



HAL
open science

Analyse lexicologique diachronique du profil phonosémantique des monomorphèmes en gr- dans l'Oxford English Dictionary

Mael Farina

► **To cite this version:**

Mael Farina. Analyse lexicologique diachronique du profil phonosémantique des monomorphèmes en gr- dans l'Oxford English Dictionary. Sciences de l'Homme et Société. 2020. dumas-03478018

HAL Id: dumas-03478018

<https://dumas.ccsd.cnrs.fr/dumas-03478018>

Submitted on 13 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire de Master 2

Mention Langues, Littératures et Civilisations Étrangères
et Régionales

Parcours Linguistique Multilingue

Analyse lexicologique diachronique du profil phonosémantique des monomorphèmes en *gr-* dans l'*Oxford English Dictionary*

Réalisé par Mael Farina

Sous la direction de Mme Chris Smith

Année universitaire 2019-2020



UNIVERSITÉ DE CAEN
NORMANDIE
Langues Vivantes Étrangères



Analyse lexicologique diachronique du profil phonosémantique des
monomorphèmes en *gr-* dans l'*Oxford English Dictionary*

REMERCIEMENTS

Avant de présenter l'étude qui suit, je tiens à exprimer ma gratitude envers les personnes et institutions qui m'ont permis de réaliser ce travail. Je remercie d'abord Mme Chris Smith, maître de conférences à l'Université de Caen Normandie et membre du Crisco, pour avoir encadré ce projet de Master 2, pour ses relectures, ses conseils et son soutien. Merci à l'ensemble des enseignants-chercheurs de l'Université de Caen, aux membres du Crisco et à tous les enseignants qui m'ont apporté leurs conseils. Merci en particulier à M. le Professeur Eric Gilbert pour ses conseils avisés. Je tiens aussi à remercier Mme la Professeure Isabelle Haïk pour m'avoir apporté des réponses lorsque j'ai décidé d'entreprendre ce projet. Je souhaite également remercier mes camarades de promotion sans qui cette année académique aurait été bien plus difficile encore. Enfin, je remercie mes amis et mes parents pour leur soutien et leur patience. Merci à ma mère pour ses relectures et ses précieux encouragements.

Résumé

Cette étude empirique fait état du comportement diachronique des monomorphèmes en *gr-* dans l'*Oxford English Dictionary*. Elle vise à dresser le profil sémantique du phonesthème *gr-*, à travers des données lexicographiques. La méthodologie s'appuie sur une analyse statistique par mots clés à travers les entrées lexicales, en vue d'identifier un petit nombre de traits phonosémantiques distinctifs, associés au marqueur submorphémique. Cette analyse est complétée d'un examen des changements métasémiques d'un nombre limité de traits, ainsi que d'une analyse des racines étymologiques à l'origine des lexèmes du corpus. Les résultats montrent la présence d'un trait subsémantique fondamental lié à la notion d'émotions négatives. Ce trait semble suivi d'autres traits subsémantiques plus faiblement attestés, associés aux notions de son et de sensations déplaisantes, qui entretiennent des relations métasémiques avec celui-ci. Les données métasémiques et étymologiques présentent d'une typologie des traits subsémantiques, dont certains seraient associés à un nombre plus limité de racines, un profil métasémique moins diversifié, et un nombre moindre de lexèmes aux origines décrites comme iconiques.

Mots clés : phonesthèmes, symbolisme phonétique, iconicité, changement sémantique, lexicologie diachronique, morphosémantique, profil cognitif, OED

Abstract

This empirical study reports the diachronic behaviour of *gr-* monomorphemes in the *Oxford English Dictionary*. It aims at tracing back the semantic profile of the phonestheme *gr-* from lexicographic data. The methodology relies on a keyword-based statistical analysis across lexical entries, in order to single out a handful of distinctive phonosemantic features that are associated with the submorphemic onset. An examination of the metasemic changes of a limited number of features, as well as an analysis of the etymological roots which the lexemes of the corpus stem from were also conducted. The results point at a core subsemantic feature associated with a notion of negative emotions. This feature seems to be followed by less statistically attested features, associated with the notions of sound and unpleasant sensations, which carry metasemic relations with the latter. Metasemic and etymological data hint at a typology of subsemantic features, since some of them seem to be associated with a more limited number of roots, a less diverse metasemic profile, as well as a smaller number of words whose origin is said to be iconic.

Key words: phonesthemes, sound symbolism, iconicity, semantic change, diachronic lexicology, morphosemantics, cognitive profile, OED

TABLE DES MATIERES

TABLE DES MATIERES	5
INTRODUCTION.....	1
PARTIE I : CADRE THEORIQUE ET EPISTEMOLOGIQUE	3
I. Prises de position théoriques.....	3
1.1. Iconicité, symbolisme phonétique et phonesthésie : état de l'art.....	3
1.2. Statut du phonesthème en théorie morphologique	6
1.3. Linguistique cognitive et sémantique incarnée	8
1.4. Bilan sur la terminologie	10
II. Nature et prise en charge des données de terrain	11
2.1. Support de la psycholinguistique	11
2.2. Linguistique computationnelle et données statistiques	14
2.3. Études de corpus.....	16
2.4. Dimension diachronique	17
PARTIE II : METHODOLOGIE DE LA RECHERCHE	20
I. L'Oxford English Dictionary comme outil de recherche	20
3.1. Sources et spécificités de l'OED	20
3.2. Aspects chronologiques et perspectives diachroniques.....	22
3.3. Collecte des données	23
II. Présentation du corpus	27
4.1. Le phonesthème <i>gr-</i> dans la littérature	27
4.2. Structure submorphémique de <i>gr-</i>	30
4.3. Sélection des mots clés.....	33
4.4. Précisions sur quelques biais inhérents à la nature du corpus.....	34
III. Comparaisons entre EBC et PAC	36
5.1. Premier regroupement en catégories	36
5.2. Elargissement aux EBC et étude des proto-sens	37
5.3. Apports de l'étude des sens métasémiques	42
5.4. Données statistiques sur l'obsolescence et les types d'emplois.....	45
PARTIE III : RESULTATS	48
I. Nature et évolution des changements sémantiques.....	48
6.1. Courbes de progression	48
6.1.1. Présentation du protocole	48

6.1.2. Analyse des EBC	50
6.1.3. Analyse des PAC	53
6.2. Schémas combinatoires	56
6.2.1. Présentation du protocole	56
6.2.2. Analyse des EBC	57
6.2.3. Analyse des PAC	65
6.2.4. Comparaison et conclusion.....	70
6.3. Co-émergences et post-attestations	72
6.3.1. Protocole de l'analyse et précisions sur les EBC	72
6.3.2. Analyse globale des PAC	73
6.3.3. Intervalles chronologiques.....	76
6.3.4. Evolution, divergence et convergence : conclusion partielle	79
6.4. Identification de l'organisation métasémique de <i>gr</i> -.....	81
6.4.1. Présentation du protocole	81
6.4.2. SVU.....	81
6.4.3. HTC.....	84
6.4.4. Bilan	85
II. Origines étymologiques des monomorphemes en <i>gr</i>-.....	86
7.1. Production lexicologique : bilan chronologique	86
7.2. Paysage étymologique des monomorphèmes en <i>gr</i> -.....	88
7.2.1. Profil global	88
7.2.2. Profil trait par trait	92
7.3. Évolutions sémantiques.....	96
7.3.1. Racines communes et évolutions morphosémantiques.....	96
7.3.2. WCAF	97
7.3.3. CSS	101
7.3.4. SVU.....	103
7.3.5. HTC	104
CONCLUSION.....	106
BIBLIOGRAPHIE	110
Outil de recherche	110
Références	110

Introduction

Lorsque l'on tente de décrire les différents types de rapports forme-sens qu'il est possible de distinguer, la composante arbitraire apparaît comme une notion élémentaire et incontournable. Et pour cause, la convention linguistique, principe fondamental qui assure la cohésion de la langue en tant que système, peut être de nature symbolique, c'est-à-dire que les relations forme-sens sont arbitraires, ou bien motivée (Marchand 1957, Haiman 1980). C'est ce concept de motivation qui sera au cœur de notre étude, puisqu'il est susceptible d'expliquer l'origine et le développement de signes linguistiques qui semblent échapper au critère fondamental de l'arbitraire du langage humain, notamment établi par Ferdinand de Saussure au début du XX^e siècle, et que l'on appelle *phonesthèmes*. Si cette motivation du signe touche à tous les domaines d'étude du linguiste, étant donné qu'elle peut se traduire sous forme phonologique (Bloomfield 1914 ; Bolinger 1940), morphologique et sémantique (Koch et Marzo 2007, Mosegaard Hansen 2008, pp. 72-77), syntaxique (Haiman 1985) ou même pragmatique, nous nous intéresserons plus particulièrement à ses manifestations phonologiques et morphologiques. L'étude approfondie de la structure phonologique et morphologique du corpus lexical que nous visons à présenter ici s'inscrit dans une recherche, particulièrement active depuis ces dernières années, portant sur des signes dont l'origine et l'évolution sont encore mal comprises. La nomenclature attachée à décrire ces unités est variée et hétérogène : ils sont tantôt qualifiés de *phonesthèmes*, d'après l'expression de Firth (1930, p. 184), de *submorphèmes* (Crystal 2008, p. 313), ou encore d'*éléments* ou de *marqueurs sublexicaux* ou *submorphémiques* (Bottineau 2008 ; Philips 2012a). Le fait que ces éléments touchent à des domaines variés souligne bien l'importance de leur étude, qui ne se limite pas au symbolisme phonétique et à la phonesthésie, mais qui englobe tous les champs de recherche liés à la motivation du signe linguistique.

Les phénomènes de type phonosymbolique ou phonesthétique sont connus depuis longtemps dans la littérature (Wallis 1653, Jespersen 1922, Firth 1930, Bloomfield 1933, Bolinger et Sears 1981, etc.), mais le statut qu'ils occupent dans les théories morphosémantiques est disputé. Comme nous le verrons au cours de cette étude, la relation forme-sens que présentent ces éléments pose de nombreuses questions. L'objectif de la présente étude est de suivre l'évolution de lexèmes porteurs du phonesthème *gr-* à travers le temps, dans un corpus lexicographique diachronique. Nous suivrons les méthodes d'analyse statistique par mots clés développées, entre autres, par Abelin 1999, Drellishak 2006 et Otis et Sagi 2008, puis

appliquées à l'analyse lexicographique dans l'*Oxford English Dictionary* par Smith (2016 et 2020). L'application de ces méthodes à la métaclasse des mots du lexique anglais en *gr-* vise à observer le comportement sémantique des lexèmes du corpus, en vue de dégager des tendances évolutives, qui participeront de la compréhension de ce phonesthème. Ce type d'étude s'inscrit dans une perspective avant tout empirique, se fonde sur les postulats de la linguistique cognitive, et s'appuie sur les progrès de la linguistique computationnelle de ces dernières années. Notre étude répond ainsi au besoin d'une meilleure compréhension de l'influence des phonesthèmes sur le lexique. En effet, si, comme nous allons le voir, la réalité des phonesthèmes a été largement démontrée sur les plans statistique et psychologique (Hutchins 1998 ; Abelin 1999 ; Bergen 2004 ; Abramova *et al.* 2013), leur émergence et leur évolution dans le lexique de l'anglais posent encore problème. En particulier, si les données psycholinguistiques indiquent une influence des phonesthèmes sur le lexique mental des locuteurs, leur rôle potentiellement structurel dans la construction du lexique institutionnel est encore incertain (Smith 2020, pp. 37-38). Dans cette optique, les études de corpus lexicographiques peuvent nous apprendre davantage sur la nature et la portée de la phonesthésie dans la langue.

Les débats relatifs à la nature et au statut des phonesthèmes, ainsi que l'avancée de l'état de l'art, seront exposés dans la première partie de cette étude, qui porte sur les aspects théoriques et épistémologiques adossés à ce sujet. Nous y aborderons notamment les notions d'iconicité, de symbolisme phonétique, de motivation et d'isomorphisme, afin d'intégrer notre problématique aux questions plus larges ayant trait à la nature du signe linguistique. La deuxième partie propose une méthodologie fondée sur une analyse lexicographique et lexicologique d'un ensemble de lexèmes potentiellement phonesthétiques, à travers l'*Oxford English Dictionary*¹. La constitution du corpus, la sélection des mots clés et la structure du phonesthème *gr-* y seront explicités. La troisième partie présente le détail des résultats issus de ce protocole, leurs implications et les interprétations qu'il est possible d'en tirer.

¹ *Oxford English Dictionary*, version en ligne pour abonnés, 3ème édition.

Partie I : Cadre théorique et épistémologique

I. PRISES DE POSITION THEORIQUES

1.1. Iconicité, symbolisme phonétique et phonesthésie : état de l'art

Ce premier volet vise à définir quelques termes qui gravitent autour du concept de phonesthésie, et à replacer notre propos dans les problématiques plus larges qui animent la recherche dans ce domaine. Définir les phonesthèmes en tant qu'éléments linguistiques est particulièrement difficile, car la terminologie ne bénéficie, à l'heure actuelle, d'aucun consensus théorique. C'est pourquoi les approches que nous tenterons de résumer ici ne vont pas toutes dans le même sens. Des relations non-arbitraires entre forme et sens sont connues depuis longtemps dans la littérature, et l'on peut même remonter aux auteurs antiques grecs et latins, comme Platon, qui parle de correspondances entre sons et significations dans son *Kratylos* (voir Jespersen 1922, p. 396). La première observation concernant les phénomènes que nous nous attacherons à décrire ici est visible dans la *Grammatica Linguae Anglicanae* (1653) de Wallis, mais c'est Jespersen (1922) qui introduit en premier l'expression de symbolisme phonétique². Il fait notamment part de certains mots qu'il décrit ainsi : « words expressive of such movement as are not to the same extent characterized by loud sounds » (1922, p. 399), avant de mentionner l'élément *-l-*, qui apparaît dans les mots en *fl-* comme *flow*, *flag*, *flake* ou *flutter*, ainsi que dans les mots en *gl-* comme *glide* (1922, p. 400). Il remarque également qu'un nombre important de ces mots sont des amalgames, comme *flurry*, issu de *fling* ou de *flow* et de *hurry*, ou encore *gruff*, qu'il présente comme un amalgame de *grum* et de *rough* (1922, p. 318).

Sapir reprend plus tard l'expression de « phonetic symbolism », afin d'en préciser les contours (Sapir 1929, p. 233). Ces observations dans le champ de la langue font, dès cette époque, écho aux travaux de Peirce, qui, pour sa part, définit le symbole comme l'objet conventionnel typique. En effet, le modèle sémiotique peircien distingue trois types de signes : les signes symboliques, qui répondent exclusivement au principe fondamental de convention, les signes iconiques, qui relèvent d'une similarité entre le signifiant et le signifié, et les signes indexicaux, qui relèvent d'une contiguïté entre la forme et le contenu ; ces deux derniers étant motivés (Anttila 1977, pp. 7-8 ; Eschbach 1983, p. xxii ; Monneret 2020, pp. 8-11). Il ne s'agit

² Voir Jespersen 1922, pp. 396-411, chap. XX "Sound-symbolism".

cependant que d'un modèle général, et Peirce lui-même explique qu'un unique signe peut facilement endosser plusieurs de ces caractéristiques, voire toutes les trois la fois (Jakobson 1980, pp. 11-12 ; p. 38). Il semble que chez Sapir, le symbolisme est soit défini comme un phénomène d'ordre référentiel, ce qui le rapproche du sens de Peirce, ou bien comme phonétique, c'est-à-dire qui met en regard une séquence phonologique et un référent, de manière plus ou moins non-arbitraire (Sapir 1929, pp. 225-226). Les jeux d'intonation, tout comme la distinction vocalique entre *tiny* et *teeny*, sont déjà mentionnés comme vecteurs d'une différence « expressive » (Sapir 1929, p. 226). Viennent ensuite les travaux de Jakobson, qui s'empare des concepts précédents, et introduit la notion d'iconicité, jusque-là développée par les sémiologues, dans les sciences du langage (Jakobson 1965 ; 1978). Il reprend d'ailleurs la distinction peircienne entre iconicité d'image et iconicité diagrammatique : la première se fonde sur des propriétés inhérentes communes au signifiant et au signifié, les unes se reflétant dans les autres ; la seconde, au contraire, se fonde sur des similarités formelles entre plusieurs signifiants, témoignant de similarités substantielles entre leurs signifiants respectifs (Haiman 1980, p. 515). Jakobson décrit le symbolisme phonétique comme un phénomène « synesthé[tique] », de nature « neuropsychologique », et insiste sur son importance dans la langue poétique (1978, p. 113).

Le terme de phonesthème, quant à lui, est défini en premier par John R. Firth, qu'il décrit comme suit : « initial and final phone groups not ordinarily recognised as having any function » (1930, p. 184). Plus précisément, il observe qu'un nombre important de mots commençant par le groupe consonantique *sl-* semblent être connotés négativement, et cite *slack*, *sludge*, *slush* et *slash*, en ajoutant ceci : « the more consistently similar sounds function in situations having a similar affective aspect, the closer their function » (1930, p. 185). Il insiste donc sur les similarités phonétiques, et leur projection sémantique, une idée qui rappelle l'iconicité diagrammatique de Peirce, mais également le principe d'isomorphisme, notamment développé chez Haiman (1980), dans le domaine grammatical, et Anttila (1977, p. 55), dans une démarche cognitive et diachronique. Les définitions subséquentes du symbolisme phonétique et du phonesthème, qui ont été construites dans des champs de recherche divers, seront présentées dans les sections ultérieures, consacrées aux domaines en question. Pour l'instant, nous nous contenterons de préciser l'une des dernières notions majeures de la recherche sur les associations forme-sens non-conventionnelles : la motivation. Ce concept est invoqué pour expliquer l'origine de ces associations, qui, pour certaines, semblent en rupture avec le principe saussurien d'arbitraire du langage (voir Saussure 1916, pp. 100-102). Haiman distingue la

motivation iconique de l'isomorphisme, car la seconde est posée comme universelle, à l'inverse de la première, dont les manifestations varieraient d'une langue à l'autre (1980, p. 515). Plus précisément, cette définition de l'isomorphisme soutient qu'une similarité formelle d'un signe à l'autre indique une similarité sémantique, de même qu'une différence formelle indique une différence sémantique (Waugh 1994, p. 54). Les phénomènes de polysémie et d'homonymie représentent donc des exceptions à ce principe (Haiman 1980, p. 516). Il ressort, dans tous les cas, que les relations entre isomorphisme et motivation manquent de clarté, tant les définitions divergentes se multiplient. Pour terminer cette présentation, notons que la notion de motivation est déjà présente chez Marchand (1957, p. 62), qui interprète certains phénomènes de rime et de variation apophonique comme le résultat d'une motivation de la forme linguistique. Il cite par exemple les paires minimales *sing/song*, *chit/chat* et *flim/flam*.

Notre objectif est avant tout empirique, et non théorique, mais, au vu de l'évolution observée à travers la littérature, on pourrait émettre l'hypothèse que les symboles peuvent montrer un niveau plus ou moins élevé d'iconicité ou d'indexicalité. Cette iconicité ou cette indexicalité résultent d'un phénomène plus profond d'analogie cognitive (Anttila 1977 ; Monneret 2020), et elles ont tendance à s'estomper au cours de l'évolution d'un signe (Flaksman 2016 ; 2017). Un symbole iconique ou indexical devient de plus en plus conventionnel, car il se conforme au système linguistique global. Les symboles les plus iconiques ou indexicaux sont généralement moins lexicalisés et grammaticalisés, avant de s'incorporer peu à peu au système. Le système a cependant tendance à constamment recréer de nouveaux signes motivés, car la motivation permet une meilleure transparence forme-sens, ce qui représente un avantage sur le plan de la reproduction sémiotique. Un signe plus transparent est en effet plus facilement accessible aux locuteurs, ce qui maximise ses chances d'être employé. D'où l'importance de cette transparence dans la production littéraire et poétique (Wolf 2001, p. 341). Toutefois, plus un signe est utilisé, plus son niveau de lexicalisation et de grammaticalisation est marqué, ce qui engendre un mécanisme de conventionnalisation du signe, et ainsi de suite. Dans une optique de théorisation, on pourrait alors parler d'un cycle de conventionnalisation du signe linguistique, dans la lignée du principe de « de-iconization » de Flaksman, qu'elle définit comme une perte progressive du caractère iconique d'un lexème, due à des changements phonétiques réguliers et à des glissements de sens (2016, p. 46 ; 2017, pp. 33-34). La présente étude n'a pas pour prétention de résoudre la question qui suit, mais on sera, à terme, en droit de se demander si les notions de polysémie et d'homonymie sont applicables au domaine du symbolisme phonétique, et, en ce qui nous concerne, de la phonesthésie.

1.2. Statut du phonesthème en théorie morphologique

La définition du phonesthème, d'un point de vue morphosémantique, pose problème pour différentes raisons, toutes décrites dans la littérature. Chacune des caractéristiques singulières du phonesthème ne soulève pas nécessairement de question en elle-même, mais c'est leur accumulation qui a rendu perplexes des générations de linguistes. Firth (1930) accorde au phonesthème un statut explicite au sein de son modèle, et Bolinger soutient l'existence d'un continuum entre les différents niveaux sémiques d'un morphème : « the transition from sub-unit to morpheme is, as regards meaning, not abrupt » (1948, p. 18). En revanche, Bloomfield (1926, pp. 155-157) rejette toute association phonosémantique, car le morphème ne peut être subdivisé en de plus petits éléments porteurs de sens, et soutient que les phonèmes sont nécessairement dénués de sens. La double articulation du langage de Martinet (1970), à son tour, rigidifie encore un peu plus le modèle d'une distinction binaire et inaltérable entre les niveaux phonologique et morphologique. De même pour l'imperméabilité des systèmes phonologique et morphologique imposée par le principe de dualité de Hockett (voir par exemple 1958, p. 574). Si l'on définit les phonesthèmes comme des unités porteuses de sens, composées d'éléments plus petits et dénués de sens, on pourrait cependant les rapprocher de la définition classique du morphème (Kwon et Round 2014, p. 3).

Entre alors le paramètre de la compositionnalité, régulièrement avancé comme critère discriminant (Bergen 2004, p. 4). L'idée est que le système linguistique développe une interconnexion entre des unités élémentaires, qui se combinent entre elles pour former des unités plus complexes, dans une optique d'efficacité sémiotique. Ce mécanisme permet en effet aux locuteurs d'interpréter de nouveaux signes par leur composition, stratégie plus efficace qu'une communication holistique (Bergen et De Beule 2006). Monneret (2020, p. 12 ; p. 17) présente d'ailleurs ce principe fondamental comme le moteur de l'iconicité dans le lexique. Le corolaire de ce principe est qu'un élément compositionnel porteur de sens n'est défini comme tel que si le segment restant une fois la forme tronquée est également analysable comme un élément porteur de sens (Blust 1988, p. 5 ; Perniss et al. 2010, p. 7). Afin d'illustrer l'écart alors évident entre morphologie et phonesthésie, on peut prendre l'exemple de la séquence phonologique /gl/, présente à l'initiale des mots *gleam*, *glow*, *glint* ou *glance*. Cette séquence est largement considérée comme phonesthétique dans la littérature, car on y associe une notion de lumière ou de vision (Drellishak 2006, Sadowski 2009, pp. 219-220). Or, les segments résiduels **-eam*, **-ow*, **-int* et **-ance* ne sont pas analysables comme des unités porteuses de sens, contrairement à une formation lexicale comme *un-interest-ing* par exemple, composée de

trois morphèmes nettement distinguables. Des contre-exemples, à la fois de morphèmes en partie non-compositionnels, et de combinaisons possibles de phonesthèmes, existent cependant.

On peut d'abord citer quelques exemples de morphèmes non-compositionnels *stricto sensu* : on aura plus de mal à décomposer des lexèmes comme *cranberry*, *raspberry*, *naughty* ou *clumsy*, que des lexèmes plus évidents comme *daily*, *rainy* ou *spittle* (Waugh 1994, p. 58). En effet, de même que pour les mots en *gl-*, les segments résiduels *rasp-*, *naught-* ou *clums-* sont difficilement analysables sans leur élément associé. Notons toutefois que si le sens de ces segments n'est pas directement accessible au locuteur en l'état, il peut être retrouvé grâce à leur étymologie, contrairement aux éléments comme **-eam* ou **-int*. On pourrait en fait parler de pseudo-morphèmes, dont l'intégrité formelle a été érodée avec le temps, mais qui constituaient, originellement, d'authentiques morphèmes indépendants, dans le même esprit que les constructions telles que *heli-copter*, dont les étymons d'origine grecque ne sont plus distingués par les locuteurs (Kolin 1979, p. 74). La situation semble donc différente des constructions phonesthétiques, mais la similarité mérite d'être soulignée, et elle attire notre attention sur l'importance de l'étymologie et de la recherche diachronique.

Par ailleurs, certaines constructions semblent faire appel à une conjugaison de deux phonesthèmes, comme *tw-irl*, les segments *tw-* et *-irl* étant tous deux réputés phonesthétiques (Kwon et Round 2014, p. 4). En outre, certains segments résiduels, associés à des phonesthèmes considérés comme tels dans la littérature, mais n'étant pas réputés phonesthétiques eux-mêmes, semblent présenter un certain degré de récurrence. Bloomfield mentionne par exemple la qualité expressive des segments résiduels *-immer* et *-icker* dans *flimmer* et *flicker*, que l'on retrouve dans *glimmer* et *snicker* ; de même pour la séquence *-ash* de *flash*, que l'on retrouve dans *slash*, *smash*, *clash* ou *crash* (1914, p. 79). Notons que si *-ash* est parfois décrit comme phonesthétique, ce n'est pas le cas de *-immer* et *-icker*, malgré leur caractère récurrent. Notons par ailleurs que les séquences phonologiques décrites comme expressives ne se limitent pas aux groupes consonantiques. Bloomfield mentionne en effet l'impression d'impact de *b-* à l'initiale, dans *bang*, *bash*, *bounce*, *biff*, *bump* et *bat*, bien que cette intuition n'ait pas été nettement confirmée depuis. De même, Bolinger (1940, p. 68) fait mention d'un gargouillis de la gorge dans *gullet*, *gorge*, *gargle* et *guttural*.

Il semble qu'aucune différence fondamentale ne puisse être établie entre les phonesthèmes et les morphèmes, mais que la distinction entre les deux soit de nature essentiellement statistique. En effet, si ces deux types d'unités porteuses de sens répondent aux

mêmes paramètres, c'est l'adéquation avec laquelle ils s'y conforment qui varie. En somme, les phonesthèmes ont globalement une plus forte probabilité de s'associer à des segments non-compositionnels, qui ne véhiculent pas de sens et qui ne sont que rarement récurrents ; ils ont également tendance à être plus iconiques que les morphèmes, et à apparaître dans des constructions lexicales non-concaténatives. Alors que les morphèmes peuvent être grammaticaux ou lexicaux, les phonesthèmes ne semblent toutefois pas subir de grammaticalisation. En effet, ils ne participent pas aux flexions grammaticales habituelles, comme les conjugaisons. En somme, ils n'agissent pas, jusqu'à preuve du contraire, comme des marqueurs grammaticaux, comme le font certains morphèmes. Ils présentent aussi une productivité lexicale plus difficile à estimer que celle des morphèmes. Enfin, on ignore encore si les phonesthèmes sont susceptibles de briser le principe d'isomorphisme (au sens de Haiman). Autrement dit, les cas d'homonymie et de polysémie sont, jusqu'à preuve du contraire, l'apanage des morphèmes. Nous reviendrons plus dans le détail sur les composantes statistiques de la définition du phonesthème dans une section ultérieure.

1.3. Linguistique cognitive et sémantique incarnée

Si Sapir décrit le symbolisme phonétique comme un vecteur d'expression, Bloomfield parle déjà, au début du XX^e siècle, de « mouvements expressifs » (1914, p. 77). Il donne l'exemple des mots *flame*, *flare*, *flicker*, *flimmer* et *flash*, qui font tous référence au feu (1914, p. 79), ou à une source de lumière en mouvement (1933, p. 257). Il mentionne d'autres attaques consonantiques, dont un autre sens de *fl-*, qui serait « mouvement dans l'air », avec des mots comme *fly*, *flap*, *flit* et *flutter*. Fait intéressant, l'élément *gl-* est décrit comme une lumière sans mouvement, dans *glow*, *glare*, *gloat* ou *glint*, ce qui met en évidence un lien sémantique avec le premier *fl-*. Il fait également mention de phonesthèmes vocaliques, comme dans le paradigme *flip*, *flap*, *flop* (1933, p. 256). S'il serait tentant de considérer l'élément *-l-* comme responsable de la similarité entre *fl-* et *gl-*, en l'association à une notion de lumière, *sl-* en est un contre-exemple. Bolinger (1940, p. 68) souligne en effet l'idée de mouvement, en l'occurrence d'inertie, qui lui serait associé dans *slow*, *sluggish*, *slothful*, *slack*, *slush*, *slosh*, *slubber* et *slog*. Or *gl-* est décrit comme une lumière inerte. On pourrait alors tenter de rapprocher l'inertie de *sl-* et celle de *gl-*, mais on perd alors le lien avec l'élément *-l-* de *fl-*. La question d'une éventuelle structure submorphémique des phonesthèmes est adressée plus rigoureusement par Philips (2002 ; 2008 ; 2012a ; 2012b). Cet aspect, qui concerne les particularités phonotactiques des phonesthèmes, sera détaillé dans la partie dédiée à la méthodologie et à la nature de notre

corpus. On peut cependant déjà mentionner l'hypothèse selon laquelle les phonesthèmes, que Philips décrit comme des marqueurs sublexicaux ou submorphémiques, pourraient être formés de plus petits éléments porteurs de sens. L'élément *g-* du marqueur *gn-*, par exemple, semble être associé à un processus physiologique de manducation (Philips 2008).

Cette dernière observation entre dans le cadre des théories sur la dimension corporelle du langage, qui le présentent dans sa « dimension motrice, sensorielle et psychologique en tant que boucle motri-sensorielle » (Bottineau 2012, p. 4). Cet appareillage théorique nécessite encore d'être confirmé par les données neurolinguistiques (voir Ahlsén 2006, notamment p. 39, sur l'hypothèse d'une « communication incarnée »³), mais il présente un intérêt pour l'analyse des éléments submorphémiques, qui sont considérés comme la manifestation d'une conceptualisation du monde par les locuteurs à partir d'une expérience sensori-motrice (Bottineau 2012, p. 9). Bottineau cite notamment le groupe consonantique *wr-*, qui dénote une « notion de torsion », dans *wring*, *wrist*, *writh*, *wrap*, *wriggle*, et, de manière moins évidente, sans doute plus imagée, dans *wrath* et *write*. Des modèles de représentation, visant à illustrer les mécanismes à l'œuvre derrière cette idée de sémantique incarnée, ont été proposés, notamment par la linguistique cognitive. Ces propositions offrent une alternative aux théories critiquées comme « désincarnées » (Lakoff 2012, pp. 773-774), généralement issus du structuralisme américain, et qui se sont longtemps intéressées à la forme en faisant abstraction du sens et du discours, comme la grammaire générative (Anttila 1977, p. 7).

Nombreux sont les modèles qui proposent une interprétation cognitive du langage et des langues, tels que les *idealized cognitive models* (ICMs) de Lakoff et Johnson (Lakoff 1987, p. 68), mais rares sont ceux qui y incluent une dimension submorphémique. Certains de ces travaux ont été utilisés, *a posteriori*, pour apporter un cadre théorique aux observations empiriques et psycholinguistiques portant sur les phonesthèmes. Stefanowitsch reprend ainsi les modèles de représentation métalinguistique d'organisation mentale, notamment développés par Langacker (1987, 1991) et Lamb (1999). Stefanowitsch considère par exemple que les phonesthèmes ne brisent pas nécessairement le principe d'arbitraire du langage, et que leur différence avec les morphèmes est ténue (2002, p. 1). Il s'appuie sur des correspondances neurologiques avancées par Lamb (1999) pour construire un modèle de représentation, où le système linguistique est présenté comme un ensemble de noyaux. Dans la lignée des unités symboliques de Langacker, il propose une interprétation des phonesthèmes et des morphèmes

³ « Embodied communication ».

comme des nœuds d'informations, tous pourvus d'un pôle phonologique et d'un pôle sémantique, c'est-à-dire d'une forme et d'un contenu (Langacker 1991, p. 16 ; Stefanowitsch 2002, p. 6). Ce principe repose sur l'idée qu'une association récurrente entre un pôle phonologique et un pôle sémantique engendre une systématisation de la relation, de sorte que l'activation d'un des deux pôles a tendance à entraîner l'activation de l'autre. Ce modèle expliquerait alors l'émergence d'une structure phonosymbolique, à l'origine de la composante phonosémantique du phonesthème.

1.4. Bilan sur la terminologie

Dans la suite de notre étude, en particulier lors de la présentation de la méthodologie adoptée et des résultats, nous tenterons d'utiliser une terminologie la plus précise et la plus consistante possible. Au vu de la diversité des termes qu'il existe à travers la littérature pour se référer aux phénomènes de constructions lexicales submorphémiques, nous avons choisi de limiter notre nomenclature à quels éléments clés. Nous parlerons de phonesthèmes ou de marqueurs submorphémiques pour désigner les groupes consonantiques initiaux comme *gr-*. Nous parlerons ensuite de contenu phonosémantique ou phononotionnel pour désigner le sens véhiculé par le phonesthème, ce que Philps nomme l'*invariant phononotionnel*, ou ce que Stefanowitsch (2002) appelle l'*association* ou la *sensation phonosymbolique*. Ces termes sont en effet jugés comme interchangeable, bien que le concept de notion, qui fait référence à la théorie des opérations énonciatives d'Antoine Culioli, renvoie plus particulièrement à une approche formelle du problème (voir par exemple Culioli 1990). Dans le modèle de Philps, les marqueurs submorphémiques sont en effet considérés comme les traces superficielles d'associations linguistiques entre une séquence phonologique et un sens ou une notion. Philps définit les phonesthèmes ainsi : « sublexical markers as surface traces of general underlying notions » (Philps 2008, p. 128). Ce concept théorique de notion est alors défini ici comme suit : « a conceptual space which may be defined as a set of complex mental representations resulting from the mind's attempts to categorize experience » (Philps 2008, p. 129). Cette définition fait directement écho au concept énonciatif de la notion, en tant que « système complexe de représentation structurant des propriétés physico-culturelles d'ordre cognitif »⁴. L'article mentionne d'ailleurs que la notion dépend des « tentatives d'un individu à catégoriser l'expérience », et poursuit en expliquant les dimensions topologiques du domaine notionnel, en se basant sur Culioli 1990. De la même manière que le modèle de représentation

⁴ Définition de la notion telle qu'elle apparaît sur le site du SIL.

métalinguistique culiolien n'a pas pour objectif de décrire des réalités neurologiques et cognitives, le modèle de Philps n'intègre la notion qu'en tant qu'outil de représentation, afin de révéler des propriétés intrinsèques aux unités submorphémiques. Son modèle s'inspire aussi de l'« espace conceptuel » de de Gärdenfors, dont les paramètres sont représentés par des éléments qualitatifs de type « lumière », « masse », « température », etc. (Philps 2012a). Nous parlerons, enfin, de structure submorphémique pour désigner la forme phonologique du phonesthème et son arrangement phonotactique, dans la lignée de Philps (2002 ; 2008 ; 2012a ; 2012b). Nous éviterons, autant que possible, de référer à des modèles contradictoires, dans un souci de clarté, bien que le choix de cette nomenclature soit en partie orienté.

II. NATURE ET PRISE EN CHARGE DES DONNEES DE TERRAIN

2.1. Support de la psycholinguistique

La réalité psycholinguistique des phonesthèmes a été démontrée par des données expérimentales (Hutchins 1998 ; Abelin 1999 ; Bergen 2004). Les études psycholinguistiques sont à mettre en regard des données statistiques, que nous présenterons dans la section suivante. C'est la thèse de Hutchins qui met en évidence le caractère prometteur des phonesthèmes, du point de vue psychologique, et sur laquelle une grande partie de la littérature ultérieure se base pour déterminer la validité potentielle d'un phonesthème. Au départ, son corpus d'étude comprend un ensemble de 46 phonesthèmes à tester. Son étude est fondée sur trois types d'analyses, suivant chacune un protocole expérimental clairement explicité. La première analyse vise à définir la variabilité de la force de la relation forme-sens au sein des différents phonesthèmes. Elle révèle que la fréquence globale ainsi que la force de l'association phonosymbolique varient toutes deux grandement d'un phonesthème à l'autre, et que la force de l'association a tendance à être inversement proportionnelle à la fréquence lexicale brute de la séquence phonologique en question. La deuxième analyse vise à estimer la productivité potentielle des phonesthèmes dans la langue, en demandant aux sujets d'évaluer la pertinence de définitions attribuées à des mots fictifs, contenant des phonesthèmes. Un groupe de 50 anglophones natifs monolingues a ainsi évalué, sur une échelle de sept points, l'adéquation entre la définition et la forme de mots fictifs, à la fois pour des phonesthèmes anglais et austronésiens. Encore une fois, les résultats montrent que le profil varie beaucoup d'un phonesthème à l'autre. Les mots phonesthétiques sont cependant deux fois plus cités comme fortement liés à leur définition que s'il s'agissait d'une répartition aléatoire. Enfin, la troisième

analyse vise à interroger l'existence d'éléments submorphémiques au niveau du phonème, qui agiraient comme des composants atomiques du phonesthème. Cette dernière expérience n'a pas livré de résultats aussi concluants que les précédentes.

La thèse d'Abelin (1999) propose une approche plus complète, qui prend en compte des données à la fois statistiques et expérimentales. Afin de déterminer quels sont les phonesthèmes les plus employés, et d'estimer la nature de leur relation forme-sens, elle présente une étude lexicologique et psycholinguistique d'attaques consonantiques à l'initiale, de syllabes à la finale, ainsi que de schémas vocaliques (Abelin 1999, p. 1). Deux éléments doivent toutefois attirer notre attention : seuls les éléments « clairement onomatopéiques » ont été sélectionnés pour l'étude lexicologique ; et, bien sûr, cette étude porte sur le suédois, bien que des parallèles avec l'anglais soient envisageables (p. 66). L'analyse lexicologique montre que les phonesthèmes les mieux attestés en suédois sont *sl-*, *sn-*, *kn-*, *kr-*, *kl-*, *sp-*, *st-* et *gl-* (pp. 74-75), que l'on retrouve tous dans le lexique anglais. Deux types d'études expérimentales ont ensuite été menées. La première consiste à présenter des mots suédois évalués comme phonesthétiques à des locuteurs d'autres langues, et de leur demander de deviner leur sens. Cette expérience part du constat que certains phonesthèmes de l'anglais et du suédois semblent très similaires, tandis que d'autres sont très différents, et interroge l'universalité potentielle de la phonesthésie. Les résultats montrent que le sens des unités lexicales les plus onomatopéiques est plus facile à retrouver à partir de la forme, en comparaison des autres types de symbolisme phonétique (p. 210). La deuxième expérience consiste à tester une population de suédophones natifs à partir de mots fictifs. Trois types d'analyse ont ainsi été conduits, en fonction de la modalité de réponse des sujets : le premier test demande au sujet d'inventer un mot construit à partir d'une signification ; dans le deuxième test, le sujet doit choisir, entre différents mots construits, celui qui correspond le mieux au sens donné ; le troisième test impose aux sujets de lier deux sens à deux mots construits (pp. 210-211). Les résultats de cette série d'expériences indiquent que les phonesthèmes du suédois agissent globalement comme des morphèmes liés, puisqu'une partie importante d'entre eux sont aussi bien attestés que ces derniers (p. 257). La mise en regard des données lexicologiques et psycholinguistiques montre une corrélation entre le sens de ces phonesthèmes et le consensus neurolinguistique sur la prépondérance des connections visuelles-auditives et tactiles-auditives dans la construction du lexique. Ces résultats sont particulièrement encourageants, mais mériteraient d'être reproduits et complétés.

Bergen (2004) développe une autre approche, notamment inspirée d'Abelin, et souligne la perfectibilité des expériences de Hutchins. Afin d'accéder au mental inconscience des sujets,

il met en place un protocole leur évitant de faire appel à leurs éventuelles connaissances épilinguistiques, et qui ne leur laisserait pas autant de temps de réflexion. Son protocole part du constat, avancé par les données distributionnelles, que la représentation statistique des phonesthèmes dans le lexique présente un écart significatif par rapport aux prévisions aléatoires. En analysant le comportement d'un échantillon de 20 locuteurs anglophones natifs, Bergen (2004, p. 14) tente de mesurer la portée psycholinguistique des phonesthèmes *gl-*, et *sn-* sur le lexique inconscient des sujets. En utilisant la méthodologie de l'amorçage morphologique⁵ de Kempley et Morton (1982), il cherche à répondre à deux interrogations de départ. Sa méthode consiste à déterminer, d'une part, le comportement lexicologique des mots contenant des phonesthèmes, et d'autre part, dans le cas où un effet positif sur la construction du lexique serait mis en évidence, si cet effet existe pour des mots ne contenant pas de séquence réputée phonesthétique (Bergen 2004, p. 13). En somme, il tente de pondérer les résultats de l'évaluation test, comprenant des attaques consonantiques statistiquement décrites comme phonesthétiques, par la mise en place d'une évaluation contrôle, comprenant des attaques consonantiques *a priori* non-phonesthétiques. Pour ce faire, un mot d'amorçage⁶ suivi d'un mot cible⁷ étaient présentés aux sujets, dans un laps de temps très court. Les sujets devaient ensuite dire si le mot cible était un véritable mot de l'anglais ou non. La manipulation consiste à varier la nature des mots d'amorçage et des mots cibles, et à mesurer le temps de réaction des sujets. En particulier, des mots fictifs contenant des phonesthèmes, et des mots fictifs sans phonesthèmes, furent comparés à des mots réels contenant des phonesthèmes et à des mots réels sans phonesthèmes. Les résultats montrent que la reconnaissance des sujets est plus rapide lorsque le mot d'amorçage et le mot cible partagent un phonesthème, que lorsqu'aucun des deux n'en contient (Bergen 2004, p. 18). De plus, des mots partageant un même phonesthème entraînent des temps de réaction largement plus courts que des mots simplement reliés phonologiquement, mais ne contenant pas de phonesthème connu, ou que des mots sémantiquement proches. Enfin, les phonesthèmes statistiquement attestés comme tels par la littérature distributionnelle entraînent des réactions significativement plus courtes que des attaques consonantiques non réputées comme phonesthétiques. Ces données indiquent que la phonesthésie est un élément important de la reconnaissance lexicale, et que les phonesthèmes font partie intégrante du lexique mental des locuteurs de l'anglais. Encore une fois, ces expériences mériteraient cependant d'être répliquées, car les résultats sont susceptibles de varier

⁵ *Morphological priming methodology*, Bergen 2004, p. 13.

⁶ « Prime word », Bergen 2004, p. 14.

⁷ « Target word », Bergen 2004, p. 14.

d'une population statistique à l'autre, d'autant plus que l'échantillon de 20 individus reste assez restreint. Il serait donc intéressant de reproduire ce type de protocole sur des populations plus larges.

2.2. Linguistique computationnelle et données statistiques

Si l'étude lexicologique d'Abelin (1999) est prometteuse, d'autres recherches statistiques ont été menées depuis, plus particulièrement sur l'anglais. Drellishak (2006) dresse, en premier, une liste très complète de phonesthèmes, qu'il classe du plus au moins statistiquement attesté. Au terme des résultats démontrés par l'expérimentation psycholinguistique, il propose une méthode de validation de séquences phonologiques déjà présentées comme possiblement phonesthétiques par Hutchins (1998). Son approche repose sur le modèle de Bergen (2004), qui postule que les phonesthèmes et les morphèmes ne nécessitent pas d'approches statistiques distinctes, mais qu'une méthode unique peut appréhender les deux phénomènes (Drellishak 2006, p. 13). Sa méthode se présente comme une alternative, moins coûteuse et chronophage, aux protocoles expérimentaux sur lesquels il s'appuie, et considère qu'un lexème testé est démontré comme phonesthétique si, d'une part, l'association son-sens qui le caractérise présente une fréquence plus significative qu'une probabilité aléatoire, et d'autre part, si cette association ne peut s'expliquer par un processus morphologique ou étymologique. Le choix d'écarter les étymons communs peut en effet être intéressant pour estimer la validité statistique d'un phonesthème potentiel, mais il gomme du même coup d'éventuels phénomènes de convergence diachronique. En outre, la distinction entre morphème et phonesthème présentée à la fois par Bergen (2004) et Drellishak (2006) ignore la complexité du statut des unités submorphémiques en théorie morphologique, en avançant le critère de compositionnalité comme systématiquement discriminant (Bergen 2004, p. 4). Les exemples d'éventuels faux positifs cités par Drellishak, notamment les morphèmes *un-* (négation) et *-viv-* (« vie ») font sens, car ils sont effectivement typiques et univoques, mais il ne faut pas oublier qu'un nombre important de morphèmes ne sont pas aussi catégoriques. Alors que le lien de la relation forme-sens symbolique de *un-* est particulièrement bien attestée dans le lexique anglais, ce n'est pas le cas de racines plus ambiguës comme *by-*, qui présente une polysémie singulièrement complexe (Stefanowitsch 2002, p. 13).

À défaut de s'appuyer sur des corpus d'usage, les données produites par Drellishak à partir de ces prérequis théoriques sont de nature lexicologique, puisqu'elles se basent sur le *Webster's Dictionary* (Drellishak 2006, p. 14). En partant de variables orthographiques, il

applique plusieurs méthodes d'analyse de la *Latent Semantic Analysis* (LSA) de Deerwester *et al.* (1990), qui consistent, par différents calculs matriciels, à faire apparaître des similarités sémantiques entre les définitions lexicographiques du dictionnaire choisi. Les mots clés des entrées lexicographiques⁸, ainsi que les mots d'entrées⁹, correspondent aux variables statistiques, que l'on range dans différentes matrices. Selon le type d'analyse appliqué, ces matrices permettent de rassembler des variables similaires (*clustering*), de classer les définitions en fonction de leurs mots clés (*document classification*), de calculer la fréquence de mots clés par définition (*relative word frequency*), ou bien de calculer l'entropie statistique des phonesthèmes potentiels (*mutual information*, MI) (pp. 17-20). Cette dernière méthode permet de déterminer la quantité d'information aléatoire produite par une variable, et, à terme, d'estimer la prévisibilité probabiliste d'une variable, ce qui est effectivement intéressant pour juger si une séquence phonologique est arbitraire ou non. Des filtres ont cependant été appliqués, afin de sélectionner les variables pertinentes (p. 21). Il s'agit surtout d'éliminer les cognats (*etymon filter*), les mots clés contenant le même groupe consonantique testé que le mot d'entrée (*pattern filter*), et les expressions figées récurrentes, utilisées comme descripteurs par les lexicographes (*stopword filter*) (pp. 22-23).

En définitive, seul le calcul de l'entropie statistique (MI) a donné des résultats probants, les phonesthèmes de Hutchins les mieux attestés étant *sn-*, *st-*, *spr-* et *-Vng-* (pp. 26-27). En effet, la *p-value*, c'est-à-dire la probabilité que le résultat soit dû au hasard, a été calculée pour chaque phonesthème identifié par la méthode MI (p. 28), ce qui donne un classement des phonesthèmes les mieux attestés (voir pp. 36-49). En somme, l'intérêt d'une telle étude est son caractère novateur, et son efficacité à pourvoir les chercheurs de listes discriminant les phonesthèmes selon leur validité statistique. Comme énoncé plus haut, la mise à l'écart des données étymologiques représente à la fois un atout et une lacune, selon la perspective du chercheur. Elle permet d'isoler les racines uniquement dues à un phénomène autre que l'étymologie, mais elle passe sous silence les processus de glissement sémantique, qui nous intéressent en l'occurrence.

Nous reviendrons plus tard sur la linguistique historique, mais on peut déjà mentionner quelques études computationnelles dans le domaine de la sémantique diachronique. Dans cette optique, Sagi *et al.* 2009 et Sagi *et al.* 2011 appliquent la LSA afin d'observer des changements sémantiques au fil du temps. L'étude de Sagi *et al.* 2011 aurait tout aussi pu être abordée dans

⁸ « definition words »

⁹ « headwords »

la section suivante sur les méthodes d'analyse de corpus, car son objectif est de tracer des changements sémantiques récurrents à partir de l'analyse statistique de lexèmes dans le Helsinki Corpus. Nous ne détaillerons pas le protocole méthodologique de l'étude, car elle ne concerne pas directement les phonesthèmes, mais elle a son importance dans le sens où ce type d'analyse permet une prise en charge rapide d'un nombre important de données, ce qui peut être très intéressant pour tester la validité des phonesthèmes. Le risque de ce type d'approche réside toutefois dans son manque de précision. En effet, des études portant sur les phonesthèmes comme celles de Sagi *et al.* 2009 ou d'Abramova *et al.* 2013, tout comme celle de Drellishak 2006, ont tendance à renvoyer une image trop générale et figée de leur objet d'étude.

2.3. Études de corpus

Notre progression nous amène au domaine de la linguistique de corpus, qui est définie ainsi par Stefanowitsch (2017, p. 24) : « any form of linguistic inquiry based on data derived from a corpus ». Si le terme de corpus peut être particulièrement polysémique, il consiste, dans ce sens, en « un vaste ensemble de textes authentiques », l'objectif de cette branche de la recherche empirique étant l'analyse de textes représentant un « usage réel de la langue » (Stefanowitsch 2017, p. 24). La linguistique de corpus et la linguistique computationnelle sont en fait imbriquées l'une dans l'autre, car ce sont souvent les méthodes statistiques qui permettent de prendre en charge une quantité de données aussi vaste que suppose une telle définition. Les méthodes utilisées sont variées, mais le principe fondamental est de tenter de repérer des schémas récurrents à travers les corpus analysés, et d'en tirer des données quantitatives (Gries 2009, p. 1226).

Dans la lignée des études expérimentales de Hutchins 1998 et Bergen 2004, l'étude d'Otis et Sagi 2008 reprend la méthode de la LSA, notamment utilisée par Drellishak 2006 dans un cadre lexicologique, et l'applique à un corpus. En utilisant l'outil Infomap, ils génèrent un espace de représentation spatial, qui projette les relations sémantiques entre les mots du corpus, de sorte que des éléments sémantiquement proches sont regroupés ensemble (Otis et Sagi 2008, p. 66). Leur analyse, appliquée au corpus *Projet Gutenberg*, consiste à estimer la probabilité de mots partageant un phonesthème à partager un contenu sémantique similaire (pp. 66-67). Un ensemble de 50 séquences phonologiques a ainsi été testé, dont deux éléments réputés non-phonesthétiques, afin de contrôler la capacité de discrimination du modèle. Sur la liste des 46 phonesthèmes de Hutchins, 27 sont estimés comme particulièrement significatifs, et aucun des deux éléments contrôles n'a montré de corrélation forme-sens, ce qui tend à confirmer la

justesse des estimations. De même que les expériences de Hutchins 1998 avaient mis au jour une corrélation inverse entre la fréquence brute d'un phonesthème dans le lexique et son attestation psycholinguistique, les calculs d'Otis et Sagi 2008 révèlent une corrélation inverse entre la fréquence d'un phonesthème dans le corpus et sa représentabilité statistique dans le modèle (p. 67). Ces résultats sont encourageants, mais les auteurs soulignent bien le fait que les attaques consonantiques plus fréquentes dans le lexique ont tendance à être interprétées comme nécessairement moins bien attestées dans le modèle. Ainsi, l'élément *gr-* figure parmi les plus faiblement attestés (validité de 0,028, Otis et Sagi 2008, p. 69), bien que Drellishak (2006, p. 43) le classe au même niveau que *gl-*, qui présente l'une des meilleures validités statistiques du modèle (0,120). Il est intéressant de noter que *sn-* et *spr-*, qui sont décrits comme *fortement confirmés*¹⁰ par Drellishak (2006, p. 36), ont des indices statistiques effectivement plus élevés que la moyenne (0,080 et 0,121 respectivement). En revanche, *st-*, qui est aussi décrit comme *fortement confirmé* par Drellishak 2006, présente une validité aussi faible que celle de *gr-* (0,028). De même, *-ng* présente une validité plutôt faible (0,035, Otis et Sagi 2008, p. 70), alors que Drellishak 2006 le range aux côtés de *sn-*, *spr-* et *st-*. Les deux modèles semblent donc se répondre, mais pas de manière systématique et uniforme. En sachant que ces deux études se basent sur des matériaux et des méthodologies différents, ces résultats restent globalement satisfaisants.

2.4. Dimension diachronique

Comme nous l'avons vu dans les sections précédentes, la validité des phonesthèmes peut être testée en synchronie, mais elle peut aussi être testée diachroniquement. L'étude de Sagi *et al.* 2009 exploite les méthodes d'analyses distributionnelles sur la synonymie afin de déterminer le contenu sémantique des phonesthèmes, dans une perspective diachronique (p. 35). Leur analyse porte plus précisément sur le marquer *gl-*, et à la racine indo-européenne **ghel* « to shine », qui en serait à l'origine (p. 36). Leur postulat de départ est que la relation forme-sens qui caractérise les phonesthèmes peut être de deux natures : elle peut s'expliquer par la racine historique des mots de la métaclasse du phonesthème, ou bien elle peut être de nature cognitive. Il est important de souligner que les deux explications sont compatibles : certains phonesthèmes seraient plutôt d'origine étymologique, tandis que d'autres auraient une véritable influence cognitive sur le lexique mental des locuteurs. L'article propose une méthode de distinction de ces deux origines potentielles, en partant de la même méthodologie qu'Otis et

¹⁰ « Strongly confirmed ».

Sagi 2008, et qui consiste à comparer les espaces sémantiques de la racine **ghel* et des mots en *gl-* sur Infomap, dans le corpus *Projet Gutenberg*. Le protocole distingue deux regroupements principaux : les mots qui descendent de la racine **ghel* et les mots contenant le phonesthème *gl-*. Sont ensuite distingués : les descendants de **ghel* qui commencent par *gl-*, les descendants de **ghel* qui ne commencent pas par *gl-*, et les mots contenant *gl-* qui ne descendent pas de **ghel*. En se basant sur le niveau de cohésion des regroupements, défini par la proportion de co-occurrences des mots de chaque ensemble au sein du modèle, les résultats indiquent que le regroupement le plus cohésif est celui des mots provenant de **ghel* et qui commencent par *gl-* (p. 38), mais les valeurs statistiques de tous les regroupements sont en fait supérieures à la norme de prévisibilité. Autrement dit, la cohésion des différents regroupements n'est pas due au hasard. Ces données confirment donc la validité de la relation entre la notion de vision/lumière et le marqueur *gl-* et elles montrent que le phonesthème *gl-* a joué un rôle dans la cohésion de la racine commune, ainsi que dans les mots non-affiliés à cette racine (p. 39). Cette méthode concluante mériterait d'être appliquée à d'autres phonesthèmes, y compris à ceux qui sont issus de plusieurs racines historiques.

La présente étude ne vise pas prioritairement à estimer la validité statistique du phonesthème *gr-*, qui a déjà été évaluée (Drellishak 2006, p. 43 ; Otis et Sagi 2008, p. 69 ; Abramova *et al.* 2013, p. 1698), mais à déterminer son contenu phonosémantique. Nous reviendrons sur la méthodologie de l'étude dans la prochaine partie, mais le principe général est de repérer des schémas récurrents dans le comportement morphosémantique diachronique de *gr-*, à travers une analyse lexicographique dans l'*Oxford English Dictionary*. Dans la lignée des travaux de Drellishak 2006 sur la recherche statistique en corpus lexicographique, mais surtout de Sagi *et al.* 2011 sur le changement sémantique, ainsi que de Flaksman (2016 et 2017) et Smith (2016 et 2020) sur l'iconicité en diachronie, notre objectif sera de dégager le comportement sémantique de *gr-* à travers le temps. Nos analyses, qui seront circonscrites à l'échelle chronologique déterminée par l'OED, tenteront de déterminer le rôle de l'étymologie dans la formation et l'évolution de la métaclasse des mots en *gr-*. La mise au jour de phénomènes récurrents non-explicables par la présence d'éventuelles racines communes sera interprétée comme un indice fort de réanalyse cognitive par les locuteurs (voir Anttila 1977 et Monneret 2020 sur l'analogie cognitive). Aucune preuve absolue de la présence effective d'un tel mécanisme ne pourra toutefois être apportée de manière catégorique, car l'interprétation de ces résultats sera limitée par la nature de notre corpus ; c'est pourquoi nous préférons parler d'indice fort. À l'inverse, dans le cas où l'étymologie s'avérerait être à l'origine de la plupart

des relations forme-sens au sein de la métaclasse des mots en *gr-*, l'hypothèse d'une réanalyse cognitive sera jugée improbable. On jugerait alors l'influence de *gr-* sur le lexique mental comme trop ténue pour être raisonnablement posée comme probable.

Partie II : Méthodologie de la recherche

I. L'OXFORD ENGLISH DICTIONARY COMME OUTIL DE RECHERCHE

3.1. Sources et spécificités de l'OED

L'*Oxford English Dictionary*, en tant que source principale sur laquelle se base notre analyse, se doit d'être présenté dans le détail, d'autant qu'il comporte à la fois des avantages et des inconvénients. L'atout majeur de ce dictionnaire est qu'il contient des informations sur l'histoire de la langue, ce qui nous concerne directement, étant donné la portée diachronique de notre étude. Les données lexicologiques qui peuvent être extraites couvrent une grande période chronologique, du VIII^e siècle à aujourd'hui. L'histoire de cette recherche lexicographique remonte au début du siècle dernier, et a permis la collecte d'une quantité impressionnante de renseignements ayant trait à la langue anglaise depuis ses premières attestations écrites jusqu'à nos jours. De plus, en peu de temps, une quantité importante de données peut être traitée à partir d'une seule source documentaire, ce qui en fait un outil particulièrement efficace. Nous aurions pu utiliser d'autres dictionnaires, comme le *Gutenberg Dictionary* ou le *Webster Dictionary* par exemple, qui est utilisé par Drellishak 2006. Dans la lignée de Smith 2016, c'est toutefois l'OED *subscription version 3rd edition* qui fut retenu afin de collecter les données brutes, en raison de la richesse de ses entrées, qui stipulent les dates des attestations écrites successives, en fonction des parties du discours ainsi que des différents sens (Smith 2016, p. 169). Précisons également que l'OED est régulièrement utilisé comme instrument de recherche dans le domaine de la linguistique historique et dans le cadre d'études portant sur les changements lexicosémantiques (Allan 2012, p. 17).

En revanche, l'un des défauts majeurs de l'OED est que, si les premières attestations écrites des entrées lexicales sont indiquées, il n'en est rien de la sortie d'usage de ces lexèmes. D'une certaine manière, cette faiblesse est assez logique, car il est quasiment impossible de détecter à quel moment précis un mot disparaît de l'usage. Cette disparition résulte avant tout de la baisse de la fréquence d'usage d'un mot, qui est progressive. Par ailleurs, il peut se passer un certain temps entre le moment où un mot n'est globalement plus utilisé, ou en tout cas bien moins qu'avant, et le moment où l'on s'en aperçoit. Ce problème n'est donc pas évident à résoudre, d'autant que les éditeurs de la première édition n'ont, sur le coup, pas réalisé l'importance d'inclure ces dates terminales. Encore une fois, cela peut se comprendre dans le

sens où c'est surtout avec l'émergence et le développement de la sémantique diachronique que ce type de question a commencé à être abordé. Même si les dernières éditions ont tenté de rectifier le tir, les sources textuelles restent, pour la plupart, non-éditées, et sont souvent très difficiles à dater, sans compter que certaines ne sont pas accessibles (Allan 2012, p. 24). La deuxième lacune méthodologique de l'OED concerne le choix des mots à l'intérieur des définitions. En effet, les entrées ne bénéficient pas d'un cadre lexicographique systématique quant aux mots qu'il convient d'utiliser ou non, ce qui autorise chaque professionnel à adopter ses propres mots clés. Ce reproche peut cependant être adressé à n'importe quel dictionnaire. Le corolaire de cette observation est que notre recherche de traits subsémantiques, basée sur des mots clés, est susceptible de s'appuyer sur des choix lexicaux qui ne sont pas nécessairement homogènes et uniformes. Il faut reconnaître cette part intuitive dans la nature de notre documentation, et la garder en tête lors de l'interprétation de nos résultats.

Si une certaine subjectivité doit être reconnue, à la fois de la part du lexicographe et du linguiste, il ne faut pas oublier que la base de données de l'OED repose sur des citations issues de documents d'origine, qui peuvent être considérées comme une forme de corpus (Stefanowitsch 2017, p. 22). La première édition de l'OED comprenait déjà environ 1,8 millions de citations (Hoffmann 2004, p. 18). Cet aspect « réel » de l'usage linguistique en contexte pallie, en partie, les défauts soulevés jusqu'ici. Dans tous les cas, les précautions qu'il s'agit de prendre ont été jugées acceptables dans le cadre de la présente étude. Gardons simplement à l'esprit que, en ce qui concerne les problèmes de datation, si les attestations écrites dont il est fait mention indiquent l'amorce de l'emploi d'un lexème, « l'absence d'attestation ne peut être interprétée comme la preuve d'une absence d'emploi » (Allan 2012, p. 25). En effet, il est fondamentalement impossible de démontrer qu'un lexème arrête d'être utilisée à un instant *t*, précisément parce qu'on ne peut prouver que l'existence d'un phénomène, et non son inexistence. C'est aussi la raison pour laquelle, si des dates de fin d'utilisation étaient proposées dans des versions futures de l'OED, celles-ci devraient être prises en considération avec esprit critique, étant donné qu'elles ne seraient qu'un indice du déclin progressif de textes attestant de l'emploi des lexèmes en question.

Un dernier mot sur la hiérarchisation des sens dans l'OED. Les entrées sont hiérarchisées à l'aide de chiffres et de lettres, quelles que soient les parties du discours. Chaque entrée est numérotée par lexème, de sorte que tous les homonymes d'un lexème sont accessibles après recherche. Ensuite, tous les glissements sémantiques de chaque lexème homonyme sont inventoriés dans leurs dossiers correspondants. Dans notre notation, un monomorphème comme

graze sera systématiquement suivi de sa nature grammaticale, de son numéro de lexème, et de son numéro de sens. Ainsi, *graze*, v.1 (1a) et *graze*, v.2 (1a) sont deux homonymes, tandis que *graze*, v.1 (1a) et *graze*, v.1 (2b) sont deux sens du même monomorphème *graze*, v.1. La lettre en petit caractère renvoie à une hiérarchisation en sous-sens. Nous ne rentrerons pas plus dans le détail, mais certains mots très polysémiques présentent un très grand nombre de sous-catégories, comme *great* par exemple.

3.2. Aspects chronologiques et perspectives diachroniques

L'objectif principal derrière toute exploration diachronique est d'observer et d'analyser les données, à la lumière des indications chronologiques dont on dispose, afin de repérer des schémas dans les changements évolutifs à l'œuvre. Ces changements peuvent être de nature morphologique ou sémantique ; nous serons ici moins intéressés par les premiers que par les seconds. Les transformations sémantiques peuvent elles-mêmes être de natures diverses, en sachant qu'un bon nombre d'entre elles est de type métonymique ou métaphorique. Nous utiliserons régulièrement le terme de métasémique pour désigner tout type de modification sémantique qui survient au cours du temps. Notre étude se présente comme une tentative de dresser le profil sémantique d'un phonesthème, à partir de l'analyse d'un corpus de monomorphèmes, et de mettre au jour les liens entre ce phonesthème en particulier et les traits sémantiques qui lui sont associés. La méthode globale s'inscrit dans la lignée des travaux de Smith (2016 et 2020) sur les mots en *fl-*, dans le sens où nous proposons de déterminer des schémas métasémiques, en nous appuyant sur des fréquences statistiques, afin d'analyser le profil diachronique des monomorphèmes en *gr-*. La finalité de cette analyse est de retracer les différentes origines étymologiques du phonesthème au cours du temps, et de rechercher des indices de schémas diachroniques d'associations phonosymboliques. En somme, de la même manière que les analyses en sémantique cognitive cherchent à reconstituer l'évolution des lexèmes, notamment grâce aux liens « entre les sens littéraux de lexèmes polysémiques, et leurs sens métaphoriques ou métonymiques », la présente étude tentera d'analyser, du moins en partie, « les mécanismes cognitifs sous-jacents qui engendrent une cartographie mentale entre les domaines conceptuels, et au sein de ceux-ci » (Allan 2012, p. 22), et ce au niveau submorphémique. L'évaluation de la nature et de l'orientation de tels changements diachroniques s'avère cependant délicate, surtout lorsque l'on s'en tient à une documentation uniquement lexicographique.

3.3. Collecte des données

Dans la mesure où aucune base donnée d'information lexicale diachronique n'est, à ce jour, disponible, la collecte des données brutes a été effectuée manuellement à partir des entrées de l'OED. Les mots contenant l'attaque consonantique *gr-* ont donc été sélectionnés à la main, selon un certain nombre de règles. Soulignons d'emblée que cet ensemble de règles peut être considéré comme un biais méthodologique étant donné que tous les mots en *gr-* n'ont pas été conservés pour la construction du corpus d'étude. Les règles en question sont toutefois loin d'être arbitraires, et l'ensemble des lexèmes en *gr-* de l'OED a bénéficié d'une forme succincte d'évaluation. D'abord, seuls les monomorphèmes ont été conservés, c'est-à-dire les unités lexicales qui ne peuvent pas être découpées en de plus petites unités minimales porteuses de sens selon le modèle compositionnel traditionnel. Cette précaution évite les redondances dues aux processus de morphogénèse plus ou moins complexes, comme la dérivation ou les mots composés. Ainsi, *grasp* a été conservé, mais *graspable*, *grasp-all*, *grasper* et *grasping* ont été délaissés. Les amalgames, qui peuvent être intéressants pour étudier la formation du lexique phonesthétique d'un point de vue cognitif (voir Smith 2014 pour la formation des amalgames cognitifs), et qui sont, de toute manière, assez difficiles à repérer, ont été gardés. De même, les mots équipés de pseudo-suffixes, c'est-à-dire des formes qui ont été suffixées à un état antérieur de l'histoire de la langue, mais qui n'agissent plus comme de véritables suffixes aujourd'hui, ont été sauvegardés. Le pseudo-suffixe *-le* est, par exemple, assez courant : *grabble* < *grab* ; *grapple* < *grape* ; *griptide* < *grip* ; *griddle* < *grid*. Notons au passage que la formation de ces lexèmes n'est pas toujours claire. L'OED indique par exemple que *grasple*, v. [1553] serait une variante de *grapple*, v., influencée par *grasp*, v., ce qui sous-entend qu'elle ne proviendrait pas directement de *grasp*, v., sans que l'on en sache davantage à son sujet. Les formes en *-el(l)* ont également été conservées : *grummel* < *grum* ; *gravell* < *grave* ; *grundel* < *grund* (OE). Notons que la plupart des mots en *-el(l)* n'est d'ailleurs pas nécessairement d'origine pseudo-suffixale, comme *grampell*, n. [1568], qui pourrait provenir du français *grampelle*, ou bien *grapnel*, qui vient de l'anglo-normand **grapnel*, lui-même un diminutif de *grapon*, cognat du français *grappin*. Il en va de même pour les mots en *-et*, qui peuvent s'apparenter à des pseudo-suffixes, comme *grummet* < *grum*, mais pas systématiquement, comme l'indique *groset*, n. [1686], déformation de *groser*, n. [1548]. Le suffixe *-er* est d'ailleurs, lui aussi, inclus dans notre corpus, bien que, en l'occurrence, il ne semble pas s'agir du même type de formation.

À l'issue du travail d'analyse effectué en M1, un manque de transparence en ce qui concerne les critères d'inclusion a été soulevé. La sélection des monomorphèmes qui seront

inclus dans l'analyse est en effet importante, car le fait de présélectionner certains mots plutôt que d'autres peut, dans l'absolu, avoir des implications statistiques. Pour commencer, tous les monomorphèmes qui renvoient de près ou de loin aux traits sémantiques décrits dans la littérature, et associés à la marque *gr-*, ont été conservés. Autrement dit, les données quantitatives que l'on pourra tirer de notre analyse de corpus n'auront pas été orientées par des choix arbitraires au sein des lexèmes *a priori* phonesthétiques. Bien sûr, il est possible qu'une partie de ces monomorphèmes s'avère non-phonesthétique, et c'est justement tout l'enjeu de la partie analytique de l'étude de distinguer les lexèmes pertinents de ceux qui ne le sont pas. Il n'en reste pas moins qu'aucune sélection *a priori* n'a été appliquée de ce côté-là. En revanche, les lexèmes en *gr-* disponibles dans l'OED n'ayant aucun rapport avec les traits sémantiques véhiculés par l'état de l'art n'ont pas tous été retenus. Cela peut poser problème dans l'objectif de révéler d'éventuels traits phonesthétiques que la littérature n'aurait pas soulevé, car ceux-ci resteraient alors inconnus. Nous avons choisi de nous appuyer sur les analyses et impressions des linguistes s'étant intéressés de près ou de loin au sujet, et nous avons pris parti de ne pas prendre en compte les unités lexicales qui sortent totalement de ce cadre. On remarquera que ce parti pris sélectif n'est pas susceptible d'affecter l'orientation de nos résultats, en ce sens que le but de notre étude n'est pas d'évaluer le potentiel phonesthétique global de *gr-* au sein de l'ensemble des lexèmes en *gr-* qui ont pu être renseignés dans l'histoire de l'OED, mais de considérer le poids statistique de tel ou tel trait phonesthétique par rapport à l'ensemble des lexèmes phonesthétiques de *gr-*. Nos calculs s'appliquent, de ce point de vue, à un ensemble clairement défini par le protocole que nous avons établi, et leur portée est uniquement interne à cet ensemble. Une étude qui aurait pour objectif de calculer la force statistique brute de tel ou tel trait par rapport à l'ensemble des lexèmes en *gr-*, qu'ils soient phonesthétiques ou non, serait contrainte d'incorporer tous les éléments de cet ensemble dans son corpus, mais ce n'est pas le cas de la présente étude. Ajoutons enfin que parmi les monomorphèmes du corpus d'analyse définitif, un certain nombre d'entre eux ne correspond pas nécessairement aux indications fournies par la littérature sur les phonesthèmes, mais s'appuient sur les analyses préliminaires qui ont été conduites au terme du projet de M1.

Si l'ensemble des lexèmes de l'OED a été succinctement évalué, afin de déterminer s'ils rentraient dans les critères présentés plus haut ou non, seule une partie d'entre eux a donc été incluse dans le corpus final. En ce qui concerne les homonymes, nous aurons l'occasion d'éclaircir la situation dans des sections ultérieures, notamment lorsque nous nous concentrerons sur l'analyse des données. Précisons simplement, pour le moment, que les

homonymes ont été retenus, et explicitement distingués les uns des autres comme des entrées bien disjointes, encore une fois dans la lignée de Smith 2016. Nous avons choisi de nous écarter du protocole de sélection préconisé par Drellishak 2006, qui met de côté les racines communes. L'évaluation du rôle de l'étymologie dans les processus évolutifs qui ont vu naître les phonesthèmes est justement l'un des axes principaux de notre recherche, ce qui motive ce choix. En effet, c'est précisément en examinant les dynamiques de divergence et de convergence à partir des racines d'origine, ainsi que les processus d'amalgame et les changements phonologiques et morphologiques à l'œuvre, qu'il nous sera possible de proposer un modèle de profil métasémique. Une fois de plus, l'objectif de notre analyse n'étant pas de déterminer la validité du phonesthème *gr-* par des calculs statistiques, mais d'étudier comment les sphères phononotionnelles s'influencent entre elles au sein du corpus, l'inclusion de mots de même origine étymologique ne pose pas de problème méthodologique, bien au contraire.

Les monomorphèmes désignés dans l'OED comme *obsolete* ont été conservés. Comme nous l'avons déjà mentionné, le dictionnaire renseigne les dates des premières attestations écrites des lexèmes, ou du moins celles qui ont été retrouvées et conservées, mais ne donne aucune information sur leur déclin, tant et si bien qu'il est impossible de connaître la durée de vie d'un lexème. Contrairement au travail de M1, nous avons fait le choix de prendre en compte, cette fois-ci, les monomorphèmes dialectaux, qualifiés dans l'OED sous différentes appellations. La nomenclature est relativement fixe, mais la nuance entre *regional* et *dialect* n'est pas évidente. Lorsque les régionalismes sont circonscrits à une zone plus précise, celle-ci est explicitée. On s'aperçoit ainsi de la grande quantité de termes employés en Écosse et dans le nord de l'Angleterre, qui ne sont généralement pas utilisés ailleurs. Cet état de fait est dû à un aspect de l'histoire de la langue que nous ne développerons pas, mais il est bon de garder en tête qu'une langue ou une zone d'étude ne sont pas des ensembles homogènes, et que tous les locuteurs d'une même communauté linguistique ne disposent pas nécessairement du même lexique mental. Nous aurons l'occasion de revenir sur la nature exacte du corpus dans la partie analyse, où nous exposerons les détails quantitatifs de sa composition. Soulignons simplement que la nomenclature n'est pas toujours évidente à interpréter, notamment parce que les indications *obsolete* et *dialect* vont souvent de pair, ce qui paraît contre-intuitif. On aurait en effet tendance à penser qu'un terme dialectal n'est, par définition, pas obsolète puisqu'il est encore utilisé. Les lexicographes ont probablement voulu dire que tel terme n'était plus utilisé dans la langue standard, mais désormais restreint au dialecte, la distinction avec un terme

purement dialectal étant que ce dernier n'est jamais apparu dans le standard. Ce n'est cependant qu'une supposition, et il est impossible de se prononcer à la place des lexicographes.

Le corpus comprend 227 monomorphèmes, pour un ensemble de 497 sens, chaque monomorphème pouvant comporter un nombre important de sens. La section analyse comprend deux types d'approche, l'une basée sur les proto-sens, c'est-à-dire qu'elle ne prend en compte que les sens premiers, ou du moins ceux désignés comme tels dans l'OED, tandis que l'autre prend en considération tous les sens des monomorphèmes, y compris les sens métasémiques. La collecte des données a été effectuée dans un tableur Excel, sous la forme d'un tableau de calcul, dans laquelle chaque colonne renseigne un type d'information donné par l'OED : le lexème d'entrée, la partie du discours, la transitivité des verbes, la définition, la date d'attestation, et le type d'usage. Ce dernier paramètre englobe des informations à la fois dialectales et liées à la fréquence d'emploi. On y trouvera des commentaires comme : *in use*, *obsolete*, *dialect*, *Scottish*, *Northern*, *regional*, *rare*, etc. Enfin, la dernière colonne renseigne les informations étymologiques. On y trouve les cognats du lexème en question, à la fois dans les langues actuelles apparentées, ainsi que dans des langues plus anciennes. Lorsque l'étymon est connu, qu'il soit attesté par les sources textuelles ou reconstruit par méthode comparative, il est indiqué, à la suite de l'ensemble des descendants qu'il a engendrés, ou des descendants de ses cognats les plus proches. Lorsqu'on a affaire à un emprunt, la langue d'origine est indiquée, si elle est connue. De manière générale, l'historique complet du mot est spécifié, ce qui nous donne accès à un compte rendu de son évolution au cours du temps.

Concernant les chiffres généraux, on compte 227 proto-sens, c'est-à-dire autant que de monomorphèmes, puisqu'à un proto-sens correspond un et un seul monomorphème, et qu'à un monomorphème correspond un et un seul proto-sens. Même si ce chiffre n'entrera pas dans nos calculs, on compte un nombre total de 270 sens métasémiques. Les conversions sont au nombre de 220 dans l'ensemble du corpus, mais sont réparties entre les proto-sens et les sens métasémiques.

II. PRESENTATION DU CORPUS

4.1. Le phonesthème *gr-* dans la littérature

Le marqueur submorphémique qui fait l'objet de la présente étude a été choisi à partir d'un ensemble de listes de phonesthèmes statistiquement attestés, essentiellement tiré des travaux fondateurs de Hutchins 1998, qui ont engagé la discussion sur la nature des phonesthèmes et leur potentielle quantification. Drellishak 2006 et Otis et Sagi 2008 ont ensuite repris ces travaux dans leurs propres articles, en y apportant des nuances et des compléments. Le choix de *gr-* relève avant tout de l'intuition et ne répond pas à une stratégie de recherche particulière, en partie parce que tous les phonesthèmes, du moins ceux présentés comme tels, sont dignes d'intérêt, et qu'une étude approfondie et exhaustive serait, de toute manière, désirable. Le marqueur *gr-* n'est effectivement pas particulièrement mieux attesté que les autres, puisqu'il est classé parmi les phonesthèmes *faiblement confirmés*¹¹ par Drellishak (2006, p. 43). On peut cependant avancer que les données statistiques brutes nécessitent une prise en charge interprétative, et qu'elles ne peuvent pas, à elles seules, définir un marqueur submorphémique dans sa globalité. Ajoutons encore que de nombreux autres paramètres sont à prendre en compte dans l'équation, que nous prendrons le temps de développer par la suite. Soulignons pour le moment que les calculs avancés jusqu'ici à travers la littérature prennent généralement en compte l'ensemble des lexèmes comprenant la séquence phonologique suspectée d'être phonesthétique. Or, il a été montré par Otis et Sagi 2008 que chaque phonesthème présentait ses propres caractéristiques, dans le sens où certains d'entre eux sont plus diffus que d'autres. Il semble que les larges assemblages comme *gr-* ont tendance à contenir de petits ensembles de lexèmes, qui partagent une forte cohérence phonesthétique entre eux, mais qui sont noyés au sein d'un ensemble globalement peu cohérent, ce qui a pour effet de baisser leur représentabilité statistique générale. Il est par ailleurs utile de rappeler que les chiffres varient d'une étude à l'autre, en grande partie parce qu'elles ne partent pas des mêmes données. En effet, Abramova *et al.* 2013 travaille à partir d'un corpus de 197 occurrences pour le phonesthème *gr-*, alors qu'Otis et Sagi 2008 se contente d'un échantillon de 66 occurrences. Sans surprise, il en ressort que les résultats obtenus par les deux études sont légèrement différents. Précisons que les chiffres exacts ne sont pas mentionnés ici dans un intérêt analytique, mais simplement pour donner une idée de l'écart entre les études. Ce type de donnée statistique quantitative ne sera

¹¹ « Weakly confirmed », Drellishak 2006, p. 43.

de toute manière pas détaillée, attendu que ce n'est pas le cœur de notre propos. On peut même considérer que notre objectif est de dépasser le cadre générique de ces analyses, qui, si elles représentent des enquêtes préalables de première importance, ont tendance à ne renvoyer qu'une image très générale et quelque peu abusive de ce qu'est réellement la phonesthésie. L'intérêt principal, sur lequel nous nous porterons, concerne avant tout l'identité sémiotique des structures phonosymboliques, ce qui s'oppose généralement à ce qu'indiquent les données statistiques brutes. On notera enfin que la longévité des lexèmes, paramètre important mais souvent négligé, doit être pris en compte afin de jauger le poids de telle ou telle notion. De plus, il est bon de noter que certains phonesthèmes qui semblent fortement ancrés dans le lexique sont finalement faiblement attestés, comme *sl-*, bien qu'il soit régulièrement cité comme un exemple typique. Enfin, il est bon de rappeler que les analyses statistiques auxquelles nous avons fait référence s'appuient sur une logique uniquement lexicographique et quantitative, sans prendre en compte les fréquences d'emploi des lexèmes. Or, l'une des hypothèses avancées sur l'origine des phonesthèmes est qu'ils émergent de l'usage des locuteurs. Blust (2003, pp. 199-200) parle par exemple d'« effet boule de neige »¹² pour illustrer l'émergence progressive des phonesthèmes dans le lexique. Une étude qui cherche à saisir la valeur statistique réelle des phonesthèmes à travers le lexique, quelle que soit la langue en question, ne peut se limiter à une analyse quantitative, qui place toutes les unités du lexique sur le même plan. Si les aspects plus qualitatifs de la présente étude sont effectivement limités, les perspectives qu'offrirait la sémantique distributionnelle et l'apport d'outils tels que Sketch Engine sont encourageantes.

En ce qui concerne la description de *gr-* dans la littérature, les analyses statistiques autant que les ouvrages généraux ou thématiques sont généralement laconiques à son sujet, lorsqu'il est mentionné. Ce manque d'intérêt est probablement à mettre sur le compte de sa faible représentabilité globale et la grande quantité de lexèmes concernés, comme expliqué plus haut. Nous tenterons cependant de présenter un compte rendu de l'état de l'art sur la question, en sachant que les indices qui suivent seront pris en compte lors de l'interprétation des données empiriques, tirées de l'OED. Waugh mentionne que *gr-* « comprend trois phonesthèmes » en un. Elle distingue les domaines suivants : quelque chose de déplaisant, que l'on observe dans les lexèmes *grim*, *grisly*, *gritty*, *gruesome*, *gruff* ou *grumpy* ; une expression de plainte, comme dans *grumble*, *groan*, *grunt*, *grieve*, *grudge*, *gripe*, et même *disgruntled* ; et un frottement indésirable, comme dans *grind*, *grate*, *grovel* ou *grub* (Waugh 1994, p. 59). Philips 2008 parle, quant à lui, d'une notion de « préhension », à laquelle cette unité submorphémique serait

¹² « Snowballing effect ».

associée, et cite les exemples de *grab*, *grapple*, *grip*, *grope* et *grub*. De la même manière que *(s)n-* est décrit comme la trace matérielle d'une valeur notionnelle de « nasalité », que Philips note sous la forme $\{(s)n-, \text{nasalité}\}$, le marqueur *gr-* est noté $\{gr-, \text{préhension}\}$ (Philips 2008, p. 128). Abramova, à son tour, associe plutôt *gr-* à un « bruit menaçant, de la colère et de l'empoigne »¹³ (Abramova et al. 2013, p. 1698). Ces commentaires véhiculent globalement une idée de sensation ou de sentiment négatif, qui transparait sous forme de son ou de frottement désagréable, ou bien de mauvaise humeur. Les exemples cités par Philips 2008 pour illustrer la notion de préhension ont en fait, eux aussi, une connotation péjorative, du moins en ce qui concerne une partie de leurs emplois. Les termes *grope* et *grub*, en particulier, ont tendance à être dépréciatifs. Allan se montre plus spécifique en parlant d'une connotation « désapprobatrice », sans plus de développement (2001, p. 136).

Il est possible de rapprocher *gr-* d'autres phonesthèmes, qui présentent une structure submorphémique comparable. Les enquêtes empiriques de Wright, en particulier, portant sur la distribution régionale des phonesthèmes à travers l'Angleterre, montrent que les attaques consonantiques /kr/, /skr/ et /gr/ sont associées à une activité de « manger de manière bruyante et vorace »¹⁴ (Wright 2012, p. 3). Or on remarque que tous les trois partagent des points communs phonologiques marquants. La séquence /kr/ correspond phonétiquement au même point d'articulation que /gr/, puisque seul le critère de voisement de la consonne vélaire permet de distinguer les deux. Elle est décrite comme associée à une notion de discordance, à la rudesse ou à une sensation de grattement¹⁵, ce qui fait parfaitement écho aux observations de Waugh, qui associe *gr-* à une sensation ou un bruit de frottement ou de friction. Ces éléments de réponse ne s'éloignent pas non plus de ce que Bloomfield décrit comme une sensation de choc bruyant, qui serait contenu dans *cr-* (1933, p. 245). L'étude de Wright rapporte que des mots tels que *granch*, *graunch*, *graunt*, *graze*, *gresh* et *grind* sont liés à une sensation d'écrasement ou de broiement¹⁶, tandis que *growdge* et *growze* véhiculent plutôt une impression de « glouglou »¹⁷. Notons au passage qu'il est plus difficile de décrire ces unités phonesthétiques en français, car la langue ne contient pas les mêmes 'impressions' ou 'sensations' phonosymboliques que l'anglais. Soulignons par ailleurs qu'un bon nombre des lexèmes mentionnés dans cet article n'apparaissent pas dans notre corpus, étant donné qu'il s'agit de régionalismes. La source

¹³ « Threatening noise, anger, grip ».

¹⁴ « Eating noisily and greedily ».

¹⁵ « jarring, harsh, grating »

¹⁶ Wright 2012 parle d'une sensation de « crunch ».

¹⁷ Wright 2012 parle cette fois-ci de « gobble ».

utilisée pour ces analyses est l'ouvrage *Survey of English Dialects: The Dictionary and Grammar* (Upton *et al.* 1994).

4.2. Structure submorphémique de *gr-*

Cette section a pour but de présenter quelques pistes concernant la structure submorphémique du phonesthème, c'est-à-dire l'agencement des éléments phonologiques qui le composent et les éventuels composants subsémantiques qui leur sont associés. Il s'avère que les marqueurs sublexicaux, comme les définit Philps 2008, entrent dans un cadre phonologique spécifique, de sorte que toutes les combinaisons ne sont pas autorisées. Les combinaisons dont il est question ici vont au-delà des structures phonotactiques générales de la langue, et concernent l'association de phonèmes indépendants, qui seraient susceptibles de porter de l'information phononotionnelle. On pourrait presque parler de subphonotaxe, en gardant à l'esprit que la forme ne va jamais sans le sens. Lorsque l'on observe les différents phonesthèmes à travers le lexique de l'anglais, deux types de structures phonologiques peuvent être distinguées : nous aurons affaire soit à une structure SC- ou à une structure CR- (Philps 2008, p. 127). La lettre C désigne une consonne de n'importe quel type, tandis que S renvoie au phonème /s/, et le R représente une consonne liquide, c'est-à-dire une vibrante ou une approximante, généralement /r/ ou /l/. La partie non-phonesthétique du lexème suit, le plus souvent, un gabarit de type -V(C), c'est-à-dire une voyelle accompagnée d'une éventuelle coda consonantique. En somme, la plupart des phonesthèmes de l'anglais sont formés d'un groupe consonantique à l'attaque, qui peut être composé d'un /s/ en première position, suivi d'une seconde consonne, quelles que soient ses caractéristiques, ou bien d'une consonne quelconque suivie d'une consonne liquide. Ce groupe consonantique initial précède une voyelle, qui peut être longue ou courte. Il est à noter que si la racine se termine en coda, la voyelle qui se trouve devant a de plus grandes chances d'être courte, tandis qu'une voyelle longue ou une diphtongue a moins de chance d'être suivie d'une coda. Ce modèle de base a été développé par Philps comme un outil pour expliquer les phénomènes de *s-* initiaux facultatifs que l'on peut observer dans les paires de mots comme *sneeze* et *neeze*, *sniff* et *niff* ou *snag* et *nag*. Ces mots, qui contiennent tous le marqueur *sn-*, ne semblent pas forcément l'exprimer sous sa forme SC-complète. Précisons tout de même que la plupart des doublons sans *s-* sont dialectaux ou informels. Contrairement au modèle de base, ces mots n'ont pas pour invariant l'élément *s-*, qui est inactif, mais l'élément *n-*, la combinaison des deux gabarits étant notée sous la forme SC-/C-. On parle alors d'alternation phonosémantique ou phononotionnelle.

Avant de passer à la structure des mots en *gr-*, tentons de résumer les aspects principaux avancés par Philips en vue d'expliquer la structure de *gn-*. L'intérêt majeur d'une comparaison entre les deux attaques consonantiques est qu'elles présentent toutes deux l'élément *g-*. Tout d'abord, Philips 2008 défend l'idée que le phénomène de consonne facultative que l'on trouve chez les mots en *sn-* n'est pas applicable à l'étude de la séquence *gn-*, car même si l'élément *g-* n'est effectivement pas prononcé, l'évolution de ce marqueur depuis le vieil anglais *gn-* à l'anglais contemporain $\emptyset n-$ ne se confond pas avec l'évolution de *sn-* en $\emptyset n-$. Le problème que pose la forme *gn-* est qu'elle ne peut pas non plus être considérée comme une simple structure SC-/C-, dans la mesure où son développement phonologique depuis un groupe consonantique dans les premiers états de la langue anglaise, en une attaque monoconsonantique dans ses états les plus récents, fait que sa forme actuelle ne peut être analysée selon le gabarit habituel. Tout dépend en fait de la dimension que l'on souhaite adopter : une perspective diachronique permet de comprendre les mécanismes évolutifs à l'œuvre derrière les formes que l'on observe à un instant *t*. Le marqueur *gn-* semble toutefois être le seul phonesthème qui pose ce type de problème formel, puisque les autres unités phononotionnelles de l'anglais peuvent s'analyser selon les gabarits énoncés plus haut. Par ailleurs, même si son histoire paraît quelque peu exotique dans le paysage phonesthétique de l'anglais, la séquence *gn-* partage des propriétés importantes avec la structure CR-. La structure phonotactique, pour l'appeler ainsi, de *gn-*, peut s'exprimer ainsi : G(V)NVC(-). La présence d'un potentiel élément vocalique entre G et N, noté (V), s'explique par le fait qu'un certain nombre de phonesthèmes est caractérisé par une structure potentiellement disruptive, de sorte que l'unité lexicale porteuse du phonesthème ne présente pas le groupe consonantique que l'on attendrait. L'ossature discontinue de *gn-* peut être observée dans des mots comme *genial* (adj.) « qui a trait ou qui se rapporte au menton ». Philips 2008 retrace l'origine historique d'un bon nombre de mots en *gn-*, qu'il compare avec l'évolution des mots en *kn-*, similaire à bien des égards. Certains lexèmes en *kn-* et en *gn-* ont d'ailleurs des ancêtres communs, puisqu'ils sont issus de mêmes formes, lorsque l'on remonte au vieil anglais, voire plus loin encore (voir Philips 2012b pour une analyse en profondeur du marqueur *kn-*).

En ce qui concerne la séquence *gr-*, il semble que ce soit le modèle C(V)R- qui se prête à l'analyse. Contrairement au schéma SC-/C-, ce dernier comprend deux éléments phonesthétiques stables, qui participent pleinement à l'élaboration de la structure phonosymbolique de *gr-*. La différence avec la simple structure CR- est que la structure de *gr-* est susceptible d'apparaître sous sa forme discontinue, sur le même principe que pour *gn-*.

Encore une fois, la littérature n'est pas particulièrement prolixue en ce qui concerne la structure phonologique de *gr-* et ses implications phonosémantiques. On peut toutefois rapporter que l'élément *g-* est décrit comme étroitement lié à la parole, à la mâchoire et aux mouvements impliqués dans l'ouverture et la fermeture de la bouche, ce qui fait écho aux résultats de Wright. Une grande quantité de mots en GVC(-) présente ainsi une signification qui a trait à des « phénomènes liés à la mâchoire, que ce soit dans leurs emplois nominaux ou verbaux », comme des « activités d'expression faciale, ainsi que des activités fonctionnelles » qui impliquent d'« ouvrir la bouche, montrer les dents, bailler », ainsi que de « mordre, avaler, respirer [...], qu'il s'agisse de sujets animés humains ou non-humains » (Philps 2008, p. 131). L'exemple de *gurgitate* (v. transitif) « avaler avidement ou de manière excessive (de la nourriture, ou, dans ses emplois plus tardifs, de la boisson) »¹⁸ est probant, car il manifeste un agencement phonesthétique discontinu, et que sa définition résonne avec certains éléments phononotionnels de *gr-*.

Parmi les mots qui entrent dans le gabarit GVC(-), ceux qui présentent les lettres ⟨b⟩, ⟨l⟩, ⟨m⟩ et ⟨p⟩ comme seconde consonne semblent fonctionner à partir de l'élément G- comme invariant, dans le sens où son comportement sémantique apparaît comme relativement stable. Parmi les exemples de mots qui vont suivre, il nous faut préciser que le concept de similarité notionnelle joue un rôle crucial, car il permet de rassembler des unités phonologiques et lexicales qui seraient jugées comme trop sémantiquement éloignées dans le cadre d'un modèle classique. C'est en partie la raison pour laquelle on préfère parler de domaine phononotionnel plutôt que de domaine phonosémantique. Le fait que c'est bien *g-* qui agit comme un invariant est démontré par la comparaison de paires minimales, dans lesquelles ⟨g⟩ est remplacé par une autre consonne. Il s'avère que le domaine phononotionnel originel véhiculé par le gabarit GVC(-) analysé est perdu lorsqu'on change ⟨g⟩ par n'importe quel autre type de consonne, ce qui montre que *g-* agit bien comme le vecteur du domaine, à l'inverse de l'autre consonne de la structure phonesthétique. On peut prendre l'exemple de *guzzles*, exposé dans l'enquête dialectale de Wright, qui représente un mot-notion revoyant à un ensemble de mots en GVC(-) associés au fait de « boire bruyamment et avidement ». Cet ensemble comprend un grand nombre de mots qui contiennent un élément *-l-*. On notera par exemple : *gobble*, *goggle*, *golk*, *gollop*, *golp*, *gubble*, *guddle*, *guggle*, *gulf*, *gulp*, *gurgle* et *guttle* (Wright 2012, p. 11). L'invariant notionnel que semble porter *g-* dans ces mots peut en effet être décrit comme « V bruyamment et avidement », ce qui explique pourquoi des mots comme *graunch*, *graze* ou *grind* dénotent tous

¹⁸ Cette définition correspond à l'entrée de l'OED *online subscription version, 3rd edition*.

le fait de « manger bruyamment et avidement ». A partir de cette démonstration, il peut être inféré que la variable *-l-* semble associée à une notion de « boire », ou du moins à une sorte de fluide, tandis que l'élément *-r-* se rapporte à un processus beaucoup moins fluide, bien plus haché ou laborieux, d'où l'association entre *GV-r-* et des bruits ou des sensations d'écrasement ou de broiement.

Pour conclure sur la nature et les implications de l'élément *g-*, rappelons que cette unité submorphémique semble faire le pont entre des mouvements physiques et des propriétés phononotionnelles, de telle sorte que les mouvements physiologiques de type occlusif puissent transparaître dans la langue sous forme d'une occlusion phonologique. En somme, il est possible que le mode d'articulation du son [g] provienne de mécanismes physiologiques en lien avec la sphère de la mâchoire, et plus précisément d'une « capacité à renvoyer à la mâchoire » (Philps 2008, p. 134).

4.3. Sélection des mots clés

Comme nous l'avons déjà précisé antérieurement, le choix a été fait de procéder à l'élimination des sens qui ne se rapportent pas directement à la nature sémantique supposée de *gr-*, que l'on trouve dans la littérature. Nous avons cependant conservé les éléments qui se rapportent aux catégories les plus importantes qui ont été dégagées lors du travail de master 1. Celles qui transparaissent dans la littérature sont : *sound/voice/utter* (SVU) (Abramova et al. 2013, p. 1698), *weep/pain/unpleasant* (WPU) (Waugh 1994, p. 59), *hold/clutch/touch* (HCT) (Philps 2008, p. 128), *mood/emotion/humour* (MEH) (Abelin 1999, p. 87), *complain/repine/discontent* (CRD) (Abelin 1999, p. 87; Waugh 1994, p. 59), et *grate/scrape/file* (GSF) (Waugh 1994, p. 59). On peut y ajouter celles que l'analyse du master 1, effectuée à partir des mots clés, a révélé comme étant de première importance, en tout cas quantitativement : *snout/teeth/animal* (STA), *low/deep/dig* (LDD), *plant/leaves/wood* (PLW) et *coarse/thick/raise* (CTR).

La récolte des données part de l'observation de mots clés, qui se rapportent aux catégories conceptuelles, à l'intérieur des définitions proposées par les lexicographes. Les mots clés qui ne présentent pas de lien sémantique direct avec ces catégories ont été laissés de côté pour l'analyse. La lecture de ces mots clés a globalement été repensée, dans la mesure où tous les termes utilisés dans une définition n'ont pas la même pertinence. Puisqu'aucune prise en charge qualitative des mots clés n'avait été effectuée dans le cadre du travail de master 1, les

valeurs quantitatives attachées aux catégories conceptuelles sont remises en question. Une lecture plus fine a donc été observée. Un certain nombre d'éléments non-pertinents a ainsi été mis de côté, comme pour la définition de *gradine*, n. 1 [1783] par exemple : « one of a series of low steps or seats raised one above the other ». Il est clair que le terme de *low* ne peut pas être considéré comme un mot clé et adossé à la catégorie conceptuelle *low/deep/dig* (LDD), car l'idée principale, de progression verticale, contenue dans cette définition, est plus saillante, et véhiculée par *steps*, *raised* et *above*. Par ailleurs, les désignations d'espèces aviaires, qui sont relativement nombreuses, ont été laissées de côté, car il est difficile de les rattacher à d'autres catégories. En effet, ces noms d'oiseaux sont trop spécifiques pour être rangés dans la catégorie *snout/teeth/animal* (STA), qui regroupe les mots faisant référence au corps animal de manière plus générale. On peut citer : *grebe*, n. [1768], *grindle*, n.2 [1610], *grouse*, n.1 [1531]. Il est cependant curieux de trouver autant de noms d'oiseaux commençant par *gr-*, que ce phénomène soit dû à une forme de symbolisme phonétique ou non. Il s'agit probablement d'origines étymologiques communes, mais cet aspect ne sera pas creusé davantage. Les espèces ichtyologiques, bien que moins nombreuses, sont écartées pour les mêmes raisons. Elles sont : *grail*, n.5 [1587], *grouper*, n.1 (1) [1615] et *grundel*, n. [1400]. Il en va de même pour les multiples références aux griffons et aux chevaux.

4.4. Précisions sur quelques biais inhérents à la nature du corpus

Certains monomorphèmes posent des problèmes spécifiques, qui reflètent leur histoire, ou qui sont dus à un manque de sources documentaires. On peut mentionner le cas de *graft*, qui, bien qu'il s'agisse originellement d'une modification morphologique de *graff*, il a été retenu pour notre analyse. Le problème de ce monomorphème est que les informations historiques fournies par l'OED indiquent que le verbe comme le nom viennent tous deux d'une variante de *graff*, sans préciser quelle partie du discours est apparue en premier. Les deux entrées sont datées de 1483. Il est fortement probable que l'une des formes soit une conversion de l'autre, mais il est tout aussi possible qu'elles soient entrées dans le lexique indépendamment l'une de l'autre, à partir de *graft*, v. et *graft.*, n. Nous avons choisi d'écarter le verbe, car sa définition renvoie à celle du nom à plusieurs reprises, alors que le contraire n'est pas vrai. Il nous faut souligner que cette décision renvoie à une forme d'arbitraire, puisque nous ne disposons pas des informations suffisantes pour trancher de manière péremptoire, bien qu'il soit impératif de prendre des choix. Lorsque de tels choix se sont montrés inévitables, le présent rapport en fait systématiquement mention, et présente les éléments qui permettent d'aboutir à la décision la

plus appropriée possible. Dans la plupart des cas, lorsque la situation impose l'élimination d'un sens au profit d'un autre bien qu'il soit impossible de faire la part des choses, les deux sens sont conservés dans le corpus. Nous avons cependant tenté d'en éliminer le plus possible, afin de limiter la perturbation des données.

L'exemple de *graft* illustre tout à fait les incohérences chronologiques que l'OED est susceptible de laisser transparaître. Et pour cause, nous avons ici un exemple de sens décrit comme provenant d'un autre sens, qui est en réalité lui-même décrit comme plus tardif. Ainsi, *graft*, v.4 [1859] est décrit comme une conversion de *graft*, n.5 [1865], qui lui est apparemment postérieur de 6 ans. Dans la mesure où il est impossible de déterminer s'il s'agit d'une simple erreur de datation, ou bien d'une information étymologique erronée, nous avons choisi de garder les deux formes. Il n'est pas non plus à écarter que l'étymologie comme la datation soient en réalité correctes, puisqu'il est tout à fait possible d'imaginer qu'une forme convertie en une autre soit attestée plus tardivement dans les textes que cette dernière, sans pour autant que les données soient fausses. Le cas de *gralloch* est intéressant de ce point de vue. Nous avons *gralloch*, v. [1848] « to disembowel (properly, a deer) » qui, nous dit-on, est issu de *gralloch*, n. [1882] « the viscera of a dead deer », bien que la première attestation du second soit postérieure à celle du premier. Il est cependant plus raisonnable d'admettre que le verbe provienne du nom que l'inverse, étant donné que c'est le mot désignant les organes de l'animal qui a dû être converti en verbe, afin d'exprimer l'action consistant à retirer ces organes. De plus, ce mot est issu d'un substantif gaélique, *grealach*, et non d'un verbe. De même, *grub*, v. (6a) [1350] « to dig ; in recent use, connoting the idea of mean or grovellingly laborious occupation », qui semble être un sens dérivé de *grub*, v. (1) [1400] « to dig superficially; to break up the surface of (the ground); to clear (ground) of roots and stumps », est en réalité attesté plus tôt dans les textes. Certains sens d'un même monomorphème ont été conservés car il est impossible de déterminer lequel d'entre eux est apparu en premier. Les deux sens de *grent* ont ainsi été gardés, car le premier sens renseigné est aussi le plus tardif, ce qui paraît paradoxal. En outre, il est difficile d'affirmer lequel de *grent*, v. (1) [1425] « to gnash the teeth » ou *grent*, v. (2) [1387] « to grunt or groan » correspond au sens originel. La section étymologique du dictionnaire ne nous est ici pas d'un grand secours puisqu'elle se contente de supposer : « onomatopoeic; compare *grin*, v.2, *grint*, v., *grunt*, v. ».

D'un certain point de vue, on pourrait considérer que ce sont les explications étymologiques qui priment sur les dates. En effet, la numérotation des sens ainsi que les justifications historiques développées par les lexicographes sont censées être relativement

structurées, tandis que la date d'un texte n'indique pas nécessairement sa réelle apparition dans le lexique. Ainsi, entre *grent*, v. (2) [1387] et *grent*, v. (1) [1425], c'est *grent*, v. (1) qui pourrait être considéré comme le sens le plus ancien, dont le second descend directement, car il est répertorié comme le premier sens dans l'OED. Afin de limiter la distorsion des données, les deux formes ont finalement été conservées.

III. COMPARAISONS ENTRE EBC ET PAC

5.1. Premier regroupement en catégories

Afin d'introduire les résultats qui suivent dans leur contexte, il est pertinent de brièvement résumer le travail préliminaire qui a abouti à la présente étude. Au terme du travail réalisé dans le cadre du master 1, un ensemble de catégories conceptuelles avait été mis en place. Cette première approche reposait sur un prélèvement systématique des mots clés les plus employés dans les définitions proposées par les lexicographes de l'OED. Ainsi, chaque mot clé était relevé, puis les plus récurrents étaient utilisés pour nommer les catégories conceptuelles obtenues. Une catégorie correspondait alors à un regroupement de mots clés considérés comme appartenant à un même domaine phononotionnel, tel qu'il est défini par Philips (2008, p. 128), c'est-à-dire un marqueur sublexical correspondant à la trace matérielle d'une notion sous-jacente. Le problème majeur d'une telle méthode concerne les critères de sélection des mots clés en question, car il n'est pas toujours évident de trancher sur l'importance subjective d'un mot dans une définition. Cet aspect avait été trop négligé, c'est pourquoi les regroupements que nous proposerons ici font l'objet d'une méthodologie bien particulière, qui repose sur l'application de filtres, comme le propose Drellishak 2006 (pp. 21-23). La conséquence directe de cette omission est qu'un très grand nombre de catégories se sont avérées hors-sujet, puisqu'elles ne contenaient aucun sens phonesthétique, en plus de leur très faible représentabilité. En effet, il s'agit le plus souvent de catégories résiduelles, des 'scories' issues du processus de regroupement conceptuel, et qui ne résultent que d'affinités fortuites.

Les 258 sens de cette première sélection ont d'abord été regroupés dans des catégories très restrictives, les *restricted categories*, pour ne pas les confondre avec les suivantes. Ces 65 groupes étaient directement issus d'une association immédiate entre les mots clés, sur lesquels presque aucun critère de sélection n'avait été appliqué, et leurs sens premiers. Le deuxième problème majeur que pose cette sélection est que les catégories ainsi obtenues sont particulièrement inégales, dans le sens où certaines correspondent à un regroupement de sept

ou huit mots clés, tandis que d'autres ne correspondent qu'à un seul d'entre eux. C'est pourquoi une seconde étape sélective était nécessaire. Les catégories comprenant un trop faible nombre d'occurrences furent laissées de côté, tandis que les catégories les plus représentatives furent regroupées en de plus grands groupes. C'est ainsi que les *broad categories* (BC) furent créées, puis analysées, à l'inverse des RC, qui ne correspondent qu'à une étape intermédiaire, mais dont l'étude était largement dispensable. Les analyses présentées dans le mémoire de master 1 sont essentiellement quantitatives, puisqu'elles s'appuient sur des données statistiques, obtenues à partir du travail de saisi. La fréquence d'apparition des catégories représente le point de départ de l'étude, à partir duquel différents types d'analyses ont été développés. En effet, ce sont surtout les catégories les plus représentatives qui ont été analysées, afin de se limiter à l'essentiel, dans les temps impartis par cette étude de master 2. Par ailleurs, une différence a été observée entre les fréquences des BC basées sur le nombre de mots clés, et celles basées sur les sens. Dans le premier cas, la fréquence de la BC repose sur le décompte des mots clés associés à cette catégorie, alors que dans le second on se place au niveau du sens. Ainsi, dans le cas où une occurrence de monomorphème était représentée par plusieurs mots clés se rapportant à la même BC, celle-ci ne comptait que pour une seule occurrence de cette BC, afin de ne pas faire gonfler les chiffres.

Il convient cependant de rappeler l'un des problèmes majeurs qui fut soulevé à l'issue de ce travail préliminaire : la différence entre des monomorphèmes distincts mais homonymes d'un côté, et les divers sens d'un même monomorphème de l'autre, n'avait pas été correctement opérée. Cela pose problème car deux monomorphèmes homonymes correspondent à deux lexèmes bien distincts et n'ont pas la même origine étymologique, tandis que les sens métasémiques d'un même monomorphème correspondent à un unique lexème et descendent d'une même racine étymologique. C'est pourquoi une séparation stricte entre ces deux phénomènes a cette fois-ci été respectée, comme nous le présenterons dans les sections ultérieures.

5.2. Elargissement aux EBC et étude des proto-sens

Le travail de master 2 repart des mêmes bases méthodologiques que celui de master 1, mais la récolte des données a été intégralement refondue. Le protocole général suit l'ordre suivant : dans le premier tableur de saisie de données, toutes les formes ont été conservées, y compris les sens secondaires. Puis dans un deuxième tableur, nous avons procédé à l'élimination de toutes les conversions dont on peut être certain qu'elles ne correspondent pas

aux sens originaux. Enfin, dans une troisième feuille, ce sont les sens secondaires qui ont été écartés. Cette méthode prend du temps mais elle permet de s'assurer qu'on ne retient bien que les sens originaux, car l'OED fait beaucoup d'erreurs à ce niveau-là. La suite du protocole dépend du type d'analyse que l'on conduit. Il convient en effet de distinguer l'étude des proto-sens de celle de l'ensemble des sens. L'étude des proto-sens ne s'appliquera pas à l'analyse diachronique de la même manière que pour les sens métasémiques, puisque l'objectif est, dans le deuxième cas, d'observer les différents glissements de sens à l'œuvre. La comparaison des deux types de données nous permettra de dresser un profil évolutif des relations forme-sens associées au marqueur submorphémique *gr-*. Le principe sous-jacent est qu'une comparaison entre un état des lieux à un point A et un point B est susceptible d'apporter des informations sur l'orientation métasémique des monomorphème en *gr-*, et donc de contribuer au profilage du comportement phononotionnel du marqueur submorphémique. En partant de cette idée directrice, nous avons procédé au regroupement des sens en de nouvelles catégories, basées sur des critères légèrement différents de ceux proposés lors de l'étude préliminaire.

La première étape de l'étude des sens premiers concerne l'élimination des sens métasémiques et des conversions. Le problème auquel on fait face d'emblée est inhérent au fonctionnement de l'OED, ainsi qu'aux éventuels erreurs de la part des lexicographes. En effet, deux indications principales nous renseignent sur l'ancienneté d'un lexème, à savoir sa place dans la numérotation hiérarchique du dictionnaire, ainsi que la date du plus ancien document attestant de son utilisation par des locuteurs. Bien souvent, le sens qui porte le plus petit numéro est le plus ancien, mais il est bon de noter que ce n'est pas systématique. Il est alors nécessaire de vérifier les dates d'attestation, afin d'éviter des aberrances chronologiques. Ceci est surtout vrai pour les entrées qui comprennent un nombre important de sens métasémiques, allant parfois jusqu'à plusieurs dizaines. Une liste des documents connus est donnée en-dessous de chaque définition, accompagnée des dates supposées, ce qui facilite grandement la tâche. Il n'en reste pas moins qu'il est parfois difficile de savoir avec certitude quel est le sens le plus ancien d'un lexème, ce qui constitue un problème majeur. Dans d'autres cas cependant, la situation est relativement transparente. Par exemple, deux sens de *graft*, n.1 avaient initialement été retenus : n.1 (1) [1393] « a shoot or scion inserted in a groove or slit made in another stock », et n.1 (2) [1555] « a twig, shoot, scion; *gen.* a branch, plant: = *graft*, n.1 ». Ici, numérotation lexicographique et datation concordent. Seul le plus ancien de ces deux sens a finalement été conservé, car le second ne constitue qu'une déviation sémantique du premier. *Graft*, n.2 [1637], quant à lui, ne comprenait qu'un seul sens, qui a donc été retenu : « a trench serving as a

fortification; a dry or wet ditch; a foss or moat ; cf. *graft*, n.2 ». Le verbe *graft*, v.1 montre que les sens d'un même lexème peuvent parfois subir des glissements sémantiques assez importants : *graft*, v.1 (1a) [1377] « to insert (a scion of one tree) into a different stock » et *graft*, v.1 (1b) [1536] « to set or fix firmly ». Aucun des deux sens n'a cependant été retenu puisqu'il s'agit d'une conversion directe du nom *graft*, n.1. *Graff*, v.2 [1532] « to dig » a, pour sa part, été retenu ici, puisqu'il s'agit d'une variante de *grave*, v.1 et non d'une conversion.

L'élimination des conversions, quant à elle, nécessite encore d'autres précautions. En particulier, un doute plane lorsque les deux mots présentent la même datation. Faut-il plutôt éliminer le nom ou le verbe ? L'idéal est de vérifier l'étymologie, puisque celle-ci nous informe sur l'évolution du lexème, sur la conversion du substantif en verbe ou le contraire. Mais dans certains cas, aucune information n'est disponible à ce sujet. De manière générale, les parties du discours présentant la même date sont conservées, car il est presque impossible de trancher. On peut citer l'exemple de *grasple*, n. et *grasple*, v., tous deux datés de 1553. Ces éléments sont à garder en tête lors de la formulation de nos conclusions, puisqu'ils parasitent les analyses et sont susceptibles de distordre les apparences. Il doit cependant être souligné qu'il s'agit d'exceptions, puisqu'en définitive, seule une très faible proportion du corpus est concernée. En plus de cela, il est parfois difficile d'estimer s'il y a conversion ou non. La section étymologique de l'OED pour *gripe*, n.1 (1a) [1393] renseigne ceci : « < *gripe*, v.1 (the early examples may belong to *grip*, n.1.) ». Il est ici logique de penser que seuls les sens tardifs constituent de véritables conversions, et donc que le sens 1, daté de 1393, n'en est pas une, puisqu'il provient d'un autre monomorphème. Il n'en reste pas moins que les lexicographes ne sont pas clairs à ce sujet. Ce cas illustre l'incertitude qui demeure dans certains cas, et qui rend le travail de saisie plus compliqué qu'il n'y paraît à première vue.

La création des *extra broad categories* (EBC) s'est articulée en deux temps. D'abord, les catégories conceptuelles du master 1 les plus représentatives ont été gardées comme base de réflexion. Il s'agit des 11 catégories suivantes : *snout/teeth/animal* (STA), *move/body/limb* (MBL), *low/deep/dig* (LDD), *sound/voice/utter* (SVU), *crush/gnash/reduce/small* (CGRS), *anger/fierce/annoy/bout* (AFAB), *sand/dust/particle* (SDP), *weep/pain/unpleasant* (WPU), *plant/leaves/wood* (PLW), *hold/clutch/touch* (HCT) et *coarse/thick/raise* (CTR). Les 14 catégories restantes, sur un total de 25, ont été soit éliminées, soit regroupées avec certaines des catégories majeures. Il s'agit des suivantes : *surface/space/frame* (SSF), *mood/emotion/humour* (MEH), *water/fluid/drain* (WFD), *make/cause/provoke* (MCP), *complain/repine/discontent* (CRD), *skin/fur/tissue* (SFT), *tool/metal/fork* (TMF), *harsh/rough/hard* (HRH), *earth/soil/dirty*

(ESD), *meal/cereal/feed* (MCF), *grate/scrape/file* (GSF), *visual/hue/colour* (VHC), *terror/horror/fright* (THF) et *sulk/surly/gloom* (SSG). La figure n° 2 présente les regroupements de catégories mineures et majeures qui ont été effectués, afin d’obtenir notre nouveau découpage, c’est-à-dire les EBC. Les trois BC qui ont été définitivement écartées sont *surface/space/frame* (SSF), *make/cause/provoke* (MCP) et *visual/hue/colour* (VHC), en partie parce qu’il s’agit de domaines trop vagues, ce qui reflète le manque de sélectivité des mots clés dont il était question plus haut. Qui plus est, ces catégories demeureraient très résiduelles.

Figure n° 2 : Regroupement des *Broad Categories* (BC) en *Extra Broad Categories* (EBC).

BC	EBC
anger/fierce/annoy/bout (AFAB)	weep/complain/anger/fright (WCAF)
weep/pain/unpleasant (WPU)	
mood/emotion/humour (MEH)	
complain/repine/discontent (CRD)	
terror/horror/fright (THF)	
sulk/surly/gloom (SSG)	
crush/gnash/reduce/small (CGRS)	crush/scrape/sand (CSS)
sand/dust/particle (SDP)	
grate/scrape/file (GSF)	
plant/leaves/wood (PLW)	plant/fluid/soil/feed (PFSF)
water/fluid/drain (WFD)	
earth/soil/dirty (ESD)	
meal/cereal/feed (MCF)	
hold/clutch/touch (HCT)	hold/clutch/touch (HCT)
low/deep/dig (LDD)	low/deep/dig (LDD)
snout/teeth/animal (STA)	snout/teeth/body/tissue (STBT)
move/body/limb (MBL)	
sound/voice/utter (SVU)	sound/voice/utter (SVU)
coarse/thick/raise (CTR)	coarse/thick/rough (CTR)

La première chose qu’il faut noter est que, si certaines EBC sont le fruit d’un regroupement de nombreuses BC, comme *weep/complain/anger/fright* (WCAF), d’autres, au contraire, conservent le même nom et le même contenu, comme *low/deep/dig* (LDD) et *sound/voice/utter* (SVU). La catégorie *weep/complain/anger/fright* (WCAF) correspond à une notion d’émotions négatives, qu’il s’agisse de colère ou de rage (*anger, fierce*), d’agacement ou de mécontentement (*annoy, discontent*), de plainte ou de mauvaise humeur (*complain, repine, sulk*), ou encore de peur (*fright*). La catégorie *plant/fluid/soil/feed* (PFSF), quant à elle, revoie aux éléments de la nature, que ce soient des plantes ou de la terre, mais aussi au

vocabulaire qui gravitent autour de la nourriture et de l'alimentation, qui comprend les nombreuses références aux céréales. Contrairement à WCAF, qui est assez homogène et consistante, PFSF paraît plus éclectique, car elle renvoie à des éléments plus disparates. Il nous faut donc garder en tête que cette catégorie ne porte pas la même charge interprétative, dans le sens où les résultats qu'elle pourrait délivrer n'auront pas tout à fait la même valeur que ceux donnés par WCAF. *Hold/clutch/touch* (HCT), *low/deep/dig* (LDD) et *sound/voice/utter* (SVU), à l'instar de WCAF, sont des catégories relativement homogènes et qui sont directement héritées du travail de l'an passé. Il va de soi que ces regroupements vont avoir tendance à orienter les chiffres, dans la mesure où les larges regroupements afficheront une fréquence d'apparition plus élevée que les plus petits, ou que les catégories inchangées. Soulignons toutefois que cette orientation est tout à fait légitime, étant donné que ces regroupements sont justifiés, puisqu'ils correspondent au repérage de traits distinctifs, qu'il s'agit justement d'examiner. La catégorie *crush/scrape/sand* (CSS) correspond à trois concepts, qui ont été jugés comme intimement liés, puisqu'ils renvoient à la notion de particule, qu'il s'agisse d'une matière granulée, ou de la sensation d'irritation qu'elle provoque, ou encore du processus de broyage qui en est à l'origine. Cette catégorie, probablement plus que les autres, illustre parfaitement le caractère phononotionnel de notre découpage conceptuel, puisque ce sont à la fois les inconvenances sensorielles et corporelles qui représentent le point de départ de tous les mots clés en lien avec la catégorie. L'EBC qui s'avère être la moins satisfaisante est *snout/teeth/body/tissue* (STBT), car le lien entre ses éléments constitutifs est moins évident, dans la lignée de *plant/fluid/soil/feed* (PFSF), qui, bien que renvoyant au domaine de la nature, comprend des notions complexes parfois très différentes les unes des autres. L'aspect bucco-nasal est probablement la partie la plus importante de STBT, mais elle a été rangée aux côtés des autres éléments corporels et somatiques, en particulier tout ce qui se rapporte au corps des animaux. Il est en fait presque impossible de clairement distinguer la sphère du corps de celle de l'animal, car il s'agit souvent de termes corporels que l'on utilise pour référer à des animaux. On peut citer *gro*, n. (1) [1230] « a kind of fur », ou *groin*, n.1 (2a) [1300] « the snout, esp. of a swine ». Même les mots qui renvoient à des parties du corps humain rappellent un comportement animal. C'est par exemple le cas de *grizzle*, v.2 (1) [1746] « to show the teeth; to grin or laugh, esp. mockingly ».

Les fréquences d'apparition des différentes occurrences de catégories à travers le corpus ont été calculées, afin de déterminer quels sont les éléments phononotionnels les plus présents (voir le tableau 3 ci-dessous). Il s'agit du premier pas, avant tout quantitatif, dans l'élaboration

d'un profil phononotionnel du marqueur submorphémique *gr-*. On peut remarquer d'emblée que la notion de mauvaise humeur et d'émotions négatives semble prédominante, puisque plus d'un quart des occurrences de catégories s'y rapportent. Les chiffres sont moins marqués pour les catégories suivantes, puisque la pente est plus douce. Les catégories HTC, LDD et STBT, en particulier, présentent exactement la même fréquence. Il est, à ce stade pourtant précoce, déjà possible d'affirmer que les résultats devraient être bien différents de ceux de l'an dernier. Le fait de ne pas avoir pris en compte les sens dérivés et d'avoir éliminé les conversions a complètement redistribué les cartes.

Figure n° 3 : Fréquences des EBC en termes de nombre d'occurrences et de pourcentages.

Extra broad categories (EBC)	Occurrences de catégories	Fréquence brute (%)
weep/complain/anger/fright (WCAF)	73	27,86%
crush/scrape/sand (CSS)	39	14,89%
plant/fluid/soil/feed (PFSF)	32	12,21%
hold/clutch/touch (HCT)	27	10,31%
low/deep/dig (LDD)	27	10,31%
snout/teeth/body/tissue (STBT)	27	10,31%
sound/voice/utter (SVU)	23	8,78%
coarse/thick/rough (CTR)	14	5,34%

Sans surprise, on peut également remarquer que les regroupements plus larges ont donné naissance à des catégories bien plus représentées que les autres. Les catégories les plus représentatives, c'est-à-dire celles qui dépassent le seuil de 10% de fréquence brute, estimé par Smith (2016, p. 176), sont les trois premières : WCAF, CSS et PFSF. Rappelons toutefois que si ces chiffres donnent une idée de la répartition globale des catégories phononotionnelles au sein du corpus, il ne s'agit que de données quantitatives brutes, qui ne reflètent rien des interactions entre les sens, ni du comportement évolutif de *gr-*.

5.3. Apports de l'étude des sens métasémiques

Après avoir créé les EBC, nous avons décidé de comparer cette répartition en catégories à un second type de répartition, qui prendrait en compte à la fois les proto-sens et les sens métasémiques. Les données de départ ont été reprises, puis toute la procédure a été réappliquée, en prenant cette fois-ci en compte tous les sens métasémiques. Les sens métasémiques sont ceux attestés après l'apparition du monomorphème, qui ne coémergent pas avec lui mais qui

apparaissent plus tardivement, et qui sont le résultat de processus métasémiques. L'objectif principal est d'analyser les changements diachroniques, puisque des glissements de sens sont susceptibles d'apparaître. La finalité de cette deuxième approche est de repérer si les mots en *gr-* ont tendance à converger sémantiquement ou non. Autrement dit, on observe le profil métasémique de la métaclasse pour voir si *gr-* a une influence d'origine cognitive sur le lexique ou non. On pourra également, dans un second temps, procéder à une analyse sémantique diachronique qualitative, afin d'identifier les types de changement sémantiques à l'œuvre (métonymie, métaphore). Cet aspect sera abordé dans une section ultérieure.

Le deuxième regroupement des monomorphèmes en catégories suit le même protocole que pour les EBC. Ces catégories basées sur le regroupement de sens secondaires sont nommées PAC (*post-attestation categories*) pour les distinguer des précédentes. Tous les sens ont d'abord été gardés pour cette analyse, même ceux qui n'étaient pas pertinents, car l'objectif est de voir si un monomorphème originellement non-phonesthétique peut acquérir un sens phonesthétique. Ce type de catégorie tire son nom du principe de récolte des données, qui, cette fois-ci, prend en compte les sens que l'on peut désigner comme post-attestés, c'est-à-dire qu'ils apparaissent ultérieurement à l'apparition du monomorphème dans les textes. Notons enfin que les conversions ont été conservées dans ce corpus-ci, car elles sont nombreuses et comprennent leur propre réseau de sens métasémiques. En effet, l'exclusion des conversions aurait occulté une grande partie des données, dans le sens où la conversion est avant tout un processus métasémique (Bauer 2018).

Figure n° 4 : Fréquences des PAC en termes de nombre d'occurrences et de pourcentages, accompagnées des fréquences ne prenant pas en compte les mots clés écartés de l'analyse.

Post-attestation categories (PAC)	Occurrences de catégories	Fréquence brute (%)	Fréquence (sans éléments écartés) (%)
weep/complain/anger/fright (WCAF)	142	23,91%	27,20%
crush/scrape/sand (CSS)	79	13,30%	15,13%
sound/voice/utter (SVU)	73	12,29%	13,98%
<i>éléments écartés</i>	72	12,12%	-
snout/teeth/body/tissue (STBT)	56	9,43%	10,73%
plant/fluid/soil/feed (PFSF)	49	8,25%	9,39%
hold/clutch/touch (HCT)	47	7,91%	9,00%
low/deep/dig (LDD)	45	7,58%	8,62%
coarse/thick/rough (CTR)	31	5,22%	5,94%

Sur le même principe que pour les EBC, les fréquences des PAC ont été calculées, afin de faire ressortir les catégories les mieux représentées à travers le corpus. Les deux catégories les plus répandues sont *weep/complain/anger/fright* (WCAF) et *crush/scrape/sand* (CSS), tout comme chez les EBC, sauf qu'ici, c'est *sound/voice/utter* (SVU) qui arrive en troisième position. Il est d'ores et déjà intéressant de remarquer que SVU passe d'une EBC faiblement représentative, puisqu'elle arrive en avant-dernière position avec une fréquence brute de 8,78%, à une PAC particulièrement bien attestée, d'une fréquence brute de 12,29%. Étant donné le nombre important de monomorphèmes qui ne rentrent dans aucune catégorie, une différence a été opérée entre la fréquence totale et la fréquence calculée sans les monomorphèmes hors-catégories. Le classement ne change pas et les fréquences restent proportionnellement les mêmes, mais les chiffres bruts donnent une meilleure idée du poids statistique des catégories. On peut ainsi considérer que plus d'un quart des catégories d'occurrences appartiennent à la phononotie de WCAF, ce qui, encore une fois, fait écho aux statistiques des EBC. Les chiffres du classement entre EBC et PAC (sans les éléments écartés) sont donc approximativement les mêmes, sauf que cette fois-ci, les catégories PFSF et HTC sont moins bien attestées, et se retrouvent dans la partie basse du tableau, tandis que SVU passe dans la partie haute. Un dernier mot sur les éléments écartés : ce chiffre indique qu'une partie des monomorphèmes qui, originellement, correspondaient à une ou plusieurs catégories, ont développé des sens non-phonesthétiques au cours du temps. La valeur de 12,12% paraît faible par rapport à l'ensemble du corpus, mais c'est une donnée qu'il faudra garder en mémoire lorsque nous tenterons de quantifier les éventuelles convergences phonesthétiques.

5.4. Données statistiques sur l'obsolescence et les types d'emplois

Profitons de cette dernière rubrique méthodologique pour préciser quelques données statistiques générales sur la nature du corpus. Notons que les éventuelles différences intrinsèques entre les deux corpus, engendrées par la dissociation des deux types d'approche, ne pourront pas être soulevées. En effet, si l'analyse basée sur les EBC et celle basée sur les PAC n'entraînent pas les mêmes données quantitatives, seuls les calculs appliqués aux EBC ont été conduits, essentiellement par manque de temps. Dans la mesure où ces calculs, particulièrement chronophages, ne font qu'apporter des détails subsidiaires sur la nature dialectale et évolutive du corpus, les données fournies par les EBC ont été jugées suffisantes. Le tableau 5 présente les données statistiques concernant la proportion de lexèmes obsolètes, archaïques, rares, dialectaux et régionaux. Ce tableau vise à détailler les différentes données statistiques de la métaclasse des mots en *gr-*.

Figure n° 5 : Types d'emploi et obsolescence des EBC, en nombre d'occurrences (tokens) et pourcentages.

EBC	Occurrences de catégories	Fréquence brute (%)	MMP Obsolètes (tokens)	MMP Obsolètes (%)	Régional/dialectal/slang (tokens)	Régional/dialectal/slang (%)
WCAF	73	27,86%	29	39,73%	21	28,77%
SVU	23	8,78%	9	39,13%	7	30,43%
CTR	14	5,34%	6	42,86%	3	21,43%
HCT	27	10,31%	12	44,44%	3	11,11%
LDD	27	10,31%	3	11,11%	12	44,44%
STBT	27	10,31%	10	37,04%	5	18,52%
CSS	39	14,89%	12	30,77%	7	17,95%
PFSF	32	12,21%	5	15,63%	7	21,88%
Total	262	100,00%	86	32,82%	65	24,81%

EBC	Rare/archaïque/spécifique (tokens)	Rare/archaïque/spécifique (%)	Total emplois non-conventionnels (tokens)	Total emplois non-conventionnels (%)
WCAF	7	9,59%	57	78,08%
SVU	1	4,35%	17	73,91%
CTR	1	7,14%	10	71,43%
HCT	3	11,11%	18	66,67%
LDD	2	7,41%	17	62,96%
STBT	2	7,41%	17	62,96%
CSS	4	10,26%	23	58,97%
PFSF	2	6,25%	14	43,75%
Total	22	8,40%	173	66,03%

Dans les entrées de l'OED, certains monomorphèmes sont dits à la fois obsolètes et archaïques, de sorte que la situation n'est pas toujours limpide. Nous avons considéré que la

mention archaïque était plus forte que celle d'obsolète, puisqu'elle implique que le monomorphème en question est encore utilisé dans quelques rares contextes, et donc qu'il n'est pas encore tout à fait obsolète. Les emplois très spécifiques, notamment ceux qui renvoient à un jargon professionnel, ont été comptabilisés dans la même catégorie que les emplois rares et archaïques. De même, les emplois idiomatiques ainsi que les éléments d'argots ont été traités comme des régionalismes, puisque ceux-ci sont généralement ancrés dans une région donnée et correspondent à une utilisation locale. Les monomorphèmes d'emplois à la fois dialectaux et spécifiques ou rares ont été rangés dans la catégorie dialectale/régionale car c'est ce caractère local qu'il est plus intéressant de souligner. Si l'opacité de certains termes employés par les lexicographes est à ajouter aux problèmes que pose l'utilisation de l'OED, il faut aussi préciser que les réalités derrière ces termes sont complexes. Comme nous l'avons déjà expliqué, l'obsolescence d'un mot n'est pas toujours une notion qui a du sens, étant donné que ce mot peut sortir de l'usage pour une partie de la communauté linguistique, et non pour une autre. C'est probablement pourquoi la frontière entre les termes *obsolete*, *archaic* et *rare* est si difficile à tracer pour les lexicographes. On remarque, en particulier, que les variétés de langue parlées en Écosse et dans le nord de l'Angleterre n'ont pas les mêmes réalités d'emplois que dans le reste du Royaume-Uni et, *a fortiori*, le reste de la sphère anglophone.

Les catégories sont rangées dans l'ordre décroissant de haut en bas, depuis la catégorie qui comprend le plus de mots d'emplois non-conventionnels à celui qui en comprend le moins. Les emplois qualifiés ici de conventionnels sont ceux dont les entrées de l'OED ne font aucune remarque particulière, en dehors du statut de *in use*. La catégorie WCAF (*weep/complain/anger/fright*) arrive, une fois de plus, en tête de file, avec 78,08% de lexèmes non-conventionnels. La suite du tableau paraît plus aléatoire, et il est impossible d'en tirer de réelles conclusions. On pourrait poser l'hypothèse selon laquelle les lexèmes phonesthétiques sont plus souvent non-conventionnels, dans la lignée de ce qu'observe Wright 2012 avec les mots d'origine dialectale. Cette enquête sur la répartition régionale des phonesthèmes de l'anglais montre en effet qu'un nombre important de lexèmes non-standard sont phonesthétiques. Ceci n'implique pas nécessairement que la plupart des lexèmes phonesthétiques sont d'origine dialectale, mais la question mérite d'être posée. Ceci est d'autant plus vrai que WCAF n'est pas la catégorie conceptuelle qui contient le plus de lexèmes dialectaux ; cette position revient à LDD (*low/deep/dig*), avec 44,44% de lexèmes régionaux, dialectaux ou argotiques. On peut aussi remarquer que la catégorie HTC (*hold/clutch/touch*) est celle qui contient à la fois le plus de lexèmes obsolètes et rares ou archaïques. En dehors de ces

quelques observations factuelles, aucun schéma particulier ne peut être repéré. Il est possible que WCAF arrive en tête parce qu'elle comprend plus de lexèmes que les autres ensembles, de sorte qu'elle se rapproche du niveau de diversité globale du corpus. Son score est cependant moins élevé que celui de l'ensemble du corpus, dont 66,03% des monomorphèmes sont non-conventionnels. Il est aussi possible qu'il s'agisse simplement d'une fluctuation aléatoire, d'autant que les chiffres sont assez serrés. En tout état de cause, même si les chiffres de notre corpus n'indiquent rien de probant, la question de la diversité des emplois doit être soulevée lorsque l'on cherche à comprendre comment les structures phononotionnelles émergent dans le lexique. À terme, l'objectif de ce type de renseignement est de permettre de dégager des schémas, en comparant le profil des différents phonesthèmes. Il est effectivement probable que les autres phonesthèmes ne présentent pas le même profil que *gr-*, d'où l'intérêt de jalonner le sillon de la recherche sur les unités submorphémiques de données disponibles, sur lesquelles il sera toujours possible de revenir *a posteriori*.

Partie III : Résultats

I. NATURE ET EVOLUTION DES CHANGEMENTS SEMANTIQUES

6.1. Courbes de progression

6.1.1. Présentation du protocole

La première section de ce chapitre, dévolu à l'évolution sémantique du marqueur submorphémique *gr-*, est consacrée à la progression diachronique des catégories conceptuelles. Les traits sémantiques, dont on suppose qu'ils constituent le profil phonosémantique de *gr-*, seront ici évalués et comparés, selon le nombre de monomorphèmes qu'ils acquièrent au cours de leur progression chronologique. En effet, les monomorphèmes en question suivent une tendance évolutive, qui se manifeste par l'acquisition de nouveaux traits sémantiques, qu'il s'agit de traquer. L'identification de la nature exacte de ces changements métasémiques sera abordée dans la dernière partie de ce volet, mais il convient d'amorcer notre analyse par l'étude des courbes de progression, qui serviront de point d'ancrage à la suite de l'exposé. Ces courbes constituent une donnée purement quantitative, dans le sens où elles ne font qu'indiquer le ratio auquel nos catégories conceptuelles intègrent de nouveaux monomorphèmes. Par ailleurs, il faut bien souligner que ces courbes sont cumulatives, c'est-à-dire qu'elles n'indiquent pas les hausses et les baisses du stock lexical des catégories, mais simplement l'accumulation de lexèmes par ces dernières. Si des courbes affichant les évolutions ascendantes et descendantes de ces catégories au cours du temps auraient été plus parlantes, le fonctionnement de l'OED ne nous permet pas d'en faire état. En effet, comme il l'a déjà été souligné dans le chapitre méthodologique, les dates fournies par l'OED concernent les premières attestations écrites des lexèmes, mais aucune information n'est disponible sur la fin de vie de ces lexèmes.

En outre, il convient de se rappeler des biais inhérents au fonctionnement de l'OED. En particulier, les datations de nombreuses entrées ont été révisées au cours du temps, de sorte qu'un bon nombre des dates pourvues ne sont qu'approximatives (Allan 2012, pp. 20-21). Certaines d'entre elles prêtent d'ailleurs à confusion, comme Smith (2016, p. 183) l'a soulevé dans le cas de *flirt*. On peut remarquer que, les sources textuelles étant plus abondantes pour les périodes plus tardives, les premiers siècles sont marqués par une plus grande pauvreté documentaire. Pour la plupart des traits sémantiques, les sens les plus anciens sont datés tous les 50 ou 100 ans, alors que les sens les plus récents sont presque datés à l'échelle de la décennie.

Précisons enfin que la section qui suit traite des EBC, tandis que la suivante concerne les PAC. Pour résumer, l'analyse des EBC, c'est-à-dire des premiers sens auxquels les monomorphèmes ont été associés au cours de leur histoire, nous donne des indications sur la formation de nouveaux monomorphèmes se rapportant aux traits sémantiques, tandis que l'analyse des PAC nous informe de la capacité des traits sémantiques à faire converger la signification de monomorphèmes n'appartenant pas originellement à la catégorie en question. Autrement dit, l'analyse des EBC est un indicateur de la productivité lexicale des traits sémantiques, alors que celle des PAC est un indice de convergence phonesthétique. Ceux deux types d'analyses sont donc complémentaires, car elles nous informent sur les deux types de paramètres sur lesquels la phonesthésie est supposée agir. Notons que le terme de productivité est ici utilisé par analogie avec le phénomène de production lexicale propre aux unités compositionnelles, bien que la phonesthésie ne soit pas considérée comme productive au sens conventionnel.

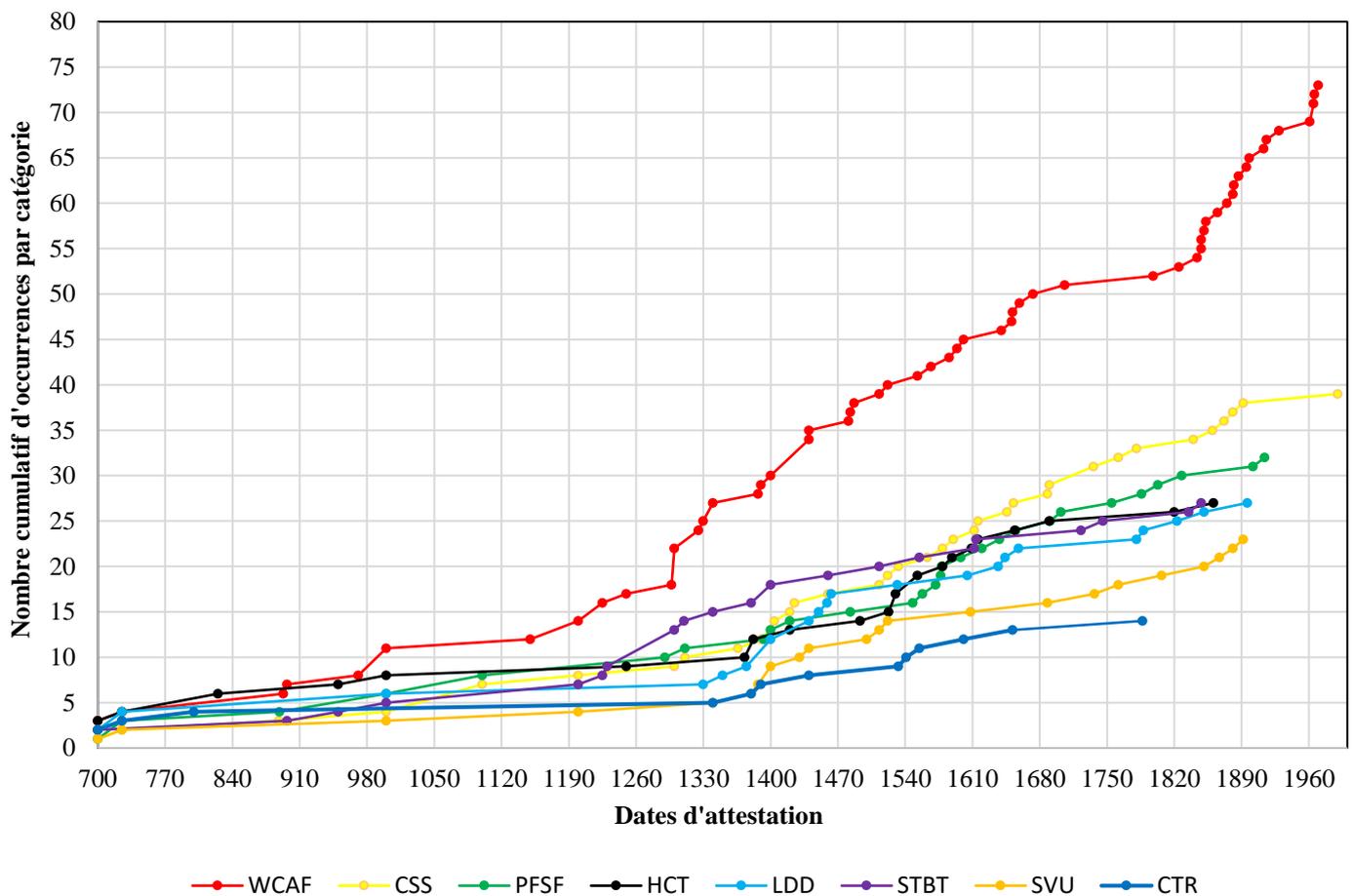
En ce qui concerne les manipulations sur Excel, nous avons d'abord pris la liste des occurrences de catégories, rangée dans l'ordre chronologique, à partir de laquelle nous avons calculé le nombre d'occurrences par sens et par date d'attestation. La courbe évolutive ne reflète donc pas le nombre d'occurrences acquises à telle date ou entre telle et telle date, mais l'accumulation d'occurrences à la date en question. Chaque date de chaque sens a ainsi été relevée, afin que le résultat revoie le plus fidèlement possible aux informations fournies par l'OED. Quelques ajustements ont cependant dû être effectués. D'abord, seules des données chiffrées peuvent être incluses dans le tableau de calcul, c'est pourquoi il a fallu nettoyer la mise en forme. Nous avons choisi de remplacer la mention *OE* par la date arbitraire de 700, car c'est un chiffre rond et qu'il s'agit de la date la plus ancienne donnée par l'OED. Si l'ensemble des courbes des traits phononotionnels part de 700, ce n'est donc pas parce que les plus anciennes attestations datent nécessairement de cette année-là, mais parce que les premiers lexèmes de toutes les catégories sont datés de la période du vieil anglais. L'appellation *OE* est en effet utilisée par les éditeurs afin d'indiquer la période globale des textes qu'il n'est pas possible de dater avec plus de précision, ce qui est très fréquent pour la période anglo-saxonne. Par ailleurs, les fourchettes chronologiques sont difficiles à intégrer à une table de calcul lorsqu'il faut attribuer une valeur fixe en abscisses à une valeur fixe en ordonnées. Nous avons choisi de partir des dates les plus hautes, car une fourchette indique que le lexème d'entrée est utilisé au moins à partir de la date haute. Ainsi, si la première attestation de *gremth*, n. est datée de [1340-70], cela indique que le premier texte qui permet de certifier l'entrée dans le lexique de ce monomorphème date au moins de 1340, autrement dit que le lexème existait déjà en cette

année-là. Or ce sont surtout les dates d'entrée dans lexicque qui nous intéressent, les dates les plus basses d'indiquant rien de particulier.

6.1.2. Analyse des EBC

La figure n° 6 ci-dessous est le diagramme de la progression diachronique des huit EBC, avec les dates en abscisses et le nombre cumulé de monomorphèmes en ordonnées. Chaque point de la courbe représente l'accumulation de proto-sens porteurs du trait sémantique pour une année donnée, tandis que la courbe elle-même représente la tendance évolutive globale du trait. En somme, à chaque point, c'est-à-dire à chaque variable de la table de calcul associée au graphe, correspond l'accumulation du nombre de sens par année indiquée par l'OED. Ainsi, si l'on prend l'exemple de WCAF (*weep/complain/anger/fright*), l'année 700 comptabilise 3 sens, puis on se retrouve avec une accumulation de 4 sens en 725, 6 en 893, et ainsi de suite, jusqu'au nombre de 73 sens en 1970.

Figure n° 6 : Progression diachronique des EBC.



Bien qu'il ait été obtenu à partir de données heuristiques, ce graphe nous donne des indications importantes, notamment parce que ce type de représentation permet une meilleure appréciation des différences et similitudes entre les huit traits sémantiques. Avant toute chose, rappelons que l'ascendance des courbes est le résultat de données cumulatives. C'est pourquoi toutes les courbes sont nécessairement ascendantes, et qu'aucune n'est descendante. L'arrêt de la productivité d'un trait se manifeste donc par un arrêt de la courbe, tandis qu'une réduction de la productivité se manifeste par un aplatissement de la courbe, un plateau. Le premier élément que l'on peut observer est que le trait WCAF (*weep/complain/anger/fright*), qui se rapporte au domaine des émotions négatives et de la mauvaise humeur, affiche une évolution plus grande et plus rapide que les autres traits. Si tous les traits démarrent approximativement de la même manière, WCAF compte déjà un nombre plus important de monomorphèmes à partir du X^e siècle. Sa courbure reste légèrement au-dessus des autres, jusqu'aux XIII^e-XIV^e siècles, où elle se démarque plus distinctement de l'agrégat de départ. La courbe ne fait ensuite qu'augmenter, sans atteindre de plateau définitif. Lorsque l'on s'intéresse au reste des

catégories, la situation est plus difficile à cerner, mais il est important de noter que cette hégémonie de WCAF semble indiquer qu'il s'agit du trait subsémantique le mieux attesté de la métaclasse. Le cas de CSS (*crush/scrape/sand*) est intéressant, car il paraît évoluer dans la médiane des différentes courbes, jusqu'au XVI^e siècle, où sa projection semble décoller, pour atteindre un plateau dans les dernières décennies. Une telle information pourrait indiquer que le trait est plus productif à partir du XVI^e siècle. Il convient cependant de rester prudent dans la formulation de nos conclusions, dans la mesure où les marges d'erreur n'apparaissent pas sur le graphe. La même question se pose pour STBT (*snout/teeth/body/tissue*), qui suit une hausse entre les XII^e et XVI^e siècles, avant de progressivement atteindre son plateau. Quant à SVU (*sound/voice/utter*) et CTR (*coarse/thick/rough*), leurs courbes paraissent nettement en-dessous de la moyenne, ce qui apparaissait déjà clairement sur la figure n° 4. PFSF (*plant/fluid/soil/feed*) et HTC (*hold/clutch/touch*) partagent, pour leur part, une courbure très similaire, qu'il est difficile d'interpréter. Ces éléments ne permettent pas de dégager de réponse catégorique, mais ils donnent un aperçu relativement fiable de la tendance générale que suivent les traits phonosémantiques de *gr-*.

La deuxième remarque importante que l'on peut formuler est que, d'après ce graphe, toutes les catégories conceptuelles ne survivent pas jusqu'à aujourd'hui. CTR, notamment, semble s'éteindre à la fin du XVIII^e siècle, avec une dernière occurrence en 1787, alors que WCAF et CSS sont encore attestées au XX^e siècle, avec pour dernières occurrences en date 1970 et 1990, respectivement. Le reste des traits semble disparaître entre la fin du XIX^e et le début du XX^e siècle, avec pour dernières occurrences : 1848 (STBT), 1861 (HTC), 1892 (SVU), 1896 (LDD) et 1914 (PFSF). Il est amusant de noter que cette période semble agir comme un filtre, au-delà duquel seuls deux des traits restants ont eu l'occasion de poursuivre leur course. Les raisons d'un tel phénomène peuvent être multiples. D'abord, il n'est pas à écarter qu'il s'agisse d'une fluctuation aléatoire. Il faut avouer qu'il serait surprenant que cinq des huit catégories analysées s'éteignent à un peu plus d'un demi-siècle d'intervalle, mais rien ne nous permet, pour le moment, d'être plus affirmatif sur ce point. Ensuite, il est envisageable que le profil sémantique de *gr-* ait simplement changé au cours du temps, de sorte que seuls les deux traits les plus importants ont survécu à l'étape du XIX^e siècle. D'une certaine manière, cela pourrait indiquer que la séquence phonologique /*gr/* a, peu à peu, acquis un sens phonesthétique qui n'était pas présent au départ. On peut même supposer qu'elle n'était pas phonesthétique du tout au départ, avant de le devenir au cours du temps, essentiellement à partir des XIII^e-XIV^e siècles. Dans ce cas de figure, l'interprétation qui convient est d'admettre que le trait

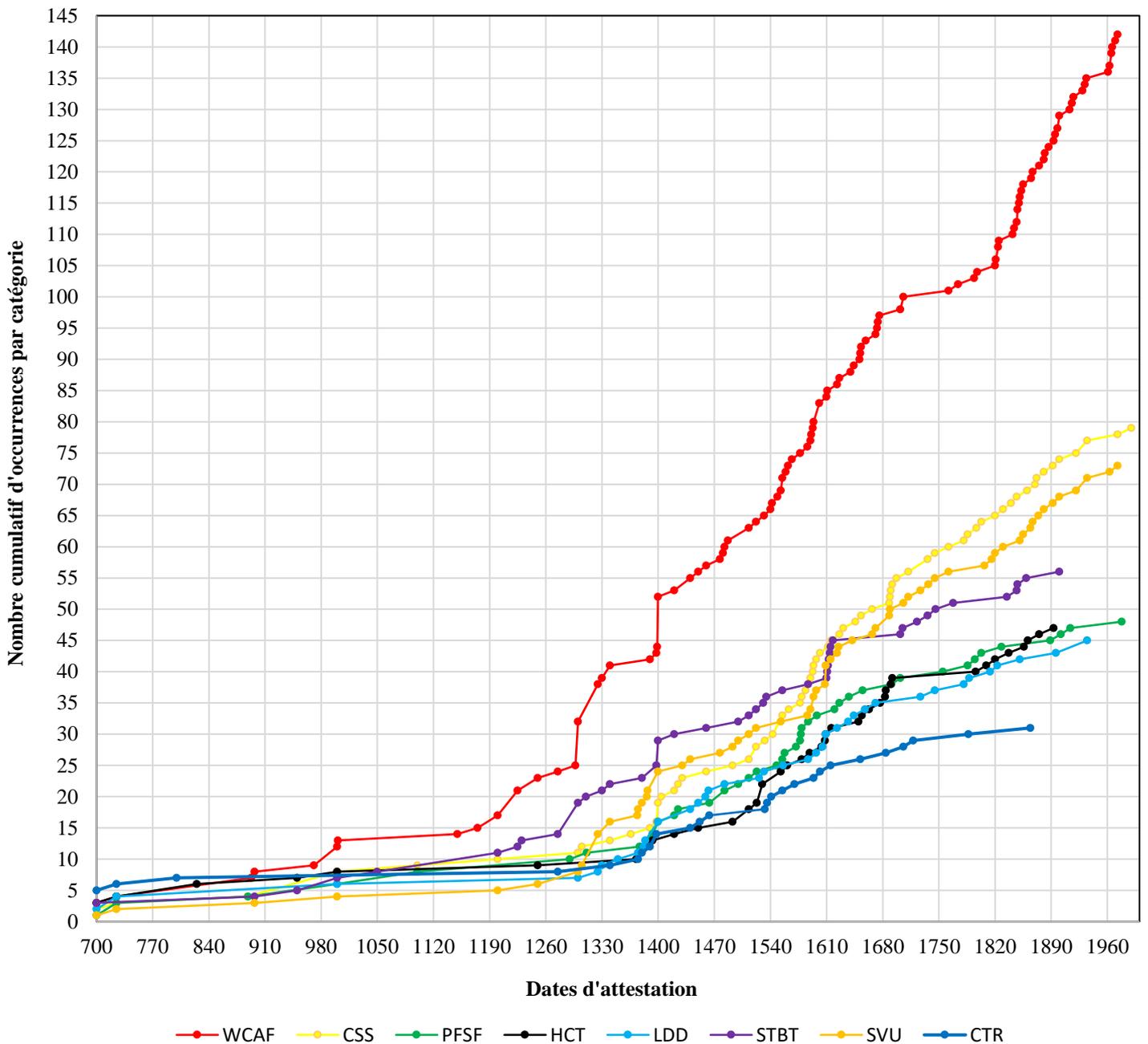
phonesthétique WCAF est apparu environ 300 ans avant CSS. Il sera particulièrement intéressant de comparer cet aspect avec la situation que l'on peut observer pour les PAC. En revanche, tous les traits de *gr-* sont présents dès le début de la période, ce qui n'est pas le cas pour tous les phonesthèmes (voir Smith 2016 et 2020 pour l'étude diachronique de *fl-*).

Dans tous les cas, ce graphe est une bonne illustration du fait que, de la même manière qu'il est difficile de garantir l'obsolescence définitive d'un lexème, il est encore plus complexe de s'assurer de la chute d'un trait conceptuel. En effet, on s'aperçoit que des périodes de vide lexical ne signifient pas nécessairement que le trait ne sera pas activé plus tard. Ainsi, entre 1706 et 1798, aucun monomorphème en *gr-* de notre corpus se rapportant à WCAF n'a été relevé, bien que la période qui précède, ainsi que celle qui suit, bénéficient d'un maillage particulièrement serré. Ceci est d'ailleurs probablement dû à une variation aléatoire, dans le sens où il est prévisible que le maillage ne soit pas régulier tout au long de la courbe.

6.1.3. Analyse des PAC

Le même procédé a été appliqué aux PAC, en sachant que la manipulation des données est beaucoup plus longue, étant donné que le corpus intégrant les sens métasémiques comprend 497 entrées, contre 227 pour l'étude des proto-sens seuls. De même que pour les EBC, la figure n° 7 ci-dessous représente l'évolution diachronique des PAC.

Figure n° 7 : Progression diachronique des PAC.



Pour commencer, l'évolution des courbes est beaucoup plus abrupte que celle des EBC, ce qui était parfaitement prévisible car, l'assemblage étant plus large, la quantité de sens pour une même échelle chronologique est bien plus importante lorsqu'on prend tous les sens en compte. La démarcation de WCAF paraît cependant plus accentuée encore, même si elle s'amorce également au cours des XIII^e-XIV^e siècles. Précisons dès maintenant que les courbes entre EBC et PAC sont nécessairement similaires, puisque l'analyse des PAC intègre une partie des données révélées par l'analyse des EBC. En effet, les PAC comprennent à la fois les proto-

sens et les sens métasémiques. Afin de mieux différencier la part que jouent les sens métasémiques, il serait intéressant de comparer ces résultats à un troisième type d'analyse, qui ne prendrait en compte que les sens métasémiques, en écartant les proto-sens. Cela permettrait de distinguer plus facilement, ne serait-ce que visuellement, le rôle des sens métasémiques de celui des proto-sens. On s'aperçoit effectivement que si le départ des courbes de la figure n° 6 est très semblable à ce qu'on observe sur la figure n° 7, c'est en grande partie parce que les premiers siècles sont dépourvus de sens métasémiques. Afin d'illustrer ce point, nous avons rangé les premières accumulations de sens de 700 à 1200 pour le trait WCAF (voir figure n° 8 ci-dessous).

Figure n° 8 : Données brutes d'exploitation de WCAF, de 700 à 1200. Le tableau de gauche correspond aux proto-sens, celui de droite aux sens métasémiques.

Dates	WCAF	Dates	WCAF
700	3	700	3
725	4	725	4
893	6	893	7
897	7	897	8
971	8	971	9
1000	11	1000	12
1150	12	1001	13
1200	14	1150	14
		1175	15
		1200	17

On remarque dès lors que la différence est relativement faible, même si elle existe. Nous ne développerons pas davantage cet aspect comparatif, mais il est important de souligner que cet écart se creuse avec le temps. Des catégories conceptuelles qui, au départ, se comportent de manière analogue aux autres, finissent par se singulariser à un moment ou à un autre. De manière générale, la proportion de sens métasémiques augmente avec le temps. Dans tous les cas, le rôle de trait distinctif fondamental que pourrait jouer WCAF semble se confirmer. Le reste des traits sémantiques offre globalement la même conjecture, en dehors de SVU. Sa courbe EBC reste, en effet, nettement en-dessous de toutes les autres, à l'exception de CTR, tandis que la PAC se hisse juste en-dessous de CSS. Cette distinction entre les deux corpus pourrait indiquer l'émergence d'un sens phonesthétique à partir des XVI^e-XVII^e siècles, où la courbe de SVU semble grimper. Les trois éléments phonosymboliques majeurs seraient ainsi identifiés comme renvoyant au domaine des émotions négatives et de la mauvaise humeur (WCAF), à

celui du grincement et des sensations irritantes (CSS), et enfin à celui du son et de la voix (SVU). Les traits PFSF, LDD et HTC, quant à eux, proposent le même profil global, et sont tout aussi proches les uns des autres pour les PAC que pour les EBC. Le trait STBT est plus compliqué à interpréter, car il apparaît en deuxième position, derrière WCAF, du XII^e au XVI^e siècle, période à partir de laquelle il est nettement dépassé par CSS et SVU. Ces deux derniers restent proches l'un de l'autre jusqu'au XX^e siècle, de manière assez régulière. L'écart est si étroit qu'il est difficile de confirmer l'ascendance de l'un sur l'autre simplement à l'aide de ce graphe. CSS se situe au-dessus de SVU jusqu'aux années 1330-1340, où SVU passe devant, jusqu'aux années 1540-1550, où CSS reprend le dessus, dans un jeu d'alternance qui ne fait pas forcément sens, et qu'il ne faudrait pas surinterpréter.

6.2. Schémas combinatoires

6.2.1. Présentation du protocole

L'objectif de la présente section est d'évaluer le potentiel combinatoire des catégories conceptuelles qui ont été présentées dans les paragraphes précédents, en vue de repérer quels sont les traits sémantiques qui se combinent le plus avec les autres. Le principe de départ n'est donc pas différent du travail de master 1, dans le sens où les traits qui agissent comme des pôles d'attraction sémantiques sont les plus susceptibles de nous en apprendre sur le contenu phonosémantique du marqueur *gr-*. En somme, le tout est de vérifier si les monomorphèmes qui acquièrent de nouveaux traits sémantiques au cours du temps ont tendance à converger vers telle ou telle catégorie conceptuelle, ce qui permettrait de dégager un cœur sémantique associé au phonesthème. L'hypothèse de départ repose sur l'idée qu'un monomorphème qui apparaît dans plusieurs catégories a plus de chances d'avoir acquis une portion de ses traits sémantiques au cours du temps. D'où l'intérêt d'une comparaison entre le profil phonosémantique des proto-sens et celui des sens acquis, qui résultent de processus métasémiques.

Nous présenterons ici les résultats de l'analyse des schémas combinatoires qui caractérisent les catégories conceptuelles, les EBC comme les PAC. Les courbes de progression cumulatives et les schémas combinatoires sont complémentaires, puisque leur objectif est d'esquisser le profil phononotionnel du marqueur submorphémique, afin d'isoler ce qui s'apparenterait à une fréquence fondamentale. Il s'agit de déterminer le nombre d'occurrences de catégories avec lesquelles chaque catégorie se combine ; la manipulation consiste à prendre la liste des occurrences de catégories, pour chacun des traits sémantiques, puis à compter le

nombre de fois que telle catégorie se combine avec une autre au sein d'un même sens. Chaque sens du corpus peut renvoyer à un, deux ou trois traits, selon les identifications qui ont été inférées des mots clés qui le définissent. En plus de cette donnée quantitative, le but de la manœuvre est de déterminer avec quels autres traits chacune des catégories se conjugue le plus. On peut ainsi établir un profil combinatoire, pour chaque trait, qui nous indique d'une part le potentiel combinatoire du trait, et d'autre part, dans quelles proportions il se combine avec l'ensemble de ses semblables.

6.2.2. Analyse des EBC

La figure n° 9 qui suit présente le profil combinatoire de chaque catégorie conceptuelle, une à une. Les tableaux de droite indiquent les données globales, c'est-à-dire la proportion d'occurrences du trait phononotiel qui ne se combine avec aucun autre trait, distinguées de celles qui se combinent avec un autre trait, et enfin de celles qui se combinent avec deux traits. Le total des combinaisons donne une idée de la force de convergence du trait. Les tableaux de gauche, quant à eux, fournissent des informations plus précises sur les occurrences de catégories avec lesquelles chaque trait se combine le plus. Les valeurs des deux tableaux sont exprimées à la fois en nombres de combinaisons par trait et en pourcentages. La fréquence brute des tableaux de gauche exprime la proportion de chaque trait par rapport au total de ses combinaisons. Ainsi, si l'on prend l'exemple de SVU, on peut lire que 47,06% des combinaisons qu'il entretient avec d'autres traits concernent WCAF, c'est-à-dire qu'il existe 8 sens se rapportant à SVU qui se rapportent également à WCAF. Il s'agit donc du trait avec lequel il se combine le plus. Les catégories sont rangées dans l'ordre décroissant de leur potentiel combinatoire global, noté dans la case bleue. La valeur de cette case bleue est calculée en divisant la quantité de combinaisons par le nombre total d'occurrences de catégorie pour le trait concerné. Enfin, à l'intérieur des tableaux de gauche, les traits combinatoires sont rangés du plus au moins fréquent.

Figure n° 9 : Tableaux des données combinatoires globales et détaillées des EBC.

SVU		
EBC	Combinaisons	Fréquence brute (%)
WCAF	8	47,06%
CSS	5	29,41%
LDD	2	11,76%
STBT	2	11,76%
CTR	0	0,00%
PFSF	0	0,00%
HCT	0	0,00%
Total	17	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	10	43,48%
1	9	39,13%
2	4	17,39%
Total combinaisons	13	56,52%

CSS		
EBC	Combinaisons	Fréquence brute (%)
SVU	5	26,32%
WCAF	4	21,05%
STBT	4	21,05%
HCT	2	10,53%
PFSF	2	10,53%
LDD	1	5,26%
CTR	1	5,26%
Total	19	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	24	61,54%
1	11	28,21%
2	4	10,26%
Total combinaisons	15	38,46%

STBT		
EBC	Combinaisons	Fréquence brute (%)
CSS	4	36,36%
WCAF	2	18,18%
SVU	2	18,18%
CTR	1	9,09%
LDD	1	9,09%
HCT	1	9,09%
PFSF	0	0,00%
Total	11	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	17	62,96%
1	9	33,33%
2	1	3,70%
Total combinaisons	10	37,04%

CTR		
EBC	Combinaisons	Fréquence brute (%)
WCAF	2	40,00%
CSS	1	20,00%
STBT	1	20,00%
PFSF	1	20,00%
SVU	0	0,00%
HCT	0	0,00%
LDD	0	0,00%
Total	5	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	9	64,29%
1	5	35,71%
2	0	0,00%
Total combinaisons	5	35,71%

WCAF		
EBC	Combinaisons	Fréquence brute (%)
SVU	8	47,06%
CSS	4	23,53%
STBT	2	11,76%
CTR	2	11,76%
LDD	1	5,88%
HCT	0	0,00%
PFSF	0	0,00%
Total	17	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	61	83,56%
1	7	9,59%
2	5	6,85%
Total combinaisons	12	16,44%

LDD		
EBC	Combinaisons	Fréquence brute (%)
SVU	2	40,00%
CSS	1	20,00%
WCAF	1	20,00%
STBT	1	20,00%
CTR	0	0,00%
HCT	0	0,00%
PFSF	0	0,00%
Total	5	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	23	85,19%
1	3	11,11%
2	1	3,70%
Total combinaisons	4	14,81%

HTC		
EBC	Combinaisons	Fréquence brute (%)
CSS	2	66,67%
STBT	1	33,33%
WCAF	0	0,00%
CTR	0	0,00%
SVU	0	0,00%
LDD	0	0,00%
PFSF	0	0,00%
Total	3	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	24	88,89%
1	3	11,11%
2	0	0,00%
Total combinaisons	3	11,11%

PFSF		
EBC	Combinaisons	Fréquence brute (%)
CSS	2	66,67%
CTR	1	33,33%
SVU	0	0,00%
WCAF	0	0,00%
STBT	0	0,00%
LDD	0	0,00%
HCT	0	0,00%
Total	3	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	29	90,63%
1	3	9,38%
2	0	0,00%
Total combinaisons	3	9,38%

La première remarque que l'on peut formuler est que le trait sémantique SVU présente une fréquence combinatoire globale très importante, avec plus de la moitié des occurrences de catégories de ce trait qui se combine avec d'autres traits. On peut également remarquer qu'il s'agit du trait avec les plus hauts taux pour les deux types de combinaisons positives (1 et 2), toute catégorie confondue. Ces données constituent un complément indispensable aux courbes de progression, globalement difficiles à interpréter, en apportant un autre regard. On s'aperçoit en effet qu'un sens qui, en termes de nombre d'occurrences de catégories, est beaucoup plus faible que la moyenne, (voir le tableau n° 3 et le graphe n° 6), semble être l'élément qui apparaît le plus en combinaison avec d'autres traits. Ce résultat n'est pourtant pas surprenant, car la notion de /son/ paraît effectivement omniprésente lorsqu'on se penche sur le marqueur *gr-*. C'est en tout cas l'intuition qu'en ont les locuteurs natifs, comme en témoignent les commentaires présents dans la littérature, et c'est également l'impression que l'on a lorsque l'on manipule les données. Restait-il encore à démontrer cette omniprésence. Dans la mesure où ces données sont issues des EBC, on peut supposer que la notion de /son/ est présente de manière sous-jacente depuis les premières attestations écrites de monomorphèmes en *gr-*, jusqu'à aujourd'hui. Son rôle dans la formation de nouveaux lexèmes est cependant mineur,

puisque sa progression brute au cours du temps est quasiment nulle (voir figure n° 6). En d'autres termes, son apparition au sein des proto-sens au cours du temps est faible, mais consistante. Ce trait n'est pas présent, dès le départ, chez beaucoup de monomorphèmes, mais de nouveaux monomorphèmes y renvoyant continuent d'apparaître tout au long de la période, de manière sporadique mais relativement régulière. Pour le moment, rien ne nous permet toutefois d'être trop catégorique quant à l'importance phonesthétique de SVU, en grande partie parce que la probabilité pour qu'une part importante des mots clés d'une définition de l'OED renvoie au domaine des sons est assez grande. En effet, tout ce qui touche au sens, et plus particulièrement à la vue, au toucher et à l'ouïe, est très courant chez beaucoup de lexèmes, que l'on est affaire à de la phonesthésie ou non. Il sera ainsi particulièrement intéressant de comparer ces données avec celles de la PAC SVU, afin de déterminer si cette omniprésence combinatoire de SVU a fini par amener des lexèmes originellement dépourvus de ce trait à l'adopter. En effet, c'est avant tout l'analyse des PAC qui nous permettra d'aborder la question de la convergence. La catégorie avec laquelle SVU se conjugue le plus est WCAF. Ceci n'est pas surprenant, étant donné la fréquence générale de ce trait à travers le corpus, et ce n'est pas par hasard s'il apparaît dans le haut du tableau de gauche pour la plupart des catégories. Plus généralement, SVU ne se combine qu'avec des traits qui sont eux-mêmes parmi les plus associatifs, c'est-à-dire WCAF, CSS et STBT. La présence de LDD, en troisième position, est plus surprenante, car elle n'a un taux global que de 14,81%. En dehors de celle-ci, SVU ne se combine jamais avec les catégories les plus isolées, que sont CTR, PFSF et HTC.

Le trait CSS arrive en deuxième position, avec un taux combinatoire global de 38,46%. Alors que CSS est le deuxième trait avec lequel SVU se combine le plus, SVU est le premier avec lequel CSS se combine le plus, ce qui fait sens. Le deuxième trait de la liste est WCAF. On observe donc un rapprochement net entre trois des huit catégories, qui apparaissent systématiquement parmi les traits les plus associatifs des uns et des autres. Il s'agit en fait des seuls traits qui occupent systématiquement les deux premières places de presque toutes les catégories. Par ailleurs, CSS présente approximativement la même proportion entre la part de combinaisons 1 et 2 que SVU, ce qui n'est globalement pas le cas des traits suivants, qui ont tendance à ne se combiner qu'une seule fois dans la plupart des cas. De manière générale, son orientation combinatoire est plus diffuse que SVU, dans le sens où même les traits avec lesquels il entretient peu de liens sont représentés, là où SVU ne se combine, en tout et pour tout, qu'avec quatre catégories sur huit.

STBT est, une fois de plus, difficile à cerner, car ce trait est à peu près aussi associatif que CSS sur le plan général, avec un taux global de 37,04%, mais il ne s'associe qu'avec un seul autre trait dans la plupart des cas. Il a même l'un des taux de combinaison 2 les plus bas, avec seulement 3,70% de ses occurrences. On peut donc considérer que ce trait se place légèrement en-dessous du trio de tête, occupé par SVU, CSS et WCAF.

CTR arrive juste derrière STBT, et ne se combine qu'avec WCAF, CSS, STBT et PFSF. Cela indique que ce trait ne se résumerait en fait qu'à un ensemble d'amplificateurs. CTR pourrait ainsi être interprété comme un amplificateur d'autres traits. Lorsque l'on reprend les définitions auxquelles ils renvoient, on s'aperçoit en effet que les entrées ayant un lien avec des émotions négatives, des sensations de crissement, des animaux ou des plantes, comportent des termes hyperboliques qui ont été interprétés comme un trait sémantique à part entière. En outre, l'intégralité des occurrences de CTR qui se conjuguent avec d'autres traits ne le font qu'avec un seul d'entre eux par entrée. Ce résultat confirme l'hypothèse d'un simple marqueur d'intensification, car il n'est là que pour renforcer le trait sémantique, unique et central, qui caractérise le lexème. Comme nous l'avons déjà précisé auparavant, la pertinence du trait CTR est à remettre en doute.

WCAF, pour sa part, se combine surtout avec les traits ayant eux-mêmes un potentiel associatif élevé. Les trois premiers traits avec lesquels il se combine le plus sont ceux qui présentent les taux globaux les plus élevés, dans le même ordre : SVU, CSS, STBT. Corolairement, les catégories qui ont un taux global plus élevé que WCAF l'intègre en première ou deuxième place, alors que celles dont le taux global est plus bas ne le présente qu'en troisième ou quatrième place. Cela pourrait indiquer que le trait WCAF se combine de préférence avec des traits qui sont déjà attractifs, et qu'inversement, les traits au plus haut potentiel associatif se conjuguent avec lui plutôt qu'avec d'autres traits. Même s'il n'occupe étonnement que la cinquième place, la plupart des autres catégories l'intègrent dans les deux premiers traits les plus associatifs. Toutefois, il convient de rappeler que la surreprésentation de WCAF à travers le corpus, par rapport aux autres traits, a tendance à orienter les résultats. En effet, si WCAF représente le trait le plus associatif pour la plupart des catégories, les liens qu'il tisse avec ces catégories ne représentent qu'une petite portion de l'ensemble de ses occurrences. C'est en grande partie pourquoi il apparaît en premier ou en deuxième dans la plupart des tableaux malgré son faible taux global. En somme, dans la mesure où ce taux global renvoie à un pourcentage du total d'occurrences de la catégorie concernée, plus la catégorie est représentative par rapport à l'ensemble du corpus, plus le nombre de liens qu'elle entretient

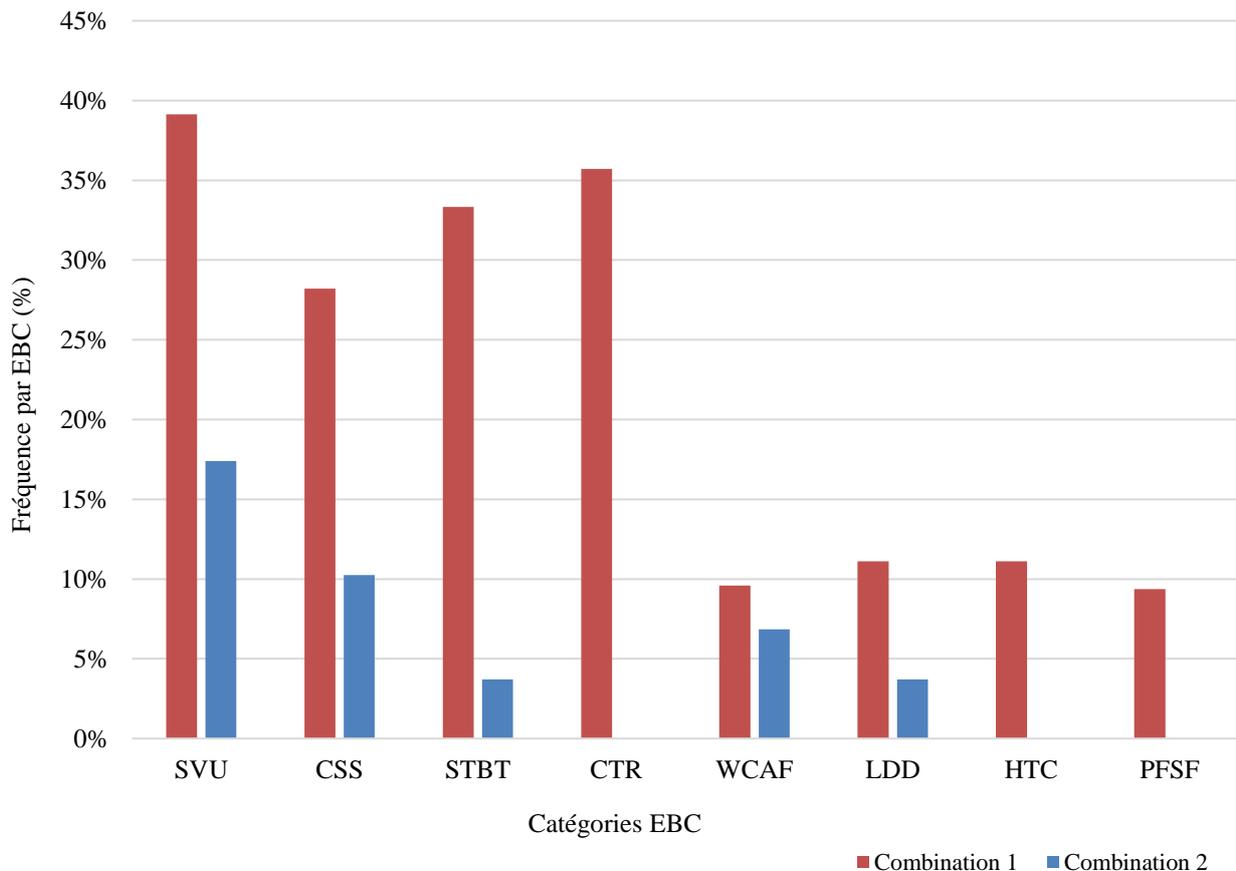
avec les autres traits est dilué. On peut en effet constater que WCAF comprend 12 occurrences associatives, contre 13 pour SVU et 15 pour CSS. Ces chiffres sont donc assez proches, malgré la différence de classement. L'ordre de ce classement fait cependant sens, car c'est justement afin d'estimer le potentiel combinatoire de chaque catégorie par rapport à sa représentativité que nous avons choisi de nous baser sur ce calcul. Établir un classement en fonction des nombres d'occurrences bruts aurait été impertinent, car plus une catégorie est étendue, plus elle a de chances d'apparaître en conjugaison avec d'autres au sein des définitions. De ce point de vue, il est étonnant qu'un trait peu représenté comme SVU, avec seulement 23 occurrences de catégories, compte 13 de ses occurrences en association avec d'autres traits, c'est-à-dire légèrement plus que WCAF, qui présente un nombre brut de 73 occurrences de catégories.

Pour finir, les traits LDD, HTC et PSF se combinent beaucoup moins que les autres. En outre, les seuls traits avec lesquels il leur arrive de se combiner sont des traits au fort potentiel. Il s'agit donc possiblement d'associations fortuites, dues à l'omniprésence des traits les plus importants. Il semblerait donc que ces sens soient plus exclusifs que les autres, puisque les sens qui les comportent ne font référence qu'à eux seuls. On peut toutefois remarquer que LDD entretient une interaction particulièrement représentative, et presque exclusive, avec SVU. Le tableau 10 ci-dessous résume les données combinatoires, à l'échelle globale, dans un même tableau, tandis que le graphe n° 11 permet une meilleure visualisation de ces données.

Figure n° 10 : Tableau de calcul du profil combinatoire global des EBC.

EBC	Combinaison 0		Combinaison 1		Combinaison 2		Total combinaisons	
SVU	10	43,48%	9	39,13%	4	17,39%	13	56,52%
CSS	24	61,54%	11	28,21%	4	10,26%	15	38,46%
STBT	17	62,96%	9	33,33%	1	3,70%	10	37,04%
CTR	9	64,29%	5	35,71%	0	0,00%	5	35,71%
WCAF	61	83,56%	7	9,59%	5	6,85%	12	16,44%
LDD	23	85,19%	3	11,11%	1	3,70%	4	14,81%
HTC	24	88,89%	3	11,11%	0	0,00%	3	11,11%
PFSF	29	90,63%	3	9,38%	0	0,00%	3	9,38%

Figure n° 11 : Graphe du profil combinatoire global des EBC.



La figure n° 11, en particulier, exacerbe le fait que la majorité des proto-sens ne se combine qu'avec un seul trait. On peut également remarquer que les traits qui ont davantage tendance à se conjuguer avec deux traits sont ceux qui ont les taux globaux les plus élevés.

L'interprétation définitive de ces données sera explicitée dans la quatrième section de ce volet, mais on peut déjà émettre quelques hypothèses. D'abord, il convient de remettre ces chiffres en perspective. L'objectif initial de cette analyse était d'évaluer le potentiel associatif des EBC, c'est-à-dire la fréquence avec laquelle les traits sémantiques du corpus se combinent avec d'autres traits. Les données observées indiquent que les traits les mieux représentés à l'intérieur du corpus ont tendance à être plus associatifs. Ceci n'est pas surprenant, car plus un trait apparaît dans un nombre important d'entrées lexicographiques, plus la probabilité que les monomorphèmes qui lui correspondent renvoient à d'autres traits est grande. Les scores de SVU et WCAF sont cependant assez étonnants. Le problème majeur de ce type d'approche est qu'il est difficile de se rendre compte de ce que l'on mesure exactement. Ces données reflètent-elles un potentiel d'attraction, ou simplement un potentiel de diffusion ? Dans ce dernier cas, la

figure n° 11 montrerait que SVU a tendance à être particulièrement diffus, alors que WCAF est beaucoup plus exclusifs dans ses apparitions.

6.2.3. Analyse des PAC

L'analyse du potentiel combinatoire des PAC suit le même protocole que pour les EBC, et la figure n° 12 ci-dessous se lit de la même manière que la figure n° 9.

Figure n° 12 : Tableaux des données combinatoires globales et détaillées des PAC.

SVU		
PAC	Combinaisons	Fréquence brute (%)
WCAF	23	38,98%
CSS	20	33,90%
LDD	11	18,64%
STBT	3	5,08%
CTR	2	3,39%
PFSF	0	0,00%
HCT	0	0,00%
Total	59	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	27	36,99%
1	33	45,21%
2	13	17,81%
Total combinaisons	46	63,01%

CSS		
PAC	Combinaisons	Fréquence brute (%)
SVU	20	37,74%
WCAF	12	22,64%
STBT	10	18,87%
LDD	4	7,55%
HCT	3	5,66%
CTR	2	3,77%
PFSF	2	3,77%
Total	53	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	37	46,84%
1	31	39,24%
2	11	13,92%
Total combinaisons	42	53,16%

LDD		
PAC	Combinaisons	Fréquence brute (%)
SVU	11	55,00%
CSS	4	20,00%
WCAF	4	20,00%
STBT	1	5,00%
CTR	0	0,00%
HCT	0	0,00%
PFSF	0	0,00%
Total	20	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	32	71,11%
1	6	13,33%
2	7	15,56%
Total combinaisons	13	28,89%

STBT		
PAC	Combinaisons	Fréquence brute (%)
CSS	10	45,45%
WCAF	5	22,73%
SVU	3	13,64%
CTR	2	9,09%
LDD	1	4,55%
HCT	1	4,55%
PFSF	0	0,00%
Total	22	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	37	66,07%
1	16	28,57%
2	3	5,36%
Total combinaisons	19	33,93%

CTR		
PAC	Combinaisons	Fréquence brute (%)
WCAF	2	18,18%
CSS	2	18,18%
SVU	2	18,18%
STBT	2	18,18%
PFSF	2	18,18%
HCT	1	9,09%
LDD	0	0,00%
Total	11	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	20	64,52%
1	11	35,48%
2	0	0,00%
Total combinaisons	11	35,48%

WCAF		
PAC	Combinaisons	Fréquence brute (%)
SVU	23	48,94%
CSS	12	25,53%
STBT	5	10,64%
LDD	4	8,51%
CTR	2	4,26%
HCT	1	2,13%
PFSF	0	0,00%
Total	47	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	106	74,65%
1	25	17,61%
2	11	7,75%
Total combinaisons	36	25,35%

HTC		
PAC	Combinaisons	Fréquence brute (%)
CSS	3	50,00%
WCAF	1	16,67%
CTR	1	16,67%
STBT	1	16,67%
SVU	0	0,00%
LDD	0	0,00%
PFSF	0	0,00%
Total	6	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	41	87,23%
1	6	12,77%
2	0	0,00%
Total combinaisons	6	12,77%

PAC	Combinaisons	Fréquence brute (%)
CSS	3	60,00%
CTR	2	40,00%
SVU	0	0,00%
WCAF	0	0,00%
STBT	0	0,00%
LDD	0	0,00%
HCT	0	0,00%
Total	5	100,00%

Combinaisons	Nombre d'occurrences	Fréquence brute (%)
0	44	89,80%
1	5	10,20%
2	0	0,00%
Total combinaisons	5	10,20%

Le profil combinatoire de *gr*- semble quelque peu différent lorsque l'on prend en compte les sens métasémiques. En effet, si l'ordre des tableaux est assez similaire entre les EBC et les PAC, les données chiffrées sont différentes. Avant tout, on peut souligner que SVU et CSS restent les traits qui comptent le plus d'associations conceptuelles, autant pour les EBC que pour les PAC. Les traits STBT et CTR d'une part, ainsi que les traits WCAF et LDD d'autre part, ont interverti leurs places. Les raisons de ces fluctuations ne sont pas évidentes ; en gardant en tête qu'il ne faut pas nécessairement attribuer de mécanisme motivé au moindre écart de valeur, il s'agit probablement d'une variation aléatoire. Quant aux résultats chiffrés, on peut remarquer qu'ils ne sont pas aberrants au vu des données des EBC, mais que les écarts sont plus marqués. Ainsi, jusqu'à 63,0% des occurrences de catégories de SVU, soit un peu moins des deux tiers, se combinent avec d'autres traits. Le pourcentage d'associations avec un seul trait est cependant un peu plus élevé. Les chiffres livrés par CSS sont également assez similaires aux données des EBC, mais plus de la moitié des occurrences sont associatives, ce qui représente sensiblement plus que les 38,46% de la figure n° 9. Le fossé qui sépare SVU et CSS du reste des traits est plus important pour les PAC, car le reste des tableaux est assez similaire entre les figures n° 9 et 12. En outre, le trio SVU-CSS-WCAF reste le plus important, car, encore une fois, si WCAF n'est pas compté parmi les traits les plus associatifs en termes de taux globaux, les autres traits le comptent parmi les plus influents.

Les traits PFSF et HTC présentent des chiffres encore une fois très bas, et ne se combinent jamais qu'avec un seul autre trait. Il paraît donc raisonnable de supposer que ces catégories conceptuelles ne sont pas phonesthétiques, au sens où leur existence à travers le lexique n'engendre pas de convergence sémantique notable. Le fait que presque aucune évolution n'apparaît entre les EBC et les PAC indique que leur existence, certes significative d'un point de vue statistique global, est probablement due à l'étymologie. Ces observations

vont dans le sens de l'hypothèse qu'avance Stefanowitsch (2002, notamment p. 15), concernant l'importance des racines communes dans la formation de ce que l'on perçoit comme de la phonesthésie. On pourrait, dans cette optique, distinguer des phénomènes résultants d'une ascendance étymologique commune, partagée par un certain nombre de monomorphèmes à l'intérieur de la métaclasse des mots en *gr-*, d'un autre type d'association forme-sens, qui, de son côté, serait motivée au niveau cognitif. Une réponse partielle à notre hypothèse de départ, concernant la distinction entre ces deux phénomènes, est donc apportée par l'analyse des combinaisons.

CTR présente un profil PAC beaucoup plus diffus, dans le sens où il se combine avec presque tous les traits, à l'exception de LDD, ce qui tranche avec le tableau des EBC. Il n'est ainsi pas absurde de supposer que ce trait sémantique a joué un rôle dans une partie des processus métasémiques à l'œuvre. Il est cependant difficile d'étayer ou de réfuter cette hypothèse simplement à partir de ces tableaux, car on voit bien que le taux global de CTR reste faible, et que, de fait, les chiffres brutes en termes de nombre d'occurrences sont d'autant plus faibles. Il se combine en fait deux fois seulement avec quasiment tous les traits sémantiques. Fait intéressant, le trait STBT affiche un taux global moins élevé que pour les EBC, mais son profil combinatoire est moins diffus, puisqu'il s'associe avec CSS dans 45,45% des cas, ce qui est sensiblement plus que les 36,36% précédents. Cela pourrait indiquer un rapprochement entre les deux traits sémantiques, puisqu'on passe de 4 à 10 associations. A l'échelle de l'ensemble du corpus, ce chiffre paraît toutefois assez faible, surtout si on le compare à l'évolution de la relation entre SVU et CSS, qui passe de 5 à 20 associations.

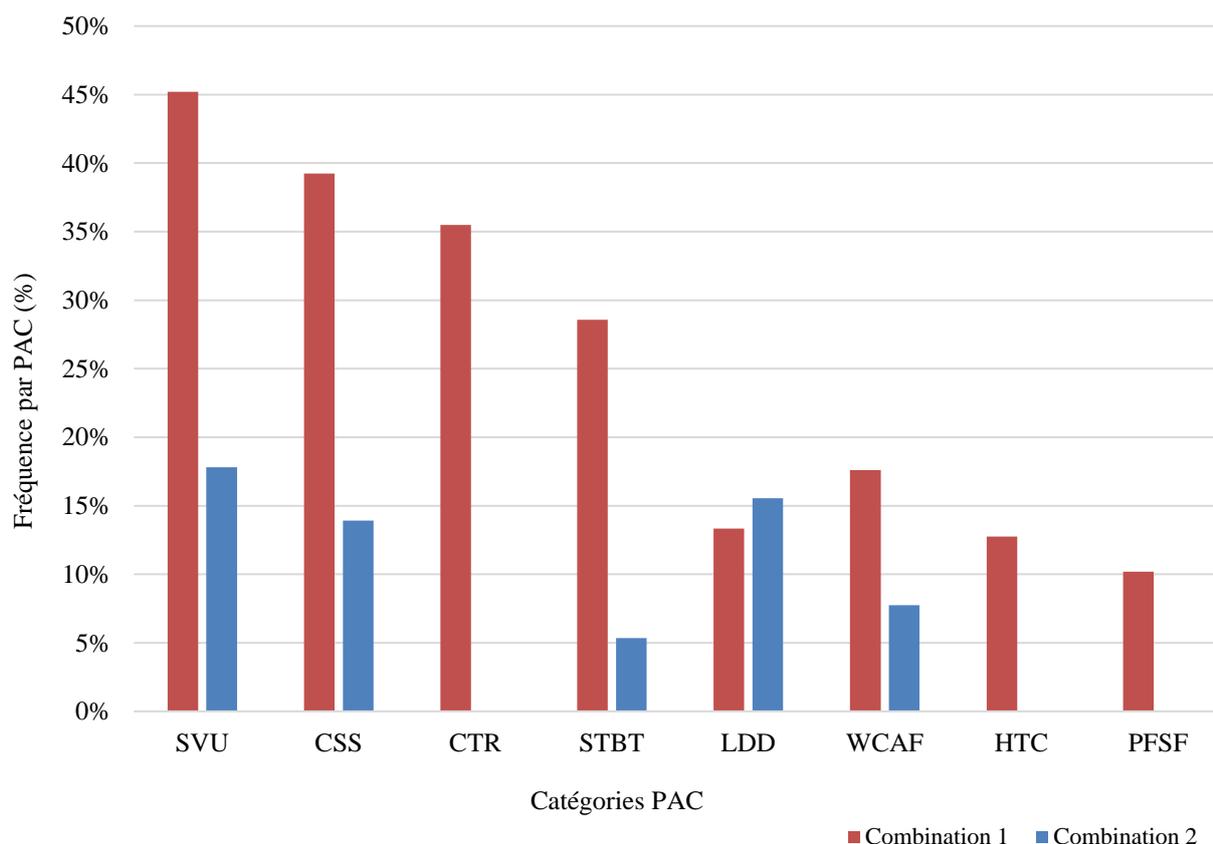
De manière générale, les liens entre SVU, CSS et WCAF sont les seuls à réellement prendre de l'importance avec l'inclusion des sens métasémiques, lorsque l'on observe les chiffres bruts. En effet, si, en termes de fréquences, les liens n'apparaissent pas forcément de façon évidente, c'est en partie parce que chaque pourcentage est le reflet de toutes les interactions à l'intérieur du corpus. Dès lors que l'on prête attention aux nombres de tokens, on peut observer des tendances au sein d'un nombre restreint de traits. Ainsi, les associations entre SVU et WCAF passent de 8 à 23, tandis que celles entre CSS et WCAF passent de 4 à 12. Bien que de tels calculs n'ont pas été conduits, il pourrait être intéressant de déterminer la variation entre les EBC et les PAC, en pourcentages, pour chaque couple de traits, ou du moins pour les associations les plus récurrentes. Cela pourrait apporter des éléments de réponse complémentaires quant à l'évolution des interactions entre traits EBC et PAC. La figure n° 13

ci-dessous résume les taux globaux par trait sémantique, sur le même principe que la figure n° 10.

Figure n° 13 : Tableau de calcul du profil combinatoire global des PAC.

PAC	Combinaison 0		Combinaison 1		Combinaison 2		Total combinaisons	
SVU	27	36,99%	33	45,21%	13	17,81%	46	63,01%
CSS	37	46,84%	31	39,24%	11	13,92%	42	53,16%
CTR	20	64,52%	11	35,48%	0	0,00%	11	35,48%
STBT	37	66,07%	16	28,57%	3	5,36%	19	33,93%
LDD	32	71,11%	6	13,33%	7	15,56%	13	28,89%
WCAF	106	74,65%	25	17,61%	11	7,75%	36	25,35%
HTC	41	87,23%	6	12,77%	0	0,00%	6	12,77%
PFSF	44	89,80%	5	10,20%	0	0,00%	5	10,20%

Figure n° 14 : Graphe du profil combinatoire global des PAC.



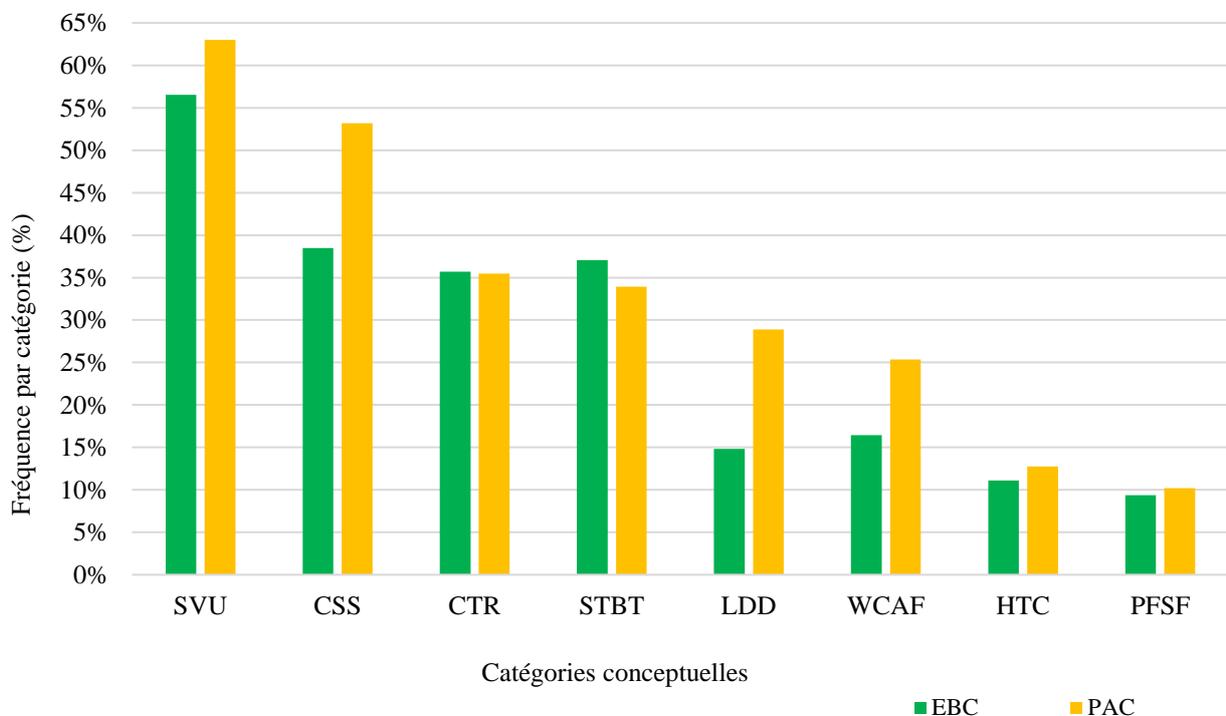
Pour clore cette section dévolue aux PAC, le graphe n° 14 permet une approche contrastive, car il met en saillance l'importance du nombre de traits associatifs par catégorie. Encore une fois, on peut clairement observer l'importance globale de CTR, malgré le fait qu'il

ne se combine qu'avec un seul autre trait à l'intérieur de ses sens porteurs. De manière générale, on ne remarque pas de différences majeures entre le taux de diffusion initial des traits sémantiques, et leur taux de diffusion avec intégration des métasens. Ce graphe fait tout de même ressortir l'étanchéité de PFSF et HTC, qui paraissent particulièrement réfractaire à toute association avec d'autres traits.

6.2.4. Comparaison et conclusion

À l'issue de cette analyse comparative, les données révèlent une tendance des traits SVU, CSS et WCAF à l'entre-soi, aux dépens de STBT et CTR, qui présentent des taux globaux dans la moyenne, mais qui ne se basent que sur des chiffres bruts assez faibles, et qui ne présentent pas d'évolution majeure entre EBC et PAC. Quant à HTC et PFSF, il semblerait que leurs profils associatifs largement moins diffus soient le reflet d'un potentiel combinatoire moindre, dont l'origine est probablement à chercher dans la documentation étymologique. La figure n° 15 ci-dessous met en contraste l'évolution entre l'analyse des proto-sens strictes et celle de l'ensemble des sens.

Figure n° 15 : Graphe comparatif des profils combinatoires des EBC et des PAC.



Ce graphe a été généré à partir des données rentrées dans les tableaux n° 10 et n° 13. Il permet de réaliser l'écart de valeurs entre les EBC et les PAC. On peut constater que, sans

surprise, les catégories SVU et CSS présentent un potentiel combinatoire, et donc possiblement un potentiel d'attraction, supérieur lorsque l'on prend les sens métasémiques en compte. Cette donnée est fondamentale, car elle sous-entend que les traits SVU et CSS pourraient être impliqués dans les processus métasémiques à l'œuvre dans l'évolution des monomorphèmes en *gr-* du corpus d'étude. Autrement dit, les associations forme-sens que la littérature spécialisée interprète comme une trace d'activité phonesthétique, et qu'elle associe au marqueur *gr-*, pourraient être définies par ces deux traits sémantiques. Il ne faut toutefois pas oublier qu'un tel graphe, au même titre que certaines des figures précédentes, n'est représentatif que des caractéristiques globales des traits sémantiques étudiés. C'est pourquoi il ne s'agirait pas de surinterpréter ce graphe-ci sans prendre en compte les précautions que nous avons énoncées plus haut. En particulier, ces données graphiques représentent des pourcentages basés sur la totalité des occurrences de catégories de chaque trait, ce qui a pour effet intrinsèque de léser les grands ensembles lexicaux comme WCAF, au profit d'autres ensembles plus modestes, en termes de fréquence brute, comme CTR ou STBT.

L'objectif de cette analyse était de comparer le profil combinatoire des traits sémantiques selon que les sens métasémiques étaient pris en compte ou non. Les données livrées par l'analyse des EBC nous renseignent sur le niveau de diffusion sémantique des traits en question, à travers les proto-sens, tandis que l'analyse des PAC nous indique comment ces données de départ évoluent au cours du temps. Plus exactement, elle nous permet de dresser un état des lieux de la situation globale du potentiel combinatoire des traits métasémiques. Nous n'insisterons jamais assez sur le caractère préliminaire de cette étude prospective, qui s'inscrit dans une perspective heuristique et exploratoire. En effet, l'une des solutions possibles pour estimer l'évolution de ces profils combinatoires au cours du temps serait de conduire des analyses par tranches chronologiques. Ce type d'approche pourrait avoir son intérêt, dans le cadre de futures recherches, mais il poserait des problèmes méthodologiques qu'il s'agirait d'aborder en tant que tels. Or, prendre le temps de poser ces problèmes et de proposer des méthodes pour les résoudre sortait largement du cadre de la présente étude. C'est pourquoi nous nous sommes contentés d'une approche plus globale d'un point de vue diachronique.

6.3. Co-émergences et post-attestations

6.3.1. Protocole de l'analyse et précisions sur les EBC

Si une analyse année par année, ou bien une étude par fourchettes chronologiques, n'a pas été menée, nous avons toutefois calculé la valeur des post-attestations. Plus précisément, l'écart chronologique entre l'apparition d'un monomorphème dans le lexique et le moment où l'un des sens métasémiques semble converger vers un trait sémantique qui lui ait associé, a été calculé pour chaque sens du corpus des PAC. Chaque sens a ainsi été inspecté, afin d'observer s'il émergeait en même temps que le monomorphème qui le porte, ou si ce dernier existait avant l'apparition de ce sens. On parle de co-émergence dans le premier cas, et de post-attestation dans le second, car le sens en question entre dans le lexique à une date postérieure à celle de l'apparition du monomorphème.

En ce qui concerne la méthodologie, il paraît judicieux de rappeler que le calcul des post-attestations ne s'applique qu'à l'intérieur des monomorphèmes, en ce sens que les homonymes ne doivent pas être comparés, étant donné qu'il s'agit de lexèmes distincts. En d'autres termes, l'étude des post-attestations n'aurait pas de sens pour les EBC, dans la mesure où tous les sens secondaires ont été éliminés. Les glissements de sens seraient pour ainsi dire invisibles. Autrement dit, les seules post-attestations visibles concerneraient des homonymes. Or il n'est pas pertinent de parler d'influence morphologique entre des monomorphèmes totalement différents, bien que partageant une même forme. La convention linguistique fait que ces mots sont fondamentalement distincts. On pourrait cependant dépasser cet état de fait et se demander si la forme d'un homonyme pourrait avoir une influence sur l'évolution sémantique d'un mot, notamment à cause du principe d'isomorphie (Haiman 1980 ; Waugh 1994). Toutefois, cette question revoie à des débats complexes ayant trait à la nature et la portée du symbolisme phonétique, dans une perspective cognitive de la construction du lexique, et sort donc du cadre empirique de la présente étude. Il est cependant important de préciser que, lorsque le lien étymologique entre deux monomorphèmes est plus éloigné ou moins bien renseigné, ceux-ci sont par contre considérés comme homonymes. Prenons l'exemple de *grill*, adj. (1) [1200] « of persons: fierce, harsh, cruel », dont la seule forme antérieure est *grill*, v.1 (1) [897] « to provoke, annoy, irritate, offend ». Bien que les deux monomorphèmes semblent entretenir une origine étymologique commune, celle-ci n'est pas attestée par les textes. En effet, *grill*, v.1 (1) proviendrait de la forme **gryllan* en vieil anglais, dont le descendant n'est attesté qu'à partir du moyen anglais, tandis que *grill*, adj. (1) pose le même problème, attendu que sa racine n'est

pas attestée au-delà du moyen anglais, même si elle pourrait remonter au vieil anglais **grylle*, dont le OE suggère qu'il pourrait s'agir d'un cognat de **gryllan*. En raison de ce déficit d'information, ces deux formes sont traitées comme des homonymes.

Par ailleurs, les monomorphèmes partageant un lien de conversion ne sont pas considérés comme des homonymes, puisque l'un provient de l'autre, et qu'ils partagent par conséquent la même origine étymologique. Ils sont donc susceptibles d'être pris en compte dans les calculs de post-attestation. Là réside toute la finesse de l'analyse. Les sens qui tombent dans la même catégorie que leur post-attestation ne sont pas compris dans l'analyse, puisque l'objectif est de comparer l'évolution sémantique des différentes catégories. C'est le cas de *grame*, v. (1) [1175] par exemple, qui provient de *grame* (adj.) [893]. Étant donné que les deux monomorphèmes correspondent à la catégorie WCAF, *grame*, v. (1) est considéré comme une co-émergence de sa catégorie, et non comme une post-attestation de *grame* (adj.).

Le calcul différentiel aurait été impossible à mener si les fourchettes de dates avaient été conservées en l'état. Typiquement, les renseignements du type « 1579-85 » n'aident vraiment pas à la manœuvre. C'est pourquoi nous avons choisi d'appliquer la même méthode de filtrage que pour les courbes de progression et les analyses combinatoires, c'est-à-dire que seule la date la plus haute est conservée, tandis que la mention *OE* est considérée comme équivalente à l'an 700. De même, la mention *c.*, comme dans « c. 893 », a tout simplement été supprimée. Cette dernière n'apporte de toute façon pas d'information pertinente pour l'étude.

6.3.2. Analyse globale des PAC

La feuille de calcul dédiée à l'analyse des post-attestations comprend la liste de toutes les occurrences de catégories du corpus, rangées par PAC, puis dans l'ordre chronologique à l'intérieur de chaque PAC. Dans un premier temps, nous avons procédé au décompte des co-émergences et des post-attestations pour chaque trait. Le résultat de ce décompte est présenté dans la figure n° 16. Dans un second temps, nous sommes passés à une lecture plus fine. Pour chaque sens de l'assemblage, la forme étudiée était recherchée dans la feuille de calcul comprenant tous les sens associés à leurs traits sémantiques, rangés dans l'ordre alphabétique. Nous ne rentrerons pas davantage dans les détails de la manipulation Excel, mais précisons simplement que l'opération consiste à vérifier le parcours lexicographique de chaque sens. Si le sens en question est un proto-sens non converti, son association au trait sémantique sera considérée comme une co-émergence. En revanche, s'il s'agit d'un sens métasémique ou d'une

conversion, l'association sera considérée comme une co-émergence si la catégorie concernée est la même que celle dans laquelle il apparaît, et comme une post-attestation dans le cas contraire. Les sens qui émergent après l'apparition de leur monomorphème ont ensuite été mis de côté, de sorte que l'on puisse disposer d'un inventaire qualitatif des sens post-attestés. Le détail du traitement de ces sens sera développé plus tard, lorsque nous aborderons les intervalles chronologiques et les interprétations que l'on peut en faire (voir figure n° 17).

D'une certaine manière, on peut considérer cette méthode d'analyse comme un complément des fréquences brutes par catégories, qui avaient été calculées à la fois pour les EBC (voir figure n° 3) et les PAC (voir figure n° 4). Les chiffres que nous présenterons par la suite ne doivent cependant pas semer la confusion par rapport à ces fréquences brutes, car il ne faut pas confondre la distinction entre EBC et PAC d'une part, et la distinction entre co-émergences et post-attestations d'autres part. En effet, la nature du corpus, dans le dernier cas, est restrictive, car il ne se base que sur l'inventaire des PAC. En particulier, il ne faut pas oublier que les conversions ont été prises en compte à l'intérieur de ce corpus. Les chiffres des co-émergences par trait sémantique sont donc susceptibles d'être sensiblement différents des fréquences brutes des EBC, bien que dans les deux cas, ce sont les sens premiers qui sont comptabilisés. La nuance tient précisément au fait que l'analyse des post-attestations prend en compte ce réseau de conversions, que nous avons choisi d'éliminer pour l'étude des EBC. En toute logique, le nombre de post-attestations par trait doit donc être plus grand que le nombre total d'occurrences de catégories. En outre, n'oublions pas qu'une partie des sens métasémiques s'avère être des co-émergences, dans le sens où ils ne divergent pas de leur trait sémantique original. En effet, seuls les sens métasémiques qui passent d'une catégorie à une autre peuvent être considérés comme de véritables post-attestations.

Figure n° 16 : Nombre global de co-émergences et de post-attestations par trait sémantique (en nombre d'occurrences et en %).

PAC	Co-émergences (occurrences)	Co-émergences (%)	Post-attestations (occurrences)	Post-attestations (%)
SVU	28	38,36%	45	61,64%
CSS	46	58,23%	33	41,77%
WCAF	87	61,27%	55	38,73%
CTR	21	67,74%	10	32,26%
STBT	40	71,43%	16	28,57%
LDD	33	73,33%	12	26,67%
HCT	43	91,49%	4	8,51%
PFSF	47	95,92%	2	4,08%

Ce tableau comprend les huit traits sémantiques, rangés dans l'ordre décroissant de leur pourcentage de post-attestations. Ce pourcentage est calculé à partir du nombre brut de post-attestations, divisé par le nombre total d'occurrences de catégories. Il indique que les traits SVU, CSS et WCAF sont ceux qui contiennent le plus de post-attestations, c'est-à-dire qu'une part plus importante des sens qui renvoient à leurs traits sémantiques respectifs sont apparus *a posteriori*. Autrement dit, il existe une grande proportion des sens qui sont associés à ces trois catégories conceptuelles, qui en sont venus à les rejoindre avec le temps, mais qui n'étaient pas présents initialement. Ce tableau est à examiner en parallèle des tableaux n° 10 et n° 13, ainsi que des graphes n° 11 et n° 14 associés, car il propose une explication possible des profils combinatoires que nous avons observés précédemment. Les traits SVU, CSS et WCAF sont, en effet, à la fois ceux qui présentent les profils combinatoires les plus forts, que ce soit par leur taux associatif global (pour SVU et CSS) ou leur omniprésence à travers le profil des autres traits du corpus (pour WCAF), et ceux qui présentent la plus grande proportion de sens post-attestés. On prendra bien note que ce sont les chiffres proportionnels, exprimés ici en pourcentages, qui sont pris en compte, et non les valeurs brutes. Ce n'est donc pas la taille générale des ensembles qui entre en ligne de compte, ou qui est susceptible d'orienter les résultats de manière directe. Et pour cause, dans le cas contraire, ce serait bien sûr WCAF qui arriverait en tête de liste. Dans la mesure où ce qui nous intéresse ici est de déterminer la capacité de ces traits à influencer la construction du lexique, via l'apparition de lexèmes nouveaux, ainsi que l'émergence et l'évolution des sens, le potentiel d'attraction des traits sémantiques est une donnée fondamentale dans l'analyse des résultats. Le tableau n° 16 permet

donc, en conjugaison des profils combinatoires, de déterminer ce potentiel d'attraction, qui, à terme, vise à délimiter les sens phonesthétiques proprement dits.

La grande proportion de sens post-attestés qu'affiche SVU permet de comprendre, en partie, pourquoi ce trait, qui apparaît parmi les EBC les moins représentés, donne suite à une PAC qui, en plus d'être particulièrement présente, semble se diffuser à travers le corpus. CSS et WCAF, qui présentent chacun un nombre important d'occurrences de catégories, ainsi qu'une courbe de progression particulièrement ascendante, semblent également comprendre un nombre important de sens post-attestés. Le phénomène de convergence des résultats est moins saillant pour ces deux traits que pour SVU, car c'est la majorité des sens de celui-ci qui sont post-attestés, et parce que les profils combinatoires de CSS et WCAF sont moins impressionnants. Mais il n'en reste pas moins que l'importance de leur potentiel d'attraction semble se confirmer. Notons que l'ordre suivi par la figure n° 16 et celui de la figure n° 13 sont très similaires, exception faite de WCAF, dont l'écart qui caractérise son comportement combinatoire a déjà été expliqué. Quant à HTC et PFSF, on peut constater que l'écrasante majorité des sens qui les définissent co-émergent avec les premières attestations de leurs occurrences de catégories. Il s'agit d'un indice de plus, qui laisse penser que ces deux catégories ont un potentiel d'attraction très limité, puisque rares sont les monomorphèmes qui acquièrent leurs traits. Pour finir sur LDD, on peut remarquer que, sur la même lancée que l'EBC, la PAC se combine presque exclusivement avec SVU, ce qui fait de LDD un trait intimement associé à SVU. On peut dès maintenant supposer que LDD n'est, en réalité, qu'un spécificateur de SVU, dans la mesure où les sons décrits son souvent graves. Ce n'est bien sûr pas le seul champ d'action de LDD, mais les données montrent qu'il s'agit d'une hypothèse à creuser. Les données métasémiques apporteront un autre regard, car les éventuelles métaphores ou métonymies permettant de passer d'un trait à l'autre pourront, si elles existent, être mises en évidence.

6.3.3. Intervalles chronologiques

Après cette approche globale, attardons-nous maintenant sur les sens post-attestés, au cas par cas. La figure n° 17 ci-dessous présente les données statistiques des sens post-attestés de chaque trait sémantique. La colonne *attestation du MMP (monomorphème)* renvoie aux dates d'attestation du monomorphème originel, c'est-à-dire au proto-sens, tandis que la colonne *attestation de la PAC* renvoie aux dates d'attestation des sens post-attestés correspondant aux monomorphèmes en question. Si le détail des sens n'apparaît pas, l'écart chronologique entre la première attestation du sens porteur du trait dans l'OED, et la première attestation du

monomorphème correspondant, a été calculé. Ce sont les post-attestations que nous nous proposons d'analyser, et non les co-émergences, précisément parce que ces dernières ne sont pas susceptibles de laisser des indices de changement sémantique, mais au contraire, elles tendent à souligner une forme de conservatisme sémantique.

Figure n° 17 : Émergence des PAC.

PAC		Attestation du MMP	Attestation de la PAC	Intervalle chronologique
SVU	Moyenne	1165,86	1604,33	443,41
	Écart-type	362,25	209,03	335,51
CSS	Moyenne	1376,73	1656,33	279,61
	Écart-type	267,08	173,72	195,07
WCAF	Moyenne	1229,29	1609,78	380,49
	Écart-type	372,27	234,74	360,56
CTR	Moyenne	1304,4	1500,00	195,60
	Écart-type	259,67	141,29	141,29
STBT	Moyenne	1141,81	1528,81	450,00
	Écart-type	392,84	245,8	364,43
LDD	Moyenne	1316,00	1644,75	328,75
	Écart-type	298,76	204,27	324,24
HTC	Moyenne	1386,00	1720,25	334,25
	Écart-type	478,89	155,06	474,76
PFSF	Moyenne	1000,00	1488,50	488,50
	Écart-type	424,26	34,65	389,62

L'objectif d'une comparaison entre ces tableaux est de faire ressortir les différences entre les multiples traits sémantiques, en termes d'émergence lexicale. Dans cette perspective, le calcul des intervalles chronologiques vise à estimer l'ampleur de l'attraction sémantique de chacun des traits au cours du temps. Deux interprétations sont cependant possibles. La première est de considérer que plus les sens acquièrent le trait de manière tardive, plus l'influence sémantique de ce trait perdure dans le temps, puisque cela suppose qu'il existe des sens qui continuent de converger vers ce trait, bien après l'apparition de leur monomorphème. La seconde interprétation considère que ce trait n'est pas aussi actif qu'un autre, puisque son influence met plus de temps à se déclencher. La raison majeure pour laquelle ces données sont si difficiles à interpréter est justement qu'une grande partie des mécanismes lexicaux et sémantiques responsables des phénomènes que l'on observe reste inconnue. Il est, en outre,

complexe de valider ou d'invalider nos hypothèses simplement à partir de cette lecture globalisante, que représente l'analyse des catégories conceptuelles.

Les intervalles chronologiques de SVU sont en moyenne de 443,41 ans, tandis que la moyenne des dates à laquelle les sens revoyant à ce trait sont attestés est de 1604,33. En effet, la plupart des dates post-attestées de SVU sont assez tardives. Notons cependant que ceci est vrai pour la plupart des catégories, ce qui n'est pas surprenant, étant donné qu'une bonne partie des sens post-attestés est d'origine métasémique. Or, comme l'indique la figure n° 7, la courbe de progression de ces sens au cours du temps est plus marquée à partir des XV^e-XVI^e siècles. De manière générale, ces données indiquent que les phénomènes de convergence, supposés par l'analyse des profils combinatoires, adviennent majoritairement sur une période plutôt tardive. Notons que les écarts-types des dates de post-attestation ne sont pas gigantesques au vu de la moyenne, ce qui est plutôt bon signe. Ce chiffre indique effectivement que cette moyenne renvoie une image plutôt fidèle de la réalité, étant donné que l'écart-type augmente lorsque les valeurs de l'échantillon sont très différentes les unes des autres. En d'autres termes, on peut ici soutenir l'idée que la plupart des changements sémantiques que l'on observe ont lieu sur une période relativement courte. Cette fenêtre de tir étroite, couplée aux dates tardives, laisse entendre que le trait SVU a dû commencer à exercer une influence phonesthétique, si influence il y a, plutôt récemment. Il est donc tout à fait envisageable que ce phénomène ne soit encore que dans sa lancée. Cette dimension n'a pas été directement abordée jusqu'ici, car c'est une variable encore plus complexe à déterminer, mais l'âge d'un phonesthème, si phonesthème il y a, est une question qui devra être posée à un moment ou à un autre. En effet, dans l'optique d'une hypothétique étude approfondie d'un ensemble plus large de marqueurs submorphémiques, on pourrait raisonnablement penser que tous les phonesthèmes existant ne sont pas apparus au même moment. Dès lors, il sera possible de repérer des schémas récurrents, qui pourront éventuellement nous indiquer l'âge absolu d'un phonesthème. Pour le moment, rien de cela n'est rendu possible, précisément parce que la masse de données accumulées n'est pas suffisante. Nous ne disposons donc pas d'assez de recul pour juger objectivement de la valeur interprétative des données qui commencent à ressortir.

La moyenne des dates des sens post-attestés de CSS est de 1656,33, c'est-à-dire dans les mêmes ordres de grandeur que pour SVU. La moyenne des intervalles chronologiques, en revanche, est plus petite, ce qui tend à montrer que les monomorphèmes dont les sens convergent vers CSS le font dans un délai plus court que dans le cas de SVU. Encore une fois, on peut soit considérer que CSS a tendance à provoquer des réactions de convergence de

manière plus efficace, soit que ce trait perdure moins dans le temps que SVU. Notons que les valeurs moyennes obtenues pour CSS sont plus représentatives de l'ensemble, car les écarts-types indiquent qu'elles sont moins dispersées. Autrement dit, les datations et les intervalles chronologiques sont plus homogènes que chez SVU. Il est cependant impossible d'affirmer que l'on a affaire à un groupement de convergences, car il est clair que les confluences sémantiques s'étalent sur plusieurs siècles, et qu'elles apparaissent progressivement.

WCAF présente un profil plus proche de celui de SVU, avec une moyenne de datations presque identique (1609,78, contre 1604,33 pour SVU). La moyenne des intervalles chronologiques est, par contre, plus petite (195,60, contre 443,41 pour SVU). CTR représente un échantillon bien plus réduit, mais on peut tout de même remarquer que la moyenne des intervalles chronologiques est assez faible. Celle du trait STBT est beaucoup plus large, mais cet écart moyen est probablement dû à la présence d'un 0 dans le tableau de calcul. En effet, il peut arriver qu'un sens soit considéré comme post-attesté, même s'il apparaît à la même date que l'émergence du monomorphème. En l'occurrence, le monomorphème *graisle* émerge en association avec les traits SVU et WCAF, via le sens *graisle*, v. (1a) [1522]. Le sens *graisle*, v. (2) [1522] lui est ainsi postérieur. Dans ce cas précis, nous nous sommes basés sur la numérotation de l'OED pour prendre une décision. En définitive, ce sont les traits HTC et PFSF qui se comportent différemment du reste des catégories, puisque le premier présente une moyenne de datations sensiblement plus tardive que les autres, tandis que celle du second est plus précoce. Les intervalles chronologiques étant globalement assez larges, et compte tenu de la taille minimale des deux échantillons, on peut considérer que ces données n'ont pas beaucoup de valeur. On aura également remarqué que les deux sens affichés par PFSF ont des amplitudes sémantiques très larges, dans le sens où *great* et *grain* peuvent être très polysémiques, surtout lorsqu'on prend en compte leurs sens obsolètes.

6.3.4. *Evolution, divergence et convergence : conclusion partielle*

Il convient de dresser un bilan des analyses successives qui ont été menées jusqu'ici. Les résultats livrés par l'étude des EBC représentent un indice de la formation de nouveaux monomorphèmes, qui co-émergent avec leur trait initial ; on pourrait parler de productivité lexicale des traits. L'étude des PAC, quant à elle, est un indice de convergence ou de divergence sémantique des sens par rapport aux traits : on pourrait parler de potentiel de convergence phonesthétique. Les questions adressées en filigrane sont : les mots en *gr-* convergent-ils ? Le profil de la métaclasse suggère-t-il une influence cognitive de *gr-* sur le lexique ? Rappelons

d'abord que 12,12% des sens sont assurément non-phonesthétiques. Cela ne veut pas dire que le reste l'est forcément, mais qu'il existe, dans tous les cas, une partie non-négligeable de l'ensemble des sens qui est hors de propos. Ceci étant dit, peut-on affirmer avec certitude que l'on observe des phénomènes de convergence ? On observe en effet une différence assez nette entre des traits sémantiques avec un profil combinatoire attractif, et qui présentent de nombreux sens post-attestés, d'un côté, et des traits au comportement combinatoire largement isolant et qui ne comprennent quasiment que des sens co-émergeants, de l'autre. On peut supposer que la surreprésentation statistiquement significative de ces derniers, associés à l'élément *gr-* dans la littérature, est due à des origines étymologiques communes. Il a en effet été soulevé (Stefanowitsch 2002) que la phonesthésie pouvait être le reflet d'un stock d'étymons renvoyant à une forme ancestrale commune. Notre hypothèse est que ce phénomène ne représente qu'une partie de la solution, et que la notion de motivation a un rôle prégnant à jouer. De ce point de vue, on peut à présent poser l'hypothèse selon laquelle les catégories WCAF, SVU et CSS seraient le fruit d'une phonesthésie motivée, tandis que les catégories HTC et PFSF résulteraient, au contraire, de racines communes. Avant de passer à une analyse diachronique plus fine, terme à terme, des changements sémantiques des mots en *gr-*, on a donc déjà en main des éléments permettant de confirmer l'hypothèse d'une évolution motivée des lexèmes de la métaclasse.

Les courbes de progression, ainsi que les données chronologiques des post-attestations, suggèrent que l'association phonesthétique de la séquence /gr/ a dû apparaître assez tôt. C'est le trait WCAF qui, vraisemblablement, a émergé en premier, suivi de CSS et SVU. L'ascension de SVU est particulièrement impressionnante, surtout lorsque l'on compare les courbes de l'EBC et de la PAC. Si l'on ajoute le fait que ce trait présente un profil combinatoire étonnamment élevé et qu'il comprend de nombreuses post-attestations, on peut se demander s'il ne s'agit pas d'un trait phonesthétique récent et en devenir, amené à se développer davantage. Quoi qu'il en soit, il est possible que les traits HTC et PFSF aient joué un rôle dans l'évolution primitive de *gr-*, c'est pourquoi ce sont les données étymologiques qui pourront, en partie, faire la lumière sur cette question. Dans la mesure où ces traits semblent se combiner de préférence avec CSS, autant pour les EBC que les PAC, on pourrait par exemple se demander s'il n'existe pas un lien métasémique entre tous ces traits.

Dans la mesure où l'on parle de convergences, on peut aussi se poser la question suivante : observe-t-on des phénomènes de divergence ? On peut imaginer que des traits qui ont été motivés – au sens de motivation phonesthétique – à un moment de l'histoire de la langue,

ont perdu cette motivation au cours du temps. Dans cette optique, les sens STBT, CTR et LDD posent question. Si leur comportement combinatoire n'est pas aussi tranché que celui des autres catégories, c'est soit parce qu'ils s'avèrent effectivement non-phonesthétiques, soit qu'il s'agit de traits phonesthétiques en perte de vitesse. Le fait que les moyennes des dates d'attestation de STBT et CTR sont plus hautes que les autres traits pourrait aller dans ce sens. Les ordres de grandeur dont nous parlons sont toutefois trop petits pour être trop catégoriques à ce propos. Enfin, bien que cet aspect n'ait pas été traité, on pourrait également s'interroger sur l'éventuelle observation d'une corrélation entre la proportion de lexèmes obsolètes ou non-conventionnels d'un trait, et son profil combinatoire, ou sa proportion de post-attestations.

6.4. Identification de l'organisation métasémique de *gr-*

6.4.1. Présentation du protocole

La présente section vise à identifier les processus métasémiques responsables des évolutions sémantiques que l'on observe chez les mots en *gr-*. Il s'agit de distinguer les évolutions sémantiques entraînées par des processus de métaphore, de métonymie, de généralisation ou de restriction sémantique. L'une des hypothèses que nous pouvons poser concerne le rôle d'une éventuellement métaphore animale. En effet, on peut supposer que le trait STBT résulterait d'une surreprésentation des descripteurs d'animaux, en raison d'une métaphore fondamentale, qui subsiste à travers tout le corpus. Cela expliquerait pourquoi ce trait se combine avec beaucoup d'autres traits, alors que sa représentation globale est parmi les plus faibles. Par manque de temps, une sélection de seulement deux catégories a cependant dû être appliquée. Les deux traits qui seront décortiqués sont SVU et HTC, car ils semblent être des archétypes antagonistes des phénomènes de combinaison conceptuelle que nous avons tentés d'identifier plus haut. En effet, alors que SVU se combine beaucoup et comporte de nombreuses post-attestations, HTC se combine peu et ne comporte presque pas de post-attestations. Nous avons choisi de mettre de côté l'étude des autres traits pour d'éventuelles recherches à venir.

6.4.2. SVU

Le trait SVU est en fait encodé sous la forme de différents types de son. On peut distinguer trois types de sons principaux, dont l'alternance résulte, pour un nombre important de cas de figure, d'évolutions métasémiques de type métaphorique. Il s'agit (1) de sons

vocaliques humains, (2) de son vocaliques non-humains (animaux), et enfin (3) de sons non-vocaliques, le plus souvent de sons désagréables provoqués par le frottement d'une surface, ce qui implique une combinaison avec le trait CSS. Les sons vocaliques humains peuvent être signifiants ou non-signifiants. La première sous-catégorie regroupe tout type de production signifiante, qu'il s'agisse de paroles, d'un chant, etc. La seconde regroupe des sons émis par des humains, mais qui ne font pas partie du système linguistique ; il s'agit, le plus souvent, de sanglots, de pleurs, ou de bruits plus ou moins imitatifs, généralement adossés à une métaphore animale. Dès lors, on comprend déjà mieux la fréquence des interactions entre SVU, CSS et STBT, autant pour les EBC que les PAC. Précisons enfin que les sons vocaliques ne sont pas toujours décrits comme explicitement humains ou animaux, et le doute peut persister. C'est notamment cette ambivalence qui démontre la profondeur de la métaphore. Lorsque les sens métasémiques contiennent un autre trait que SVU, sans association avec celui-ci, les entrées en question dénotent généralement une activité susceptible de dégager un son désagréable ou discordant, même si le son n'apparaît pas *verbatim* dans la définition.

Le choix a été fait de ne pas détailler les évolutions métasémiques de chaque sens, mais de revenir sur les principaux types de métasémies à l'œuvre. De manière générale, nous tenterons de nous concentrer sur les liens métasémiques récurrents, ainsi que sur les évolutions qui entraînent un lien avec d'autres traits sémantiques. SVU comprend un total de 16 métaphores, pour 6 métonymies. Commençons par les premières, avant d'aborder les secondes. Chronologiquement, le premier sens de SVU est *groan*, v. (1a) [700] « to breathe with a deep-toned murmur; to utter a low deep sound expressive of grief or pain ». Il s'agit donc de produire un son non-signifiant. Le premier sens métasémique répertorié, *groan*, v. (3a) [1609] « to utter with groans », dénote la production d'un discours éventuellement signifiant dont le contenu est parasité de sons. Il s'agit donc d'une métaphore : on compare le discours au son, car il en est parasité. Vient ensuite *groan*, v. (1c) [1642] « to breathe (one's life, soul) away or out in groaning; similarly, to groan one's heart out », qui revient donc à l'idée initiale de son non-signifiant. Le sens *groan*, v. (3b) [1762] « to bewail, lament » est une métaphore du sens précédent, puisque l'on compare la qualité de l'émotion à la qualité du son ou de la respiration bruyante, qui est décrite comme négative, plaintive, en souffrance. Dernière métaphore de ce monomorphème, le sens *groan*, v. (1d) [1816] « to talk in a groaning voice, grumble », où l'on compare la voix au son plaintif. Le proto-sens *grunt*, v. (1a) [725] « of a hog: to utter its characteristic low gruff sound ; also of other animals and of persons (with conscious allusion to the pig): to utter a sound resembling this » renvoie à un bruit ou un cri animal, tandis que le

sens métasémique *grunt*, v. (2a) [1325] « to utter a similar sound, expressive of discontent, dissent, effort, fatigue, etc.; to grumble, murmur » représente un glissement de sens métaphorique. Le locuteur est effectivement comparé à l'animal. Le point de comparaison n'est ici pas clair. Le porc serait, d'une manière ou d'une autre, associé à une mauvaise humeur. Dans *grunt*, v. (1b) [1340] « to groan », le son de l'humain est comparé à celui de l'animal. Poursuivons avec le mot *grede*, v. (1a) [1000] « to cry, cry out, shout; to wail », qui représente un bruit ou un cri humain non-signifiant. Le sens dérivé *grede*, v. (1c) [1300] « to utter (a cry), sing (a song) » est plus difficile à identifier que les cas précédents, car il s'agit simplement d'un léger glissement de sens, où l'on passe d'un son humain non-signifiant à un son humain signifiant, éventuellement complété d'une composante lyrique. Il s'agit probablement d'une métaphore, où la parole rappelle un cri. Une seconde métaphore du même type apparaît avec le sens *grede*, v. (3) [1305] « to announce with a loud voice; to proclaim, publish; to proclaim (a person) to be (something) », où le contenu sonore est amplifié. L'un des ponts métasémiques possibles entre la notion d'animal et celle de manifestation somatique est exemplifié par le proto-sens *growl*, v.3 (1a) [1380] « of the bowels: to rumble », qui appartient à la fois à STBT et à SVU. Son premier sens métasémique est *growl*, v.3 (1b) [1671] « of an animal: to utter a low guttural sound, expressive of rising anger », où c'est la notion de son qui semble être le point de comparaison. Le grognement animal est ainsi métaphoriquement comparé à des gargouillements d'estomac. En outre, ces gargouillements sont visiblement associés à de la colère. Cette première métaphore débouche sur une seconde, qui semble typique de *gr-*, qui consiste en une comparaison humain-animal, avec *growl*, v.3 (2a) [1706] « of persons: to murmur angrily ». Cette métaphore humain-animal est également présente à travers l'évolution de *groin*, v.1 (1a) [1400] « of animals: to grunt; to growl », qui donne, plus tard, *groin*, v.1 (2) [1400] « of persons: to grumble, murmur », ce qui laisse, encore une fois, transparaître le point de comparaison de mauvaise humeur. Dans ces contextes précis, WCAF semble agir comme un pont métasémique entre STBT et SVU. Le trait CSS a, lui aussi, tendance à être assez présent dans les sens métasémiques, comme le laissaient déjà présager les données des sections précédentes. Le proto-sens *graisle*, v. (1a) [1522] « to make a harsh noise, esp. repeatedly; to creak, rattle » en est un exemple typique. Ainsi, le sens métasémique *graisle*, v. (2) [1522] « to gnash or grind (one's teeth); also in extended use of a boar grating or scraping with its tusks », dérivé d'un son non-vocalique, débouche sur un son somatique, en l'occurrence le fait de grincer des dents. La métaphore consiste à comparer le son des dents au bruit physique naturel non-humain. On peut, en outre, observer, dans un second temps, une extension au domaine animal.

Le sens *grunt*, v. (3a) [1400] « to grind (the teeth) », quant à lui, ne s'inscrit pas dans la même lignée métasémique. Nous avons là affaire à une métonymie, dans la mesure où l'on passe du son, désigné par le proto-sens *grunt*, v. (1a) [725] « of a hog: to utter its characteristic low gruff sound ; also of other animals and of persons (with conscious allusion to the pig): to utter a sound resembling this », à l'activité articulatoire, ou bien du son à un autre type de son, selon le point de vue adopté. Notons que ce changement sémantique introduit les traits CSS et STBT. De manière générale, les métonymies ont tendance à dévier vers d'autres traits, et ce davantage que les métaphores. On a, par exemple, l'introduction de WCAF dans *growl*, v.3 (3) [1774] « to be dully painful », sens métasémiquement dérivé de *growl*, v.3 (2a) [1706] « of persons: to murmur angrily », lui-même une métaphore du proto-sens *growl*, v.3 (1a) [1380] « of the bowels: to rumble ». Il s'agit d'une métonymie car on substitue la source du mécontentement à l'expression orale du mécontentement. Pour revenir à CSS, on peut remarquer que la métaphore de *graisle*, décrite au paragraphe précédent, est ensuite dérivée par métonymie, puisque le sens *graisle*, v. (1b) [1786] « to grate or scrape against something, esp. repeatedly » consiste à remplacer le son par l'activité qui produit ce son. Pour terminer sur SVU, précisons que métaphores et métonymies peuvent très bien s'enchaîner au cours du temps. On peut ainsi mentionner le cas de *grit*, v. (1) [1762] « to produce a grating sound such as is caused by the crushing of grit; to move with such a grating noise », proto-sens qui dérive sur *grit*, v. (4a) [1797] « to grind or grate (the teeth) », métaphore du précédent, puisqu'on compare le son des dents au bruit physique naturel non-humain. Plus tard, un son humain, éventuellement signifiant, en est tiré : *grit*, v. (4b) [1900] « to utter with gritting of the teeth ». Il s'agit cette fois d'une métonymie où le bruit des dents est substitué à la production de paroles qui y est associée.

6.4.3. HTC

HTC comprend un ensemble de 11 métaphores, 3 métonymies, 3 restrictions et une généralisation, ce qui n'est pas excessivement différent de SVU. Les changements métasémiques en question sont cependant bien différents lorsqu'on entre dans le détail. Le sens *greet* v.1 (1) [725] « in various senses which did not survive beyond Old English: to approach, come up to; to begin upon, begin to treat or handle, take in hand » présente un exemple de restriction ou spécification sémantique, via le sens *greet* v.1 (2) [893] « to assail, attack », puisque l'on passe d'un type d'interaction humaine à un autre, plus restreint. Un deuxième changement métasémique, de type métonymique, peut ensuite être observé, où l'on substitue

une approche sociale (verbale ou non-verbale) à l'approche physique : *greet* v.1 (3c) [700] « to salute with words or gestures; transferred to receive at meeting or arrival with some speech or action (whether friendly or otherwise) in lieu of salutation ». Le problème de datation introduit cependant une dose d'incertitude supplémentaire. Les métaphores de HTC fonctionnent à peu près toutes sur le même principe : elles concernent presque toutes la notion de préhension, qui passe d'un emploi propre, physique, à un emploi figuratif, imagé. Le sens *gripe* v.1 (1a) [OE-971] « to make a grasp or clutch, to seek to get a hold (literal and figurative) » en est un exemple typique. Le sens originel dénote une action de préhension, avec déjà la possibilité d'une dénotation figurative, métaphore classique du sens littéral. Le sens *gripe* v.1 (7) [1559] « to grieve, afflict, distress », qui, comme on peut le voir, apparaît dans le lexique bien après le proto-sens, est métaphorique, car la douleur émotionnelle est assimilée au geste de préhension ; il s'agit d'une application figurative typique de HTC. L'idée de préhension peut représenter l'oppression sociale, comme dans *gripe* v.1 (6) [1644] « to oppress by miserly or penurious treatment; to 'pinch', 'squeeze'. (said also of poverty) », le point de comparaison étant sûrement l'idée de compression, puisqu'il s'agit de s'emparer de la chose 'préhensible' et de l'immobiliser, dans un rapport de force physique ou moral. Enfin, la métaphore peut être de nature somatique, voire clinique, dans *gripe* v.1 (8b) [1702] « to produce pain in the bowels as if by constriction or contraction; to cause 'gripes' ». Dans ce cas précis, les douleurs intestinales évoquent la compression présente lorsque l'on empoigne un objet avec la main. Fait intéressant, cette douleur physique est métonymiquement substituée à l'expression vocale plaintive du malade dans *gripe* v.1 (10) [1932] « to complain, 'grouse' ». Notons que cette métonymie semble être doublée d'une généralisation à d'autres situations que la gêne physique ; il s'agit alors d'une plainte en général.

6.4.4. Bilan

Pour résumer ce volet, on peut souligner que les profils métasémiques de SVU et HTC sont assez différents. D'abord, SVU comprend un nombre plus important de métonymies que HTC. Ensuite, SVU ne présente aucun autre type de changement que la métaphore et la métonymie, contrairement à HTC qui présente également quelques exemples de généralisation et de restriction. Enfin, les métaphores qui apparaissent chez HTC sont presque toutes de la même nature ; il s'agit en fait d'une métaphore filée, dans laquelle la notion de préhension joue le rôle de point de comparaison systématique. Une douleur somatique ou psychologique pourra ainsi être comparée à cette notion, ce qui débouche sur une idée de compression ; le corps et les

émotions sont malmenés, serrés, comprimés. Les métaphores de SVU sont plus diversifiées, sans doute parce que la notion, plus générique, de son, s'adapte à davantage de situations que celle, plus restreinte, de préhension. Une partie de la réponse à la question des interactions combinatoires entre les traits SVU, CSS et WCAF, est apportée par cette analyse métasémique. En effet, il apparaît que des sons extérieurs, naturels ou purement physiques, sont susceptibles de déclencher une série de sens métasémiques véhiculant l'idée de sons humains, signifiants ou non. Les sons extérieurs en question, généralement désagréables, d'où leur association avec CSS, ont tendance à donner naissance à une notion de plainte, lorsque le son humain, qui se résume souvent à un grognement, finit par devenir signifiant. L'association avec la notion animale est à creuser davantage, via l'analyse détaillée du profil métasémique de STBT. On peut toutefois déjà remarquer l'omniprésence d'une métaphore humain-animal. Il est ainsi courant qu'un proto-sens débouche sur une série de sens métasémiques, à travers l'enchaînement de métaphores et de métonymies de différents types : on passera d'un son non-vocalique à un son vocalique non-humain (animal), qui dérivera lui-même vers un son humain, souvent sur une forme de grognement plaintif, avant de produire des mots proprement dits.

II. ORIGINES ETYMOLOGIQUES DES MONOMORPHEMES EN *GR-*

7.1. Production lexicologique : bilan chronologique

Après avoir traité des changements métasémiques apparus au cours de l'évolution des monomorphèmes en *gr-*, il en ressort que le comportement métasémique de ces mots explique, en partie, le phénomène de phonesthésie associé au marqueur *gr-*. Une recherche plus axée sur les origines étymologiques des monomorphèmes du corpus nous servira à délimiter la part que jouent les processus diachroniques traditionnels dans la formation de cette association phonesthétique. Plus particulièrement, l'objectif de ce volet est de confronter nos observations précédentes aux données historiques, afin de détecter d'éventuels schémas. Commençons par un bilan global de la chronologie suivie par les mots en *gr-*. Les figures n° 18 et 19 ci-dessous montrent le nombre de proto-sens par tranche chronologique. La figure n° 18 présente les chiffres bruts pour chaque tranche chronologique, tandis que la figure n° 19 affiche les chiffres cumulés, afin de donner une idée de l'évolution progressive du lexique jusqu'à notre époque. La figure n° 20 permet une visualisation graphique du tableau 18. Les tranches chronologiques en question correspondent à des intervalles de 300 ans. Ce chiffre est largement arbitraire, mais permet d'avoir une vision d'ensemble, à la fois globale sans être trop large.

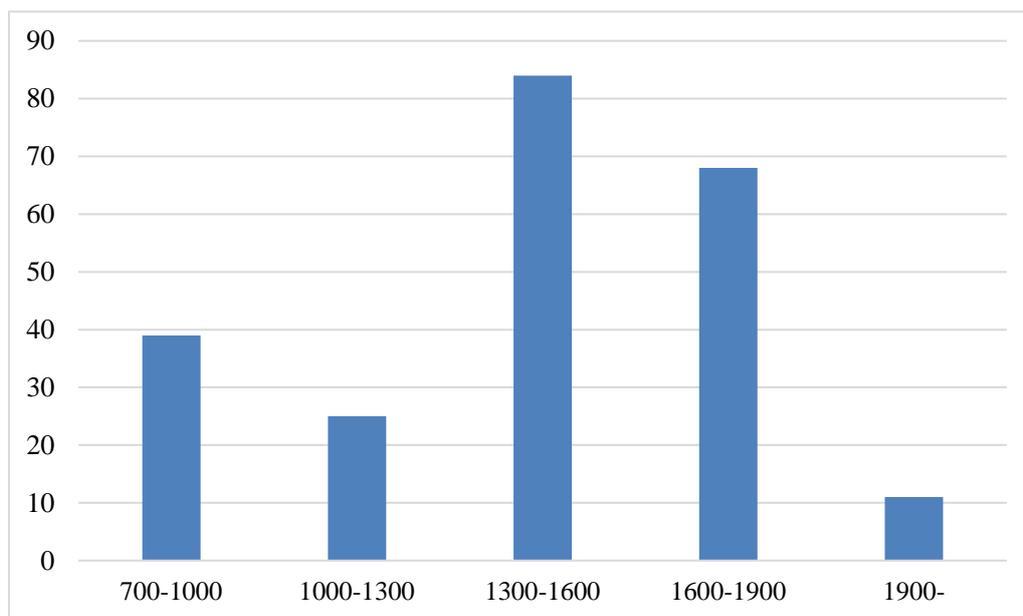
Figure n° 18 : Tableau de la répartition des proto-sens par période chronologique de 300 ans.

Tranches chronologiques	Occurrences	Fréquences
700-1000	39	17,18%
1000-1300	25	11,01%
1300-1600	84	37,00%
1600-1900	68	29,96%
1900-	11	4,85%
Total	227	100,00%

Figure n° 19 : Fréquences cumulées du nombre de proto-sens par période chronologique de 300 ans.

Tranches chronologiques	Fréquences cumulées
700-1000	17,18%
700-1300	28,19%
700-1600	65,20%
700-1900	95,15%
700-	100,00%

Figure n° 20 : Histogramme groupé du nombre d'apparitions de monomorphèmes par tranche chronologique de 300 ans.



Ces documents donnent à voir la production évolutive des monomorphèmes en *gr-*. Sur le même principe que pour l'analyse quantitative des EBC, ces données nous donnent accès à la production lexicale à l'œuvre au sein du corpus. Le graphe 20 confirme nos prévisions selon lesquelles le milieu et la fin du Moyen Âge ont connu une forte production lexicale. La période 1300-1600 compte en effet pour 37% des sens de l'ensemble. Il est d'ailleurs étonnant de constater que la tranche 1600-1900 paraît moins productive que la précédente. La période shakespearienne est en effet connue pour ses nombreuses innovations lexicales, et c'est à cette

époque qu'apparaissent de nombreux mots devenus très communs aujourd'hui (voir par exemple Milroy 2007, p. 26). Les périodes plus récentes, au contraire, ont surtout vu gonfler l'inventaire de mots savants, correspondant à des constructions complexes, et souvent issus de racines grecques et latines, ce qui ne concernent pas les monomorphèmes. Il est, par contre, réellement surprenant de constater que la période 700-1000 est plus productive que la période 1000-1300, en sachant la rareté des sources qui la caractérise. En dépit d'une documentation textuelle fragmentaire, il semblerait donc que les premiers siècles de notre ère d'étude aient été singulièrement fructueux. En ce qui concerne les fréquences cumulées, il est intéressant de remarquer que les deux tiers du corpus, tel que nous l'avons constitué, sont déjà apparus au tournant du XVII^e siècle, comme l'indique la figure n° 19. En somme, toutes ces données expliquent partiellement les variations observées au travers des courbes de progression. Les plateaux et les ruptures de courbes qui apparaissent au-delà du XIX^e siècle peuvent ainsi s'expliquer par le fait que l'évolution du lexique n'est tout simplement pas terminée, et ne le sera d'ailleurs jamais.

7.2. Paysage étymologique des monomorphèmes en *gr-*

7.2.1. Profil global

La figure n° 21 ci-dessous montre la proportion générale de chaque origine étymologique au sein de l'ensemble du corpus. La figure n° 22 rassemble ces données, et permet de résumer les catégories étymologiques les plus représentatives. Les mêmes valeurs sont représentées sous forme de secteur sur la figure n° 23. Il s'agit des étymons d'origines germanique d'une part, et latine ou romane d'autre part. La catégorie *autre* comprend toutes les autres origines.

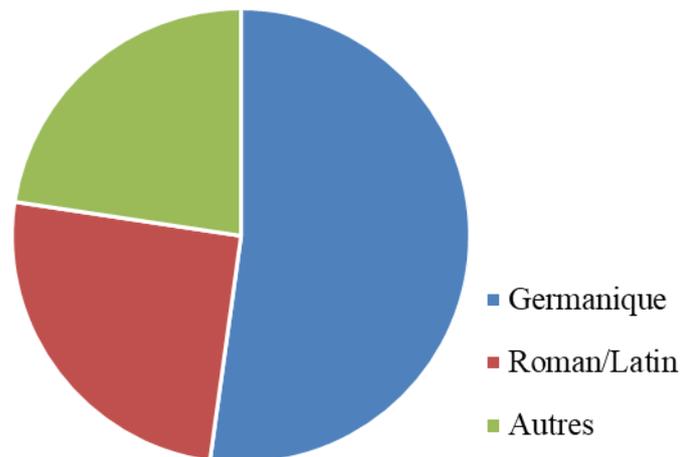
Figure n° 21 : Tableau des différentes origines étymologiques des monomorphèmes en *gr-*.

Origine	Occurrences	Fréquence
Vieil anglais	57	25,1%
Ancien français	35	15,4%
Origine inconnue	32	14,1%
Origine germanique	17	7,5%
Français contemporain	13	5,7%
Vieux norrois	10	4,4%
Proto-germanique	9	4,0%
Latin	8	3,5%
Onomatopée	8	3,5%
Moyen anglais	7	3,1%
Néerlandais contemporain	7	3,1%
Echoïque	6	2,6%
Moyen néerlandais	5	2,2%
Anglo-normand	3	1,3%
Bas-allemand	2	0,9%
Gaélique	2	0,9%
Imitatif	2	0,9%
Anglais contemporain	1	0,4%
Allemand contemporain	1	0,4%
Flamand contemporain	1	0,4%
Gallois	1	0,4%
Espagnol	1	0,4%
Italien	1	0,4%
Portugais	1	0,4%
Total	230	101,3%
Représentabilité	198	86,1%

Figure n° 22 : Tableau de la répartition globale des grandes familles étymologiques.

Origine	Occurrences	Fréquence
Germanique	120	52,9%
Roman/Latin	58	25,6%
Autres	52	22,9%
Total	230	101,3%

Figure n° 23 : Graphe de la répartition globale des grandes familles étymologiques.



Dans ces tables, les pourcentages se basent sur le nombre total de proto-sens, qui est, rappelons-le, de 227. Le total des origines étymologiques est cependant de 230, car 3 des sens du corpus présentent deux renseignements étymologiques. Il s'agit de : *gruff*, adj. (1a) [1533], qui provient d'une racine néerlandaise ou bas-allemande ; *gribble*, n.1 (a) [1578], qui provient d'une racine germanique, mais qui est également dit onomatopéique ; et *grab*, v. [1589], qui compte les mêmes informations que *gribble*. Nous avons donc trois variables du tableau de calcul qui apparaissent en doublons avec une autre. Nous avons choisi de conserver ces renseignements, car les incertitudes de l'OED sur l'origine exacte de ces lexèmes doivent être

considérées comme des informations en soi. La proportion des origines dans les calculs devait, par contre, se calquer sur le nombre total de sens, car il s'agit de trois variables étymologiques en plus, et non de trois sens supplémentaires. Précisons que les 14,1% de mots d'origine inconnue proviennent de l'incapacité des lexicographes à apporter une quelconque information sur l'étymologie de certains mots. Il s'agit d'une part assez importante, qui fait baisser la représentabilité du calcul, c'est-à-dire son exactitude, à 86,1%.

En ce qui concerne les données globales, on peut constater que les étymons d'origine germanique sont majoritaires, puisqu'ils constituent environ la moitié de l'ensemble. Les étymons romans et latins représentent à peu près un quart du corpus, ce qui peut paraître beaucoup pour une langue germanique, mais qui est en fait assez peu, compte tenu de la nature particulière du lexique de l'anglais. En effet, si l'anglais appartient bien à la famille des langues germaniques, un bon nombre de ses caractéristiques, et notamment son lexique, le rapproche des langues romanes (Milroy 2007, p. 9). On estime qu'environ 80% du vocabulaire est aujourd'hui non-germanique, et qu'une grande portion est d'origine latine ou romane, surtout française (Milroy 2007, p. 29). Dans la mesure où les lexèmes obsolètes de notre assemblage sont pris en compte, il n'est toutefois pas étonnant d'observer une proportion plus importante de lexèmes germaniques (voir les figure n° 24 et 25 ci-après pour plus de précisions). Ensuite, on remarque que ce sont le vieil anglais et l'ancien français qui ont joué le rôle le plus important dans la constitution lexicale de notre ensemble. Une fois de plus, rien de surprenant à cela, car la plupart des étymons germaniques de l'anglais sont d'origine anglo-saxonne, tandis qu'une plus faible part revient aux mots norrois (Milroy 2007, p. 30). Nous n'entrerons pas ici dans les détails de l'histoire de la langue, mais précisons simplement que la proportion de termes d'origine danoise dépend des régions, le nord de l'Angleterre et l'Écosse étant davantage concernés. Sa part est ici de 4,4%, ce qui ne paraît ni particulièrement haut ou bas. Le reste des chiffres est relativement anecdotique. On peut tout de même noter une part non-négligeable de lexèmes dits onomatopéiques, échoïques et imitatifs, surtout lorsqu'on combine les trois. De ce point de vue, on peut considérer que 7% des proto-sens sont signalés comme motivés dans le dictionnaire. La nature et l'origine de cette motivation restent cependant inconnues, car, dans la plupart des cas, la rubrique étymologique n'explicite pas les raisons de telles catégorisations.

Dans l'ensemble, il semble que la majorité des monomorphèmes potentiellement phonesthétiques, comprenant le marqueur *gr-*, présente un profil étymologique assez prévisible, bien qu'il faille noter une proportion de racines germaniques plus élevée que la moyenne. Cette

observation a tendance à confirmer l'hypothèse selon laquelle les étymons germaniques sont plus susceptibles d'entraîner une motivation phonesthétique que les autres racines.

Figure n° 24 : Tableau du profil d'usage par origine étymologique, en termes de nombre de sens et de pourcentages.

Oringine étymologique	Obsolète		Régional/dialectal/slang		Rare/archaïque/spécifique		Total non-conventionnel	
Latin	6	75,00%	0	0,00%	4	50,00%	10	125,00%
Onomatopée	6	75,00%	1	12,50%	1	12,50%	8	100,00%
Néerlandais contemporain	5	71,43%	4	57,14%	1	14,29%	10	142,86%
Moyen anglais	5	71,43%	0	0,00%	1	14,29%	6	85,71%
Vieux norrois	7	70,00%	4	40,00%	1	10,00%	12	120,00%
Moyen néerlandais	3	60,00%	1	20,00%	1	20,00%	5	100,00%
Proto-germanique	5	55,56%	3	33,33%	1	11,11%	9	100,00%
Origine germanique	9	52,94%	6	35,29%	4	23,53%	19	111,76%
Ancien français	18	51,43%	5	14,29%	5	14,29%	28	80,00%
Vièl anglais	28	49,12%	11	19,30%	8	14,04%	47	82,46%
Anglo-normand	1	33,33%	0	0,00%	0	0,00%	1	33,33%
Français contemporain	4	30,77%	2	15,38%	2	15,38%	8	61,54%
Origine inconnue	8	25,00%	13	40,63%	1	3,13%	22	68,75%
Echoïque	1	16,67%	3	50,00%	0	0,00%	4	66,67%
Imitatif	0	0,00%	1	50,00%	1	50,00%	2	100,00%
Gallois	0	0,00%	1	100,00%	0	0,00%	1	100,00%
Anglais contemporain	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Allemand contemporain	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Bas-allemand	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Flamand contemporain	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Gaèlique	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Espagnol	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Italien	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Portugais	0	0,00%	0	0,00%	0	0,00%	0	0,00%
Total	106	46,70%	55	24,23%	31	13,66%	192	84,58%

Figure n° 25 : tableau du profil d'usage par groupe étymologique, en termes de nombre de sens et de pourcentages.

Oringine étymologique	Obsolète		Régional/dialectal/slang		Rare/archaïque/spécifique		Total non-conventionnel	
Germanique	63	52,50%	29	24,17%	17	14,17%	109	90,83%
Roman/Latin	28	48,28%	7	12,07%	11	18,97%	46	79,31%
Autres	15	28,85%	19	36,54%	3	5,77%	37	71,15%
Total	106	46,70%	55	24,23%	31	13,66%	192	84,58%

Les deux tableaux ci-dessus témoignent de la diversité d'usage des mots en *gr-*, quelle que soit leur origine étymologique. Les fréquences du tableau n° 24 se basent sur le nombre total de proto-sens de l'origine étymologique en question, qui sont tirés de la figure n° 21. Ainsi, sur les 57 proto-sens d'origine anglo-saxonne, 28 d'entre eux sont obsolètes, 11 sont d'un

emploi régional, dialectal ou argotique, et 8 sont rares, archaïques ou spécifiques, pour un total de 47 sens non-conventionnels, c'est-à-dire 82,46%. Les totaux de lexèmes non-conventionnels peuvent être supérieurs à 100%, car certains sens sont décrits par plusieurs vecteurs, comme nous l'avons déjà expliqué dans la partie méthodologique. L'intérêt de telles données est qu'on s'aperçoit qu'une part importante de mots remontant au vieil anglais (49,12%) et à l'ancien français (51,3%), pour prendre les deux origines les plus communes, n'existent plus dans le lexique actuel. Ces chiffres ne sont cependant pas en rupture avec ceux de la figure n° 5, qui traite du niveau d'obsolescence générale du corpus. Notons au passage que la différence entre les deux types de tableaux est que la figure n° 5 se base sur les occurrences de catégories des EBC, alors que les tableaux n° 24 et 25 ci-dessus se basent sur le nombre de sens brut. Fait intéressant, on peut d'ailleurs observer que les chiffres de la figure n° 5 sont globalement plus élevés que ceux des figures n° 24 et 25. L'une des explications possibles est que les catégories ont tendance à se combiner davantage au sein de monomorphèmes plus récents, ou en tout cas encore en utilisation. Il ressort en effet d'une telle comparaison que les lexèmes obsolètes ont tendance à être le foyer de moins de combinaisons de traits. Cela tend à confirmer nos précédents résultats dans l'idée que le profil combinatoire des catégories conceptuelles reflète l'impact des traits phononotionnels sur la construction du lexique.

Il serait, quoi qu'il en soit, absurde de retirer les lexèmes obsolètes du corpus, car bien qu'ils ne soient plus employés aujourd'hui, ils ont participé à la construction du lexique actuel de l'anglais, tout au long de son évolution. Enfin, on peut constater que les fréquences des origines germanique et romane globales sont très similaires à celles des fréquences des origines anglo-saxonne et française, ce qui s'inscrit dans la lignée des données lexicographiques générales. En effet, comme nous l'avons déjà précisé, la majorité des lexèmes germaniques provient du vieil anglais, tandis que la plupart des mots romans descend de l'ancien français. Une fois de plus, rien d'absolument inattendu. On peut toutefois noter que 70% des lexèmes d'origine norroise sont obsolètes, et que 40% d'entre eux sont régionaux ou dialectaux, ce qui semble cohérent avec le fait que, la plupart du temps, ces étymons sont exclusivement employés dans les régions septentrionales, ou, au mieux, sont en perdition dans le reste de la communauté.

7.2.2. Profil trait par trait

Revenons, pour ce point, sur les catégories conceptuelles. Le profil étymologique des traits phonosémantiques fait suite aux analyses des courbes de progression, des profils combinatoires, et surtout de l'organisation métasémique des monomorphèmes en *gr-*. En

particulier, l'hypothèse d'une explication étymologique au comportement isolant des traits HTC et PFSF, doit être mise à l'épreuve. Nous aborderons dans le détail les évolutions diachroniques qu'ont connu les monomorphèmes du corpus au cours de la troisième section, mais il convient de commencer par une présentation du paysage étymologique pour chacun des traits. La figure n° 26 qui suit présente un tableau des origines étymologiques de ces traits. Les fréquences affichées sont calculées à partir du nombre d'occurrences d'entrées étymologiques. Dans la mesure où certains sens ont reçu plusieurs estampillages étymologiques, ce nombre est susceptible de légèrement varier du nombre d'occurrences de catégories de l'EBC associée. La fréquence dite totale a justement été calculée à partir du nombre d'occurrences de catégories pour chaque trait. On peut ainsi voir que seuls les traits CSS et HTC ont un nombre d'entrées étymologiques légèrement supérieur à leur nombre d'occurrences de catégories.

Figure n° 26 : Tableau des différentes origines étymologiques par catégorie.

WCAF		
Origine	Occurrences	Fréquences
Origine inconnue	20	27,4%
Vieil anglais	19	26,0%
Ancien français	8	11,0%
Origine germanique	6	8,2%
Français contemporain	3	4,1%
Vieux norrois	3	4,1%
Latin	2	2,7%
Onomatopée	2	2,7%
Moyen anglais	2	2,7%
Néerlandais contemporain	2	2,7%
Moyen néerlandais	2	2,7%
Proto-germanique	1	1,4%
Anglo-normand	1	1,4%
Gaélique	1	1,4%
Imitatif	1	1,4%
Total	73	100,0%
Représentabilité	53	72,6%

CSS		
Origine	Occurrences	Fréquences
Vieil anglais	10	25,6%
Ancien français	7	17,9%
Onomatopée	5	12,8%
Français contemporain	4	10,3%
Origine inconnue	3	7,7%
Néerlandais contemporain	3	7,7%
Echoïque	3	7,7%
Bas-allemand	1	2,6%
Imitatif	1	2,6%
Anglais contemporain	1	2,6%
Flamand contemporain	1	2,6%
Italien	1	2,6%
Total	40	102,6%
Représentabilité	37	94,9%

PFSF		
Origine	Occurrences	Fréquences
Vieil anglais	8	25,0%
Ancien français	7	21,9%
Origine inconnue	5	15,6%
Latin	3	9,4%
Origine germanique	2	6,3%
Proto-germanique	2	6,3%
Français contemporain	1	3,1%
Vieux norrois	1	3,1%
Onomatopée	1	3,1%
Espagnol	1	3,1%
Portugais	1	3,1%
Total	32	100,0%
Représentabilité	27	84,4%

HCT		
Origine	Occurrences	Fréquences
Vieil anglais	11	40,7%
Ancien français	6	22,2%
Origine germanique	3	11,1%
Moyen anglais	2	7,4%
Origine inconnue	1	3,7%
Français contemporain	1	3,7%
Proto-germanique	1	3,7%
Onomatopée	1	3,7%
Néerlandais contemporain	1	3,7%
Anglo-normand	1	3,7%
Total	28	103,7%
Représentabilité	27	96,4%

LDD		
Origine	Occurrences	Fréquences
Vieil anglais	5	18,5%
Proto-germanique	4	14,8%
Origine germanique	3	11,1%
Moyen anglais	3	11,1%
Moyen néerlandais	3	11,1%
Origine inconnue	2	7,4%
Vieux norrois	2	7,4%
Néerlandais contemporain	2	7,4%
Français contemporain	1	3,7%
Anglais contemporain	1	3,7%
Allemand contemporain	1	3,7%
Total	27	100,0%
Représentabilité	25	92,6%

STBT		
Origine	Occurrences	Fréquences
Vieil anglais	7	25,9%
Ancien français	6	22,2%
Vieux norrois	4	14,8%
Echoïque	2	7,4%
Origine inconnue	1	3,7%
Origine germanique	1	3,7%
Français contemporain	1	3,7%
Latin	1	3,7%
Moyen anglais	1	3,7%
Bas-allemand	1	3,7%
Gaélique	1	3,7%
Flandand contemporain	1	3,7%
Total	27	100,0%
Représentabilité	26	96,3%

SVU		
Origine	Occurrences	Fréquences
Vieil anglais	6	26,1%
Ancien français	4	17,4%
Echoïque	4	17,4%
Origine inconnue	3	13,0%
Onomatopée	2	8,7%
Imitatif	2	8,7%
Origine germanique	1	4,3%
Latin	1	4,3%
Français contemporain	0	0,0%
Total	23	100,0%
Représentabilité	20	87,0%

CTR		
Origine	Occurrences	Fréquences
Vieil anglais	4	29%
Latin	3	21%
Français contemporain	2	14%
Ancien français	1	7%
Origine germanique	1	7%
Proto-germanique	1	7%
Néerlandais contemporain	1	7%
Bas-allemand	1	7%
Origine inconnue	0	0%
Total	14	100,0%
Représentabilité	14	100,0%

Tout comme pour la figure n° 21, la représentabilité de l'échantillon est calculée à partir de la proportion de sens dont l'étymologie est inconnue. Les seuls traits qui tombent en-dessous de la barre des 90% sont WCAF, SVU et PFSF, ce qui peut poser problème étant donné que WCAF et SVU sont particulièrement intéressantes d'un point de vue interprétatif. Sans surprise, le vieil anglais et l'ancien français sont parmi les origines les plus fréquentes de presque tous les traits. Les proportions sont toutefois très différentes d'un trait à l'autre. Ainsi, à peu près un quart des catégories WCAF, CSS, PFSF, STBT et SVU est issu du vieil anglais. CTR monte à un tiers, tandis que HTC en comprend jusqu'à 40%. Cette dernière observation est intéressante, car on voit bien que HTC présente un profil étymologique assez restreint, avec une majorité de sens issus du vieil anglais ou de l'ancien français. Le trait LDD a, lui aussi, un profil singulier. Il est en effet très orienté vers les racines germaniques, avec une majorité de sens issus du vieil anglais, d'origine germanique générale, ou remontant au proto-germanique. Il n'est, soit dit en passant, pas évident d'interpréter ce que l'OED désigne comme germanique, sans plus de précision, mais il s'agit, la plupart du temps, de mots ayant de nombreux cognats germaniques, sans qu'il soit possible de distinguer un étymon bien précis pour autant. On distingue ce cas figure de la mention *proto-germanique* qui indique que la chaîne étymologique a pu être reconstituée jusqu'à la proto-langue. Autre singularité du corpus, le trait STBT semble contenir plus de sens d'origine norroise que les autres. En dehors de ce trait, les termes scandinaves sont assez rares.

On remarque, par ailleurs, que les descripteurs ayant trait à la motivation apparaissent de manière inégale selon les catégories, ce qui ne pouvait être mis en avant qu'au travers de tels calculs. On s'aperçoit ainsi que le trait CSS comporte 12,8% de sens onomatopéiques, tandis que SVU comprend 17,4% de sens d'origine échoïque. SVU est particulièrement intéressant de ce point de vue, car c'est le trait qui comprend la plus grande proportion de sens échoïques, mais aussi imitatifs (8,7%), ainsi que la deuxième plus grande proportion d'onomatopées, avec CSS (8,7%). On peut ainsi considérer que 34,8% des proto-sens de SVU présentent déjà une forme de motivation cognitive, quelle qu'en soit la nature exacte. Ces éléments peuvent être des indices de phonesthésie. Toutefois, on ignore encore, à ce point, si ces indices de motivation lexicale ont un rapport direct avec le phénomène que nous tentons d'évaluer ici. Il est en tout cas probable que cette large proportion de processus onomatopéiques, échoïques et imitatifs a joué un rôle dans la lexicogénèse des mots en *gr-*. Il est ainsi envisageable que ces particularités de SVU aient eu un rôle à jouer dans la formation de l'association phonosémantique adossée au marqueur submorphémique *gr-*.

7.3. Évolutions sémantiques

7.3.1. Racines communes et évolutions morphosémantiques

L'étude approfondie de l'ensemble des mécanismes morphologiques et sémantiques, responsables de l'évolution des monomorphèmes en *gr-* du corpus, s'avère particulièrement complexe. Nous ne nous concentrerons donc que sur les aspects qui nous intéressent ici, à savoir les éléments qui pourront éventuellement nous permettre de retracer l'émergence du phonesthème *gr-*, et d'identifier ses caractéristiques subsémantiques. Si les analyses conduites jusqu'ici représentent une première approche, l'analyse des racines étymologiques vise à confirmer ou à infirmer les hypothèses que nous avons émises plus haut. En particulier, l'idée selon laquelle le trait SVU pourrait être un développement subsémantique plutôt récent, assez lent, et très productif, est à tester. De même, la dissection du tissu étymologique des traits WCAF et CSS, qui sont les plus représentatifs du corpus, et donc de meilleurs candidats au poste d'éventuels traits subsémantiques fondamentaux, paraît indispensable. L'analyse de HTC permettra, quant à elle, de mettre les résultats en perspective, en proposant une réponse alternative. Comme nous l'avons montré, ce trait a plus de chances de contenir un nombre limité mais omniprésent de racines, qui expliqueraient la sensation de phonesthésie ressentie par les locuteurs. Seules ces quatre catégories conceptuelles seront abordées, essentiellement pour des raisons de temps de développement, mais aussi parce qu'il s'agit des traits les plus prometteurs.

L'analyse partira de chaque catégorie, mais les ponts qui existent entre ces catégories seront soulevés lorsque les occasions se présenteront, notamment pour juger des interactions entre traits sémantiques. L'objectif de cette section est d'analyser l'évolution diachronique des monomorphèmes du corpus de plus près, à l'échelle du sens, en vue de repérer des indices d'iconicité, et d'estimer l'importance de ces indices par rapport au poids qu'a pu jouer l'étymologie. Les informations étymologiques sont celles présentées par l'OED. Elles sont globalement stables, dans le sens où le corpus ne semble pas comporter d'informations contradictoires. Dans l'idéal, l'appel à d'autres outils de recherche serait désirable, mais, malheureusement, comme nous l'avons déjà expliqué, peu d'outils présentent la même exhaustivité et la même rigueur que ce dictionnaire, surtout sur des périodes chronologiques aussi longues. Dans les sections qui suivent, nous commencerons par décrire les racines communes, avant d'aborder la question de l'iconicité telle qu'elle est présentée dans l'OED.

7.3.2. WCAF

Les proto-sens du trait WCAF remontent, pour la plupart, à huit racines qu'il est possible de suivre. Avant d'entrer dans le détail, rappelons que, comme l'indiquait déjà la figure n° 26, le trait WCAF comporte une proportion conséquente de mots dont l'origine est inconnue ou incertaine (27,4%). Nous aurons l'occasion de revenir sur ces sens, mais il convient d'insister sur la distorsion possible des résultats que nous présentons ici. Rappelons cependant que tout travail à partir de langues ou de racines anciennes est irrémédiablement gouverné par des biais de sources, qu'il s'agit de garder en tête. Lorsque l'on met de côté les racines isolées, c'est-à-dire qui n'ont donné naissance qu'à un seul monomorphème tout au long de la période, on se retrouve avec un ensemble de huit racines, qui sont à l'origine de la plupart des sens du trait WCAF. Revenons sur ces racines de la plus productive à la plus rare.

On peut d'abord distinguer la racine proto-germanique (PG) **grīs-/*grus-*, dont la variation morphologique est due à une alternance apophonique caractéristique des langues indo-européennes, que la recherche philologique traditionnelle connaît sous l'appellation d'ablaut. Ainsi, les différents degrés vocaliques de cette racine ont donné naissance à une multitude d'étymons, dont il est possible de retracer le parcours, et qui expliquent la diversité des formes qui subsistent aujourd'hui. Nous ne détaillerons pas le parcours diachronique, morphologique et sémantique de chacun des sens du corpus, car cela nous prendrait bien plus de temps que nécessaire, mais l'on peut s'attarder à décrire quels sens sont étymologiquement liés. Les processus d'évolution sémantiques sont, par contre, très difficiles à suivre sur le long terme car, dans la plupart des cas, le sens, même général, des racines anciennes ou reconstruites, demeure inconnu. La racine **grīs-/*grus-* est à l'origine des monomorphèmes *grure*, n. [OE/900] « fright, terror; something frightful, fearful agony », *grisly*, adj. [1150] « causing horror, terror, or extreme fear; horrible or terrible to behold or to hear », *grise*, adj. [1300] « terrible; fearful », *grizzel*, adv. [1898] « horribly », *grisle*, n.[1225] « horror; terror », *grisol*, adjective [1440] « horrible; grisly », *grise*, v. (2) [1250], « to shudder or tremble with terror; to be full of horror, greatly afraid », *grist*, n.1 [1000] « gnashing of teeth; hence, anger », *grusnen*, v. [1325] « to cry out with fright », *grue*, v.1 (1a) [1330] « to feel terror or horror, shudder, tremble; quake; to shrink from something; to be troubled in heart; to thrill », *grye*, v. [1390] « to shudder, feel horror » et *groose*, v. [1673] « to shiver, shudder ».

L'une des racines les plus productives est la forme PG **gram-/*grem-*, à l'origine de mots assez distincts morphologiquement parlant. Les forms qui en descendent sont : *grame*,

adj. (1) [893] « (a) angry; vexed; furious; (transferred of heat) fierce. (b) grieved, sorrowful », *grim*, adj. (1a) [971] « of persons or animals: fierce, cruel, savage or harsh in disposition or action », *gremth*, n. [1340-70] « anger; rage », *greme*, v. (1a) [893] « to anger, grieve; to vex », *grame*, n. (1) [1000] « anger, wrath, ire », *greme*, n. (1) [1300] « anger; wrath », *grindel*, adj. [1400] « fierce; angry », *grim*, v. [1400-50] « to be angry, look fierce » et *grime*, v. [1483] « to cover with grime, to blacken, befoul; also (figurative) to grime the face of ». Les processus évolutifs de ces différents lexèmes ne sont toutefois pas exactement les mêmes, et les chemins suivis peuvent être très différents. Cette racine est un cas d'école des disparités morphologiques que peuvent entraîner les variations apophoniques. En effet, le degré ouvert **gram-* a donné les formes *grame*, adj. (1) [893] (< OE *gram/grom*) et *grame*, n. (1) [1000] (< OE *grama*), tandis que le degré fermé **grem-* a donné les formes *gremth*, n. [1340-70] (< OE **grēmð(u)*), *greme*, v. (1a) [893] (< OE *grēmian*) et *greme*, n. (1) [1300] (< vieux norrois *grēmi*), mais aussi les formes *grim*, adj. (1a) [971] (< OE *grim(m)*), *grime*, v. [1483] et *grindel*, adj. [1400]. L'évolution de *-e-* en *-i-* n'est pas explicitée par l'OED, mais les deux voyelles sont si proches phonétiquement que leur rapprochement n'a rien d'étonnant. Il est par ailleurs important de souligner que les origines exactes de *grime*, v [1483] et *grindel*, adj. [1400] n'ont pas pu être retracées avec le même degré de confiance que pour les autres racines, mais qu'elles pourraient provenir du moyen néerlandais **grimen* et du vieux norrois *grimm-r*, respectivement. On fera remarquer que le monstre du récit héroïque anglo-saxon *Beowulf* porte le nom de Grendel, et que, bien que ce mot n'apparaisse pas dans la version diachronique de l'OED, la ressemblance morphologique avec *grindel* est frappante. Nous reviendrons brièvement sur les créations littéraires par la suite.

Le terme normand *groucier* « to murmur, to grumble » a également généré un nombre important de mots en *gr-*, mais ils sont presque tous issus de *grutch*, v. (1a) [1200-25] « to murmur, complain, repine; to grudge ». Parmi les variantes morphologiques de ce dernier, on compte, dans l'ordre chronologique d'apparition : *grudge*, v. (1a) [1440] « to murmur; to utter complaints murmuringly; to grumble, complain; to be discontented or dissatisfied », *grunch*, v. (a) [1487] « to grumble, express discontent; to grudge, to object », *grouch*, n. (1) [1895] « grumbling; a complaint or grumble; a grumbling, sulky mood; a fit of ill temper or sulkiness » et *grouch*, v. [1916] « to grumble ». Au vu de la date d'apparition de cette racine, on peut affirmer qu'il s'agit d'un emprunt normand post-conquête. De manière générale, on constate une forte disparité entre les racines romanes et germaniques, qui apparaît encore plus évidente qu'à travers les données statistiques globales. En effet, on s'aperçoit que la quantité de mots

d'origine romane a tendance à être surreprésentée, étant donné qu'une seule racine, introduite à un instant *t*, plutôt récemment, peut donner naissance à un nombre important de monomorphèmes. A l'inverse, on voit bien que les racines germaniques, souvent très anciennes, se développent généralement sous diverses formes à différentes étapes de l'histoire de la langue.

On distingue ensuite une deuxième racine PG, **ghrēud-* « to weep, to cry out, to rage », qui pourrait remonter au proto-indo-européen (PIE) **ghrēd-/ghrōd-*. Le conditionnel est de mise, pour les raisons habituelles en linguistique historique, mais aussi parce que la nomenclature de l'OED n'est pas toujours facile à suivre. Pour l'entrée de *greet*, v.2 (1a) [725] « to weep, cry, lament, grieve », on nous parle d'une racine du « vieil aryen »¹⁹ **ghrēd-/ghrōd-*. Cette appellation correspond probablement à une terminologie désuète, dont nous avons supposé qu'il s'agit du PIE. Cette racine pourrait être liée au sanskrit *hrād* « résonner », ce qui introduit une dimension sonore. On serait bien sûr tenté de faire un rapprochement étymologique direct entre ce sens de WCAF et le trait SVU, mais le niveau de preuve est ici trop bas pour pouvoir l'affirmer. La réalité du terrain est en fait encore plus complexe, car la racine à l'origine de *greet*, v.2 (1a) [725] est incertaine. Il pourrait s'agir du vieil anglais (OE) *græt-/grét-* ou *gréot-*, deux descendant du PG **ghrēud-*. Le terme *grede*, v. (1a) [1000] « to cry, cry out, shout; to wail » vient du OE *grédan*, qui pourrait être un cognat de *grétan*, qui, lui, descend de la même racine **ghrēud-*. Il est donc possible que ce sens soit un cognat de *greet*, v.2 (1a). Le terme *grot*, n.2 [1325] « weeping, lamentation » fait également partie de la même famille, mais par le biais du groupe germanique septentrional, puisqu'il est issu de la racine norroise *grát-r*, elle aussi descendante de **ghrēud-*. Quant à *grot*, n.5 [1970] « an unpleasant, dirty, or ugly person » et *grody*, adj. [1965] « disgusting, revolting, 'gross'; dirty, unhygienic, squalid; unattractive, slovenly, sloppy », ceux-ci proviennent, en définitive, de *grot*, n.2.

On tombe ensuite sur des racines plus rares. Les mots *grill*, v.1 (1) [897] « to provoke, annoy, irritate, offend », *grill*, adj. (1) [1200] « fierce, harsh, cruel » et *grill*, n.1 (a) [1300] « ill-will, vexation; harm, mischief » remontent à la même racine reconstruite **gryl-* du OE. Cette racine germanique ne doit pas être confondu avec la racine romane *gril-*, qui a notamment donné *grill*, v.2 (1c) [1825] « to torment with heat, to 'broil' » (< français *griller*).

¹⁹ « Old Aryan », *sic*.

Enfin, dernière racine d'origine germanique, la forme **grai-/grī-/grin-* est à l'origine du verbe proto-germanique **granjan*, ancêtre du sens *grin*, v.2 (1) [OE]. De même, le verbe **grainôjan*, ancêtre de *groan*, v. (1a) [700], provient de la même racine.

On distingue également une deuxième racine française, issue d'un emprunt à la suite de la conquête normande, sous la forme *gr(i)ef-/grev-*. Cet étymon, du latin *gravis* « heavy, grave », apparaît sous les formes *grief* ou *gref* en ancien français, et a donné l'anglais *grief*, n. (1) [1225] « hardship, suffering; a kind, or cause, of hardship or suffering » et *grief*, adj. (A) [1300] « grievous, grave; troublesome, oppressive ».

L'ultime racine étymologique qu'il est possible de distinguer, à partir de l'OED, est celle du latin *grassārī* « to roam about, to lie in wait, to run riot, to rage », qui a donné *grassant*, adj.1 [1601] « of something or someone unwelcome, harmful, etc.: widespread, prevalent, endemic; flourishing or spreading unchecked; spec. (of a disease) raging » et *grassate*, v. [1652] « of a disease: to rage; later more generally, of anything unwelcome, harmful, etc.: to flourish or spread unchecked ».

La catégorie WCAF ne comporte pas de lexèmes dits iconiques à proprement parler, en tout cas tels que décrits par l'OED. On a cependant, d'une part, un nombre important de sens dont l'origine est inconnue ou incertaine, et d'autre part, un nombre tout aussi conséquent de formes qui semblent avoir subi des processus morphologiques complexes, dont l'origine est possiblement iconique. Ce n'est bien entendu pas parce que l'origine d'un mot est inconnue qu'elle est forcément iconique, mais une étude plus approfondie de ces lexèmes permettrait de lever le mystère sur l'importance et la nature exacte de l'iconicité à l'intérieur du corpus. L'OED précise notamment que *grum*, adj. (a) [1640] « of persons, and their aspect and mode of speaking: gloomy, morose, surly, glum » pourrait être issu d'un amalgame entre différentes formes, puisque sa forme rappelle *grim*, *glum*, *gruff* et *grumble*. On a également un parallèle en danois : *grum* « cruel ». Si cet amalgame était confirmé, on pourrait parler de motivation iconique du lexique, puisque le phonesthème *gr-* permettrait alors de créer de nouveaux lexèmes, contenant l'un de ses traits subsémantiques, WCAF en l'occurrence. Il semblerait que *grump*, v. [1875] « to sulk » soit issu du même type de morphogénèse. L'OED s'en tient cependant à une comparaison avec *grump* n. et *glump*, v. (intransitif, dialectal, 1804), sans préciser s'il s'agit réellement d'un amalgame ou non. Les monomorphèmes *graunch*, v. [1881] « to make a crunching or grinding sound; transitive, to cause to make such a sound; hence, to damage » et *grent*, v. (2) [1387] « to grunt or groan » sont, quant à eux, décrits comme

onomatopéiques. *Graunch*, v. est comparé à *granch*, v., tandis que *grent*, v. (2) est comparé à *grin*, v.2, *grint*, v. et *grunt*, v. Un autre indice de l'éventuelle productivité d'un élément submorphémique est son apparition dans des néologismes, c'est-à-dire des lexèmes créés de toutes pièces. Quelques créations littéraires sont à noter. Si le cas de *Grendel* n'est pas renseigné dans l'OED, on peut mentionner l'exemple de *Grundy*, n.3 [1798] « the surname of an imaginary personage (Mrs. Grundy) who is proverbially referred to as a personification of the tyranny of social opinion in matters of conventional propriety », qui provient d'une pièce de Morton, dans laquelle le personnage de Mrs. Grundy est une caricature de la voisine moralisatrice et indiscreète. L'aspect désagréable du personnage est probablement la raison de son association avec WCAF. Le terme de *grok*, v. (a) [1961] « to understand intuitively or by empathy; to establish rapport with » peut paraître inopiné, car s'il est question d'émotions, celles-ci sont ici plutôt positives. Nous avons toutefois préféré le garder, car cet aspect méritait d'être soulevé, d'autant plus que cette forme est une pure création de l'auteur Robert A. Heinlein. Une analyse de l'œuvre de science-fiction *Stranger in a Strange Land* (1961) pourrait nous en apprendre sur les raisons éventuelles d'une telle création. Enfin, le cas de *gremlin*, n. (1) [1929] est un peu particulier, car bien que le terme ait été repris par Roald Dahl, il n'en est pas l'inventeur. L'origine du mot est incertaine, et pourrait venir d'une racine irlandaise.

7.3.3. CSS

Le nombre de racines communes qu'il est possible de distinguer pour CSS est beaucoup plus réduit que pour WCAF : on compte trois racines principales. La première remonte au OE *gréot-/grytt-/grút-*, issue de la forme reconstruite **greutom* (PG), elle-même dérivée de **ghreud-* (PG). Cette forme a donné une variété de sens, assez proches les uns des autres, mais de manière moins nette que pour les sens étymologiques de WCAF, où la parenté sémantique était systématiquement évidente. On distingue d'abord des parcours diachroniques continus, depuis le PG à l'anglais contemporain, en passant par le OE : *grit*, n.1 (1a) [OE] « formerly: sand, gravel, small stones; now: minute particles of stone or sand, as produced by attrition or disintegration », *grit*, v. (1) [1762] « to produce a grating sound such as is caused by the crushing of grit; to move with such a grating noise », *grittle*, v. [1736] « to break (corn) or grind it coarsely », *grot*, n.1 [888] « a fragment, particle, atom; every grot = every whit », *groats*, n. (1a) [1100] « hulled, or hulled and crushed grain of various kinds, chiefly oats, but also wheat, barley, and maize » et *grout*, n.1 (1) [725] « coarse meal, peeled grain ». Apparaissent ensuite des racines plus délicates à retracer, dont les informations sont moins solidement étayées : le

sens *gravel*, n. (2a) [1366] « a material consisting of coarse sand and water-worn stones of various sizes, often with a slight intermixture of clay, much used for laying roads and paths » est décrit comme provenant de l'ancien français *gravele*, dont la racine serait un étymon du OE *gréot*, donc un descendant du PG **ghreud*. De même pour *greave*, n.3 [1579] « The sandy shore of a river », qui vient du français *grève*, de la même origine que *gravele*. Enfin, dernière racine française récente, *gruel*, n. (1) [1311], de l'ancien français *gruel*, qui serait issu du latin *grūtum*, d'ascendance germanique. Pour la plupart de ces lexèmes, l'OED ne nous informe guère plus que de la mention évasive « compare OE *grút* grout n.1 », pour prendre le dernier exemple.

La deuxième racine est celle du PG **grinstu-*, elle-même dérivée du PG **grind-*, descendant possible du PIE **ghrendh-*. Cet étymon a la paternité des monomorphèmes suivants : *grind*, v.1 (1a) [1000] « to reduce to small particles or powder by crushing between two hard surfaces; esp. to make (grain) into meal or flour in a mill », *grist*, n.2 [1000] « the action of grinding; an act or spell of grinding », *grist*, n.1 [1000] « gnashing of teeth; hence, anger » et *grist*, v.1 [1460] « to gnash or grind the teeth », issu du précédent. Les deux derniers lexèmes sont particulièrement intéressants car ils représentent probablement le pont sémantique entre les traits CSS et WCAF, qui se combinent régulièrement.

La dernière racine notable n'est pas germanique, mais française. La forme *grat-* a donné *grater*, n.1 (1) [1390-1] « an instrument with a rough indented surface used for grating or rasping », *grattoir*, n. [1872] « a flint scraping tool in which the working edge is at the end of the blade or flake and lies across its long axis; an 'end-scraper' » et *grate*, v.1 (1a) [1400] « to scrape, file, abrade; to rub harshly, scarify, excoriate ».

Les lexèmes iconiques sont : *grouse*, v. [1200-25] « to devour or munch with a crunching sound », *grint*, v. (1a) [1300] « to grind or gnash the teeth; usually *to grint with the teeth* », *grush*, v. [1420] « to crush; to make a deep wound in; to gash; also intransitive for reflexive: to crumble », *grent*, v. (1) [1425] « to gnash the teeth », *graisle*, v. (1a) [1522] « to make a harsh noise, esp. repeatedly; to creak, rattle; (in modern use) to crackle », *grash*, v. [1563] « gnash », *granch*, v. [1736] « of the teeth: to gnash », *grundy*, n.2 [1840] « granulated pig-iron », *graunch*, v. [1881] « to make a crunching or grinding sound; transitive, to cause to make such a sound; hence, to damage » et *grinch*, v. [1892] « to make a harsh grating noise ». On peut remarquer que la notion de son est très présente parmi ces sens.

7.3.4. SVU

Sur les quatre racines à l'origine des monomorphèmes en *gr-* associés au trait SVU, trois sont également en lien avec WCAF. Notons toutefois que le parallèle est mince, car ces racines communes ne reposent que sur la combinaison conceptuelle de deux ou trois sens. Nous avons d'abord **grai-/*grī-/*grin-* (PG), qui a donné *groan*, v. (1a) [700] « to breathe with a deep-toned murmur; to utter a low deep sound expressive of grief or pain ». Il s'agit en fait du seul sens commun, du point de vue étymologique, entre WCAF et SVU, pour cette racine-ci. De même, le seul sens commun pour **græt-/*grét-/*ghrēud-* est *grede*, v. (1a) [1000] « to cry, cry out, shout; to wail ». Le normand *groucier* a donné deux sens communs : *grutch*, v. (1a) [1200-25] « to murmur, complain, repine; to grudge » et *grudge*, v. (1a) [1440] « to murmur; to utter complaints murmuringly; to grumble, complain; to be discontented or dissatisfied ». Viennent ensuite les termes iconiques, très nombreux chez ce trait.

Parmi les lexèmes qui ne correspondent à aucune racine retraçable sur le long terme, mais qui présentent des composantes iconiques, nous avons : *grunt*, v. (1a) [725] « of a hog: to utter its characteristic low gruff sound; also of other animals and of persons (with conscious allusion to the pig): to utter a sound resembling this », *gruntle*, v. (1) [1400] « to utter a little or low grunt; said of swine, occasionally of other animals; rarely of persons », *growl*, v.3 (1a) [1380] « of the bowels: to rumble », *grent*, v. (2) [1387] « to grunt or groan », *graisle*, v. (1a) [1522] « to make a harsh noise, esp. repeatedly; to creak, rattle; (in modern use) to crackle », *grumph*, v. [1807] « to grunt; said both of animals and human beings », *grumph*, n. [1737] « a grunt, whether from an animal or a human being », *grrr*, int. [1867] « representing the sound of growling, as by a dog or similar animal; (hence also in later use) expressing anger or (more usually) mock anger », *graunch*, v. [1881] « to make a crunching or grinding sound; transitive, to cause to make such a sound; hence, to damage » et *grinch*, v. [1892] « to make a harsh grating noise ». Les lexèmes *grassil*, v. [1513] « to make a harsh noise; to creak, rattle; also quasi-trans., to make a harsh noise with (the teeth or tusks), to gnash » et *grumble*, v. (1a) [1608] « of persons and animals: to utter dull inarticulate sounds; to mutter, mumble, murmur; to growl faintly » sont susceptibles d'être également iconiques, compte tenu du rapprochement sémantique avec les sens cités plus haut, même si l'OED ne le spécifie pas explicitement. La caractéristique principale de tous ces lexèmes est la composante animalesque, qui est très présente. Par ailleurs, on peut remarquer que la majorité de ces sens sont relativement tardifs. Cette observation vient corroborer les données combinatoires et métasémiques, et tend à

confirmer notre hypothèse selon laquelle une partie des valeurs potentiellement phonesthétiques de SVU sont apparues récemment.

7.3.5. HTC

Comme le comportement général de HTC le laissait présager jusqu'ici, ce trait résulte de l'évolution d'un petit groupe de racines communes, qui forme le vivier des formes phonesthétiques en *gr-* en rapport avec la notion de préhension. On compte trois racines principales, particulièrement productives, et deux autres racines plus discrètes. La proportion relativement importante de lexèmes d'origine romane pour HTC est en fait due à l'emprunt à l'ancien français du mot *grapon/grapin*, qui apparaît sous de multiples formes dans les textes. L'emprunt le plus ancien répertorié est *grapnel*, n. (1a) [1373] « an instrument with iron claws intended to be thrown by a rope for the purpose of seizing and holding an object, esp. an enemy's ship ». Nous trouvons ensuite : *grape*, n.2 [1493] « a hook », *grape*, v. [1523] « grapple », *grapple*, n. (1) [1530] « an implement for grappling or laying hold », *grasple*, n. [1553] « grapple n. (1) », *grasple*, v. [1553] « grapple, v. 4 (8a) », *gropple*, v. [1861] « grope » et *grubble*, v. (1) [1690] « to grope; to grubble up: to scrape together ». Le pseudo-suffixe *-le* est particulièrement productif ici.

La racine *grip-* (OE) « grasp, clutch » est également très productive. Cette forme du OE provient du PG occidental **grippjan*, elle-même issue du PG **ghreib-/*ghroib-*. Les formes que l'on observe dans notre corpus sont, de la plus ancienne à la plus récente : *grip*, n.1 (1a) [OE] « firm hold or grasp; the action of gripping, grasping, or clutching; esp. the tight or strained grasp of the hand upon an object (cf. handgrip n.); also, grasping power », *grip*, v.1 (1a) [950] « to grasp or seize firmly or tightly with the hand; to seize with the mouth, claw, beak or other prehensile organ », *gripe*, v.1 (1a) [OE-971] « to make a grasp or clutch, to seek to get a hold », *gripple*, adj. (1) [1000] « griping, niggardly, usurious » et *gripple*, n.2 [1530] « a hook to seize things with ». L'OED ne précise malheureusement pas si cette racine est un cognat de *grapon/grapin*, mais il paraît raisonnable de le supposer, compte tenu du rapprochement morphologique et sémantique des monomorphèmes et des sens qui découlent de l'une et de l'autre.

La troisième et ultime racine majeure, en termes de nombre de sens descendants, est **graip-/*grīp-* (PG). Celle-ci nous a donné les sens suivants : *grope*, v. (1) [825] « to use the hands in feeling, touching, or grasping; to handle or feel something », *grope*, n.1 (1) [OE]

« grasp; figurative grasp of a subject », *grob*, v. [1654] « to grope », *grasp*, v. (1a) [1382] « to make clutches with the hand » et *grappe*, v. (a) [1382] « to feel with one's hands, to grope ». Tous ces sens sont des synonymes proches les uns des autres. Toujours sur le même principe, on est en droit de se demander si les racines **ghreib-/*ghroib-* et **graip-/*grīp-* descendent d'une même racine PG, qui descendrait elle-même d'une racine PIE éventuellement partagée par les formes *grapon/grapin*. Une investigation plus approfondie, faisant appel à d'autres outils de recherche étymologique, apporterait un complément utile.

Les deux racines mineures sont *grab-*, d'origine néerlandaise ou bas-allemande, et *grin-*, présente en OE. La première a donné les formes *grabble*, v. (1a) [1579-80] « to feel or search with the hands, to grope about ; sometimes to grope and grabble » et *grab*, v. [1589] « to grasp or seize suddenly and eagerly; hence, to appropriate to oneself in a rapacious or unscrupulous manner », tandis que la deuxième est l'ancêtre des sens *grin*, n.1 (1a) [825] « a snare for catching birds or animals, made of cord, hair, wire, or the like, with a running noose » et *grane*, n. [1250] « a snare, trap; a noose ». Insistons sur le fait que la forme *grin-* dont il est question ici n'est pas la même que la racine PG **grai-/*grī-/*grin-*, que l'on trouve chez WCAF. Il n'est bien sûr pas à écarter que les deux formes soient apparentées, mais il nous reste encore à le prouver. Pour terminer sur les racines, on peut souligner que *greet*, v.1 (1) [725] « to approach, come up to; to begin upon, begin to treat or handle, take in hand » est issu du vieil anglais *grætan*, qui remonte à la forme PG occidentale **grōto-z/*grôtâ*, qui descend de la même racine PIE **ghrōd-/*ghrēd-* que chez SVU. Aucun lexème d'origine iconique n'est explicitement mentionné pour ce trait sémantique, mais on peut citer deux lexèmes potentiels. D'abord *grisp*, v. [1420] « to grasp, to grope », amalgame de *grip*, v.1 [950] (HTC et STBT) et de *grasp*, v. [1382] (HTC). Ensuite, *graze*, v.2 (1a) [1616] « to touch (a surface) lightly in passing; esp. to roughen or abrade (the skin or a part of the body) in rubbing or brushing past; also, to suffer slight abrasion of (a part of one's body) », dont l'origine est incertaine, mais qui pourrait être un amalgame entre *glaze/glace*, v. et *grate*, v.1.

Conclusion

Au terme de cette étude, un certain nombre de points méritent d'être soulignés. D'abord, il semble que les différents types d'analyses qui ont été menés soient en accord les uns avec les autres, et qu'ils apportent des éléments de réponse convergents. En effet, un trait phonosémantique fondamental lié à la notion d'émotions négatives semble transparaître à l'issue du traitement des données. Les fréquences brutes et les courbes de progression montrent ainsi une représentation bien au-dessus de la moyenne du trait WCAF. Le deuxième trait le plus représenté, CSS, présente cependant un profil combinatoire plus intéressant que WCAF, à l'instar du trait SVU qui, pour sa part, affiche une fréquence brute et un profil évolutif général bien plus faibles. Rappelons toutefois que les écarts de taille des différents ensembles de sens liés à ces traits sémantiques ont tendance à perturber les données. En effet, l'échantillon de 142 occurrences de WCAF ne présente pas la même représentabilité que les 31 occurrences de CTR, pour prendre les deux cas extrêmes en exemple. D'un côté, un trait davantage présent à travers le corpus comme WCAF aura plus de chances de se combiner avec d'autres traits, mais, d'un autre côté, ces combinaisons représenteront une proportion moindre de son total d'occurrences. La différence majeure entre l'analyse des EBC et des PAC est le changement de comportement assez radical de SVU, qui passe d'une catégorie conceptuelle très peu attestée à la troisième plus attestée du corpus. L'analyse des schémas combinatoires propulse également SVU sur le devant de la scène, avec les taux globaux les plus hauts, à la fois pour les EBC que les PAC. De plus, on remarque une interaction très soutenue entre SVU, CSS et WCAF. Quant à l'analyse des co-émergences et post-attestations, celle-ci indique une prévalence de SVU, suivi de près par les traits CSS et WCAF. En somme, ces analyses préliminaires indiquent une importance majeure de trois traits sémantiques du corpus, associés aux notions d'émotions négatives (WCAF), de sensations ou de bruits déplaisants (CSS) et de sons ou de voix (SVU).

L'analyse des évolutions métasémiques n'a pas été réalisée pour tous les traits, mais seulement pour SVU et HTC, en se basant sur les résultats livrés par les analyses précédentes. On a pu remarquer un nombre important de processus de métaphorisation et de métonymisation dans le profil de SVU, et ce davantage que pour HTC. Il semblerait ainsi que la notion de son soit particulièrement propice au changement sémantique sur le long terme. Cette observation est à mettre en parallèle des données étymologiques, qui révèlent un nombre très important de lexèmes, associés à SVU, d'origine iconique. Par ailleurs, étant donné que trois des racines à l'origine des lexèmes associés à SVU sont également à l'origine de lexèmes associés à WCAF,

nous avons formulé l'hypothèse selon laquelle ces deux traits étaient intrinsèquement liés depuis longtemps. Ce lien n'a d'ailleurs cessé d'être entretenu, comme l'attestent les données combinatoires. Le trait CSS compte également un nombre important de lexèmes d'origine onomatopéique, imitative ou échoïque, souvent en lien avec la notion de son. Il est intéressant de noter que ces traits, qui comprennent beaucoup de sens iconiques²⁰, sont aussi ceux qui ont les plus hauts taux combinatoires, ce qui constitue un indice fort de convergence phonesthétique. On se retrouve donc avec, d'un côté, des catégories conceptuelles dont les sens communs sont issus de racines étymologiques communes, et de l'autre, des catégories conceptuelles dont les sens communs ne peuvent être expliqués par une origine commune. Il paraît raisonnable de supposer que ces dernières sont le fruit d'une analogie cognitive, qui aurait eu une influence sur la construction des lexèmes du corpus. Ces catégories sont, au vu des données à la fois combinatoires, métagématiques et étymologiques, SVU et CSS. Au contraire, ces mêmes données indiquent que les catégories HTC et PFSF sont des exemples typiques de traits sémantiques dont la validité statistique n'est due qu'à un ensemble de racines communes particulièrement productives, mais qui ne semblent pas issues du même phénomène d'analogie. Le reste des traits est plus ambigu, notamment pour des raisons étymologiques et combinatoires, mais nous y reviendrons. Ces résultats nécessiteraient cependant d'être confirmés par d'autres méthodologies. En particulier, et c'est là la limite d'une telle étude, la production de données à partir d'une lecture uniquement lexicographique ne paraît pas suffisante pour expliquer tous les tenants et aboutissants des processus à l'œuvre.

Parmi les éléments qu'il est possible d'améliorer, dans le cadre d'éventuelles recherches à venir sur le sujet, les manipulations statistiques nécessiteraient plus de rigueur. Une compréhension plus poussée des étapes de calculs comme celles proposées par Drellishak 2006, par exemple, serait souhaitable. Une partie des manipulations qui ont été effectuées sont par ailleurs incomplètes. Il serait par exemple intéressant d'appliquer les méthodes d'analyse métagématique mises en place au cours de cette étude à l'ensemble des catégories conceptuelles. En effet, si nos résultats tentent à montrer une différence entre deux types de traits sémantiques, les uns présentant davantage de liens métonymiques et métaphoriques que les autres, seule une analyse exhaustive de tous les sens du corpus permettrait de s'en assurer. Cet aspect est crucial, car c'est cette analyse de l'évolution sémantique des lexèmes phonesthétiques qui permet une observation fine des potentiels paramètres de l'émergence de *gr-*. De même, la recherche de

²⁰ Le terme d'*iconique* fait ici référence aux indices de formations onomatopéiques, imitatives et échoïques décrits plus avant.

racines communes à partir des informations étymologiques nécessiterait d'être menée à terme pour tous les traits, afin de disposer d'une description complète de toutes les racines de *gr-*.

Ensuite, une recherche plus approfondie des mécanismes d'évolution morphosémantique permettrait de retracer l'apparition des racines et leur évolution avec plus de précision, afin mieux comprendre le rôle de l'étymologie dans l'émergence des traits phonosémantiques. Les données étymologiques et métasémiques pourraient ainsi être mises en regard. En particulier, les lexèmes décrits comme iconiques dans l'OED méritent d'être analysés davantage. Il est intéressant de noter que, dans le cas de *gr-*, la notion d'émotions négatives n'est pas celle qui présente le plus de lexèmes iconiques, bien que l'on serait tenté de penser qu'elle agit comme un trait phonosémantique fort, au vu de sa représentabilité globale. D'une certaine manière, il est possible que le trait WCAF ne soit pas à proprement parler phonesthétique, mais qu'il résulte essentiellement d'un ensemble de racines communes, sur le même principe que HTC. Les différences fondamentales qui existent entre WCAF et HTC ne permettent toutefois pas d'en être certain. En effet, la taille des deux ensembles est difficilement comparable, et l'on a bien montré que WCAF remontait à huit racines distinctes, ce qui paraît beaucoup pour expliquer l'émergence d'un trait prétendument homogène sur le plan étymologique. HTC, en comparaison, comprend trois racines majeures, à l'origine de la quasi-totalité des lexèmes de cette catégorie, ce qui est peu. Il nous faut toutefois ajouter que les diverses racines de WCAF pourraient être d'ascendance commune, mais l'OED ne nous renseigne pas davantage sur ce sujet. C'est d'ailleurs en partie pour cette raison que l'utilisation d'autres outils de recherche étymologiques paraît indispensable, car, même si l'OED est très complet, le recours à un seul outil n'est jamais suffisant pour explorer toutes les facettes d'un problème. De plus, les lexèmes d'origine inconnue sont relativement nombreux, et dans des proportions diverses selon les traits sémantiques. Dans la mesure où une origine inconnue est susceptible d'indiquer une construction inhabituelle, qui n'entre pas dans les cadres philologiques conventionnels, le rôle de ces lexèmes dans la formation des phonesthèmes doit être questionné.

Des analyses distributionnelles à partir de l'outil Sketch Engine représenteraient une source de données complémentaire à notre analyse, car elles permettraient de sortir du cadre lexicographique, dans une perspective contextuelle. En effet, les données empiriques de nature lexicographique que nous avons analysées doivent être mises en perspective, notamment grâce à l'analyse d'un ou de plusieurs corpus historiques. Un corpus tel que le *Early English Books Online Corpus (EEBO)* serait particulièrement intéressant. La question de la nature du trait

WCAF, par exemple, pourrait être élucidée si l'on prenait en compte les contextes d'apparition des lexèmes qui lui sont associés (voir par exemple Rychlý et Kilgarriff 2007, Lenci 2008, Rupert *et al.* 2015). Encore une fois, les travers liés à la nature des sources, et en l'occurrence à la comparaison de textes à la fois médiévaux, modernes et contemporains, représentent un enjeu méthodologique non-négligeable.

BIBLIOGRAPHIE

Outil de recherche

The Oxford English Dictionary, version en ligne pour abonés, 3ème édition. Oxford: Oxford University Press. <https://www-oed-com.ezproxy.normandie-univ.fr/>

Références

- Abelin, Åsa. 1999. *Studies in Sound Symbolism* (Gothenburg Monographs in Linguistics 17). Göteborg University.
- Abramova, Ekaterina, Raquel Fernandez & Federico Sangati. 2013. Automatic Labeling of Phonesthemic Senses. *UC Merced* (Proceedings of the Annual Meeting of the Cognitive Science Society) 35(35), 1696–1701.
- Ahlsén, Elisabeth. 2006. *Introduction to Neurolinguistics*. Amsterdam / Philadelphia: John Benjamins.
- Allan, Kathryn. 2012. Using OED Data as Evidence. In Kathryn Allan & Justyna A. Robinson (eds.), *Current Methods in Historical Semantics* (Topics in English Linguistics), 17–39. Berlin/Boston: De Gruyter Mouton.
- Allan, Keith. 2001. *Natural Language Semantics*. Oxford: Blackwell.
- Anttila, Raimo. 1977. *Analogy* (Trends in Linguistics, State-of-the-Art Reports 10). The Hague: De Gruyter Mouton.
- Bauer, Laurie. 2018. Conversion as Metonymy. *Word Structure* 11(2).
- Bergen, Benjamin K. 2004. The Psychological Reality of Phonesthemes. *Linguistic Society of America* (Language) 80(2), 290–311.
- Bergen, Benjamin K. & Joachim De Beule. 2006. On the Emergence of Compositionality. In Angelo Cangelosi, Andrew D.M. Smith & Kenny Smith (eds.), *The Evolution of Language*, 35–42. Rome.
- Bloomfield, Leonard. 1914. *An Introduction to the Study of Language*. Londres: G. Bell and Sons.
- Bloomfield, Leonard. 1926. A Set of Postulates for the Science of Language. *Linguistic Society of America* (Language) 2(3), 153–164.
- Bloomfield, Leonard. 1993. *The Teaching of Languages*. Landmarks of American Language & Linguistics.
- Blust, Robert. 1988. *Austronesian Root Theory: An Essay on the Limits of Morphology*. Amsterdam: John Benjamins.
- Blust, Robert. 2003. The Phonestheme *ŋ*- in Austronesian Languages. *Oceanic Linguistics* (Project Muse) 42(1), 187–212.

- Bolinger, Dwight L. 1940. Word Affinities. *Duke University Press* (American Speech) 15(1), 62–73.
- Bolinger, Dwight L. 1948. On Defining the Morpheme. *Word* 4(1), 18–23.
- Bolinger, Dwight L. & Donald A. Sears. 1981. *Aspects of Language*. New York: Houghton Mifflin Harcourt.
- Bottineau, Didier. 2008. The Submorphemic Conjecture in English: Towards a Distributed Model of the Cognitive Dynamics of Submorphemes. *Lexis* (Lexical Submorphemics) 2, 17–40.
- Bottineau, Didier. 2012. Submorphémique et corporéité cognitive. *Miranda* 7.
- Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.
- Culioli, Antoine. 1990. *Pour une linguistique de l'énonciation : Opérations et représentations*. Paris: Ophrys.
- Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer & Richard Harshman. 1990. Indexing by Latent Semantic Analysis. *Journal of the Association for Information Science and Technology* 41(6), 391–468.
- Drellishak, Scott. 2006. *Statistical Techniques for Detecting and Validating Phonesthemes*. University of Washington.
- Eschbach, Achim (ed.). 1983. *Charles S. Peirce, 'Studies in Logic' by Members of the Johns Hopkins University, 1883*. Vol. 1. Amsterdam/Philadelphia: John Benjamins.
- Firth, John Rupert. 1930. *Speech*. London: Ernest Benn.
- Flaksman, Maria. 2016. Iconic Words in Proto-Germanic. In Maria Flaksman & O.I. Brodovich (eds.), *Phonosemantics : in commemoration of Professor Dr. Stanislav Voronin's 80th anniversary* (Anglistics of the XXI Century), vol. 2, 39–51. Saint Petersburg: Saint Petersburg State University.
- Flaksman, Maria. 2017. Iconic Treadmill Hypothesis: The Reasons Behind Continuous Onomatopoeic Coinage. In Angelika Zirker (ed.), *Dimensions of Iconicity* (Iconicity in Language and Literature 15). John Benjamins.
- Gries, Stefan Th. 2009. What Is Corpus Linguistics? *Language and Linguistics Compass* 3(5), 1225–1241.
- Haiman, John. 1980. The Iconicity of Grammar: Isomorphism and Motivation. *Linguistic Society of America* (Language) 56(3), 515–540.
- Haiman, John. 1985. Symmetry. *Iconicity in Syntax: Proceedings of a symposium on iconicity in syntax, Stanford, June 24–26, 1983* (Typological Studies in Language), 73–96. John Benjamins.
- Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: The McMillan Company.

- Hoffmann, Sebastian. 2004. Using the OED Quotations Database as a Corpus: A Linguistic Appraisal. *ICAME Journal*(28).
- Hutchins, S.S. 1998. *The Psychological Reality, Variability, and Compositionality of English Phonesthemes*. Emory University.
- Jakobson, Roman. 1965. A la recherche de l'essence du langage. *NRF Gallimard* (Diogène) 51.
- Jakobson, Roman. 1978. *Six Lectures on Sound and Meaning*. Cambridge/London: The MIT Press.
- Jakobson, Roman. 1980. *The Framework of Language* (Michigan Studies in the Humanities). University of Michigan Press.
- Jespersen, Otto. 1922. *Language: Its Nature* (Development and Origin). London: Allen&Unwin.
- Kempey, S. T. & John Morton. 1982. The Effects of Priming with Regularly and Irregularly Related Words in Auditory Word Recognition. *British Journal of Psychology* 73(4), 441–454.
- Koch, Peter & Daniela Marzo. 2007. A Two-Dimensional Approach to the Study of Motivation in Lexical Typology and Its First Application to French High-Frequency Vocabulary. *John Benjamins* (Studies in Language) 31(2), 259–291.
- Kolin, Philip C. 1979. The Pseudo-Suffix *-oholic*. *Duke University Press* (American Speech) 54(1), 74–76.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago and London: The University of Chicago Press.
- Lakoff, George. 2012. Explaining Embodied Cognition Results. *Department of Linguistics, University of California* (Topics in Cognitive Science) 4(4), 773–785.
- Lamb, Sydney M. 1999. *Pathways of the Brain*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites*. Stanford University Press.
- Langacker, Ronald W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar* (Cognitive Linguistic Research). Berlin/New York: De Gruyter.
- Lenci, Alessandro. 2008. Distributional Semantics in Linguistic and Cognitive Research. *Italian Journal of Linguistics*.
- Liberman, Anatoly. 2008. *An Analytic Dictionary of English Etymology: An Introduction*. Minneapolis: University of Minnesota Press.
- Marchand, Hans. 1957. Motivation by Linguistic Form: English Ablaut and Rime Combinations and Their Relevancy to Word-Formation. *Studia Neophilologica* 29(1), 54–66.

- Martinet, André. 1970. *La linguistique synchronique* (Collection Sup). Presses Universitaires de France.
- Milroy, James. 2007. The History of English. In David Britain (ed.), *Language in the British Isles*, 9–33. Cambridge: Cambridge University Press.
- Monneret, Philippe. 2020. Le symbolisme phonétique et la fonction iconique de l’analogie. *Signifiances* (Signifying) 3(1), 1–19.
- Mosegaard Hansen, Maj-Britt. 2008. *Particles at the Semantics / Pragmatics Interface: Synchronic and Diachronic Issues* (Current Research in the Semantics/Pragmatics Interface 19). Amsterdam: Elsevier.
- Otis, Katya & Eyal Sagi. 2008. Phonesthemes: A Corpus-Based Analysis. *Proceedings of the Annual Meeting of the Cognitive Science Society* 30(30), 65–70.
- Permiss, Pamela, Robin L. Thompson & Gabriella Vigliocco. 2010. Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology* (Language Sciences) 31(227), 1–15.
- Philps, Dennis. 2002. Le concept de marqueur sub-lexical et la notion d’invariant sémantique. *De Boeck Supérieur* (Travaux de Linguistique) 2(45), 103–123.
- Philps, Dennis. 2008. Submorphemic Iconicity in the Lexicon: A Diachronic Approach to English ‘gn- Words.’ *Lexis [online]* (Lexical Submorphemics) 2.
- Philps, Dennis. 2012a. Le concept de « marqueur sub-lexical » : bilan d’un ballon d’essai. *Anglophonia/Sigma* 16(32), 183–202.
- Philps, Dennis. 2012b. Submorphemes: Backtracking from English ‘kn- words’ to the Emergence of the Linguistic Sign. *Miranda: Multidisciplinary peer-reviewed journal on the English-speaking world* 7, 1–20.
- Ruppert, Eugen, Manuel Kaufmann, Martin Riedl & Chris Biemann. 2015. A Web-Based Visualization for Graph-Based Distributional Semantic Models. *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, 103–108. Beijing: Association for Computational Linguistics and The Asian Federation of Natural Language Processing.
- Rychlý, Pavel & Adam Kilgarriff. 2007. An Efficient Algorithm for Building a Distributional Thesaurus (and Other Sketch Engine Developments). *Association for Computational Linguistics* (Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion) Proceedings of the Demo and Poster Sessions, 41–44.
- Sadowski, Piotr. 2009. *From Interaction to Symbol* (Iconicity in Language and Literature 8). Amsterdam & Philadelphia: John Benjamins.
- Sagi, Eyal, Armelle Boussidan & Sabine Ploux. 2009. Phonaesthetic and Etymological Effects on the Distribution of Senses in Statistical Models of Semantics. In Yves Peirsman, Yannick Versley & Tim Van de Cruys (eds.), *Proceedings of the Cognitive Science Conference 2009 Workshop on Distributional Semantics beyond Concrete Concepts*, 35–40. Amsterdam.

- Sagi, Eyal, Stefan Kaufmann & Brady Clark. 2011. Tracing Semantic Change with Latent Semantic Analysis. In Kathryn Allan & Justyna A. Robinson (eds.), *Current Methods in Historical Semantics* (Topics in English Linguistics 73), 161–183. Berlin/Boston: Mouton de Gruyter.
- Sapir, Edward. 1929. A Study in Phonetic Symbolism. *Journal of Experimental Psychology* 12(3), 225–239.
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. (Ed.) Charles Bally & Albert Sechehaye. Paris: Payot.
- Smith, Chris A. 2014. The Phonesthetics of Blends: A Lexicographic Study of Cognitive Blends in the OED. *De Gruyter* (Exploration in English Language and Linguistics (ExELL)) 2(1), 12–45.
- Smith, Chris A. 2016. Tracking Semantic Change in *fl*- Monomorphemes. *Journal of Historical Linguistics* 6(2), 165–200.
- Smith, Chris A. 2020. Approche cognitive diachronique de l'émergence du phonesthème *fl* : réanalyse phonosymbolique et transmodalité dans le Oxford English Dictionary. *Signifiances* (Signifying) 3(1), 36–62.
- Stefanowitsch, Anatol. 2002. Sound Symbolism in Usage-Driven Model. *Rice University* [unpublished MS].
- Stefanowitsch, Anatol. 2017. *Corpus Linguistics: A Guide to the Methodology*. Language Science Press [ms. non-publié].
- Upton, Clive, David Parry & J. D. A. Widdowson. 1994. *Survey of English Dialects: The Dictionary and Grammar*. London: Routledge.
- Wallis, John. 1653. *Grammatica Linguae Anglicanae*.
- Waugh, Linda R. 1994. Degrees of Iconicity in the Lexicon. *Journal of Pragmatics* 22(1), 55–70.
- Wolf, Werner. 2001. The Emergence of Experiential Iconicity and Spatial Perspective in Landscape Descriptions in English Fiction. In Olga Fischer & Max Nänny (eds.), *The Motivated Sign: Iconicity in language and literature*, vol. 2, 323–350. Amsterdam/Philadelphia: John Benjamins.
- Wright, David. 2012. Scrunch, Growze, or Chobble?: Investigating Regional Variation in Sound Symbolism in the Survey of English Dialects. *University of Leeds* 17.