. Elles permettent de passer de mots longs, dont la structure peut être grammaticalisée, à des mots courts. Ces forces, nous avons argumenté, sont des manifestations du principe du moindre effort. Les troncats sont donc les formes résultantes de la compétition entre différentes contraintes et ce sont les règles de construction de lexèmes qui disposent de ces différentes informations. Il est intéressant de se demander comment quantifier ces forces expérimentalement. La troncation est un moyen de faire évoluer la langue car il est possible de former par la suite de nouveaux lexèmes sur la base des troncats avec de nouvelles règles de construction de lexèmes dont une des manifestations sera la réduction des sens qui leur est intrinsèque.

En conclusion, la dérivation réduit le sens des mots, et la troncation réduit la forme des mots.

RÉFÉRENCES

- Antoine, F. (2000). Dictionnaire français-anglais des mots tronqués (Vol. 105). Peeters Publishers.
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis : A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Language and speech, 47(1), 31–56.
- Bates, D., Mächler, M., BenBolker, & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Bonami, O., Boyé, G., & Kerleroux, F. (2009). L'allomorphie radicale et la relation flexion-construction. Aperçus de morphologie du français, 103–125.
- Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations : Syntactic projections and the lexicon. The nature of explanation in linguistic theory, 31–67.
- Dister, A. (1998). Aspects phonétiques et morphologiques de l'apocope en français contemporain. DOCUMENT RESUME, 53.
- Dubois, J., Dubois-Charlier, F., & Dubois-Charlier, F. (1999). La dérivation suffixale en français. Nathan.

- Fradin, B., & Kerleroux, F. (2009). L'identité lexémique. Presses universitaires de Vincennes.
- Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally : Uniform information density as an optimal strategy for language production. In Proceedings of the annual meeting of the cognitive science society (Vol. 30).
- Fridrichová, R. (2012). La troncation en tant que procédé d'abréviation de mots et sa perception dans le français contemporain. Université Palacký d'Olomouc.
- Fridrichová, R. (2013). Quelques observations sur les mots tronqués dans le français contemporain. Romanica Olomucensia(1), 1–13.
- Grimm, S., & McNally, L. (2013). No ordered arguments needed for nouns. In Proceedings of the 19th amsterdam colloquium (pp. 123–130).
- Grimshaw, J. (1990). Argument structure. the MIT Press.
- Haspelmath, M., & Sims, A. (2013). Understanding morphology. Routledge.
- Jamet, D. (2009). A morphophonological approach to clipping in english. can the study of clipping be formalized? Lexis. Journal in English Lexicology(HS 1).
- Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. Language, 533–578.
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort : Language users optimise a miniature lexicon for efficient communication. Cognition, 165, 45–52.
- Kanwal, J. K. (2018). Word length and the principle of least effort : language as an evolving, efficient code for information transfer (Thèse de doctorat non publiée). The University of Edinburgh.
- Kerleroux, F. (1999). Sur quelles bases opère l'apocope? Sillexicales 2 : la morphologie des dérivés évaluatifs, 95–106.

- Luo, B., Jiang, T., Liu, Y., Wang, J., Lin, A., Wei, X., & Feng, J. (2013). Brevity is prevalent in bat short-range communication. Journal of Comparative Physiology A, 199(4), 325–333.
- Mahfoudh-Hubert, H. B. (2009). Un collaborateur n'est pas toujours vu comme un collabo. Entre sens et signification : Constitution du sens : points de vue sur l'articulation sémantique-pragmatique, 157.
- McCarthy, J. J., & Prince, A. (1996). Prosodic morphology 1986.
- Ochs, E. (1993). Constructing social identity : A language socialization perspective. Research on language and social interaction, 26(3), 287–306.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. Proceedings of the National Academy of Sciences, 108(9), 3526–3529.
- Plénat, M. (1994). L'extramétricité des voyelles initiales. French Generative Phonology : Retrospective and Perspective, Salford, University of Salford, AFLS & ESRI, 239–257.
- Plénat, M. (1999). Prologomènes à une étude variationniste des hypocoristiques à redoublement en français. Cahiers de grammaire, 24, 183–219.
- Raccah, P.-Y. (2015). Quelques infos sur les nœuds-pap et sur les collabos.
- Roy, I., & Soare, E. (2013). Event-related nominals. Cambridge Scholars Publishing.
- RStudio Team. (2015). Rstudio : Integrated development environment for r [Manuel de logiciel]. Boston, MA. Consulté sur <http://www.rstudio.com/>
- Saint-Dizier, P. (2006). Rôles thématiques. In D. Godard, L. Roussarie, & F. Corblin (Eds.), Sémanticlopédie : dictionnaire de sémantique. <http://www.semantique-gdr.net/dico/> : GDR Sémantique & Modélisation, CNRS. Consulté sur http://www.semantique-gdr.net/dico/index.php/R%C3%B4les_th%C3%A9matiques
- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In Proceedings of challenges in the management

- of large corpora 3 (cmlc-3). Lancaster : IDS. Consulté sur <http://rolandschaefer.net/?p=749>
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In Proceedings of the eight international conference on language resources and evaluation (lrec'12) (pp. 486–493). Istanbul, Turkey : European Language Resources Association (ELRA). Consulté sur <http://rolandschaefer.net/?p=70>
- Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. Biology letters, 6(4), 469–471.
- Shannon, C. E. (1948). A mathematical theory of communication. Bell system technical journal, 27(3), 379–423.
- Smith, K. (2004). The evolution of vocabulary. Journal of theoretical biology, 228(1), 127–142.
- Tovena, L. M., & Kihm, A. (2008). Event internal pluractional verbs in some romance languages. Recherches linguistiques de Vincennes(37), 9–30.
- Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. Memory & Cognition, 7(6), 411–419.
- Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wesley press.

Annexes

ANNEXE A

LISTES DE SENS

A.1 COMPILATION/COMPIL

1. Action de compiler [*Ev*]
 - [...] organiser une **compilation** dans le but d'aider le financement des associations [...]
2. Recueil de textes et de documents pris à divers ouvrages. [*NoEv*]
 - Merci à Emmanuel Valeur pour la **compilation** musicale.
3. En informatique, action de compiler du code source grâce à un compilateur. [*Ev*]
 - [...] je lui balance mon dossier latex La **compilation** tourne nickel et nous sort un beau pdf [...]

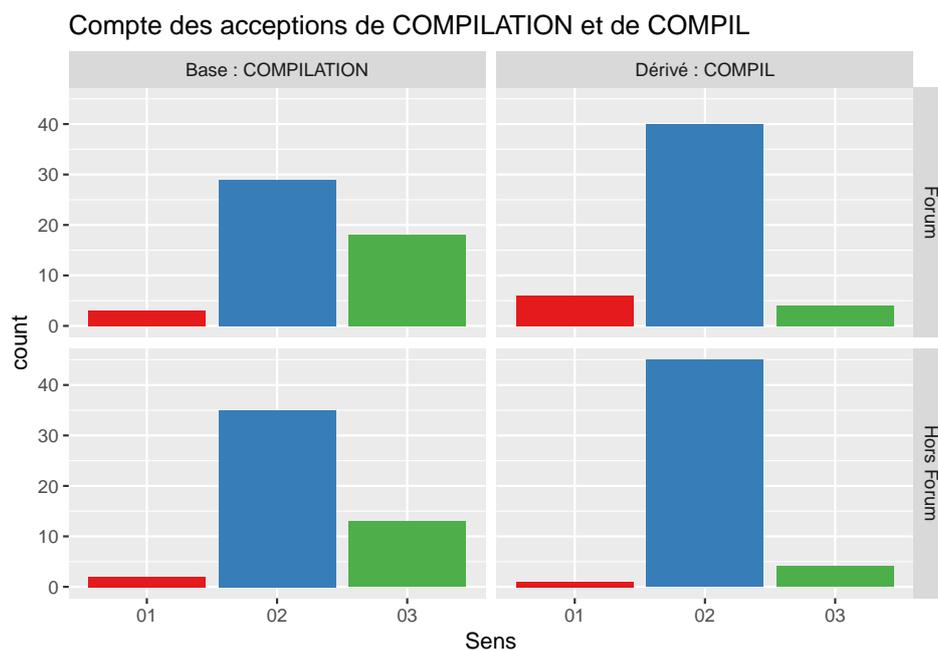


FIGURE A.1 – Distribution des sens de COMPILATION et de COMPIL

A.2 DÉCORATION/DÉCO

1. Action de décorer. [*Ev*]
 - [...] portant affectation de crédits pour des travaux de **décoration** dans les bâtiments d'enseignement [...]
2. Ce qui sert à décorer. [*NoEv*]
 - [...] Luxueux, volumineux, emplis de voilage, de poufs, de **décorations** diverses (statues colossales, objet précieux, des serveurs partout) [...]
3. Au théâtre, décor. [*NoEv*]
 - Non attesté.
4. Récompense honorifique. [*NoEv*]
 - [...] peut-être aussi par les épaulettes et la **décoration** du soldat.

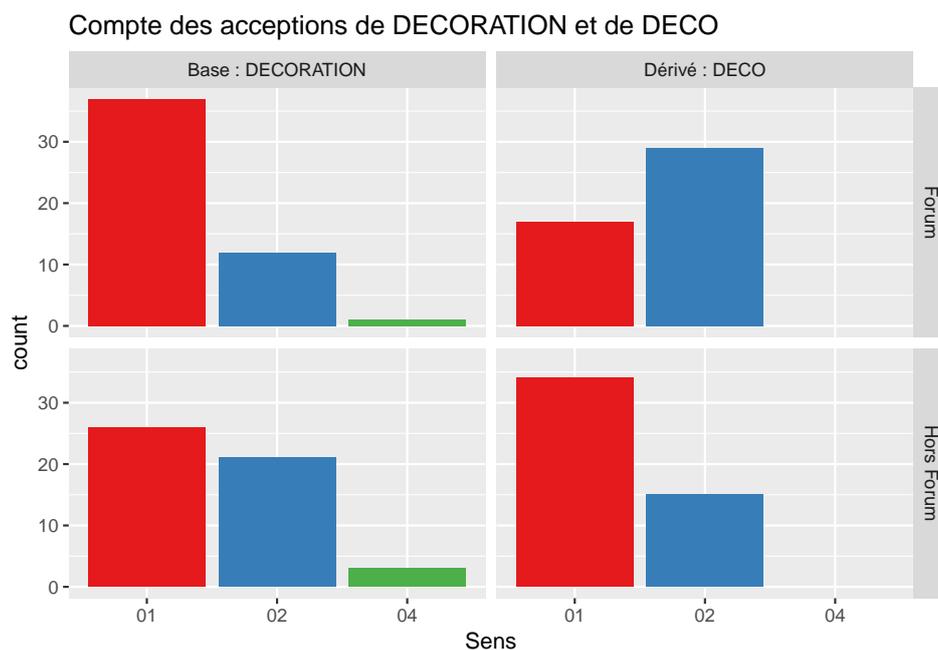


FIGURE A.2 – Distribution des sens de DÉCORATION et de DÉCO

A.3 DÉMONSTRATION/DÉMO

1. Action de montrer en quoi consiste, comment fonctionne X [art, sport, communication] [*Ev*]
 - [...] *Je suis à votre disposition pour toute **démonstration** [...]*
2. Action de montrer pour vaincre (armée, manifestation de masse) [*Ev*]
 - [...] *Manifestations obligées, **démonstrations** de force impé-
ratives mais aussi discussions [...]*
3. Signe extérieur qui manifeste les sentiments, intentions d’une per-
sonne. [*NoEv*]
 - [...] *elle faisait la **démonstration** silencieuse du caractè-
re grotesque de cet idéal [...]*
4. Action de démontrer par le raisonnement [*Ev*]
 - [...] ***démonstration** par l’absurde pourtant je pensais les belges
très second degré [...]*
5. Ce qui sert à démontrer quelque chose. [*NoEv*]

- [...] votre **démonstration** qui n'en est pas une est donc infondée et minable mais cela n'est pas une surprise [...]
- 6. Logiciel, version de démonstration [*NoEv*]
 - Je n'ai testé que la version de **démonstration** [...]
- 7. Enregistrement musical [*NoEv*]
 - [...] Il manque la **Demo** et 2 chansons d'un split japonais [...]
- 8. Production faite par un demomaker [*NoEv*]
 - Je me régale avec mon guerrier , ma **démo** aussi (surtout que j'ai le marcheur du Vide depuis hier) [...]

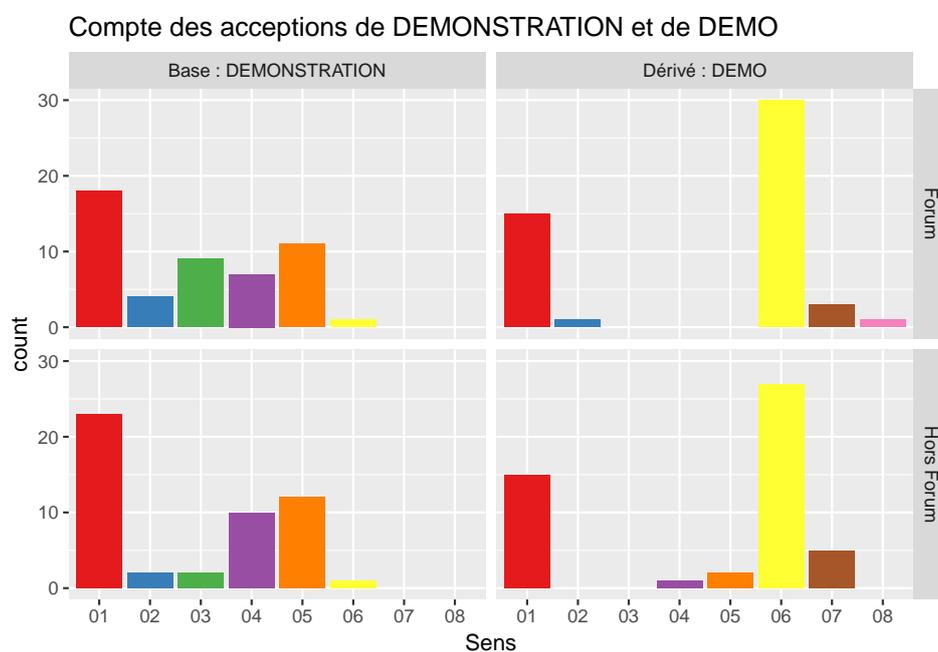


FIGURE A.3 – Distribution des sens de DÉMONSTRATION et de DÉMO

A.4 INFORMATION/INFO

1. Action d'informer ou recherche que l'on fait. [*Ev*]
 - [...] car aujourd'hui quasiment aucune prévention ni **information** n'est faite...
2. Tout média qui apporte des informations. [*NoEv*]

- *Les mutations de l'information et des médias locaux et régionaux [...]*
- 3. Renseignement sur quelque chose. [*NoEv*]
 - *Vous cherchez d'autres informations au sujet d'un master à distance [...]*
- 4. Faits, événements nouveaux, événements d'intérêt général. [*NoEv*]
 - *Reste à savoir comment l'information circule dans ce dernier cas [...]*
- 5. Instruction judiciaire, enquête faite. [*Ev*]
 - *Le commissaire aux comptes examine les éléments d'information communiqués [...]*
- 6. Connaissances réunies sur un sujet déterminé. [*NoEv*]
 - *[...] elle contient un traitement automatisé d'informations nominatives [...]*

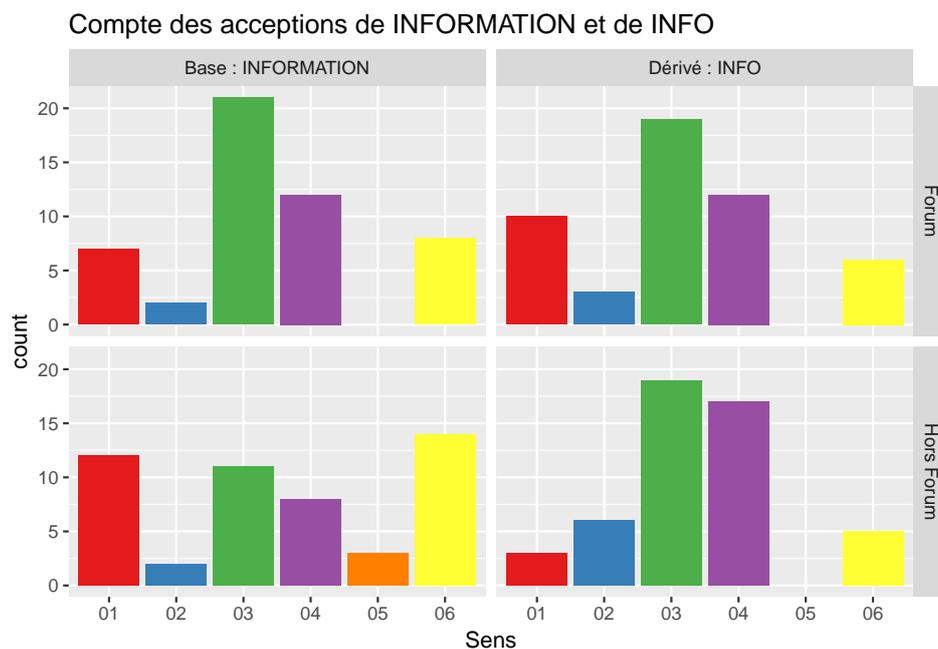


FIGURE A.4 – Distribution des sens de INFORMATION et de INFO

A.5 INTOXICATION/INTOX

1. Action d'intoxiquer ou résultat de cette action. [*Ev*]
 - [...] les **intoxications** aiguës par le pétrole chez l'enfant au Gabon [...]
2. État mental provoqué par l'obsession d'une idée. [*NoEv*]
 - [...] **intoxication** générée par la médiacratie et ses mensonges [...]
3. Action de propagande par la désinformation. [*Ev*]
 - L'**intoxication** représente une réaction contre cette violence de la modernité [...]
4. Canular qui cherche à faire passer pour vrai ce qui est faux. [*NoEv*]
 - Combien de fois l'accusation devra-t-elle être répétée pour que cette **intoxication** - si intoxication il y a, fasse finalement effet ?

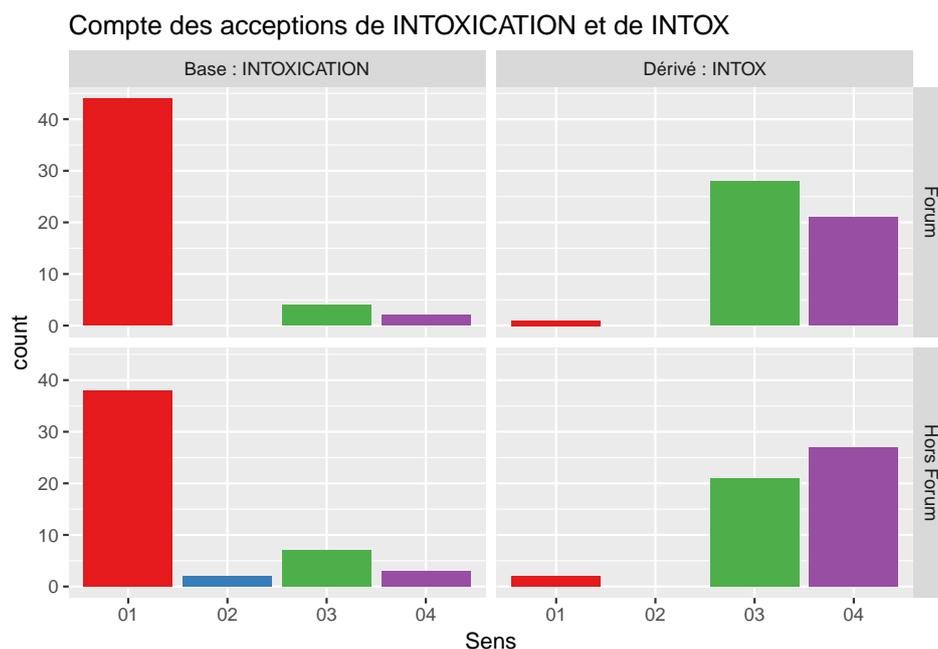


FIGURE A.5 – Distribution des sens de INTOXICATION et de INTOX

A.6 MANIFESTATION/MANIF

1. Action de manifester (opinion, qualité physique ou morale, etc.) ou de se manifester. [Ev]
 - [...] *je ne vois pas de **manifestation** de mauvaise foi ni d'arbitraire [...]*
2. Action, fait de révéler [Synonyme de révélation]. [Ev]
 - [...] *le Seigneur le fera disparaître par le souffle de sa bouche, l'anéantira par la **manifestation** de sa Venue.*
3. Signe et forme sensible d'un principe immatériel, invisible. [NoEv]
 - *J'ai tout de suite pensé qu'il était venu me voir, me dire au revoir, je ne sais pas. Cela fait peut-être 8 ou 10 ans. Je n' ai jamais oublié cette nuit, mais depuis, aucune nouvelle **manifestation** ne s' est produite [...]*
4. Phénomènes, signes ou symptômes, traduisant une maladie. [NoEv]
 - *J'aurai plutôt penché pour des **manifestations** de pbls digestifs [...]*
5. Présentation publique, rassemblement de personnes, pour des festivités. [NoEv]
 - *Jérôme s'est déjà retrouvé à distribuer des tracts de la **manifestation** vélocipédique [...]*
6. Rassemblement de personnes dans le but de faire connaître, de défendre une opinion. [NoEv]
 - [...] *les **manifestations** anti-Français se sont déroulées dans le cadre de la guerre civile [...]*

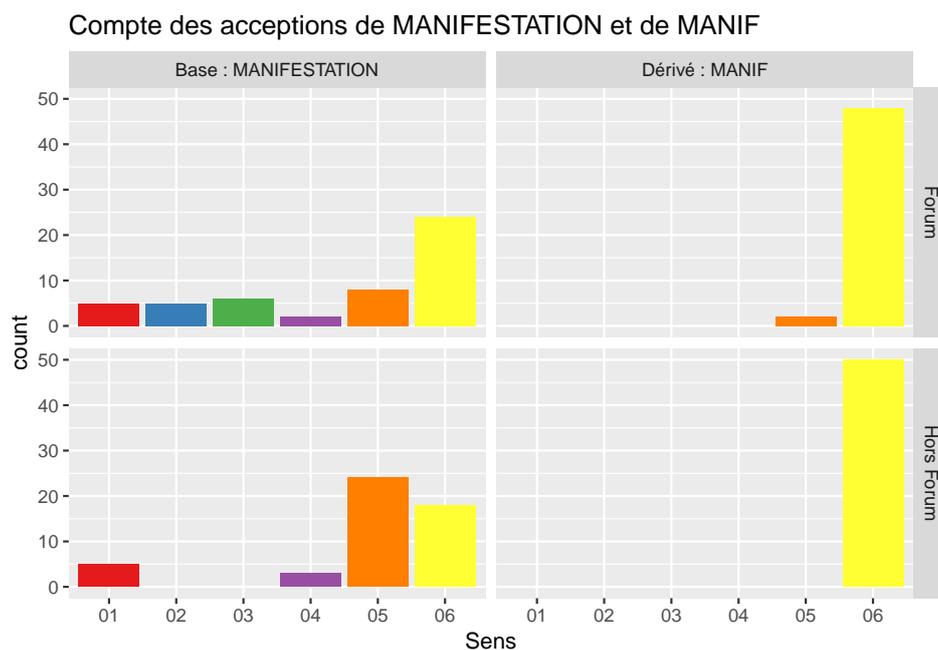


FIGURE A.6 – Distribution des sens de MANIFESTATION et de MANIF

A.7 PROMOTION/PROMO

1. Action d'élever simultanément plusieurs personnes à un même grade. [Ev]
 - [...] *i.e. les incompétences et la **promotion** de généraux ayant faillis ce genre de choses que je trouve plus intéressant à débattre [...]*
2. Ensemble des candidats entrés la même année dans certaines grandes écoles. [NoEv]
 - [...] *Jean-Pierre Liégeois qui m'invite à présenter mon travail devant une **promotion** des élèves de L' ENA [...]*
3. Nomination, accession à un grade, une dignité, une fonction supérieure. [Ev]
 - [...] *entrée dans le métier , formation carrière , salaires , **promotion** horaires , temps et conditions de travail [...]*
4. Accession d'une personne, d'un groupe social à de meilleures conditions de vie. [Ev]

- *La classe préparatoire TSI , un exemple de **promotion** sociale [...]*
5. Ensemble des moyens mis en œuvre pour promouvoir un bien, un service ou un événement. [*NoEv*]
 - *il ne s'agit pas ici de faire la **promotion** du naturisme et de ses valeurs [...]*
 6. Action de faire construire des immeubles en vue de les vendre ou de les louer. [*NoEv*]
 - *[...] acteur majeur de la **promotion** immobilière du Grand Ouest [...]*
 7. Opération commerciale afin d'attirer l'attention sur une offre ou un produit. [*NoEv*]
 - *[...] Profitez des **promotion** pour acheter d'importantes quantités que vous congèlerez. [...]*

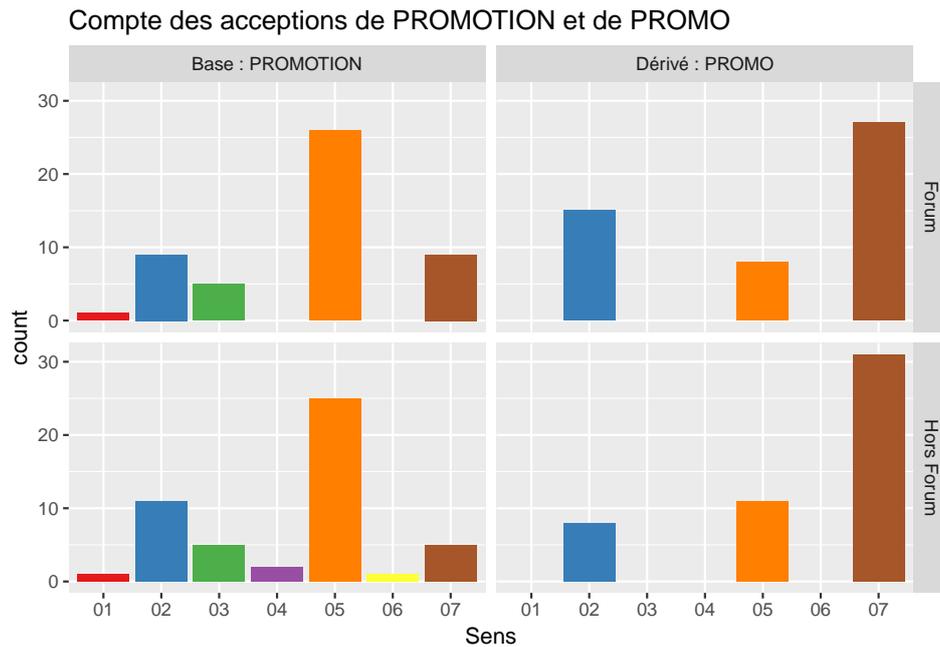


FIGURE A.7 – Distribution des sens de PROMOTION et de PROMO

A.8 RÉCRÉATION/RÉCRÉ

1. Détente, distraction, délassement qui succède à un travail, à une occupation sérieuse. [NoEv]
 - *Le petit Sarkozy, qui ne connaît rien à l'économie, a depuis son tout jeune âge, une frousse terrible de ce mot qu'il pense être le contraire de **récréation** : la récession.*
2. Temps accordé aux élèves pour se délasser, pour se détendre ; cette détente elle-même. [NoEv]
 - *Nous allions bientôt devenir la risée du collège entier (l'heure de la **récréation** de l'après-midi arrivait).*
3. Détente accordée aux religieux. [NoEv]
 - *Le concile de Paris, en 1212, condamne à son tour les échecs et interdit tout particulièrement aux gens d'Église, qui en faisaient volontiers leur **récréation**, de s'y adonner.*
4. Exercice présenté sous une forme attrayante, divertissement intellectuel. [NoEv]
 - *Mon professeur de philosophie nous avait en effet expliqué que l'histoire des androgynes avait fait couler un peu trop d'encre et n'était sûrement qu'une sorte de " **récréation** " avant l'argumentation de Socrate.*

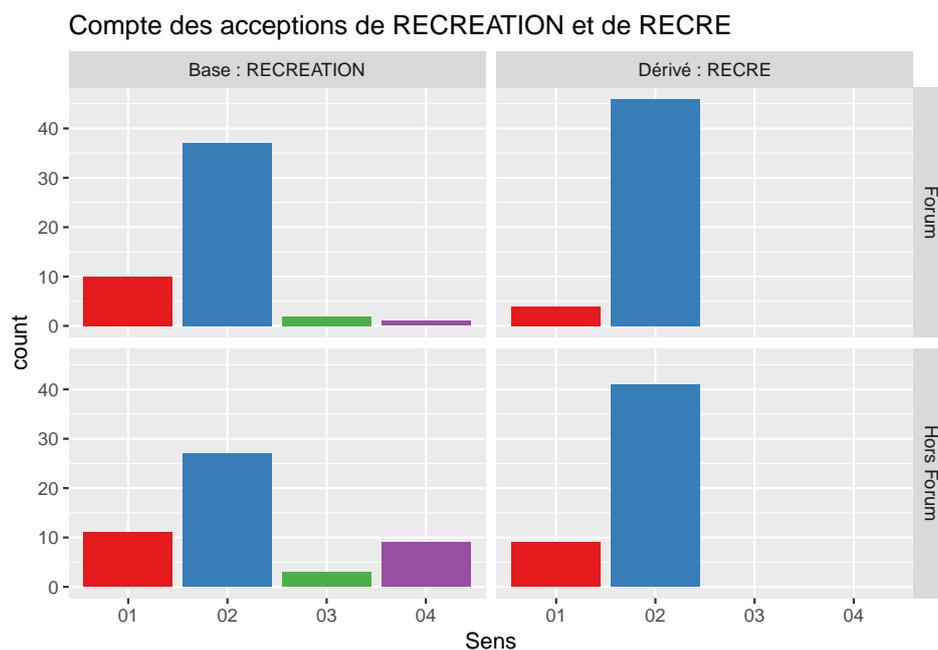


FIGURE A.8 – Distribution des sens de RÉCRÉATION et de RÉCRÉ

A.9 RÉDACTION/RÉDAC

1. Action et résultat de rédiger un ouvrage. [*Ev*]
 - [...] par l'organisation de séminaires, de rencontres, la **rédaction** d'amendements qui ont été relayés et soumis aux parlementaires.
2. Travail, service, responsabilité du ou des rédacteurs en chef. [*NoEv*]
 - [...] Alain Genestar, directeur de la **rédaction** de Paris Match
 - [...]
3. Ensemble des rédacteurs d'un journal. [*NoEv*]
 - Le contenu des blogs n'engage pas la **rédaction** de la Tribune de Genève.
4. Locaux des rédacteurs d'un journal. [*NoEv*]
 - [...] on infiltre toutes les **rédactions** de journaux des mois avant l'élection pour s'assurer de la bonne diffusion de ses messages
 - [...]

5. Exercice d'expression écrite, à partir d'un sujet par des élèves. [*NoEv*]
 - [...] *Hors Sujet a la redaction du Brevet [...]*

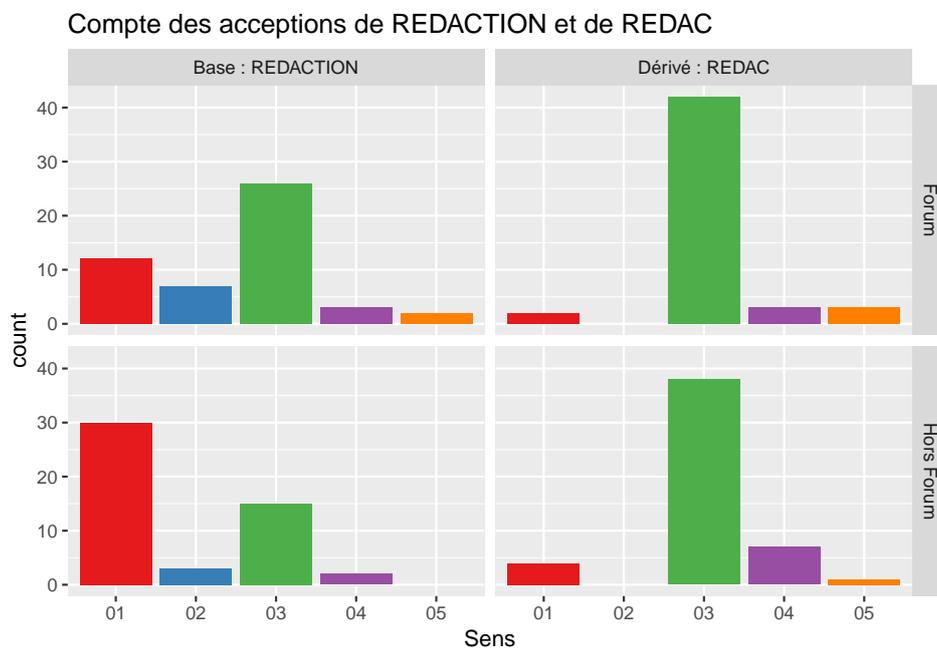


FIGURE A.9 – Distribution des sens de RÉDACTION et de RÉDAC

A.10 SONORISATION/SONO

1. Action d'ajouter une bande son à un support visuel. [*Ev*]
 - *MOYENS TECHNIQUES* . *Matériel de prise de vues , machine-rie ... 10 000,00 Matériel de prise de son ... 3 000,00 Eclairage ... 4 000,00 Montage et sonorisation , post-production ...*
2. Fait de voiser une consonne. [*Ev*]
 - *Non attesté.*
3. Ensemble des moyens permettant d'assurer la diffusion du son. [*NoEv*]
 - [...] *L'album bénéficia des meilleures techniques de sonorisation en plus d'un excellent ingénieur du son [...]*
4. Matériel de sonorisation et diffusion . [*NoEv*]

— [...] vous pouvez vous en inspirer, mais n'y voyez pas la solution miracle pour avoir la **sonorisation** du stade de France [...]

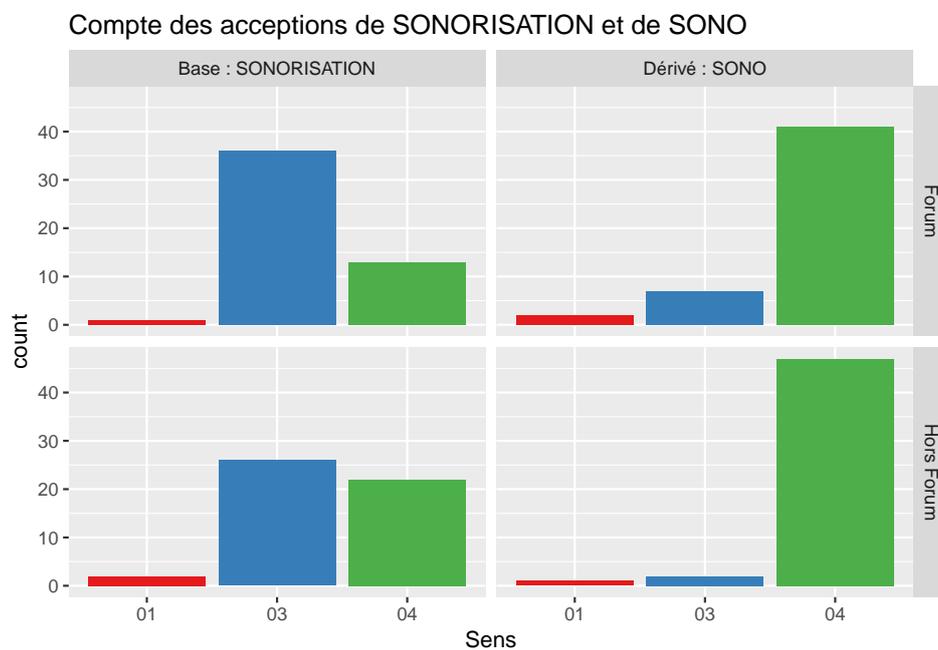


FIGURE A.10 – Distribution des sens de SONORISATION et de SONO

ANNEXE B

LISTES DE SENS POUR LES DIMINUTIFS

B.1 GLANDER/GLANDOUILLER

1. Ramasser, manger des glands.
— *Non attesté.*
2. Perdre son temps à des simulacres d'occupation.
— *[...] ils étaient, sur leur mob, a **glander** dans epinay/orge , ou il ne se passe jamais rien [...]*
3. Se promener sans but précis.
— *[...] c'est fou ce que l'on peut apprendre de choses sur sa boîte en **glandant** dans les bureaux de l'agence [...]*
4. Faire.
— *Ma question initiale était non pas de savoir ce qu' ils **glandent** durant leur période indemnitaire [...]*
5. Attendre en vain.
— *[...] le dermato l'a fait **glander** pendant 4 ans avec des pseudo creme (qui marche jamais sois dit en passant) [...]*

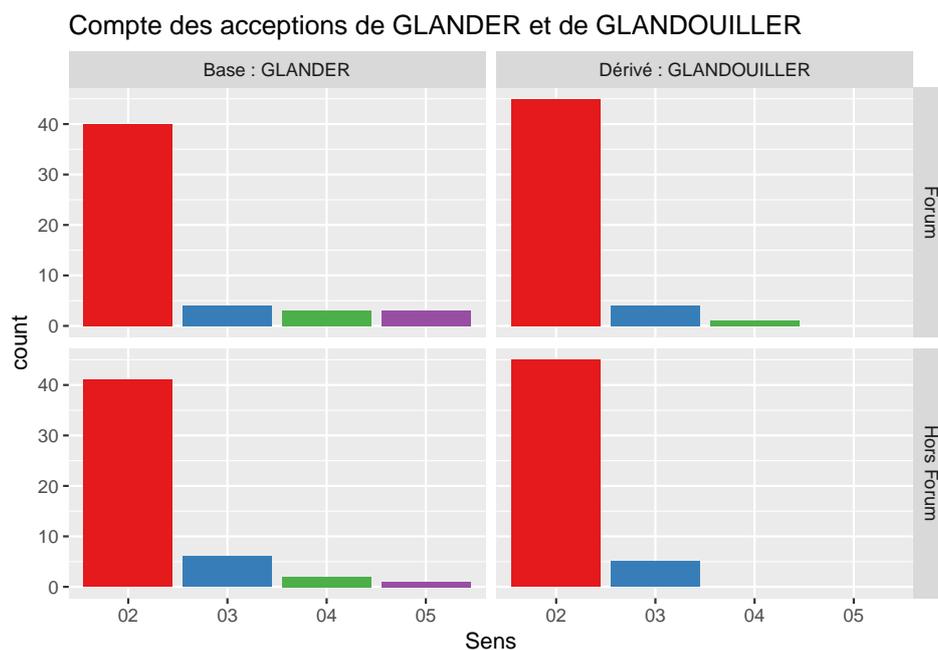


FIGURE B.1 – Distribution des sens de GLANDER et de GLANDOUILLER

B.2 GRATTER/GRAT(T)OILLER

1. Frotter en entamant légèrement pour nettoyer, polir, etc.
 - *Il va falloir que j'envisage une arthroscopie pour me **gratter** le cartilage mort.*
2. Racler dans un mouvement de va-et-vient
 - *Un livre rempli de pastilles odorantes qu'il suffit de **gratter** pour sentir l'odeur des objets de la maison.*
3. Frotter le corps ou une partie du corps qui démange ou qui irrite.
 - *Ne **grattez** pas et ne faites pas saigner une verrue.*
4. Remuer, farfouiller, fouiller.
 - *Et cela pourrait n'être pas fini ; il suffirait de **gratter** un peu les déclarations ou les non déclarations ISF de nos éminences...*
5. Parler à quelqu'un de ce qui lui plaît et à quoi il est sensible, caresser ses faiblesses.

- [...] *toi faut pas te **gratter** beaucoup pour réveiller ta vraie nature.*
- 6. Chercher à se le rendre favorable.
 - *Non attesté.*
- 7. Provoquer une sensation désagréable de démangeaison ou d'irritation.
 - *Mon cuir chevelu me **gratte** atrocement, j'ai tout le temps chaud, le front gras,...*
- 8. Recueillir péniblement et peu à peu (tout ce qui peut être utilisé, une somme d'argent).
 - *Nous raconter qu'il faut **gratter** des millions ici et là pour économiser et dépenser depuis des semaines, bientôt des mois, 1,2 million d'€ PAR JOUR au Mali, ça ne marche pas.*
- 9. Gagner (de petits profits), souvent de manière illicite.
 - *Car toutes ces mesures ont un dénominateur commun : **gratter** encore quelques avantages aux salariés.*
- 10. Dépasser, surclasser (un concurrent, un véhicule) en le gagnant de vitesse.
 - *Pour l'instant, le tracé est un poil trop long, mais il est sûrement possible de **gratter** de la longueur sur les lignes droites.*
- 11. Ne pas obtenir ce que l'on désire ou espère.
 - *Non attesté.*
- 12. Frotter discrètement, faire un léger bruit avec les ongles pour demander à entrer ou avertir d'une présence.
 - *Non attesté.*
- 13. Jouer un peu, ou plus ou moins bien, d'un instrument.
 - *Enfin, le copain du grenier, lui, **grattait** la guitare avec un bonheur certain qui se déclinait parfois en compositions originales.*
- 14. Travailler, avoir un emploi rétribué.
 - [...] *alors je retourne sur mes immeubles **gratouiller** du ciment, peinture...*
- 15. Chatouiller.

— *Moi je sais, si ce soir tu me **grattes** le sexe, tu vas gagner un bâton !*

16. Tricher à l'aide de gratte lors d'un examen.

— *J' avance sur la 3^{ème} étapes. Je suis aller a la piscine avec et j' ai **gratter** tranquille.*

17. Creuser, approfondir sa recherche, son enquête.

— *Ils parlent précieux et font des gestes méticuleux comme une baronne en représentation dans un salon de thé. <Gratte un peu et tu es surpris d'y voir du Alzheimer et du Parkinson empoussiérés.*

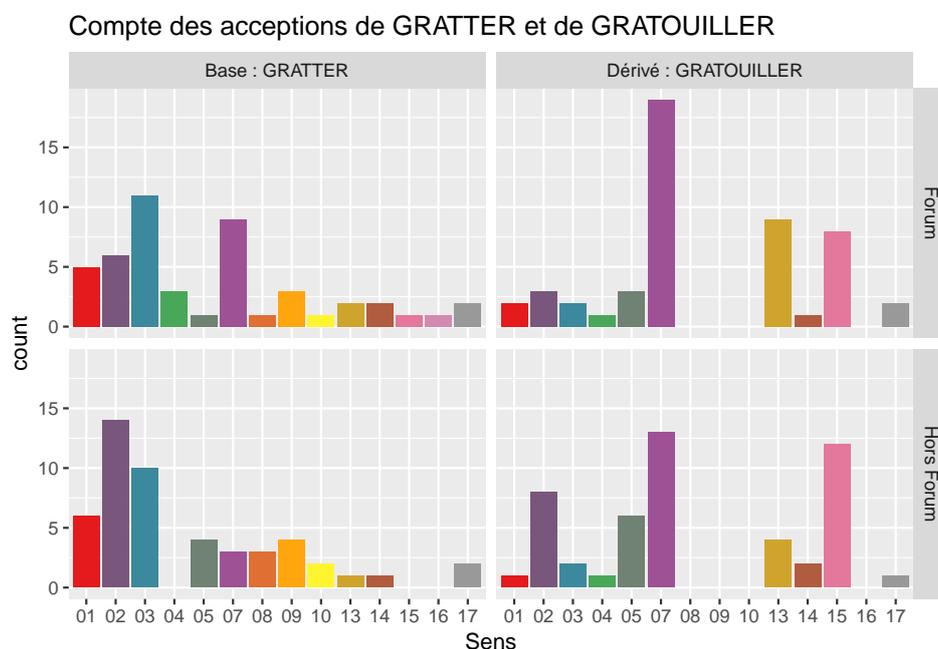


FIGURE B.2 – Distribution des sens de GRATTER et de GRATOUILLER

B.3 MÂCHER/MÂCHOUILLER

1. Action de broyer un aliment à l'aide des dents et de la mâchoire.

— *Les oiseaux ne **mâchent** pas ce qu' ils mangent [...]*

2. Réduire en mordillant avec les dents et/ou en humectant de salive.

- [...] les feuilles de coca par exemple , **mâchées** dans certaines régions pauvres [...]
- 3. Mordiller, mâchonner.
 - [...] on ne me suce plus , on me **mâche** [...]
- 4. Préparer pour faciliter la tâche à quelqu'un.
 - [...] cherche tu trouveras je ne vais pas te **mâcher** le travail [...]
- 5. Prononcer indistinctement.
 - [...] et je **mâche** mes mots [...]
- 6. S'exprimer de façon claire et sans ménagement [si forme négative].
 - [...] Jean-Louis Triaud n' a pas **mâché** ses mots au micro [...]
- 7. Penser sans cesse à quelque chose.
 - Enfermé dans ses ratiocinations **mâchouillant** de mornes pensées [...]
- 8. Couper quelque chose sans faire une section nette, déchirer.
 - [...] on **mâche** dans la mouillure de la feuille de salade [...]
- 9. User ou marquer de façon irrégulière.
 - [...] il s' agit peut-être d' un docu déjà **mâché** [...]
- 10. Meurtrir.
 - Les combos lourds pour **mâcher** et broyer ses adversaires et les combos légers [...]

B.4 MORDRE/MORDILLER

1. Serrer fortement de manière à entamer avec les dents.
 - [...] ils risquent de se **mordre** les doigts, les archives leur serviront de preuve !
2. Être jeté à terre, battu, tué, subir un échec, avoir des problèmes.
 - Dans les anciens combats classiques, on faisait "**mordre** la poussière" à son ennemi [...]
3. Serrer fortement de manière à blesser avec les dents, piquer.

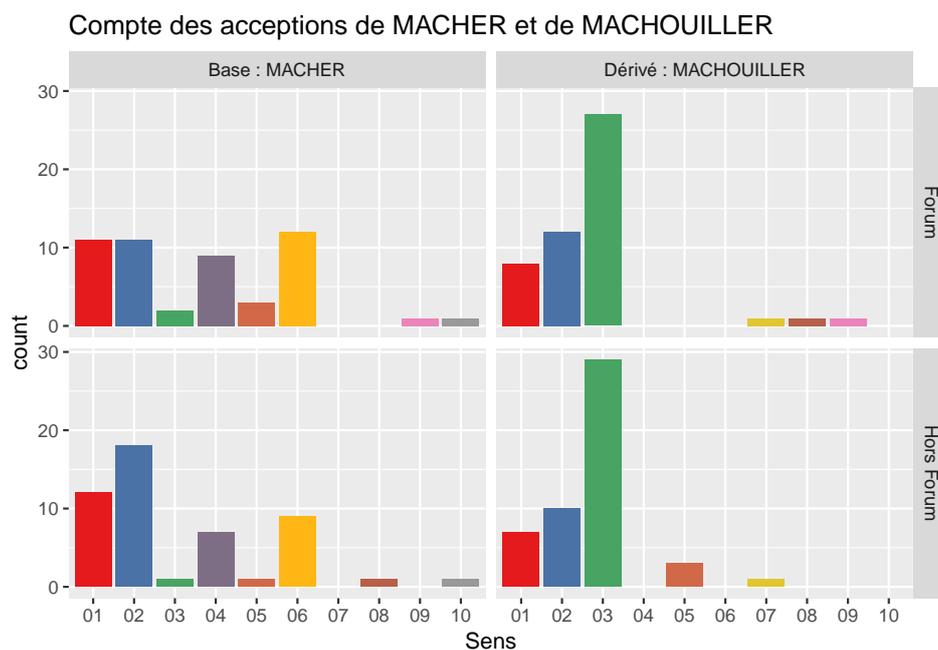


FIGURE B.3 – Distribution des sens de MÂCHER et de MÂCHOUILLER

— [...] mais j ai peur qu il se fasse **mordre** et apres qu il devienne mechant avec les autre chiens [...]

4. Serrer fortement entre ses dents au point de se couper, de se blesser.

— Le menuisier a attrapé la gueule de bois en abusant du buffet. Le charcutier a épousé un boudin, quel andouille ! Mon boucher s’est **mordu** la langue en taillant une bavette. Le boulanger s’est fait rouler dans la farine, il est dans le pétrin . L’épicier a pris une amende en passant à l’orange.

5. Mordre légèrement et à diverses reprises.

— Je concentre maintenant mes efforts sur son clitoris, que je lèche , **mords**, titille tandis que mes doigts s’engouffrent dans sa grotte que j’excite de l’intérieur.

6. S’enfoncer dans, s’accrocher à (quelque chose), trouver prise dans (quelque chose) en s’y enfonçant, en entamant.

— Mais les vis vont **mordre** le bois et ça va se stabiliser à la longue.

7. Entamer en coupant.

- [...] à **mordre** dans le fromage de la sécu et à se concentrer pour dominer encore plus le marché.
- 8. Porter atteinte, attaquer en usant, en consommant. [Le suj. désigne des agents, des phénomènes érosifs, corrosifs]
 - Les bretelles de mon sac à dos commençaient à me **mordre** cruellement la chair [...]
- 9. Ronger, faire profondément souffrir, tourmenter. (sentiments)
 - Non attesté.
- 10. Attaquer de façon désagréable, avec agressivité, ironie, tourmenter.
 - [...] le docu qui aboie sans **mordre** [...]
- 11. Regarder, voir.
 - [...] tu te doutes mon bon Rustenko que l' un **mordillera** le trophée et l' autre... ben soulèvera le plateau avec un p'tit sourire forcé.
- 12. Prendre du goût.
 - [...] on connaît vite tout le monde, mais c'est pour ça, c'est ce côté super convivial que j'ai, j'ai **mordu** dedans et puis dans le handicap ça fait partie des choses.
- 13. Se courber en avant pour un pédalage forcé.
 - Pas attesté.
- 14. Venir à l'appât et se faire prendre.
 - Un cocktail classique mais je suis heureux de constater qu'il existe des alliances et joueurs qui ne **mordent** pas à l'hameçon.
- 15. Empiéter sur quelque chose.
 - Son offensive **mord** sur l' électorat UMP.

B.5 SAUTER/SAUTILLER

1. Se détacher un bref moment de l'endroit où l'on est. Changer de position, d'un seul élan du corps.
 - Je **saute** assez haut dc ptetre qe ds quelqe année jpourré mettre un cake à PAUL !

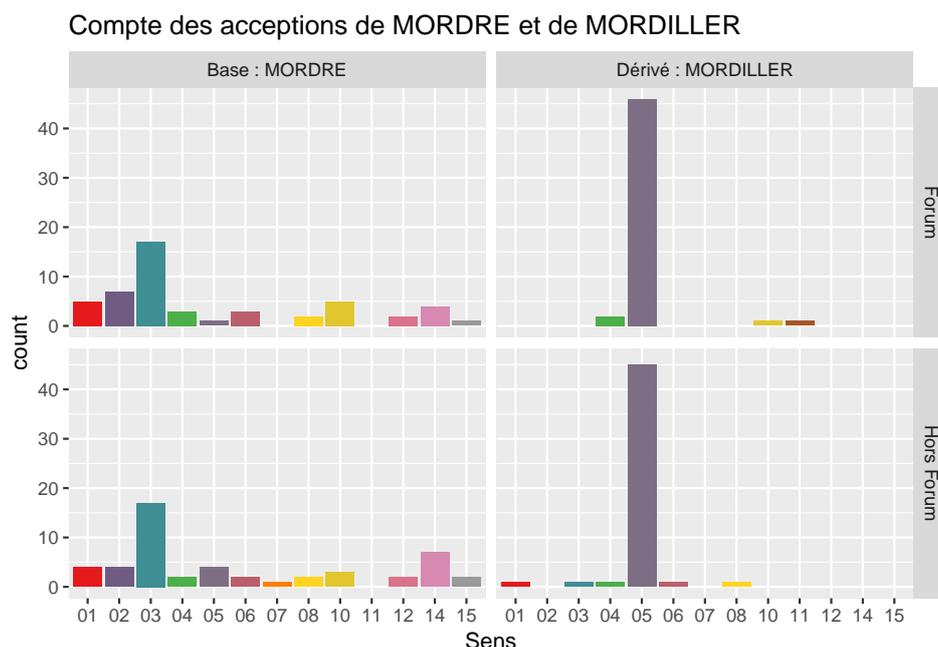


FIGURE B.4 – Distribution des sens de MORDRE et de MORDILLER

2. S'installer d'un bond, quitter d'un bond.

— *Chvez s'en occupe et le petit président de notre pays saute sur son tabouret pour récupérer une "image".*

3. Réagir vivement sous l'effet d'un choc émotif.

— *[...] elles n'ont pas compris qu' on l'ait laissée passée et ont sauté au plafond d' entendre ce qu'elle à dit.*

4. Se déplacer au sol par des bonds successifs.

— *[...] dans un ciel jaune, des lapins sautent au milieu d'étoiles en forme de carotte.*

5. Faire une série de bonds.

— *[...] 6000 pétitions de soutien recueillies, Madame BACHELOT va saute de joie !!!*

6. Se précipiter.

— *Sautez vers le point le plus loin vers la droite, ramassez les objets à proximité et dirigez vous vers la sortie nord ouest.*

7. S'élancer vers le bas, dans le vide.
 - [...] vous aurez probablement envie de **sauter** par la fenêtre après coup !
8. Exécuter des sauts par jeu ou par pratique sportive. Pratiquer un saut de l'une des disciplines de l'athlétisme ou de la gymnastique.
 - [Quand on te donne 3 essais pour **sauter** en hauteur faudrait être con pour s'imposer de ne sauter que 2 fois [...]
9. Danser.
 - Le Titre d'après est un de mes Préférés, la Musique y est d'Enfer, et les Paroles me font **sautiller** [...]
10. Passer d'un sujet à un autre. Omettre certains éléments d'une série, d'un ensemble ordonné.
 - Bon sang, de quoi parlons-nous ? Vous **sautez** du coq à l'âne.
11. Changer brusquement et fortement de direction. Varier en intensité, très brusquement. Être mobile, instable, vaciller.
 - [...] j'en déduis que j'ai un système magnétique vu comme l'aiguille **sautille**...
12. Se jeter pour agresser.
 - [...] j'ai hâte de voir le 1er client me **sauter** dessus et venir m'araguer avec ce reportage.
13. Se battre.
 - Sarkozy va tenter de mordre, de **sauter** à la gorge et de défoncer à l'arme lourde.
14. Être évident.
 - L' évidence **saute** aux yeux : ce programme économique a été conçu par des fonctionnaires qui n'ont jamais mis¹ les pieds dans une entreprise.
15. S'élancer pour s'emparer de, s'emparer ou accepter sans attendre.
 - Non attesté.
16. Franchir (un obstacle) en faisant un bond.

- *Après quelques ratés dans les départs à droite, parce que je me tortille dans ma selle et/ou que Pouné **saute** la barre au lieu de partir simplement, nous y arrivons.*
- 17. Omettre, en prononçant, en écrivant, ou en lisant, des éléments d'un énoncé ou d'un texte.
 - *Pour les non-matheux ou pour **sauter** la démonstration, il faut cliquer là [...]*
- 18. Dépasser brusquement un coureur à pied dans la foulée duquel on se tenait.
 - *Contador reste l'immense favori et je ne pense pas que Schleck le fera "**sauter**" demain, quelques soit sa stratégie.*
- 19. Avoir faim.
 - *Non attesté.*
- 20. Arrêter quelqu'un par surprise.
 - *Non attesté.*
- 21. Baiser.
 - *[...] by by chérie, va te faire **sauter** par ton anglais (si ce n' est pas déjà fait) ! aucun respect [...]*
- 22. Faire cuire (un aliment) en remuant vivement la poêle.
 - *[...] étaler une première couche de pommes de terre, saler légèrement, répartir des cèpes secs déjà **sautés** à la poêle.*
- 23. Imprimer un mouvement de bas en haut à des choses qui reposent sur une surface, qui sont rattachées à une masse.
 - *J'ai l'impression de tirer un âne mort et je **sautille** sur ma selle à chaque tour de roue.*
- 24. Être ballotté, agité, Être agité de tremblements.
 - *Non attesté.*
- 25. Faire sortir une chose du logement dans lequel elle est retenue. Projeter une chose à distance, en donnant un coup.
 - *Les murs résonnaient de leurs rires égrillards, de leurs provocations verbales, de leurs propos sans queue ni tête, des déflagrations des bouchons qui **sautaient**.*

26. Être éjecté, projeté.
— *Non attesté.*
27. Tuer en frappant ou en tirant un coup de feu à la tête.
— *Et on suppose que Moriarty a fait **sauter** la mamie [...]*
28. Provoquer un court-circuit.
— *Ce soir deux fusibles ont **sauté** dans un gouvernement en court-circuit.*
29. Éclater, être brisé, Détruire avec une charge explosive, Être détruit, exploser.
— *Après la guerre, les services de déminage ont fait **sauter** des munitions à proximité de ce blockhaus [...]*
30. La mettre dans l'impossibilité de payer, la mettre en faillite.
— *Non attesté.*
31. Perdre son emploi, sa place. Virer quelqu'un.
— *La meilleure façon de le virer, c'est de mal jouer ce soir, de perdre et domenech **sautera** /**sauter** [...]*
32. Remettre subrepticement les cartes dans l'ordre où elles étaient avant la coupe.
— *Non attesté.*
33. Battre fortement.
— *Non attesté.*
34. Parvenir d'une place inférieure à une place plus élevée, sans passer par les degrés intermédiaires.
— *Les sauts de classe sont difficiles mais parfois c'est la meilleure solution. Voyez bleuest est modeste, sa fille a **sauté** et ça se passe bien !*

Compte des acceptions de SAUTER et de SAUTILLER

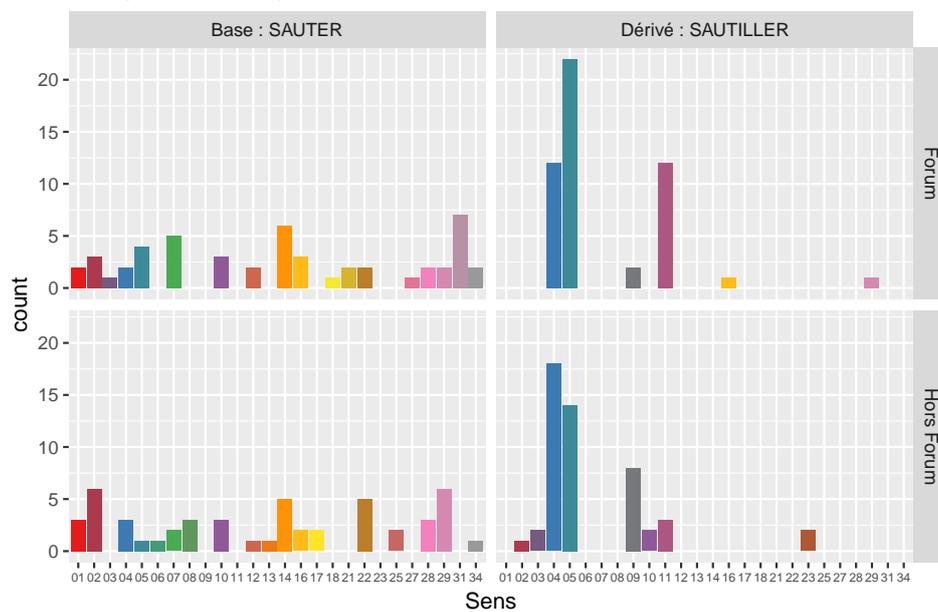


FIGURE B.5 – Distribution des sens de SAUTER et de SAUTILLER

