

UN CAS D'EXTRAGRAMMATICALITÉ PARTICULIER : LES AMALGAMES LEXICAUX FANTAISISTES

INTRODUCTION

Si Sablayrolles (2000 : 224) dénombre une dizaine de termes tels que mot-valise, mot-centaure, mot-gigogne ou mot-tiroir pour désigner les amalgames lexicaux, ces unités n'ont paradoxalement pas beaucoup bénéficié du regard des linguistes. Seules quelques études – qui seront détaillées *infra* – concernent la place de ces unités lexicales en français. Plusieurs facteurs sont susceptibles d'expliquer ce manque d'intérêt. Galisson (1987 : 58) ou Sablayrolles (1997 : 30) soulignent notamment leur caractère éphémère. La collecte de données s'en trouve alors entravée bien que l'utilisation d'outils informatiques¹ confère plus de facilité dans cette tâche².

L'amalgamation lexicale mérite pourtant que l'on s'y intéresse, peut-être justement parce qu'elle relève de l'extragrammatical. En effet, cette matrice de création lexicale est particulièrement productive³. Des amalgames lexicaux sont apparus dans des lexiques de spécialité comme la publicité, le commerce ou encore la physique et la chimie. L'intérêt didactique⁴ de l'amalgamation lexicale est également un élément à prendre en considération.

Dans cet article, nous ne reviendrons pas sur la nature extragrammaticale des amalgames lexicaux puisque nous adhérons à l'étude conduite par Fradin, Montermini et Plénat (2009). En revanche, nous désirons dresser un état des lieux des recherches actuelles et initier une description pour mieux cerner le phénomène. Cette description s'appuie sur un corpus original constitué de 7052 amalgames lexicaux issus de dictionnaires fantaisistes. Ce corpus, de nature similaire à celui de Galisson (1987) mais de taille bien plus importante, doit fournir autant d'informations sur les structures morpho-phonologiques de ces unités que possible, dans le cadre d'une création lexicale *in vitro*. Les résultats obtenus sont ensuite confrontés à un second corpus composé d'amalgames lexicalisés, voire lexicographiquement attestés, issus de différentes sources. Ce second corpus est, de fait, quantitativement moins important que le premier, dans la mesure où nous avons déjà pointé la difficulté à rassembler des données attestées. La mise en regard de ces deux corpus per-

* Adresse de l'auteur: LDI, Université de Cergy-Pontoise, 33 boulevard du Port, 95011 Cergy-Pontoise, France. Mél: arnaud.leturgie@gmail.com

¹ La base de données néologique Neologia du laboratoire LDI (CNRS/universités Paris 13 et Cergy-Pontoise) fournit un nombre important de données pour la collecte d'amalgames lexicaux.

² Cependant, il n'existe pas, à notre connaissance, un outil permettant l'identification automatique des amalgames lexicaux.

³ Fradin, Montermini et Plénat (2009).

⁴ Cf. Galisson (1987) et Léturgie (à paraître) sur ce sujet.

met de mettre en lumière des correspondances et des divergences qui offrent une description générale de l'amalgamation lexicale.

Dans un premier temps les principales approches théoriques de l'amalgamation lexicale seront présentées. Cela sera l'occasion d'évaluer les différences qui les séparent et les lacunes qu'elles peuvent présenter. Puis, une approche prototypiste du phénomène sera introduite. Cette approche permet notamment de s'affranchir des divergences qui divisent les études antérieures afin d'offrir un cadre de recherche plus souple. Dans un second temps, ce sont les corpus sur lesquels nous nous appuyons qui seront présentés. La mise en relation d'un corpus *in vitro* et d'un corpus *in vivo* illustre un rapport similaire à celui entretenu par les morphologies grammaticale et extragrammaticale. Enfin, la recherche d'un prototype pour le français sera amorcée. La description proposée dans cet article doit permettre de mieux percevoir l'amalgamation de façon générale, notamment pour apporter de la documentation sur l'extragrammaticalité du phénomène.

1. L'AMALGAMATION LEXICALE : DIFFÉRENTES APPROCHES THÉORIQUES

Comme annoncé en introduction, plusieurs études ont été menées sur l'amalgamation lexicale en français. Mais ces études sont assez anciennes (milieu des années 80) et n'ont pas fait l'objet de recherches suivies. De plus, les méthodes d'identification et d'analyse des amalgames lexicaux proposées dans ces études diffèrent et ne reposent pas sur les mêmes critères. Les quatre approches les plus significatives pour le français présentent effectivement des différences significatives qu'il faut souligner. Les trois premières sont apparues en trois années d'intervalle et suivent des méthodes bien différentes. Alors que la première approche de Grésillon (1984) est, encore aujourd'hui, l'étude centrale du phénomène pour le français, celles de Clas (1987) et Galisson (1987) sont moins souvent citées, particulièrement pour la dernière. Plus récemment, Fradin (1997 et 2000) et Fradin, Montermini et Plénat (2009) ont évoqué les aspects sémantiques et phonologiques des amalgames lexicaux en français.

Précisons d'emblée que l'aspect sémantique de l'amalgamation lexicale ne sera pas traité dans cet article. Les corpus sur lesquels sont basées les recherches présentées ici n'ont pas encore fait l'objet d'un dépouillement de type sémantique. Afin de livrer une étude aussi précise que possible, nous limitons volontairement notre approche à la morphologie et à la phonologie.

Ces quatre approches vont être présentées ici. Puis, un modèle théorique assez récent proposant de procéder par prototypie, et que nous avons décidé d'adopter, sera introduit.

1.1. Approches pour le français

Puisque l'étude que nous proposons concerne le français, seules les approches concernant cette langue seront développées ici. Les quatre approches retenues sont les plus significatives mais ne proposent pas nécessairement de situer la place de

l'amalgamation lexicale au sein de la morphologie. En effet, Clas (1987) et Galisson (1987) traitent le phénomène sans insister sur son caractère extragrammatical alors que Grésillon (1984) qualifie de « monstres » les amalgames lexicaux pour traduire ce caractère. Enfin, Fradin (2000) puis Fradin, Montermini et Plénat (2009) installent l'amalgamation lexicale au sein de l'opposition entre morphologie grammaticale et morphologie extragrammaticale.

1.1.1. Une approche globale : Grésillon (1984)

La première étude d'importance est celle conduite par Grésillon (1984). C'est encore aujourd'hui l'étude de référence pour les chercheurs qui s'intéressent au phénomène de l'amalgamation lexicale en français. Elle détermine des critères définitoires permettant d'identifier et de catégoriser les amalgames lexicaux. Selon elle, une unité lexicale doit être dotée de plusieurs propriétés (p. 12-13) pour être considérée comme un amalgame. Cependant, c'est la présence d'un segment homophone, ou quasi-homophone aux bases-sources qui incarne « la condition *sine qua non* de la formation du mot-valise » (p. 15). Elle envisage donc quatre schémas de formation (p. 24-25) :

- avec segment homophone = *enfantaisie* < enfant + fantaisie ;
- avec troncation = *infiniverti* < infini + inverti ;
- avec segment homophone et troncation = *éphémérde* < éphémère + merde ;
- avec enchâssement = *enseignement* < enseignement + saigne.

Cette approche a l'avantage de décrire la formation des amalgames d'un point de vue morpho-phonologique et les schémas proposés s'avèrent assez complets pour être appliqués à des corpus de grande ampleur. Pourtant, le corpus principal sur lequel repose ce travail est composé d'une soixantaine de termes allemands créés par Heine, poète allemand du XIX^e siècle, complété d'un corpus annexe de nature très hétérogène⁵. Or, étant donné la difficulté de tirer beaucoup de données à partir du corpus réduit sur lequel elle s'appuie, ces schémas semblent manquer de détails. Mais si Grésillon ne prend pas la peine de détailler les combinaisons morphologiques des amalgames qu'elle étudie, c'est peut-être que ces combinaisons étant imprédictibles – ce qui constitue d'ailleurs une des propriétés qui fait des amalgames lexicaux des unités morphologiquement extragrammaticales – il suffit de mentionner le phénomène de troncation. Néanmoins, situer le lieu de la troncation dans les bases-sources ne semble pas superflu. L'aspect phonologique est mieux documenté puisque les schémas de recouvrements homophoniques sont présentés (p. 19-20).

Un des problèmes à soulever est que les critères semblent devoir convenir aussi bien pour le français que pour l'allemand et l'anglais. Et on imagine qu'il doit en

⁵ Le corpus de Grésillon (1984) est constitué de termes issus de différentes sources telles que le journalisme, le commerce, la publicité ou l'édition, mais aussi de termes forgés particulièrement pour l'étude, ainsi que de répertoires de mots fantaisistes. Trois langues (allemand, français et anglais) sont utilisées et traitées indifféremment.

être de même pour toutes les autres langues. Seulement, des propriétés importantes – qui ne sont pas toujours partagées – de ces langues sont à prendre en compte, comme la structure accentuelle des unités lexicales. L’accentuation va jouer un rôle considérable dans l’opération de segmentation puis d’amalgamation des bases-sources dans des langues comme l’anglais⁶ ou l’hébreu⁷.

1.1.2. Typologie morphologique : Clas (1987)

Clas (1987) s’appuie sur un corpus de termes techniques, essentiellement issus de la chimie ou de la physique. Son analyse repose totalement sur la segmentation morphologique des unités lexicales. Il dresse donc une typologie des amalgames lexicaux en considérant 6 schémas de troncation. C’est effectivement selon la place où intervient la troncation que l’amalgame lexical est catégorisé. Il peut donc s’agir d’unités formées par:

- aphérèse d’une base-source = *robotique* < robot + informatique ;
- apocope d’une base-source = *publipostage* < publicité + postage ;
- aphérèse de la première base-source et apocope de la seconde base-source = *domotique* < domicile + robotique ;
- aphérèse des deux bases-sources = *nylon* < vinyl + coton ;
- apocope des deux bases-sources = *télex* < teleprinter + exchange ;
- aphérèse ou apocope d’une base-source et syncope de la deuxième base-source = *upérisation* < ultra + pasteurisation.

Bien qu’il soit morphologiquement plus détaillé que celui de Grésillon (1984), le modèle de Clas (1987) ne prend pas en considération tout l’aspect phonologique. Les cas d’homophonie indiqués par Grésillon ne peuvent être identifiés par l’approche de Clas alors que ces cas paraissent représenter plus fidèlement le concept d’amalgame lexical. Clas n’évoque finalement que la dimension concernée par l’extragrammaticalité. C’est en effet la morphologie, dégagée de ses interactions avec la phonologie, la syntaxe ou la sémantique qui est envisagée dans sa typologie. Le problème est que, ce faisant, le chercheur ne peut pas livrer tous les schémas nécessaires à une description de l’amalgamation lexicale. Nous verrons au §3 que ce modèle doit être adapté pour devenir réellement opératoire.

1.1.3. Approche ludo-didactique : Galisson (1987)

L’étude proposée par Galisson (1987) est assez atypique⁸. En effet, le corpus sur lequel il s’appuie est, à l’image de celui que nous proposons aujourd’hui⁹, composé entièrement d’amalgames fantaisistes puisqu’il repose sur 3 répertoires fantaisistes,

⁶ Bertinetto (2001) ou Gries (2004).

⁷ Bat-El (1996).

⁸ Cf. Léturgie (à paraître) pour plus de détails.

⁹ Cependant, le traitement qu’il applique à ce corpus est très différent de notre méthodologie.

dont son propre *Distractionnaire* (1986), ce qui représente environ un millier d'unités. Galisson distingue les unités qui relèvent de l'insertion de celles qui relèvent de l'imbrication, ces deux matrices connaissant des variations :

- insertion sans changement graphique = *ministère* < mini + stère + ministère ;
- insertion avec changement graphique du mot originel = *mélancolis* < mélancolie + colis ;
- insertion avec altération du mot originel = *néofrite* < néophyte + frite ;
- insertion avec substitution d'une partie du mot originel = *merfidie* < mère + perfidie ;
- imbrication = *ringargariser* < ringard + gargariser.

Les schémas proposés semblent être assez souples pour s'adapter à un corpus de grande ampleur. Cependant, la description du phénomène demeure assez pauvre et il n'est fait aucune mention de l'aspect morpho-phonologique de l'amalgamation lexicale. La typologie résultante est donc moyennement satisfaisante et l'approche préconisée par Grésillon (1984) semble encore la plus indiquée.

1.1.4. Les études récentes de Fradin

Enfin, plus récemment, Fradin (1997) et Fradin, Montermini et Plénat (2009) ont également tenté de décrire le phénomène. Fradin (1997) centre son analyse sur la sémantique des amalgames lexicaux. Fradin (2000) et Fradin et *al.* (2009) laissent plus de place à la description morpho-phonologique de ces unités lexicales.

Fradin (2000) et Fradin et *al.* (2009) offrent une description assez précise du phénomène en s'appuyant sur des données antérieures assez anciennes (dont Grésillon 1984). L'agencement phonologique est particulièrement observé, notamment en ce qui concerne les cas d'homophonie des bases-sources. L'analyse proposée par Fradin (2000 : 28-30) et Fradin et *al.* (2009) est très aboutie et a l'avantage de reposer sur de nombreuses études antérieures. Les schémas de construction phonologique sont très nombreux et détaillés, à tel point que nous ne disposons pas de la place nécessaire pour les reproduire ici.

Ces différentes approches ne pointent pas toutes dans la même direction. Ainsi, afin de ne pas se limiter à suivre telle ou telle approche, Renner (2006) préconise une description de l'amalgamation lexicale par prototypie.

1.2. Approche prototypiste

Les divergences théoriques qui viennent d'être illustrées par ces quatre approches ne concernent pas que la recherche francophone. Brdar-Szabó/Brdar (2008) notent également une absence de consensus entre les chercheurs.

There are so many things about [lexical blending] that remain to be investigated and so many issues on which there is no real consensus among researchers, from a satisfying definition to an extensional account of the phenomenon including an inventory of various subtypes. Brdar-Szabó/Brdar (2008 : 172)

Renner (2006) remarque lui aussi une multiplicité de points de vue sur le phénomène. Il propose néanmoins d'utiliser ces différentes approches afin de dégager une définition prototypique de l'amalgamation lexicale. En synthétisant les quatre contraintes les plus fréquentes parmi les études réunies, il dévoile un prototype général d'amalgame lexical. Cette approche est tout particulièrement intéressante dans la mesure où elle confronte non seulement des points de vues sur un même phénomène mais qu'elle s'inscrit dans une perspective contrastive puisqu'elle repose sur des études en français, anglais, espagnol, allemand et bulgare.

Il s'agira donc ici¹⁰ de s'appuyer sur cette étude afin d'illustrer les différentes façons d'appréhender ce phénomène. À partir d'une définition dite du « plus petit dénominateur commun » commune à toutes les théories, plusieurs critères défini- toires sont exposés afin de témoigner des divergences entre les différentes approches. Trois critères partagent les spécialistes sur la question et nous allons les évoquer brièvement ici à titre de rappel.

1.2.1. « Le plus petit dénominateur commun »

Revenons en tout premier lieu sur ce que Renner appelle « PPDC »¹¹ : la définition minimale implicitement acceptée par les linguistes lorsqu'il s'agit de définir l'amalgamation lexicale. Ainsi, « l'amalgame est issu de l'assemblage d'au moins deux lexèmes, processus au cours duquel au moins un des lexèmes perd une partie de son signifiant »¹². C'est une définition qui est, comme le souligne Renner, suffisante pour de nombreux linguistes¹³ ayant traité la question. Cependant, elle est insuffisante selon d'autres chercheurs et d'autres critères entrent en jeu pour déterminer de l'appartenance ou non à la catégorie des amalgames lexicaux.

1.2.2. Contrainte morphologique : TRONC INT

La première contrainte évoquée par Renner (2006 : 139) est la troncation interne (TRONC INT). Il s'agit de la caractéristique la plus courante et elle est d'ailleurs adoptée par de nombreux linguistes¹⁴. Cette contrainte « veut que le schéma de troncation corresponde à une apocope de la première base-source et/ou une aphérèse de la deuxième base-source » (Renner, *ibid.*). En d'autres termes, les mots composés par deux bases-sources apocopées ou aphérésées comme *chloroforme* (< chlore + for- myle) ; *pixel* (< picture + element) ou *nylon* (< vinyl + cotton) n'appartiendraient pas à la catégorie des amalgames lexicaux¹⁵. Cette contrainte en appelle une autre : la

¹⁰ Cette partie est identique à ce qui est présenté dans Leturgie (2011 : 77-79).

¹¹ Pour « plus petit dénominateur commun » (Renner 2006 : 139).

¹² *Ibid.*

¹³ Sont cités Adams (1973 : 137, 146), Algeo (1977 : 49-50), Clas (1987 : 348) et López-Rúa (2002 : 44, 46 ; 2004 : 65-66).

¹⁴ Renner cite Davy (2000 : 67), Paillard (2000 : 81), Bauer et Huddleston (2002 : 1635), Bauer (2003 : 47), Gries (2004 : 645-647) et Booij (2005 : 20). Ajoutons Bat-El (2006 : 66).

¹⁵ Il s'agirait de « clipped-compounds » (Bat-El 2006 : 66), soit de composés par troncation.

double troncation. Les amalgames dont les deux bases-sources sont tronquées, comme *alicament* (< aliment + médicament), respectent la contrainte DBLE TRONC.

1.2.3. *Contrainte sémantique : COORD*

La seconde caractéristique définitoire suppose une contrainte sémantique, la coordination (COORD). Renner explique que selon Kubozono (1990 : 1-2), Plag (2003 : 122) et Grésillon (1984 : 120-121) il existe une nuance entre les « blends »¹⁶ de type *motel* (< motor + hotel) et les « proper blends » de type *alicament* (< aliment + médicament) ou *modem* (< modulator + demodulator). Cette contrainte repose sur un lien de coordination entre les signifiés des bases-sources. Ainsi, les « blends » ne répondent pas positivement à l'énoncé-test « un X est à la fois un Y et un Z », tandis que les « proper blends » y répondent positivement. En effet, un *modem* est à la fois un modulateur et un démodulateur, tout comme un *alicament* est à la fois un aliment et un médicament. En revanche, un *motel* n'est pas à la fois un motor-car et un hôtel.

1.2.4. *Contrainte morpho-phonologique : INTERP*

La troisième et dernière caractéristique définitoire mise en évidence par Renner est d'ordre morpho-phonologique : l'interpénétration (INTERP). Selon certains linguistes¹⁷ il est nécessaire qu'un segment de l'amalgame soit commun aux bases-sources. Ce critère exclut donc des mots comme *brunch* (< breakfast + lunch) ou *caméscope* (< caméra + magnétoscope) de la catégorie, contrairement à des mots tels que *motel* pour lesquels il existe un segment commun. Le segment *ot* commun à *motor* et à *hotel* est dit ambimorphémique¹⁸.

1.2.5. *Un prototype pour le français?*

L'approche proposée par Renner (2006) est celle qui a été adoptée dans le cadre de notre recherche et de cette étude. Il s'agit pour lui de procéder à une approche prototypiste de l'amalgamation lexicale. En réunissant les études concernant le phénomène dans autant de langues que possible, on peut établir un certain nombre de contraintes récursives qui sont à respecter pour considérer l'unité comme un amalgame lexical, et ainsi déterminer un niveau de prototypie. Cette approche constitue un modèle complexe et suffisamment puissant pour déterminer ce niveau de prototypie, mais il faut, à notre sens, affiner ce modèle pour chaque langue. Comme il a été évoqué *supra*, certains facteurs ne concernent que certaines langues et, afin de proposer le meilleur candidat, il faut considérer chaque langue à partir de ce modèle général. C'est ce que nous proposons de faire dans le §3 en utilisant deux corpus complémentaires.

¹⁶ Selon la terminologie de Plag (2003).

¹⁷ Grésillon (1984 : 15), Bankov, Dimitrov et Dragnev (1989 : 10) et Piñeros (1999 : 1).

¹⁸ Piñeros (1999 : 14 ; 2004 : 225).

2. LES CORPUS

Étant donné qu’au regard des études antérieures sur le phénomène, on peut ne pas être entièrement satisfait des méthodes d’identification et de classification des amalgames lexicaux, nous proposons d’adopter une approche s’appuyant sur un corpus original. En effet, les corpus sur lesquels reposent les études précédentes présentent soit des natures très hétérogènes, soit comportent sur un nombre trop restreint d’items. Le caractère extragrammatical de l’amalgamation lexicale ne doit pas parasiter la cohérence du corpus. Ainsi, Grésillon (1984) et Fradin (1997) s’appuient sur des corpus trop hétérogènes pour tirer des résultats généraux. Galisson (1987) s’appuie sur un corpus intéressant, constitué de mots fantaisistes mais qui ne démontre rien de la langue courante. Clas (1987) utilise des données attestées mais uniquement issues de lexiques scientifiques qui risquent donc de présenter des particularités propres à la terminologie scientifique.

L’utilisation parallèle de deux corpus distincts et complémentaires a été avancée en introduction. Cette section précise la nature de ces corpus. Le premier, que nous avons baptisé COAL¹⁹ est le plus important quantitativement puisqu’il est composé d’environ 30 fois plus d’amalgames que le second. Mais ces corpus ne s’opposent pas uniquement par le nombre de données qu’ils renferment. COAL est constitué d’amalgames lexicaux tirés de dictionnaires fantaisistes, c’est-à-dire de termes forgés *in vitro*, dans le cadre d’un projet lexicographique. Le second corpus est un agglomérat de termes issus de différentes sources – qui vont être détaillées *infra* – et qui ont été forgés *in vivo*, c’est-à-dire dans le cadre de situations discursives.

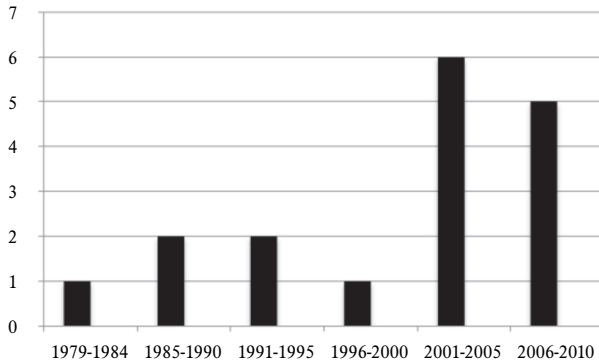
Le rapport entre ces deux corpus est, dans une certaine mesure, comparable à l’opposition entre grammaticalité et extragrammaticalité. Les amalgames créés *in vivo* répondent à un besoin de communication, tandis que les termes contenus dans COAL sont factices et n’ont pas vocation à apparaître ailleurs que dans les dictionnaires d’où ils sont tirés.

2.1. Des dictionnaires fantaisistes aux amalgames lexicaux

COAL est un corpus construit par l’extraction de 7052 amalgames lexicaux de 17 dictionnaires fantaisistes. Les dictionnaires d’amalgames lexicaux utilisés pour constituer COAL appartiennent à une catégorie de dictionnaires particuliers : les dictionnaires fantaisistes²⁰. Ce type de dictionnaire est singulier à plusieurs titres. Tout d’abord le nombre de dictionnaires publiés au cours des 30 dernières années est particulièrement important. Depuis 1979, 17 dictionnaires de mots-valises fantaisistes ont paru selon la chronologie illustrée par le graphique 1. Ainsi en 9 ans (2001 à 2010), pas moins de 11 dictionnaires de mots-valises ont vu le jour en France, ce qui est considérable pour ce type de littérature.

¹⁹ Pour COrpus d’Amalgames Lexicaux.

²⁰ Cf. Léturgie (2010).



Graphique 1 : répartition chronologique des dictionnaires d'amalgames depuis 1979

Bien que ces différents dictionnaires répondent principalement à une fonction ludique, ils ne s'y limitent pas tous. Certains rédacteurs de ces dictionnaires nourrissent en effet un objectif pédagogique, et cherchent à utiliser l'amalgamation lexicale comme outil de sensibilisation au lexique du français. C'est notamment le cas de Galisson et Porcher (1986) qui ont conçu le *Distractionnaire* comme un outil didactique à destination des locuteurs étrangers. Nous avons, par ailleurs, eu l'occasion de constater que des ateliers de création de mots-valises sont régulièrement organisés dans les classes de l'enseignement primaire. Il est intéressant de constater que le caractère extragrammatical de l'amalgamation lexicale n'entre pas en conflit avec leur emploi dans un cadre didactique²¹.

COAL est une base de données de 7052 items dont les bases-sources ont été identifiées. Chaque amalgame est catégorisé selon son schéma de construction en suivant les modèles de Clas (1987) et de Grésillon (1984) et étiqueté en fonction des contraintes de Renner (2006).

2.2. Un corpus *in vivo*

En regard de COAL, il importe de constituer un corpus de mots forgés *in vivo*. Ce corpus, auquel nous ferons référence par le sigle CALA²², comporte aujourd'hui 214 termes issus de différentes sources. La source primaire utilisée pour cette étude est le corpus français de Grésillon (1984 : 176-186) pour 130 items, puis une partie de la base de données *Neologia* pour 56 items et enfin le *Grand Robert* 2009 pour 28 items. CALA n'a pas encore bénéficié d'une analyse sur l'ensemble de ses items et on estime à un peu plus de 250 le nombre d'items qu'il réunira.

²¹ Cf. Léturgie (à paraître) pour plus de détails au sujet de l'utilisation didactique d'amalgames lexicaux.

²² Pour Corpus d'Amalgames Lexicaux Attestés.

Le corpus de Grésillon (1984) a fait l'objet d'un dépouillement afin de ne conserver que les items issus de certaines catégories discursives. Ont donc été conservés les termes marqués comme venant de l'audiovisuel, du commerce, de l'édition, du journalisme, et de la publicité ; les amalgames issus de textes littéraires ou produits par des malades mentaux ont été écartés. Les amalgames bénéficient des mêmes renseignements que dans COAL.

3. À LA RECHERCHE D'UN PROTOTYPE POUR LE FRANÇAIS

Maintenant que les corpus ont été présentés, il est possible de leur appliquer les différentes théories introduites au §1. Les deux corpus seront dans un premier temps interrogés par les approches classiques, c'est-à-dire les quatre approches principales pour le français. Cette première analyse va faire apparaître un schéma d'amalgamation prototypique qu'il faudra corrélérer aux contraintes proposées par Renner (2006).

3.1. Interrogation des corpus par les approches classiques

Seuls les modèles de Clas (1987) et de Grésillon (1984) vont être appliqués aux corpus. La théorie de Galisson (1987) n'apporte pas de données supplémentaires à ces deux études et va donc être mise de côté. L'analyse menée par Fradin (2000) nécessite une longue recherche et ne peut être présentée dans cet article. Cependant, l'analyse des structures phonologiques des amalgames fantaisistes est un travail que nous souhaitons conduire dans notre thèse.

Les corpus vont livrer des informations sur les structures morphologiques, puis morpho-phonologiques dominantes en français. Nous allons voir que, comme on l'avait postulé, les résultats fournis par COAL coïncident avec ceux de CALA.

3.1.1. Approche morphologique

La première analyse est d'ordre morphologique. En adaptant la typologie de Clas (1987), on obtient des données sur la morphologie des amalgames lexicaux. Rappelons que Clas (1987) repose sur un corpus de termes techniques, essentiellement de physique ou chimie. Les 6 premiers schémas sont ceux originellement présentés par Clas (1987) tandis que le dernier concerne les cas d'interpénétrations parfaites. Le modèle original ne laissant pas de place à l'aspect phonologique dans la formation des amalgames, il a fallu ajouter ce schéma puisque 1424 items de COAL suivent ce patron.

type d'amalgamation	exemple in vivo	CALA		exemple in vitro	COAL	
		#	%		#	%
(1) apocope + aphérèse	domotique < dom[icile] + [rob]otique	40	18,7%	camembour < cam[embert] + [cal]embour	811	11,5%
(2) apocope + apocope	telex < tel[eprinter] + ex[change]	2	0,9%	∅	0	0,0%
(3) aphérèse +aphérèse	nylon < vinyl + coton	0	0,0%	∅	0	0,0%
(4) apocope simple	publipostage < publi[cité] + postage	87	40,7%	scrypte < s[cript] + crypte	3408	48,3%
(5) aphérèse simple	robotique < robot + [informa]tique	46	21,5%	toasticomane < toast + [tox]icomane	1081	15,3%
(6) apocope/aphérèse + syncope	upérisation < u[ltra] + p[ast]e[u]risation	13	6,1%	escameloter < esca[m]oter + [ca]melote	328	4,7%
(7) interpénétrations parfaites	∅	26	12,1%	bovidéaliste < bovidé + idéa-liste	1424	20,2%
		214	100%		7052	100%

Tableau 1 : analyse du corpus selon la méthode de Clas (1987)

Le tableau 1 révèle que la majorité des amalgames des deux corpus est formée par apocope simple. Il faut en réalité entendre que c'est la base-source gauche qui est apocopée et que la base-source droite apparaît entièrement. On remarque que les autres schémas sont plutôt proportionnels entre les deux corpus. On remarque cependant que les cas (2) et (3) sont absents de COAL alors que le patron (2) apparaît à deux reprises dans CALA. Si le schéma (3) n'apparaît pas, c'est que le corpus de Clas (1987) n'a pas encore été intégré à CALA. Il est intéressant de constater que COAL ne comporte pas d'unités adoptant la structure (2) ou (3). Cela signifie que la contrainte TRONC INT est respectée par 100% des amalgames fantaisistes. Ce que l'on note également c'est que les 2 termes qui correspondent à la structure (2) proviennent du *GR* et sont des termes techniques empruntés à l'anglais (*algol* < algorithmic language et *fortran* < formular translation). Le schéma ne semble donc pas productif en français dans CALA non plus.

Le problème de cette analyse est que toute considération sur la phonologie des amalgames est écartée. Or, la présence de segments homophones joue un rôle de première importance. Rappelons que Grésillon (1984 : 15) considère que la présence

d'un segment (quasi-)homophone comme « la condition *sine qua non* de la formation du mot-valise ». Fradin, Montermini et Plénat (2009 : 35) signalent qu'« un survol des données disponibles [...] permet de supposer que les mots-valises maximisent en leur sein le nombre et l'étendue des tronçons communs aux lexèmes dont ils sont issus. » C'est pour cette raison qu'il est nécessaire de conduire une analyse d'ordre morpho-phonologique.

3.1.2. Approche morpho-phonologique

En reprenant les schémas d'amalgamation de Grésillon (1984), il est possible de quantifier la proportion d'items créés par troncation ou par enchâssement tout en mesurant la part d'homophonie au sein de chaque matrice. Le segment homophone identifié ici correspond à la contrainte INTERP. Il s'agit donc de ce que Piñeros (1999 et 2004) appelle segment ambimorphémique : un segment homophone et situé au point de jonction des bases-sources. Les segments homophones situés à une autre place dans l'amalgame ne sont pas pris en compte ici. Il serait pourtant intéressant de vérifier la place de ce type de segment à la manière de ce qu'à fait Fradin (2000 : 28-30).

type d'amalgamation	exemple in vivo	CALA		exemple in vitro	COAL	
		#	%		#	%
(8) avec segment homophone	enfantaisie < enfant + fantaisie	26	12,1%	bovidéaliste < bovidé + idéaliste	1427	20,2%
TOTAL homophonie		26	12,1%		1427	20,2%
(9a) avec troncation (aphérèse + apocope)	amarouché < am[ouraché] + [effa]rouché	20	9,3%	urigence < uri[ner] + [ur]gence	191	2,7%
(9b) avec troncation (apocope)	bavardhurler < bavarder + hurler	21	9,8%	coquinlicot < coquin + coquelicot	421	6,0%
(9c) avec troncation (aphérèse)	infiniverti < infini + [in]verti	14	6,5%	ripoublicain < ripou + [répu]blicain	335	4,8%
TOTAL troncation		55	25,7%		947	13,4%
(10a) avec segment homophone et troncation (aphérèse + apocope)	falsifis < fal[sifier] + [sal]sifis	20	9,3%	racifisme < ra[ci]sme + [pa]cifisme	621	8,8%
(10b) avec segment homophone et troncation (apocope)	éphémerde < éphé[mère] + merde	68	31,8%	buthym < bu[ti]n + thym	2982	42,3%
(10c) avec segment homophone et troncation (aphérèse)	bisouterie < bisou + [bijou]terie	32	15,0%	rénalcitrant < rénal + [récal]citrant	747	10,6%
TOTAL troncation + homophonie		120	56,1%		4350	61,7%
(11a) avec enchâssement	ubiamourquité < ubiquité + amour	11	5,1%	obchèqueieux < obséqueux + chèque	77	1,1%
(11b) avec enchâssement et homophonie	enseignement < en[seigne]ment + saigne	2	0,9%	désautobéis- sance < désobéis- sance + auto	251	3,6%
TOTAL enchâssement		13	6,1%		328	4,7%
		214	100%		7052	100%

Tableau 2 : analyse du corpus selon la méthode de Grésillon (1984)

Le tableau 2 met plusieurs éléments en évidence. Tout d'abord le rapport entre (8) et (9) est presque inverse entre COAL et CALA. Les schémas de troncation sans homophonie sont préférés au schéma (8) dans CALA alors que les amalgames fantaisistes respectent la contrainte INTERP. Globalement, c'est malgré tout le schéma (10b) qui s'impose dans les deux corpus. Ce schéma correspond simplement au schéma (4) avec respect de la contrainte INTERP, c'est-à-dire que la base-source droite partage au moins un phonème commun avec la base-source gauche.

Les schémas (11a) et (11b) sont, à l'image de ce qu'on vient d'observer entre (8) et (9), en miroir. L'enchâssement sans homophonie est préféré par CALA tandis que COAL compte trois fois plus de termes répondant à (11b).

Les résultats recensés dans le tableau 2 confirment la tendance dégagée par le tableau 1 : la majorité des amalgames lexicaux sont construits par apocope de la base-source gauche et comportent un segment ambimorphémique. On s'aperçoit que cette dernière propriété est importante puisque dans le cas des troncations sans homophonie, le schéma (9c) est à peine plus fréquent que (9a) pour CALA ou que (9c) pour COAL.

3.2. Mesure de prototypie des corpus

À partir des données qui ont été réunies par l'analyse des nomenclatures de 17 dictionnaires fantaisistes d'amalgames lexicaux et de leur comparaison avec des données attestées en français, il est possible de déterminer un prototype d'amalgame lexical. À l'aide des tableaux 1 et 2, on peut établir que le prototype répond aux propriétés morpho-phonologiques suivantes :

- l'unité est formée par amalgamation de deux lexèmes ;
- la base-source gauche subie une apocope ;
- la base-source droite reste intacte ;
- un segment homophone est présent au point de jonction des bases-sources.

Ces propriétés correspondent partiellement au schéma (4) et totalement au schéma (10b). Les unités qui ne respectent pas l'ensemble de ces propriétés (comme les termes construits par les différents schémas qui viennent d'être présentés) ne sont pas exclues de la catégorie des amalgames lexicaux mais s'échelonnent sur un degré de typicité moindre.

Si on transpose ces observations avec les contraintes proposées par Renner (2006), on obtient un respect de deux contraintes sur trois puisque nous écartons sciemment la contrainte COORD dans l'analyse proposée ici. COAL montre que dans 61,7% des cas, les contraintes TRONC INT et INTERP sont respectées, et 29%²³ si on ajoute la contrainte DBLE TRONC. En ce qui concerne CALA, 56,1% des amalgames respectent les contraintes TRONC INT et INTERP. Si la contrainte DBLE TRONC est ajoutée, il ne reste que 21,4% du corpus ce qui est inférieur à la proportion d'amalgames

²³ On additionne le schéma (8) et le (10a).

ne respectant que TRONC INT²⁴. Ce résultat peut surprendre puisqu'on semble s'éloigner du prototype proposé par Renner (2006). En réalité, nous pensons que le français a tendance à favoriser le schéma (10b) et que, même si l'unité résultante semble moins parfaitement construite qu'en suivant le schéma (8), il s'agit bien du prototype pour le français dans la mesure où la majorité des amalgames collectés dans les deux corpus correspondent au schéma (10b).

CONCLUSION

L'extragrammaticalité morphologique des amalgames lexicaux n'a pas fait l'objet de commentaires autres que les observations de Fradin (2000) et Fradin *et al.* (2009). Nous pensons qu'il est nécessaire de continuer à documenter le phénomène pour développer ces observations, et que cette documentation passe notamment par l'analyse d'un corpus tel que COAL. S'il ne permet pas de transférer ses résultats directement au lexique général du français, l'approche contrastive que nous proposons semble être un bon moyen de confirmer ou d'infirmer certaines propriétés des amalgames lexicaux qui ne constituaient finalement que des tendances sur des ensembles réduits de mots.

Les modèles théoriques présentés au §1 sont d'excellents outils d'analyse. On voit qu'en confrontant les données obtenues par application de ces théories aux deux corpus, un consensus apparaît. Étant donné que les résultats de COAL coïncident avec ceux de CALA, on voit véritablement un prototype, vérifié par un nombre conséquent de données, se dégager. Il n'est pas absurde de considérer le schéma le plus productif comme étant le plus prototypique dans la mesure où on s'appuie sur plus de 7000 items, ce qui représente une base de données suffisante. Nous désirons cependant continuer cette analyse en tenant compte de l'aspect sémantique des amalgames lexicaux. L'approche la plus complète est, avec celle de Grésillon (1984), celle de Fradin (1997, 2000). Il faut donc vérifier si les modèles théoriques de ces chercheurs s'adaptent à nos deux corpus et les résultats que l'on peut en tirer. Une dernière étape dans la mise au jour d'un prototype serait de conduire une étude contrastive sur plusieurs langues, à l'image de ce qu'a fait Renner (2006). Nous postulons en effet que le schéma le plus productif ne sera pas nécessairement le même dans toutes les langues, notamment du fait de la structure accentuelle des unités lexicales ; la phonologie jouant, comme on l'a vu, un rôle crucial dans la formation des amalgames.

Enfin, alors que leur extragrammaticalité morphologique – et notamment leur imprédictibilité – rend les amalgames lexicaux difficiles à détecter automatiquement, la description que nous proposons doit, à terme, permettre de fournir assez d'éléments pour élaborer des outils de détection automatique. Cette détection s'avère complexe du fait de l'extragrammaticalité de l'amalgamation lexicale et ce n'est qu'en accumulant un grand nombre de données qu'il sera possible de développer ce genre d'outils.

²⁴ 25,7% des amalgames de CALA ne respectent que la contrainte TRONC INT alors que le pourcentage tombe à 13,4% dans COAL.

Références bibliographiques

- ADAMS, Valerie (1973) *An Introduction to Modern English Word-Formation*. Harlow : Longman.
- ALGEO, John (1977) "Blends, a structural and systemic view." *American Speech* 52, 47-64.
- BANKOV, Dimitri/Dimitar DIMITROV/Vladimir DRAGNEV (1989) "Contribution à l'étude des mots-valises ou "mixonymes." *Supostavitelno Ezikoznanie* 14 (5), 8-11.
- BAT-EL, Outi (1996) "Selecting the best of the worst: the grammar of Hebrew blends." *Phonology* 13 (3), 283-328.
- BAT-EL, Outi (2006) "Blends." In K. Brown (éd.), 66-70.
- BAUER, Laurie (2003) *Introducing Linguistic Morphology*, 2nd ed., Edinburgh : Edinburgh University Press.
- BAUER, Laurie/Rodney HUDDLESTON (2002) "Lexical Word-formation." In : R. Huddleston/G. Pullum (éds), 1621-1721.
- BERTINETTO, Pier Marco (2001) "Blends and syllable structure: A four-fold comparison." In : Mercé Lorente et al. (éds), 59-112.
- BOOIJ, Geert (2005) *The Grammar of Words*. Oxford : Oxford University Press.
- BOOIJ, Geert/Jaap VAN MARLE (éds) (1990) *Yearbook of Morphology* vol. 3, Dordrecht : Foris Publications.
- BRDAR-SZABÓ, Rita/Mario BRDAR (2008) "On the marginality of lexical blending." *Jezikoslocije* 9/1-2, 171-194.
- BROWN, Keith (éd.) *Encyclopedia of Language and Linguistics*, 2nd ed., Vol. 2. Oxford : Elsevier.
- CLAS, André (1987) "Une matrice terminologique universelle : la brachygraphie gigogne." *Méta* 32 (3), 347-355.
- CORBIN, Danielle et al. (éds) (1997), *Sillexicales 1, Mots possibles et mots existants*. Villeneuve d'Ascq : URA 382 - CNRS/Université de Lille 3.
- DAVY, Dennis (2000) "Shortening Phenomena in Modern English Word Formation: An Analysis of Clipping and Blending." *Franco-British Studies* 29, 59-76.
- DOLESCHAL, Urusula/Anna M. THORNTON (éds) (2001) *Extragrammatical and Marginal Morphology*. München : Lincom Europa.
- DYKSTRA, Anne/Tanneke SCHOONHEIM (éds) (2010) *Proceedings of the XIV Euralex International Congress*. Ljouwert : Afûk/Fryske Academy.
- FRADIN, Bernard (1997) "Les mots-valises : une forme productive d'existants impossibles ?" In : D. Corbin et al. (éds), 101-110.
- FRADIN, Bernard (2000) "Combining forms, blends and related phenomena." In U. Doleschal /A. M. Thornton (éds), 11-59.
- FRADIN, Bernard/Fabio MONTERMINI/Marc PLÉNAT (2009) "Morphologie grammaticale et extragrammaticale." In : B. Fradin et al. (éds), p. 21-45.
- FRADIN, Bernard/Françoise KERLEROUX/Marc PLÉNAT (2009) *Aperçus de morphologie du français*. Saint-Denis : Presses Universitaires de Vincennes.
- GALISSON, Robert (1987) "Les dictionnaires de parodie comme moyens de perfectionnement en langue française." *Études de linguistique appliquée* 67, 57-118.
- GRÉSILLON, Almuth (1984) *La Règle et le monstre : le mot-valise. Interrogations sur la langue, à partir d'un corpus de Heinrich Heine*. Tübingen : Niemeyer.
- GRIES, Stefan Th. (2004) "Shouldn't it be breakfunch? A quantitative analysis of blend structure in English." *Linguistics* 42 (3), 639-667.

- HUDDLESTON, Rodney/Geoffrey PULLUM (éds) (2002) *The Cambridge Grammar of the English Language*. Cambridge : Cambridge University Press.
- KUBOZONO, Haruo (1990) "Phonological Constraints on Blending in English as a Case for Phonology-Morphology Interface." In G. Booij/J. Van Marle (éds), 1-20.
- LÉTURGIE, Arnaud (2010) "Une pratique lexicographique émergente : les dictionnaires détournés." In : A. Dykstra/T. Schoonheim (éds), 1340-1346.
- LÉTURGIE, Arnaud (2011) "À propos de l'amalgamation lexicale en français." *Langages* 183, 75-88.
- LÉTURGIE, Arnaud (à paraître) "Des dictionnaires fantaisistes d'amalgames lexicaux comme dictionnaires d'apprentissage." *Nouvelles perspectives en lexicographie*.
- LÓPEZ RÚA, Paula (2002) "On the Structure of Acronyms and Neighbouring Categories: A Prototype-Based Account." *English Language and Linguistics* 6 (1), 31-60.
- LÓPEZ RÚA, Paula (2004) "The Categorical Continuum of English Blends." *English Studies* 85 (1), 63-76.
- MERCÉ, Lorente et al. (éds) (2001) *La gramàtica i la semàntica en l'estudi de la variació*. Barcelona : Promociones y Publicaciones Universitarias.
- PAILLARD, Michel (2000) *Lexicologie contrastive anglais-français. Formation des mots et construction du sens*. Gap : Ophrys.
- PEREIRO, Myriam /Henry DANIELS (éds) (2006) *Le Désaccord*. Nancy : Publications de l'AMAES.
- PIÑEROS, Carlos-Eduardo (1999) "Word-Blending as a Case of Non-Concatenative Morphology in Spanish." *Rutgers Optimality Archive* 343-0999, <<http://roa.rutgers.edu>>, consulté le 13 septembre 2010.
- PIÑEROS, Carlos-Eduardo (2004) "The Creation of Portmanteaus in the Extragrammatical Morphology of Spanish." *Probus* 16 (2), 203-240.
- PLAG, Ingo (2003) *Word-Formation in English*. Cambridge : Cambridge University Press.
- RENNER, Vincent (2006) "Dépasser les désaccords : pour une approche prototypiste du concept d'amalgame lexical." In : M. Pereiro/H. Daniels (dirs), 137-147.
- SABLAYROLLES, Jean-François (1997) "Néologismes : une typologie des typologies." *Cahiers du C.I.E.L. 1996-1997*, 11-48.
- SABLAYROLLES, Jean-François (2000) *La néologie en français contemporain. Examen du concept et analyse de productions néologiques récentes*. Paris : Champion.
- TOURNIER, Jean (1985) *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris : Champion.

Dictionnaires d'amalgames lexicaux

- CHIFLET, Jean-Loup (1999) *Le Cafard laqué. Les mots-portemanteaux*. Paris : Mots & Cie.
- CHIFLET, Jean-Loup (éd.) (2002) *Mais que fait l'Académie ? Le dictionnaire des mots qui devraient exister*. Paris : Mots & Cie.
- CHIFLET, Jean-Loup/Nathalie KRISTY (1992) *Le Dictionnaire des mots qui n'existent pas*. Paris : Presses de la cité.
- CHIFLET, Jean-Loup/Nathalie KRISTY (2003) *Le Mokimanké : le dico des mots qui existent enfin !* Paris : Mots & Cie.
- CLÉMENT, Yves-Marie/Gérard GRÉVERAND (1993) *Pianissimots, petit dictionnaire des mots-valises*. Levallois Perret : Many.

- CRÉHANGE, Alain (2004) *Le Pornithorynque est un salopare. Dictionnaire de mots-valises*. Paris : Mille et une nuits.
- CRÉHANGE, Alain (2006) *L'Anarchiviste et le Biblioteckel. Dictionnaire de mots-valises*. Paris : Mille et une nuits.
- FINKIELKRAUT, Alain (2006 [1979]) *Petit dictionnaire illustré. Les mots qui manquent au dico*. Paris : Seuil.
- FOURNEL, Paul/Henri CUECO (2007) *Les Animaux d'amour et autres sardinosaures*, Bordeaux : Le Castor Astral.
- GALISSON, Robert/Louis PORCHER (1986) *Distractionnaire*. Paris : Clé international.
- GUILLAUME, Jean-Claude (2008) *La valise diplodocumatique*. Saint-Denis : Édilivre-Aparis.
- LE TELLIER, Hervé/Xavier GORCE (2007) *Les Opossums célèbres*. Bordeaux, Le Castor Astral.
- LOUBIÈRE, Sophie/Olivier LATYK (2003) *Éléphanfare. Le jeu des mots-valises pour les petits*. Paris : Albin Michel Jeunesse.
- MINDA, Alexandre (2004) *Mots-valises à poignées*. Paris : L'Harmattan.
- ROUBAUD, Jacques (2008) *Sardinosaures & Cie*. Bourges : Mille univers.
- SERGUINE, Jacques (1988) *Contradictionnaire*. Paris : Ramsay.
- THIBAUD, Jean-Jacques (2005) *Nouveaucabulaire*. Paris : Recherche midi.

Résumé

UN CAS D'EXTRAGRAMMATICALITÉ PARTICULIER : LES AMALGAMES LEXICAUX FANTAISISTES

L'amalgamation lexicale est un procédé d'innovation lexicale assez peu décrit en français et dont la morphologie ressort de l'extragrammatical. Partant du constat que les différentes classifications des amalgames lexicaux sont établies sur des corpus restreints, nous proposons de dresser un état des lieux de ces classifications en confrontant plusieurs modèles théoriques avec deux corpus complémentaires pour fournir une description morpho-phonologique de ces unités lexicales afin de faire émerger un prototype pour le français. Nous avons procédé, dans un premier temps, au recensement de 7 052 amalgames lexicaux tirés de 17 dictionnaires fantaisistes. Puis, un second corpus d'amalgames attestés en discours a été constitué afin d'être comparé au premier. À l'aide des données récoltées, il est possible de voir apparaître un schéma prototypique de l'amalgamation lexicale en français.

Povzetek

POSEBEN PRIMER IZVENSISTEMSKOSTI : IZMIŠLJENI LEKSIKALNI AMALGAMI

Leksikalno amalgamiranje je postopek izmišljanja besed, ki je v francoščini dokaj slabo raziskan in ki sodi na področje izvensistemske morfologije. Ugotavljamo, da so se različne tipologije leksikalnih amalgamov do sedaj opirale na korpuse manjšega obsega. V članku tako podamo pregled dosedanjih tipologij, več teoretskih modelov pa soočimo z gradivom dveh korpusov, s čimer pridemo do morfofonološkega opisa takšnih leksikalnih enot, kar nam omogoča izluščiti prototip francoskih leksikalnih amalgamov. Najprej smo sestavili korpus 7052 leksikalnih amalgamov, ki smo jih našli v 17 slovarjih izmišljenih besed. Nato smo zgradili korpus amalgamov, ki smo jih našli v dejanski rabi in ga primerjali s prvim. Na podlagi pridobljenih podatkov je moč določiti prototipsko shemo leksikalnega amalgamiranja v francoščini.