

# Prédire la structure des amalgames lexicaux du français ?

Léturgie, Arnaud

Université de Cergy-Pontoise  
Laboratoire LDI (CNRS UMR 7187)  
arnaud.leturgie@gmail.com

## 1 Introduction

L'objectif fixé dans cet article est de s'interroger sur la mise au point d'une méthode permettant de représenter au mieux la structure des amalgames lexicaux (ou mots-valises)<sup>1</sup> du français. Jusqu'à présent, aucune approche théorique ne permet de formuler une grammaire des amalgames lexicaux pour le français.

En effet, un rapide examen des études portant sur ce type d'unité lexicale en français permet de s'apercevoir que la description de leurs structures n'a presque pas été détaillée. Quelques chercheurs ont effectivement tenté de développer des typologies autour des amalgames lexicaux. Dès 1984, Grésillon offre une approche morpho-phonologique permettant de déterminer des patrons, reposant à la fois sur le type de troncation des lexèmes-sources et sur la présence ou l'absence d'un segment homophone commun aux lexèmes-sources. À l'aide de cette typologie, il est possible de classer des unités lexicales relevant de l'amalgamation en plusieurs groupes.

Cependant, le modèle de Grésillon (1984), au même titre que celui de Clas (1987) ou de Galisson (1987), présente des lacunes trop importantes pour en permettre une utilisation dans le cadre d'une approche générative. Autrement dit, ces modèles ne permettent pas de prédire la structure de l'output résultant de l'amalgamation de deux lexèmes-sources. Rien n'explique donc que l'amalgamation de *confiture* et de *compote* donne *confipote* plutôt que *compoture* ou *confitupote*. Il est donc nécessaire d'observer l'output préalablement, pour en déterminer ensuite la structure.

En dégagant des schémas phonologiques détaillés, Fradin (2000) apporte de premiers éléments de réponse afin d'établir une grammaire des amalgames lexicaux. Mais, précédemment, c'est Bat-El (1996) qui propose une grammaire des amalgames lexicaux hébreux. Comme nous le montrerons, cette approche, qui repose sur la théorie de l'optimalité, permet à Bat-El de déterminer quel output sera, en hébreu, privilégié sur les autres possibilités à l'aide de différentes contraintes. Est-il alors possible d'adapter cette grammaire au français ?

Afin de tenter d'apporter une réponse à cette question, les descriptions théoriques évoquées jusqu'à présent vont être rapidement commentées afin d'en cerner les apports et limites. Ceci permettra de dresser un état des lieux des travaux les plus significatifs sur l'amalgamation en français afin notamment d'élaborer l'identité du prototype d'amalgame lexical en français. Puis, une méthode d'adaptation de la grammaire de Bat-El (1996) au français sera présentée. Ceci permettra d'expérimenter une application de cette méthode sur quelques amalgames lexicaux posant divers problèmes.

## 2 Les différentes approches de l'amalgamation lexicale en français

Au cours des années 1980, plusieurs linguistes se sont essayés au développement de typologies des amalgames lexicaux. Ce besoin de classement est sans doute né du caractère insaisissable des amalgames. Il s'agit effectivement d'un type de néologisme assez particulier dans la mesure où il fait intervenir deux lexèmes, mais dont la segmentation ne respecte pas les règles de la morphologie<sup>2</sup>. Les lexèmes-sources ne subissent effectivement pas de troncation morphématique et la structure des outputs peut sembler

chaotique de prime abord. Afin donc de bénéficier d'un cadre théorique minimal, les premières typologies ont vu le jour.

## 2.1 Trois typologies formelles initiatrices

En quelques années, le français s'est doté de trois typologies formelles assez différentes pour un même phénomène. Les différences entre ces trois typologies permettent d'ailleurs d'étendre le champ de visée de chacune d'entre elles puisqu'elles se complètent les unes par rapport aux autres. En effet, alors que Grésillon (1984) propose une description morpho-phonologique des amalgames lexicaux, Clas (1987) développe l'aspect morphologique uniquement, tandis que Galisson (1987) s'appuie sur l'aspect graphique.

Ces trois approches reposent sur des données très fortement divergentes, ce qui exerce nécessairement une influence sur la façon de catégoriser les amalgames lexicaux. Ainsi, Grésillon (1984) prend pour point de départ une soixantaine d'amalgames allemands créés par le poète H. Heine et se dote d'un corpus secondaire trilingue – allemand, anglais et français – composé d'amalgames issus de sources très diverses telles que le journalisme, la littérature, la publicité, des répertoires de mots-valises, des créations de malades mentaux ou encore des amalgames créés par elle-même pour les besoins de son étude. Ceci diffère radicalement du corpus utilisé par Clas (1987) qui comporte des termes techniques provenant pour l'essentiel du domaine de la chimie, de l'électronique ou de la physique. Enfin, Galisson (1987) utilise comme seule source son propre dictionnaire d'amalgames lexicaux fantaisistes, le *Distractionnaire*. Cette hétérogénéité dans les données observées conduit donc nécessairement à des divergences théoriques parmi lesquelles des divergences de sélection d'appartenance à la catégorie. En effet, alors que Galisson (1987) présente *ministère* comme étant l'amalgame de *mini+stère+ministère*, Grésillon (1984) le considère comme un « mot-dévalisé »<sup>3</sup>, un type très particulier de structure à la limite de l'amalgamation. Clas (1987) n'est pas confronté à ce type de structure, apparemment inexistante dans les technolèctes qu'il analyse.

Les trois typologies formelles vont maintenant être brièvement détaillées et commentées.

### 2.1.1 Grésillon 1984

Grésillon (1984) entreprend la première étude d'importance sur les amalgames lexicaux en français. C'est d'ailleurs encore aujourd'hui l'étude fondamentale pour toutes les recherches sur le sujet, les bibliographies de ces recherches en témoignent. Elle détermine des critères définitoires permettant d'identifier et de catégoriser les amalgames lexicaux selon quatre schémas (p. 24-25) :

- avec segment homophone (en<fant>•fantaisie → *enfantaisie*)<sup>4</sup> ;
- avec troncation (infini•<in>verti → *infiniverti*) ;
- avec segment homophone et troncation (consomm<ation>•astuce → *consommastuce*) ;
- avec enchâssement (en<seigne>ment•saigne → *enseignement*)<sup>5</sup>.

Les schémas proposés peuvent bénéficier de davantage de détails, notamment sur la place de la troncation dans les lexèmes-sources. De par la nature extrêmement hétérogène du corpus utilisé, l'approche de Grésillon (1984) n'est que partiellement satisfaisante. En effet, le fait de traiter indifféremment des amalgames forgés spontanément, par des enfants ou des malades mentaux, et des amalgames forgés artificiellement dans le cadre de textes littéraires, journalistiques ou commerciaux (autant pour les noms de marques et de produits que pour les slogans publicitaires) ne semble pas judicieux. Il est préférable d'utiliser plusieurs corpus en parallèle et de comparer ensuite les données obtenues. Par ailleurs, le fait que le chercheur crée une partie du corpus lui-même est un risque supplémentaire de fausser les données. Les mécanismes présidant à la formation d'amalgames lexicaux commerciaux, par exemple, sont sans doute différents de ceux qui concernent les amalgames provenant de textes littéraires ou encore

d'amalgames forgés par un linguiste averti dans la mesure où l'aspect pragmatique du nom de marque ou de produit jouera un rôle crucial.

Le fait de traiter plusieurs langues simultanément est une réelle valeur ajoutée à l'analyse. Cependant, Grésillon (1984) ne remarque jamais les différences entre les trois langues. La morphologie de ces langues n'étant pas identique, il y a sûrement des particularités à noter. C'est pourquoi, étant adaptée tant au français qu'à l'allemand ou à l'anglais, la typologie de Grésillon (1984) peut rencontrer des limites quant à la description formelle détaillée des amalgames lexicaux de l'une de ces trois langues individuellement.

### 2.1.2 Clas 1987

Clas (1987) dresse 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'un lexème-source (robo<t>•<informa>tique → *robotique*) ;
- apocope d'un lexème-source (publi<ité>•postage → *publipostage*) ;
- aphérèse du lexème-source gauche et apocope du lexème-source droit (dom<icile>•<rob>otique → *domotique*) ;
- aphérèse des deux lexèmes-sources (<vi>nyl•<cot>on → *nylon*) ;
- apocope des deux lexèmes-sources (tel<eprinter>•ex<change> → *télex*) ;
- aphérèse ou apocope d'un lexème-source et syncope du deuxième lexème-source (u<ltra>•p<ast>e<u>risation → *upérisation*).

Bien qu'il soit morphologiquement plus détaillé que celui de Grésillon (1984), le modèle de Clas (1987) souffre d'un manque de considération 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. En outre, Clas n'opère pas de distinction entre ce qui relève de l'amalgamation lexicale et ce qui relève de l'affixation par fracto-morphèmes. Des lexèmes tels que *robotique* ou *télex* présentent en effet un segment de type affixal (les éléments de type suffixal *-tique* et préfixal *tél(é)-*) qui s'insèrent dans des paradigmes constructionnels réguliers où le segment conserve un sens identique (*informatique*, *domotique*, *téléphone*, *télévision*), ce qui n'est pas caractéristique de l'amalgamation mais plutôt de l'affixation<sup>6</sup>.

### 2.1.3 Galisson 1987

L'étude proposée par Galisson (1987) est assez atypique dans la mesure où sa démarche scientifique tend majoritairement à la promotion du dictionnaire d'amalgames lexicaux qu'il a publié l'année précédant son article. Le *Distractionnaire* (1986) est effectivement un dictionnaire de mots-valises fantaisistes destiné aux apprenants du français langue étrangère. Galisson souhaitait élaborer un outil ludo-didactique permettant aux apprenants d'enrichir leur vocabulaire tout en se familiarisant avec la consultation de dictionnaires classiques. Nous ne développerons pas cet aspect ici.

Le corpus sur lequel Galisson s'appuie est donc entièrement composé d'amalgames fantaisistes puisqu'il repose sur 3 répertoires fantaisistes dont son propre *Distractionnaire* (1986), ce qui représente environ un millier d'unités. Galisson s'attache à façonner une typologie essentiellement basée sur les modifications graphiques. L'observation des phénomènes morphologiques et phonologiques n'est pas entreprise. Formellement, il distingue les unités qui relèvent de l'insertion de celles qui relèvent de l'imbrication, ces deux matrices pouvant connaître diverses variations :

- insertion sans changement graphique (mini•stère•ministère → *ministère*) ;

- insertion avec changement graphique du mot originel (mélan<colie>•colis → *mélancolis*) ;
- insertion avec altération du mot originel (néo<phyte>•frite → *néofrite*) ;
- insertion avec substitution d'une partie du mot originel (mère•<per>fidie → *merfidie*) ;
- imbrication (rin<gard>•gargariser → *ringargariser*).

Ces schémas semblent être assez souples pour s'adapter à de nombreuses structures différentes. Cependant, la description du phénomène demeure assez faible. Galisson ne livre effectivement aucun critère d'identification des mots-valises alors qu'il assimile au phénomène des lexèmes – tels que *ministère* cité *supra* – qui ne relèvent pas à proprement parler de l'amalgamation lexicale. En réalité, c'est le schéma d'imbrication qui semble s'approcher le plus du phénomène d'amalgamation lexicale. De plus, il n'est fait aucune mention de l'aspect phonologique de l'amalgamation lexicale. L'approche de Galisson (1987 : 72) ne repose que sur le « dénombrement des lettres qu'affiche » chacun des lexèmes-sources au sein de l'amalgame. Galisson traite indifféremment des unités relevant de l'amalgamation lexicale et des unités telles que *ministère*, qui ne correspondent pas à l'amalgamation de deux lexèmes pour en former un troisième. Au final, Galisson est plus intéressé par les éventuelles applications didactiques de son *Distractionnaire* que par l'observation rigoureuse de l'amalgamation lexicale, preuve en est le corpus qu'il utilise – totalement exempt de mots-valises attestés – et le fait qu'il a « conduit [son] étude en ignorant les travaux des linguistes sur les mots-valises » (*op. cit.* : 61). La typologie résultante est donc moyennement satisfaisante et l'approche préconisée par Grésillon (1984) semble encore la plus indiquée. D'ailleurs, le travail de Galisson sur le sujet n'est presque jamais invoqué dans les recherches actuelles sur l'amalgamation lexicale.

## 2.2 Description phonologique (Fradin 2000)

Jusqu'ici, les trois typologies formelles examinées proposent uniquement de dresser des classements des différents amalgames lexicaux observés en corpus. On s'aperçoit que les analyses sont menées au niveau de l'output seulement et que rien ne permet d'expliquer ni la façon dont les tronctions ont lieu ni l'ordre dans lequel les lexèmes-sources se combinent.

Fradin propose une étude qui se démarque de ces typologies en recourant à des schémas phonologiques des amalgames lexicaux. L'objectif est de fournir un ensemble de schémas de combinaison des lexèmes-sources à partir d'un segment homophone commun. Cette approche permet de formuler 9 schémas différents, que Fradin (2000 : 28-30) regroupe en 4 familles. Fradin recense donc 9 types d'amalgames lexicaux différents, certains types pouvant connaître des variations, selon la place occupée par le segment homophone<sup>7</sup>. Ceci permet ainsi de dégager des schémas tels que les cinq suivants<sup>8</sup> :

- Schémas de la famille A : les  $\Sigma$  se situent soit à droite, soit à gauche des lexèmes-sources.
  - Type 1  $A\Sigma\Sigma B \rightarrow A\Sigma B$  (métro<pole>•politique → *métropolitique*) ;
  - Type 2  $A\Sigma B\Sigma \rightarrow AB\Sigma$  (potir<on>•marron → *potimarron*) ;
  - Type 3  $\Sigma A \Sigma B \rightarrow \Sigma AB$  (pas d'exemple).
- Schémas de la famille B : l'un des  $\Sigma$  se situe dans l'un des lexèmes-sources.
  - Type 4  $A\Sigma B\Sigma B' \rightarrow A\Sigma B'$  (hippie•<épi>démie → *hippidémie*) ;
  - Type 5  $A\Sigma A' \Sigma B \rightarrow A\Sigma B$  (céli<bataire>•battante → *célibattante*).

Comme on le constate, ce type de représentation de la structure des amalgames lexicaux a l'avantage d'être particulièrement complet. Ces schémas permettent de situer le segment homophone et d'indiquer le matériau phonologique tronqué. Les cas qui ne présentent pas de segment homophone sont également traités par Fradin (2000 : 30).

### 2.2.1 Une grammaire des amalgames lexicaux français ?

La différence fondamentale entre les typologies présentées au §2.1 et celle de Fradin (2000) réside dans le fait qu'à partir du schéma de combinaison des lexèmes-sources, il est possible de déduire la structure finale de l'amalgame lexical. Cependant, comme Fradin (2000 : 28) le précise, son approche n'a pas pour objectif de dégager une grammaire des amalgames lexicaux.

It must not be taken as an account of blending, that is as an explicitation of grammatical mechanisms capable of generating – or recognising – blends.

Autrement dit, le modèle de représentation qu'il propose ne doit pas être considéré comme une grammaire des amalgames lexicaux du français. Fradin (*ibid.*) précise d'ailleurs que les schémas définis dans son article ne sont que des outils descriptifs.

It should be stressed that these patterns are not rules of grammar, but mere descriptive devices. They say nothing about the phonological constraints which ultimately determine the combination of the base-lexemes. For example, they do not predict why starting from molecule 'molecule' and agriculteur 'farmer' we get moléculteur which follows pattern 4, instead of \*agrimoléculteur, which would follow pattern 6.

Cependant, il s'agit probablement de l'étude sur l'amalgamation lexicale en français qui s'approche le plus de la description détaillée des différentes structures possibles pour ces unités, permettant donc, en poursuivant l'analyse, d'élaborer une grammaire. En réunissant suffisamment de données, il est possible de dégager une échelle de typicité permettant de déterminer un prototype d'amalgame lexical pour le français. Partant de ce prototype et en le comparant à des unités plus ou moins typiques, il devient possible de formuler des contraintes relatives, par exemple, à l'ordre d'amalgamation des lexèmes-sources. Au final, des règles de grammaire peuvent être définies. Fradin (2000 : 30) présente la famille A comme étant la plus courante.

It should be noted that all these patterns are not equivalent. Pattern family A, including the variants, is by far the most frequent.

Cette affirmation est corroborée par Léturgie (2011 : 100) qui propose les caractéristiques suivantes pour définir la structure prototypique d'amalgame lexical du français :

- i. l'unité est formée par amalgamation de deux lexèmes ;
- ii. le lexème-source de gauche subit une apocope et le lexème-source de droite reste intact ;
- iii. un segment homophone est présent au point de jonction des bases-sources.

En confrontant des unités typiques à des amalgames moins typiques, il serait potentiellement possible d'observer des régularités et d'en déduire des règles de grammaire.

### 2.2.2 Quelques limites

La typologie de Fradin (2000) connaît néanmoins des limites. Ainsi, selon Fradin (2000 : 28-29), il n'existe pas en français d'amalgames formés sur le schéma  $\Sigma A \Sigma B \rightarrow \Sigma AB$ . Fradin & al. (2009 : 38) précise qu'il s'agit de l'amalgamation de lexèmes dissyllabiques possédant une syllabe commune qui est impossible.

On note néanmoins que le mot-valisage de deux dissyllabiques possédant une syllabe en commun paraît extrêmement difficile : *copain* + *cobaye* ne semble pouvoir donner ni *copainbaye*, ni *cobayepain*. Tout se passe comme s'il fallait deux syllabes consécutives pour qu'un lexème soit identifié. Si tel était le cas, l'impossibilité de \**Robespierrecard* ou de \**banchérubin*, où les deux syllabes de *Rocard* et de *bambin* sont dissociées, s'expliquerait sans mal.

Il s'agit d'un cas rare mais pas inexistant. Des amalgames lexicaux tels que *doudouleur* (doudou, douleur), *clocloporte* (Cloclo, cloporte), *mémélasse* (mémé, mélasse) ou *tataverne* (tata, taverne) sont effectivement possibles. Cependant il importe de préciser que dans tous ces cas, l'amalgamation devient

possible parce que le premier lexème-source comporte une reduplication. Dans ce cas, l'affirmation de Fradin & al. (*ibid.*) tient, puisque chacun des lexèmes-sources dissyllabiques retrouve ses deux syllabes dans l'amalgame. Autrement dit, l'identification des lexèmes-sources se ferait par la consécutive de deux syllabes au moins. C'est ce qui expliquerait donc l'impossibilité de *\*Robespierrecard* ou de *\*banchérubin* selon Fradin & al. Comment expliquer alors *guilimaue* (guili, guimauve), qui, *a priori*, ne semble pas poser de problème d'interprétation et d'identification des lexèmes-sources<sup>9</sup> ? *Imbibécile* (imbibé, imbécile), bien que résultant de l'amalgamation de lexèmes-sources trisyllabiques adopte une structure relativement similaire. Néanmoins, les deux lexèmes-sources sont ici identifiables selon les critères de Fradin & al. (2009) puisque deux syllabes se suivent consécutivement pour chacun des termes.

Jusqu'ici, les exemples cités sont tirés d'un corpus d'amalgames lexicaux fantaisistes<sup>10</sup>. Cependant, l'explication avancée par Fradin & al. (2009) ne permet pas non plus d'expliquer la structure de *bedondaine* (bedon, bedaine) qui est tiré du *Grand Robert 2009*.

Par ailleurs, rien n'indique que les schémas proposés par Fradin (2000) sont exhaustifs. Il ne prétend d'ailleurs aucunement à l'exhaustivité. Cela dit, il paraît difficile d'imaginer des structures supplémentaires sans produire des amalgames peu acceptables. En réalité, les limites rencontrées par Fradin (2000) sont liées au corpus utilisé. Il prend effectivement appui sur le corpus de Grésillon (1984) et sur des données collectées par le laboratoire LLI (aujourd'hui LDI). Le problème vient donc de la nature du corpus, très hétérogène comme on l'a déjà signalé, ainsi que de la quantité de données réunies, trop peu importante pour permettre l'appréhension des structures des amalgames lexicaux dans leur diversité et dans leur récursivité. Les possibilités d'amalgamation étant très nombreuses, il est indispensable de se doter d'un corpus aussi riche que possible, afin d'examiner un maximum de cas différents. Nous reviendrons sur ce problème au §3 et allons à présent observer l'approche suggérée par Bat-El (1996).

### 2.3 Une grammaire des amalgames lexicaux hébreux (Bat-El 1996)

Toutes les typologies observées jusqu'à présent s'ancrent du côté de la morphologie extragrammaticale. Bat-El (1996) tente cependant d'élaborer une grammaire des amalgames lexicaux de l'hébreu – où les mots-valises sont assez nombreux dans le langage courant par rapport au français par exemple – afin de montrer que si ces unités sont considérées comme relevant de l'extragrammatical, elles ne doivent pas pour autant être négligées. Bat-El (1996 : 317) conclut d'ailleurs en replaçant l'amalgamation lexicale dans une position moins marginale au sein de la morphologie<sup>11</sup>.

It has been shown that Optimality Theory, which engages ranked and violable constraints, can account for blending, and thus promote it from its extragrammatical status to a more central position in linguistic research.

C'est effectivement en utilisant la théorie de l'optimalité et en adaptant certaines de ses contraintes à l'amalgamation en hébreu que Bat-El (1996) parvient à proposer une grammaire de ces unités lexicales. L'approche qu'elle propose, reposant uniquement sur la phonologie, est séduisante et semble fonctionner de façon assez régulière pour constituer une véritable grammaire. L'avantage que propose ce modèle est indéniablement la possibilité de déterminer l'output à partir des lexèmes-sources. Le fait de posséder une grammaire des amalgames lexicaux permet donc d'adopter une approche générativiste du phénomène, ce qui n'était jusqu'alors pas possible. À l'aide des contraintes définies dans le cadre de la théorie de l'optimalité, des règles d'amalgamation émergent et permettent au chercheur d'expliquer l'output résultant de l'amalgamation des lexèmes-sources composant l'input.

Cependant, l'analyse de Bat-El (1996) rencontre certaines limites. En effet, la chercheuse ne précise pas la nature du corpus utilisé pour produire son modèle d'analyse. On ne sait rien des exemples sélectionnés pour illustrer les différentes contraintes ni d'éventuels items qui résisteraient aux règles en présentant des irrégularités. Il sera donné l'occasion de voir au §4.2.2 que les amalgames du français contreviennent parfois aux contraintes retenues par Bat-El (1996). Dans la partie finale de son article, Bat-El (1996) avance que ce modèle est adaptable à toutes les langues, ce qu'elle montre partiellement avec l'anglais.

Nous supposons néanmoins que certaines contraintes ne vont pas s'adapter sans mal à certaines langues, comme le français par exemple. Les différentes contraintes proposées par Bat-El (1996) vont être présentées brièvement afin d'en percevoir la compatibilité avec le français.

### 2.3.1 Contraintes non dominées

Bat-El (1996) établit une hiérarchie des contraintes utilisées afin de sélectionner le meilleur candidat à l'amalgamation de deux lexèmes. Ainsi, les contraintes non dominées, au nombre de quatre, sont hiérarchiquement supérieures aux contraintes dominées. Les contraintes non dominées ne sont pas hiérarchisées, elles priment simplement sur les contraintes dominées, qui elles connaissent une hiérarchisation.

Les contraintes non dominées sont les premières à satisfaire pour qu'un output possible soit sélectionné comme étant le meilleur candidat. La première de ces contraintes est celle de linéarité (LINEARITY)<sup>12</sup>.

LINEARITY : un amalgame lexical possède une structure compatible avec celles de ses lexèmes-sources et vice versa.

Ainsi, la structure de l'amalgame doit respecter celles de ses lexèmes-sources : par exemple, *sondôlatrie* (sondage, idolâtrie) vs *\*sondôtriela*. Cette contrainte exclut par ailleurs la métathèse des procédés d'amalgamation, alors que de nombreux amalgames lexicaux sont construits à l'aide de ce procédé. *Manbidule* (mandibule, bidule), *métaformose* (métamorphose, Formose) ou *agfanistan* (Agfa, Afghanistan), sont par exemple obtenus par métathèse. Il semble qu'il y ait néanmoins amalgamation de deux lexèmes avec une troncation d'un des termes de l'input. Pourtant, selon la contrainte de linéarité, *\*mandidule*, *\*métamormose* ou *\*aghanistan* seraient de meilleurs candidats.

La seconde contrainte non dominée est celle de l'alignement des bords (ALIGNMENT OF EDGES).

ALIGNMENT OF EDGES : le premier segment du premier lexème-source et le dernier segment du second lexème-source sont préservés.

Cette contrainte implique que l'amalgame comporte la première syllabe du premier lexème-source et la dernière du second. *\*Vinson* serait donc mieux formé que *nylon* (vinyl, nylon), et tous les cas de bi-apocope ou de bi-aphérèse ne seraient pas les meilleurs candidats possibles dans la mesure où une des deux syllabes mentionnées dans la contrainte n'apparaît pas dans l'output. Ceci dit, de nombreux linguistes, cités par Renner (2006) notamment, considèrent que ce schéma de troncation, proposé par Clas (1987) par exemple, ne produit pas des unités relevant de l'amalgamation lexicale.

La troisième contrainte non-dominée est celle de contribution minimale (MINCONTRIB).

MINCONTRIB : chaque lexème-source doit apporter au moins une syllabe à l'amalgame lexical.

Afin d'assurer que les amalgames dissyllabiques soient formés d'au moins une syllabe de chaque lexème-source, la contrainte de contribution minimale pénalise les outputs tels que *algol* (algorithmic, language) qui ne comportent qu'une consonne de l'un des lexèmes-sources, au profit de *tigron* (tigre, lion) par exemple.

La quatrième et dernière contrainte non dominée est celle de maximalité syllabique ( $\partial$ MAX)

$\partial$ MAX : chaque syllabe dans les deux lexèmes-sources doit avoir un correspondant dans l'amalgame lexical.

Cette contrainte favorise les candidats dont le nombre de syllabes n'est pas inférieur au nombre de syllabes du lexème-source le plus long. Par exemple, *délicaresse* (délicatesse, caresse) est optimal face à *\*décaresse* ou *\*catesse* puisqu'il comporte autant de syllabes que le lexème-source le plus long (quatre syllabes pour /de.li.ka.tes/).

Lorsque ces contraintes ne permettent pas de sélectionner le candidat optimal entre plusieurs outputs possibles – ce qui est fréquent –, d'autres contraintes s'ajoutent. Ces contraintes sont hiérarchiquement inférieures et connaissent une hiérarchie entre elles.

### 2.3.2 Contraintes dominées

Tout d'abord, la contrainte de segment identique (DESIGNATED IDENTICAL SEGMENT CONSTRAINT ou DISC) concerne l'homophonie entre les lexèmes-sources.

DISC : un amalgame lexical doit avoir au moins une consonne qui a un correspondant dans l'un des deux lexèmes-sources.

Bat-El (1996) fait reposer cette contrainte sur une correspondance consonantique parce que, comme elle le précise (*ibid.* : 292), le segment partagé par les lexèmes-sources est majoritairement une consonne en hébreu.

However, as most studies on blending have observed, a phonological consideration is often involved, since in many blends the two base elements share one or more identical segments. It seems that in Hebrew the shared segment must be a consonant (though this might be a preference rather than a must).

En français, cette contrainte fonctionne très mal. Des amalgames lexicaux tels que *bosphormidable* (Bosphore, formidable), *rurbain* (rural, urbain), *abribuste* (abrisbus, buste) ou *grascier* (gracier, scier) présentent bien un segment homophone commun aux lexèmes-sources. Mais ce segment peut être de nature vocalique, voire syllabique. Cette contrainte demande donc à être adaptée pour le français de manière à prendre en compte un segment homophone sans préciser la nature phonétique de ce segment.

La contrainte de dépendance syllabique ( $\partial$ DEP) est dominée par la contrainte de segment identique (DISC).

$\partial$ DEP : toutes les syllabes de l'amalgame lexical doivent avoir un correspondant dans les lexèmes-sources.

Cette contrainte entre en relation avec la contrainte non dominée  $\partial$ MAX. Alors que la contrainte  $\partial$ MAX détermine la longueur syllabique minimale de l'amalgame lexical, la contrainte  $\partial$ DEP détermine la longueur syllabique maximale de l'amalgame. Ainsi, *\*gangstérisme* est privilégié par rapport à *gangsterrorisme* (gangster, terrorisme) parce qu'il ne possède pas plus de syllabes que le lexème-source le plus long (trois syllabes pour /te.ʁo.ʁism/). Pourtant, c'est bien le second candidat qui est attesté en français. On observe le même phénomène entre *numériclature* (numérique, nomenclature) et *\*numenclature*. Le second candidat est optimal selon la contrainte  $\partial$ DEP parce qu'il possède autant de syllabes que le lexème-source le plus long, tandis que l'amalgame attesté dans le *Grand Robert 2009* compte une syllabe de plus que /no.mã.kla.tyʁ/.

Ces deux contraintes dominées peuvent être complétées par des contraintes résiduelles (qui ne sont pas hiérarchisées) permettant de traiter tous les amalgames lexicaux. La contrainte de maximalité de segment (SEGMAX) est donc hiérarchiquement inférieure aux trois premières contraintes.

SEGMAX : tous les segments dans les lexèmes-sources ont un correspondant dans l'amalgame lexical.

Cette contrainte permet, lorsque les autres contraintes sont également respectées ou violées, de sélectionner le meilleur candidat selon le nombre de phonèmes tronqués. Ainsi, entre *sarkolâtre* (Sarkozy, idolâtre) et *\*sarkozidolâtre*, le meilleur candidat se révélerait être *\*sarkozidolâtre* dans la mesure où un phonème seulement est tronqué (le phonème [i], qui, de plus, n'est pas comptabilisé puisqu'il est commun aux deux lexèmes-sources) contre cinq pour *sarkolâtre*.

Enfin, la contrainte de contact syllabique ( $\partial$ CONT) permet de sélectionner le candidat optimal lorsque les contraintes précédentes sont insuffisantes.



$\partial$ CONT : l'attaque d'une syllabe ne doit pas être d'une sonorité supérieure au dernier segment de la syllabe qui la précède directement.

La contrainte  $\partial$ CONT, à l'aide d'une échelle mesurant l'indice de sonorité, permet donc de déterminer que le candidat optimal est celui qui obtient l'indice le plus bas. Bat-El (1996 : 303) propose l'échelle suivante pour mesurer l'écart de sonorité.

Voyelle (5) > glide (4) > liquide (3) > nasale (2) > fricative (1) > occlusive (0)

Les chiffres indiquent le degré de sonorité et permettent, à l'aide d'une équation, de mesurer le contact syllabique résultant de l'amalgamation de deux lexèmes. Ainsi, /saʁ.ko.<zi•i.do>.latʁ/ présente un contact entre les syllabes [ko] et [latʁ]. Ceci se traduit par la formule 5-(5-3), puisque Bat-El (*ibid.*) explique que le degré de sonorité est obtenu en soustrayant l'indice du segment final de la syllabe précédente (ici la voyelle /o/ du segment /ko/) à l'indice de l'attaque de la syllabe au point de contact (ici la liquide [l] du segment [latʁ]). Le résultat est ensuite soustrait à l'indice maximal, autrement dit la sonorité vocalique : 5. *Sarkolâtre* obtient donc un indice de 3, ce qui signifie que la contrainte est violée trois fois. \**Sarkozidolâtre* (/saʁ.ko.zi•<i>.do.latʁ/) obtient lui un indice de 0 (le contact se produisant entre les phonèmes /i.d/, soit 5-(5-0)). Cet indice correspond au meilleur contact syllabique possible, c'est-à-dire un contact voyelle-occlusive qui permet de ne pas violer la contrainte du tout.

À travers les exemples utilisés pour illustrer ces différentes contraintes, nous avons souligné à plusieurs reprises que les candidats optimaux n'étaient pas toujours les amalgames attestés dans les données que nous avons constituées. Nous allons désormais présenter ces données pour ensuite appliquer la grammaire proposée par Bat-El (1996) à un petit ensemble d'amalgames lexicaux de structures différentes.

### 3 Une application au français ?

Afin d'adapter cette grammaire au français, il faut tout d'abord vérifier si les contraintes proposées par Bat-El (1996) s'appliquent à la morphologie du français. Nous allons donc confronter les contraintes introduites précédemment à un petit corpus d'amalgames lexicaux. Ce corpus est constitué à l'aide de deux ressources qui vont être présentées brièvement.

Nous avons souligné à plusieurs reprises la difficulté éprouvée par les chercheurs à produire des modèles de description formelle des amalgames lexicaux basés sur des corpus soit restreints soit hétérogènes. Nous proposons donc l'utilisation parallèle de deux corpus complémentaires. Le premier, COAL – pour COrpus d'Amalgames Lexicaux – 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 de projets lexicographiques. Le second corpus, CALA – pour Corpus d'Amalgames Lexicaux Attestés –, 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.

Les amalgames de CALA, 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.

#### 3.1 COAL

COAL est un corpus construit par l'extraction de 7039 amalgames lexicaux de 17 dictionnaires fantaisistes dont la liste est donnée en bibliographie de cet article. Les lexèmes-sources de cette base de données ont été identifiés et chaque amalgame est catégorisé selon son schéma de construction en suivant les typologies de Grésillon (1984), de Clas (1987) et de Fradin (2000), puis étiqueté en fonction des contraintes de prototypie proposées par Renner (2006). Ceci permet d'obtenir un nombre suffisant d'informations formelles sur les structures de ces amalgames lexicaux.

### 3.2 CALA

En regard de COAL, il importe de constituer un corpus de mots forgés *in vivo*. Ce corpus, CALA, 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.

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.

## 4 Résultats de l'analyse

À l'aide de ces deux corpus, nous avons cherché à savoir si le modèle théorique proposé par Bat-El pouvait s'appliquer au français et dans quelle mesure. Pour cela, des amalgames lexicaux issus des deux corpus ont été sélectionnés et les différentes contraintes évoquées au §2.3 leur ont été appliquées. La sélection des termes soumis aux contraintes résulte de plusieurs paramètres.

Tout d'abord, il fallait tester des amalgames lexicaux plus ou moins proches du modèle prototypique présenté au §2.2.1. À partir des observations formulées par Fradin (2000) et Léturgie (2011) présentées *supra*, quatre amalgames typiques et quatre autres non typiques ont été sélectionnés. Parmi ces deux catégories, deux termes proviennent de COAL, deux autres de CALA. Cette répartition assure tout d'abord une possibilité de sélection des termes en fonction de leur proximité au prototype. Le fait de posséder un corpus tel que COAL, quantitativement important, permet de disposer d'un grand nombre d'amalgames de structures variées. Parallèlement, l'utilisation de CALA nous autorise à formuler des remarques sur l'amalgamation en français de façon générale, sans s'en tenir à des termes fantaisistes.

Les tableaux d'analyse figurant dans les sections suivantes sont établis à partir de ceux présentés par Bat-El (1996). Ils ont été adaptés afin de réunir le maximum d'informations sur les différentes contraintes testées. Le système de notation de violation des contraintes est celui de la théorie de l'optimalité. L'astérisque indique une violation et le point d'exclamation signale une violation fatale, c'est-à-dire une violation qui exclut le candidat de la sélection. La flèche précédant l'amalgame indique qu'il s'agit du candidat optimal. Les contraintes sont présentées dans l'ordre hiérarchique. Toutes ne sont pas renseignées dans la mesure où certaines sont systématiquement respectées et qu'il n'y a donc pas de commentaire à formuler. C'est le cas des contraintes LINEARITY, ALIGNMENT OF EDGES ou MINCONTRIB.

Afin de proposer une analyse la plus exhaustive possible, nous avons adapté le système de mesure de l'optimalité. Ainsi, alors que l'analyse des contraintes violées ou respectées par un candidat doit s'arrêter si une violation fatale apparaît, nous avons continué l'analyse pour vérifier les différentes transgressions possibles par rapport aux autres candidats. Les cellules grisées indiquent donc que les contraintes n'ont pas besoin d'être testées puisqu'une violation fatale les précède. Cependant, nous avons souhaité les faire apparaître afin de livrer le maximum d'observations possible.

Les tableaux comportent la transcription phonétique des lexèmes. Les deux phonèmes directement à la suite des amalgames représentent les sons mesurés pour la contrainte  $\partial$ CONT. Le premier candidat de chaque tableau correspond à l'amalgame original qui a été extrait des corpus. Les lignes suivantes présentent des candidats possibles, sans prétendre à l'exhaustivité. Pour des raisons de commodité, nous avons conservé la désignation originale des contraintes.

### 4.1 Amalgames lexicaux typiques

Sur l'ensemble des amalgames lexicaux typiques testés, tous les candidats optimaux se révèlent être les amalgames tirés du corpus. Ceci tend à accréditer la grammaire proposée par Bat-El (1996), qui

fonctionnerait sans peine pour le français. On observe malgré tout l'apparition de tendances concernant les troncations.

#### 4.1.1 Amalgames fantaisistes

Parmi les amalgames lexicaux fantaisistes, on s'aperçoit que les deux exemples testés fonctionnent très bien avec les contraintes de la grammaire de Bat-El (1996).

/a.sã.sœʁ, sã.syʁ/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ a.sã.<sœʁ•sã>.syʁ ã.s				5-(5-1) *	***!
a.<sã.sœʁ>•sã.syʁ a.s				5-(5-1) *	***!
a.sã.sœʁ•<sã>.syʁ ʁ.s		*!	*!	5-(3-1) **!*	**
sã.syʁ•<a.sã>.sœʁ ʁ.s		*!		5-(3-1) **!*	***!
sã.<syʁ>•a.sã.sœʁ ã.a		*!	*!	5-(5-5) **!*	***!

Tableau 1 : Structures possibles de l'amalgamation des lexèmes *ascenseur* et *censure*

L'output le plus probable pour l'amalgamation d'*ascenseur* et de *censure* semble être *ascensure*. C'est effectivement l'amalgame issu de COAL qui respecte le mieux les contraintes. Les trois candidats opposés à *ascensure* violent la contrainte DISC puisqu'aucun phonème n'est partagé au point de contact des lexèmes-sources. Or, l'homophonie est l'un des traits prototypiques des amalgames lexicaux en français, ce qui tend à éliminer la possibilité de former ces trois candidats plutôt que *ascensure*. En ce qui concerne la troncation phonétique des lexèmes-sources, on s'aperçoit que les candidats des deux premières lignes respectent également les cinq contraintes. On peut donc imaginer que dans ce cas de figure, l'orthographe jouera un rôle dans l'identification du matériau tronqué. L'amalgame issu du corpus est orthographié *ascensure*, permettant une identification du lexème-source *ascenseur* plus efficace que ne le permettrait la graphie *acensure* ou /a-/ semblerait fonctionner comme préfixe privatif.

Dans le cas de *délicatesse*, on s'aperçoit que c'est la contrainte de ∂MAX qui est violée par les deux candidats non optimaux.

/de.li.ka.tɛs, ka.ʁɛs/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ de.li.<ka.tɛs>•ka.ʁɛs i.k				5-(5-0)	***
de.li.ka.<tɛs•ka>.ʁɛs a.ʁ				5-(5-3) *! **	***
de.<li.ka.tɛs>•ka.ʁɛs e.k	*!	*!		5-(5-0)	****!***
ka.<ʁɛs•de.li.ka>.tɛs a.t	*!*	*!		5-(5-0)	****!***

Tableau 2 : Structures possibles de l'amalgamation des lexèmes *délicatesse* et *caresse*

À nouveau, c'est l'amalgame issu de COAL qui est le candidat optimal. Il respecte la contrainte ∂MAX ainsi que la contrainte DISC. Le nombre de syllabes permet de mieux identifier les lexèmes-sources et semble aider à l'interprétation de l'amalgame. Rares sont les amalgames qui violent cette contrainte et un rapide examen des données que nous possédons ne nous a pas permis d'illustrer la violation de la contrainte ∂MAX avec un amalgame tiré des corpus. En ce qui concerne la troncation phonétique des lexèmes-sources, on observe cette fois-ci une différence sur le respect de la contrainte ∂CONT. Alors que la première segmentation proposée à la première ligne ne transgresse pas du tout la contrainte ∂CONT, la troncation testée à la seconde ligne du tableau ne correspond pas au meilleur contact syllabique possible et ∂CONT est violée trois fois. Ceci est intéressant dans la mesure où cette observation conforte le trait

typique qui demande qu'un amalgame résulte de la troncation du lexème-source de gauche alors que celui de droite reste intact, ce qu'il est possible de constater dans les deux tableaux suivants.

#### 4.1.2 Amalgames attestés

En ce qui concerne les amalgames attestés, tous les candidats sélectionnés par application des contraintes correspondent aux termes issus du corpus. *Omnipräsident* est préféré aux deux autres possibilités qui violent la contrainte DISC.

/ɔm.ni.pʁe.zã, pʁe.zi.dã/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ ɔm.ni.<pʁe.zã>•pʁe.zi.dã i.p			*!	5-(5-0)	*
ɔm.ni.pʁe.<zã•pʁe>.zi.dã e.z			*!	5-(5-1)*	***!
ɔm.ni.pʁe.<zã•pʁe.zi>.dã e.d		*!		5-(5-0)	***!**
pʁe.<zi.dã>•ɔm.ni.pʁe.zã e.ɔ		*!	*!	5-(5-5) *!****	***!**

Tableau 3 : Structures possibles de l'amalgamation des lexèmes *omnipräsident* et *président*

Comme précisé au §2.3.2, cette contrainte est particulièrement importante en français puisqu'elle suppose l'existence d'un segment homophone entre les lexèmes-sources. Ce segment homophone est tenu comme une condition de la structure d'un amalgame lexical par certains linguistes tels que Grésillon (1984 : 15) ou Piñeros (1999 et 2004). Notons cependant que la contrainte ∂DEP est violée par *omnipräsident* puisque l'output comporte une syllabe de plus que son lexème-source le plus long. Enfin, la segmentation phonétique optimale correspond à une apocope du lexème-source gauche.

Le cas de *préfectorlupitude* est sensiblement similaire à *omnipräsident*. Le candidat optimal correspond à nouveau à l'amalgame tiré de CALA.

/pʁe.fek.tyʁ, tyʁ.ly.pi.tyd/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ pʁe.fek.<tyʁ>•tyʁ.ly.pi.tyd k.t			***!	5-(0-0)*!****	
pʁe.fek.tyʁ•<tyʁ>.ly.pi.tyd ʁ.l			***!	5-(3-3)*!****	
pʁe.<fek.tyʁ>•tyʁ.ly.pi.tyd e.t		*!		5-(5-0)	*!**
tyʁ.ly.<pi.tyd•pʁe>.fek.tyʁ y.f		*!		5-(5-1) *!	*!*****
pʁe.fe<k.tyʁ•tyʁ.ly>pi.tyd e.p		*!		5-(5-0)	*!*****

Tableau 4 : Structures possibles de l'amalgamation des lexèmes *préfecture* et *turlupitude*

On remarque encore une fois que c'est la contrainte DISC qui est violée par les autres candidats. La seule violation de *préfectorlupitude* se situe au niveau de la contrainte ∂DEP.

#### 4.2 Amalgames lexicaux non typiques

Le cas des amalgames lexicaux non typiques (cf. §2.2.1) est plus problématique. Seuls deux amalgames issus des corpus respectent les contraintes de la grammaire. Ce constat ouvre des interrogations sur la définition de l'amalgamation lexicale, comme nous l'abordons en conclusion.

#### 4.2.1 Amalgames fantaisistes

Les deux amalgames fantaisistes non typiques testés respectent la grammaire de Bat-El (1996). *Chatil*<sup>13</sup> et *gribou*, deux termes issus de COAL, sont les deux candidats sélectionnés par application des contraintes de typicité.

/fat, fə.nil/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ fat<fə.n>il a.t		*		5-(5-0)	***!
fə<fə.>nil t.n		*		5-(0-2)*!*****	**
fə<t.fə>.nil a.n		*		5-(5-2) *!*	***!
fə.<nil>.fat ə.f		*		5-(5-1) *!	***!

Tableau 5 : Structures possibles de l'amalgamation des lexèmes *chat(te)* et *chenil*

/grip.su, ibu/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
→ gri<p.su.i>.bu i.b		*		5-(5-0)	***
gr<ip.su.>.bu i.b		*		5-(5-0)	****!
i.bu.<gri>.p.su p.s		*	*!	5-(0-1) *!*****	***
grip.<su.i>.bu p.b		*		5-(0-0) *!****	***

Tableau 6 : Structures possibles de l'amalgamation des lexèmes *grippe-sou* et *hibou*

La contrainte DISC est violée par tous les candidats, mais les contraintes ∂DEP et ∂CONT sont respectées par les deux amalgames fantaisistes, tandis que les autres candidats les violent. Si la contrainte de segment identique (DISC) est violée, c'est que les lexèmes-sources de ces amalgames ne partagent pas de segment homophone, ce qui en fait des amalgames non typiques.

#### 4.2.2 Amalgames attestés

Les amalgames attestés non typiques vont en revanche poser problème à la grammaire de Bat-El (1996). Dans les deux exemples testés, le candidat optimal se révèle être différent de l'amalgame tiré du corpus.

Ainsi, *bedondaine* viole la contrainte DISC quand *bedon* ou *bedaine* la respectent. Le problème est que ces deux candidats, qui sont les seuls à respecter la contrainte DISC, ne peuvent être interprétés comme des amalgames lexicaux puisqu'ils sont équivalents à l'un de leurs lexèmes-sources. Ils sont par ailleurs les seuls à ne pas violer la contrainte ∂DEP en ne dépassant pas deux syllabes.

/bœ.dɔ̃, bœ.den/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
bœ.dɔ̃•<bœ>.den	ɔ̃.d	*!	*!	5-(5-0)	**
bœ.den•<bœ>.dɔ̃	n.d	*!	*!	5-(2-0) *!	**
bœ.<den•bœ>.dɔ̃	œ.d			5-(5-0)	***! **
→ bœ.<dɔ̃•bœ>.den	œ.d			5-(5-0)	***! *
bœ.<den>•bœ.dɔ̃	œ.b	*!	*!	5-(5-0)	***!
bœ.<dɔ̃>•bœ.den	œ.b	*!	*!	5-(5-0)	**

Tableau 7 : Structures possibles de l'amalgame des lexèmes *bedon* et *bedaine*

Cet exemple n'est pas anodin puisqu'il rappelle le schéma non productif souligné par Fradin (2000) : Type 3  $\Sigma A \Sigma B \rightarrow \Sigma AB$ . Comme précisé au §2.2.2, l'amalgame de deux lexèmes dissyllabiques à initiales communes n'est pas impossible, mais il s'agit probablement d'un cas limite que la grammaire de Bat-El (1996) n'admet pas. Pourtant, il semble impossible d'accepter *bedon* ou *bedaine* comme des amalgames lexicaux étant donné qu'ils existent déjà en tant que lexèmes simples. Il faudrait donc amender la grammaire des amalgames pour l'adapter à ce type de structures en précisant une règle concernant les amalgames de deux lexèmes dissyllabiques à initiales communes.

Le cas de *confipote* est, lui aussi, singulier. L'output optimal d'après les contraintes est *compoture*. En effet, les deux autres candidats proposés, dont *confipote*, violent la contrainte DISC.

/kɔ̃.fi.tyɔ̃, kɔ̃.pɔt/	∂MAX	DISC	∂DEP	∂CONT	SEGMAX
kɔ̃.fi.<tyɔ̃•kɔ̃>.pɔt	i.p	*!		5-(5-0)	****! *
→ kɔ̃.pɔ<t•kɔ̃>.fi.tyɔ̃	ɔ̃.t			5-(5-0)	****! *
kɔ̃.fi.ty<ɔ̃•kɔ̃>.pɔt	y.p	*!		5-(5-0)	***

Tableau 8 : Structures possibles de l'amalgame des lexèmes *confiture* et *compote*

Pourtant, il semble que le segment [pɔt] soit moins proéminent dans la structure /kɔ̃pɔtyɔ̃/ que dans la structure /kɔ̃fipɔt/. Ceci tient à la place de ce segment qui, en position finale dans l'amalgame tiré du corpus, s'interprète comme le lexème simple /pɔt/. Ce segment détient donc une place sémantique importante dans l'interprétation de l'amalgame *confipote* qui réunit les signifiants de *confiture*, *compote* et *pote*. La *confipote* serait interprétable en un mélange de confiture et de compote, à partager avec ses potes, tandis que *compoture* ne bénéficie pas de cette interprétation possible. L'aspect sémantique, que nous n'avons pas évoqué jusqu'à présent pour nous concentrer sur la morphologie des amalgames, occupe un rôle crucial, d'autant plus que *confipote* est un nom de produit, ce qui implique des considérations pragmatiques de première importance<sup>14</sup>. En éludant totalement ce rôle, la grammaire de Bat-El (1996) risque de poser des problèmes quant à la sélection d'un candidat plutôt que d'un autre. Malgré l'absence de segment homophone, *confipote* n'est pas mal formé. Il présente simplement un caractère moins typique que le candidat qui respecte la contrainte de segment identique.

## 5 Conclusion

Comparativement aux premières typologies, les approches proposées par Fradin (2000) et par Bat-El (1996) offrent des avantages indéniables. Cependant, elles rencontrent également des limites.

Si Bat-El (1996) part du postulat que les amalgames lexicaux sont construits à partir de procédés phonologiques uniquement, elle s'appuie sur les données observées en hébreu. Le fait est qu'en français,

comme l'ont montré les exemples analysés, l'aspect sémantique joue un rôle important. Il reste à évaluer ce rôle et à formuler des contraintes qui permettent de déduire des règles sémantiques permettant de compléter les règles formelles.

Nous voyons aussi une problématique nouvelle se manifester à travers l'étude proposée ici. Les amalgames non typiques doivent-ils être assimilés aux amalgames typiques ? Autrement dit, faut-il poser deux catégories d'unités lexicales comme l'a fait Cusin-Berche (1999) et comme le propose Sablayrolles (à paraître) : les mots-valises d'un côté et les compocations de l'autre ? La mesure par prototypie – répartissant les amalgames lexicaux sur une échelle du plus typique au moins typique (cf. notamment Renner (2006) pour l'amalgamation lexicale en anglais) – est séduisante puisqu'elle permet de ne pas disperser en de multiples catégories des unités qui partagent de nombreux points communs, mais elle empêche en même temps d'élaborer une grammaire des amalgames lexicaux. Ces unités lexicales peuvent, grâce au travail fondateur de Bat-El (1996) notamment, bénéficier d'un cadre qui permet la construction d'une grammaire formalisant des contraintes d'optimalité, ce qui résulterait en la possibilité de développer des outils de génération ou de détection automatique, non seulement des amalgames lexicaux, mais aussi de leurs lexèmes-sources.

### Références bibliographiques

- Bat-El, O. (1996). Selecting the best of the worse: the grammar of Hebrew Blends. *Phonology*, 13, 283-328.
- Clas, A. (1987). Une matrice terminologique universelle : la brachygraphie gigogne. *Meta*, 32 (3), 347-355.
- Cusin-Berche, F. (1999). Le lexique en mouvement : création lexicale et production sémantique, *Langages*, 136, 5-26.
- Fradin, B. (2000). Combining forms, blends and related phenomena. In Doleschal, U., Thornton, A. M. (eds). *Extragrammatical and Marginal Morphology*. München : Lincom Europa, 11-59.
- Fradin, B. (2003). *Nouvelles approches en morphologie*. Paris : P.U.F.
- Fradin, B., Montermini, F., Plénat, M. (2009). Morphologie grammaticale et extragrammaticale. In Fradin, B. & al. *Aperçus de morphologie du français*. Saint-Denis : Presses universitaire de Vincennes, 21-45.
- Galisson, R. (1987). Les dictionnaires de parodie comme moyens de perfectionnement en langue française. *Études de linguistique appliquée*, 67, 57-118.
- Grésillon, A. (1984). *La Règle et le monstre : le mot-valise. Interrogations sur la langue, à partir d'un corpus de Heinrich Heine*. Tübingen : Niemeyer.
- Léturgie, A. (2011). Un cas d'extragrammaticalité particulier : les amalgames lexicaux. *Linguistica*, 51, 87-104.
- Piñeros, C.-E. (1999). Word-Blending as a Case of Non-Concatenative Morphology in Spanish. *Rutgers Optimality Archive*, 343-0999. <<http://roa.rutgers.edu/files/343-0999/343-PINEROS-0-0.PDF>>, consulté le 10 mars 2012.
- Piñeros, C.-E. (2004). The Creation of Portmanteaus in the Extragrammatical Morphology of Spanish. *Probus*, 16 (2), 203-240.
- Renner, V. (2006). Dépasser les désaccords : pour une approche prototypiste du concept d'amalgame lexical. In Pereiro M., Daniels, H. (dirs). *Le Désaccord*. Nancy : Publications de l'AMAES, 137-147.
- Sablayrolles, J.-F. (à paraître). Quelques remarques sur une typologie des néologismes : amalgamation ou télescopage : un processus aux productions variées (mots-valises, détournements...) et un tableau hiérarchisé. *CINEO 2011, 2<sup>e</sup> Congrès international des langues romanes*.
- Tournier, J. (1985). *Introduction à la lexicogénétiq ue de l'anglais contemporain*. Paris : Honoré Champion.

### Dictionnaires d'amalgames lexicaux

- Chiflet, J.-L. (1999). *Le Cafard laqué. Les mots-portemanteaux*. Paris : Mots & Cie.

- Chiflet, J.-L. (éd.). (2002). *Mais que fait l'Académie ? Le dictionnaire des mots qui devraient exister*. Paris : Mots & Cie.
- Chiflet, J.-L., Kristy, N. (1992). *Le Dictionnaire des mots qui n'existent pas*. Paris : Presses de la cité.
- Chiflet, J.-L., Kristy, N. (2003). *Le Mokimanké : le dico des mots qui existent enfin !*. Paris : Mots & Cie.
- Clément, Y.-M., Gréverand, G. (1993). *Pianissimots, petit dictionnaire des mots-valises*. Levallois Perret : Many.
- Créhange, A. (2004). *Le Pornithorynque est un salopare. Dictionnaire de mots-valises*. Paris : Mille et une nuits.
- Créhange, A. (2006). *L'Anarchiviste et le Biblioteckel. Dictionnaire de mots-valises*. Paris : Mille et une nuits.
- Finkielkraut, A. (2006). *Petit fictionnaire illustré. Les mots qui manquent au dico*. Paris : Seuil.
- Fournel, P., Cueco, H. (2007). *Les Animaux d'amour et autres sardinosaires*, Bordeaux : Le Castor Astral.
- Galisson, R., Porcher, L. (1986). *Distractionnaire*. Paris : Clé international.
- Guillaume, J.-C. (2008). *La valise diplodocumatique*. Saint-Denis : Édilivre-Aparis.
- Le Tellier, H., Gorce, X. (2007). *Les Opossums célèbres*. Bordeaux, Le Castor Astral.
- Loubière, S. (2003). *Éléphanfare. Le jeu des mots-valises pour les petits*. Paris : Albin Michel Jeunesse.
- Minda, A. (2004). *Mots-valises à poignées*. Paris : L'Harmattan.
- Roubaud, J. (2008). *Sardinosaires & Cie*. Bourges : Mille univers.
- Serguine, J. (1988). *Contradictionnaire*. Paris : Ramsay.
- Thibaud, J.-J. (2005). *Nouveaucabulaire*. Paris : Cherche midi.

---

<sup>1</sup> Le terme *amalgame lexical* est utilisé à la place de *mot-valise* afin de rassembler les mots-valises et ce que F. Cusin-Berche (1999) a appelé les compositions – résultats de lexèmes s'amalgamant sans partager un segment homophone, tels que *brunch* – au sein d'un ensemble unique. Dans la mesure où la littérature sur le sujet n'opère pas systématiquement une distinction entre ces différents types de formations qui s'opposent sur une caractéristique (le segment homophone commun) souvent tenue comme condition nécessaire de la formation des mots-valises, il semble préférable de réunir sous une étiquette plus large l'ensemble des unités créées par amalgamation lexicale.

<sup>2</sup> À ce sujet, cf. Fradin, Montermini et Plénat (2009).

<sup>3</sup> Il s'agit de la terminologie employée par Grésillon (1984 : 25).

<sup>4</sup> Entre parenthèses apparaissent des exemples d'amalgames correspondants au type identifié par le chercheur. Ces exemples sont directement cités des études commentées. Les lexèmes-sources sont présentés dans l'ordre dans lequel ils apparaissent dans l'output. Ils sont séparés par le point et les segments entre chevrons correspondent au matériel tronqué de chaque lexème-source. Le soulignement indique les segments homophones. Le mot en italiques pointé par la flèche est l'amalgame lexical.

<sup>5</sup> Il s'agit de l'analyse de Grésillon (1984), *enseignement* pouvant également être analysé en *en<seignement>•saignement → enseignement*. Cette dernière analyse masque cependant l'enchâssement.

<sup>6</sup> Au sujet de la composition par fracto-morphèmes, cf. Tournier (1985). Fradin (2003 : 213) parle de « composition cachée » et assimile ces éléments à des formes abrégées de lexèmes complets, tels que *télévision* ou *logiciel* pour les segments *télé-* et *-ciel*, conservant leur contenu sémantique.

<sup>7</sup> Les lexèmes-sources sont indiqués par les lettres A et B, le segment homophone étant signalé par Σ.

<sup>8</sup> Les exemples cités sont ceux fournis par Fradin (2000 : 28-30). Tous les schémas ne sont pas reproduits.

<sup>9</sup> En réalité, le fait de trouver des segments identiques en dehors des segments tronqués joue un rôle important dans l'interprétabilité de l'output. Il faudrait donc prendre ces segments en compte dans le cas de *guilmauve* notamment, dans la mesure où le segment *-mauve* n'apparaît que dans *guilmauve* tandis que le segment *guili-* n'apparaît quant à lui que dans *guili-guili*. La présence de ces segments contribue donc grandement à l'interprétation de l'amalgame.

<sup>10</sup> Ce corpus est présenté au §3.1.



---

<sup>11</sup> Bat-El (2000) soutient même que des procédés prétendus comme étant extragrammaticaux, dont l'amalgamation lexicale, ne le sont pas.

<sup>12</sup> Les contraintes proposées par Bat-El (1996) ont été traduites pour des raisons de commodité. Nous signalons néanmoins systématiquement la désignation originale utilisée par Bat-El.

<sup>13</sup> Bien que *chatil* puisse être perçu comme une dérivation du lexème *chat* par suffixation de *-il* inféré de *chenil*, nous l'avons considéré comme un amalgame dans la mesure où il provient de la base de données COAL, et a donc été forgé en tant qu'amalgame lexical. Cependant, *chatil* semble bien faire partie des exemples les moins typiques que nous avons pu réunir.

<sup>14</sup> C'est d'ailleurs ce qui était avancé au §2.1.2 et qui confère au corpus de Grésillon (1984) un caractère trop hétérogène pour élaborer une typologie générale de l'amalgamation lexicale en français.