

POLYSÉMIE ET TRONCATION DES NOMS EN *-ION* EN FRANÇAIS

Rémi ANSELME

Laboratoire Dynamique Du Langage, UMR5596, Université Lumière Lyon 2

Olivier BONAMI

Université de Paris, LLF, CNRS

Heather BURNETT

Université de Paris, LLF, CNRS

RÉSUMÉ

Cet article présente une étude de corpus des effets sémantiques de la troncation des noms en -ion en français. Nous documentons trois effets : la distribution statistique des sens des noms tronqués est significativement différente de celle de leurs bases, elle est moins équilibrée, et les sens événementiels sont défavorisés. Ces observations suggèrent que les généralisations de Kerleroux (1999) sont valides si on leur donne une forme graduelle plutôt que catégorique.

ABSTRACT

This article presents a corpus study of the semantic effects of truncation of nouns ending in -ion in French. We document three effects : the statistical distribution of the meanings of truncated nouns is significantly different from that of their bases, they are less evenly distributed, and eventive meanings are disfavoured. These observations suggest that Kerleroux (1999)'s generalizations are accurate, but only if they are stated in a continuous rather than categorial way.

1. INTRODUCTION

La troncation est définie par le Trésor de Langue Française 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 ». Supprimer une ou plusieurs syllabes en début de mot porte le nom d'aphérèse, comme en (1a) où la forme *bleme* est utilisée pour la forme *problème*, et en fin de

mot le nom d'apocope comme en (1b) où la forme *réus* est employée pour la forme *réunions*. Lorsque la troncation a lieu au milieu du mot, on parle alors de syncope. Cet article s'intéressera uniquement à l'apocope et le terme *troncation* sera utilisé de manière générique¹.

- (1) a. «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 [...]»²
- b. « 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. »³

Nous partons de l'hypothèse que la troncation est une opération dérivationnelle qui construit un nouveau lexème, le *troncat* (Antoine, 2000), à partir d'un lexème base, sans modifier le type dénotatif : base et troncat ont la capacité à dénoter des entités de même type. Le caractère dérivationnel de la troncation se justifie par la capacité des troncats à devenir des bases pour un autre procédé constructionnel (par ex. VÉLOCIPÈDE > VÉLO > VÉLOTISTE) et par le fait que les troncats sont susceptibles d'acquérir un sens lexical distinct de celui de leur base (SONORISATION « action de sonoriser > SONO « matériel de sonorisation ») La troncation s'applique à différentes classes de lexèmes, mais dans cet article nous nous concentrons sur la troncation des noms déverbaux en *-ion*.

Comme les autres procédés dérivant des noms d'événements à partir de verbes, la dérivation en *-ion* peut produire des résultats sémantiques variés. À la suite de Grimshaw (1990), on peut distinguer les noms d'événements à proprement parler, qui se subdivisent en noms d'événements simples (comme DÉCISION dans *Paul a pris une décision*), les noms d'événements complexes (comme DESTRUCTION dans *la destruction de la ville par les bombes*), et les noms résultatifs, qui ne dénotent pas des événements (comme DISSERTATION dans *les dissertations sont sur la table*). Crucialement, un même nom est souvent ambigu entre plusieurs de ces types d'interprétations. Cette multiplicité des interprétations et leurs relations contribuent à la polysémie de ces noms déverbaux et rend leur dérivation intéressante à analyser du point de vue sémantique.

¹ Les recherches rapportées ici ont été effectuées dans le cadre du mémoire de master 2 du premier auteur. Nous remercions Lucie Barque, les participants à la journée d'études *Questions de morpho-sémantique* (Toulouse, 8 juillet 2019), et deux relecteurs anonymes pour leurs commentaires et suggestions. Cette recherche a reçu le soutien partiel du projet Demonext (ANR-17-CE23-0005) ainsi que d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du programme « Investissements d'Avenir » portant la référence ANR-10-LABX0083.

² « Blème » : [consulté le 24 mars 2019]. http://forum.ados.fr/love/Amour/bleme-sujet_24513_1.htm

³ « Réu » : [consulté le 24 mars 2019]. <https://www.forum-na.fr/viewtopic.php?pid=7841>

Cet article étudie les effets de la troncation sur l'interprétation des noms en *-ion*. La section 2 résume la catégorisation des noms déverbaux existante (Roy & Soare, 2013). On s'intéressera particulièrement à la généralisation de Kerleroux (1999 : 87) selon laquelle les noms avec une interprétation d'événement complexe sont non-troncables. Cette partie servira à mettre en avant qu'il existe des contraintes sur la troncation qui sont d'ordre syntaxique mais aussi d'ordre sémantique, comme la réduction sémantique opérant lors de la dérivation réduisant les lexèmes (Raccah, 2016). La section 3 présentera la stratégie que nous avons adoptée pour quantifier la répartition des sens des noms déverbaux de manière manuelle à partir d'un corpus. Nous montrerons que les généralisations de Kerleroux et Raccah ne peuvent être maintenues sous une forme catégorique, mais qu'elles sont valides si on leur donne une forme graduelle. La section 4 discute les conséquences de nos résultats pour l'étude de la morphologie constructionnelle en général, suggérant que les distributions statistiques de sens doivent être placées au centre de l'attention.

2. INTERPRÉTATIONS DES NOMS EN *-ION* ET TRONCATION

La troncation est une opération dérivationnelle qui peut s'appliquer aux noms déverbaux en *-ion*. Cependant, tous les noms en *-ion* ne sont pas aussi facilement troncables, et en particulier la troncabilité est sensible au type de lecture du nom, certains sens des noms polysémiques rechignant à la troncation (Kerleroux, 1999). À la suite de Grimshaw (1990) et de Roy & Soare (2013), on admettra l'existence de trois types de sens pour les nominalisations déverbales⁴. Les *noms d'événements complexes* (« Argument-Structure nominals » chez Grimshaw) comme en (2a), dénotent des événements (ou plus généralement des éventualités), héritent de la structure argumentale du verbe, et sont normalement employés avec l'article défini singulier. Les *noms d'événement simples*, illustrés en (2b), dénotent également des événements, mais ont les propriétés syntaxiques des noms ordinaires ; en particulier ils ne prennent typiquement pas d'argument et se combinent librement avec tout déterminant. Enfin, les *noms résultatifs*, illustrés en (2c), désignent un objet (concret ou abstrait) qui est produit par l'action correspondant au sens événementiel du nom. Ils ont eux aussi typiquement une syntaxe nominale ordinaire.

- (2) a. La réunion de toutes les signatures en une semaine est un exploit.
- b. Chaque réunion dure au moins deux heures.
- c. Cette construction majestueuse mesure plus de 30 mètres.

⁴ Il existe dans la littérature un autre point de vue développé à propos des différentes classes de noms, selon lequel 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 catégorique entre noms d'événements simples et complexes proposée par Grimshaw. Nous renvoyons le lecteur à l'article de Grimm & McNally (2013) pour une argumentation détaillée basée sur une étude de corpus.

C'est sur cette classification que Kerleroux (1999) se base pour soutenir que la troncation peut s'appliquer aux noms d'événements simples et aux noms résultatifs. Cette généralisation expliquerait les contrastes observés dans ses exemples repris ici en (3) et (4).

- (3) a. L'introduction du Lynx dans le Vercors provoque des tollés.
 b. *L'intro du Lynx dans le Vercors provoque des tollés.
 c. L'intro de ta dissertation nous a bien fait rire.
 d. L'introduction de ta dissertation nous a bien fait rire.
- (4) a. La manifestation de la vérité aura pris cinquante ans.
 b. *La manif de la vérité aura pris cinquante ans.
 c. La manif des étudiants aura duré longtemps.
 d. La manifestation des étudiants aura duré longtemps.

En (3) et (4), nous retrouvons des noms déverbaux en *-ion* qui en fonction de leur interprétation peuvent ne pas être troncables. INTRODUCTION lorsqu'il réfère à l'événement complexe d'introduire ne peut pas être tronqué (3.b) alors que lorsqu'il fait référence à un nom résultatif, le troncat INTRO est possible. Ces observations sont similaires avec MANIFESTATION, qui lui peut être tronqué lorsqu'il a une interprétation de nom d'événement simple. L'interprétation des différents noms déverbaux en *-ion* peut aussi être mise en relation avec leurs différents sens, autrement dit, chaque sens spécifique d'un nom peut correspondre à une interprétation. Tous les sens d'un nom associé ne sont pas troncables. Une conséquence logique est que les troncats sont donc attendus comme moins polysémiques, voire monosémiques.

Kerleroux (1999 : 93) parle « d'opération de singularisation des dénominations » pour faire référence à ce que Raccah (2016) considère comme une restriction, une modification du domaine de référence en parlant de cette perte de la polysémie à travers la troncation. Pour mettre en avant cette restriction, Raccah analyse des occurrences de formes tronquées en les comparant avec les formes non-tronquées. En guise d'exemple, nous retrouvons le contraste qui existe entre INTOXICATION et INTOX lorsque la forme est précédée du déterminant indéfini *une* et suivie de l'adjectif *alimentaire*. Le contraste d'interprétation en (5) est justifié par le fait que le domaine de référence de INTOX est limité au domaine intellectuel.

- (5) a. une intoxication alimentaire
 b. une *intox alimentaire

Ainsi, il n'est pas possible de parler d'une intoxication alimentaire avec sa forme tronquée. De même, pour COMPILATION, le domaine après troncation n'inclut pas tous les domaines de la base. Il n'est donc pas habituel de parler de **compil d'article* pour parler d'une compilation d'articles, le domaine de COMPIL étant restreint au domaine musical. Bien que la restriction soit une implication de la troncation, il n'est pas pour autant toujours évident de caractériser la restriction opérée par la troncation pour certaines formes.

Raccah reconnaît l'utilité d'une étude de corpus pour vérifier ses hypothèses. La subjectivité du locuteur entre 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 communs à l'ensemble des locuteurs d'une langue, mais bien personnels, et peuvent néanmoins être partagés par un groupe d'individus. Une des premières conséquences de parler de subjectivité du locuteur se retrouve à travers de simples recherches sur internet, à travers les forums ou sur Twitter par exemple. Les exemples de troncats (6) à (9), tous collectés sur Twitter fin 2018, ne devraient pas exister selon les travaux précités. Dans les faits, ils sont bel et bien attestés. Dans certains cas, le domaine de restriction des troncats n'est pas celui qui est prédit, dans d'autres cas les troncats ont une interprétation d'événement complexe.

- (6) « Bon c'est bien mignon les soirées etc mais quand tu finis dans les toilettes à 3h de matin en hyper ventilation et une *intox* alimentaire alors que tu boss le lendemain c'est pas si drôle »
- (7) « Vous m'excusez, mais je préfère avoir les liens des études directement plutôt qu'une *compil* d'articles de presse. »
- (8) « Je me lance lance dans la *redac* de l'article, gros blocage sur l'intro #lablague »
- (9) « Pour la *manif* de la vérité je ne soutiens personne je laisse la justice agir. C'est ce que vous devriez faire en tant qu'elu »

Ces observations posent un problème pour les généralisations proposées par Kerleroux (1999) et Raccah (2016), qui ne semblent donc plus tenir en tant que contraintes catégoriques. Néanmoins, si les contre-exemples existent, ils ne sont pas abondants. Il est donc possible qu'une version affaiblie, non-catégorique, des contraintes suggérées soit valide : plutôt que de les interdire, la troncation défavorise certaines interprétations. C'est cette possibilité que nous évaluons dans cet article.

3. UNE ÉTUDE QUANTITATIVE DE LA POLYSÉMIE DES TRONCATS

Nous nous proposons de procéder à une analyse quantitative des sens des lexèmes permettant de valider ou d'infirmer les tendances suggérées par la littérature. Il est habituel, en parlant des sens des lexèmes, de ne faire qu'une analyse qualitative en identifiant les sens qu'un lexème peut avoir, et en argumentant quant à la manière dont les processus morphologiques opèrent avec ces sens. Comme la prévalence des sens n'est pas quantifiée, on laisse entendre que les locuteurs disposent au même degré de tous ces sens lorsqu'ils utilisent un lexème, ce qui est peu vraisemblable. Notre méthodo-

logie pour éviter ce problème ne fait pas de présupposition sur les sens que doivent posséder les troncats.

Nous avons admis que les dictionnaires nous renseignaient de manière adéquate sur la liste des sens des lexèmes étudiés, et avons décompté manuellement en corpus les emplois des troncats et de leur base qui relèvent de tel ou tel sens. Ces comptes ont ensuite été utilisés pour calculer différents indicateurs quantitatifs opérationnalisant nos questions de recherche.

3.1. Questions de recherche

Notre travail vise à répondre à plusieurs questions de recherche. 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 à établir un possible rôle des caractéristiques sociolinguistiques du contexte d'énonciation. Pour cela, comme nous avons vu que la troncation était accompagnée d'instructions et de contraintes, la première question à laquelle nous souhaitions répondre était de savoir s'il existait un lien de dépendance entre la distribution des sens d'un lexème, c'est-à-dire le nombre de sens disponibles pour un lexème et les proportions d'utilisation de chaque sens, et le fait que ce lexème soit un troncat ou non. Autrement dit, il s'agit d'avoir la possibilité de mettre en évidence des contraintes sémantiques. Dans le même cadre, nous avons réfléchi 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. La troncation étant généralement perçue comme restreinte à un contexte informel, on peut s'attendre à l'existence d'une interaction entre sens et niveau de formalité du discours : certains sens sont peut-être uniquement présents dans des contextes informels et d'autres uniquement dans des contextes formels.

Dans un deuxième temps, nous avons voulu savoir s'il est possible de mettre en évidence une différence entre bases et troncats en termes de dispersion des sens véhiculés. Les observations de Kerleroux et Raccah suggèrent que la troncation forme des lexèmes sémantiquement plus spécialisés que leurs bases. Une manifestation possible de cette propriété est que certains sens soient simplement exclus pour le troncat, comme le suggère la généralisation de Kerleroux. Une version moins radicale de cette généralisation supposerait simplement que la distribution statistique des sens soit plus contrastée : là où la répartition statistique des emplois est équilibrée pour les bases, elle serait plus déséquilibrée pour les troncats, certains sens étant fortement défavorisés. On verra dans la section 3.4 comment cette intuition peut être opérationnalisée en utilisant la notion d'entropie.

Finalement, une dernière question a émergé. Il s'agissait de savoir s'il était possible de mettre en évidence que certains types de 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 et autres sens est important

pour comprendre un des mécanismes sous-jacents de la troncation. Kerleroux (1999) a proposé que les événements complexes ne permettaient pas l'apocope à cause de plusieurs propriétés qui leur étaient intrinsèques. Dire que les sens événementiels sont désavantagés par la troncation permet d'appuyer l'argument qu'il existe une répercussion de contraintes dérivationnelles au niveau sémantique.

3.2. La méthode

La méthodologie que nous avons élaborée a pour but de permettre de répondre ces questions grâce à une analyse de corpus. Pour cela, nous avons choisi de travailler avec le corpus web FRCOW (Schäfer, 2015 ; Schäfer & Bildhauer, 2012). Nous avons utilisé la version du 10 décembre 2017 qui comporte 8 789 873 699 occurrences de mots correspondant à 27 642 351 mots-formes et 403 981 lexèmes. Le corpus est distribué avec un étiquetage morphosyntaxique et une lemmatisation établis automatiquement et dont la qualité n'a pas été évaluée. Enfin, chaque énoncé est associé à des métadonnées, indiquant notamment le pays d'origine de la page web, et le fait que la page soit ou non un forum de discussion. Il s'agit aussi d'une catégorisation qui a été faite automatiquement.

Nous nous sommes restreints 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. Kanwal (2018) avait travaillé sur une liste de 30 troncats de ces noms déverbaux. Ayant de plus accès aux troncats du dictionnaire d'Antoine (2000) et à la liste de Raccah (2015), nous avons élaboré une liste de 10 paires (base, troncat) présents dans nos trois sources et que nous présentons dans le Tableau 1. La petite taille de la liste se justifie par l'annotation manuelle chronophage qui sera détaillée par la suite. 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 tronqués.

À partir des sens récupérés, nous avons fait un premier travail d'élaboration de listes de sens en regroupant les sens trop proches sémantiquement pour pouvoir facilement être distinguées en corpus. Dans un deuxième temps, nous avons fait le même travail avec le Wiktionnaire francophone dans son état du printemps 2019.

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

Tableau 1. – Liste des paires de bases-troncats utilisées

Les définitions dans le TLFi sont organisées de manière hiérarchique, les sens jugés proches étant regroupés, et l'ordre reflétant généralement une antériorité historique. Le Wiktionnaire adopte à l'inverse une structure plate, les sens les plus utilisés étant présentés en premier. De plus, les définitions données par le Wiktionnaire rendent compte de plus de sens ayant émergé récemment, là où « la rédaction du TLF [Trésor de la Langue Française] est terminée depuis 1994 » (<http://atilf.atilf.fr/>). Ceci explique certainement au moins en partie que les définitions pour les troncats soient plus riches dans le Wiktionnaire que dans le TLFi. De plus, nous avons de manière anecdotique ajouté certaines définitions provenant de pages Wikipedia détaillées pour les troncats, nous permettant d'enrichir nos listes de sens.

Pour illustrer cette étape, nous reproduisons ci-dessous la liste de sens retenue pour la paire de lexèmes (PROMOTION, PROMO), avec des exemples issus de FRCOW⁵ :

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

⁵ Nous avons pris le parti de ne pas corriger l'orthographe des exemples attestés dans FRCOW.

- d. Accession d'une personne, d'un groupe social à de meilleures conditions de vie
*La classe préparatoire TSI , un exemple de **promotion** sociale [...]*
- e. Ensemble des moyens mis en œuvre pour promouvoir un bien, un service ou un événement
*Il ne s'agit pas ici de faire la **promotion** du naturisme et de ses valeurs [...]*
- f. Action de faire construire des immeubles en vue de les vendre ou de les louer
*[...] acteur majeur de la **promotion** immobilière du Grand Ouest [...]*
- g. Opération commerciale afin d'attirer l'attention sur une offre ou un produit
*Profitez des **promotions** pour acheter d'importantes quantités que vous congèlerez. [...]*

Nous avons divisé le corpus en deux parties correspondant aux exemples issus ou non de forums de discussions, puis effectué dans nos deux sous-corpus des tirages aléatoires d'occurrences en contexte des bases et des troncats. Nous souhaitions travailler avec un nombre assez important d'occurrences, mais comme l'annotation était manuelle nous nous sommes limités à 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 occurrences du sous-corpus pour la condition formelle (hors forum)⁶.

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 réalisée par le premier auteur. Toutes les analyses ont été faites avec R (R Core Team, 2017) en utilisant l'environnement Rstudio (Rstudio Team, 2015).

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 nous demande-

⁶ 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.

rons si un type de sens, le sens événementiel, est défavorisé ou bien donné comme absent par la troncation.

3.3. Les distributions de sens sont-elles différentes pour les bases et les troncats ?

Nous appelons *distribution des sens* la manière dont les sens se répartissent pour un même mot. Autrement dit, ce qui nous intéresse avec les distributions de sens c'est de savoir quels sont les sens sur-représentés pour un lexème, et ceux qui sont sous-représentés. La Figure 1 résume sous la forme de quatre diagrammes à bâtons la distribution des sens pour la paire (PROMOTION, PROMO). L'axe horizontal distingue la situation des deux lexèmes, alors que l'axe vertical oppose les sous-corpus de forums et de non-forums.

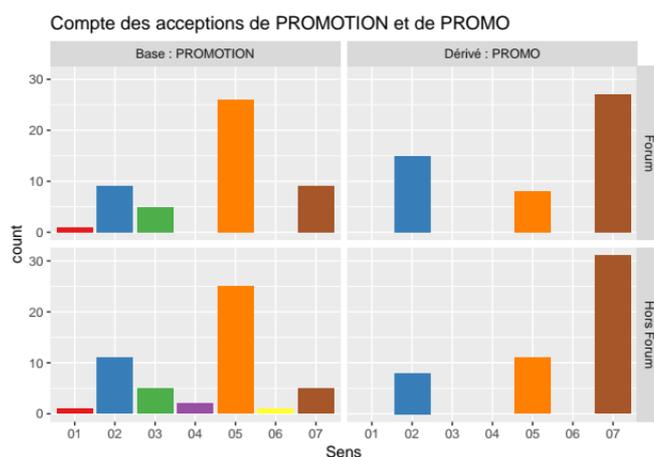


Figure 1. – Distribution des sens de PROMOTION / PROMO

Chacun des quatre diagrammes présente l'effectif dans notre échantillon des occurrences relevant de chacun des sept sens possibles pour la paire (PROMOTION, PROMO). Dans ce cas particulier, nous observons que les sens 1, 3, 4 et 6 ne sont pas représentés dans notre échantillon pour le lexème 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 un de ces sens. On ne rencontre donc PROMO qu'avec les sens 2, 5 et 7, le sens 7 étant le plus représenté autant en condition forum qu'en condition non-forum. En ce qui concerne PROMOTION,

tous les sens sont représentés dans la condition non-forum, mais dans la condition forum les sens 4 et 6 ne sont pas représentés⁷.

L'observation visuelle de la figure suggère qu'il y a peu ou pas de différences entre les distributions des sens dans les forums vs. non-forums, alors qu'il existe une différence importante entre ces distributions pour la base et le troncat. Pour confirmer cette impression et la généraliser à l'ensemble des paires examinées, nous avons réalisé une série de tests du χ^2 de Pearson. Nous avons simulé les p -valeurs grâce à une simulation de Monte-Carlo avec 1000 répliques pour diminuer le risque d'erreur. Nous présentons dans un premier temps nos résultats pour la paire (PROMOTION, PROMO).

En premier lieu, nous nous sommes intéressés au compte des sens d'occurrence pour PROMOTION en forum et en non-forum. Les résultats ($\chi^2(6)=4,3625$, $p=0,7223$) ne montrent pas une différence significative et ne permettent donc pas de rejeter l'hypothèse nulle d'indépendance. Nous ne sommes donc pas en mesure de conclure à un effet sur la distribution des sens pour le lexème PROMOTION de son emploi dans un forum ou dans un non-forum. Nous aboutissons à la même conclusion en ce qui concerne le troncat PROMO ($\chi^2(2)=2,88$, $p=0,2517$).

En second lieu, nous nous sommes demandé si le type de lexème (tronqué ou non) avait un effet sur la distribution de sens. En comparant les échantillons dans les non-forums, nous constatons une différence significative ($\chi^2(6)=33,696$, $p<0,001$) nous permettant de rejeter l'hypothèse d'indépendance. Autrement dit, la distribution des sens est dépendante du fait que l'on examine des occurrences de la base PROMOTION ou du troncat PROMO. Nous aboutissons à la même conclusion en comparant les échantillons d'occurrences trouvées dans les forums ($\chi^2(2)=26,029$, $p<0,001$).

Nous avons répliqué nos analyses statistiques à l'ensemble de nos paires de lexèmes. Nous présentons ces résultats de test de χ^2 de Pearson dans le Tableau 2. Nous avons grisé 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 un effet de la forme d'un lexème sur sa distribution de sens. Pour 8 paires (base, troncat) sur 10, leur distribution de sens est complètement indépendante du genre textuel (forum ou non). Pour les deux dernières paires, seule la forme présente ou absente dans un forum a un effet sur la distribution des sens.

⁷ Dans le cas de PROMOTION et PROMO, tous les sens ont été trouvés au moins une fois dans une condition, mais ce ne fut pas le cas pour certains autres lexèmes dont nous avons annoté les sens. La taille de nos échantillons étant relativement petite, on ne peut pas en conclure que ces sens sont simplement absents du corpus, seulement qu'ils ne sont pas fréquemment rencontrés.

	Effet de Forum / Hors Forum						Effet de Troncat / Base					
	Troncat			Base			Forum			Hors Forum		
	χ^2	DL	<i>p</i> -val	χ^2	DL	<i>p</i> -val	χ^2	DL	<i>p</i> -val	χ^2	DL	<i>p</i> -val
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

Tableau 2. – Résultats des tests d'indépendance du χ^2

L'ensemble des résultats de ces tests statistiques suggère que les sens des lexèmes étudiés ont la même distribution dans les forums et dans le reste du web. En tout cas, si différence il y a, elle est trop subtile pour être observable avec un jeu de données de cette taille. Les données ne soutiennent donc pas l'hypothèse d'un sens social associé à la troncature, qui se manifesterait par une distribution différente des formes tronquées selon les contextes sociolinguistique. Cette situation 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. Il est aussi possible que la taille des données soit insuffisante, ce qui rend difficile la détection d'un effet subtil. Malheureusement, les coûts de l'annotation manuelle rendent inenvisageables de changer d'ordre de grandeur. Finalement, une dernière explication possible est de dire tout simplement que la distinction entre forum et hors forum n'est pas suffisamment corrélée avec la distinction formel / informel. Ainsi, il faudrait pouvoir trouver d'autres moyens de mesurer le degré de formalité en corpus.

En revanche, les tests statistiques apportent un soutien clair à notre première hypothèse, selon laquelle 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 lexème dépend de son caractère tronqué ou non.

3.4. Les sens sont-ils plus variés pour les bases que pour les troncats ?

Nous avons établi l'existence d'une différence entre les distributions de sens pour les bases et les troncats, mais nous n'avons encore rien dit sur la nature de cette différence. Nous allons maintenant tenter de caractériser cette différence, et plus spécifiquement de tester l'hypothèse selon laquelle la distribution des sens est moins équilibrée pour les bases que pour les troncats. On s'attend à ce que ce soit le cas pour deux raisons : d'une part, si Kerleroux a raison de penser que certains sens sont exclus pour les formes tronquées, le nombre d'occurrence devrait être nul pour ces sens pour un troncat, ou au moins plus faible qu'il ne l'est pour la base. D'autre part, si Raccach a raison de soutenir qu'il y a une spécialisation du domaine de référence pour les troncats, cette spécialisation devrait se manifester dans la distribution des sens, au moins pour autant que la répartition des sens capte la distinction pertinente. Par exemple, INTOXICATION une fois tronqué donne INTOX, et le troncat a un domaine de référence restreint : on parle d'*intox* dans le domaine journalistique mais pas d'une *intox alimentaire*.

Nous utilisons l'entropie de Shannon (Shannon, 1948) appliquée aux distributions de sens pour mesurer le caractère plus ou moins équilibré de ces distributions. Etant donné une variable aléatoire X dont les valeurs sont x_1, \dots, x_n l'entropie de X est donnée par :

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$$

L'entropie est une grandeur positive qui mesure l'incertitude associée à une distribution de probabilités. S'il n'y a pas d'incertitude, l'entropie est de zéro ; une entropie de 1 correspond à l'incertitude liée au choix entre deux événements équiprobables, tels que le fait que le lancer d'une pièce équilibrée donne « pile » ou « face ». S'il y a un grand nombre de possibilités et que leurs probabilités sont équilibrées, l'entropie peut être élevée, jusqu'à l'infini dans le cas limite.

Dans notre cas, les valeurs de la variable aléatoire sont les différents sens qu'une occurrence d'un lexème peut recevoir, et nous évaluons empiriquement la probabilité de chaque sens à partir de sa fréquence dans nos échantillons. Imaginons par exemple qu'un de nos lexèmes X ait quatre sens, avec une distribution totalement équilibrée : dans un échantillon de 12 occurrences, chaque sens est rencontré exactement 3 fois. L'entropie de X prend alors la valeur $H(X) = 2$, qui est le maximum possible pour une variable aléatoire à 4 valeurs.

$$\begin{aligned} H(X) &= - \left(\left(\frac{1}{4} \times \log_2 \frac{1}{4} \right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4} \right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4} \right) + \left(\frac{1}{4} \times \log_2 \frac{1}{4} \right) \right) \\ &= 2 \end{aligned}$$

Supposons maintenant que le lexème Y a lui aussi 4 sens, mais que l'un des sens est rencontré 9 fois dans un échantillon de 12 occurrences, les autres sens étant chacun représenté une seule fois. L'entropie est alors plus basse :

$$H(Y) = -\left(\left(\frac{9}{12} \times \log_2 \frac{9}{12}\right) + \left(\frac{1}{12} \times \log_2 \frac{1}{12}\right) + \left(\frac{1}{12} \times \log_2 \frac{1}{12}\right) + \left(\frac{1}{12} \times \log_2 \frac{1}{12}\right)\right) \\ \approx 1,42$$

Cette valeur d'entropie basse reflète une situation avec une incertitude qui a diminué par rapport au cas précédent. Finalement, notre dernier exemple correspond à un lexème Z supposément polysémique, toujours à 4 sens, mais dont un seul sens est attesté dans notre échantillon. Nous obtenons l'entropie suivante :

$$H(Z) = -(1 \times \log_2 1) \\ = 0$$

Une valeur d'entropie nulle est donc associée à une distribution sans incertitude. Nous résumons les trois cas graphiquement dans la Figure 2, avec trois distributions statistiques possibles résultant de l'annotation des sens d'un lexème.

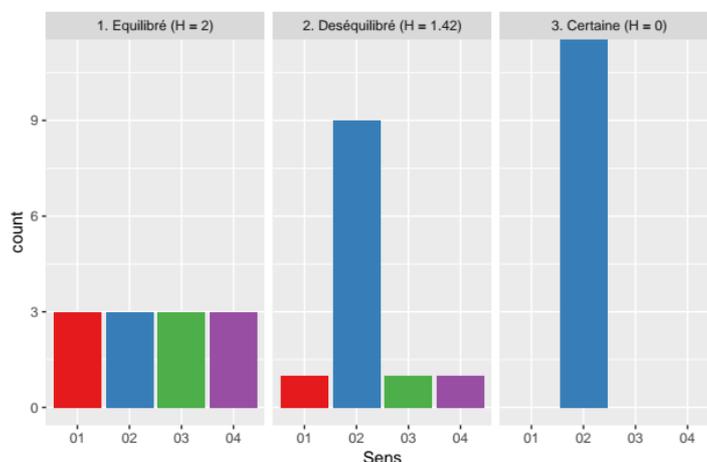


Figure 2. – Différents types de distributions statistiques

La notion d'entropie ayant été définie et illustrée, revenons à notre question de recherche. S'il y a une spécialisation des sens au moment de la troncature, nous pouvons nous attendre à ce que la distribution des sens d'un troncat ait une entropie plus basse que celle de sa base. Les bases non-spécialisées doivent regrouper plus de sens, répartis plus ou moins

équitablement. Cela équivaut à une entropie élevée. Pour les troncats, nous nous attendons à avoir moins de sens attestés et/ou une répartition plus déséquilibrée, et donc à ce que l'entropie soit relativement 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 3. Nous avons grisé les paires de cases qui correspondent à une entropie plus basse pour le troncat que pour la base.

	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

Tableau 3. – Mesure d'entropie pour les distributions de sens

À partir du tableau, nous observons 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. Dans le seul cas qui fait exception, qui est celui de la paire (INTOXICATION, INTOX), les valeurs sont très proches. Dans les forums, dans 7 cas sur 10 nous avons une diminution de l'entropie, et le seul cas où il existe une différence importante qui va dans un sens inattendu est celui de la paire (INTOXICATION, INTOX). 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 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 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 ($t(9) = -3,976$; $p < 0,001$), 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 troncats⁸.

Nous pouvons donc conclure que, comme le prédit la littérature, il y a bien une réduction de la diversité des sens lorsque nous passons d'un lexème non-tronqué au troncat correspondant. Il est important de noter que nous ne rencontrons dans aucun des cas une entropie nulle : la réduction est donc partielle et non-catégorique, mais elle est bien réelle.

3.5. Les sens événementiels sont-ils défavorisés pour les troncats ?

Ayant établi que la troncation opère une spécialisation sémantique, il nous reste à mettre à l'épreuve l'hypothèse plus spécifique de Kerleroux (1999), selon laquelle les noms d'événements complexes ne sont pas apocopables, contrairement aux noms d'événements simples et aux noms résultatifs. Nous avons déjà établi dans la section 2 que l'hypothèse ne pouvait pas être maintenue sous une forme catégorique. Nous pouvons cependant reformuler l'hypothèse sous une forme graduelle : on s'attend à ce que les sens correspondant à un emploi comme nom d'événement complexe soient moins fréquents pour les troncats que pour leurs bases.

Il ne nous a malheureusement pas été possible de tester cette hypothèse directement. En effet, les critères permettant de distinguer les noms d'événements simples et complexes ne sont pas discriminants en corpus, la distribution des noms d'événements complexes étant incluse dans celle des noms d'événements simples. Par exemple, les noms d'événements complexes sont réputés ne se rencontrer qu'accompagnés de l'article défini singulier, mais les noms d'événements simples peuvent aussi être rencontrés avec cet article. Face à une occurrence indéfinie ou plurielle, on est donc en mesure de trancher, mais pas face à une occurrence au défini singulier, ce qui est le cas le plus fréquent.

Au vu de cette situation, nous nous sommes contentés d'examiner la distribution des sens correspondant à une dénotation de type événement ou objet concret. En effet, si Kerleroux a raison en ce qui concerne les noms d'événements complexes, et si par ailleurs l'emploi des lexèmes comme nom d'événements simples n'est pas affecté par la troncation, on s'attend à ce que

⁸ Nous avons en outre cherché à vérifier s'il existait une réduction de sens entre deux bases lorsque nous passions 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.

la proportion totale de noms d'événements (simples ou complexes) soit plus faible pour les troncats que pour leurs bases.

Nous avons donc 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, le plus souvent définis dans nos dictionnaires selon le patron « Action de V », 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 1, 3 et 4, et de l'autre les sens 2, 5, 6, et 7 (voir les annexes d'Anselme 2019 pour des listes exhaustives des sens accompagnés de définitions). Le tableau 4 donne une idée des effectifs.

	Hors Forum		Forum	
	Sens non-événementiel	Sens événementiel	Sens non-événementiel	Sens événementiel
Base	269	181	279	171
Troncat	363	86	360	86

Tableau 4. – Compte des sens événementiels et non-événementiels

À partir du Tableau 4, 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 du compte des sens non-événementiels entre la base et le troncat, aussi bien en forum qu'en dehors 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.

Pour valider statistiquement nos résultats, nous avons construit un modèle de régression logistique à effets mixtes grâce au package lme4 et la fonction glmer (2014) sous R qui cherchait à prédire la forme d'un lexème (pouvant avoir deux valeurs, tronqué ou non tronqué) en fonction du genre textuel (forum ou non-forum) et du type de sens (événementiel ou non). Nous avons ajouté un effet aléatoire pour l'influence que les différents lexèmes pouvaient avoir, correspondant aux paires (base, troncat), pour rendre compte d'une probable variabilité en fonction de la paire de lexèmes considérée. Nous avons centré nos effets fixes et aléatoires. Les résultats de notre modèle nous indiquent que le type de sens a un effet clairement significatif sur le fait que la forme soit tronquée ou non ($p < 0,001$). Une fois de plus, nous ne documentons pas d'effet du genre textuel. Le modèle ne rend donc

pas compte d'un effet de la formalité sur la forme d'un lexème. Il n'y a pas non plus d'interaction documentée entre type de sens et genre textuel.

Nos résultats sont cohérents avec nos attentes. En effet, les sens événementiels étant désavantagés, il est possible d'inférer qu'une forme avec un sens événementiel a plus de chance d'être une forme longue, alors qu'une forme avec un sens non-événementiel a proportionnellement plus de chance d'être une forme tronquée. En supposant que la distribution des autres sens n'est pas affectée de manière systématique par la troncature, nous pouvons donc confirmer l'hypothèse de Kerleroux selon laquelle la troncature défavorise les noms d'événements complexes.

4. VERS UN MODÈLE PROBABILISTE DE MORPHOLOGIE CONSTRUCTIONNELLE

L'objectif de cet article est principalement empirique : il s'agissait pour nous de reformuler la généralisation de Kerleroux sous une forme non-catégorique et d'établir sa validité à partir d'une étude de corpus. Dans cette dernière question, nous voulons cependant souligner que la méthode que nous avons développée suggère une vision de la morphologie constructionnelle qui s'éloigne des modèles théoriques actuellement dominants.

Pour cette discussion nous partons de Fradin et Kerleroux (2009), qui constitue un point de référence utile en ce qu'il adresse (entre autres) certains des phénomènes discutés ici, et en particulier l'interaction de la polysémie avec l'économie des règles de construction de lexèmes (RCL). Fradin et Kerleroux s'intéressent à diverses situations où une RCL s'applique à une unité réputée polysémique, et notent que, dans cette situation, la règle ne s'applique pas à l'aveugle à l'ensemble des sens mais opère une sélection parmi ceux-ci. Cela les conduit à proposer un grain très fin d'identité lexémique, où ce qui est souvent traité comme un lexème polysémique est reconceptualisé comme un réseau de lexèmes distincts quoique sémantiquement similaires. La Figure 3 donne une image de l'économie des RCL dans une telle vision : les associations entre base et dérivé ne valent que pour un sens de base spécifique et un sens spécifique du dérivé ; les RCL ne sont en aucun sens polysémiques.

Il convient de noter que cette vision de la morphologie constructionnelle vise à capter une conception catégorique de la relation entre base et dérivé : une RCL est applicable à un lexème particulier (doté d'un sens spécifique) ou elle ne l'est pas. Elle est donc de fait opérante pour rendre compte de généralisations telle que celle de Kerleroux (1999) sous sa forme catégorique d'origine : il existe deux lexèmes MANIFESTATION, et la règle d'apocope ne peut pas s'appliquer à celui de ces deux lexèmes qui est un nom d'événement complexe.

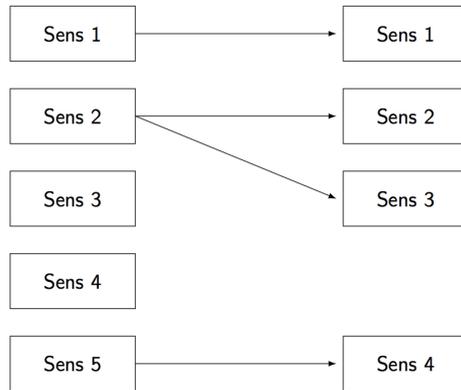


Figure 3. – RCL comme une relation entre sens spécifiques

Un résultat majeur de cet article est d'avoir mis à jour des généralisations sur lesquelles un tel modèle de la morphologie constructionnelle est, par construction, muet. Chacune des trois questions de recherche que nous avons abordées est formulée en considérant la distribution statistique des différents sens associés à une même forme. Les généralisations sont des généralisations sur la manière dont ces distributions sont affectées par la morphologie constructionnelle. Les RCL de Fradin et Kerleroux ne sont pas à même de modéliser de telles généralisations, dans la mesure où elles se focalisent sur le niveau atomique d'une relation entre un sens de base et un sens dérivé, et sont donc aveugles à l'économie plus générale des distributions de sens.

Si ces généralisations ont une valeur, elles suggèrent qu'un modèle complet du système constructionnel doit s'autoriser à modéliser la relation entre la collection de sens associés à une même forme de base et la collection de sens associés à une forme dérivée. Cette association doit prendre une forme probabiliste : on relie la probabilité que la base prenne tel ou tel sens avec la probabilité que le dérivé prenne tel ou tel sens. La Figure 4 donne une l'image d'une telle relation.

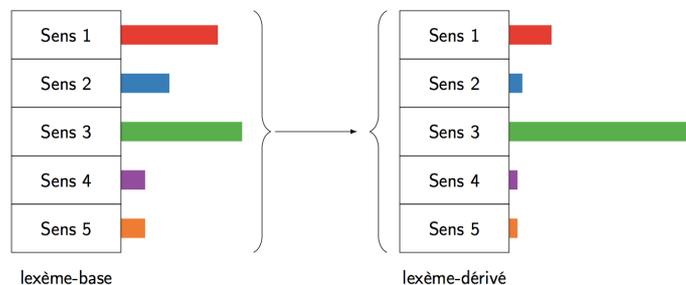


Figure 4. – Relations entre distributions de probabilités des sens

Dans un tel modèle, la situation qui est au centre de l'attention de Fradin et Kerleroux correspond au cas limite où le dérivé potentiel associé à l'un des sens de la base se voit assigner une probabilité nulle. Si nous avons ici remis en cause l'une des généralisations catégoriques sur lesquelles ils s'appuyaient, un modèle probabiliste n'est donc en rien contradictoire avec la prise en compte d'une telle situation là où elle est avérée. Le point important est que la prise en compte de la distribution des sens nécessite que la morphologie constructionnelle manipule collectivement des sens multiples, ce qui est directement contraire aux hypothèses de Fradin et Kerleroux.

Nous ne prétendons pas ici élaborer un modèle probabiliste de la construction de lexèmes. Une telle construction présente de nombreuses difficultés, qui sont en partie masquées par notre focalisation sur un procédé qui n'introduit aucune opération sémantique sur le sens de la base. Nous nous contentons de mettre sur la table un besoin qui devra être comblé par des travaux futurs.

5. CONCLUSION

Dans cet article, nous avons proposé une méthode pour explorer la polysémie des procédés constructionnels. À partir de la sélection d'un échantillon aléatoire d'attestations de bases et de dérivés, nous avons annoté manuellement les sens lexicaux rencontrés. Cette façon de faire nous a permis de faire des analyses statistiques sur les distributions de sens rencontrés, et nous a permis de valider sous une forme révisée une généralisation initialement proposée par Kerleroux (1999) : la troncation a un effet sur la distribution des sens des noms en *-ion*. Nous avons enfin discuté la manière dont un modèle théorique de la morphologie constructionnelle pourrait prendre en compte une telle généralisation.

La méthode employée dans la présente étude est certainement perfectible. En particulier, la collecte des sens à partir du dictionnaire plutôt que d'une étude indépendante est critiquable, et leur annotation en corpus est une tâche particulièrement difficile, sans doute entachée d'erreurs locales dont on espère qu'elles ne créent pas de biais. Cependant, elle ouvre des perspectives et pose des questions qui mériteraient d'être examinées à grande échelle. Ainsi, dans la section 3.4, on a établi que la distribution des sens d'un troncat est moins équilibrée que celle de sa base – en un certain sens que captent les mesures d'entropie, les dérivés sont moins polysémiques que leurs bases. Il est naturel de se demander s'il s'agit là d'une propriété spécifique à la troncation, ou si elle se généralise à une famille d'opérations de dérivation, voire à toutes. Anselme (2019) présente une étude pilote des verbes diminutifs qui va dans le sens d'une réponse positive. Ce type de verbes présente aussi une distribution moins équilibrée initiant une réduction de sens. Si de nouvelles études venaient à confirmer cette hypothèse, on pourrait spéculer sur les causes de cette baisse de la polysémie. Peut-être voit-on ici

simplement un effet secondaire du fait que les dérivés tendent à être plus récents que leurs bases, puisqu'on sait par ailleurs que les lexèmes plus anciens ont tendance à être plus polysémiques (Berdicevskis, 2020). Mais peut-être aussi la dérivation permet-elle un déséquilibre de la distribution des sens ayant pour conséquence une potentielle optimisation de la communication par affinage de sens, les mots les plus polysémiques étant sujets à plus de changements sémantiques (Hamilton et al., 2016).

Un autre aspect qui mériterait élaboration est la recherche d'une explication de la défaveur des troncats pour les sens événementiels : s'agit-il d'une pure convention grammaticale, ou cette défaveur est-elle motivée ? On pourrait spéculer qu'une motivation pourrait être trouvée en s'appuyant sur l'existence de corrélations non-catégoriques entre formes et sens des mots (Dautriche *et al.*, 2016). En français, la grande majorité des noms d'événements sont des dérivés par suffixation, et sont donc porteurs d'un indice phonologique explicite de leur sémantique à travers la phonologie du suffixe. Cet indice est utile pour signaler un type dénotatif qui n'est pas prototypique pour un nom (Croft, 1991). La troncature effaçant le suffixe, elle fait disparaître cet indice, et le troncat a donc une phonologie qui ne fait pas l'objet d'une association systématique avec une sémantique événementielle : il ressemble à un nom ordinaire, et est donc plus volontiers employé avec une sémantique ordinaire. Quoique cette hypothèse ait une certaine plausibilité initiale, sa confirmation empirique est loin d'être acquise, et devra attendre une future étude.

RÉFÉRENCES

- ANSELME R. (2019). *La troncature en français*. Mémoire de master 2. Université Paris-Diderot, Paris.
- ANTOINE F. (2000). *Dictionnaire français-anglais des mots tronqués*, vol. 105. Louvain : Peeters Publishers.
- BATES D., MÄCHLER M., BOLKER B., WALKER S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
- BERDICEVSKIS A. (2020). Older English words are more polysemous. In : *The Evolution of Language : Proceedings of the 13th International Conference*, 14-21.
- BONAMI O., BOYÉ G., KERLEROUX F. (2009). L'allomorphie radicale et la relation flexion-construction. In : B. Fradin, F. Kerleroux, M. Plénat (éds), *Aperçus de morphologie du français*. Saint-Denis: Presses Universitaires de Vincennes, 103-125.
- CROFT W. (1991). *Syntactic Categories and Grammatical Relations*. Chicago : The University of Chicago Press.

- DAUTRICHE I., MAHOWALD K., GIBSON E. & PIANTADOSI S.T. (2016). Wordform Similarity Increases With Semantic Similarity: An Analysis of 100 Languages. *Cognitive Science* 2016, 1-21.
- FRADIN B. & KERLEROUX F. (2009). L'identité lexémique. In : B. Fradin, F. Kerleroux, M. Plénat (éds), *Aperçus de morphologie du français*. Saint-Denis: Presses Universitaires de Vincennes, 85-102.
- GRIMM S., & MCNALLY L. (2013). No ordered arguments needed for nouns. *Proceedings of the 19th Amsterdam colloquium*, 123-130.
- GRIMSHAW J. (1990). *Argument structure*. Cambridge : MIT Press.
- HAMILTON W.L., LESKOVEC J., & JURAFSKY D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, 1489-1501.
- 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 ? In : D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat, M. Roché (éds), *La morphologie des dérivés évaluatifs. Forum de morphologie (2^e rencontres). Actes du colloque de Toulouse (29-30 avril 1999)*, Silexicales 2, U.M.R. 8528 du C.N.R.S. (SILEX) / Université de Lille III, 95-106.
- OCHS E. (1993). Constructing social identity: A language socialization perspective. *Research on language and social interaction* 26 (3), 287-306.
- RACCAH P.Y. (2015). Quelques infos sur les nœuds-pap et sur les collabos. Non-publié.
- ROY I., SOARE E. (2013). Event-related Nominals. In: G. Iordăchioaia, I. Roy & K. Takamine (eds), *Categorization and category change*. Newcastle : Cambridge Scholars Publishing, 123-152.
- R CORE TEAM (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Consulté sur <https://www.R-project.org/>
- RSTUDIO TEAM. (2015). Rstudio: Integrated development environment for R [Manuel de logiciel]. Boston, MA. Consulté sur <http://www.rstudio.com/>
- SCHÄFER R. (2015). Processing and querying large web corpora with the COW14 architecture. *Proceedings of Challenges in the Management of Large Corpora*, 28-34.
- SCHÄFER R., & BILDHAUER F. (2012). Building large corpora from the web using a new efficient tool chain. *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, 486-493.
- SHANNON C.E. (1948). A mathematical theory of communication. *The Bell system technical journal* 27 (3), 379-423.