



**HAL**  
open science

# Vers une morphologie de l'évidence : d'une morphologie de l'input à une morphologie de l'output

Georgette Dal

► **To cite this version:**

Georgette Dal. Vers une morphologie de l'évidence : d'une morphologie de l'input à une morphologie de l'output. Linguistique. Université Lille 3, 2004. tel-01509249

**HAL Id: tel-01509249**

**<https://hal.univ-lille.fr/tel-01509249v1>**

Submitted on 3 Jun 2017

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

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

Public Domain

**Université Lille 3  
2003-2004**

**Georgette Dal**

**Vers une morphologie de l'évidence :  
d'une morphologie de l'input à une morphologie de l'output**

*Document de synthèse présenté dans le cadre de l'habilitation à  
diriger des recherches le 29 juin 2004*

Jury :

Madame Anne-Marie Berthonneau, Professeure, Université Lille 3 [garante]

Monsieur Geert Booij, Professeur Vrije Universiteit Amsterdam

Monsieur Bernard Fradin, Directeur de Recherche, CNRS [pré-rapporteur]

Monsieur Christian Jacquemin, Professeur, Université de Paris Sud

Monsieur Marc Plénat, Directeur de Recherche, CNRS [pré-rapporteur et président]

*Je dédie cette habitation à diriger des recherches  
à la mémoire de Danielle Corbin,  
sans qui la morphologie du français  
ne serait pas ce qu'elle est aujourd'hui,  
et sans qui je n'aurais jamais rencontré  
la morphologie.*

# Remerciements

Je remercie les nombreux collègues, dont certains sont bien plus que cela, de m'avoir poussée à présenter cette habilitation à diriger des recherches : je pense à Philip Miller, à Anne-Marie Berthonneau, à Marc Plénat, à Bernard Fradin, à Françoise Kerleroux, à Dany Amiot, à Nelly Flaux, à Pierre Corbin, à Marleen Van Peteghem, à Danièle Van de Velde, et j'en oublie certainement.

Un merci tout particulier :

- à Anne-Marie Berthonneau, pour la patience qu'elle a eue à m'écouter lui raconter mes affres morphologiques, et pour avoir accepté de m'accompagner dans l'écriture de cette note de synthèse,
- à Anne-Marie Berthonneau, de nouveau, à Geert Booij, Bernard Fradin, Christian Jacquemin et Marc Plénat, pour avoir accepté de faire partie de mon jury d'habilitation,
- à Pierre Corbin, co-morphologue émérite, pour les conversations de morphologie impromptues que nous avons eues au détour d'une photocopieuse, ou dans le couloir menant au parking souterrain de l'Université à des heures indues,
- à Bernard Fradin, Dany Amiot, Marc Plénat pour les questions, parfois saugrenues, que j'ai pu leur poser, et surtout les réponses qu'ils m'y ont apportées,
- à Katia Paykin, pour sa relecture très minutieuse.

Un grand merci enfin à mon mari, Bruno, et à nos enfants, Valentin et Capucine, pour leur aptitude à comprendre ce que représente le travail de chercheur, et à supporter mon manque total de disponibilité depuis un certain temps sans jamais s'en plaindre.

# Table des matières

<b>Introduction</b> .....	1
<b>1. Sous-bassement théorique : le modèle constructionnel de Danielle Corbin</b> .....	5
1.1. Principe d'associativité .....	5
1.2. Mot construit et règle de construction de mots .....	8
1.3. Les outils .....	9
<b>2. D'une morphologie de la non évidence à une morphologie de l'évidence</b> .....	11
2.1 La grammaire du suffixe <i>-et(te)</i> .....	11
2.1.1. Pour un dépassement de l'évidence observable .....	11
2.1.2. Problèmes théoriques et empiriques.....	15
2.1.2.1. Le marquage d'appropriation : le bon niveau de généralité ? .....	15
2.1.2.2. Le problème des stades virtuels.....	17
2.1.2.3. Le principe de compositionnalité du sens des mots construits par rapport à leur structure.....	19
2.1.3. La grammaire du suffixe <i>-et(te)</i> : bilan provisoire .....	24
2.2. Le projet MorTAL .....	25
2.2.1. Présentation générale .....	25
2.2.2. Présentation de la base de données .....	32
2.2.3. Bilan linguistique .....	38
2.2.3.1. Phénomènes constructionnels non centraux .....	39
2.2.3.2. Etymologie ou construction en français ? .....	43
2.2.3.3. Le problème des choix multiples.....	47
2.2.3.4. Bilan linguistique global .....	53
2.3. La catégorisation des bases de construction de lexèmes.....	54
2.3.1. D'un principe d'unicité catégorielle à un principe d'unicité sémantique ..	54
2.3.2. Les catégories vues par les règles de construction de lexèmes .....	60
2.4. Lexique construit et relations paradigmatiques .....	66
2.4.1. Lexique construit et théorie de l'optimalité .....	69
2.4.2. Lexique construit et analogie .....	75
<b>Conclusion : quelle morphologie aujourd'hui ?</b> .....	87
<b>Bibliographie</b> .....	93
<b>Annexe 1. Liste de mes travaux depuis ma thèse</b> .....	103
<b>Annexe 2. Liste des travaux générés par le projet MorTAL</b> .....	109
<b>Index des auteurs</b> .....	111

# Introduction

En morphologie dérivationnelle (ou morphologie *constructionnelle* pour utiliser l'étiquette qui s'impose désormais, au moins dans le domaine français<sup>1</sup>), l'heure du bilan sonne tous les dix ans environ. Pour ne citer que quelques exemples :

- en 1974, Peters Matthews consacrait le premier chapitre de son ouvrage à justifier pourquoi re-constituer en objet d'étude la morphologie en général, la morphologie dérivationnelle en particulier. Il est vrai qu'à l'époque, l'étude des mots était « momentarily out of fashion » (Matthews 1974 : 3)<sup>2</sup> ;
- onze ans plus tard, Jean Molino se demandait où en était la morphologie dans le numéro 78 de *Langages* qu'il dédiait au « retour de la morphologie » ;
- en 1994, Bernard Fradin faisait un point circonstancié sur les développements récents du domaine, tandis que Mark Aronoff intitulait l'un de ses ouvrages *Morphology now* ;
- en 2003, Bernard Fradin publiait son ouvrage *Nouvelles approches en morphologie*, tandis qu'avec Françoise Kerleroux, il coordonnait un numéro de la revue *Langages* intitulé *Quoi de neuf en morphologie ?*.

Ce type d'interrogation n'est pas propre au domaine (*cf.* par exemple le titre de Laks éd. (2000), *Où en est la phonologie du français ?*). En morphologie constructionnelle, cette inquiétude permanente s'explique cependant par la crise viscérale qui touche le domaine, menacé sur plusieurs de ses fronts, après avoir régné en maître jusqu'à la période du structuralisme (pour un point historique très documenté, je renvoie au premier chapitre de Beard 1995). Avatar de la syntaxe pour certains, aux temps pionniers de la grammaire générative (*cf.* Lees 1963) comme dans ses plus récents développements, appendice de la phonologie pour d'autres, la morphologie n'a en effet cessé de prouver sa légitimité<sup>3</sup>.

En cette année 2004, sans prétendre que les recherches que j'ai effectuées depuis ma thèse que j'ai soutenue en 1994 permettent assurément de dresser un bilan qui vaille

- 
1. Comme le note P. Corbin (sous presse : n. 2), une partie des morphologues français utilisent désormais de façon tout à fait naturelle cette dénomination. Sauf erreur, elle apparaît pour la première fois dans D. Corbin (2001a : 42, n. 3). Corbin & Paul (2000 : 51, n. 1) parlent, eux, de « théorie constructionnelle ». Pour une justification de ce choix dénommatif, *cf.* D. Corbin (à paraître, chap. 1, § 2.2.2.) auquel renvoie P. Corbin (*ibid.*).
  2. La reconnaissance de sa légitimité 17 ans plus tard, par au moins une partie de la communauté, explique, elle, que ce chapitre introductif, intitulé « Why study morphology ? » dans la première édition de l'ouvrage, devienne « What is morphology ? » dans l'édition revue de 1991.
  3. Y compris, comme me le fait remarquer Pierre Corbin, dans des travaux où son existence est supposée aller de soi (*cf.* Molino 1985, Aronoff 1994, Fradin & Kerleroux 2003a, pour ne citer que quelques exemples).

pour la discipline dans son ensemble, je dirais que mon parcours personnel constitue un lointain reflet des changements épistémologiques perceptibles en morphologie constructionnelle pendant la même période.

En dix ans en effet, et avec un certain retard par rapport à d'autres domaines de la linguistique, à la phonologie notamment, dont le passage des règles aux représentations remonte aujourd'hui à plus de vingt ans, la morphologie constructionnelle est passée d'une morphologie mettant la seule focale sur l'input<sup>4</sup> à une morphologie qui, sans négliger l'input, prend également en considération l'output. Ce changement de cap doit notamment aux travaux menés dans le cadre de la théorie de l'optimalité et à l'essor des formalismes déclaratifs comme LFG ou HPSG, qui permettent d'énoncer les conditions de bonne formation sur les représentations. Il s'accompagne d'un abandon progressif des règles conçues sur le mode à la fois procédural et séquentiel, telles que les envisageait la grammaire générative standard – par règle, désormais, on entend de plus en plus souvent 'expression d'une régularité' sans prétendre dire quoi que ce soit du mode opératoire : cf. Plag (1999 : 17-22) et, plus nettement encore, Fradin (2003a : 264 *sqq.*) –, voire de l'abandon de l'idée même de dérivation (cf. les travaux de Luigi Burzio, par exemple 2001<sup>5</sup>).

En même temps que, dans le domaine de la morphologie constructionnelle, les règles au sens de la grammaire générative standard disparaissent peu à peu, une partie de la machinerie permettant de passer d'un niveau de dérivation à l'autre devient caduque, si bien que l'on s'achemine désormais vers ce que j'appellerai une morphologie de l'évidence<sup>6</sup> : non que les régularités dégagées soient désormais triviales et ne nécessitent plus un travail méticuleux de description, mais elles sont davantage écologiques<sup>7</sup>, en ceci qu'elles regagnent en intuitivité.

Or, dans le même temps, mes recherches se dégagent peu à peu du cadre théorique de Danielle Corbin, qui considérait que les règles s'appliquaient à une base pour former un dérivé, moyennant, le cas échéant, la reconstruction d'étapes intermédiaires d'ajustements formels et/ou sémantiques entre une structure sous-jacente et une forme de surface, pour aller explorer des choses aussi variées que l'analogie ou la théorie de l'optimalité.

C'est ce que montrera cette note de synthèse. Après avoir rappelé dans ses grandes lignes le cadre théorique élaboré par Danielle Corbin, je dresserai dans un second temps

- 
4. J'utilise ici délibérément les dénominations d'*input* et d'*output*, auxquelles on pourrait préférer celles d'*entrée* et *sortie*, parce que ce sont elles qui sont employées actuellement dans bon nombre de travaux, y compris dans le domaine français. Par ailleurs, l'abandon de la notion de dérivation en morphologie qui se dessine actuellement rend impropre le couple *base / dérivé*.
  5. Le titre de Burzio (2001), « Zero derivation », proclame cet abandon à la manière d'un slogan. Par ce titre, Luigi Burzio entend réagir à la proposition faite par Rubach (2000) d'aménager la théorie de l'optimalité en y réintroduisant la notion de dérivation, alors même que le fondement de la théorie de l'optimalité est précisément d'évacuer cette notion.
  6. Cette appellation constitue un clin d'œil au titre de la première partie de D. Corbin (1987a), « Contre une morphologie de l'évidence », dont je reparlerai au §.1.1. *infra*.
  7. J'emprunte ici, en l'adaptant, le terme *écologique* aux psychologues, pour qui l'étude d'un comportement est écologique quand l'expérimentation place l'individu dans un contexte le moins artificiel possible.

un bilan de mes recherches de ces dix dernières années en revisitant mes travaux les plus anciens. Je profiterai de cet exposé pour établir dans ma conclusion un cahier des charges minimal pour mes recherches futures.



# 1. Sous-bassement théorique : le modèle constructionnel de Danielle Corbin

L'ensemble des travaux que j'ai menés depuis ma thèse soutenue en octobre 1994 à l'Université de Lille 3 sous la direction de Danielle Corbin se situe dans le domaine de la morphologie constructionnelle, dont la tâche, d'après Fradin (2003a : 5), « consiste à calculer les corrélations qu'entretiennent la forme phonique et le sens des mots construits »<sup>1</sup>. C'était du reste déjà le cas de ma thèse consacrée au suffixe *-et(te)* français dont mon ouvrage de 1997 constitue une version sensiblement remaniée (cf. annexe 1, [0]<sup>2</sup>), ainsi que des travaux qui l'ont précédée, qu'ils soient universitaires (mémoires de maîtrise en 1987 et de DEA en 1988, dirigés l'un et l'autre par Danielle Corbin) ou publiés (cf. Dal 1991 et 1993 ; D. Corbin & al. 1993).

À cette homogénéité de domaine s'ajoute une homogénéité théorique. En effet, pour d'évidentes raisons autobiographiques qui apparaissent dans le bref rappel ci-dessus, la plupart de mes travaux prennent pour cadre théorique le modèle de morphologie constructionnelle élaboré par Danielle Corbin, ou se situent par rapport à lui. Pour comprendre leur évolution, il est par conséquent nécessaire de commencer par exposer rapidement ce modèle. Pour ce faire, je me fonderai principalement sur D. Corbin (1991a), qui en constitue la dernière version publiée<sup>3</sup>, en complétant le cas échéant par des renvois à des travaux ultérieurs de Danielle Corbin.

## 1.1. Principe d'associativité

Le principe fondateur qui guide le modèle, que ce soit dans sa version initiale de 1987, dans sa version réactualisée de 1991, ou dans les derniers travaux publiés de Danielle Corbin, est celui de l'associativité. Selon ce principe, « le sens d'un mot construit<sup>4</sup> est construit en même temps que sa structure morphologique, et compositionnel-

- 
1. Cf. également Zwicky (1992 : 327) : «A theory of morphology specifies what sorts of systems can associate the meanings of words to their pronunciation?».
  2. Les numéros en caractère gras entre crochets renvoient à ceux de mes travaux cités dans l'annexe 1 de la présente note de synthèse. Les travaux numérotés entre [1] et [24] figurent dans le volume annexe, sauf [23].
  3. Pour la version actuelle du modèle, je renvoie à D. Corbin (à paraître). Même si on peut voir les infléchissements que prend le modèle dans les derniers travaux publiés de Danielle Corbin, je fais l'hypothèse, peut-être à tort, que les principes qu'elle faisait siens en 1991 demeurent d'actualité dans son ouvrage à paraître.
  4. En 1991 et dans ses travaux antérieurs, Danielle Corbin parle de *mot* construit. Ultérieurement, sans renoncer définitivement à cette étiquette (par ex. D. Corbin 2001b), elle parlera d'*unité lexicale* construite (D. Corbin 1997a), dénomination à laquelle on pourra préférer celle de *lexème* construit,

lement par rapport à celle-ci »<sup>5</sup> (D. Corbin 1991a : 9). Par ce choix de la première heure, qui s'inscrit dans la lignée des travaux de Dell (1970) et (1979), de Halle (1973) ou d'Aronoff (1976), le modèle entendait se démarquer de la conception dominante en grammaire générative standard pour les modèles dits 'dissociatifs'. Selon ces derniers, la structure des mots construits et leur sens sont construits de façon disjointe, la structure étant première, et le sens étant affecté aux structures par des règles interprétatives (D. Corbin 1991a : 9), comme cela se passait – et se passe encore – en syntaxe dans le paradigme chomskien. La conséquence en est que, comme les autres théories associatives, et comme d'ailleurs une partie des modèles actuels, le modèle revendique l'autonomie de la morphologie par rapport à la syntaxe (D. Corbin 1991b : 35). Dans toutes les théories, la reconnaissance de l'autonomie (ou de la non autonomie) d'un composant morphologique par rapport au composant syntaxique fonctionne en effet comme une prémisse, ayant valeur d'axiome<sup>6</sup>. Ainsi peut-on opposer la posture théorique du modèle de Danielle Corbin avec celle, par exemple, de Lieber (1992 : 1), dont le prologue commence en ces termes :

This book is an attempt to articulate a theory of word formation based on the premise that there is no separate component of morphology in the grammar.

Malgré cela, la démarche adoptée par Danielle Corbin s'inscrit dans le paradigme de la grammaire générative standard :

- parce qu'elle se veut *générative* : elle veut rendre compte des mots construits attestés mais aussi donner les moyens de prédire les mots construits possibles<sup>7</sup>, et rien qu'eux ;
- parce qu'elle distingue deux niveaux d'analyse : celui de la structure morphologique, parfois dite 'profonde' (D. Corbin 1991b : 37), et celui de la forme observable, la seconde pouvant ne pas être le fidèle reflet de la première (D. Corbin 1991a : 12).

Cette stratification de l'analyse au niveau formel se retrouve au niveau sémantique, puisque le sens dérivationnellement construit d'un mot construit est réputé ne pas toujours coïncider avec son ou ses sens lexicalisés (D. Corbin 1991a : 11-12 ; 1991b : 37).

Le modèle pose, par principe, qu'au prix d'un travail d'abstraction par rapport au niveau de l'observable, aucune donnée ne doit résister à l'analyse. Aussi, pour pouvoir

---

pour les raisons qu'expose Fradin (2003a : 104). Dans ce paragraphe, j'utiliserai préférentiellement l'étiquette *mot* construit puisque c'est celle qu'utilise Danielle Corbin dans les travaux sur lesquels je me fonde. Par la suite, j'utiliserai, autant que faire se peut, celle de *lexème* construit.

5. On s'en tiendra ici à une compréhension intuitive du principe de compositionnalité tel que l'énonce Danielle Corbin (on comprendra pourquoi au § 2.1.2.3). J'aurai l'occasion de revenir sur ce principe à plusieurs reprises dans l'exposé de mes travaux, dans la mesure où sa remise en cause progressive tel qu'il est formulé est centrale dans le déplacement de mes recherches vers une morphologie de l'évidence.
6. Il est du reste intéressant de constater que les deux camps argumentent à partir des mêmes données pour asseoir leur option théorique. Ainsi, Fu & al. (2001) font état des prises de position opposées adoptées par Jane Grimshaw et Hagit Borer à partir de l'observation des mêmes noms de procès. Malgré cela, on imagine mal de ne pas se situer dans ce débat, au moins implicitement.
7. Par *possible*, j'entendrai désormais 'construit conformément aux règles de la grammaire, mais non répertorié dans les dictionnaires'. Cette définition correspond à peu près à celle qu'on trouve dans D. Corbin (1997b : 79), et, surtout, à l'usage effectif qui est fait de l'appellation dans les travaux menés dans le cadre théorique de Danielle Corbin.

appliquer le principe de compositionnalité et réduire l'ensemble des données observables apparemment récalcitrantes met-il à disposition un certain nombre de mécanismes que Danielle Corbin situe à un niveau post-dérivationnel (D. Corbin 1991a : 12), permettant de passer du niveau dérivationnellement prédictible au niveau observable.

Au niveau formel, le modèle distingue trois types de mécanismes :

– mécanismes de troncation, qu'on trouve déjà dans Aronoff (1976) et Booij (1977) même si Danielle Corbin ne les utilise pas exactement aux mêmes fins (D. Corbin 1987a : 341-370). Elle s'en sert pour expliquer « l'absence superficielle de segments dont la présence est prédite par une structure morphologique conforme à l'interprétation » (D. Corbin 1991a : 12) : par exemple, les segments attendus /ɛz/ <sup>8</sup> et /jɔ̃/ absents de *japonerie* et *séparatiste* si, comme D. Corbin (*ibid.* : 13), on les dérive de *japonais* et *séparation* ;

– mécanismes d'allomorphie, auxquels Danielle Corbin recourt pour expliquer aussi bien les mots construits en français que les mots hérités, dès lors qu'ils sont analysables comme construits en français. Par exemple, dans D. Corbin (1991b : 41), elle affecte à l'adjectif français *floral* la structure [ [fleur]<sub>N</sub> (al)<sub>af</sub> ]<sub>A</sub>, tout en rappelant qu'il constitue un calque du latin *floralis*. Elle explique la distorsion entre la forme observable de cet adjectif et sa forme prédictible à partir de cette structure (notée <sup>+</sup>*fleural*) au moyen d'une allomorphie consistant en une postériorisation vocalique ;

– mécanismes d'intégration paradigmatique, également appelée « marquage de classe » dans des travaux ultérieurs (D. Corbin 2001a : 50), permettant « d'expliquer la présence superficielle de segments affixoïdes <sup>9</sup> non nécessaires à l'interprétation » (D. Corbin 1991a : 14), et dont le rôle est de rendre compte des relations formelles qu'entretiennent certains mots construits, sans que les uns dérivent des autres. L'un des exemples récurrents dans les travaux de Danielle Corbin est l'adjectif *antigrippal*, doublet d'*antigrippe* dans les dictionnaires, dont le sens compositionnel se calcule à partir de l'application d'*anti-* à *grippe*, sans que *-al* intervienne sémantiquement <sup>10</sup>.

Au niveau sémantique, les mécanismes utilisés pour expliquer une distorsion entre un sens dérivationnellement construit et un sens lexicalisé sont pour l'essentiel la métonymie, la synecdoque, la métaphore et la catachrèse (*i.e.* la conservation d'une dénomination, quand cette dernière paraît sémantiquement immotivée : *cf.* l'exemple de *pommade* développé dans D. Corbin 1991b : 42). Notamment sous la pression des noms en *-et(te)* et en *-asse* (Dal 1991), le modèle prévoit que chacun de ces transferts de dénomination puisse s'appliquer en amont ou en aval des règles de construction de mots, ce qui peut avoir pour effet de masquer un sens dérivationnellement prédictible (D. Corbin 1991a : 11). Par exemple, D. Corbin (1993 : 201) explique le sens de *cartonnier* dans

---

8. Par convention, les barres obliques encadreront désormais des séquences phoniques, les chevrons des séquences graphiques.

9. Dans les faits, ces segments sont exclusivement suffiximorphes en français. J'ignore s'il existe dans le monde des langues qui présenteraient de tels segments ailleurs qu'en finale.

10. Dans D. Corbin (2001a : 51), s'inspirant de la théorie de l'optimalité dans laquelle des contraintes peuvent entrer en conflit (*cf. infra*, § 2.4.1.), elle considère que la forme *antigrippe* représente la « victoire de la contrainte de conformité sémantique, alors que la forme *antigrippal* représente celle de la contrainte de conformité catégorielle ».

*guêpe cartonnière* (“guêpe qui construit son nid avec une substance rappelant le carton”) comme résultant d’une opération de métaphorisation du nom *carton* préalablement à la suffixation par *-ier*. Le sens de *couturière* dans *fauvette couturière* (“fauvette qui construit un nid dont les feuilles semblent cousues”) s’explique, lui, par une opération de métaphorisation sur le nom *couturier*<sup>11</sup>. D. Corbin & al. (1993) systématisent l’observation pour expliquer les nombreuses références attestées du nom *lunette(s)*.

Pour résumer ce qui précède en un slogan, je dirai que le souci constant du modèle est d’homogénéiser l’hétérogène, et de régulariser l’irrégulier, étant entendu qu’il ne faut jamais se fier aux apparences : le titre de la première partie de D. Corbin (1987a), « Contre une morphologie de l’évidence », est à cet égard éloquent.

## 1.2. Mot construit et règle de construction de mots

Etant admis ces préliminaires, le modèle pose qu’un mot est construit quand il satisfait les deux conditions suivantes (D. Corbin 1991b : 37)<sup>12</sup> :

- les constituants de sa structure morphologique doivent être catégorisables et associables à un sens de façon reproductible ;
- son sens prédictible et sa structure morphologique profonde doivent être engendrés par une règle de construction de mot (notée RCM<sup>13</sup>), de telle sorte que le premier soit compositionnel par rapport à la seconde.

À son tour, une RCM se définit par l’association de trois composants :

- (i) une opération structurelle instaurant un rapport catégoriel entre la base et le dérivé. Ce rapport catégoriel est réputé unique dans D. Corbin (1987a), puis réputé souffrir des entorses à partir de 1991, en particulier dans le domaine des suffixes évaluatifs et des préfixes spatio-temporels (D. Corbin 1991a : 21 ; 1991b : 39, n. 12). Il est beaucoup moins intangible dans D. Corbin (2001a) ;
- (ii) une opération sémantique, conférant un même sens fondamental à tous ses produits. Par exemple, dans D. Corbin (1993 : 71) :

tous les adjectifs dénominaux en français, quel que soit le procédé qui ait servi à les construire [...] ont un comportement sémantique semblable, à un certain niveau de généralité : dans le

- 
11. Dans Dal (1991 : 228-229), j’attribue pareillement la référence de *mélasse* à un « résidu sirupeux de la cristallisation du sucre » (*Nouveau Petit Robert* 2001, désormais *PRO1*) à l’application successive d’une métaphore au nom *miel*, puis du suffixe *-asse* à ce sens métaphorique. Je recourrais à *bécasse* pour mettre en évidence l’ordre inverse : d’abord suffixation, puis transfert de dénomination (métonymie, dans le cas présent).
  12. D. Corbin (1991b) énonce trois conditions, mais la troisième, qui prévoit le recours aux mécanismes dont il s’est agi plus haut en cas de distorsion entre une forme et/ou un sens observable(s) et une forme et/ou un sens prédictible(s), n’en est pas une. Elle donne les moyens de rendre valide la seconde condition.
  13. La dénomination *règle de construction de mots* (et le sigle correspondant) est calquée sur celle des *word formation rules* (abrégée en WFR) d’Aronoff (1976).

cadre canonique d'un syntagme nominal, ils servent à dire qu'il y a une certaine relation entre le référent de leur nom recteur et celui de leur base nominale

Ainsi conçue, l'opération sémantique associée à une RCM est une sorte de plus grand dénominateur commun sémantique à tous les procédés qui relèvent de son paradigme morphologique (*cf.* iii) ;

(iii) un paradigme morphologique, rassemblant un ensemble de procédés constructionnels qui ne sont pas concurrents entre eux, y compris sémantiquement (D. Corbin 1991b : 39).

Le modèle détaille les composants des RCM. Mais il ne dit rien du statut de ces règles, si ce n'est que, généralement, elles constituent le bon niveau de généralisation (D. Corbin 1993 : 71). Il ne précise en particulier pas si elles correspondent à une réalité psychologique chez les locuteurs, ni si les étapes intermédiaires éventuellement restituées entre le niveau profond et le niveau de surface disent quelque chose de la procédure effectivement suivie, quand on forme ou quand on entend un lexème construit. En revanche, et contrairement à d'autres modèles comme, par exemple, celui de Jackendoff (1975), le modèle conçoit explicitement les règles sur le mode génératif, et non sur le mode statique. Alors que, dans Jackendoff (*ibid.*), les règles sont des règles de redondance appariant des unités du lexique<sup>14</sup>, dans le modèle de Danielle Corbin, il s'agit de règles de production orientées, qui vont des unités moins complexes vers les unités plus complexes. Enfin, et ce sera le dernier point que je soulignerai, étant donné le principe de compositionnalité sur lequel elles reposent, les RCM mettent en jeu une procédure exclusivement syntagmatique<sup>15</sup>, excluant de leur champ tout ce qui n'est pas bases et procédés constructionnels : elles sont en ce sens comparables aux règles de réécriture en syntaxe.

### 1.3. Les outils

Pour en terminer avec l'exposé du modèle théorique qui a servi de cadre de référence, de façon explicite ou implicite, à l'ensemble de mes travaux, je dirai quelques mots sur les outils auxquels il recourt pour valider les hypothèses qu'il formule.

---

14. J'utilise ici le terme *lexique* de façon non contrainte. Pour un point sur les notions de lexiques restreint, étendu et extensif, *cf.* Fradin (2003a : 184 *sqq.*). En un mot, le lexique étendu est « l'ensemble des unités non fléchies correspondant aux unités atomiques dont a besoin la grammaire » (*ibid.* : 184), qu'elles soient simples ou complexes, le lexique extensif l'ensemble des unités fléchies obtenues en appliquant les règles flexionnelles aux éléments du lexique étendu, le lexique restreint le lexique exempt de toute redondance. Pour une discussion sur la supériorité des règles statiques par rapport aux règles génératives, *cf.* Zwicky (1992 : 341-342).

15. Il est inutile à mon présent propos de chercher à expliciter si cette procédure syntagmatique relève de ce que Fradin (2003a) appelle la « morphologie morphématique combinatoire », ou de la « morphologie lexématique classique ». J'aurai l'occasion de revenir sur des deux types d'approche à plusieurs reprises dans la suite de ce travail.

Dans les faits, ces outils sont assez peu nombreux, puisqu'ils reposent sur un travail d'introspection de la part du descripteur, et sur l'attestation dans les dictionnaires (D. Corbin 1987a : 46 *sqq.*).

C'est en effet pour l'essentiel par un travail d'introspection que le descripteur accepte (ou rejette) une forme comme possible (ou impossible). Ce travail d'introspection prend appui sur des analyses argumentées quand il s'agit de formes intermédiaires intervenant au cours d'une analyse, quitte à ce que ces formes soient jugées contre-intuitives. Je citerai ici le nom *°buande* et l'adjectif *°buandier* qui en dérive, dont font l'hypothèse Corbin & Corbin (1991) pour expliquer le nom *buandier* ("personne qui lave le linge") à partir du verbe *buer* ("faire la lessive")<sup>16</sup>.

Ce travail par introspection n'est pas propre au modèle de Danielle Corbin, même s'il y est particulièrement sollicité. Il était en effet à la base de la plupart des travaux en grammaire générative standard, avant l'avènement des très gros volumes de données textuelles sous forme informatique, qui, comme le notent Habert & Zweigenbaum (2002 : 88), remettent en question « la validité de l'intuition d'acceptabilité ou inversement l'impression que tel énoncé est "impossible" »<sup>17</sup>. Cependant, comme le suggèrent *°buande* et l'adjectif *°buandier* ci-dessus, l'introspection ne se confond pas avec une simple intuition du descripteur, reposant sur sa compétence de locuteur de tous les jours.

Quant aux dictionnaires, Danielle Corbin écrit d'eux qu'ils « sont un instrument irremplaçable pour le linguiste, et en particulier pour le morphologue » (D. Corbin 1992 : 191), même si « on aurait tort [...] de prendre les données lexicographiques pour des analyses linguistiques » (*ibid.*).

En revanche, le modèle recourt peu, pour ne pas dire pas, aux corpus textuels, en particulier aux corpus sur supports électroniques, pour des raisons évidentes. Leur essor est trop récent pour qu'ils aient pu être utilisés massivement dans les années 1990. Il s'ensuit que les mots construits sont la plupart du temps considérés hors emploi ; quand ils sont contextualisés, et ils le sont rarement, ils sont placés dans des énoncés non authentiques, qui sont le fait du descripteur. On comprend également que le descripteur n'ait pas d'autre moyen que l'introspection pour vérifier si les mots construits qu'il pose comme possibles ou comme impossibles le sont vraiment. Il est à cet égard plausible que Danielle Corbin aurait révisé certaines de ses analyses des années 1990, si la consultabilité d'Internet avait été ce qu'elle est aujourd'hui<sup>18</sup>.

---

16. Je citerai désormais les verbes bases de construction de lexèmes en utilisant leur forme dénominative conventionnelle (l'infinitif, pour le français), étant entendu qu'il est communément admis que les procédés constructionnels s'appliquent sur des lexèmes, donc sur des entités abstraites de leurs marques flexionnelles (D. Corbin 1987a, Fradin 2003a, etc.). L'infixe d'infinitif n'est donc pas à prendre en considération dans les exemples cités. La remarque vaut pour les verbes construits.

17. Comme le montrent les travaux menés par Marc Plénat, seul ou en collaboration, l'essor et la consultabilité des gros volumes textuels sur support électronique ont également permis de mettre en évidence des phénomènes jusqu'alors restés inaperçus. Je renvoie par exemple à Plénat (2000) sur le suffixe *-esque*, à Plénat (2002) sur le suffixe *-issime* et à Hathout & al. (2003), qui montrent le bien-fondé de ce qu'ils appellent une « approche extensive » de la suffixation par *-able* en français.

18. *A priori*, l'enquête constituait l'un des moyens de vérification possibles avant l'avènement des grands volumes textuels informatisés, à ceci près qu'en matière de lexique construit, les résultats sont peu fiables (*cf.* D. Corbin 1987a : pp. 599-606). Plénat (2000 : 51 *sqq.*) est moins pessimiste quant à la fiabilité des enquêtes.

## 2. D'une morphologie de la non évidence à une morphologie de l'évidence

Avant de commencer ce travail de synthèse, j'aurais dit que les travaux menés depuis ma thèse n'avaient d'autres points communs qu'une unité de domaine, la morphologie constructionnelle, et une unité théorique, le cadre développé par Danielle Corbin. La prise de recul qu'impose l'écriture d'une habilitation à diriger des recherches m'a permis de me rendre compte que mes travaux étaient guidés par un autre fil rouge, et qu'au gré du temps, ils étaient passés insensiblement d'une morphologie de la non évidence à une morphologie de l'évidence, s'écartant ce faisant peu à peu de la prise de position théorique de Danielle Corbin rappelée ci-dessus. En 1987, Danielle Corbin refusait de se rendre à l'évidence en matière de morphologie dérivationnelle. Or, en morphologie comme ailleurs, peut-on raisonnablement nier l'évidence ?

La présentation que je ferai dans ce qui suit tiendra compte de cette évolution. Je commencerai par exposer ceux d'entre mes travaux qui appliquent, sans le discuter, le parti pris préconisé par Danielle Corbin, en les revisitant : il ne s'agira donc pas d'un simple résumé de mes travaux, mais d'une relecture pour les plus anciens d'entre eux. J'ordonnerai ensuite les travaux suivants selon les contre-évidences qu'ils interrogent.

Mon ouvrage de 1997, *Grammaire du suffixe -et(te)*, ainsi que le projet MorTAL et les publications qui lui sont associées mettent en œuvre, chacun à leur manière, le modèle élaboré par Danielle Corbin sans remettre en cause ses fondements. Aussi commencerai-je par eux, en dressant pour chacun un bilan final faisant apparaître les problèmes rencontrés <sup>1</sup>.

### 2.1. La grammaire du suffixe -et(te)

#### 2.1.1. Pour un dépassement de l'évidence observable

*Grammaire du suffixe -et(te)*, publié en 1997 dans la collection "Etudes de sémantique lexicale" de l'Institut National de la Langue Française, est la version largement remaniée de ma thèse soutenue en 1994 (le remaniement m'a demandé un an). Il porte, comme l'indique son titre, sur le suffixe -et(te) français, à l'exclusion des autres procédés formant des lexèmes à valeur évaluative ou diminutive.

---

1. Appliquent également sans réserve le modèle de Danielle Corbin [1] et [2]. Je ne les présenterai pas ici, dans la mesure où ils n'apportent guère d'éléments nouveaux par rapport à mon présent propos.

Ce travail constitue un véritable plaidoyer pour une morphologie de la non évidence. Il applique sans réserve le modèle de Danielle Corbin – du moins le croit-il, on verra pourquoi –, à ceci près qu’il porte précisément sur l’un des cas d’entorse au principe d’unicité catégorielle défendu par Danielle Corbin (*cf.* ci-dessus, § 1.1.), sur lequel je reviendrai au § 2.3. Il se donne en effet comme objectif de montrer qu’à partir d’observables divers quant à leur forme, leur sens et leur référence, il est possible de proposer une analyse unificatrice de cette diversité, sans recourir à la notion fourre-tout d’idiosyncrasie. Il suffit pour cela :

- de désintriquer, dans le sens lexicalisé des dérivés en *-et(te)* du français, ce qui ressortit à l’application de ce suffixe et ce qui ressortit à l’application d’autres opérations linguistiques, de nature constructionnelle (j’écrivais *dérivationnelle* à l’époque) ou de nature proprement sémantique ;
- de recourir aux mécanismes formels mis à disposition par le modèle pour expliquer les cas de distorsion entre une forme attendue et une forme attestée.

La démonstration est menée à partir du corps de propositions suivant, élevé au rang d’axiomatique (on peut lui en faire reproche). Je le reprends ici tel qu’il figure dans mon ouvrage (pp. 79-89) <sup>2</sup> :

P<sub>1</sub> : L’hétérogénéité des rôles assumables par *-et(te)* est réductible à condition de dépasser le strict niveau de l’évidence observable.

P<sub>2</sub> : La grammaire du français ne comprend qu’un suffixe *-et(te)* mono-instructionnel.

P<sub>3</sub> : Le suffixe *-et(te)* est fondamentalement un marqueur d’appropriation.

P<sub>3,1</sub> : Tout dérivé issu de la suffixation par *-et(te)* possède les traits classifiants immédiats de sa base.

P<sub>3,1'</sub> : Tout dérivé issu de la suffixation par *-et(te)* appartient à la catégorie lexicale de sa base.

P<sub>3,2</sub> et P<sub>3,2'</sub> : Les dérivés ne possédant pas les traits classifiants immédiats de leur base apparente (P<sub>3,2</sub>) ou n’appartenant pas à la catégorie lexicale de cette dernière (P<sub>3,2'</sub>) sont les produits de l’application ordonnée du suffixe *-et(te)* et d’une (de plusieurs) autre(s) opération(s).

Il n’y a pas grand-chose à dire de la proposition P<sub>1</sub>, qui constitue davantage une posture théorique de principe qu’une véritable proposition, sauf à remarquer qu’elle reprend, sans le remettre en cause, le dépassement de l’évidence en matière d’analyse de mots construits que préconise le modèle de Danielle Corbin, ni, dans le fond, de P<sub>2</sub> qui mène à terme la proposition précédente. La proposition P<sub>3</sub> et les propositions liées sont, elles, plus intéressantes, parce qu’elles contraignent les analyses proposées dans l’ouvrage. Je ne discuterai pas ici du rapport de cause à effet que j’avais cru voir à l’époque entre P<sub>3</sub> et les propositions connexes. [6], que je présenterai rapidement au § 2.1.2.2., remet en cause ces propositions connexes pour les noms déverbaux sans remettre en cause P<sub>3</sub>.

---

2. Fradin (2001 : 87, n. 2) remarque que les propositions P<sub>3,1</sub> et P<sub>3,1'</sub> « rappellent fortement les contraintes de la Théorie de l’Optimalité en phonologie [...], telles que : minimiser l’écart entre l’input et l’output ». Ce rappel, s’il est exact, était cependant fortuit, cette théorie étant encore confidentielle quand a été rédigé l’ouvrage (en tout cas, je ne la connaissais pas à l’époque).

Je ne donne pas dans [0] de définition explicite de la notion d'appropriation, mais tire argument de l'abondance de noms en *-et(te)* dans le discours publicitaire (par ex. *Craquinettes, Danette, Viizirette*) pour asseoir mon hypothèse (p. 80 *sqq.*), l'idée étant que la périphérie peut dire quelque chose du centre<sup>3</sup>. Si le discours publicitaire utilise de façon massive les noms de marque en *-et(te)*, c'est que ce dernier fonctionne comme « stimulus ostensif » (Sperber & Wilson 1986), engageant l'acheteur à s'approprier le produit nommé. Ainsi, quand j'écris dans [0] que *-et(te)* est un marqueur d'appropriation, j'entends signifier que ce suffixe indique que le locuteur peut s'approprier ce à quoi réfère le dérivé. Ma définition, implicite, rejoint donc celle, explicite, de Fradin (2003b : 56), elle-même inspirée de Dressler & Merlini-Barbaresi (1994) :

[O]n veut dire par là qu'en employant un diminutif, le locuteur signifie que ce dont il parle est dans la sphère de ce qui lui est familier ou intime, ou relève d'un espace qu'il contrôle sans effort ou dans lequel il n'a ni à exhiber ses défenses ni à faire bonne figure.

à ceci près que, dans [0], je considère que la diminution est un cas d'espèce d'appropriation, et parle d'« appropriation physique ». Un locuteur peut en effet plus facilement s'approprier un objet petit, ou un objet plus petit que ne l'est le référent de sa base. Je ne discuterai pas ici de cette option, qui importe finalement peu pour mon présent propos.

Pour valider P<sub>3</sub>, et en me limitant aux noms, il faut donc réussir à montrer que, dans des lexèmes construits aussi divers que ceux sous (1)<sup>4</sup>, le rôle de *-et(te)* se limite à un marquage d'appropriation.

- (1) a **ciselet** : « Petit ciseau émoussé servant aux bronziers, aux graveurs, aux orfèvres »  
**chevrette** : « 1♦ Petite chèvre. [...] 2♦ (1611) Femelle du chevreuil. [...] »  
**laneret** : « Mâle du lanier »  
**maisonnette** : « Petite maison »<sup>5</sup>  
**opérette** : « Petit opéra-comique dont le sujet et le style, légers et faciles, sont empruntés à la comédie »
- b **baïonnette** : « Arme pointue qui s'ajuste au canon du fusil et que l'on peut retirer à volonté »  
**binette** : « Instrument servant au binage de la terre » (**1. binette**)  
**castorette** : « Peau traitée de manière à évoquer la fourrure du castor »

3. L'idée que des phénomènes langagiers marginaux peuvent nous apprendre quelque chose du fonctionnement central de la langue est au cœur de plusieurs articles de Doleschal & Thornton eds (2000). C'est également le thème des recherches sur l'utilisation de la morphologie constructionnelle dans les noms de marques que mène Anne-Cécile Dubocage, dont j'ai dirigé les mémoires de maîtrise et de DEA, et dont je dirige la thèse actuellement, avec Anne-Marie Berthonneau.

4. Sauf mention contraire, toutes les définitions citées dans cette note de synthèse sont extraites du *PRO1*.

5. On sera peut-être surpris de voir que, dans le *PRO1*, *maisonnette* reçoit la marque « Vieilli », autrement dit qu'il est « encore compréhensible de nos jours, mais [...] ne s'emploie plus naturellement dans la langue parlée courante » (*PRO1*, « Tableau des termes, signes conventionnels et abréviations du dictionnaire »). Or, une requête sur Google en avril 2004 par le mot *maisonnette* (au singulier et au pluriel) ramène 30 630 pages (21 400 + 9 230), ce qui fait beaucoup pour un mot peu utilisé.

**causette** : « Fam. Petite causerie, entretien familial »  
**chardonneret** : « Oiseau chanteur, au plumage coloré (*passériformes*) »  
**chenillette** : « 1♦ Plante dont la gousse enroulée ressemble à une chenille. 2♦ Petit véhicule automobile sur chenilles »  
**cousette** : « Fam. Jeune apprentie couturière »  
**mésangette** : « Piège, cage à trébuchet pour prendre les petits oiseaux »  
**oubliette** : « Général au plur. Cachot où l'on enfermait les personnes condamnées à la prison perpétuelle »  
**pâquerette** : « Petite plante dicotylédone (*composées*), annuelle ou vivace et dont certaines variétés sont appelées *marguerites* »  
**trempe** : « Fam. *Faire trempe* : prendre un bain hâtif sans entrer complètement dans l'eau »

L'entreprise est relativement facile pour des dérivés comme ceux sous (1a) :

- dans *ciselet* et *chevrette* 1, le suffixe joue son rôle central de marqueur de diminution (l'appropriation est donc physique) ;
- on peut hésiter, hors contexte, entre une appropriation physique et une appropriation affective pour *maisonnette* ;
- dans *opérette*, étant donné le sens de la base, l'appropriation est celle d'un objet de culture, plus proche du locuteur que ne l'est l'opéra<sup>6</sup> ;
- je considère dans [0] que la référence à un mâle pour *laneret*, à une femelle pour *chevrette* 2 est due, elle, à une inférence à partir du rôle de *-et(te)* comme indicateur dimensionnel, les femelles d'animaux étant généralement plus petites que les mâles, sauf, précisément, chez les rapaces. Or, le lanier est un aigle<sup>7</sup>.

Elle l'est moins pour des dérivés comme ceux sous (1b), pour lesquels les analyses développées dans [0] sont, pour le coup, à rebours de l'évidence observable. *Contra* ce qu'on pourrait inférer de ces exemples, et *contra* la plupart des analyses concurrentes, je montre en effet :

- que les noms sous (1b) héritent de leur base leur valeur métonymique ou synecdochique (*baïonnette*, *castorette*, *chardonneret*, *chenillette* 2, *mésangette*, *pâquerette*), instrumentale (*binette*), processive (*causette*, *trempe*), métaphorique (*chenillette* 1), agentive (*cousette*), locative (*oubliette*) ;
- que cette valeur héritée peut être le fait d'une opération dérivationnelle, éventuellement non visible dans la forme observable du dérivé, ou de l'application d'un transfert de dénomination.

6. Encore qu'*opérette* semble un emprunt à l'italien *operetta*, via l'allemand *Operette*, d'après le *Robert électronique*, désormais *RE*.

7. D'après les recherches que j'ai faites en ligne, il existe en autourserie et en fauconnerie trois autres noms de mâles en *-et*, tous trois attestés dans le *Trésor de la langue française*, désormais *TLF* :

**mouchet** : « Épervier mâle »

**sacret** : « Mâle du sacre. *C'était la femelle qui portait le nom de sacre, le mâle s'appelait sacret; il n'y a d'autres différences entre eux que dans la grandeur* (BAUDR. *Chasses* 1834) »

**tiercelet** : « Mâle de certains oiseaux de proie, plus petit d'un tiers que la femelle; *en partic.*, épervier mâle, faucon mâle ».

Ainsi, et pour ne prendre que trois exemples (on trouvera dans [24] une justification argumentée du premier d’entre eux) :

– j’attribue la composante métonymique du sens de *chardonneret* (le chardonneret est friand de chardons) à l’application du suffixe *-ier* antérieurement à la suffixation par *-et(te)*. La référence du nom *chardonnier* à un animal friand de chardons n’est pas reconnue par les dictionnaires<sup>8</sup>. Elle est cependant étayée par le paradigme de noms en *-ier* comme *bousier*, *fourmillier*, etc., référant à des animaux se nourrissant de ce à quoi réfère la base ;

– j’attribue la composante métonymique de *baïonnette* à une authentique métonymie, intervenue avant la suffixation par *-et(te)* (les premières baïonnettes ont été fabriquées à Bayonne<sup>9</sup>) ;

– j’impute la valeur instrumentale de *binette* à une étape reconstruite °*binoir*, antérieure à la suffixation par *-et(te)*.

Les analyses qui précèdent font l’objet d’une argumentation serrée dans [0], qu’il ne s’agit pas de reprendre ici. Il existe cependant trois problèmes qu’élude l’ouvrage, alors même qu’ils sont centraux dans l’hypothèse défendue et, plus généralement, dans le cadre théorique dans lequel il se situe<sup>10</sup>. Tous ont partie liée avec le dépassement de l’observable en matière d’analyse du lexique construit : il s’agit de la pertinence du niveau de généralité atteint par le commun dénominateur ‘marquage d’appropriation’ (§ 2.1.2.1.), de la légitimité des stades virtuels (§ 2.1.2.2.), et de l’inaptitude du principe de compositionnalité à rendre compte de certaines données factuelles (§ 2.1.2.3.).

## 2.1.2. Problèmes théoriques et empiriques

### 2.1.2.1. Le marquage d’appropriation : le bon niveau de généralité ?

Sans renoncer aujourd’hui entièrement à l’idée que ce qui unit les différents emplois de *-et(te)* soit un marquage d’appropriation à condition, ce qui reste à faire, qu’on con-

---

8. J’ignore s’il convient de dire qu’elle correspond à une application référentielle du sens construit du nom *chardonnier*, ce dernier étant susceptible de référer à toute catégorie d’entités entretenant une relation pragmatique avec les chardons (il serait donc référentiellement sous-spécifié : c’est l’hypothèse de Corbin & Corbin 1991), ou qu’elle correspond à un lexème virtuel, *lexème* étant ici pris au sens de Fradin (2003a : 258-259), Kerleroux (sous presse), *i.e.* “unité morphologique catégorisée, sous-spécifiée pour la flexion mais entièrement spécifiée pour le sens”.

9. Le dictionnaire de Wartburg enregistre du reste un nom *baïoune*, qu’il donne comme synonyme de *baïonnette*.

10. Il est plus facile pour moi aujourd’hui de pointer mes insuffisances de l’époque, dont une partie m’était apparue déjà quand j’ai rédigé ma thèse, sans que j’ai pu ou su le formuler. Une partie des remarques qui vont suivre doit aux objections faites par Bernard Fradin dans son compte rendu de [0] (*cf.* Fradin 2001), ainsi qu’à la vague de travaux sur *-et(te)* qui en a découlé, qu’ils soient publiés ou destinés à l’être (Fradin 2003b et à paraître ; Fradin, Hathout & Meunier 2003 ; [23]), ou oraux (communication à la journée d’étude sur la productivité organisée dans le cadre de l’opération 5 du GDR 2220 le 12 mai 2003 : *cf.* [30]).

traîne la notion <sup>11</sup>, on peut se demander s'il s'agit là du bon niveau de généralité, ou s'il n'est pas préférable de descendre d'un ou plusieurs niveaux, étant donnée la double observation suivante faite par Bernard Fradin dans plusieurs de ses travaux sur *-et(te)* :

– nulle part [0] ne se pose la question de la productivité du suffixe *-et(te)*, ni de la variabilité éventuelle de sa productivité selon les cas de figure rencontrés. Or, tous les schémas dégageables à partir de l'observable n'ont pas la même productivité au sens de Baayen (1992) <sup>12</sup>. Par exemple, Fradin & al. (2003 : 73) obtiennent une productivité cumulée de 0,0056 (*-et* : 0,0038 ; *-ette* : 0,0018) pour ce qu'ils appellent le type R1 qui, à partir d'un nom d'objet, forme un nom d'objet dont la taille est inférieure à celle de l'objet nommé par la base (ex. *cloche* / *clochette*) ; elle est en revanche de 0,0010 (*-et* : 0,0010 ; *-ette* : 0), donc plus de cinq fois moindre, pour le type R4 permettant de former des noms de jeunes animaux (ex. *porc* / *porcelet*) ;

– au sein d'un même type, on observe des différences de productivité selon que le lexème construit l'est au moyen de *-et* ou de *-ette* : on le voit ci-dessus avec le type R4, qui n'apparaît que dans des hapax masculins dans le corpus sur lequel Fradin & al. (2003) ont fait le calcul ; les noms d'événements (type R9 dans la typologie de Fradin & al.) <sup>13</sup>, en revanche, sont très majoritairement, voire exclusivement des noms en *-ette*, qu'ils soient ou non hapaxiques (par ex. *causette*, *giclette*).

On peut dès lors se demander ce qu'atteint une généralité qui nécessite qu'on redescende d'un ou plusieurs niveaux pour faire de bonnes prédictions sur les patrons productifs, si on admet que la productivité constitue le problème empirique central (Plag 1999 : 1), voire exclusif (Baayen & Lieber 1991 : 801-802 ; Lieber 1992 : 1), des re-

- 
11. Anne-Marie Berthonneau me suggère la notion de 'sphère personnelle', également à l'œuvre dans la possession, indépendamment d'un lien de propriété effective avec le référent du 'possédé' (sur la notion qui constitue un cas d'espèce de 'déixis personnelle', cf. notamment Pamies 2002).
  12. La productivité au sens de Baayen (1992) se calcule en corpus. Elle correspond au quotient du nombre d'hapax porteurs du procédé étudié par le nombre d'occurrences porteuses du procédé dans le corpus considéré. Pour un point sur la notion de productivité, cf. [15] et, plus largement, [41], qui constitue le premier aboutissement d'une réflexion menée depuis septembre 2000 dans le cadre de l'action 5 « Bases de données constructionnelles et productivité des procédés morphologiques en corpus » du GDR 2220 « Description et modélisation en morphologie ». C'est moi qui ai assuré la coordination de cette action.
  13. Pour le type R9, Fradin & al. (2003) parlent de noms d'événement (p. 62). L'étiquette est certainement une commodité, dans la mesure où le premier des deux exemples qu'ils citent, *giclette*, ne satisfait pas les deux critères que Godard & Jayez (1996 : 43) imposent à ce type de noms : (i) pouvoir être tête d'un GN complément d'une préposition temporelle ( ??*pendant la giclette*), (ii) pouvoir entrer dans des constructions nominales de durée ( ??*une giclette de cinq secondes*). *Causette* passe avec plus de succès ces tests, comme le montre l'exemple suivant, relevé sur la Toile en avril 2004 : « Sandrine le maternelle : Miloud pose sa tête sur ses cuisses pendant la causette. Il ne bouge pas. ». Il n'est cependant certainement pas un nom d'événement complexe (Grimshaw 1990), puisqu'il accepte le déterminant indéfini (*une causette*), n'accepte pas l'expression de l'agent du verbe *causer* (aucune occurrence sur la Toile de *\*la causette par X*, où *X* dénomme un agent), peut être pluralisé (ex. relevé sur la Toile : « Tous les jours, c'est le même scénario. Car au sein de la rédaction du journal [...], il y a des chahuts, des causettes en passant par les nombreuses blagues et autres anecdotes »), toutes propriétés refusées par le nom d'événement complexe. Pour un résumé des propriétés des noms d'événement complexe, cf. Fradin (2003a). Pour une mise en œuvre de ces propriétés, cf. Kerleroux (1999). Je renvoie également à l'ouvrage à paraître de Danièle Van de Velde, dont le chapitre 3, consacré aux « nominalisations événementielles », met en évidence de nouvelles propriétés remarquables pour les noms d'événements complexes.

cherches en morphologie. On peut aussi se demander si la généralité proposée dans [0], qui confond *-et* et *-ette* en un seul suffixe, est valide, puisque, d'après la seconde observation, chacun des patrons  $R_i$  dégagés privilégie de façon plus ou moins nette une forme au détriment de l'autre. Fradin (2003b : 54-55) fait l'hypothèse, sans l'explorer plus avant, que les formes *-et* et *-ette* ne sont pas de simples variantes d'un même suffixe. On trouve la même hypothèse dans Temple (2002). La démonstration reste à faire. Bernard Fradin, Nabil Hathout et moi-même nous y employons actuellement pour le chapitre « Morphologie évaluative : la suffixation en *-ette*, *-et* » de l'ouvrage en préparation *Aperçus de morphologie du français* (cf. [23]).

Dit de façon plus abrupte, le dépassement de l'observable en la matière est de l'ordre de l'esthétique, mais ne sert concrètement pas à grand-chose, s'il est dépourvu du moindre pouvoir explicatif et prédictif. On rejoint là la problématique des signifiés de puissance dont l'abstraction est parfois telle qu'ils deviennent infalsifiables : je pense ici au traitement du nom *boîte* dans Cadiot (1994), que je trouve en même temps séduisant et trop abstrait.

### 2.1.2.2. Le problème des stades virtuels

Fradin (2001 : 91) écrit à propos de [0] que « l'une des questions lancinantes que pose ce livre concerne la légitimité des dérivations proposées lorsqu'elles mettent en jeu des formes non attestées ». C'est, de fait, une question que je m'étais posée déjà à l'époque où j'ai rédigé ma thèse, et à laquelle tente de répondre partiellement [6], écrit deux ans avant que Bernard Fradin ne souligne le problème.

Dans [0], et comme le remarque très justement Bernard Fradin, deux principes méthodologiques, non explicités, sont à l'œuvre pour essayer de contraindre les étapes intermédiaires convoquées dans les analyses. C'est particulièrement vrai pour les noms que je refusais à l'époque de construire sur des verbes, mais cela vaut également pour tous les cas de restitution d'étapes intermédiaires, comme pour *chardonneret* cité *supra*. Ces principes peuvent être énoncés ainsi :

– principe d'adéquation sémantique : quand il existe plusieurs étapes intermédiaires envisageables, choisir la plus appropriée sémantiquement :

- même si elle met en jeu une étape non attestée : par ex. *soufflet* que je ne construis pas dans [0] sur le nom *souffleur*, bien qu'il soit attesté, mais sur <sup>o</sup>*souffloir*, que je reconstruis parce qu'il est sémantiquement plus approprié ;
- même si elle pose plus de problèmes formels que la solution concurrente, qu'elle soit attestée (par ex., dans [0], je dérive *poussette* de *poussée*, quand ce nom exprime l'action de pousser, dans le domaine sportif, et non du nom processif *pousse*, parce que, comme *poussée*, *poussette* présente une valeur d'accompli absente de *pousse*<sup>14</sup>), ou qu'elle soit posée comme possible (c'est le cas de *soufflet* ci-dessus).

14. Cette valeur se marque, notamment, par l'utilisation du déterminant indéfini, dans *faire une poussette*, expression que l'on utilise au billard, au tennis de table, en sport automobile ou cycliste.

– principe de prime à l’existant : à adéquation sémantique égale, choisir l’étape mettant en jeu un mot (un sens, une référence) attesté(e) : par ex. *caqueret*, que je construis sur *caqueur* et non sur °*caquoir*, malgré le parallèle que l’on peut être tenté de faire avec *soufflet* ci-dessus.

Sans que [0] le dise explicitement, ces principes méthodologiques tentent, autant que faire se peut, de contraindre les étapes intermédiaires. Mais, comme le note Fradin (2001), ils laissent dans l’ombre le problème du devenir des éventuelles marques morphologiques accompagnant ces étapes, disparues de la forme de surface des dérivés.

Cette question, laissée en suspens dans [0], constitue le fondement de [6], qui revient sur l’inaptitude présumée de *-et(te)* à former des noms déverbaux. Il m’est en effet apparu en 1999 :

- que l’hypothèse défendue dans [0] selon laquelle *-et(te)* est inapte à former des noms déverbaux était entièrement indexée à la proposition P<sub>3,1'</sub> et qu’on pouvait en même temps défendre l’idée que *-et(te)* est un marqueur d’appropriation et qu’il forme des dérivés n’appartenant pas à la catégorie lexicale de leur base. En d’autres termes, il m’est apparu que l’analyse menée dans [0] de noms comme *binette*, *causette*, etc. ne démontrait pas le bien-fondé de la proposition P<sub>3,1'</sub>, mais qu’elle était induite par elle ;
- que les étapes reconstituées dans [0] étaient la plupart du temps *ad hoc* d’un point de vue formel, puisque je n’avais pas toujours d’explication à la disparition ou à l’allomorphie des étapes suffixales intercalaires proposées<sup>15</sup>.

La conclusion de [6] est que les noms apparemment déverbaux sont, de fait, construits sur des verbes, comme le donne à voir leur forme observable. Je maintiens toutefois dans [6] – et continue de penser aujourd’hui – qu’ils héritent leur valeur processive (*causette*), agentive (*cousette*), instrumentale (*sonnette*), locative (*oubliette*), objectale<sup>16</sup> (*sucette*) de leur base. Pour ce faire, m’inspirant de Pustejovsky (1995), je considère en 1999 que *-et(te)* a accès à la structure argumentale des verbes auxquels il s’applique, et au typage de leurs arguments. Pour expliquer les différences observables dans le lexique attesté<sup>17</sup> entre le nombre de noms déverbaux relevant de chacun des sous-types sémantiques distingués (noms d’instruments, noms de procès, etc.), je propose la hiérarchisation suivante dans [6 : 71]<sup>18</sup> :

Quand il opère sur un verbe *V*, le suffixe *-et(te)* sélectionne de façon privilégiée l’argument ‘instrument’ de ce verbe, s’il existe. À défaut, il sélectionne son argument ‘objet’, s’il existe. À défaut, il évalue le procès exprimé par *V*, et construit un nom. À défaut, il sélectionne son argu-

---

15. Le problème se pose différemment selon les bases convoquées : ainsi, la disparition de la finale /e/ de *poussée* devant /Et/ peut être vue comme un cas d’espèce d’haplologie, d’autant plus qu’on la retrouve par ailleurs (par ex. *liberté* / *libertaire*). En revanche, celle de /war/ de °*souffloir*, par exemple, est beaucoup plus difficile à expliquer.

16. Par *objectal*, j’entends ici “qui correspond à ce sur quoi s’applique le procès”. L’adjectif *thématique* (“qui assure le rôle de thème”) me paraissait ambigu.

17. La recherche reste à faire pour les lexèmes hors dictionnaires. La typologie des patrons proposés dans Fradin & al. (2003) et Fradin (2003b), en particulier dans le type L1g, écrase malheureusement ces différences.

18. Les renvois aux pages des documents annexés à la présente note de synthèse correspondent à la contre-pagination que j’ai effectuée, et non à la pagination des documents originaux.

ment ‘lieu’, s’il existe. (La sélection d’un argument de niveau n ne bloque pas nécessairement la sélection d’un argument de niveau n-i.)

Le suffixe *-et(te)* marquerait donc des préférences sémantiques dans les arguments de la structure sémantique des verbes auxquels il s’applique, sans toutefois qu’elles s’excluent mutuellement. On le voit avec *poussette*, qui a une double référence attestée à une action, et à divers objets que l’on pousse (voiture d’enfant ou chariot pour transporter des charges diverses). La préférence, nette dans le lexique attesté, de *-et(te)* pour l’argument ‘instrument’ s’expliquerait, elle, par la proximité entre le marquage de maniabilité que joue *-et(te)* dans ce cas et son rôle central de marquage de diminution. Plus une entité est petite, plus elle est maniable. Je rejoins donc ici ce que j’écrivais dans Dal (1997 : 188), à ceci près que je ne transite plus par une étape en *-oir(e)* intermédiaire :

[L]orsqu’il figure dans un nom à valeur instrumentale, le suffixe *-et(te)* sélectionne de façon privilégiée (mais non exclusive) des bases instrumentales en *-oir(e)*. Cette accointance peut s’expliquer. En effet, dans le cas particulier d’un nom dont le sens comporte une composante instrumentale, le suffixe *-et(te)* exerce son rôle générique de marquage d’appropriation sous la forme d’un marquage de maniabilité.

### **2.1.2.3. Le principe de compositionnalité du sens des mots construits par rapport à leur structure**

Dans [0], j’annonce, dès l’introduction (p. 11), que j’adopterai le principe de la compositionnalité du sens des mots construits par rapport à leur structure, tel que le pose le modèle de Danielle Corbin (*cf. supra*, § 1.1.) :

Je tenterai de rendre compte de façon conjointe de la forme et du sens des mots en *-et(te)*. En effet, que ce soit dans sa version primitive de 1987 ou dans sa version la plus récente [...], le modèle élaboré par D. CORBIN se fonde sur le principe suivant : les règles dérivationnelles construisent conjointement les sens et les structures des mots construits, de sorte qu’à une structure donnée soit associé un sens compositionnel et un seul. Aussi ce principe me guidera-t-il dans la recherche de l’instruction sémantique unique associée à la suffixation par *-et(te)*.

On a déjà vu que ce principe revient comme un leitmotiv dans le modèle de Danielle Corbin. Il est cependant remarquable que, sauf erreur de ma part, cette dernière ne définit nulle part la notion de structure morphologique sur laquelle il repose, ni comment on obtient la structure d’un mot construit donné. Ainsi D. Corbin (1991 : 11) comporte-t-il un paragraphe intitulé « Qu’est-ce que le sens d’un mot construit ? », un autre intitulé « Qu’est-ce que la forme observable d’un mot construit ? », mais aucun qui définit ce qu’est la structure d’un mot construit. Me fondant sur ma propre expérience et sur la compréhension intuitive que j’ai de la notion, je définirai la structure morphologique d’un mot construit donné comme une construction théorique, qui résulte d’une hypothèse faite à partir du sens du mot considéré et des éléments qu’on identifie à partir de

sa forme observable. Ainsi assigne-t-on la structure [ [maison]<sub>N</sub> (ette) ]<sub>N</sub><sup>19</sup> à *maisonnette* parce qu'on y identifie formellement les séquences <maison> et <ette> (/mɛzɔ̃n/ et /et/), et qu'on estime que le sens attesté de ce nom, « petite maison », peut raisonnablement être considéré comme résultant de la combinaison du sens qu'on attribue à chacun des constituants ainsi isolés. Dans un cas simple comme celui-ci, on entre dans la structure par la forme observable, et on utilise le sens comme vérification. Ce cas illustre ce que les psycholinguistes appellent « transparence »<sup>20</sup>.

Il peut cependant y avoir des cas où l'on doute que les constituants repérables dans la forme observable d'un mot construit soient un bon reflet de sa structure. Le doute peut avoir deux causes : soit la forme observable du mot compte un ou plusieurs segments de plus (ou de moins) que ce à quoi on pourrait s'attendre ; soit la structure qu'on obtient à partir de la forme observable du mot, même confortée par le sens, contrevient à des hypothèses qu'on a faites par ailleurs. Ce sont ces deux cas de figure que je vais examiner.

– Hors contexte, pour faire une hypothèse sur la structure d'un mot construit donné, on ne dispose que de deux types de données : sa forme observable et le sens que présente le mot soumis à analyse. Quand la forme observable est jugée non fiable, on utilise donc le sens. Ainsi, si Danielle Corbin considère dans ses travaux que l'adjectif *antigrippal*, « qui permet de lutter contre la grippe », ne résulte pas de l'application de *anti-* à *grippal* ni de celle de *-al* à *antigrippe*, c'est parce qu'elle estime qu'aucune de ces deux possibilités ne rend compte adéquatement du sens de cet adjectif. Dans ce cas, c'est le sens et uniquement le sens qui permet de faire une hypothèse sur la structure. Autrement dit, le sens est alors en même temps un point de départ (on obtient la structure à partir du sens) et un point d'arrivée (le sens se calcule compositionnellement par rapport à la structure), ce qui ne va pas sans poser un problème de circularité. Le grief que je fais ici au principe de compositionnalité tel qu'il est mis en œuvre dans le modèle de Danielle Corbin dans le cas observé est récurrent, et dépasse largement ce cadre. Ainsi Godart-Wendling & al. (1998), dans le panorama historique qu'ils dressent du principe de compositionnalité sémantique, soulignent-ils pareillement que ce principe présuppose la détermination préalable des parties qu'on identifie dans le tout par un moyen qui ne soit pas le sens, tandis que Nazarenko (1998 : 3) et Poirier (1998 : 101) notent à la fois le flou et l'omniprésence de ce principe, dans les études proprement linguistiques, et dans celles qui relèvent du traitement automatique des langues<sup>21</sup> ;

19. La décomposition proposée ici, comme toutes celles qui recourent à des schémas crochetés, se situe clairement dans une morphologie morphématique combinatoire (cf. Fradin 2003a : 141 ; 149), puisqu'elle cherche à segmenter les mots construits en unités atomiques. Dans D. Corbin (1993 : 199), Danielle Corbin envisage que la structure morphologique puisse résulter d'opérations sur des éléments. On note là un glissement vers une morphologie lexématique, selon laquelle les unités complexes résultent de l'application d'une fonction à un lexème (Fradin 2003a : 79), quoique les tenants durs de cette approche récuse l'idée même de structure, sauf dans le cas des composés (cf. Anderson 1992 : chap. 10 et 11).

20. Cf. [15 : 203, n. 12] : « D'un point de vue sémantique, la notion de transparence, surtout utilisée par les psycholinguistes, correspond à celle de (sentiment de) compositionnalité : un mot cesserait d'être transparent – deviendrait opaque –, dès qu'on ne percevrait plus de rapport entre sa forme et son sens, même si, au moment de sa construction, ce rapport existait [...] ».

21. Pour un point sur le principe de compositionnalité sémantique en général, je renvoie à Nazarenko éd. (1998). Plusieurs articles de Booij & al. eds (2000) traitent de son utilisation dans le domaine du lexique construit : Hoeksema (2000) lui est entièrement consacré, mais on trouve aussi des mentions

– Le nom *séparatiste*, déjà cité au § 1.1. au titre de la troncation, me servira à illustrer le second cas évoqué. On a vu que Danielle Corbin le construit sur le nom *séparation*<sup>22</sup>. Or, il n'existe ni d'opposition formelle majeure à le construire formellement sur le radical /separat/ de *séparer*, ni d'obstacle sémantique majeur à le construire sémantiquement sur ce verbe, puisqu'on peut considérer qu'à leur différence catégorielle près, *séparer* et *séparation* expriment le même procès. Si elle le construit sur un nom, c'est à cause des hypothèses qu'elle fait par ailleurs sur la suffixation par *-iste*, réputée s'appliquer à des noms, dont des noms eux-mêmes suffixés par *-ation* (le *RE* en donne 68 : *intégrationniste*, *sécessionniste*, etc.), alors qu'il ne semble pas pouvoir s'appliquer à des verbes. La structure morphologique assignée à *séparatiste* n'est donc que partiellement fondée sur sa forme et sur son sens observables<sup>23</sup>. Le choix de cette structure, et non de celle qui transparaît de sa forme, relève de ce que j'appellerai "l'économie générale du système", économie dont on peut du reste se demander par rapport à quoi elle est réalisée. Je laisse ce point en suspens, et y reviendrai dans le § 2.4.

Ces précisions étant faites, admettons, malgré ce qui vient d'être dit, le principe de compositionnalité tel que le définit Danielle Corbin, et revenons-en au suffixe *-et(te)*. Ce principe nous dit que le sens des mots construits est compositionnel par rapport à leur structure, par conséquent, qu'il doit l'être par rapport au sens de la base (D. Corbin 1993).

De fait, on peut admettre que c'est le cas pour des noms comme ceux sous (1a) (*ciselle*, *chevrette*, *lanerette*, *maisonnette*, *opérette*) auxquels, pour faire bonne mesure, j'ajouterai les noms sous (2), accompagnés de la définition que leur donne le *PRO1* :

- (2) **balayette** : « Petit balai à manche court. — Spécialt Ce balai, utilisé pour nettoyer la cuvette des W.-C. »  
**camionnette** : « Véhicule automobile utilitaire de faible tonnage »  
**clochette** : « Petite cloche »  
**coffret** : « Petit coffre »  
**jardinet** : « Petit jardin »  
**merlette** : « Femelle du merle »  
**tonnelet** : « Petit tonneau, petit fût »

En effet, il n'y a aucune raison de penser que la forme observable de ces mots ne reflète pas fidèlement leur structure. Il est donc concevable que le sens dérivationnellement construit de *coffret*, par exemple, soit compositionnel par rapport aux séquences <coffr> et <et> (/kɔfr/ et /ɛ/) que donne à voir sa forme observable, et qu'il en aille de même pour *balayette* par rapport à <balai> et <ette> (/balɛ/ et /ɛt/, modulo le yod). Dit

---

de ce principe dans Zwanenburg (2000), Iacobini (2000), Lieber & Mugdan (2000), pour n'en citer que quelques-uns.

22. Sauf erreur, elle n'envisage pas qu'il puisse s'agir d'un emprunt à l'anglais *separatist*, alors que d'après le *RE* et le *TLF*, le premier emploi connu du mot date de 1650, « en parlant d'une secte anglaise » (*RE*), et que *separatist* apparaît en anglais en 1608.  
23. Je renvoie à D. Corbin (1992) pour un cas exemplaire de confrontation de données pour parvenir à dégager la structure morphologique des verbes français *déchiffonner* et *décloisonner*.

plus simplement, un coffret est bien un petit coffre, une balayette est bien un petit balai (ce qui peut impliquer qu'il n'en soit plus un : un balai à manche court ne remplit plus les mêmes fonctions que le balai prototypique ; la même remarque vaut par exemple pour *camionnette*<sup>24</sup>). Le rôle de *-et(te)* peut être décrit dans ce cas comme servant à mettre en évidence un écart par rapport au prototype associé au référent du lexème base. Cette hypothèse figure explicitement dans Dal (1997 : 19) pour des cas comme ceux sous (1a) et (2) :

Comme dans le cas de *brémotte* ou de *ciselet*, une façon de décrire le rôle de *-et(te)* dans *bi-quet(te)* ou *vachette* consiste à poser qu'il sert à exprimer la perception d'un écart par rapport au prototype-meilleur représentant de la catégorie extralinguistique que dénomme la base.

On la retrouve, formulée autrement, dans Fradin (2003b : 65) :

Dans tous ces cas, la suffixation en *-ET* indique que le degré d'une dimension  $DMS_i$  que possède le référent du lexème-dérivé est inférieur au degré que présente l'objet de référence [...], en l'occurrence le référent du N-base (objet ou événement).

Dans de nombreux cas cependant, ce principe de compositionnalité du sens des mots construits par rapport à leur base est pris en défaut, indépendamment du problème de la détermination de leur structure<sup>25</sup> :

- (3) a **alouette** : « Petit oiseau à plumage gris ou brunâtre (*passériformes*) »  
**arabette** : « Genre de plante dicotylédone (*Cruciféracées*), scientifiquement appelée *arabis*, petite plante vivace, très florifère, dont on cultive plusieurs espèces dans les jardins » (*RE*)
- b **binette** : « Instrument servant au binage de la terre » (**1. binette**)  
**chardonneret** : « Oiseau chanteur, au plumage coloré (*passériformes*) »  
**sonnette** : « Petit instrument métallique qui sonne pour avertir »

En (3a), bases et dérivés sont coréférentiels, ou l'étaient pour *alouette* avant que ne disparaisse le nom *aloe*, et *-et(te)* sert à mettre en évidence une propriété du référent de

24. C'est la raison pour laquelle, dans [0],  $P_{3,1}$  n'impose pas aux dérivés en *-et(te)* d'être des hyponymes stricts de leur base, la diminution pouvant impliquer un changement de catégorie référentielle. L'héritage par le mot en *-et(te)* des traits classifiants de sa base serait le cas par défaut. En cas de contradiction entre l'instruction de *-et(te)* et une propriété définitoire de la base, base et dérivé seraient cohyponymes. C'est le cas de *camion* : « Gros véhicule automobile transportant des marchandises ou des personnes » vs *camionnette* : « Véhicule automobile utilitaire de faible tonnage ».

25. De nouveau, je m'étais aperçue du problème en rédigeant ma thèse et en la retravaillant pour l'ouvrage qui s'en est suivi, mais je ne pouvais pas le faire apparaître, étant donnée l'option théorique que j'avais prise à l'époque. Une partie des problèmes qui suivent est exposée dans [24], qui constitue une première révision de Dal (1997). [24] devait paraître en 1999 dans le numéro 16 de la revue *Lexique*, initialement coordonné par Danielle Corbin. La maladie l'ayant empêchée de mener à terme ce projet, la parution de *Lexique* 16 a été retardée d'année en année, si bien que j'ai préféré retirer [24], qui ne m'a plus semblé d'actualité. Je le joins néanmoins aux publications annexées à la présente note de synthèse, dans sa version datée du 25 juillet 1999 (les seuls changements que j'ai effectués concernent la mise en page). Je remercie Pierre Corbin de m'avoir incidemment remis en mémoire ce travail dont j'avais oublié l'existence même.

la base ou, ce qui revient au même, du dérivé. De fait, si l'on en croit les dictionnaires, *alouette* n'a jamais nommé une variété petite d'aloë, quand le mot *aloë* était encore en usage. Il indique que l'oiseau appelé *aloë* était petit, pour un oiseau. De même, *arabette* ne désigne pas une variété petite d'arabis, mais indique que la plante appelée *arabis* est petite, pour une plante (elle mesure de 2 à 15 centimètres, selon les variétés). Le suffixe *-et(te)* marque alors la diminution par rapport à la classe à laquelle appartient le référent de la base – classe des oiseaux pour *aloë*, classe des plantes pour *arabis* –, si bien qu'il n'est pas possible de considérer que leur sens soit une fonction du sens de la suffixation par *-et(te)* appliquée à leur base. Mon analyse rejoint ici tout à fait celle de Fradin (2003b).

Cet isomorphisme entre forme et référence s'observe dans d'autres lieux de la morphologie constructionnelle, par exemple en français avec les noms *peuplier*, *mangoustanier* (*mangoutier*), *magnolier*, *microcoulier*, *palétuvier*, etc., dans lesquels la finale *-ier* indique que ces lexèmes désignent des arbres sans qu'ils soient pour autant construits (D. Corbin 1991a : 14 ; 1997c : 7 ; 2001a : 50), ou en néerlandais avec des noms comme *Dominicaner* et *Franciscaner*, dans lesquels la finale *-er* marque iconiquement le fait que *Dominicaan* et *Franciscaan* sont des noms de personne (Booij 2002 : 124)<sup>26</sup>. On l'observe également en syntaxe, avec *petit* dans des séquences comme *petite maisonnette* ou *petit nain*, certes pléonastiques, mais largement utilisées si on en croit la Toile (854 occurrences de *petite(s) maisonnette(s)* au 15 avril 2004, et 2 604 de *petit(s) nain(s)*).

Le principe de compositionnalité est d'autant plus inapte à rendre compte de lexèmes comme ceux sous (3a) qu'on remarque en outre que *-et(te)* cumule son rôle de marqueur dimensionnel avec d'autres rôles, ce que ne prévoit pas le modèle, comme le notent, là aussi [24], ainsi que Fradin (2003b) et Fradin & al. (2003) :

– l'alouette est un oiseau présent dans le folklore français (qu'on se souvienne de la gentille alouette que l'on voulait plumer, enfant), ce qui s'explique par le fait qu'on la trouve dans toutes les régions de France<sup>27</sup>. En nommant *alouette* l'oiseau nommé antérieurement *aloë*, on a pu également souhaiter marquer l'appartenance de cet oiseau à l'environnement quotidien du locuteur ;

– l'arabette, encore appelée *corbeille d'argent*, est une plante de culture facile, que l'on trouve également à l'état sauvage dans les prairies<sup>28</sup>. Elle appartient par conséquent elle aussi à notre environnement. Comme je le signale dans [0], *arabette* est en outre une dénomination vernaculaire. Le choix de *-et(te)* est donc motivable de multiples façons, et il n'y aurait aucun sens à choisir une des raisons plutôt qu'une autre.

26. Pour van Marle (1993 : 257), ce phénomène d'« adaptation morphologique » (*morphological adaptation*) doit être décrit en termes de propriétés de formes de surface, et non en termes de règles. Pour une discussion sur l'inaptitude des modèles basés sur règles à rendre compte de ce type de données, cf. *infra*, § 2.4.

27. Cf. [natura2000.environnement.gouv.fr/especes/A246.html](http://natura2000.environnement.gouv.fr/especes/A246.html). Je fais l'hypothèse que c'était déjà le cas quand le nom *alouette* est apparu à la fin du XII<sup>e</sup> siècle.

La corrélation entre les thèmes choisis par les chansons ou comptines animalières et la familiarité des animaux chantés est évidente, si l'on se fonde sur <http://www.momes.net/comptines/comptines-oiseaux.html>. (coucou, poule, coq, hirondelle, canard, hibou, pie, rossignol, moineau).

28. Accessoirement, elle a le privilège d'avoir été le premier végétal dont le génome complet ait été étudié (il a fait l'objet d'une publication dans la revue *Nature* en 2000).

En (3b), le problème est différent, bien que, pour *chardonneret*, on puisse également considérer qu'il y ait cumul de rôles (cf. [24]). Il est en effet impossible de considérer que ces noms sont des diminutifs de leur base, quelle que soit la base dont on fait l'hypothèse, dans la mesure où qui dit diminution dit degré standard par rapport auquel diminuer. Or, il n'y a aucun sens à considérer qu'une propriété dimensionnelle soit associée à l'argument 'instrument' des verbes *biner* et *sonner*, si, comme dans [6], on en dérive *binette* et *sonnette*. Quant au référent de noms ou de sens virtuels posés pour les besoins de la démonstration, ils sont par définition dépourvus de propriétés dimensionnelles prototypiques. C'est le cas pour *°binoir*, *°sonnoir* ainsi que pour *chardonnier* si on accepte sa référence reconstruite à un oiseau friand de chardons, puisque ni *-oir(e)*, ni *-ier* ne disent quoi que ce soit de la dimension des référents des mots qu'ils construisent. Si ces derniers sont sous-spécifiés pour la dimension, le suffixe ne peut pas marquer la diminution par rapport à eux. La seule solution, pour ce type d'exemples, est celle que préconisent Fradin & al. (2003) et Fradin (2003) : *-et(te)* ne sert pas à corrélérer une base et un dérivé, mais à donner une indication sur le dérivé, indépendamment de tout calcul de sens effectué à partir du lexème identifiable en position de base.

Ainsi, ici non plus, le principe de compositionnalité tel que le définit Danielle Corbin, et tel que j'annonce que je l'appliquerai dans [0] sans toujours m'apercevoir que je ne l'applique pas, n'est opératoire.

### 2.1.3. La grammaire du suffixe *-et(te)* : bilan provisoire

Les observations qui précèdent rendent en bonne partie invalide le traitement du suffixe *-et(te)* que j'ai proposé dans [0] ([6] me paraît beaucoup plus d'actualité). Plus généralement, elles montrent que le dépassement de l'observable est peut-être une mesure salutaire en matière d'analyse du lexique construit, à condition toutefois de ne pas contraindre les données de telle sorte qu'elles entrent dans le cadre qu'on s'est fixé au préalable : la théorie doit rendre compte des données, et non l'inverse.

Elles montrent également que le principe d'associativité, sur lequel repose entièrement l'édifice théorique élaboré par Danielle Corbin, n'est pas entièrement adéquat pour rendre compte de l'ensemble des dérivés en *-et(te)*, indépendamment de la structure qu'on leur affecte, ce qui ne va pas sans être paradoxal étant donné la défense même de ce principe que [0] annonçait en introduction. Il reste à déterminer si cette non-conformité d'une partie des lexèmes en *-et(te)* avec ce que Fradin (2003b : 55) appelle un « traitement constructionniste » de la morphologie tient au fait que ce suffixe relève de la morphologie évaluative et, si oui, pourquoi cette dernière échappe au modèle constructionniste.

Je ne récusé cependant pas la méthodologie générale de ce travail, en particulier la minutie avec laquelle j'ai examiné les données. Dans [0], j'annonce que les analyses qui seront proposées dans l'ouvrage seront « délibérément falsifiables » (p. 11). Et, de fait, comme le souligne Fradin (2003b : 57, n. 8), c'est la raison pour laquelle elles donnent prise à la critique. Je maintiens également aujourd'hui qu'il est crucial de faire le départ, dans le sens des lexèmes en *-et(te)*, entre ce qui est du ressort du suffixe et ce qui est

construit indépendamment de lui. Ne pas procéder de la sorte reviendrait à faire assumer à la suffixation par *-et(te)* une gamme d'instructions sémantiques couvrant à peu près tout ce qu'on rencontre en morphologie, et en sémantique lexicale. Par là même, on nierait toute spécificité au suffixe, ce qui n'est pas satisfaisant.

Ainsi, à défaut de décrire de façon définitive la grammaire du suffixe *-et(te)*, [0] aura au moins constitué une étape dans sa mise au jour.

## 2.2. Le projet MorTAL

### 2.2.1. Présentation générale <sup>29</sup>

Le projet MorTAL, acronyme de « MORphologie pour le Traitement Automatique des Langues », a bénéficié d'un financement pour 3 ans de la part du ministère de l'éducation nationale, de la recherche et de la technologie français, au titre des « Actions concertées incitatives ». Il a officiellement commencé le 1<sup>e</sup> janvier 2000 pour se terminer, tout aussi officiellement, le 31 décembre 2003. Comme cela apparaîtra, nous continuons toutefois à y travailler. Il a été mené en collaboration avec Christian Jacquemin, Nabil Hathout et Fiammetta Namer. J'en ai assuré la coordination, d'un point de vue gestionnaire et scientifique.

L'objectif initial du projet était d'implémenter le modèle de Danielle Corbin pour produire automatiquement une base de données lexicales pour le français contemporain d'environ 70 000 unités lexicales correspondant aux nomenclatures réunies du *TLF* et du *RE*, qui soit utilisable dans des tâches de traitement automatique des langues, puis de tester cette base.

MorTAL constitue l'aboutissement de FRANLEX, « LEXique pour le FRANçais », qui, plus qu'un projet, était une plate-forme de discussion électronique, initiée par Christian Jacquemin et par moi-même, à la suite de notre rencontre au séminaire de morphologie robuste qui s'est tenu en novembre 1996 à Grenoble (*cf.* [1]). Il nous a alors semblé qu'il pouvait être intéressant de faire se rencontrer des chercheurs appartenant au champ de la morphologie théorique et à celui du traitement automatique des langues (désormais TAL). Les uns et les autres travaillent en effet sur le même objet, la langue, et voient en elle un objet susceptible de donner lieu à des calculs. Cependant, leurs angles d'attaque respectifs, dont [19] donne un aperçu, sont apparemment inconciliables : pour les uns, l'étude de la langue constitue une fin en soi, pour les autres, elle est un outil. Il s'ensuit des incompréhensions réciproques. Ainsi, pour un morphologue qui vient du terrain de la théorie, il est difficile de comprendre qu'on puisse acquérir des informations morphologiques automatiquement sans connaissances linguistiques préalables (*cf.* par ex. la constitution d'une base de données morphologiques dans le domaine médical par apprentissage automatique de Grabar & Zweigenbaum 1999 et 2003). L'incompréhension est d'abord totale – du moins l'a-t-elle été pour moi –, puis, une fois qu'on a compris ce que cela signifie, il faut admettre que l'on puisse se dispenser de théorie pour faire de la morphologie. Or, c'est proprement inadmissible pour

---

29. On trouvera des présentations, plus ou moins détaillées, du projet dans [8], [9], [10], [12] et [19].

quelqu'un qui se situe sur le terrain de la théorie, puisque cela revient à reconnaître l'inutilité de son travail. La difficulté, dont je viens de rendre compte ici du point de vue de la théorie, vaut également du point de vue du TAL, pour les raisons strictement opposées. Le dialogue était donc *a priori* difficile. L'utilisation des mêmes termes de métalangage correspondant à des concepts différents ajoute à la difficulté. Ainsi pour le TAL, « <E> / ite », où <E> correspond à une chaîne vide, est-elle une « règle de substitution de suffixes » (Grabar & Zweigenbaum 2003 : 110). Or, si on ignore qu'on appelle *suffixe* en TAL toute chaîne finale de caractères indépendamment de son statut, et que *règle* ne dit rien d'autre que "régularité observable", le dialogue est rompu avant d'avoir commencé. Les mêmes malentendus peuvent naître à partir de l'utilisation du nom *dictionnaire*, compris comme référant aux dictionnaires de langue générale comme le *RE*, le *TLF*, etc., par les linguistes, et comme référant au stock de mots dont dispose en mémoire un système informatique par les talistes.

Vouloir faire dialoguer des chercheurs appartenant à ces deux communautés linguistiques sur le thème de la morphologie constructionnelle avait donc tout d'une gageure. Malgré cela, autour de FRANLEX s'est assez vite constitué un groupe de réflexion électronique réunissant une quinzaine de chercheurs des deux communautés, parmi lesquels je citerai de façon non exhaustive Danielle Corbin, Bernard Fradin, Benoît Habert, Nabil Hathout, Françoise Kerleroux, Fiammetta Namer et Marc Plénat. Il a débouché sur un premier travail en commun (*cf.* [7]), puis sur le projet MorTAL, qui l'a rendu caduc.

Comme nous l'avions constaté dans FRANLEX, le français est particulièrement déficient en bases de données constructionnelles, contrairement à d'autres langues comme l'allemand, l'anglais ou le néerlandais, qui disposent de la base CELEX, mise au point à l'université de Nijmegen, à partir de 1986<sup>30</sup>. Or, comme le notent entre autres Anderson (1992 : chap. 14) et Daille & al. (2002 : 219 *sqq.*), plusieurs domaines d'application du TAL ou liés à lui<sup>31</sup> peuvent tirer bénéfice d'informations sur le lexique construit, indépendamment de l'étiquetage catégoriel des mots inconnus à seule fin de ne pas bloquer une analyse syntaxique. Je citerai à titre d'exemples :

- la recherche d'information, dans laquelle des informations constructionnelles résolvent en partie les problèmes liés aux variantes de termes (Jacquemin & Tzoukerman 1999) : par exemple *élèvent des moutons / élevage de moutons, saison pluvieuse / pluviosité de la saison, substrat synthétique / synthétiser des substrats*, que je reprends aux documents en ligne rédigés par Christian Jacquemin sous <http://www.univ-lille3.fr/www/Recherche/silex/mortal/index.html> ;
- la fouille de textes, encore appelée *analyse de l'information dans les textes*, qui rend possible la structuration et la classification des documents en factorisant de l'information ;

---

30. Cf. <http://www.kun.nl/celex/>

31. On me pardonnera si, du point de vue des spécialistes des différents domaines cités, tous ne relèvent pas à proprement parler du TAL : ainsi la recherche d'information, dont la convergence avec le TAL « n'est pas simple » (Jacquemin 2000 : 328), et qui, jusque très récemment, formait un domaine à part, avec une communauté de chercheurs bien distincte.

- les systèmes écrit-parole (*text-to-speech*) dans lesquels certains choix de prononciation requièrent des informations structurelles (*heure* vs *gageure* en français ; *fathead* vs *father* en anglais) ;
- la compréhension automatique de textes, puisqu'on sait que près d'un tiers des formes non reconnues dans les systèmes TAL sont des lexèmes construits (Froissart & Lallich-Boidin 1996)<sup>32</sup>. Par exemple, l'adjectif *délectable*, que n'attestent ni le *PRO1*, ni le *RE*, ni le *TLF* sauf, pour ce dernier, sous la forme d'une remarque dans l'article **délecter**<sup>33</sup>, est présent sur 28 460 pages de la Toile (la requête a été effectuée *via* Google, et limitée aux pages : France) ;
- la génération automatique de textes, en particulier pour les domaines de spécialité qui utilisent une terminologie recourant massivement à des lexèmes construits ;
- etc.<sup>34</sup>

[19] cherche à mettre en évidence quel bénéfice la recherche d'information et la fouille de textes peuvent tirer d'une ressource comme celle que nous avons produite dans le projet MorTAL. [8] et [9] présentent l'apport des règles constructionnelles implémentées dans le projet pour la compréhension et la génération automatiques des lexèmes construits hors dictionnaires.

Le projet MorTAL a donc d'abord une visée applicative. Lorsque nous l'avons conçu, nous avons en outre fait l'hypothèse qu'il pourrait avoir un intérêt théorique :

- **Premièrement**, si nous réussissions à implémenter le modèle de Danielle Corbin en obtenant de bons résultats, nous garantissons du même coup son adéquation théorique, puisque nous étions contraints de traduire les principes qu'il défend et les traitements de procédés particuliers menés dans ce cadre en des règles formalisées entièrement explicites. Alors que, dans un modèle théorique, le linguiste peut se retrancher derrière les mots en laissant des zones d'ombre ([0] en est la preuve), on ne peut pas user de ce type de subterfuge avec une machine ;
- **Deuxièmement**, nous faisons se rencontrer compétence et performance. En effet, comme la plupart des modèles théoriques, celui de Danielle Corbin s'attache à décrire la compétence dérivationnelle des locuteurs (D. Corbin 1987a : 46 *sqq.*), *i.e.* le savoir intuitif qu'ils ont sur le lexique construit de leur langue indépendamment de sa mise en œuvre. Or, indépendamment du fait qu'on peut se demander si cette compétence n'est pas une fiction permettant de cautionner ce qui fait parfois figure de jeu – par exemple

---

32. Ce chiffre n'a rien d'étonnant. Il y a trente ans, Dubois & Dubois (1971 : 138) estimaient en effet que les mots suffixés représentaient 68,2% des termes enregistrés dans la lettre A du *Petit Larousse*. Ce pourcentage, qui ne veut pas dire grand-chose si on ne se donne pas une définition contrainte préalable de ce qu'on retient au titre des suffixés, donne malgré tout une idée de la part du lexique construit en français.

33. « **Rem.** On rencontre ds la docum. *délectable*, adj. Qui peut être détecté ». Alors que *délectable* n'appartient pas à la nomenclature du *TLF*, ce dernier l'utilise comme terme définitoire sous l'entrée **tube** : « *Tube à rayons cathodiques, tube cathodique.* Tube électronique convertisseur de signal électrique en image dans lequel le faisceau d'électrons émis et modulé est dirigé vers un écran pour produire une image visible ou détectable par d'autres moyens ».

34. Pour une revue plus complète des domaines d'application du TAL, *cf.* Sproat (1992 : chap. 1) et Fuchs éd. (1993 : 13-18).

°*désimmortalisabiliser* et °*institutionnalisationnalisation*, posés comme possibles dans D. Corbin (1987a) –, il est aussi intéressant de se préoccuper de performance quand on travaille en morphologie. S’agissant de °*désimmortalisabiliser*, que Danielle Corbin glose par “rendre désimmortalisable” (p. 790), les résultats mis au jour dans [8] me conduisent à affirmer qu’il est impossible. En effet, moins de 2% des 700 verbes de forme *Xiser* que nous avons relevés dans les principaux dictionnaires actuels de langue générale peuvent être mis en relation avec un adjectif de forme *Xable*. En voici la liste :

- (4) *amabiliser, comptabiliser, culpabiliser, fiabiliser, malléabiliser, navigabiliser, perméabiliser, probabiliser, rentabiliser, respectabiliser, responsabiliser, stabiliser, viabiliser*<sup>35</sup>

Au mieux, trois de ces verbes, *fiabiliser, navigabiliser* et *respectabiliser*, sont interprétables aujourd’hui comme construits sur un adjectif lui-même assurément suffixé par *-able*. Le second d’entre eux, *navigabiliser*, qui est certainement celui pour lequel la reconnaissance d’une base suffixée par *-able* est la plus consensuelle, ne compte curieusement aucune occurrence pertinente sur la Toile (requête faite en avril 2004 avec les formes *navigabiliser, navigabilise, navigabilisent*), ce qui remet en cause sa légitimité. Je fais en effet l’hypothèse, que je dois à ma collaboration avec Fiammetta Namer, Nabil Hathout, Stéphanie Lignon et Ludovic Tanguy dans le cadre du projet WeSConVa (« Web, Suffixation et Concurrence des déVerbaux d’Action ») subventionné par l’ILF (Institut de linguistique française), que ce ne sont pas les dictionnaires qui donnent leur légitimité aux mots, mais l’usage effectif qu’on en fait (cf. [31]). Ainsi, un mot attesté dans les dictionnaires qui n’a aucune occurrence sur la Toile est, à cet égard, suspect. Nous imputons dans [8] cette extrême rareté au fait que les adjectifs que construit la suffixation par *-able* expriment des propriétés latentes, donc des propriétés inhérentes au référent de leurs noms recteurs<sup>36</sup>. Cette caractéristique des adjectifs en *-able* est développée dans [12 : 155], cité ici à partir de la version française de l’article :

Sémantiquement, les adjectifs en *-able* ont ceci de commun qu’ils expriment la possession par le référent de leur nom recteur d’une propriété latente qui, en tant que telle, est supposée valoir indépendamment de sa mise en œuvre effective : une côte peut être dite *abordable* même si elle n’a jamais été abordée ; un homme politique peut être dit *ministrable* même s’il n’a jamais été ministre. Une propriété n’est toutefois latente que pour autant qu’elle est susceptible de s’actualiser : on prédit ainsi que les propriétés qu’expriment les adjectifs en *-able* doivent mettre en jeu à un certain niveau la réalisation d’un procès permettant de les actualiser.

Le procès permettant d’actualiser la propriété décrite par l’adjectif en *-able* peut être exprimé par sa base (*abordable*), ou être récupéré dans la rubrique sémantique de cette dernière (*ministrable* : la rubrique sémantique de *ministre* doit contenir une information disant que, pour devenir ministre, il faut qu’il y ait nomination préalable). Or, dans les verbes en *Xiser* attestés dérivés d’adjectifs (par ex. *adverbialiser, africaniser, banaliser*,

35. Le *TLF* et le *RE* dérivent tous deux *viabiliser* du radical de *viabilité*.

36. Dans [8], nous parlons de propriété endogène. Le terme *inhérent* est meilleur, parce que la propriété a besoin d’une cause extérieure pour être actualisée. Les propriétés exprimées par les adjectifs en *-able* sont donc à la fois marquées [+inhérent] et [+exogène] : cf. *infra* § 2.2.2.

*catholiciser*, etc.), la base adjectivale précise le caractère attribué au référent du thème du verbe à l'issue du procès qu'elle exprime ce dernier. Il s'ensuit qu'elle doit pouvoir exprimer une propriété acquérable<sup>37</sup>, ce qui est incompatible avec le sens qu'affecte le suffixe *-able* aux adjectifs qu'il construit.

La pénurie, dans le lexique attesté, de verbes de forme *Xabiliser* où *Xable* exprime une propriété latente serait ainsi due à une incompatibilité sémantique entre le sens construit par la suffixation par *-able* et les contraintes que fait peser *-is-* envers ses propres bases. Elle est corroborée par la recherche en ligne que nous avons faite pour [8]. À l'automne 1999, nous avons en effet interrogé automatiquement le moteur de recherche yahoo.fr à l'aide de 1287 verbes de forme *Xabiliser* inventés pour les besoins de la démonstration au moyen du système GéDériF, créé par Fiammetta Namer dans le cadre du projet MorTAL. Par exemple :

- (5) *abordabiliser, cassabiliser, cicatrisabiliser, commercialisabiliser, copiabiliser, digitalisabiliser, domesticabiliser, dosabiliser*

Seules huit de ces 1287 créations, soit approximativement 0,5%, ont eu à l'époque des résultats positifs, alors même que l'adjectif en *-able* correspondant était soit attesté dans les dictionnaires, soit largement présent sur la Toile (le nombre entre parenthèses donne le nombre de pages web comportant le dérivé à l'époque où la requête a été faite) :

- (6) *commutabiliser* (238), *instabiliser* (2), *notabiliser* (1), *portabiliser* (1), *potabiliser* (7), *sociabiliser* (12), *variabiliser* (1), *vulnérabiliser* (2)

On remarque en outre que trois seulement de ces verbes sont construits sur un adjectif assurément suffixé par *-able* : *commutabiliser* (< *commutable* < *commuter*), *portabiliser* (< *portable* < *porter*), *variabiliser* (< *variable* < *varier*), et que deux d'entre eux, *portabiliser* et *variabiliser*, sont hapaxiques sur la Toile.

Ainsi, si tant est que la différence entre compétence et performance ne soit pas à abolir, la performance, ici représentée par les attestations sur la Toile, est venue infléchir des résultats énoncés au nom de la compétence des locuteurs. Cette expérience n'aurait pas pu être menée à grande échelle sans les outils TAL développés dans MorTAL ;

– **Troisièmement**, nous étions contraints de rendre compte de l'ensemble des lexèmes construits *a priori* porteurs d'un procédé particulier, sans en exclure d'entrée de jeu. En morphologie théorique, en effet, quand on analyse un procédé particulier, on travaille en général sur un ensemble jugé représentatif à partir duquel sont formulées des hypothèses. Il peut cependant arriver que, malgré soi, on exclue d'emblée des données, ju-

---

37. Cette hypothèse s'oppose à celle de Fradin (2003a : 296), pour qui la non-attestation de °*virginiser*, qu'il pose comme possible, tient au fait que ce que décrit la base exprime un état premier. Pour lui, ce verbe n'existe pas « pour des raisons conceptuelles ou pragmatiques », mais il est linguistiquement disponible (p. 297). Pour ma part, je dirais plutôt qu'il serait possible dans un monde contre-factuel, dans lequel la propriété qu'exprime *vierge* ne serait pas une propriété inhérente, mais que, dans notre monde, il n'est pas possible.

Nous exploitons cette incompatibilité sémantique entre les bases indiquant des propriétés inhérentes et la suffixation par *-is-* dans [14] pour montrer que cette dernière n'équivaut pas au verbe causatif *rendre* (cf. *infra*, « Conclusion »).

gées hors champ, parce qu'on aborde rarement une étude, entièrement vierge de tout préjugé. Or, comme le note très justement Franckel (1997 : 49), cette façon de faire peut pervertir les résultats obtenus. De nouveau, la machine ne permet pas un tel faux-fuyant, puisque, pour elle, tous les verbes commençant par la suite graphique <dé>, par exemple, sont potentiellement des verbes construits au moyen du préfixe *dé-*. Pour alimenter la base, il nous a donc fallu faire des choix explicites et motivés pour chaque unité lexicale rencontrée, que nous la gardions ou que nous la rejetions.

La première originalité du projet MorTAL était donc de combler le manque en matière de bases de données constructionnelles du français souligné plus haut, en faisant travailler ensemble linguistes et talistes. Sa seconde originalité était de chercher à implémenter le modèle de Danielle Corbin pour construire cette base. En effet, les rares systèmes traitant d'informations constructionnelles requièrent le plus souvent peu, voire pas, de connaissances linguistiques, et les règles qu'ils mettent en œuvre sont apprises automatiquement à partir des données qu'ils traitent, et non posées en amont. Pour un état des lieux sur ces systèmes et les méthodes qu'ils utilisent, je renvoie à [8] et à [9], ainsi qu'à Sproat (1992 : chap. 3) et Daille & al. (2002).

Dans les faits, nous avons construit notre base de données avec un seul système, appelé DériF ("Dérivation du Français"), mis au point par Fiammetta Namer à partir des descriptions linguistiques des procédés constructionnels que je lui soumettais. Nous avions initialement prévu un second système, appelé DéCor ("Dérivation pour les Corpus"), mis au point par Nabil Hathout. Ce second système, dont on trouve une présentation détaillée dans [12], ne requiert pas de connaissances linguistiques préalables. Après une période de tâtonnement, nous avons estimé que la gestion de deux systèmes aussi antithétiques pour produire une même ressource serait trop difficile, et nous avons renoncé à faire produire des données à DéCor, sauf à titre de comparaison avec celles produites par DériF<sup>38</sup>. La coexistence de ces deux systèmes posait en outre le problème de l'hétérogénéité des données produites, puisque DéCor ne fournit que des appariements des lexèmes analysés avec une base présumée.

Ainsi, après cette phase de tâtonnement, le rôle de chacun des partenaires du projet s'est stabilisé :

– j'essayais d'établir une grammaire minimale de chacun des procédés que nous avons décidé d'implémenter. J'en faisais part à Fiammetta Namer, nous en discutons par

---

38. Les résultats de DéCor ont été comparés avec ceux de DériF que nous avons considérés comme un étalon. Pour chaque suffixe, deux valeurs ont été calculées : le rappel (*i.e.* le rapport entre le nombre de résultats corrects de DéCor et le nombre total de résultats produits par DériF), la précision (*i.e.* le rapport entre le nombre de résultats corrects de DéCor et le nombre total de ses résultats). Il ressort des comparaisons effectuées que la qualité linguistique des appariements produits par DéCor dépend fortement de l'homogénéité de l'ensemble des bases potentielles. Ainsi, pour le suffixe *-able*, les résultats sont relativement bons (rappel et précision sont supérieurs à 95%), car les bases sont en très grande majorité des verbes, formellement différenciables des noms. Pour le suffixe *-ité*, ils le sont un peu moins (rappel et précision sont de l'ordre de 90%), car, même si les bases sont majoritairement des adjectifs, les formes de ces derniers sont assez proches des formes nominales. Enfin, les résultats sont insuffisants pour le suffixe *-is-* (rappel et précision sont de l'ordre de 50%) du fait de la répartition plus équilibrée entre les bases adjectivales et nominales que l'on peut difficilement départager sans informations linguistiques.

courriel, et elle traduisait ces hypothèses linguistiques sous forme de règles informatiques, qu'elle appliquait ensuite sur le corpus de référence, constitué, je le rappelle, des nomenclatures du *TLF* et du *RE*. Elle me retournait les résultats de son implémentation, je vérifiais manuellement les données produites, et elle corrigeait ses règles en fonction de mes remarques. Les cas jugés récalcitrants étaient traités par liste d'exceptions, pas toujours parce qu'il s'agissait à proprement parler d'exceptions, mais parce que leur traitement informatique était jugé trop coûteux eu égard au nombre de lexèmes concernés, ou qu'il était tout simplement impossible. C'est l'option que nous avons prise pour, par exemple, les adjectifs en *-able* dérivés de noms, qui représentent moins de 2% des dérivés en *-able* attestés dans le *TLF* et le *RE*. Nous avons ainsi traité par liste d'exceptions les quatre adjectifs *cardinalable*, *ministrable*, *papable* et *présidentiable*, dans lesquels la base désigne un personnage social qui n'acquiert sa fonction qu'au terme d'une élection ou d'une nomination. Nous pressentions que le schéma qu'ils instancient était productif – cette intuition est confirmée par Hathout & al. (2003) qui ont recueilli sur la Toile (*candidat*) *députable*, (*étudiante*) *doctorable*, (*premier danseur*) *étoilable*, (*haut fonctionnaire*) *préfectable*, etc. –, mais nous n'avions pas les moyens de les traiter automatiquement, puisque nous ne disposions pas d'informations sémantiques sur les bases (cf. § 2.2.2.).

Nous avons procédé de la sorte à l'implémentation complète des suffixes *-able*, *-ité*, *-ifi-*, *-is-*, des préfixes *dé-* et *re-*, et entamé celle du préfixe *in-*. Fiammetta Namer a poursuivi seule l'implémentation de ce dernier préfixe, et a enchaîné avec le suffixe *-eur* et la conversion. À ce jour, notre base compte quelque 18 000 unités lexicales. Le tableau ci-dessous, repris au rapport de fin de contrat envoyé au ministère le 25 janvier 2004, liste les procédés traités, et précise entre parenthèses le nombre de lexèmes concernés :

<i>dé-</i> (1819)	<i>in-</i> (873)	<i>re-</i> (1062)	autres préfixes (208)	
<i>-able</i> (4117)	<i>-ifi-</i> (244)	<i>-is-</i> (2152)	<i>-ité</i> (3227)	<i>-eur</i> (3520)
conversion A→V (409)	conversion N→V (1050)	conversion A→N (295)	conversion N→A (5)	

Tableau 1. Etat d'avancement de la base de données au 25 janvier 2004.

À la date du 25 janvier 2004, DériF était en outre en mesure d'effectuer l'analyse d'un millier de noms composés issus de la 10<sup>e</sup> version de la classification internationale des maladies (cf. Namer 2003a).

– Christian Jacquemin et Nabil Hathout étaient chargés de l'évaluation de la base. Voici ce que nous en avons écrit dans notre rapport de fin de contrat :

Notre objectif de réaliser une évaluation des données dans des tâches réelles de Recherche d'Information ou de Fouille de Textes de façon à apprécier le gain obtenu grâce à notre base n'a été atteint qu'en partie. Ch. Jacquemin a utilisé les résultats de notre analyseur en recherche d'information, dans des tâches de recherche de variantes (cf. le rapport de cette recherche en ligne sous <http://www.univ-lille3.fr/www/recherche/silex/dal/index.html> où on trouve aussi un outil d'interrogation permettant d'interroger linguistiquement un échantillon de notre base). Nous avons également cherché à réaliser une évaluation de la base de données MorTAL par des

utilisateurs potentiels extérieurs disposant de moteurs de recherche. Une première collaboration avec la société Lexiquest n'a pas pu être menée à terme. Nous avons en effet défini avec Gaëlle Recourcé et Hugues de Mazancourt un protocole d'évaluation de l'apport de la base de données MorTAL pour l'expansion de requêtes. Les expériences prévues n'ont pas pu être réalisées du fait des difficultés que cette société a rencontrées. Suite à cet échec, nous avons approché Mohand Boughanem de l'équipe SGI de l'IRIT (CNRS, UPS & INPT) qui a développé le moteur de recherche Mercure. Nous avons convenu à nouveau d'une collaboration visant à évaluer la base de données pour l'expansion de requêtes. Le corpus et le jeu de tests à utiliser sont ceux de la dernière campagne CLEF. Seules les étapes préliminaires de l'évaluation ont été réalisées, à savoir remplacer les index de formes tronquées habituellement utilisés avec Mercure par des index de lemmes. Cette première expérience, effectuée par Cécile Laffaire, ingénieure d'étude à l'IRIT, a montré que l'utilisation de ces index ne dégrade pas les performances du système. L'étape suivante, qui consiste à étendre les requêtes au moyen de la base de données MorTAL, devrait être réalisée dans le courant de cette année.

Dans la suite de cette présentation, je me limiterai à exposer le contenu de notre base (§ 2.2.2.), et à donner un aperçu des problèmes que nous avons rencontrés (§ 2.2.3.) puisque je n'ai pas participé à la phase d'évaluation du projet. S'agissant du système sur lequel repose la base, DériF, je n'en parlerai que pour autant que ce sera utile à mon propos. On en trouvera une présentation technique, écrite par Fiammetta Namer, dans [12].

## 2.2.2. Présentation de la base de données

Le système DériF prend pour entrée la forme citationnelle des lexèmes qu'il traite (ce qui correspond au *lemme* des talistes), accompagnée d'une étiquette catégorielle restreinte aux trois catégories lexicales dites 'majeures' susceptibles de servir de bases aux procédés constructionnels en français : noms, verbes, et adjectifs<sup>39</sup>. Il ne dispose pas d'informations sémantiques. Les sorties de l'analyseur sont dotées de plusieurs champs, calculés automatiquement :

– le **premier champ** réside en un schéma crocheté et étiqueté, appelé « arbre d'analyse » par Fiammetta Namer. Il représente la structure morphologique de l'unité traitée, et fait apparaître la ou les opérations constructionnelles qui ont présidé à sa construction, pour autant qu'elle(s) ai(en)t été implémentée(s) ;

---

39. Cette assertion est tendanciellement vraie seulement. Je citerai le cas du préfixe *re-*, mentionné dans [20 : 310, n. 28], compatible avec toute base marquant un procès ou un état, ou inscrite dans un scénario supposant un procès ou un état, indépendamment de son appartenance catégorielle. Par exemple, *rebonjour*, qui suppose qu'on ait déjà dit "bonjour" à la même personne, en général dans la même journée, *re carottes râpées*, qui suppose une première occurrence, rapprochée dans le temps, de carottes râpées dans un scénario (ex. trouvé sur la Toile : « matin : 1 café, 60g de pain, beurre allégé ; midi : carottes râpées, endives, poissons vapeur, fromage, yaourt 0% ; dîner : re carottes râpées), *re-promenade*, qui ne peut se dire que si on a fait une promenade préalable (*ibid.* « Cette année c'est un petit déjeuner entre nous, puis promenade et repas de midi au resto, et re-promenade et enfin expo. »), *re pas le moral*, qui indique qu'à peu de temps de là, on n'avait déjà pas le moral (*ibid.* « 21 matin, re pas le moral »). Le préfixe, dans ce cas, prend pour base l'élément saillant du scénario dont le dérivé marque l'itération. Sur le préfixe *re-* français, je renvoie à Franc-  
kel (1997), Jalenques (2000) et (2002) ainsi qu'à Amiot (2002).

– le **second champ** donne la famille constructionnelle de l'entrée décrite, énoncée à rebours, de l'unité la plus complexe à l'unité la moins complexe. Chaque étape est dotée d'une étiquette catégorielle ;

– le **dernier champ** consiste en une glose reflétant le résultat sémantique de l'application de la dernière opération constructionnelle. Cette glose, exprimée en langue (semi-) naturelle, est susceptible d'être réécrite de façon plus formelle, sous forme de fonctions logiques, pour certaines applications (*cf.* [19]). Nous avons choisi de ne pas remonter, dans la glose, au-delà de la dernière opération intervenue, parce que nous avons supposé que cette information suffirait à la plupart des demandes des applications TAL auxquelles se destine la base. Si toutefois l'utilisateur avait besoin d'aller à un niveau de profondeur plus important, il lui suffirait de combiner la glose associée à l'entrée concernée et celle de sa base, et de réitérer l'opération jusqu'à parvenir au primitif de l'unité en entrée.

Voici deux exemples, empruntés à [10], qui illustrent ce qui précède (le contenu de la glose des adjectifs en *-able* sera commenté *infra*) :

```
chavirable ==> [ [ chavirer VERBE] able ADJ] (chavirable/ADJ,
chavirer/VERBE)
::(<Prep> lequel/Que l') on peut chavirer
```

Figure 1. Analyse par DériF de *chavirable*.

```
imperturbable ==> [ in [[ perturber VERBE] able ADJ] ADJ] (im-
perturbable/ADJ, perturbable/ADJ, perturber/VERBE)
::non perturbable
```

Figure 2. Analyse par DériF d'*imperturbable*.

Les entrées de notre base de données ne comportent pas de transcription phonétique, contrairement à CELEX, par exemple, qui donne des indications phonologiques complètes<sup>40</sup>. La raison tient au peu de place qu'occupait la phonologie dans le modèle de Danielle Corbin, et, par conséquent, au peu d'attention que je lui ai portée quand nous avons convenu du format à donner aux entrées de notre base. Elle tient aussi sans doute au fait que, comme le souligne Fradin (1994 : 2) à la suite de Richie & al. (1992), les travaux en TAL se consacrent majoritairement aux textes sous leur forme écrite. Or, il est de plus en plus net que des contraintes phonologiques sont à l'œuvre dans le lexique construit, y compris en français, même si l'accent de mots y est fixe (pour une langue accentuelle comme l'anglais, l'existence de contraintes phonologiques est reconnue depuis longtemps, puisqu'une opération constructionnelle peut s'accompagner d'un changement accentuel). Comme le montrent les recherches menées à l'ERSS sous l'impulsion de Marc Plénat, il existe en effet en français des contraintes de « bonne formation » (Roché 2003) :

40. *Cf.* le détail de la rubrique *phonologie* sous l'URL de CELEX : « Phonology : phonetic transcriptions (using SAM-PA notation or Computer Phonetic Alphabet (CPA) notation) with : syllable boundaries ; primary and secondary stress markers ; consonant-vowel patterns ; number of phonemes/syllables ; alternative pronunciations ; frequency per phonetic syllable within words ».

– contraintes dissimilatives mises au jour il y a plus d’un siècle et “oubliées” jusque récemment, qui expliquent par exemple le peu d’accointance du suffixe *-et(te)* pour des bases terminées par /t/ et les stratégies palliatives mises en place pour éviter cette rencontre (Plénat & Roché 2003)<sup>41</sup>;

– contraintes d’optimalité prosodique des bases ou des dérivés<sup>42</sup>, dont je reparlerai dans quelques lignes ainsi qu’au § 2.2.3.1. à propos du suffixe *-oche*.

La rareté de dérivés de forme *Xisabilité* en corpus que nous avons constatée dans [8] (16 sur les 833 dérivés testés) est également passible d’une explication de ce type. La liste figure sous (7), le nombre entre parenthèses donne le nombre de pages web comportant le dérivé lorsque la requête a été faite (automne 1999) :

- (7) *axiomatisabilité* (2), *décentralisabilité* (2), *délocalisabilité* (1), *don juanisabilité* (54), *factorisabilité* (1), *généralisabilité* (7), *homogénéisabilité* (1), *inutilisabilité* (2), *localisabilité* (1), *parallélisabilité* (2), *réalisabilité* (5), *réutilisabilité* (186), *révisabilité* (2), *stabilisabilité* (9), *universalisabilité* (1), *utilisabilité* (77)

Il ne s’agit pas, comme pour les *Xabiliser* plus haut, d’une incompatibilité sémantique entre bases et dérivés : *-ité* est compatible avec des bases en *-able*, *-able* l’est avec des bases en *-is-* (cf. [8]), mais bien d’un problème de prosodie. Quel que soit le nombre de syllabes que comporte *X*, le dérivé, trop long, s’écarte des contraintes de bonne formation du mot prosodique idéal en français, qui est de deux syllabes (Plénat & Roché 2003 : 285). Dans [8], nous imputons ce constat à un problème de performance. Aujourd’hui, j’irai jusqu’à dire que, plus la forme d’un dérivé s’écarte de cet idéal dissyllabique, moins elle est possible, donc moins elle est grammaticale<sup>43</sup>. Il s’agit donc bien d’un problème de compétence si, de nouveau, le concept est pertinent.

Si l’on se fonde sur le format de sortie des éléments de notre base, l’approche combine à la fois morphologie morphématique combinatoire (premier champ), comme toutes celles qui recourent aux schémas crochetés (cf. *supra* § 2., n. 19), et morphologie lexématique, puisque le deuxième champ fait apparaître les sorties et entrées de chaque opération constructionnelle impliquée dans la formation du lexème. Le contenu de ce

---

41. Si l’on se fonde sur Hasselrot (1957 : 12), l’apparition en latin de l’ancêtre du suffixe *-et(te)* français est, elle aussi, due à un phénomène morphophonologique. Le suffixe *-itt-* était absent en latin, sauf dans quelques cognomina féminins : *Gallitta*, *Jul(l)itta*, *Pollitta*, *Livillitta*, le premier de la liste étant aussi le plus ancien. Or, les suffixes dévolus à l’époque à la formation d’hypocoristiques étaient *-ill-* et *-in-*. *Gallilla* posait le problème de la consécution de deux /l/ dans deux syllabes adjacentes, *gallina* était attesté avec le sens “petite poule”. Aussi aurait-on inventé *gallitta*, qui aurait fait office de modèle pour les bases en /l/, ce qui fait écrire à Bengt Haselrot que « [l]e véritable domaine du suffixe *-itta* en latin classique, ce sont les surnoms dont le radical se termine par *l* ».

42. Il est intéressant de constater cette re-découverte, en phonologie, de principes connus dès le XIX<sup>e</sup> siècle. Ainsi, Durand & Laks (2000 : 34) soulignent pareillement les convergences entre la théorie de l’optimalité et les lois néogrammaticiennes de Grammont de la fin de XIX<sup>e</sup> siècle.

43. Les contre-exemples à cette affirmation ne manquent pas, en particulier dans les technolèctes (langue médicale, rapports techniques, etc.) : (7) suffit à le montrer. Cependant, sans avoir jamais effectué de recherches en ce sens, je suis convaincue qu’il existe deux types de dérivés : certains sont destinés à être prononcés, d’autres, proprement imprononçables, sont réservés à l’écrit. Il n’y a donc aucune raison pour que ces derniers satisfassent les contraintes du mot prosodique optimal.

deuxième champ distingue notre base des analyseurs qui ne donnent en général qu’une décomposition structurelle (Sproat 1992 : chap. 3). Voici, à titre de comparaison, les informations que comportent deux entrées choisies au hasard dans la base CELEX :

```
73\abnormally\28\C\1\N\N\N\N\Y\abnormal+ly\Ax\SA\N\N\N\#\N\N\ASAA\(((
ab) [A|.A], ((norm) [N], (al) [A|N.] ) [A] ) [A], (ly) [B|A.] ) [B] \N\N\N

1177\alterable\4\C\1\N\N\N\N\Y\alter+able\3x\SA\N\N\N\#\N\N\SA\((alte
r) [V], (able) [A|V.] ) [A] \N\N\N
```

Figure 3. *Abnormally* et *alterable* dans CELEX.

L’énoncé, en clair, des différentes étapes lexématiques dont le lexème analysé est le terme a été exploité par Christian Jacquemin pour constituer automatiquement des familles constructionnelles, entendues ici comme “ensembles de mots partageant un même primitif” (l’algorithme permettant de constituer ces familles figure sous : <http://www.univ-lille3.fr/www/Recherche/silex/mortal/rapport-familles.html>).

Par exemple (les trois premières familles sont empruntées à [19]) :

<b>famille1</b>	code; coder; décodabilité; décodable; décodage; décodé; décodeur
<b>famille2</b>	rapide; rapidement; rapidité
<b>famille3</b>	séquençage; séquence; séquencer; séquenceur
<b>famille4</b>	fraise; fraisage; fraisement; fraiseur

L’acquisition automatique de familles présente en effet des visées applicatives, tout particulièrement en recherche d’information où, par exemple, une requête formulée au moyen du verbe *décoder* doit pouvoir ramener tous les documents contenant ce verbe à une de ses formes fléchies, mais également au moins les noms *décodage*, *décodeur*, *décodabilité* et *code*. Comme nous le signalons dans [19 : 276], le bruit<sup>44</sup> éventuellement engendré par une extension de requête à tous les lexèmes de la famille du terme utilisé pour la requête initiale est limité par le fait que les requêtes comportent en général plusieurs termes, ce qui restreint les possibles. J’emprunte à [19] l’exemple de *fraiseur* (« Ouvrier, ouvrière qui exécute des travaux de fraisage »), et le développe ici.

Admettons qu’une requête contienne ce nom, et qu’elle soit étendue aux lexèmes de sa famille *fraisier*, *fraisage*, *fraisement* et *fraise*. Du fait de la polyréférence du nom *fraise*, elle ramènera vraisemblablement des documents relatifs à la boucherie, où on appelle *fraise* la membrane comestible qui enveloppe les intestins du veau et de l’agneau, à l’horticulture et au maraîchage, sur la culture et la vente des fruits qu’on appelle *fraise*, à la pâtisserie, où on utilise ces mêmes fruits, à l’histoire, à cause de la collerette que portaient hommes et femmes, chez les nobles, au XVI<sup>e</sup> siècle et au début du XVII<sup>e</sup> siècle, à différents métiers manuels, où on recourt à l’outil nommé *fraise*. Si, en revanche, on effectue la requête au moyen des termes *fraiseur* et *ouvrier* et que l’extension de requête conserve le terme *ouvrier*, il est probable qu’on ramènera moins de documents non pertinents. À titre indicatif, j’ai tenté l’expérience sur la Toile en avril 2004, en me limitant aux pages : France, au moyen de la fonction “recherche avancée”

44. On appelle *bruit* en TAL l’ensemble des documents non pertinents ramenés lors d’une requête.

de Google. Une requête par *fraiseur(s)* seul ramène 7 280 pages (je ne tiens pas compte ici des éventuelles redondances). Si on étend la requête au nom *fraise(s)*, on ajoute 300 000 pages. Si, maintenant, on effectue la recherche initiale au moyen des deux termes *ouvrier(s)* et *fraiseur(s)*, on ramène 4 473 pages, tandis qu’une extension au nom *fraise(s)* conservant le terme *ouvrier(s)* ramène 2 269 pages. On observe avec intérêt que, pour cette dernière requête, selon que *fraise* est au singulier ou au pluriel, le domaine d’activité est tendanciellement différent : horticulture et maraîchage pour *ouvrier(s) + fraises*, mécanique pour *ouvrier(s) + fraise* <sup>45</sup>.

Les gloses que nous fournissons dans le troisième champ sont, elles aussi, originales, puisque les analyseurs que nous connaissons ne donnent jamais de glose. On le voit avec la figure 3 ci-dessus. Nos gloses ont été conçues pour être utilisables dans des tâches réelles de TAL. Aussi, plutôt que d’essayer de subsumer d’éventuelles gloses multiples sous une glose unique, trop abstraite pour avoir un quelconque intérêt pratique, avons-nous opté, quand nécessaire, pour des gloses cumulatives. Je renvoie à la glose de *chavirable* mentionnée dans la figure 1, où les deux segments figurant de part et d’autre de la barre oblique (“<Prep> lequel” et “Que l’”, où les chevrons encadrant ‘Prep’ indiquent que la préposition est facultative) sont prévus pour rendre compte à la fois d’adjectifs comme *chavirable* ou *portable*, glosables par “que l’on peut chavirer (porter)”, et d’adjectifs comme *skiable* ou *circulable*, glosables par “sur lequel on peut skier” pour le premier, “sur lequel, dans lequel on peut circuler” pour le second <sup>46</sup>. [19] donne des exemples d’utilisation possible de la glose en recherche d’information et en fouille de textes.

Dernièrement, Fiammetta Namer a ajouté un quatrième champ, recensant, pour chaque étape, l’ensemble des traits sémantiques appris automatiquement à partir des implémentations préalablement effectuées. Les données que produit DériF ont donc désormais l’aspect suivant (la figure 4 ci-dessous reproduit une partie de la figure 1 de Namer 2003a : 33 <sup>47</sup> ; les “xxx” indiquent que d’autres traits pourront être ajoutés, en fonction des futures implémentations) :

- 
45. Toutes les pages ramenées ne sont pas pertinentes : la requête prend par exemple dans ses fichiers plusieurs pages dans lesquelles les ouvriers ramènent leur fraise, et d’autres où ils les sucent.
46. Cf. sur la Toile (i) “sur lequel on peut circuler” : « Du 16 au 27 juin 2003, de 6h à 13h, la largeur circulable sur chaussée fera l’objet d’un rétrécissement sur le couloir bus, *i.e.* face sur laquelle on peut circuler » ; (ii) “dans lequel on peut circuler” : « Ils n’ont pas respecté, non plus, le travail d’hommes qui se lèvent en pleine nuit, samedi et dimanche inclus, pour rendre la ville circulable ».
47. Je remercie Fiammetta Namer de m’avoir donné une version électronique de son article, et de m’avoir autorisée à en reproduire ici des schémas.

<b>1) explicabilité, NOM =&gt; [ [ [ expliquer VERBE] able ADJ] ité NOM] (explicabilité/NOM, explicable/ADJ, expliquer/VERBE) :: "faculté d'être explicable"</b>	
explicabilité	(abstrait, propriété, xxx)
explicable	(xxx, inhérent, exogène, prédicatif)
expliquer	(xxx, transitif, [agent, thème])
<b>2a) introuvable, ADJ =&gt; [ in [ [ trouver VERBE] able ADJ] ADJ] (introuvable/ADJ, trouvable/ADJ, trouver/VERBE) :: "non trouvable"</b>	
trouvable	(xxx, inhérent, exogène, prédicatif)
trouver	(xxx, transitif, [agent, thème])
<b>2b) désossable, ADJ =&gt; [ [ dé [os NOM] VERBE] able ADJ] (désossable/ADJ, désosser/VERBE, os/NOM) :: "que l'on peut désosser"</b>	
désossable	(xxx, inhérent, exogène, prédicatif)
désosser	(xxx, transitif, [agent, thème])

Figure 4. Configuration actuelle des analyses produites par DériF.

Ainsi :

– le marquage ‘abstrait, propriété’ d’*explicabilité* lui vient de la suffixation par *-ité* (‘propriété’ spécifie ici le trait ‘abstrait’, par opposition au binôme ‘abstrait, action’ associé par exemple aux dérivés issus de la suffixation par *-ation*) ;

– le trait ‘inhérent’ des adjectifs en *-able* (1. *explicable*, 2a. *trouvable*, 2b. *désossable*) leur vient de leur sens construit par la suffixation par *-able*, et est corroboré, on l’a vu, par leur inaptitude à être suffixés par *-is-*. Le trait ‘exogène’ qui leur est associé indique que l’actualisation de la propriété qu’ils expriment nécessite l’intervention d’une cause extérieure. Le trait ‘prédicatif’ leur vient de leur aptitude à être suffixés par *-ité*, puisque ce dernier exige des bases prédicatives (cf. Temple 1996 ; Fradin 2003a ; Kerleroux sous presse), comme le montre le contraste entre *la cérébralité d’une peinture* et *\*la cérébralité d’un lobe* ;

– le marquage ‘transitif [agent, thème]’ des verbes *expliquer*, *trouver* et *désosser* est appris automatiquement à partir de leur aptitude à être suffixés par *-able* (on rappelle que nous ne disposons pas de connaissances sémantiques préalables sur les bases : du point de vue du TAL, on part donc de l’effet, ici l’aptitude à être suffixés par *-able*, pour remonter à la cause). Ce marquage nécessite un nettoyage manuel, puisque *-able* peut aussi sélectionner des verbes monovalents dont le procès met en jeu un lieu, même s’il ne se réalise pas nécessairement sous la forme d’un argument obligatoire (*skier (sur)*, *circuler (sur, dans)* ; cf. le « SP loc » dont fait l’hypothèse Fradin 2003a : 279). Le marquage sera également à revoir selon le double résultat suivant dégagé dans cette dernière étude : certains verbes bases d’adjectifs en *-able* ne distribuent pas le rôle d’agent (*cicatrisable*, *pourrissable*) ; plutôt que ‘thème’, le second argument des bases pour les cas prototypiques devrait être marqué ‘-agent, -cible’<sup>48</sup>.

La figure 5, extraite de Namer (2003a : 34), récapitule les traits sémantiques appris automatiquement pour les bases et pour les dérivés, dans l’état actuel d’avancement de

48. On trouve des objections à ce marquage dans Hathout & al. (2003), ainsi que dans [20 : 309, n. 26].

la base de données. Je n'en commenterai pas plus avant le contenu, que l'on doit à Fiammetta Namer :

affixe	base	dérivé
<b>a-, é-, -is(er), -ifi(er)</b>	A = (xxx, acquérable, xxx, predicatif) ( <i>tendre, court</i> )	V = (causatif, transitif, [cause, thème]) ( <i>attendrir, écourter</i> )
<b>dé1-</b>	A = (transitoire, xxx,xxx, predicatif) ( <i>las</i> )	V = (causatif ...)   (résultatif, intransitif, [thème]) ( <i>délasser</i> )
<b>-oir</b>	V = (xxx, xxx,xxx) ( <i>hacher, mourir</i> )	N = (concret, inanimé, lieu/instr) ( <i>hachoir, mouvoir</i> )
<b>-able</b>	V = (xxx, transitif, [agent,thème]) ( <i>laver, encastrer</i> ) V = (xxx, intransitif, [agent, sur(lieu)]) ( <i>skier</i> )	A = (xxx, inhérent, exogène, prédicatif) ( <i>lavable, encastrable, skiable</i> )
<b>-aie</b>	N = (concret, inanimé, végétal) ( <i>bananier</i> )	N = (concret, inanimé, collectif) ( <i>bananeraie</i> )
<b>-aille</b>	N = (concret, xxx, xxx) ( <i>flic, fer, tripe</i> )	N = (concret, xxx, collectif) ( <i>flicaille, ferraille, tripaille</i> )

Figure 5. Acquisition automatique de traits sémantiques (Namer 2003a).

Le recensement systématique des traits sémantiques appris de l'implémentation présente un double intérêt : applicatif pour la génération automatique de termes, par exemple, et théorique. À terme, il devrait en effet faire apparaître l'ensemble des traits sémantiques que manipule la morphologie constructionnelle, en entrée et en sortie. À ma connaissance, aucun recensement de ce type n'a jamais été entrepris. Or, il s'agit là d'un élément pouvant entrer dans le cadre d'une réflexion générale sur les catégories telles que les voit la morphologie constructionnelle (*cf. infra*, § 2.3.2.).

Enfin, comme je l'ai déjà dit, les règles qui ont été implémentées pour analyser les lexèmes de notre corpus de référence se sont révélées également utilisables pour analyser des lexèmes hors dictionnaires, et, de façon plus intéressante encore, pour générer des lexèmes construits, avec un taux de réussite plutôt satisfaisant (*cf. [8 : 441 sqq.]*). Nous avons donc conçu un analyseur et un générateur, ce que nous n'avions pas prévu.

### 2.2.3. Bilan linguistique

Il existe maintes façons de dresser le bilan du projet : par rapport aux objectifs initiaux que nous nous étions fixés, par rapport à la synergie qu'il a créée, par rapport aux publications auxquelles il a donné lieu, par rapport au modèle théorique sur lequel il reposait.

Je passerai très vite sur les trois premières évaluations :

– par rapport à notre objectif initial de constituer une base de données constructionnelles des quelque 70 000 lexèmes attestés dans le *TLF* et le *RE*, notre bilan est plutôt positif, même si notre base a une couverture moindre que celle que nous avons prévue, et même si la phase d'évaluation n'a pas (encore) donné les résultats souhaités ;

– par rapport à la synergie qu'il a créée, de mon point de vue, l'expérience a été riche. Tout linguiste sait qu'un même objet d'étude peut recevoir des éclairages différents, mais, lorsqu'il affirme cela, il pense en général aux éclairages relevant, eux aussi, de la théorie. Or, il est aussi intéressant de quitter le terrain de la théorie, et de confronter un point de vue théorique avec un point de vue applicatif. Cela aide à reconsidérer autrement un objet que l'on pensait familier, et à réviser des vérités que l'on pensait intangibles. En outre, la collaboration établie grâce au projet MorTAL s'est révélée fructueuse puisque Fiammetta Namer, Nabil Hathout et moi-même travaillons actuellement sur d'autres projets ;

– par rapport aux publications auxquelles il a donné lieu, je renvoie à l'annexe 2, qui contient l'ensemble des publications qu'a générées le projet, de façon directe ou indirecte.

Je m'arrêterai un peu plus longuement sur le bilan linguistique, en centrant mes remarques sur les problèmes liés à l'implémentation du modèle de morphologie constructionnelle de Danielle Corbin.

### 2.2.3.1. Phénomènes constructionnels non centraux

Le système sur lequel repose notre base, DériF, a été conçu dans une approche combinatoire de la morphologie. Les analyses qu'il propose reposent en effet sur deux moteurs, l'un dédié aux lexèmes suffixés, l'autre aux lexèmes non suffixés (initialement, ce second moteur était réservé aux lexèmes préfixés, mais il traite désormais en plus des formes converties et composées). Le premier moteur a priorité sur le second jusqu'à ce que l'analyse soit terminée, ou jusqu'à ce qu'elle soit bloquée parce qu'elle parvient à une étape impossible. La figure 6, extraite de Namer (2003a), donne un aperçu du fonctionnement effectif de DériF, quand il doit analyser un lexème comme *indéboulonnable* contenant un préfixe et un suffixe présumés. Se fondant sur la forme de cet adjectif, le système commence par activer la suffixation par *-able* (encadré 'able' dans le schéma) et, de façon conjointe, la préfixation par *in-* quand elle s'applique à un adjectif en *-able* (encadré 'pre\_Xable'). Ce couplage effectué par Fiammetta Namer dans DériF de la suffixation par *-able* et de la préfixation par *in-*, et la priorité donnée à une décomposition structurelle en *in-* + *Xable*, se fonde sur la récurrence constatée d'adjectifs porteurs de ces deux affixes (*inconsommable*, *incontestable*, *incontrôlable*, etc.), et sur le fait que, dans la majorité des cas, le dérivé de forme *inXable* est analysable comme le résultat de la préfixation par *in-* d'un adjectif en *-able*<sup>49</sup>. DériF analyse donc *indéboulon-*

49. Le système est aujourd'hui capable de traiter des adjectifs structurellement ambigus, comme *importable* "que l'on ne peut pas porter" (par ex. sur la Toile : « Le caleçon long à fleurs, importable quelle que soit sa morphologie [...] ») vs "que l'on peut importer" (*ibid.* : « Le paternalisme japo-

*nable* comme construit sur *déboulonnable*, et renvoie ce dernier à la suffixation par *-able*. À son tour, *déboulonnable* est analysé comme construit sur *déboulonner*. Comme aucun suffixe n'y est détecté dans l'état d'avancement du système (ce ne serait pas le cas si *-on* avait été implémenté), *déboulonner* est envoyé au moteur 2, qui le décompose structurellement en *dé-* + *boulon*, ce dernier choix étant motivé par les décisions que nous avons prises quand nous avons implémenté la préfixation par *dé-* (cf. *infra* § 2.2.3.3.). Dans la figure 6, la fenêtre du bas récapitule les analyses effectuées automatiquement :

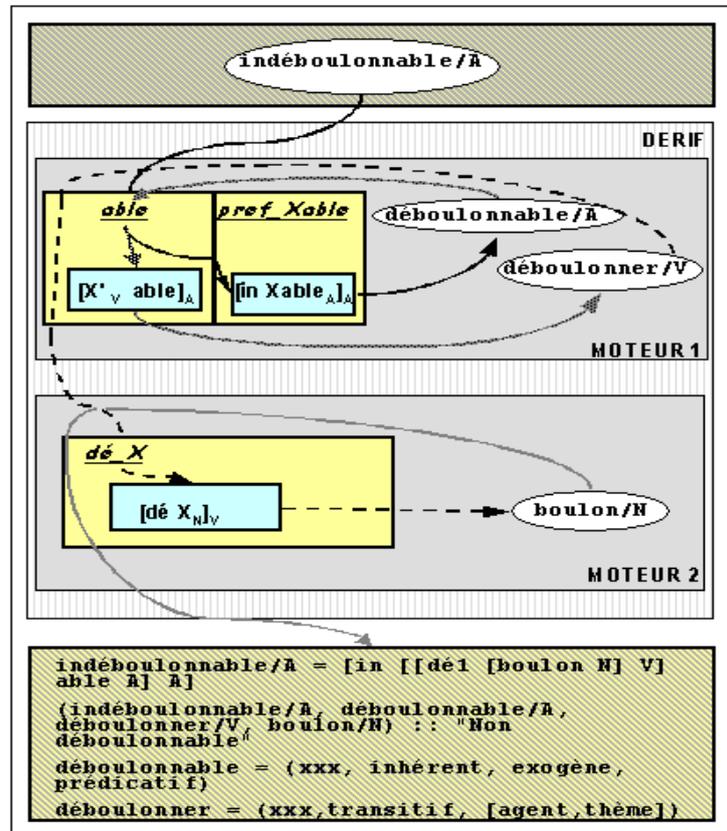


Figure 6. Analyse par DériF d'*indéboulonnable*.

Le mode de fonctionnement choisi pour DériF, dont je viens de donner un aperçu, tient au fait que les opérations constructionnelles prototypiques du français sont de type affixal, et que, bien que D. Corbin (1987 : 101-139) argumente longuement « contre une morphologie de la concaténation », la dérivation affixale fait, elle aussi, office de prototype dans ce modèle.

Il est par conséquent prévisible que DériF rencontre des difficultés avec une morphologie non concaténatoire. Namer (2003a) expose les solutions qu'elle a trouvées pour la conversion et la composition savante du domaine médical. Je renvoie à ce travail,

---

nais est-il importable ? »). Il ne le pouvait pas lorsque nous avons écrit [8]. Pour le détail de la procédure suivie par DériF dans ce cas, je renvoie à Namer (2003a).

puisque je n'y ai participé que superficiellement, par des discussions informelles avec Fiammetta Namer sur la conversion.

En tout état de cause, DériF se trouverait dépourvu face à des phénomènes constructionnels dans lesquels D. Corbin (1997c : 8) voit des « procédés de “déformation” d'unités lexicales », et que Dressler (2000) considère comme extragrammaticaux ou marginaux<sup>50</sup>, comme, entre autres, la reduplication, ou la suffixation en *-oche* dite « substitutive » (Plénat & Roché 2003 : 289), qui produit des dérivés dans lesquels on reconnaît peu, voire pas la base (les exemples sous (8) et (9) sont tous attestés dans notre corpus de référence) :

- (8) *bobonne* (< *bonne*), *chouchou* (< *chou*), *doudou* (< *doux*), *foufou* (< *fou*), *guéguerre* (< *guerre*), *mémère* (< *mère*), *pépère* (< *père*)
- (9) *cantoche* (< *cantine*), *cinoche* (< *cinéma*), *élastoche* (< *élastique*), *fanoche* (< *fâné*), *fastoche* (< *facile*), *patoche* (< *patte*), *péloche* (< *pellicule*), *pistoche* (< *piscine*), *taxoche* (< *taxi*), *téloche* (< *télé(vision)*), *valoche* (< *valise*)

Pour les premiers, on pourrait imaginer que DériF soit enrichi d'une fonction lui permettant de repérer des séquences dupliquées, à partir de leur forme graphique puisque nous ne disposons pas d'informations phonologiques (comme <mé> dans *mémère*, *modulo* le changement d'accent), mais le format des sorties serait inadéquat. La structure de *mémère*, si la notion de structure est pertinente pour des lexèmes comme celui-là, n'est en effet pas représentable au moyen d'un schéma crocheté et étiqueté. La remarque vaut pour les accourcissements (par ex. *prof* pour *professeur*, *métallo* pour *métallurgiste*), dont on rencontre des exemplaires dans les dictionnaires, par ex. dans le *PRO1* :

- (10) *apéro* (< *apéritif*), *beauf* (< *beau-frère*), *bénéf* (< *bénéfice*), *blair* (< *blaireau*), *fac* (< *faculté*), *homo* (< *homosexuel*), *info* (< *information*)

Pour les exemples comme ceux sous (9), le problème n'est pas tant celui de représentabilité de leur structure au moyen d'un schéma crocheté que celui de la reconnaissance automatique de leur base. Si *-oche* tronquait un segment bien identifiable, le problème serait techniquement surmontable. Le problème est que ce n'est pas le cas, et que l'opération de suffixation par *-oche* du français ne doit pas être vue du côté du troncet mais du côté du résultat<sup>51</sup>. Plénat & Roché (2003) la décrivent en effet comme résultant, d'un point de vue prosodique, de contraintes éventuellement contradictoires :

---

50. Pour un inventaire et une présentation commentée de ces phénomènes, je renvoie à Fradin (2003a : 206 *sqq.*). Pour ma part, je conteste l'étiquette *extragrammatical*, puisque cela suppose que chacun s'entende au préalable sur ce qu'est la grammaire. Or, je prends le parti de considérer qu'un schéma utilisé de façon récurrente est nécessairement grammatical. Pour une prise de position analogue, au niveau des lexèmes, cf. [3].

51. C'est aussi le cas pour les accourcissements. On peut d'ailleurs se demander si ce n'est pas toujours le cas, et si, même pour les phénomènes constructionnels centraux, il est nécessaire de parler de troncet en mettant l'accent sur l'élément tronqué.

- contrainte de fidélité entre le nombre de syllabes de la base et le nombre de syllabes du dérivé : dans ce cas, *-oche* se substitue à la dernière rime de la base ;
- contrainte de mise en conformité avec un schème dissyllabique. Dans ce cas, une base longue peut se trouver sensiblement accourcie dans le dérivé, à tel point qu’elle peut ne plus être reconnaissable ; une base monosyllabique, en revanche, est en général reconnaissable. Si elle se termine par une fricative sourde, on observe toutefois l’apparition récurrente d’une épenthèse, par ex. *vache* /vaʃ/ → *vachetoche* /vaʃtɔʃ/ que citent Plénat & Roché (*ibid.* : 290).

Le linguiste est capable de dire que *cantoche* en (9) constitue une version familière de *cantine*, de même que *cinoche* en est une par rapport à *cinéma*, tout simplement parce qu’en tant que locuteur du français, il connaît ces dérivés. Face à *patoche*, pourtant attesté dans le *PRO1*, il aura peut-être un peu plus de mal, s’il n’en connaît pas le sens. En admettant qu’il ait des indices lui permettant de savoir qu’il s’agit d’un nom féminin : est-ce une variante familière de *patte* ou de *pâte*, avec mise au gabarit dissyllabique du dérivé ? De *pâtée*, de *patine*, de *patate*, avec contrainte de fidélité et troncature de la rime finale de la base ? De *pataugeoire*, de *patinoire*, de *pâtisserie*, de *patrouille*, avec accourcissement pour que le dérivé soit un dissyllabe ? D’autre chose ? Une enquête rapide sur la Toile montre que ces interrogations ne sont pas que des élucubrations de linguiste <sup>52</sup> :

**patoche** = “pattes”

Momo: euhh je voulais dire djaffé... pfff.. j`m`emmêle **les patoches** là...

**patoche** = “patate”

Elle est couramment appelée patate ; la côte du Coutançais dît des patardes et on peut entendre, ça et là, mais surtout dans le sud, **des patoches**. ...

**patoche** = “patinoire”

Club de hockey sur glace de gap - les rapaces ... Composition Equipe 1977 et 1978, KH, 2, Oct 14, 2003, 10:36 AM. Tout le monde **à la patoche** samedi

**patoche** = “pâtisserie”

J’ai remarqué aussi qu’avec l’âge j’avais tendance à m’affûter plus facilement, je dois brûler plus facilement les graisses en cours de saison, alors je peux bien me lâcher sur une **patoche** (pâtisserie) de tant à autre, cela ne me pose pas de problème de conscience.

**patoche** = “partition” <sup>53</sup>

Excellent ce site ! C’est terrible de trouver des accords [...], des **patoches** de chansons qu’on n’arrive pas à trouver à l’oreille.

**patoche** = ?

Ne t’inquiète pas je sors **à la patoche** ce soir, ch’ui pas mort (quoique ...)

52. Sur la Toile, *patoche* est aussi le surnom d’un certain nombre de Patrice et de Patrick, et peut également désigner des patins : « ... Blasphèmes ! Que nenni ! Arrière malandrin ! L’ai déjà vu sur des **patoches** l’gamin et il se débrouille ‘achement bien. ... ».

53. Sur la Toile, *patoche* “partition” compte moins d’occurrences que *partoche*, mais il en compte suffisamment pour qu’il ne s’agisse pas d’une erreur. Comme me le suggère Nabil Hathout que je remercie, il est possible que *patoche* soit refait à partir de *partoche*, et non directement à partir de *partition*.

Face à des dérivés comme ceux sous (9), qui sont tous attestés dans notre corpus de référence, une machine se trouve, elle, encore plus démunie : pour *patoche*, par exemple, tout lexème de son référentiel commençant par les suites graphiques <pat>, <pât>, voire <part> (cf. *patoche* “partition”), est une base potentielle.

Dans les faits, nous ne nous sommes pas trouvées confrontées au problème, puisque nous nous en sommes tenues à des procédés de la morphologie conventionnelle. Mais dès qu’on s’attaque à des données s’écarter de la morphologie affixale prototypique, on s’aperçoit des limites d’une morphologie posant que le sens d’un mot construit est compositionnel par rapport à sa structure, soit que la notion de structure est inopérante (réductions, raccourcissements), soit que l’attribution d’une structure nécessite un examen minutieux des données, au-delà de ce que l’on peut demander à une machine.

### 2.2.3.2. Etymologie ou construction en français ?

Un problème lancinant quand on s’intéresse au lexique construit est celui de la frontière entre étymologie et construction en synchronie (D. Corbin 1987a : 96 *sqq.*), et des critères que l’on fixe pour établir cette frontière. S’agissant du français, Huot (1997 : 141) estime que 80% des lexèmes du lexique conventionnel du français viennent du latin. On a vu plus haut que Dubois & Dubois (1971) évaluent, eux, à près de 70% la part des mots suffixés enregistrés sous la lettre A du *Petit Larousse*. Il est donc assez plausible de supposer qu’une bonne part des lexèmes apparemment dotés d’une structure construite dans les dictionnaires est en fait constituée d’héritages du latin.

Le problème se pose quand il s’agit de mesurer la productivité d’un procédé de construction de lexèmes donné. Dans [15 : 211-212], commentant la mesure de la productivité en corpus que propose Harald Baayen (cf. *supra*, § 2., n. 12), je me demande s’il faut retenir les mots hérités au titre de cette mesure. Si l’on décide de les retenir, il faut alors déterminer si tous, qualitativement, ont le même poids vis-à-vis de la productivité selon qu’ils sont analysables comme construits dans la langue considérée ou non.

Le problème s’est également posé à nous dans le cadre du projet MorTAL, tout particulièrement quand nous avons implémenté les suffixes *-able* et *-ité*, puisqu’un pourcentage non négligeable des lexèmes de notre référentiel porteurs de l’un ou l’autre de ces suffixes peut être mis en relation avec un lexème latin, en *-abilis* ou en *-itas, itatis*. Voici, à titre d’exemple, l’analyse que propose le *PROI* des 10 derniers *Xable* (11) et des 10 premiers *Xité* (12) qu’il atteste :

- (11) **construits en français** : *valable* (< *valoir*), *vendable* (< *vendre*), *vérifiable* (< *vérifier*), *véritable* (< *vérité*), *vitriifiable* (< *vitriifié*), *vivable* (< *vivre*), *volatilisable* (< *volatiliser*)  
**hérités du latin (bas latin, latin médiéval)** : *variable* (< *variabilis*), *vénérable* (< *venerabilis*), *vulnérable* (< *vulnerabilis*)
- (12) **construits en français** : *absoluité* (< *absolu*), *absorptivité* (< *absorptif*), *acceptabilité* (< *acceptable*), *accessibilité* (< *accessible*), *acerbité* (< *acerbe*)  
**hérités du latin (bas latin, latin médiéval)** : *absurdité* (< *absurditas*), *acidité* (< *aciditas*), *activité* (< *activitas*), *actualité* (< *actualitas*), *acuité* (< *acuitas*)

Une première réponse consiste à poser que ce qui prévaut n'est pas la construction effective dans la langue considérée mais l'analysabilité dans cette langue. C'est, par défaut ou explicitement, à cette solution que se rangent la plupart des analyses constructionnelles menées sur la période contemporaine<sup>54</sup>. C'est du reste celle qui motive l'analyse de *floral* citée au § 1.1. En effet, si l'on décide de rendre compte de la compétence constructionnelle des locuteurs et non de la culture qu'ils peuvent avoir (ou ne pas avoir) sur le français et son rapport au latin, *acidité* entretient la même relation sémantique et formelle avec *acide* qu'*absoluité* avec *absolu*. D'un point de vue applicatif, on peut également supposer qu'en recherche d'information par exemple, l'expression *l'acidité des terrains* est une variante de *les terrains sont acides*, et qu'une requête au moyen de *humidité* doit pouvoir ramener des documents comportant l'adjectif *humide*, même si *humidité* est un héritage du latin *humiditas*, ou, plus exactement, même s'il est donné comme tel. Sauf à avoir accès au texte comportant la première occurrence du terme, et à pouvoir affirmer que le terme français est une traduction d'un terme latin, on ne peut jamais garantir qu'un lexème d'une langue donnée est un héritage, surtout s'il est conforme au système de la langue cible.

Admettons donc que le critère retenu soit celui de l'analysabilité. On peut de la sorte analyser comme construits en français des adjectifs en *-able* comme ceux sous (13) et des noms en *-ité* comme ceux sous (14), même si, du point de vue de l'étymologie, il s'agit d'héritages du latin, ou qu'on peut penser qu'ils en sont :

(13) *admirable* (< *admirer* / *admirabilis*), *comparable* (< *comparer* / *comparabilis*), *conso-  
lable* (< *consoler* / *consolabilis*), *détestable* (< *détester* / *detestabilis*), *évitable* (< *éviter* /  
*evitabilis*)

(14) *absurdité* (< *absurde* / *absurditas*), *aquosité* (< *aqueux* / *aquositas*), *aridité* (< *aride* /  
*ariditas*), *assiduité* (< *assidu* / *assiduitas*), *atrocité* (< *atroce* / *atrocitas*)

Voici, à titre d'exemples, les descriptions attribuées à *admirable* et *absurdité* par DériF en 2000 :

```
admirable ==> [ [ admirer VERBE] able ADJ] (admirable/ADJ, admi-  
rer/VERBE)  
::(<Prep> lequel/Que l') on peut admirer
```

Figure 7. Analyse par DériF d'*admirable*.

```
absurdité ==> [ [ absurde ADJ] ité NOM] (absurdité/NOM, absurde/ADJ)  
::Propriété de ce qui est absurde
```

Figure 8. Analyse par DériF d'*absurdité*.

54. Cf. par exemple Plénat (2002). L'affirmation n'est pas toujours exempte de contradictions. Par exemple, Huot (2001), conçu « dans une optique descriptive résolument synchronique » (p. 6), décompose *cantate*, qui est un emprunt à l'italien *cantata*, en [cant + at(e)]. Pour un compte rendu de l'ouvrage, cf. [17].

Dans des cas simples comme ceux sous (13) et (14), le critère d’analysabilité est lié à la possibilité de relier en synchronie les lexèmes apparemment porteurs du procédé étudied avec un lexème correspondant à un mot-forme<sup>55</sup> du français, selon le schéma majoritaire qui a été dégagé lors de l’implémentation du procédé concerné : base verbale pour la suffixation par *-able*, base adjectivale pour la suffixation par *-ité*.

Cependant, cela ne résout pas le cas de lexèmes comme ceux sous (15) et (16), puisqu’aucun mot-forme du français ne correspond aux séquences *fri-*, *mallé-*, *permé-*, *pot-*, *prob-* de (15), et que la même remarque vaut pour les séquences *alacr-*, *anfractuos-*, *crud-*, *céc-* de (16). S’agissant de *alter-* en (16), le cas est différent. Les dictionnaires attestent bien un verbe *altérer*, mais le suffixe *-ité* ne s’applique jamais à des verbes :

(15) *friable, malléable, perméable, potable, probable*

(16) *alacrité*<sup>56</sup>, *altérité, anfractuosité, crudité, cécité*

- Ces quelques exemples sont illustratifs des questions que nous nous sommes posées :
- jusqu’où doit-on effectuer la décomposition ?
  - jusqu’à quel point l’étymologie doit guider l’analyse constructionnelle ?
  - faut-il poser des bases non autonomes ?
  - si oui, quels critères fixer pour juger de la pertinence de ces bases ?

Dans MorTAL, nous avons considéré que qui peut le plus peut le moins. Nous avons donc décidé de décomposer des exemples comme ceux sous (15) et (16), pour peu que nous retrouvions dans leur sens attesté le sens général des dérivés en *-able* et en *-ité* (expression d’une potentialité pour les premiers, d’une propriété pour les seconds), et que le constituant restant, une fois ôté le suffixe, correspondait à un lexème du latin ou du grec. Leur base, ou ce qui a été identifié comme tel, a été catégorisée FWD (“foreign word”), étant admis qu’il ne s’agit là que d’une étiquette indiquant que la séquence en question ne se trouvait pas dans le référentiel de départ. Nous avons donné un équivalent de cette séquence au moyen d’un mot de notre référentiel (par ex. figure 9 : *permé-* “~ ‘traverser’”). Une fois qu’elle avait été identifiée comme “FWD”, la séquence venait augmenter le référentiel avec l’étiquette catégorielle de son équivalent sémantique, si bien que, si nous la rencontrions une nouvelle fois dans un mot potentiellement construit, l’analyseur ne se trouvait plus bloqué.

Les figures 9 et 10 donnent l’analyse proposée par DériF de *perméable* et d’*alacrité* :

```
perméable ==> [ [ permé- FWD] able ADJ] (perméable/ADJ, permé-/FWD ~
'traverser')
:: (<Prep> lequel/Que l') on peut ~ 'traverser'
```

Figure 9. Analyse par DériF de *perméable*.

55. J’emprunte le concept de ‘mot-forme’ à Fradin (2003a : 85 *sqq.*), qui l’adapte lui-même du concept de ‘word-form’ de Lyons (1977), en lui donnant une acception plus restrictive. Pour Fradin (*ibid.*), un mot-forme est un cas d’espèce de mot-type, *i.e.* un atome syntaxique, pleinement indépendant.

56. Le *PRO1* définit *alacrité* par « Enjouement, entrain ».



<p>alacrité ==&gt; [ [ alacr- FWD] ité NOM] (alacrité/NOM, alacr-/FWD ~ 'allègre')</p> <p>::Propriété de ce qui est ~'allègre'</p>
--

Figure 10. Analyse par DériF d'*alacrité*.

Nous avons été guidées dans notre décision par l'existence, dans le modèle de Danielle Corbin, de ce qu'elle appelle « unités infralexicales », *i.e.* unités inaptes à occuper une position syntaxiquement atomique, même si elles ont le même type de sens que des unités lexicales (D. Corbin 1997a : 80), et, au sein de ces unités infralexicales, de ce qu'elle appelle « archéoconstituants », *i.e.* « constituants empruntés par le français au latin et au grec (ex. *brachy-* 'court', *anthropo-* 'homme', *-cide* 'qui tue) » (D. Corbin 2001a : 44 ; *cf.* aussi D. Corbin 2001b : 131, n. 11).

De fait, il est possible de construire en français des lexèmes à partir de constituants auxquels ne correspondent pas de mots-formes en syntaxe. (17) donne ainsi un ensemble de lexèmes construits dont la forme comporte le segment *-lud-*. Ils ne figurent dans aucun des principaux dictionnaires du français actuel (*RE*, *TLF* et *PRO1*) et, pourtant, ne posent aucun problème d'interprétation. Tous sont attestés en corpus (*Toile*, *Le Monde*, *Libération*) :

- (17) *ludobus*, *ludoculturel*, *ludologie*, *ludologue*, *ludomanie*, *ludonirique*, *ludopathe*, *ludophage*, *ludophile*, *ludophobe*, *ludospace*

La question, pour la théorie, est toutefois celle de l'attribution d'un statut particulier à ce type de séquences, et de leur rapport aux règles de formation de lexèmes. Danielle Corbin leur en affecte explicitement un, celui d'archéoconstituant. Elles reçoivent également un statut dans Huot (2001), qui voit en elles des racines liées, ainsi que dans Rey-Debove (2001), pour qui ce sont des morphèmes lexicaux liés, ces derniers constituant l'objet de la « morphologie profonde »<sup>57</sup>. Or, ces objets, quelque nom qu'on leur donne, ne sont pas des lexèmes, si, comme Fradin (2003a : 102), on définit le lexème comme « une unité abstraite à quoi se rapporte un mot-forme », faute de mot-forme correspondant. S'ils ne sont pas des lexèmes, ils ne peuvent pas se voir appliquer de règles de construction de lexèmes, puisque ces dernières, selon le même ouvrage et, plus généralement, selon les tenants de la morphologie lexématique, prennent pour unité de base le lexème. Par ailleurs, l'assignation d'un statut particulier à des séquences comme *lud-* est fondée sur une décomposition qui ressortit à l'analyse en constituants immédiats, ce qui revient à leur assigner le statut de morphème (Huot 2001 et Rey-Debove 2001 l'écrivent explicitement). Or, il est de plus en plus communément admis, depuis Anderson (1992), que l'analyse en constituants immédiats et, avec elle, le morphème ne sont pas pertinents pour l'analyse du lexique construit. On se trouve donc avec, sur les bras, des unités indésirables, tout en devant rendre compte du fait que ces unités peuvent être mobilisées pour former de nouveaux lexèmes. Fradin (2003a : 197-198) considère que ces « formants d'origine savante » peuvent entrer dans la construction de lexèmes par

57. *Cf.* p. 533 : « La morphologie profonde n'est pas la morphologie dérivationnelle traditionnelle : c'est le lieu lexical des mots dont les unités constituantes ne sont jamais libres ».

« le jeu de mise en série », sans vraiment préciser quel statut il leur donne, si ce n'est que ce sont des unités hétérolexicales (p. 197). Leur statut exact et leur prise en charge par la morphologie constructionnelle restent par conséquent à établir <sup>58</sup>.

### 2.2.3.3. Le problème des choix multiples

Un autre problème auquel nous avons été confrontées pour la réalisation de notre base a été celui des choix multiples, et des critères à établir pour favoriser une analyse plutôt qu'une autre. Il ne s'agit pas ici de cas comme *importable* déjà cité dans la note 49, dont la forme correspond à deux lexèmes différents, obtenus par des chemins constructionnels différents, mais réellement de cas où il nous a semblé qu'un dédoublement en deux analyses menant au même sens était inadéquat. Le cas s'est particulièrement posé lors de l'implémentation du préfixe *dé-*, et, pour d'autres raisons, lors de l'implémentation des suffixes *-is-* et *-ifi-* <sup>59</sup>.

S'agissant de *dé-*, nous avons considéré qu'il peut s'appliquer à des noms (*os* > *dé-sosser*), à des adjectifs (*jaune* > *déjaunir*), et à des verbes (*coincer* > *décoincer*). Dans ce dernier cas, formellement, *dé-* s'applique au radical verbal, mais, sémantiquement, il sélectionne ce qui, dans le verbe, marque une propriété, donc, si on veut, le participe passé, puisque la propriété 'coincé' correspond en français à une forme de verbe. Contrairement à D. Corbin (1992), je ne pense pas qu'il soit nécessaire de poser un adjectif *coincé*, issu de ce qu'elle appelle une opération de « transcatégorisation » <sup>60</sup> du participe passé du verbe : une base verbale convient également, d'autant plus que le participe, comme l'infinitif, relève de la flexion inhérente (Booij 2002 : 20). Or, on sait que les procédés constructionnels peuvent prendre pour base des expressions fléchies, s'il s'agit de flexion inhérente (Booij 2002 : 41 ; 80 *sqq.*) <sup>61</sup>.

58. Bernard Fradin (communication personnelle) propose de traiter *lud-* comme radical supplétif du lexème JEU et de le mentionner, à ce titre, sous la rubrique phonologique de ce dernier. L'une des rubriques de JEU stipulerait qu'il faut choisir ce radical pour construire des lexèmes à partir de JEU. Cette solution, qui recoupe celle que propose D. Corbin (1997c, supplément : II) pour *galact-* et *lact-*, qu'elle donne comme des formes supplétives de *lait*, est envisageable pour *lud-* dans la mesure où ses valeurs sémantiques recouvrent celles de *jeu* (activité ; support de l'activité), à condition de mentionner un second radical supplétif, *joc-* (ex. *jocothérapie* "thérapie par le jeu"). Il faudrait toutefois vérifier si elle est extensible à tous les archéoconstituants, par exemple à *brachy-* 'court' que cite D. Corbin (2000a : 44), ou encore à *-graph-* auquel ne correspond pas à proprement parler de lexème en français (D. Corbin 1997a : 80 le glose par "laisser une trace visible sur un support").

59. Il s'est également posé, dans une moindre mesure, lors de l'implémentation du suffixe *-able*. Pour ce dernier, je renvoie à [12 : 155-156], pour la solution que nous avons préconisée lorsque nous avions le choix entre une base verbale et une base nominale.

60. Plus tard dans ses travaux, elle parlera de *déflexivation*. Cf. D. Corbin (2001b : 131, n. 10) : « J'appelle *déflexivation* l'opération qui consiste à construire une unité lexicale à partir d'une forme fléchie (en français, infinitif et participes). La *déflexivation* reproduit la forme de sa base et donne au mot construit une catégorie autre (nom ou adjectif) ».

61. Sur la distinction entre flexion contextuelle et flexion inhérente, cf. Booij (1996). En un mot, la flexion contextuelle est requise par le contexte syntaxique, sans elle-même être porteuse d'information particulière, la flexion inhérente donne, elle, de l'information. La flexion des adjectifs en français selon la catégorie du nombre est du premier type, celle des noms du second : ainsi, dans *les grands arbres*, la marque de pluriel dont est porteur *grands* est dépendante d'*arbres*, mais ne dit

Dans les trois cas, rejoignant Gerhard-Krait (2000), nous avons été amenées à constater que, de façon massive, *dé-* apparaissait comme un marqueur de dissociation (cf. [19 : 279]), entre ce que désigne la base (objet ou propriété) et le référent du thème du verbe : *désosser un poulet*, *déjaunir du linge*, *décoincer une clé d'une serrure*. Dans le cas d'une construction sur un nom, l'entité dissociée peut correspondre à la base du verbe (*désosser un poulet*), ou à son thème (*débarquer des marchandises*), avec des cas d'ambiguïté, dans lesquels la base peut être interprétée comme site ou comme cible (Gerhard-Krait *ibid.* : 213). C'est le cas dans *décartonner* : "enlever qch d'un carton" / "enlever le carton de qch", ou dans *dessabler* : "enlever qch du sable" / "enlever le sable de qch"<sup>62</sup>.

Les figures 11 à 13 font apparaître les gloses que nous avons attribuées à chacun des cas distingués :

```
désosser ==> [ dé1 [os NOM ] VERBE] (désosser/VERBE, os/NOM)
:: (Enlever | (Faire) sortir de) os
```

Figure 11. Analyse par DériF de *désosser*.

```
déjaunir ==> [ dé1 [jaune ADJ ] VERBE] (déjaunir/VERBE, jaune/ADJ)
:: (Supprimer | Faire perdre) le caractère jaune
```

Figure 12. Analyse par DériF de *déjaunir*.

```
décoincer ==> [ dé1 [coincer VERBE ] VERBE] (décoincer/VERBE, coincer/VERBE)
:: (Enlever ce qui a pour effet de | Annuler l'état lié au processus) de coincer
```

Figure 13. Analyse par DériF de *décoincer*.

Nous avons été confrontées à un problème de choix avec des cas comme *débarrier* ou *décapuchonner*, pour lesquelles deux analyses étaient envisageables. L'une, conformément à la figure 11, construit *débarrier* sur *barre*, *décapuchonner* sur *capuchon* ; l'autre, conformément à la figure 13, construit *débarrier* sur *barrer*, *décapuchonner* sur *capuchon*, *barrer* et *capuchonner* étant eux-mêmes issus de la conversion de noms. Chacune de ces analyses possibles conduit à un verbe exprimant en substance le procès de faire en sorte qu'une entité, qui avait une barre ou un capuchon, ne l'ait plus à l'issue du procès, la différence étant que, dans le second cas, la présence de la barre et du capuchon est présentée comme résultant de la réalisation d'un procès. (18) reprend les deux analyses possibles pour chacun de ces verbes :

---

rien du nombre de propriétés 'grand', en revanche, le *s d'arbres* nous dit quelque chose du réel. L'affirmation selon laquelle la dérivation s'applique nécessairement antérieurement à la flexion (cf. [11]) ne vaut donc que pour la flexion contextuelle.

62. Le *RE* définit *décartonner* par « Enlever le carton (de qqch.) ». Les seules attestations de ce verbe que j'aie trouvées sur la Toile mettent en jeu l'autre sens, "sortir d'un carton", par ex. « Décartonner les yaourts avant de les ranger ». *Dessabler* est défini par « ôter le sable de ». Les deux sens possibles sont, eux, attestés sur la Toile : « Mettre les tellines à dessabler dans de l'eau de mer » vs « je me dis qu'il serait grand temps d'aller dessabler la moto ».

- (18) a *barre* → *débarrer* “enlever la/une barre”  
 b *barre* → *barrer* “pourvoir d’une barre” → *débarrer* “enlever ce qui a pour effet de barrer / annuler l’état lié au procès de barrer, *i.e.* de pourvoir d’une barre”  
 a’ *capuchon* → *décapuchonner* “enlever le/un capuchon”  
 b’ *capuchon* → *capuchonner* “pourvoir d’un capuchon” → *décapuchonner* “enlever ce qui a pour effet de capuchonner / annuler l’état lié au procès de capuchonner, *i.e.* de pourvoir d’un capuchon”

Nous avons systématiquement opté pour une base nominale (analyses a et a’ *supra*), pour plusieurs raisons. Admettons que *dé-* soit sensible au trait [ $\pm$  acquis] du référent du nom en position de base et que l’analyse varie en fonction de ce trait. Il s’ensuit qu’on assignera une analyse différente à *dégraisser* selon qu’il prend pour argument *une viande* ou *la vaisselle* :

- dans le cas de la viande, la graisse, qui est une partie naturelle de la viande, est inhérente, et le verbe est dénominal (*graisse* → *dégraisser*) ;
- dans l’autre, la graisse, que la vaisselle n’a pas de façon inhérente, du moins on le suppose, est acquise, et le verbe est déverbal (*graisse* → *graisser* → *dégraisser*).

Il nous a semblé suspect que le sens et la structure du verbe en *dé-* puissent varier en fonction de son argument interne, et des propriétés ontologiques qu’entretiennent le référent de cet argument et celui du nom en position de base. Plus exactement, si l’on se place du point de vue du locuteur, il nous a semblé que dissocier l’analyse selon le trait [ $\pm$  acquis] du référent du nom en position de base était surestimer sa compétence, et qu’il n’était pas déraisonnable de penser que ce que le locuteur entend dire avec le verbe *dégraisser*, c’est qu’avant le procès, il y avait de la graisse, et qu’ensuite, il n’y en a plus, sans se soucier du caractère inhérent ou acquis de la graisse en question. La remarque vaut pour d’autres verbes, par exemple *délarde*, qu’il semble contre-intuitif de dédoubler en deux verbes, dotés chacun d’une structure différente selon que l’on *délarde* un porc qui vient d’être abattu, ou une volaille que l’on a entourée de lard pour la faire cuire <sup>63</sup>.

Par ailleurs, le choix que nous avons fait présentait également l’avantage de ne pas restituer d’étapes verbales intermédiaires absentes de notre référentiel, ou attestées avec des sens qui ne correspondaient pas à notre analyse. Ainsi, *délainer* (« Enlever la laine de (une peau de mouton, de chèvre), de sorte qu’elle soit utilisable par l’industrie textile », *RE*) ne peut-il être construit que sur *laine*, puisque la laine de la peau de mouton est une partie naturelle. Si nous dédoublions l’analyse, il fallait poser un second verbe *délainer*, construit sur *lainer*, lui aussi attesté dans notre référentiel, mais avec un sens inadéquat (« **1.** Rendre moelleux (un tissu de laine). **2. Par anal.** Saupoudrer (un papier de tenture) de laine hachée, pour imiter le velouté d’une étoffe », *RE*), ou bloquer dans ce cas une dérivation sur le verbe.

63. Cf. aussi *décapsuler* (un organe vs une bouteille), *décolorer* (ses cheveux vs la laine), *décorner* (un bœuf vs une page de livre), *dégivrer* (la planète Mars vs une vitre), *délustrer* (la peau des truites vs un vêtement), etc. La solution proposée évite aussi de se demander quel rapport entretiennent des lattes et un plancher (*délatter un plancher*) : en tant qu’artefact, un plancher doit avoir fait l’objet d’un processus de lattage, mais en tant que plancher, les lattes lui sont inhérentes. La remarque vaut pour un mât et un bateau (*démâter un bateau*), une tresse et une corde (*détresser une corde*), etc.

Si nous dédoublions l'analyse quand deux candidats au rôle de base, un nom et un verbe, étaient envisageables, et si nous voulions être cohérentes, il fallait aussi que nous la dédoublions pour un cas comme *désépaissir*, pour lequel on peut hésiter entre une base adjectivale, *épais*, et une base verbale, *épaissir*. Or, qu'il s'applique à *une chevelure* ou *une sauce*, le sens du verbe est le même : il indique qu'il s'agit d'enlever la propriété 'épais' du référent du thème du verbe, qu'elle soit inhérente ou acquise<sup>64</sup>.

Le fait que le verbe en *dé-* soit en relation d'antonymie avec un verbe issu de la conversion d'un nom (*débarrer / barrer ; débarricader / barricader ; décapuchonner / capuchonner ; desseller / seller*), ou avec un verbe préfixé par *en-* (*déchaîner / enchaîner ; décrasser / encrasser ; dégluer / engluer ; dessabler / ensabler*) ne s'oppose pas à notre solution. Acceptons, suivant Gerhard-Krait (2000 : 267 *sqq.*) qui emprunte elle-même la notion à Aurnague & Plénat (1996) et (1997), que ce qui unit le référent de la base du verbe en *dé-* et le référent de son argument interne soit une relation d'attachement habituel (*barre / porte ; capuchon / stylo ; sable / mollusques ; selle / cheval ; etc.*)<sup>65</sup>, et admettons que l'on veuille former un verbe. Il n'existe alors que deux possibilités : soit le verbe exprime le procès d'associer les deux référents (*barrer une porte ; barricader une entrée ; capuchonner un stylo ; seller un cheval ; encrasser ses poumons ; ensabler sa moto ; etc.*), soit il exprime le procès de les dissocier (*débarrer une porte ; décapuchonner un stylo ; décrasser ses poumons ; dessabler sa moto ; desseller un cheval*). Il est par conséquent prévisible que les verbes en *dé-* et les verbes en *en-* ou les verbes issus d'une conversion expriment des processus inverses, sans que les premiers dérivent pour autant des seconds.

Enfin, s'agissant de doublets, dans les dictionnaires, comme *décadrer / désencadrer ; déchaîner / désenchaîner ; décrasser / désencrasser ; dégluer / désengluer ; déneiger / désenneiger ; dessabler / désensabler ; dévaser / désenvaser*, etc., dans lesquels chaque terme exprime le procès d'enlever ce que désigne le nom repérable dans sa forme observable, ou de sortir quelque chose de ce que désigne ce nom, nous avons construit le premier sur le nom, conformément à l'analyse exposée ci-dessus, et le second sur le verbe préfixé par *en-*. L'existence de ces doublets ne nous a pas paru suffisante pour conclure que la forme sans *en-* était une version accourcie de la forme avec *en-*, au contraire même<sup>66</sup> : si *dé-* devait tronquer *en-*, il le ferait toujours, et le lexique attesté ne donnerait pas à observer ces doublets. Les figures 14 et 15 appliquent l'analyse au doublet *dégluer / désengluer* :

64. S'agissant de *désépaissir*, notre choix rejoint l'analyse proposée par Gerhard-Krait (2000 : 330). Elle construit en revanche *débarrer* et *décapuchonner* sur des verbes (p. 344).

65. La relation peut aussi être contingente (*barricade / entrée ; sable / moto*). Elle s'inscrit alors dans un scénario (par exemple, course Paris/Dakar pour le sable et la moto).

66. *Développer* et *dépêtrer* s'opposent apparemment à l'hypothèse, dans la mesure où le segment *-velop-* du premier, *-pêtr-* du second semblent provenir des verbes en *en-* correspondants. S'agissant de *développer*, le RE le donne comme construit sur l'ancien français *voloper* "envelopper, entourer". Quand à *dépêtrer*, on peut faire jouer l'effet de série, et considérer qu'il est le résultat d'un calcul analogique effectué à partir des doublets qu'on examine. En tout état de cause, il ne suffit pas à remettre en cause l'analyse que nous avons proposée.

```
dégluer ==> [ dél [glu NOM ] VERBE] (dégluer/VERBE, glu/NOM)
::(Enlever | (Faire) sortir de) glu
```

Figure 14. Analyse par DériF de *dégluer*.

```
désengluer ==> [ dél [engluer VERBE ] VERBE] (désengluer/VERBE, en-
gluer/VERBE)
::(Enlever ce qui a pour effet de | Annuler l'état lié au pro-
cès) de engluer
```

Figure 15. Analyse par DériF de *désengluer*.

Le même choix a été fait pour des doublets comme *dégourdir / désengourdir*, *modulo* la catégorie adjectivale du primitif :

```
dégourdir ==> [ dél [gourd ADJ ] VERBE] (dégourdir/VERBE, gourd/ADJ)
::(Supprimer |Faire perdre) le caractère gourd
```

Figure 16. Analyse par DériF de *dégourdir*.

```
désengourdir ==> [ dél [engourdir VERBE ] VERBE] (désengourdir/VERBE,
engourdir/VERBE)
::(Enlever ce qui a pour effet de | Annuler l'état lié au pro-
cès) de engourdir
```

Figure 17. Analyse par DériF de *désengourdir*.

Ce qui précède montre que l'implémentation du préfixe *dé-* a été l'occasion de réfléchir assez sérieusement aux problèmes linguistiques que pose ce préfixe, qui constitue un thème récurrent dans les travaux de Danielle Corbin. Les analyses que nous avons faites s'éloignent sur plusieurs points de ses propres analyses. Elles n'ont en particulier pas le même degré de sophistication (*cf.* par exemple D. Corbin 1992). Il me semble toutefois que leur simplicité même constitue un garant de leur adéquation si, de nouveau, on veut rendre compte de la compétence des locuteurs. Le détour que fait par exemple D. Corbin (1992 : 209) par un adjectif *cloisonné* entre le nom *cloison* et le verbe *décloisonner* pour expliquer ce dernier est motivé par la volonté de calquer l'analyse sur celle des verbes en *dé-* qu'elle construit sur des adjectifs eux-mêmes issus d'une transcatégorisation d'un participe passé d'un verbe éventuellement converti (par ex. *déchiffrer*). Malgré les arguments qu'elle avance, il nous est apparu que l'analyse pouvait être simplifiée, et qu'elle gagnait d'ailleurs à l'être.

Le second problème de choix, que je commenterai rapidement, concerne l'implémentation des suffixes *-is-* et *-ifi-*. Comme il se pose dans les mêmes termes pour les deux suffixes, je me centrerai ici sur le cas de *-is-*.

On admettra que, comme son homologue anglais *-ize* (*cf.* Plag 1999), le suffixe français *-is-* peut s'appliquer à des adjectifs (*adverbialiser, africaniser*) et à des noms (*bémoliser, cabaliser, hospitaliser, macadamiser*). Dans des cas simples comme les exemples que je viens de citer, une machine est capable de décider quelle base choisir. Il suffit qu'elle se fonde sur la forme graphique du dérivé à analyser, avec, éventuellement, quelques règles d'allomorphie (par ex. si le verbe se termine par la suite gra-

phique <aniser>, une règle stipule d'aller regarder si le référentiel ne contient pas une base adjectivale possible terminée par <ain>).

L'identification automatique d'une base pour des verbes comme *algébriser*, *allégoriser*, *anarchiser*, *astrologiser*, *automatiser*, *bureaucratiser*, *dogmatiser*, *harmoniser* s'est révélée, elle, plus délicate (les définitions qui suivent sont extraites du *RE*) :

(19) **algébriser** : « ♦1. **Vx.** Faire de l'algèbre, des mathématiques. ♦2. **V. tr. [a] (XXe).** Donner un caractère algébrique à... *Algébriser la topologie.* **Pron.** *Plusieurs domaines des mathématiques tendent à s'algébriser* ». **[b] (Au sens fig. d'algèbre).** Rendre formel, précis »

**allégoriser** : « ♦1. Expliquer allégoriquement (qqch.). [...] ♦2. Représenter (des choses, des notions abstraites) de manière allégorique. **Absolt.** S'exprimer par allégories »

**anarchiser** : « ♦1. **V. tr. Vx.** Rendre anarchique ; mettre dans l'anarchie. ♦2. **V. intr. (1793).** Se conduire en anarchiste »

**astrologiser** : « Faire de l'astrologie en tirant des conclusions hasardeuses »

**automatiser** : « Rendre automatique. [...] **(Compl. n. de personne).** Rendre semblable à un automate »

**bureaucratiser** : « Transformer par la mise en place d'une bureaucratie »

**dogmatiser** : « Ériger en dogme »

**harmoniser** : « ♦1. Mettre en harmonie, en accord (plusieurs choses; une chose avec une autre). [...] ♦2. **Mus.** Combiner (une mélodie) avec d'autres parties ou avec des suites d'accords, en vue de réaliser un ensemble harmonique »

Pour chaque verbe de (19), notre référentiel offrait plusieurs candidats :

– un nom, plus ou moins visible dans la forme du verbe : respectivement *algèbre*, *allégorie*, *anarchie*, *astrologie*, *automate*, *bureaucrate*, *dogme*, *harmonie*. On remarque qu'il s'agit dans la moitié des cas d'un nom terminé par /i/ ;

– un nom dont le suffixe aurait été tronqué (*anarchiste*) ;

– un adjectif, le plus souvent en *-ique* (*algébrique*, *allégorique*, *anarchique*, *astrologique*, *automatique*, *bureaucratique*, *dogmatique*, *harmonique*), mais pas exclusivement (*harmonieux*).

D'un point de vue sémantique, les deux constructions, l'une sur un nom, l'autre sur un adjectif, sont recevables pour la quasi-totalité des cas. D'un point de vue formel, *catholiciser*, *mysticiser*, *organiciser*, *publiciser*, *techniciser* montrent qu'une finale en /ik/ peut résister devant le suffixe *-is-*, sous la forme /is/. *Dialectiser*, *érotiser*, *techniser*<sup>67</sup> montrent toutefois que /ik/ peut aussi disparaître devant *-is-*. L'alternative qui s'offrait à nous était donc la suivante : soit dédoubler chaque forme verbale en *-is-* en deux homonymes, dotés chacun d'une structure différente (c'est l'option faite par Roger 2003 dans des cas similaires), soit opter pour une analyse et une seule. C'est ce dernier choix que nous avons fait, en optant systématiquement pour l'analyse le plus en accord

67. Selon le *RE*, *techniciser* est plus courant que *techniser* (une vérification rapide sur la Toile le confirme). Voici comment le *RE* décrit *techniser* dans sa rubrique étymologie : « Av. 1964, date où le dér. *technisation* est attesté ». On peut donc se demander si *techniser* n'est pas une « formation régressive » de *technisation*, lui-même version accourcie de *technicisation*, de façon à éviter la cooccurrence de deux séquences phoniques proches dans le même dérivé (/tek.ni.si.za.sj□/).

avec la forme observable du dérivé à analyser. Nous avons donc systématiquement opté pour une base nominale. Cette solution n'est pas entièrement satisfaisante, puisque nous avons dû corriger manuellement des analyses de ce fait erronées (par ex. *dialectiser*, qui n'est pas construit sur *dialecte*, mais sur *dialectique*). Elle se fonde toutefois sur le fait qu'à un certain niveau, un nom et l'adjectif en *-ique* qui en dérive ont le même sens<sup>68</sup>. Cette intuition, que nous n'avons pas creusée pour le projet MorTAL, est à la source de [20] (cf. *infra* § 2.3.), et est corroborée par le fait que, de façon récurrente, un nom et l'adjectif qui en dérive peuvent être utilisés comme supplétifs l'un de l'autre (cf. [16], [21] et §§ 2.4.1. et 2.4.2. *infra*).

#### 2.2.3.4. Bilan linguistique global

Tous les problèmes dont il vient de s'agir ne se situent pas au même niveau : certains montrent les limites du modèle de morphologie constructionnelle de Danielle Corbin (phénomènes constructionnels non centraux), d'autres concernent tous les modèles, quels qu'ils soient : c'est le cas du rapport entre étymologie ou emprunt et analysabilité en synchronie, ainsi que des problèmes de choix auxquels nous avons été confrontées. Cependant, les décisions que nous avons prises ont été, pour la plupart, guidées par une volonté de simplicité. Ne disposant pas de sens pour analyser automatiquement le lexique construit, notre seul guide a été la forme. Or, il m'apparaît, avec le recul, que les résultats auxquels nous sommes parvenues, même s'ils sont discutables et perfectibles, ont le mérite de ne pas nécessiter une machinerie lourde comme celle à laquelle j'ai pu recourir dans [0]. Par ailleurs, face à un lexème construit qu'il rencontre pour la première fois, le locuteur est aussi démuni qu'une machine, et, dans un premier temps, c'est la forme, et seulement elle, qui lui donne des indices sur son sens. On peut donc supposer que la situation dans laquelle nous nous sommes trouvées pour implémenter le modèle de Danielle Corbin n'est pas si éloignée que cela de celle dans laquelle se trouvent les locuteurs dans leur rapport au lexique construit.

Les deux pans de ma recherche qui viennent d'être présentés s'inscrivaient, initialement au moins, dans le cadre du modèle de morphologie constructionnelle de Danielle Corbin. Les deux thèmes dont il va s'agir dans les §§ 2.3. et 2.4. remettent en cause, chacun à leur manière, des choix effectués par ce modèle, au moins dans sa version initiale : il s'agit du principe d'unicité catégorielle, que j'ai déjà évoqué rapidement dans ce qui précède, et qui pose plus généralement le problème de la catégorisation des bases de construction de lexèmes, et de la place des relations paradigmatiques entre lexèmes dans un modèle d'inspiration générative comme l'est celui de Danielle Corbin.

### 2.3. La catégorisation des bases de construction de lexèmes

---

68. Dans les termes de Croft (1991) (cf. *infra*, § 2.3.2.), ils appartiennent à la même classe sémantique, mais n'ont pas la même fonction pragmatique (référence *vs* modification). Booij (2002 : 108) cite pareillement des cas en néerlandais où l'adjectif et le nom ont le même sens (par ex. *muzikaal aanleg*, *muziekaanleg* "disposition pour la musique").

Deux de mes travaux, [5] et [20], portent explicitement sur le problème de la catégorisation des bases de construction de lexèmes. Le premier date de 1997, et est paru dans les actes d'un colloque intitulé « Linguistique et informatique : théories et outils pour le traitement automatique des langues ». Le second, à paraître dans *Lexique* 16, est tout récent, puisque je l'ai écrit parallèlement à cette note de synthèse. Nous avons également sans cesse été confrontées à la question de la catégorisation des bases dans le projet MorTAL, dans la mesure où les problèmes de choix multiples dont j'ai donné un aperçu dans le § 2.2.3.3. mettaient le plus souvent en jeu des bases appartenant à des catégories différentes. Les décisions que nous avons prises supposaient par conséquent que nous nous situions préalablement par rapport à cette problématique<sup>69</sup>. L'amendement apporté à [0] en 1999 était, lui aussi, lié à un problème de catégorisation des bases.

On a vu dans le § 1.2. *supra* qu'en 1991, Danielle Corbin définit une RCM comme résultant de l'association de trois composants, dont une opération structurelle instaurant un rapport catégoriel entre bases et dérivés. On a également vu qu'en 1987, elle considère que ce rapport catégoriel est unique, qu'à partir de 1991, elle accepte qu'il souffre des entorses dans deux domaines bien circonscrits, celui des suffixes évaluatifs et celui des préfixes spatio-temporels, et qu'en 2001, il devient beaucoup moins strict. De fait, D. Corbin (2001a) se termine sur la conclusion que l'instruction catégorielle dont sont porteurs les procédés constructionnels est déterminée par leur instruction sémantique, combinée au type sémantique des bases (p. 65). On passe ainsi d'un principe d'unicité catégorielle à un principe d'unicité sémantique, ce qui constitue plus qu'un simple changement de point de vue. Dès lors qu'on déplace la focale des catégories mises en relation par les procédés constructionnels au sens qu'ils construisent, on se débarrasse d'un certain nombre de problèmes auxquels achoppait le principe d'unicité catégorielle, appliqué au pied de la lettre. On explique en outre en partie pourquoi certaines catégories peuvent être mises en relation, et pourquoi certaines ne le peuvent pas. [5] relate ce changement de perspective dans le modèle de Danielle Corbin (§ 2.3.1.). [20] reconsidère le problème de la catégorisation des lexèmes impliqués dans la construction du lexique construit à la lumière de la théorie des catégories de Williams Croft (§ 2.3.2.).

### 2.3.1. D'un principe d'unicité catégorielle à un principe d'unicité sémantique

La caractérisation des procédés de construction de lexèmes passe systématiquement par un appariement catégoriel, y compris, non sans paradoxe, chez les linguistes comme Ingo Plag pour qui les bases sont catégoriellement sous-spécifiées (j'en reparlerai ci-dessous) : ainsi en français *-ité* est-il réputé former des noms à partir d'adjectifs, *in-* des adjectifs à partir d'adjectifs, *-ation* des noms à partir de verbes, etc. Cette pratique, qu'on observe aussi loin que l'on remonte dans le temps (*cf.* par exemple la façon dont

---

69. Par exemple, l'analyse des dérivés en *dé-* dans le projet MorTAL suppose le rejet préalable de l'hypothèse que fait D. Corbin (1992) d'une catégorisation unique des bases.

Jespersen 1942 : 466 décrit la préfixation par *un-* en anglais), conjuguée au fait que certains procédés ne semblent s'appliquer qu'à une catégorie de bases et une seule (c'est le cas des exemples ci-dessus) a conduit Aronoff (1976 : 48) à formuler l'hypothèse de la base unique (« Unitary Base Hypothesis », désormais UBH) :

We will assume that the syntacticosemantic specification of the base, though it may be more or less complex, is always unique.

Quelques années plus tard, Scalise (1984 : 137 *sqq.*) ajoute l'hypothèse de la sortie unique (« Unitary Output Hypothesis », désormais UOB) <sup>70</sup> :

[B]esides the Unitary *Base* Hypothesis, it is necessary to postulate a Unitary *Output* Hypothesis. This second hypothesis does not allow a particular phonological form to be considered a single affix if it produces outputs with different category labels or different semantics.

avant d'évaluer la validité de l'UBH. Pour lui, l'hypothèse vaut, au mieux, pour les suffixes, et, s'agissant de ces derniers, l'unicité n'est pas tant celle de la catégorie syntaxique de l'entrée que celle de ses traits syntaxiques, formulés dans le cadre de la grammaire générative. Constatant qu'en italien, un nombre non négligeable de suffixes peuvent s'appliquer à la fois à des adjectifs et à des noms d'une part, à des adjectifs et à des verbes d'autre part, il subsume en effet les bases des premiers sous le trait [+ N], celles des seconds sous le trait [+ V] <sup>71</sup>, et propose la version modifiée suivante de l'UHB (MUHB) :

A suffix may be attached only to bases that form a syntactic class specifiable in terms of a single category feature in X bare theory terms.

La MUBH prédit ainsi qu'un suffixe donné s'attache soit à des adjectifs et à des noms, soit à des adjectifs et à des verbes. Elle exclut en revanche la possibilité d'une application conjointe à des noms et à des verbes, puisque ces deux catégories de mots ne partagent pas de trait syntaxique dans la théorie X-barre (Scalise *ibid.* : 139) <sup>72</sup>. Pour Sergio Scalise, les "suffixes" qui s'opposent à la MUBH soit n'en sont pas (*-ic-* en italien, qui serait un infix), soit ne sont pas productifs avec l'une des catégories de bases (*-ista*, *-ismo*), soit doivent être dédoublés en deux suffixes homonymes (*-ata*). On note qu'Aronoff (1976 : 48) recourt pareillement au dédoublement homonymique pour le suffixe *-able*, selon qu'il s'applique à des verbes (*acceptable*, *doable*) ou à des noms (*fashionable*, *sizable*).

Le cas des suffixes évaluatifs est à part. Conscient de la redondance qu'il y aurait à distinguer trois règles, une pour chaque appariement catégoriel effectué, Scalise (*ibid.* : 132) considère que les suffixes évaluatifs sont transparents eu égard aux catégories

70. L'hypothèse est en filigrane dans Aronoff (1976 : 49-50).

71. Je rappelle que, dans la veine de Chomsky (1970), chaque catégorie est définie en fonction des traits  $[\pm N]$  et  $[\pm V]$  : le nom est marqué  $[+N, -V]$ , le verbe  $[-N, +V]$ , l'adjectif  $[+N, +V]$ , l'adposition  $[-N, -V]$ .

72. Scalise (*ibid.*) remarque aussitôt que cette interdiction vaut pour les suffixes mais pas pour les préfixes, ce qui explique qu'il exclue ces derniers de la MUHB.

mises en relation (ils forment des noms sur des noms, des adjectifs sur des adjectifs, des verbes sur des verbes), ainsi d'ailleurs qu'aux traits de sous-catégorisation des bases, qu'on retrouve dans les dérivés. Ces caractéristiques, corroborées par le fait que les suffixes évaluatifs sont périphériques par rapport aux autres règles de construction de mots, en font, pour lui, des entités à mi-chemin entre les règles constructionnelles et les règles flexionnelles.

En 1987, Danielle Corbin conjoint l'UHB d'Aronoff (1976) et l'UOB de Scalise (1984) en une seule hypothèse, qu'elle appelle la « contrainte d'unicité catégorielle ». Elle l'énonce de la façon suivante (1987a : 482) :

Chaque RCM ne s'applique qu'à une catégorie de base et une seule, et n'attribue aux mots qu'elle construit qu'une catégorie et une seule.

On remarque d'abord que seul est conservé l'étiquetage catégoriel des bases et des dérivés, et qu'il n'est plus question de sens (contrairement à l'UOB, et, moins clairement, à l'UHB, dans laquelle l'adjectif *syntacticosemantic* est flou). Par ailleurs, dans le modèle de Danielle Corbin, à une RCM donnée peuvent correspondre des préfixes et des suffixes. Il s'ensuit que l'unicité catégorielle vaut pour ces deux types d'affixes, et non, comme dans Scalise (1984), seulement pour les suffixes. D. Corbin (1987a : 580) rejette d'ailleurs la MUBH, sans vraiment s'en expliquer.

Tel qu'il était énoncé en 1987, le principe d'unicité catégorielle ne souffrait pas d'exception : de fait, comme je le souligne dans [5], il ne le pouvait pas, sauf à perdre son statut de principe. Il en résulte, dans certains cas, une démultiplication d'une même forme affixale en des procédés homonymes, même dans des cas comme ceux sous (20) que je reprends à [5]. Malgré la proximité sémantique des dérivés (procès mené à terme en a, dépassement par rapport à une échelle normée en b, marquage d'évaluation en c), D. Corbin (1987a) détriplexe chaque forme affixale de (20), selon le rapport catégoriel dans lequel elle est impliquée :

- (20) a *lourd / alourdir* : A > V  
       *mener / amener* : V > V  
       *coude / accouder* : N > V
- b *capacité / surcapacité* : N > N  
       *doué / surdoué* : A > A  
       *administrer / suradministrer* : V > V
- c *fille / fillette* : N > N  
       *propre / propre* : A > A  
       *voler / voleter* : V > V

Comme je l'écris dans [5], dans une théorie dissociative, qui disjoint forme et sens, des décisions de ce type ne feraient pas problème : rien ne s'opposerait alors qu'une instruction sémantique donnée trouve des échos multiples dans des formes affixales homomorphes. Ces décisions sont en revanche davantage problématiques dans une théorie comme celle de Danielle Corbin, qui opte pour une logique strictement associative.

Dans une telle perspective, il n'est pas satisfaisant qu'une même opération sémantique se trouve clonée dans des procédés constructionnels homomorphes.

Un second effet pervers de la contrainte d'unicité catégorielle, appliquée strictement, consiste en des analyses contre-intuitives, dont le but est de réduire un rapport catégoriel observable à un rapport catégoriel inévitable, voire de ramener à l'unité un rapport catégoriel multiple jugé apparent. À titre d'illustrations, je citerai le suffixe *-et(te)*, dont D. Corbin (1986) démontre l'inaptitude à former des noms à partir de verbes (on a vu que [0] se range à cette hypothèse), ou le préfixe *dé-*, que D. Corbin (1992) n'autorise à s'appliquer qu'à des adjectifs quand il forme des verbes. On comprend dès lors les étapes adjectivales qu'elle restitue, même si certaines semblent forcées<sup>73</sup>.

À partir de 1991, Danielle Corbin admet que la contrainte d'unicité puisse souffrir des entorses dans le domaine des suffixes évaluatifs et dans celui des préfixes spatio-temporels, rejoignant en cela en partie Scalise (1984). Parmi les suffixes, elle cite *-asse* et *-et(te)* (1991a : 21), parmi les préfixes (*ibid.*), *archi-*, *co-*, *inter-*, *intra-*, *post-*, *pré-*, *pro-*, *sous-*, *sur-*, *trans-*, et clôt sa liste par un « etc. ». Dans les deux cas, le revirement tient à des raisons sémantiques<sup>74</sup>.

Après avoir subi de nouveaux assauts dus à l'évolution de la réflexion de Danielle Corbin, mais aussi aux monographies qui ont été écrites dans le cadre théorique qu'elle a élaboré (par ex. Aliquot-Suengas 1996 sur les suffixes collectifs), la contrainte d'unicité catégorielle cède le pas dans les derniers écrits de Danielle Corbin à un principe d'unicité sémantique (*cf.* D. Corbin 2001a). Non qu'elle considère désormais que la formation des lexèmes est catégoriellement non contrainte, mais elle fait l'hypothèse que les contraintes catégorielles se font l'écho de l'instruction sémantique des procédés constructionnels. Les configurations catégorielles dans lesquelles entrent ou n'entrent pas les procédés sont désormais assujetties au sens, et rien ne s'oppose plus à ce qu'un procédé donné entre dans plusieurs configurations, si son sens le lui permet.

Ce changement de perspective, dans les derniers travaux de Danielle Corbin, est intéressant à plusieurs égards :

– du point de vue des analyses menées, il n'est plus nécessaire de restituer des stades virtuels plus ou moins acrobatiques pour satisfaire un rapport catégoriel présumé unique, ou de démultiplier les procédés en autant d'homomorphes qu'on observe de rapports catégoriels instaurés, même quand on pressent une similitude sémantique entre eux ;

---

73. Critiquant le principe d'unicité catégorielle en tant que conduisant parfois à reconstituer des stades virtuels, Rainer (1997 : 233-234) écrit que « [s]i l'unicité de la catégorie lexicale de la base existe bel et bien comme tendance statistique dans la formation des mots, on ne saurait en faire un principe dont la non-observation ferait trembler les fondements mêmes de la discipline ». Il resterait à vérifier qu'il s'agit bien d'une tendance statistique.

74. Pour les suffixes évaluatifs, Dal (1991) ne me semble pas étranger à ce revirement. Pour les préfixes, les recherches de Dany Amiot sur le préfixe *pré-* à l'époque ont, de leur côté, certainement également participé de l'assouplissement de la contrainte d'unicité catégorielle. Les différents travaux qu'elle a réalisés depuis confirment le fait que les préfixes spatio-temporels sont transcatégoriels (*cf.* notamment Amiot 1997 et 2003).

– du point de vue de l’adéquation des analyses avec la compétence des locuteurs, le changement de perspective est compatible avec le constat que, paradoxalement peut-être, le sens a plus de réalité du point de vue du locuteur que l’appartenance catégorielle stricte exprimée en termes de ‘nom’, ‘adjectif’ et ‘verbe’. En effet, même s’il existe, dans les langues du monde, des propriétés stables pour le nom et le verbe prototypiques (cf. Croft 1991, qu’utilise [20] : j’en reparlerai dans le § 2.3.2.), les catégories lexicales sont des constructions théoriques métalinguistiques instables dès qu’on s’écarte des cas centraux, alors que le sens, au moins organisé en grandes classes (objets, individus, propriétés, actions), est une donnée de l’expérience de chacun<sup>75</sup>. Pour affirmer cela, je me fonde sur la difficulté que le linguiste éprouve à étiqueter des expressions dès qu’elles s’écartent d’un fonctionnement prototypique<sup>76</sup>. Ainsi, l’adjectif non prédicatif *pétrolière* en (21a) est-il encore un adjectif à part entière, ou seulement un “pseudo-adjectif”, pour reprendre une étiquette qui a cours ? Le nom propre *Picasso* en position d’attribut sans déterminant en (21b) précédé de l’adverbe *très* est-il encore un nom, ou bascule-t-il dans la catégorie des adjectifs ? Il s’agit là de questions de linguiste, que le locuteur ne se pose certainement pas. Pour ce dernier, en (21a), *pétrolière* exprime l’existence d’une relation entre deux référents, celui de *crise* et celui de *pétrole*, en (21b), *Picasso* exprime une propriété, celle à laquelle on réduit stéréotypiquement la peinture de Picasso, et, dans l’usage effectif de sa langue, il lui importe peu de doter les séquences concernées d’une étiquette catégorielle :

- (21) a Affaiblie par la crise pétrolière des années 70, Lamborghini a connu plusieurs changements de propriétaires avant d’accéder de nouveau à la stabilité.  
 b Un Christ Pantocrator me subjugué : il fait très moderne, très Picasso.

Le fait que le contour des catégories lexicales change régulièrement (par ex. l’adjectif, sorti de la catégorie des noms en 1770 par l’Abbé Girard, pour y revenir quelques temps, pour en ressortir de nouveau, sans que le consensus soit aujourd’hui encore tout à fait établi, ou le participe, rentré, lui, dans la catégorie du verbe au début du XX<sup>e</sup> siècle après des siècles d’autonomie comme catégorie de mots à part entière<sup>77</sup>) montre également combien les catégories sont des constructions théoriques, éloignées des préoccupations des locuteurs dans leur rapport quotidien au lexique, construit ou non construit ;

– du point de vue de la théorie, ce glissement d’une unicité catégorielle vers une unicité sémantique a une conséquence sur la conception même des RCM et leur architecture telle que je l’ai rappelée dans le § 1.2. Admettre, avec D. Corbin (2001a : 58), que,

---

75. Dans le domaine de l’acquisition, il semble que les locuteurs immergés dans un pays dont ils ne parlent pas la langue, et qui sont en situation de devoir l’apprendre par leurs propres moyens, se créent assez rapidement un système d’expressions organisées selon ces grandes classes sémantiques, indépendamment de leur catégorie. Je me fonde ici sur des données avancées par Clive Perdue, lors d’une conférence qu’il a faite le 7 mai 2004, dans le cadre du séminaire d’équipe de SILEX.

76. Cette difficulté se rencontre dans les systèmes TAL, quand il s’agit pour eux d’étiqueter des corpus (cf. par ex. Clément 2001 : chap. 3).

77. Le participe constitue une des 8 classes de la première classification élaborée dans la grammaire occidentale, attribuée à Denys le Thrace (2<sup>e</sup>, 1<sup>e</sup> siècles avant J.-C.).

« sauf intervention d'une contrainte plus forte, l'instruction sémantique dont les affixes sont porteurs à la fois détermine leur instruction catégorielle, et donc la catégorie des mots qu'ils servent à construire », revient à admettre qu'il existe une relation de biunivocité entre un procédé et une règle. Dès lors, l'opération structurelle instaurant un rapport catégoriel entre la base et le dérivé (composant 1) n'a plus de raison d'être, puisque chaque procédé détermine ses propres configurations catégorielles, à partir de sa propre instruction sémantique ; l'opération sémantique supposée valoir pour un ensemble de procédés (composant 2) perd de sa légitimité, puisque « chaque affixe est doté d'une identité sémantique propre » (D. Corbin 2001a : 43) ; la notion de paradigme morphologique (composant 3), dont faisait l'hypothèse Danielle Corbin encore en 1993, devient caduque.

S'agissant de ce dernier composant, il y avait en tout état de cause dès le départ une contradiction fondamentale à rassembler des procédés en paradigmes et à revendiquer pour eux une individualité sémantique, comme le faisait déjà D. Corbin (1991b). Par ailleurs, et comme je le fais observer dans [20 : 298, n. 8], poussée jusqu'au bout, la logique des regroupements en paradigmes est contradictoire avec l'hypothèse d'une compositionnalité du sens des mots construits par rapport à leur structure. En effet, de deux choses l'une :

– ou la structure morphologique est indifférente à la manière dont se réalisent formellement les règles de construction de mots, et les fonctions sémantiques associées aux règles n'en tiennent pas compte (on déconnecte alors forme et sens) : il est dans ce cas légitime de poser l'existence de la RCM<sub>PROP</sub> ('règle de construction de mots formant des noms de propriété'), de la RCM<sub>ÉVAL</sub> ('règle de construction de mots formant des mots à valeur évaluative'), etc., et d'associer à chacune un ensemble de formes susceptibles de les incarner (suffixes *-ité*, *-erie*, *-esse*, etc. pour la première, suffixes *-asse*, *-et(te)*, *-in*, etc. pour la seconde). Comme le remarque Fradin (2003a : 113-114), ce point de vue n'est pas si éloigné que cela de celui de Beard (1995 : 42 *sqq.*), et de son opération d'épel (*spelling operation*), qui relève du composant phonologique, et dont le rôle est d'épeler la bonne forme affixale ;

– ou ces réalisations formelles participent de la structure morphologique, et le regroupement des procédés en paradigmes perd de sa pertinence. On distingue dans ce cas une règle par procédé : RCL<sub>age</sub>, RCL<sub>erie</sub>, RCL<sub>asse</sub>, etc, où 'RCL' se lit 'règle de construction de lexèmes'. C'est, me semble-t-il, à cette dernière solution qu'aboutit le changement de perspective dans le modèle de Danielle Corbin relaté dans [5], que je viens d'exposer ici <sup>78</sup>, et c'est également cette solution qu'adoptent, implicitement ou explicitement, une bonne partie des morphologues actuellement (par ex. Plag 1999 : 85 ; Fradin 2003a : 114). Il faut toutefois que la théorie enregistre, d'une façon ou d'une autre, le fait que plusieurs règles de construction de lexèmes partagent une partie de leurs instructions sémantiques, dans la mesure où on en a besoin pour expliquer les cas de supplétion affixale mis au jour par plusieurs des chercheurs de l'ERSS : par ex. *avionneur* vs *\*avionnier* (Roché 1997, repris et développé dans [21 : 331 *sq.*]),

78. Pour une remise en cause de la notion de paradigme constructionnel dans le modèle de Danielle Corbin, cf. également Roger (2003).

*bougeoir* vs \**bougier* (Roché *ibid.*), *moliéresque* vs \**moliérien* (Lignon 2000)<sup>79</sup>, etc. Quant à justifier l'existence même de RCL proches sémantiquement, l'explication est relativement simple. Elle tient au fait qu'il existe de multiples façons de présenter une propriété, ou de présenter la relation qu'entretiennent les référents de deux noms, ou encore de marquer son point de vue. On comprend dès lors qu'on ait voulu y voir une preuve de l'existence de règles comme les RCM<sub>PROP</sub> et RCM<sub>EVAL</sub> citées ci-dessus. En revanche, il n'existe que peu de façons de présenter un instrument ou un agent par rapport à un procès. On note d'ailleurs, qu'en dehors de quelques grandes règles comme celles qui précèdent, jamais Danielle Corbin, ni d'ailleurs aucun des morphologues proposant des règles comparables, n'a posé l'existence d'une RCM<sub>INSTR</sub> ('règle de construction de mots formant des mots d'instruments'), ou d'une RCM<sub>AGENT</sub> ('règle de construction de mots formant des mots d'agents'), ce qu'on comprend : comme leur paradigme aurait été réduit à un singleton, il était plus rapide de parler, pour le français, du suffixe *-oir(e)* ou du suffixe *-eur*.

Même si le changement de perspective n'est pas sans incidence sur la conception même des RCM telles que les décrivait Danielle Corbin, il y a un gain, pour la théorie, à abandonner le principe d'unicité catégorielle, même assoupli, pour le remplacer par un principe d'unicité sémantique. D. Corbin (2001a : 65-66) le souligne d'ailleurs explicitement, après avoir rappelé que le principe d'unicité catégorielle était entre autres motivé par le « désir de mettre de l'ordre dans un domaine où la description se contente souvent de calquer sans recul l'observation la plus immédiate ».

### 2.3.2. Les catégories vues par les règles de construction de lexèmes

Dans le changement de perspective qui vient d'être relaté, l'étiquetage catégoriel des lexèmes impliqués dans la formation du lexique construit exprimé en termes de 'noms', 'adjectifs' et 'verbes' demeure, même s'il est désormais assujéti au sens. La seconde partie de [20] aborde cette question, en se demandant comment les règles de construction de lexèmes appréhendent les catégories.

Les étiquettes de 'nom', 'adjectif' et 'verbe' font en effet partie du bagage métalinguistique quotidien. Pourtant, quand on les utilise, on ne sait pas toujours exactement de quoi on parle, alors même que le classement des mots de la langue en parties du discours est une préoccupation qui remonte à l'Antiquité grecque (pour une remarque simi-

79. Sur la concurrence entre suffixes, cf. aussi Plénat (2000 : 46 *sqq.*).

On observe que les paradigmes morphologiques dont faisait l'hypothèse Danielle Corbin n'étaient pas toujours en mesure d'expliquer ces cas de supplétion, dans la mesure où *-ier* et *-eur*, *-ier* et *-oir(e)* ne relevaient pas de la même RCM (rapports catégoriels et opérations sémantiques différents). S'agissant des supplétions possibles *-ier/-eur* et *-ier/-oir(e)*, elles indiquent que les notions sémantiques d'agent et d'instrument, communes à une partie des noms en *-ier* et en *-eur*, et à une partie des noms en *-ier* et en *-oir(e)* ont davantage de saillance cognitive du point de vue du locuteur que les catégories que mettent typiquement en relation ces différents suffixes. Cette remarque confirme l'observation que j'ai faite plus haut selon laquelle le sens a davantage de réalité pour le locuteur que les catégories mises en relation.

laire, cf. Baker 2003 : 1). Ainsi m'a-t-il semblé intéressant d'exploiter la théorie des catégories de Croft (1991), que j'ai découverte à l'occasion de la lecture de Fradin (2003a) et des comptes rendus que j'en ai faits (cf. [18] et surtout la première partie de [20]), pour tenter de répondre à la question initiale.

Croft (1991), que reprend dans ses grandes lignes Croft (2003), se situe dans une perspective typologique des langues. Il analyse les catégories du nom (désormais 'N'), de l'adjectif ('A') et du verbe ('V') selon deux dimensions. Il leur assigne :

- une fonction pragmatique (*propositional act function* : Croft 2003 : 184), chaque catégorie présentant le concept encodé selon un mode spécifique : respectivement référence, modification, prédication ;
- une classe sémantique (respectivement objets, propriétés, actions), déterminable à partir de la valence, la stativité, la persistance et la gradabilité : par exemple, la classe sémantique 'objets' se définit par la conjonction de propriétés 'valence 0', 'stativité', 'persistance', 'non-gradabilité'.

Si l'on croise les fonctions pragmatiques et les classes sémantiques, on parvient à un tableau à neuf cases (cf. tableau 2 ci-dessous emprunté à Croft 2003 : 185), dont trois sont dites « non marquées ». Les catégories 'N', 'A', 'V' prototypiques correspondent, dans la perspective typologique évoquée, aux corrélations non marquées suivantes :

- nom = <objet+référence> ,
- adjectif = <propriété+modification> ,
- verbe = <action+prédication> .

	Reference	Modification	Predication
<i>Objects</i>	UNMARKED NOUNS	genitive, adjectivizations, PPs on nouns	predicate nominals, copulas
<i>Properties</i>	desadjectival nouns	UNMARKED ADJECTIVES	predicate adjectives, copulas
<i>Actions</i>	action nominals, complements, infinitives, gerunds	participles, relative clauses	UNMARKED VERBS

Tableau 2 (Croft 2003 : 185).

Toutefois, comme le montre ce même tableau, les expressions linguistiques qu'on a coutume d'appeler 'noms' et 'adjectifs'<sup>80</sup> sont à proprement parler orthogonales aux trois classes sémantiques de Williams Croft. Le tableau 2 montre en effet :

- (colonne « Reference ») que le nom peut appartenir (i) à la classe sémantique des objets (« unmarked nouns »), par ex. *roi* que cite Fradin (2003a : 17) en guise d'illustration, (ii) à la classe sémantique des propriétés (« desadjectival nouns »), par ex. *blancheur* (*ibid.*) et (iii) à la classe sémantique des actions (« action nominals ») : par ex. *réduction* (*ibid.*) ;

80. Pour les langues, naturellement, qui connaissent l'adjectif, ce qui n'est pas le cas de toutes, par ex. le Mohawk (Baker 2003 : 4 *sqq.*), dans lequel un mot comme *hutsi* ('noir') relèverait de la catégorie des verbes. Toutes les langues du monde connaissent en revanche les catégories du nom et du verbe (Croft 2003 : chap. 3), même si elles ne les définissent pas de la même façon (*ibid.* : 183).

– (colonne « Modification ») que les adjectifs traversent pareillement les trois classes sémantiques posées, si on accepte de voir dans les participes des cas d'espèce d'adjectifs. Les exemples que cite Fradin (2003a : 17) pour illustrer cette colonne sont respectivement *royal*, *blanc*, et *réduit / réductible*, pour la dernière case.

S'agissant de ce qu'on appelle 'verbe', le cas est un peu différent puisqu'on ne le trouve comme tel que dans la dernière case. Parallèlement, la fonction pragmatique de prédication (dernière colonne) est, elle aussi, différente des deux autres, puisqu'elle peut être instanciée par des noms (case <objet+prédication>) si j'interprète bien les « predicate nominals » du tableau, des adjectifs (case <propriété+prédication>), et, naturellement, des verbes (case <action+prédication>).

Le schéma (1) ci-dessous, emprunté à [20], transpose ce qui précède dans une représentation inspirée de celles qu'utilise le lexique hiérarchique, à ceci près que ne peuvent être appariés que des traits repérés par le même chiffre ou par la même lettre, de façon à éviter les inférences illicites (par ex. <objet+prédication>, pour les adjectifs). Les appariements s'effectuent en prenant un trait dans chaque dimension. Ainsi, pour les noms, 'référence', noté '123', vaut pour les objets (1), les propriétés (2) et les actions (3) ; pour les adjectifs, 'modification', noté 'ac', vaut pour les propriétés (a) et les objets (c). Les traits en gras correspondent aux catégories non marquées. Le trait en pointillés marque la réserve que j'ai faite ci-dessus pour les « predicate nominals ». Pour le verbe, je me suis limitée au cas prototypique :

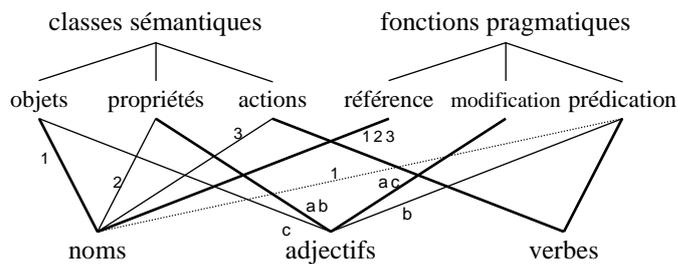


Schéma 1. Représentation des catégories de Croft à la manière du lexique hiérarchique.

En syntaxe, une appréhension globale des catégories, sans distinction des deux dimensions qu'elles cristallisent, constitue parfois le bon niveau de généralité. Par exemple, en français, tout nom demande un déterminant, tout adjectif peut figurer au sein d'un syntagme nominal. Dans le domaine de la morphologie constructionnelle, en revanche, les catégories ne constituent jamais le bon niveau de généralité puisqu'aucun procédé constructionnel ne s'applique jamais à l'ensemble des éléments d'une catégorie, même les moins exigeants d'entre eux.

On remarque en outre que, si on s'en tient aux catégories que donnent à voir les données observables sans chercher à restituer d'étape intermédiaire pour établir une catégorisation unitaire, il est fréquent qu'un même affixe s'applique à plusieurs catégories de bases, certains pouvant même former plusieurs catégories de dérivés. On l'a déjà vu dans les pages qui précèdent :

– avec le suffixe *-able*, qui s'applique à des noms et à des verbes (*ministrable*, *portable*),

– avec les suffixes *-ifi-* et *-is-*, qui s’appliquent à des noms et à des adjectifs (*choséfifier*, *cocufier* ; *macadamiser*, *humaniser*),

– avec les préfixes *a-*, et *dé-*, tous deux compatibles avec des adjectifs, des noms et des verbes (*alourdir*, *accouder*, *amener* ; *déjaunir*, *désosser*, *décoincer*), *dé-* étant en outre apte à former des verbes, mais aussi, même si le cas est plus rare, des adjectifs (*déloyal*), et des noms (*déshonneur*).

Je pourrais encore citer le préfixe *in-*, compatible avec des adjectifs pour former des adjectifs (*pur* > *impur*), mais aussi avec des noms pour former des noms (*harmonie* > *inharmonie*). Les dictionnaires attestent une grosse quinzaine de noms en *in-* qu’il est difficile d’analyser autrement que construits sur des noms<sup>81</sup>. Les corpus électroniques montrent pour leur part la disponibilité du schéma, par ex. *inadversité*, que j’ai relevé dans *Libération* :

Et la Suisse, entre soulagement et inquiétude, a vu, samedi à Key Biscane, la confirmation que l’inadversité dans laquelle elle enfilait ses victoires comme des perles avait duré.

Des remarques similaires à celles que je viens de faire ont conduit certains linguistes (par ex. Serbat 1989 ou Plag 1999) à considérer que les procédés constructionnels s’appliquent à des bases catégoriellement sous-spécifiées :

– pour Serbat (1989 : 4), les bases de construction de lexèmes sont le « support d’une ‘notion’ avant que celle-ci ne soit catégorisée syntaxiquement » ;

– pour Plag (1999 : 122), « the syntactic category of the base is underspecified » (cf. aussi pp. 131 sq.).

Cette solution pose un certain nombre de problèmes que je liste dans [20 : 315 sq.]. En particulier, considérer que les bases de construction de lexèmes ne sont pas étiquetées revient à leur récuser le statut de lexèmes, puisque, par définition, ces derniers possèdent une étiquette catégorielle. Or, un consensus se dégage depuis Anderson (1992) pour considérer que le lexème est l’unité de base de la morphologie, avec, le cas échéant, des amendements à la définition initiale de la notion (je renvoie ici aux travaux de Françoise Kerleroux et de Bernard Fradin, seuls ou en binôme, qui visent à montrer que le lexème est sous-spécifié pour la flexion, mais qu’il est entièrement spécifié pour le sens<sup>82</sup>). Par ailleurs, considérer que les bases de construction de lexèmes ne sont pas catégorisées implique que les entrées et les sorties des règles de construction de lexèmes ne soient pas du même type, puisque, à ma connaissance, personne n’a jamais avancé l’hypothèse d’une sous-spécification catégorielle des dérivés. En outre, cela suppose qu’on décatégorise les sorties des règles de construction de lexèmes quand, à leur tour, elles sont utilisées comme bases (par ex. *pauvre* > *appauvrir* > *appauvrissement*).

---

81. Par ex. *incivilisation*, *incommunication*, *inconfort*, *inhabitude*, *invraisemblance*, etc. On note que tous sont apparus après le milieu du XVIII<sup>e</sup> siècle (respectivement 1795, 1786, 1893, 1910, 1763 pour les exemples qui précèdent). *Inharmonie* date de 1765. Je fais l’hypothèse, provisoire, que le préfixe *in-* français marque l’absence d’une propriété ou d’un procès attendus. Une étude fine de ce préfixe reste à faire.

82. Cf., entre autres, Fradin (2003a), Fradin & Kerleroux (2003b), Kerleroux (sous presse).

Comme d'autres, je considère donc dans [20] que les bases de construction de lexèmes sont pourvues d'une étiquette catégorielle, donc que ce sont bien des lexèmes.

Je fais en revanche l'hypothèse que les règles de construction de lexèmes sont capables de distinguer, dans les catégories, ce qui relève de leur dimension sémantique et ce qui relève de leur dimension pragmatique, aussi bien pour les catégories des bases que pour celles des dérivés. On ne pourrait pas expliquer, sinon, les appariements originaux que sont capables de produire certaines règles constructionnelles, comme <objet+modification> pour ce que la syntaxe appelle 'pseudo-adjectifs' et la morphologie 'adjectifs relationnels' (par ex. (*activité*) *pétrolière*, (*compte*) *bancaire*, (*chèque*) *postal*), ou comme <propriété+référence>, pour ce qu'on a coutume d'appeler 'noms de propriété'<sup>83</sup>.

J'examine dans [20] un certain nombre de contre-arguments à cette hypothèse, et j'essaie notamment d'expliquer les préférences catégorielles que l'on peut observer pour tel ou tel affixe. Par exemple, si l'on se fonde sur les données chiffrées issues des analyses menées dans le cadre du projet MorTAL, le préfixe *dé-* s'applique préférentiellement à des noms (402 lexèmes). Viennent ensuite les verbes (256 lexèmes), puis, loin derrière, les adjectifs (38 lexèmes). Dans le cas de *dé-*, cette préférence pour le nom s'explique facilement : on conçoit plus facilement la notion de dissociation quand elle implique un objet que quand elle implique une propriété. Or, la classe sémantique des objets correspond typiquement au nom. De façon générale, je fais l'hypothèse dans [20] qu'une fois déterminée la base sémantiquement appropriée, les procédés constructionnels choisissent prioritairement des bases correspondant à des corrélations non marquées (principe 1). L'appartenance de la base à une corrélation non marquée constituerait donc un cas par défaut :

**Principe 1 :** Quand il existe plusieurs bases sémantiquement appropriées, choisir préférentiellement celle dont la catégorie correspond à une corrélation non marquée.

J'ajoute un second principe dans [20] pour expliquer qu'il ne s'agit parfois que de préférences, et que, dans certains cas, soit sélectionnée une corrélation marquée :

**Principe 2 :** Quand il existe deux bases sémantiquement appropriées pour construire un lexème donné, et que l'une est analysable comme construite sur l'autre, choisir préférentiellement la moins complexe des deux.

Ce second principe permet élégamment d'expliquer les choix que nous avons faits dans le projet MorTAL quand nous avons décidé de construire sur des noms les verbes en *-is-* pour lesquels deux types catégoriels de base étaient envisageables, un nom et un adjectif qui en dérive (cf. *supra*, § 2.2.3.3.). De fait, tout en correspondant à la corrélation marquée <propriété+référence>, *harmonie*, *idéologie*, *mélancolie* ont un degré de complexité moindre que les adjectifs qui leur correspondent (respectivement *harmonieux/harmonique*, *idéologique*, *mélancolique*). Ce principe, qu'il reste à tester sur davantage de données, évite dans certains cas de recourir aux artefacts des stades virtuels, à seule fin de préserver une catégorie de base présumée.

---

83. L'appariement <propriété+référence> se retrouve en français dans des lexèmes non construits (par ex. *misère* ou *harmonie*), mais ils sont rares.

Le dernier contre-exemple qu'examine [20] est celui de la conversion dite « catégorielle » dans Fradin (2003a : 157), puisqu'on la définit en général comme une opération constructionnelle permettant de construire une catégorie de lexèmes à partir d'une autre catégorie de lexèmes, sans adjonction de matériau constructionnel (par exemple en français, l'adjectif chromatique *orange* à partir du nom de fruit homomorphe). Par définition donc, elle présuppose un écrasement de la bidimensionnalité des catégories telles que les définit Williams Croft, qu'il s'agisse de celles des bases ou de celles des dérivés. L'une des solutions, que je ne fais qu'esquisser dans [20], consiste à passer en revue chaque corrélation possible et impossible du côté des bases comme du côté des dérivés. On peut d'ores et déjà constater que :

– dans certains cas, la conversion joue sur la seule dimension pragmatique sans affecter la dimension sémantique : par ex. *garder* <action+prédication> → *garde* “action de garder” <action+référence><sup>84</sup> ; *commode* <propriété+prédication> → *commode* “meuble” <propriété+référence> ;

– dans d'autres, elle joue sur la dimension sémantique, sans affecter la dimension pragmatique : par ex. *calme* <propriété+prédication> → *calmer* <action+prédication> ;

– dans d'autres encore, elle affecte les deux dimensions : par ex. *garder* <action+prédication> → *garde* “personne qui garde” <objet+référence>, si l'on accepte de voir dans les individus des cas particuliers d'objets.

Il ne s'agit là que d'une esquisse, qui demande à être travaillée. Elle montre cependant que même le type de données le plus hostile à la proposition que j'avance est passible d'une explication, indiquant que les règles construction font le départ, dans les catégories, entre leur dimension sémantique et leur dimension pragmatique.

En considérant que les règles de construction de lexèmes n'appréhendent pas les catégories comme des tous, une bonne partie des apories auxquelles conduit une catégorisation des bases en termes de 'N', 'A', 'V' se dissolvent d'elles-mêmes. À cet égard, la théorie des catégories de Williams Croft, pour qui les catégories résultent de la cristallisation d'une classe sémantique et d'une fonction pragmatique, s'avère une approche prometteuse, puisqu'aux trois catégories classiques, elle substitue potentiellement neuf appariements possibles.

Il reste que, si l'approche croftienne permet une granularité plus fine, aussi bien pour les bases que pour les dérivés, elle ne suffit pas à rendre compte de tous les phénomènes relevant de la morphologie constructionnelle, ce dont on n'a pas à lui faire reproche puisque ce n'est pas à cette fin qu'elle a été conçue. En particulier, la partition en trois classes sémantiques, 'objets', 'propriétés', 'actions', doit être complétée d'une typologie sémantique, qui peut être interne à chacune de ces classes (ce qui a été dit à propos du projet MorTAL § 2.2.1. montre que la classe des propriétés doit au moins être subdivisée en propriétés inhérentes et propriétés acquises pour rendre compte de

---

84. Le nom *garde*, quand il désigne l'action de garder, appartient à la même classe sémantique que sa base verbale. L'opération de conversion n'est toutefois pas neutre sémantiquement, dans la mesure où, comme le note Fradin (2003a : 160), la structure argumentale du nom converti comporte souvent un argument en moins que celle du verbe base, cet argument correspondant en général à l'agent.

l'incompatibilité des premières avec les suffixations par *-ifi-* et par *-is-* : \**jeunifier*, \**mangeabiliser*), voire leur être transversale. Ainsi, il semble que la morphologie constructionnelle ait besoin d'un trait 'part' ("partie" ; cf. Jackendoff 1993), qui vaille pour les objets et les propriétés, pour rendre compte du marquage de dissociation qu'effectue le préfixe *dé-*. Le projet MorTAL et l'acquisition automatique des traits sémantiques dont il s'est agi plus haut (cf. *supra* § 2.2.2.) fournissent à cet égard des éléments de réponse précieux.

La proposition que j'ai faite dans [20] n'est donc qu'un premier jalon vers une définition des catégories telles que les voit la morphologie constructionnelle. En tout état de cause, il me semble qu'une réflexion de fond sur le rapport entre la morphologie constructionnelle et les catégories est à mener.

#### 2.4. Lexique construit et relations paradigmatiques

L'ensemble des modèles qui préconisent un traitement constructionniste (Fradin 2003b : 55 *sq.*) de la morphologie se rejoignent en ceci qu'ils reposent sur une procédure fondamentalement syntagmatique : dans ces modèles, le sens et la forme d'un lexème construit sont décrits comme résultant de l'application de fonctions, sémantiques et formelles, à un lexème base<sup>85</sup>. Ils ne sont par conséquent pas outillés pour rendre compte des relations paradigmatiques éventuellement repérables entre lexèmes, en dehors des relations de bases à dérivés, sauf à mettre en place des dispositifs particuliers. C'est le cas, dans le modèle de Danielle Corbin, du mécanisme d'intégration paradigmatique, encore dit 'marquage de classe', que j'ai déjà évoqué au § 1.1., et qui est destiné à expliquer une distorsion entre un sens construit et une forme observable, le sens ayant préséance sur la forme.

L'objectif que s'assignent les modèles constructionnistes en général est de rendre compte de la façon dont est construit le lexique construit. Au fil du temps et des lectures que j'ai pu faire, il m'est cependant apparu qu'en procédant de la sorte, ils oublient la seconde spécificité du lexique construit, qui est qu'il s'agit de lexique, et non de syntaxe. Or, il est établi que chaque locuteur dispose en mémoire d'un certain nombre de lexèmes<sup>86</sup> – c'est ce qu'il est convenu d'appeler le 'lexique mental' –, et que, parmi ces lexèmes, certains sont construits. En effet, après une période où on a pensé que les

85. La distinction entre les facettes sémantique et formelles des règles constructionnelles n'est pas toujours aussi clairement exprimée que dans Fradin (2003a). Pour ce dernier, les règles de construction de lexèmes disposent de façon simultanée de toutes les informations stockées dans les lexèmes d'entrée : à ce titre, elles peuvent régir des fonctions susceptibles de s'appliquer sur chacune des rubriques d'un lexème d'entrée donné (rubriques graphémique, phonologique, syntaxique, morphologique, sémantique). Pour le détail, je renvoie à [20 : 297 *sqq.*]. Pour la différence, que je ne fais pas ici, entre règles et opérations morphologiques, cf. Fradin (2003a : 116 *sqq.*), qui reprend la distinction de Zwicky (1992).

86. Aitchison (1994 : 7) évalue à 50 000 le nombre de mots qu'un locuteur anglais adulte éduqué peut comprendre ou potentiellement utiliser, soit la taille d'un dictionnaire mono-volume déjà bien fourni (à titre d'indication, le *PRO1* compte environ « 60 000 mots », d'après la préface du *Nouveau Petit Robert*, par Josette Rey-Debove et Alain Rey). Il en déduit que le lexique mental doit être organisé en mémoire, d'autant plus que l'accès peut être quasi-instantané (quelques centaines de millisecondes, selon les cas).

lexèmes complexes<sup>87</sup> étaient tous décomposés en mémoire, quelle que soit leur fréquence, sous la forme de règles symboliques (cf. par ex. Taft & Forster 1975), la tendance, chez les psycholinguistes, est actuellement aux modèles combinant règles symboliques et stockage en mémoire (*dual route models*) voire, pour les plus extrémistes d'entre eux, stockage en mémoire seulement (*single route models*) :

– Le linguiste est, par nature, un être économe : du point de vue de la théorie, moins un traitement linguistique est redondant, plus il est jugé élégant, donc souhaitable. On note d'ailleurs que le coût du système est parfois l'ultime argument auquel on recourt, en cas d'indécidabilité entre deux traitements linguistiques concurrents (cf. par ex. D. Corbin 1976). Aussi, pour le linguiste habitué à raisonner en termes de coûts et d'économie, l'option pour un stockage en mémoire est, en première analyse, plus coûteuse que l'option pour une décomposition systématique, selon des règles symboliques. À cet égard, Frauenfelder & Schreuder (1992) montrent que, du point de vue du locuteur, il convient de distinguer deux types d'économie : une économie de stockage et une économie de traitement. Dans la perspective du stockage, le principe qui prévaut est celui d'une réduction du nombre d'entrées en décomposant ce qui peut l'être, accompagnée, le cas échéant, d'une organisation rationnelle du lexique construit en familles. Dans la perspective du traitement, le principe est celui d'une mise en mémoire sans décomposition avec, ici aussi, des principes d'organisation rationnelle. Il est en effet plus facile, partant plus rapide, d'accéder à un mot entier que de devoir en retrouver les constituants, puis devoir activer la règle dont il est le produit, en évitant les chausse-trappes en cas d'analyses *a priori* concurrentes ainsi que les leurres que sont les pseudo-affixes et les pseudo-bases (on retrouve là les problèmes qu'a dû résoudre DériF dans le cadre du projet MorTAL). Aucune expérimentation décisive n'ayant mis selon eux en évidence la primauté d'un de ces deux principes sur l'autre, Frauenfelder & Schreuder (1992) optent pour une solution de compromis, et préconisent, à la suite d'autres psycholinguistes, un *dual route model*. En situation de reconnaissance, les deux chemins d'accès, l'un par décomposition, l'autre par recherche globale en mémoire, seraient ainsi activés de façon concomitante<sup>88</sup>, la victoire de l'un ou de l'autre dépendant notamment de la fréquence en termes de *token* et de la transparence. Ce premier grand groupe de modèles prend, le cas échéant, également en considération la fréquence relative. C'est par exemple le cas de Hay (2001), qui fait l'hypothèse que sont stockés en bloc dans le lexique mental les lexèmes construits dont la fréquence est supérieure à celle de leur base, par ex., en anglais, *illegible vs legible*, et que font l'objet d'une décomposition les lexèmes construits moins fréquents que leur base, par ex. *illiberal vs liberal*. Comme exemple de modèle à deux voies, je citerai également le modèle de Pinker & Prince (1994), davantage axé sur la flexion. Selon ce modèle (p. 326) :

---

87. Je parle ici de lexème *complexe* et non de lexème *construit*, dans la mesure où l'hypothèse du double accès, en bloc et par décomposition, vaut chez certains psycholinguistes à la fois pour la flexion et pour la dérivation.

88. D'après Ullman (2001), le lexique mental et la grammaire mentale ne sont pas gérés par les mêmes zones du cerveau : le premier est stocké dans le lobe temporal, tandis que la seconde réside dans le cortex frontal.

Regulars are computed by an implementation of a classic symbolic rule of grammar, which concatenates an affix with a variable that stands for the stem. Irregulars are memorized pairs of words, but the linkages between the pair members are stored in associative memory structure with certain connectionist-like properties (cf. Bybee 1988).

– Les modèles de Bybee (1985) et, quelques années plus tard, de Skousen (1989), complété par Skousen (1992), peuvent être cités comme instances de *single route models*, optant pour un stockage massif en mémoire. Skousen (1989) constitue le premier modèle analogique explicitement formulé. Derwing & Skousen (1994) en résument les grandes lignes directrices : en deux mots, il est basé sur des supracontextes au comportement homogène, assorti de ce qu'on appelle un 'effet de gang' (*gang effect*)<sup>89</sup>.

Ce passage, en psycholinguistique, des modèles reposant sur des règles symboliques à des modèles de compromis voire à des modèles récusant l'idée même de règle correspond à une tendance générale en psychologie, qui s'accélère depuis deux petites décennies : ainsi, jusqu'au début des années 1980, les psychologues en général faisaient l'hypothèse que ce qu'ils appellent l'« apprentissage implicite » s'effectuait en abstrayant des règles à partir de l'exposition à des occurrences particulières (de situations, de comportements, etc.). À partir de 1980, cette interprétation abstractionniste de l'apprentissage a peu à peu cédé le pas soit à une interprétation uniquement exemplariste, soit à une interprétation mixte, combinant mise en mémoire d'occurrences particulières (de situations, de comportements, etc.) et abstraction de schémas à partir de l'exposition aux occurrences particulières<sup>90</sup>.

En tout état de cause, et pour en revenir à mon propos initial, il n'est par conséquent pas déraisonnable de penser que cette partie mémorisée du lexique que constitue le lexique mental puisse influencer la construction des lexèmes, en dehors de la relation privilégiée de base à dérivé. On rejoint là l'hypothèse des correspondances multiples dont fait l'hypothèse Luigi Burzio dans plusieurs de ses travaux (cf. en particulier Burzio 1998 et 2000).

Trois de mes travaux, dont deux sont sous presse, s'intéressent aux relations paradigmatiques observables dans le lexique construit, et explorent deux pistes alternatives aux règles conçues sur le mode des règles de la grammaire générative : le premier, [21], écrit en collaboration avec Fiammetta Namer, s'inspire de la théorie de l'optimalité pour expliquer une étrangeté dont un modèle basé sur règles peut difficilement rendre

---

89. Sur l'effet de gang, cf. aussi Fehring (2003), qui voit par exemple dans le umlaut des noms désadjectivaux de certains dialectes allemands un umlaut analogique, copié sur celui que reçoivent le comparatif et le superlatif de l'adjectif base.

90. Pour un point récent sur la question de la mémoire et de l'apprentissage implicites, cf. Nicolas & Perruchet éds (1998). On note avec intérêt l'association parfois faite entre *implicite* et *procédural* d'une part, *explicite* et *déclaratif* d'autre part (cf. l'intitulé de l'atelier 1 des journées internationales d'Orsay sur les sciences cognitives organisées les 11 et 12 décembre 2002 : « Corrélations différenciées de l'apprentissage scolaire avec les apprentissages implicite/procédural et explicite/déclaratif »). Or, comme je l'écris dans l'introduction de cette note de synthèse à la suite de Fradin (2003a), la tendance, en linguistique théorique, est actuellement aux modèles déclaratifs énonçant les contraintes sur les entrées et les sorties, au détriment des modèles procéduraux (cf. également [20 : 304 sq.]).

compte ; les deux suivants, [16] et [22], traitent de l'analogie et de sa réémergence sous différentes formes dans les théories constructionnelles actuelles.

#### 2.4.1. Lexique construit et théorie de l'optimalité

La notion de règle est fondamentalement binaire (un objet linguistique est ou n'est pas conforme à la règle) et, de ce fait, ne peut qu'engendrer des exceptions, règles et exceptions formant « une sorte de couple infernal » (Touratier 1988 : 7-8).

Depuis une dizaine d'années cependant se dégage une autre tendance, consistant à rejeter purement et simplement la notion même de règle pour lui préférer celles de principes (dans le courant génératif chomskien), ou de contraintes, la théorie phare en la matière étant la théorie de l'optimalité.

L'objectif de [21] est précisément d'évaluer, au moyen d'exemples empruntés au français, les services que peut rendre cette dernière en matière de description du lexique construit, même si, comme le souligne Plénat (2000 : 44), elle n'a en définitive pas modifié la partition binaire des modèles basés sur règles.

La théorie de l'optimalité, désormais OT, dont Prince & Smolensky (1993) constitue le texte fondateur, repose sur quelques principes simples, dont certains renouent avec la tradition (*cf. supra*, § 2, n. 42) :

- les langues doivent satisfaire un certain nombre de contraintes universelles de bonne formation ;
- à l'intérieur d'une langue donnée, des contraintes peuvent entrer en conflit ;
- certaines contraintes peuvent être enfreintes ;
- les contraintes sont ordonnables, selon que leur infraction est ou n'est pas fatale ;
- les différences entre les langues du monde peuvent être ramenées à des différences dans l'ordonnement des contraintes.

Là où, classiquement en grammaire générative, les règles sont conçues comme des montages en série s'appliquant à un input donné pour former un output donné *via* un certain nombre d'ajustements expliquant les distorsions éventuelles entre forme profonde et forme de surface, l'OT génère en parallèle un ensemble de candidats au moyen de la fonction « GEN » (*generator*), puis les évalue au moyen de la fonction « EVAL » (*evaluator*) en recourant au système de contraintes préalablement dégagées<sup>91</sup>. Le schéma (2), emprunté à Mc Carthy & Prince (1993 : 86), résume cette procédure :

---

91. Plénat (1996) considère que le fonctionnement de l'évaluateur, lui, n'est pas entièrement parallèle, dans la mesure où « les formes candidates sont évaluées contrainte par contrainte jusqu'à ce que l'une d'entre elles satisfasse une contrainte que ne satisfasse aucune de ses concurrentes » (p. 586). Aussi préconise-t-il de substituer à l'OT la FPOT (*Full Parallel Optimality Theory*), dans laquelle « toutes les contraintes pertinentes pèsent simultanément sur le destin de chaque candidat ». Marc Plénat fait l'hypothèse que, du point de vue de la performance effective des locuteurs, la FPOT est certainement plus réaliste que l'OT, et certainement plus facile à mettre en œuvre.

An Optimality-Based Grammar, Schematically

$GEN(in_i) = \{cand_1, cand_2, \dots\}$

$EVAL(\{cand_1, cand_2, \dots\}) \rightarrow cand_k$  (the output, given  $in_i$ )

Schéma 2. Procédure de construction d'un output à partir d'un input donné dans le cadre de l'OT.

Ainsi résumée <sup>92</sup>, l'OT apparaît davantage comme une métathéorie que comme une théorie restreinte à un domaine en particulier, mais son domaine d'application privilégié, qui est aussi le domaine pour lequel elle a été conçue, est celui de la phonologie. De fait, les contraintes universelles qui ont été dégagées relèvent toutes de la phonologie, et, quand elle est utilisée en morphologie constructionnelle, c'est la plupart du temps pour expliquer des faits de phonologie (pour des exemples, cf. Plag 1999). Or, la révision de certaines données linguistiques dans un cadre inspiré de l'OT <sup>93</sup> en ceci qu'il recourt à la notion de contraintes, éventuellement transgressables, permet aussi d'expliquer des incongruités constructionnelles, qui ne sont pas (pas uniquement ?) phonologiques.

[21] examine plus précisément deux cas exemplaires d'échangisme : un cas d'échangisme affixal en français emprunté à Roché (1997), avec *avionneur* là où on attend la forme *avionnier*, un cas d'échangisme de bases, avec les noms en *-ité* français comportant un toponyme en position de base là où on attend plutôt un adjectif ethnique. Je ne reprendrai, en le développant, que ce dernier cas ici (pour le premier, je renvoie à [21 : 331 sq.] ainsi, de façon plus générale, qu'aux travaux déjà cités au § 2.3.1. sur le phénomène d'échangisme suffixal réalisés par des chercheurs de l'ERSS).

La plupart des noms de propriété en *-ité* du français sont analysables comme dérivant, à la fois formellement et sémantiquement, de bases adjectivales, qu'ils aient effectivement été construits sur la période française (par ex. *acerbité*), ou qu'ils soient les résultats de l'évolution phonétique de noms construits en latin (par ex. *absurdité*). À titre d'indication, sur les 3227 lexèmes en <ité> qu'a traités le système DériF dans le cadre du projet MorTAL, seuls une dizaine sont analysables comme dérivés de noms. En voici la liste, avec une définition extraite du *RE*, sauf mention contraire. J'ai ajouté, entre crochets, l'analyse que reçoit chacun de ces noms dans le dictionnaire consulté, quand elle existe :

(22) **angélicité** : « **Rare et littér.** Caractère angélique (Huysmans, Barrès, in T.L.F.) ». [1895, Huysmans ; du rad. du lat. *angelus* (→ Ange), et suff. *-ité*]

**avant-gardité** : « **Rare.** Caractère d'une production littéraire ou artistique d'avant-garde » [1964 ; de *avant-garde*]

**bouddhéité** : « **Relig.** Caractère de l'état de bouddha; essence du bouddha » [1930, dict. pratique des connaissances religieuses ; de *bouddha*, et suff. *-ité*. → *Déité*]

**déité** : « **Littér.** Divinité mythique; dieu ou déesse » [lat. chrét. *deitas*]

92. Pour des résumés plus circonstanciés, cf. Archangeli (1997) et Boersma & al. (2000).

93. La démarche adoptée dans [21] s'inspire seulement de l'OT parce qu'elle ne vise pas à l'universalité. Je signale par ailleurs que, dans ses derniers travaux (cf. D. Corbin 2001a et 2001b), Danielle Corbin a commencé à appliquer au lexique construit des principes inspirés de cette théorie (cf. aussi *supra*, § 1.1., n. 10).

**édilité** : « **Didactique ou administratif**. ♦1. **Hist. rom.** Magistrature de l'édile ; exercice de cette magistrature. ♦2. (1838). **Mod. Rare.** Magistrature municipale. Service municipal chargé de l'entretien des rues, des édifices, dans les grandes villes » [lat. class. *edilitas*]

**fraternité** : « ♦1. (1690). **Rare.** Parenté, lien de sang entre frères et soeurs. ♦2. (V. 1206). Lien existant entre les hommes considérés comme membres de la famille humaine; sentiment profond de ce lien. ♦3. Lien particulier établissant des rapports fraternels. 4. (XX<sup>e</sup>; de l'angl. *fraternity*). Nom de certaines communautés religieuses ou laïques » [lat. *fraternitas*]

**métallité** : « Qualité de ce qui est un métal; réunion des propriétés qui caractérisent les métaux » (LITTRÉ) » (TLF, s.v. **métall(o)-**)

**odorité** : « **Didact.** Caractère d'une substance qui détermine des sensations olfactives. *L'odorité d'une substance est liée à sa volatilité* » [v. 1970 ; dér. sav. du lat. *odor*]

**sororité** : « ♦1. **Anglic.** Groupement de femmes, équivalent d'une fraternité (4.), pour les femmes ou les jeunes filles. ♦2. (Du lat. *soror*). Lien, solidarité existant entre les femmes considéré comme spécifique par rapport à la fraternité qui unit les hommes » [1970 ; lat. *soror*, et *fraternité*, d'après l'angl. *sorority*]

**ustensilité** : « **Didact.** Caractère d'ustensile; capacité à servir comme ustensile » [Mil. XX<sup>e</sup> ; de *ustensile*]

Je ne commenterai pas cette liste, sauf à constater que trois de ces lexèmes (*déité*, *édilité* et *fraternité*) ont un correspondant latin, que *déité* a pu faire office de modèle pour *bouddhéité* et *angélicité*, que *sororité* fait, lui aussi, écho à *fraternité*, et que les dérivés restants, *avant-gardité*, *métallité*, *odorité* et *ustensilité*, sont en tout état de cause rares.

Des noms comme *africanité*, *germanicité*, *romanité*, *magyarité*, relevés dans les dictionnaires et sur la Toile, dans lesquels la base coïncide formellement avec un adjectif ethnique *modulo* une éventuelle allomorphie, peuvent être considérés comme des instances du schéma général de la suffixation en *-ité*. Cependant, si l'on se fonde sur les données extraites de la Toile, un autre patron est également disponible. Il consiste à appliquer le suffixe *-ité* directement sur le toponyme, alors même, on vient de le voir, que *-ité* a *a priori* peu d'accointance avec les bases nominales. Je citerai *belgicité* ou *portugalité*, dans lesquels une construction sur un adjectif serait inapte à rendre compte des segments <ic> et <al> qu'on y repère. Les dictionnaires n'attestent qu'un nom en *-ité* dont la base coïncide formellement avec un toponyme. Il s'agit de *francité*, dont il est impossible de dire s'il instancie le premier patron, avec troncation du suffixe *-ais* de *français*, ou le second patron, avec suffixation directe du toponyme<sup>94</sup>.

Pour essayer de voir un peu clair dans ce type de données et comprendre dans quel cas l'échangisme de bases est possible, Fiammetta Namer et moi-même avons monté l'expérience suivante : nous avons sélectionné une grosse centaine de noms propres référant à des continents, des pays ou des régions. Nous leur avons associé leur(s) adjectif(s) ethnique(s) attesté(s), qu'il(s) en soi(en)t ou non dérivé(s) (par ex. pour *Hongrie* : *hongrois* et *magyar*), puis nous avons généré automatiquement l'ensemble des noms en

94. Guilbert (1975 : 186) le dérive sans hésitation de *français*, et note par ailleurs que « la dérivation du nom à partir de l'adjectif dans cette variante fournit de nombreux néologismes : *lybianité*, *chiliani-té* ».

-*ité* dont la base était tour à tour l'adjectif (les adjectifs) ou le toponyme, le cas échéant, leurs supplétifs, en effectuant toutes les opérations phonologiques qui nous ont semblé pertinentes (allomorphies, haplogogies, troncations), sans préjuger du caractère bien formé ou non du résultat. Après cette première étape, qui simule la fonction GEN de l'OT, nous avons mis en œuvre le système WaliM (cf. Namer 2003b), de façon à vérifier et à quantifier la présence de chacun de ces noms sur la Toile.

Le tableau 3, emprunté à [21 : 333 sq.], donne un échantillon représentatif des résultats que nous avons obtenus au 15 septembre 2003, sur google France (“+” et “-” signalent la présence et l'absence du candidat généré sur la Toile) :

africanité	+
afriquité	-
africité	+
albanaisité	-
albanité	+
allemandité	+
allemandité	-
allemanité	+
germanité	+
germaniquité	-
germanicité	+
antillaisité	-
antillanité	+
antillité	-
belgité	+
belgicité	+
bretonnité	+
bretagnité	-
bretanité	-

finlandaisité	-
finlandité	+
finnoisité	-
finnité	+
gréicité	+
gréquité	-
héliénité	+
hongroisité	-
hongrité	-
magyarité	+
ivoirianité	+
ivoirité	+
portugaisité	-
portugalité	+
lusitanianité	-
lusitanité	+
québécoisité	-
québéquité	-
québécoicité	+

Tableau 3. Echantillon de noms en *-ité* à composant toponymique ou ethnique trouvés sur la Toile.

De façon générale, nous avons constaté :

– que, lorsque l'adjectif ethnique est suffixé par *-ain*, *-ien* ou *-éen*, le suffixe *-ité* le sélectionne préférentiellement : en plus d'*africanité*, *ivoirianité* ci-dessus, nous avons relevé sur la Toile *guadeloupéanité*, *marocanité*, *sicilianité*, préférés à *guadeloupité*, *maroc(qu)ité*, *sicilité*, absents. Il ne s'agit cependant là que d'une tendance, puisque la base du nom en *-ité* peut également coïncider formellement avec le toponyme, modulo d'éventuels ajustement formels, de façon exclusive (hors tableau : *acadité*, qui compte sept occurrences sur google contre zéro pour *acadianité*), ou en plus de l'autre possibilité (cf. tableau 3 : *africité*, *ivoirité*<sup>120</sup>). S'agissant des noms en *-ité* pouvant être mis en correspondance avec un adjectif en *-ien* construit sur un toponyme dont la dernière syllabe comporte un /n/, qu'il soit final : par ex. *Ukraine* (/y.kren/), suivi d'une voyelle : par ex. *Jordanie* (/ʒɔr.da.ni/), ou revête la forme d'une voyelle nasale : par ex. *Iran* (/i.rã/), ils méritent une mention particulière. On peut en effet les analyser comme dérivés d'un adjectif en *-ien* avec troncation : dans cette hypothèse, *lusitanité* (tableau 3), *estonité*, *iranité*, *jordanité*, *mauritanité*, *méditerranéité*, *palestinéité*, *ukrainité* (hors ta-

120. Et ce, pour *africité*, bien qu'*africanité* appartienne au lexique conventionnel. On contrastera quand même la fréquence de ce dernier sur google France au 15 septembre 2003 (1460 occurrences), avec celle d'*africité* (une seule occurrence comme nom commun).

bleau) sont des formes accourcies de *lusitaNIANité*, *estoNIANité*, *iraNIANité jordanIANité*, *mauritaNIANité*, *méditerranéANité*, *palestiNIANité*, *ukraiNIANité*, ou comme dérivés directement du toponyme base de l'adjectif, avec haplogogie quand le toponyme se termine lui-même par /i/ : dans cette hypothèse, *lusitanité* dérive de *Lusitanie*, *estonité*, *iranité*, *jordanité*, *mauritanité*, *méditerranité*, *palestinité*, *ukrainité* de, respectivement, *Estonie*, *Iran*, *Jordanie*, *Mauritanie*, *Méditerranéen*, *Palestine*, *Ukraine*. Nous n'avons trouvé aucun argument décisif permettant de trancher entre ces deux hypothèses, mais on trouve là une illustration de la tendance dissimilative déjà mentionnée ci-dessus au § 2.2.2., qui proscrit la consécution, dans une fenêtre étroite, de deux phonèmes identiques ;

– que, lorsque l'adjectif ethnique se termine par <ais> ou par <ois><sup>121</sup>, le nom résultant de la concaténation du suffixe *-ité* et de cet adjectif est (quasi-)systématiquement absent d'Internet. On le voit dans le tableau 3, dans lequel *albanaisité*, *finlandaisité*, *hongroisité*, *portugaisité*, *québécoisité* sont notés “-”, et, hors tableau, avec *chinoisité* (*sinoisité*), *françaisité*, *irlandaisité*, *islandaisité*, etc., tous absents de la Toile. Dans ce cas, deux stratégies s'observent : soit on recourt à un adjectif supplétif de l'adjectif en *-ais* (*-ois*), s'il existe (cf. tableau 3 : *magyarité*), soit on applique *-ité* directement au toponyme (cf. tableau 3 : *albanité*, *finlandité*, *portugalité*, *québécoisité* ; hors tableau : *aveyronnité*, *camarguité*, *hollandité*, *japonité*, *martinicité*, *réunionnité*, *rwandité*, etc.), ou à son supplétif (tableau 3 : *finnité* ; hors tableau : *sinité*). Le cas d'*antillanité* est intéressant : *antillaisité* est jugé mauvais, *antillité* (/ã.ti.ji.te/) n'est pas meilleur, certainement à cause de la rafale de /i/ et /j/ qu'il implique. Le patron qui a été choisi est le patron majoritaire des noms en /anite/, et le choix a certainement été favorisé par le fait qu'à notre *antillais* français correspond un *antillano* espagnol, qu'on retrouve d'ailleurs dans *antillanisme*, lui-même calque de l'espagnol *antillanismo* (je signale aussi que l'adjectif ethnique français correspondant au toponyme *Antilles* était *antillien* au XIX<sup>e</sup> siècle) ;

– que, lorsque l'adjectif ethnique (ou son supplétif) ne résulte d'aucune des suffixations qui précèdent, les choix sont multiples, sans être pour autant mutuellement exclusifs : *-ité* peut s'appliquer à cet adjectif (cf. tableau 3 : *allemandité*, *germanicité*<sup>122</sup>, *belgité*, *bretonnité*, *hellénité* ; hors tableau : *asiaticité*, *afghanité*, *hindouité*, *kurdité*, etc.). Il peut également s'appliquer au toponyme (cf. tableau 3 : *belgicité* et *allemanité*, que nous analysons dans [21] comme construits sur *Allemagne* avec une dépalatisation comme dans *Boulogne / boulonnais* ; hors tableau : *auvergnité*). On citera encore ici tous les cas d'indécision constructionnelle comme *bulgarité*, *croatité*, *kabylité*, *mongolité*, *rhénanité*, *russité*, *serbité*, *slovénité*, *turquité*, *yougoslavité*, etc., pour lesquels il est impossible de déterminer si la base est un adjectif ethnique, ou un toponyme en /i/.

121. Ces finales correspondent la plupart du temps à des suffixes constructionnels, mais elles peuvent aussi n'être que suffiximorphes. C'est le cas dans *portugais*, qui constitue une francisation du portugais *português* d'après le TLF.

122. Le cas d'*allemandité* (1 occ.), *germanicité* (2 occ.), et, plus loin, *allemanité* (4 occ.) est intéressant, en ceci qu'il montre que la présence, dans le lexique conventionnel, de *germanité* ne bloque pas la formation d'autres candidats, même s'ils sont moins fréquents.

Nous tirons un certain nombre d'enseignements dans [21] de l'expérience que nous avons menée. J'insisterai ici sur l'intérêt que constitue désormais une ressource comme la Toile pour mener ce type d'enquête (je rejoins entièrement sur ce point Marc Plénat dans plusieurs de ces travaux : *cf. supra*, § 1., n. 17), ainsi que sur le constat qu'un modèle basé sur des règles au sens de la grammaire générative standard est inapte à rendre compte des phénomènes que nous avons observés, à moins d'énoncer, sur le mode déclaratif, des contraintes pesant sur les entrées pour éviter des dérivés prosodiquement malvenus. Les données rassemblées mettent en outre en évidence les relations paradigmatiques qu'entretiennent un adjectif ethnique et le toponyme qui lui correspond, même s'il n'est pas construit sur lui (*portugais, Portugal*), puisque le toponyme peut suppléer l'adjectif si nécessaire (*portugalité vs \*portugaisité*). On en déduit que, du point de vue du locuteur, il n'existe pas de grande différence sémantique entre les deux, du moins quand il s'agit d'en dériver un lexème construit en *-ité* (il serait vain de chercher une différence dans l'instruction sémantique de *-ité* selon qu'il s'applique à *marocain* ou à *Portugal*). Rétrospectivement, on peut ainsi se demander si *japonerie*, cité dans le § 1.1. au titre des troncations dans le modèle de Danielle Corbin, est effectivement un cas de troncation, ou si la base de ce dérivé n'est pas simplement le toponyme *Japon*.

Faute de place, nous n'avons pas exploité dans [21] un autre type de données intéressantes, illustrant de nouveau la concurrence possible entre un adjectif ethnique et le toponyme qui lui correspond. Il s'agit des noms en *-itude* marquant l'appartenance à un groupe ethnique identifié, et dont certains cohabitent avec un nom en *-ité* construit sur la même base. En première approximation, la productivité du suffixe *-itude* est moindre en français que celle du suffixe *-ité*<sup>123</sup>. Cependant, nous avons relevé sur la Toile des lexèmes construits comme *alsacitude* (2 occurrences au 12 mai 2004 *via* google, pages : France), *belgitude* (288 occ., *ibid.*), *brésilianitude* (1 occ., *ibid.*), *bretonnitude* (79 occ., *ibid.*), *finitude* (3 occ., *ibid.*), *hongritude* (1 occ., *ibid.*), *ibéritude* (1 occ., *ibid.*), *irlan-ditude* (2 occ., *ibid.*), *norvégitude* (3 occ., *ibid.*), *suissitude* (123 occ., *ibid.*). Comme tout à l'heure pour certains noms en *-ité*, on se trouve, pour une partie d'entre eux, face à des cas d'indécidabilité constructionnelle. Plutôt que de trancher en les décrétant construits sur des adjectifs en recourant à des mécanismes de troncation *ad hoc*, parce qu'on a préalablement décidé que la construction sur un nom était impossible, je laisse la question ouverte, et suggère qu'on a pu préférer, pour certains d'entre eux au moins, choisir le candidat jugé optimal, même si ce dernier enfreint un rapport catégoriel présumé.

---

123. Je n'ai pas fait la recherche, mais je me fonde sur le fait que, dans le lexique attesté, *-itude* apparaît de façon majoritaire joint à des adjectifs monosyllabiques en /t/, ce qui restreint le nombre de candidats possibles (il constitue par conséquent un contre-exemple à la tendance dissimilative mentionnée à plusieurs reprises). Par ailleurs, la plupart des noms en *-itude* qu'attestent les dictionnaires sont apparus en français avant le XVI<sup>e</sup> siècle, une partie d'entre eux correspondant en outre à un nom latin en *-itudo*. Les seules innovations du XX<sup>e</sup> siècle qu'enregistre le TLF sont *féminitude*, *finitude*, *francitude* et *négritude*.

## 2.4.2. Lexique construit et analogie

L'objectif de [16] et de [22] est de reconsidérer le concept d'analogie dans le domaine de la formation des lexèmes, à la lumière des déplacements qui s'opèrent en linguistique ces dernières années. [22], que j'ai écrit avant [16], dresse un historique de la notion, en mettant la focale sur l'utilisation du concept dans la description du lexique construit, [16] à la fois condense et complète cet historique, et cherche à réunir des preuves en faveur du concept en tant que présidant à la construction des lexèmes.

Ma découverte de l'analogie a été fortuite. Elle est due à une remarque que m'a faite Nabil Hathout en juillet 2002 en réaction à une version antérieure de mon article sur la productivité morphologique (cf. [15]), lorsque nous nous sommes fait des lectures réciproques des articles constitutifs de [41] au sein de l'action 5 du GDR de morphologie. J'escamotais alors l'analogie dans une note, qui a disparu de la version finale, en disant qu'elle était hors champ pour la notion de productivité. Nabil Hathout a commenté mon propos en disant que toutes les règles étaient des analogies. Je n'ai pas creusé la question pour [15], mais l'ai reprise pour les Journées de Morphologie organisées par l'ERSS à l'université de Toulouse-Le Mirail sur le thème « Lexique et morphologie » les 5 et 6 décembre 2002, où j'ai présenté une communication intitulée : « L'analogie : un concept opératoire dans le domaine de la formation des mots ? ».

J'ai alors découvert un champ immense, dont je ne soupçonnais pas l'existence, et pour cause. Comme je le montre dans [22 : 345 sqq.], après avoir été considérée comme un principe fondamental dans le domaine de la formation des lexèmes de la fin du XVII<sup>e</sup> siècle jusqu'à la première moitié du XX<sup>e</sup> siècle, la notion d'analogie devient confidentielle dans le paysage morphologique international tandis qu'émerge le générativisme. Ainsi, il est notable que le terme *analogie*, ou ses équivalents dans d'autres langues, est absent de la plupart des index thématiques des ouvrages de morphologie d'inspiration générative, par ex. Scalise (1984), Di Sciullo & Williams (1987), Lieber (1992). Il est pareillement absent de l'index de D. Corbin (1987a) et, plus généralement, de l'ensemble des travaux de Danielle Corbin. Et, quand la notion est évoquée par les tenants d'une morphologie générative, c'est pour traiter de cas jugés rétifs aux règles de construction de lexèmes (cf. par ex. Bauer 1983 : 249) ou, à la rigueur, pour expliquer le 'changement analogique' conçu comme la régularisation d'une forme jugée aberrante<sup>124</sup>. Alors que, jusque là, analogie impliquait régularité<sup>125</sup>, le concept est désormais associé à celui d'irrégularité (non-prédictibilité), mais aussi à celui de non-productivité. On assiste en effet pendant la même période à un second glissement consistant à opposer analogie et productivité, lequel perdure encore aujourd'hui (cf. par ex. Fradin 1994 : 16, et, plus récemment, Dressler & Ladányi 2000 ou Booij 2002 : 10-11). Ce glissement est d'autant plus curieux que la définition de Schultink (1961 : 113), qui

---

124. Pour des travaux récents sur la notion de changement analogique, cf. Lass (1997), Joseph (1998) et Joseph & Janda eds (2003).

125. Rappelons seulement la controverse, chez les Grecs anciens puis chez les latins, entre analogistes, qui voyaient la langue comme fondamentalement régulière, et anomalistes, qui soutenaient la thèse inverse.

fait office de standard en matière de productivité, y compris chez les morphologues générativistes, associe intimement les deux concepts :

Onder produktiviteit als morfologisch fenomeen verstaan we dan de voor taalgebruikers bestaande mogelijkheid door middel van het morfologisch procédé dat aan de vormbetekeniscorrespondentie van sommige hun bekende woorden ten grondslag ligt, onopzettelijk een in principe niet telbaar aantal nieuwe formatives te vormen

[Par productivité en tant que phénomène morphologique, nous entendons la possibilité, pour les usagers d'une langue, de former de façon non intentionnelle un nombre en principe infini de nouvelles formations, au moyen d'un procédé morphologique, fondée sur une mise en correspondance formelle et sémantique avec des mots qu'ils connaissent (GD)].

Etant donné ce qui précède, et étant donnée mon option théorique de départ, il était prévisible que je ne me sois jamais intéressée à l'analogie avant la remarque que m'a faite Nabil Hathout.

J'explique dans [22] et [16] l'occultation du concept pendant les années dures du générativisme :

– le point de vue adopté n'est plus celui du locuteur « inventeur de sa langue » (Bréal 1890 [1982] : 328), mais celui du linguiste descripteur, parfois même prescripteur de la langue : or, le concept d'analogie se situe nettement du côté du locuteur ;

– le concept d'analogie, qui repose sur une formation d'après l'existant selon le principe de la quatrième proportionnelle<sup>126</sup>, est incompatible avec les règles de la grammaire générative standard (i) par le type de procédure mis en œuvre (paradigmatique *vs* syntagmatique), (ii) par le niveau d'intervention (surfacique *vs* profond), (iii) par le type d'unité linguistique postulé (mot ou, si on préfère lexème, *vs* morphème : *cf.* van Marle 2000 : 226).

Il me semble cependant que la plupart des travaux qui opposent règles et analogie, quelle que soit l'option qu'ils prennent, ne situent pas toujours l'opposition au bon niveau :

– si on considère que les règles constructionnelles telles que les conçoivent les modèles d'inspiration générative ont une existence psychologique ou, pour le moins, une « plausibilité cognitive » (Koenig 1999 : 8), alors l'opposition est en première analyse licite, aussi bien du côté de la réception d'un lexème construit que de celui de sa production. Dans cette première hypothèse entrent en effet en opposition un chaînage syntagmatique mettant en jeu une ou plusieurs règles symboliques consécutives et une compréhension / production paradigmatique à partir d'un modèle. En tout état de cause cependant, la « plausibilité cognitive » implique que l'on entérine l'existence du lexique mental, ce qui affaiblit l'opposition : il n'est par exemple pas cognitivement plausible qu'un psy-

---

126. Le principe de la quatrième proportionnelle renoue avec la conception aristotélicienne du concept, qui le définit comme une égalité de rapports. La définition du concept en mathématiques ( $a/b = c/d$ ) revient à Euclide (-3<sup>e</sup> siècle), au sein d'une théorie de la proportion reprise à Eudoxe (-408, -355). Une application connue de ce principe en mathématiques est la « règle de trois », ou « règle de proportionnalité », définie comme la recherche, dans une proportion, du quatrième nombre, les trois autres étant connus.

chologue décompose ou recompose l'adjectif *psychologique* à chaque fois qu'il le rencontre ou qu'il l'utilise, ou qu'un morphologue travaillant sur le lexique construit fasse de même pour les adjectifs *dérivationnel* ou *constructionnel*. Par ailleurs, comme on l'a vu plus haut dans le point introductif au § 2.4., aucune expérience décisive n'a démontré que les régularités repérables entre ce qu'on a coutume d'appeler une base et un dérivé existent en mémoire sous la forme de règles symboliques, ou si elles sont récupérées par le biais d'instances servant de modèles (les résultats de Clahsen & al. sous presse sont à cet égard mitigés) ;

– si, par *règle constructionnelle*, on entend 'expression d'une régularité observable' sans rien dire de la procédure suivie (Fradin 2003a : 264 *sqq.*), alors l'opposition n'a pas de raison d'être, dans la mesure où rien ne s'oppose à ce que la régularité observée soit le fruit d'un processus d'analogie. Il n'y a plus, dans cette seconde interprétation, d'incompatibilité foncière à ce que le linguiste décrive la régularité qu'il observe au moyen de patrons abstraits (je renvoie aux patrons que dégagent de la sorte Fradin *ibid.* ainsi que Plag 2003 : 185 *sqq.*), et à ce que, de son côté, le locuteur perçoive cette régularité à partir de calculs analogiques *via* des couples d'instances qu'il possède dans son lexique mental. Dans cette seconde interprétation, la seule obligation pour le linguiste est toutefois qu'il abandonne le modèle des règles orientées, puisque, dès lors qu'il les oriente, il change de perspective et entend dire quelque chose de la procédure suivie<sup>127</sup>. Pour la même raison, il ne peut plus non plus opter pour des règles à proprement parler 'dérivationnelles', faisant l'hypothèse d'étapes intermédiaires entre une entrée et une sortie, puisqu'il ne prétend plus entrer dans la boîte noire que constitue le cerveau. On note en passant que le modèle analogique dispense, lui aussi, d'orientation. J'en reparlerai plus bas.

L'ostracisme dont a été frappée l'analogie pendant les années dures du générativisme pèse toutefois encore sur elle aujourd'hui. Ainsi, tout en adoptant la définition qui précède du nom *règle*, Plag (1999) affirme plus qu'il ne démontre que les formations par analogie doivent être distinguées des instanciations de règles de formation de lexèmes. On note qu'en 2003, Ingo Plag est nettement moins affirmatif, puisqu'il écrit en substance que la ligne de partage entre règles et analogie est en définitive difficile à établir, pour autant qu'elle doive l'être (pp. 38-39). Les reproches qu'on a pu faire à la notion sont, eux aussi, parfois caricaturaux. Ainsi l'affirmation toute récente de Ten Hacken (2003 : 12), qui écrit : « Sans s'appuyer sur une règle déclarant explicitement les contraintes et possibilités du processus, l'analogie se contente d'avoir un exemple », avant de conclure (*ibid.* : 20) : « Heureusement on n'est pas forcé d'abandonner les règles pour expliquer la formation de mots ». Or, les travaux portant sur l'analogie en général, quel que soit le domaine d'application concerné, ont montré qu'il existe bel et bien un calcul analogique, et que l'analogie entre deux domaines s'accompagne nécessairement

---

127. Comme je le souligne à la fin de la première partie de [20], Bernard Fradin n'insiste pas suffisamment sur les conséquences de l'abandon qu'il prône de l'abandon des règles conçues sur un mode opératoire, en particulier sur l'absence d'orientation qui en résulte. Les patrons constructionnels que propose Plag (2003) à la fin de l'ouvrage (pp. 185 *sqq.*) sont eux aussi également non orientés. Je reviendrai sur les conséquences de cette nouvelle façon de reconsidérer la morphologie dans la conclusion.

d'« opérations d'ancrage » (Charconnet 2001). L'analogie exploite donc, dans un contexte donné, ce qui est perçu comme une similitude de propriétés entre les éléments mis en relation (*cf.* le modèle de Skousen dont j'ai parlé plus haut). Rien n'interdit par conséquent que, dans le domaine du lexique construit, une analogie soit établie sur la base de la perception de ressemblances, qui peuvent être de différents types : phonologique, sémantique, syntaxique, etc., chacune de ces ressemblances pouvant elle-même jouer au niveau de a comparé à c, ou de b comparé à d, où a, b, c, d désignent les termes de l'équation de la quatrième proportionnelle. L'analogie, en tout état de cause, « ne se contente pas d'un exemple ».

On assiste actuellement à une véritable renaissance voire à une explosion de la notion, qui bénéficie des progrès effectués dans les sciences cognitives en général, dont les neurosciences, la psychologie cognitive et le traitement automatique (je renvoie par exemple au tout récent travail d'Yves Lepage que j'ai découvert sur la Toile en décembre dernier : *cf.* Lepage 2003). L'analogie fait également l'objet de recherches fécondes en psychologie depuis un quart de siècle (pour une revue détaillée, *cf.* Sander 2000), et est à la source du rapprochement entre psychologie et intelligence artificielle, dans leurs soucis convergents de modéliser le traitement d'informations (*cf.* en particulier Gineste 1997). Il n'est donc plus possible désormais de balayer le concept d'un revers de main, d'autant plus qu'il est compatible avec les tendances générales actuellement à l'œuvre en matière d'étude du lexique construit. C'était déjà vrai il y a un peu plus de dix : en 1992 déjà, Stephen Anderson constatait que l'analogie méritait davantage de considération que celle qu'on lui accordait à l'époque (p. 369), et qu'elle était compatible avec le rejet préconisé par lui de la notion de morphème. Il s'essayait même à reformuler certaines des lois d'analogie dégagées par Kurylowicz (1949) dans les termes de la théorie morphologique qu'il défendait. C'est encore plus vrai aujourd'hui, avec l'abandon massif d'une conception des règles vues comme un ensemble de procédures ordonnées à partir de matériau en entrée au profit de modèles déclaratifs mettant en relation des formes de surface (pour une revue de ces modèles, *cf.* [16]), et la réhabilitation du locuteur dans la description du processus linguistique. Cette réhabilitation a d'ailleurs à voir avec la possibilité qu'offrent la Toile et les corpus électroniques en général d'utiliser des données authentiques, et non plus des données artefactuelles produites par le linguiste. La ligne de partage s'effectue désormais non plus entre ce que les règles prédisent comme possible et impossible (ou, plus exactement, entre ce que le linguiste dit que les règles prédisent comme possible ou impossible), mais entre ce que les locuteurs produisent réellement et ne produisent pas (ou produisent rarement).

Le contexte général actuel est donc tout à fait favorable à la réémergence du concept d'analogie dans le domaine du lexique construit. Il ne reste plus qu'à essayer de rassembler des preuves en sa faveur. [16] s'y emploie, en commençant par remarquer que la tâche n'est pas facile. La plupart du temps en effet, aucun argument décisif ne permet de dire si un lexème construit donné est le produit de l'application de règles ordonnées à partir de matériau en entrée, ou si c'est par analogie avec des mots existants qu'il a été

produit <sup>128</sup>. Par exemple, le même nom *choucrouterie*, que j'ai relevé dans les archives de *Libération*, peut être utilisé par les partisans d'une morphologie basée sur règles, ou par ceux d'une morphologie basée sur l'analogie : les premiers diront qu'il illustre l'aptitude de *-erie* à s'appliquer à des noms d'artefacts pour former des noms d'activité susceptibles de référer par ailleurs à des lieux (cf. Temple 1996) ; les seconds verront en lui la solution de l'équation *croissant:croissanterie = choucroute:x*, sans qu'ils jugent nécessaire de le décomposer par ailleurs en *choucroute + erie*. Et la remarque vaut pour la grande majorité des lexèmes construits.

Les seules données que l'on puisse utiliser sont donc celles dont le traitement ne va pas de soi dans une conception basée sur des règles conçues sur un mode syntagmatique, en ceci qu'elles connaissent un accident, soit formel, soit sémantique, ou que le type de régularité qu'elles exhibent ne peut pas être pris en compte par une morphologie de ce type. Il existe des données éparses, qui montrent l'influence du lexique attesté sur le lexique en construction :

– par exemple, la géminée que comporte le verbe *amerrir*, qu'il hérite d'*atterrir* (cf. Dressler & Ladányi 2000), ou le suffixe *-ustre* de *lacustre* visiblement emprunté à *palustre* ("marécageux"), lui-même hérité du latin *paluster* <sup>129</sup> ;

– par exemple encore, la séquence /arj/ de noms comme *entrepre(u)nariat* <sup>130</sup>, *interprétariat*, *podestariat* ou *vedettariat*, visiblement copiée sur celle de noms comme *actionnariat*, *missionnariat*, *notariat* ou *secrétariat*, qui, eux, peuvent être en relation avec un lexème en *-aire* ;

– par exemple enfin, les innovations lexicales en *Xennat* nommant des périodes, comme le montre ci-dessous la petite enquête que j'avais faite à l'occasion de l'écriture de [16], et que j'ai finalement décidé de ne pas inclure dans l'article.

Les principaux dictionnaires contemporains (*TLF*, *RE* et *PRO1*) attestent quatre noms de forme *Xennat* nommant des périodes : le plus ancien, *triennat*, est daté de 1752, le second, *septennat*, de 1873, le troisième, *quinquennat*, du milieu du XX<sup>e</sup> siècle, le quatrième, *sexennat*, n'apparaît que sous l'entrée *-at* du *TLF*, sans plus de précision. Les CD-Rom de *Libération* (désormais *libé*), *Le Monde (LM)*, *Le Monde diplomatique (LMD)* ainsi qu'une requête sur la Toile via le moteur de recherche google faite au 16 juillet 2003 m'ont permis d'augmenter la liste des cinq noms suivants :

128. Cf. à ce propos la remarque qui ouvre le long paragraphe que Bauer (2001) consacre à l'analogie. Même s'il la fait à propos de la flexion, ce qu'il dit est transposable à la construction d'unités lexicales : « How is the plural of a word like *arb* 'a dealer in stocks of companies facing take-over bids' [...] to be generated? Should it be generated by comparison with the plural of an already existing known word, like *barb*, or should it be generated by the application of a suitable sequence of rules which adds an appropriate element and determines its precise form on the basis of phonological compatibility with the root? In a case like *arb* there is no observable difference in the output: it will be *arbs* in any case, and this will be true in the vast majority of instances ».

129. Huot (2001) cite à plusieurs reprises *lacustre* pour illustrer la suffixation en français. Or, d'après Diez (1874 [1973]), le suffixe *-uster* de *paluster* est déjà « pétrifié en latin ». La fabrication de *lacustre* sur le modèle de *palustre* au XVI<sup>e</sup> siècle a vraisemblablement été favorisée par une proximité de sens (adjectifs exprimant une propriété liée à une étendue d'eau), et par la structure phonique initiale en CVC.

130. Sur les hésitations qu'ont les locuteurs quant à l'orthographe et la prononciation de ce mot, je renvoie à [16 : 224, n. 12], où je montre qu'il est pris dans plusieurs réseaux de calculs analogiques.

- *biennat*, dont *LM* fournit deux occurrences (dans le même document), et google trois <sup>131</sup>,
- *quadriennat*, hapaxique dans les trois journaux considérés, mais présent à soixante-sept reprises sur google,
- de façon tout à fait intéressante, *sexennat*, avec un /t/ intercalaire, dont *Libé* <sup>132</sup> et *LMD* fournissent six occurrences (dans six documents signés par six auteurs différents), et google cinquante-quatre,
- *octennat*, hapaxique dans le corpus considéré,
- *novennat*, absent de mon corpus journalistique, mais qui compte onze occurrences sur google, dont quatre pertinentes, toutes ayant trait à un projet de loi sur le changement de durée du mandat des sénateurs.

Le tableau 4 reprend ces résultats (pour les corpus journalistiques, j’ai en outre mentionné le nombre de documents (‘doc.’) et d’auteurs (‘aut.’) pour les nombres d’occurrences inférieurs à 10) :

	Dictionnaires	<i>Libé</i>	<i>LM</i>	<i>LMD</i>	Google (16/07/03)
biennat	-	-	2 occ. (1 doc.)	-	3 occ.
triennat	1752	-	-	1 occ.	625 occ.
quadriennat	-	-	-	1 occ.	67 occ.
quinquennat	milieu XX <sup>e</sup>	151 doc.	124 occ.	24 occ.	10 872 occ.
sexennat	pas de date	6 occ. (6 doc., 3 aut.)	-	3 occ. (2 doc., 2 aut.)	219 occ.
sexennat	-	3 occ. (3 doc., 3 aut.)	-	3 occ. (3 doc., 3 aut.)	54 occ.
septennat	1873	487 doc.	311 occ.	52 occ.	7 418 occ.
octennat	-	-	-	-	1 occ.
novennat	-	-	-	-	4 occ.

Tableau 4. Dérivés de forme <Xennat> nommant des périodes dans et hors dictionnaires.

On commencera par remarquer qu’une analyse basée sur règles est bien embarrassée avec ces noms. En effet, on pourrait voir dans leur finale <at> un avatar du suffixe *-at* qui, à partir de noms de personnes, forme des noms de fonctions (par ex. *mandarinat*, *inspectorat*), susceptibles, pour certains d’entre eux, de nommer la période pendant lesquelles s’exerce la fonction en question (c’est le cas d’*inspectorat*) <sup>133</sup>. Mais, dans le cas présent, on ne voit pas à quelle base nominale il se serait appliqué, sauf à poser, pour les seuls besoins de l’analyse, des bases de la forme <Xen->, avec X = nom de nombre de forme latine, et <en> allomorphe de *an* “an” (les adjectifs en <Xennal> attestés dans les

131. Les chiffres que je donne ici pour google sont des chiffres nettoyés. Je n’ai par exemple retenu que les trois occurrences non redondantes de *biennat* nommant des périodes sur les soixante-quinze occurrences de <biennat(s)> dans google.

132. Jean-François Boyer, journaliste de *Libé*, produit aussi bien *sexennat* (à quatre reprises) que *sexennat* (une fois).

133. Mon mémoire de maîtrise était consacré à l’étude des formes suffixales *-at* du français (cf. Dal 1987).

dictionnaires sont donnés, eux, comme des héritages du latin, *centennal* excepté). Ces bases auraient ainsi la double particularité de devoir rester à un niveau infralexical, et de ne pas être conformes au type sémantique de bases que réclame normalement le suffixe *-at*. Or, il existe des solutions alternatives, qui varient selon les néologismes considérés. Pour le *triennat* qu'attestent les dictionnaires, elle consiste à voir en lui la réfection du nom *triennal*, qui, lui aussi, a nommé une période de trois ans selon le *TLF*, et à décrire le rôle de *-at* comme ayant été d'intégrer explicitement *triennat* dans le paradigme des noms de périodes, sur le modèle des noms en *-at* ci-dessus (c'est, formulée autrement, la solution que préconise le *TLF*, s.v. **triennat**<sup>134</sup>). Si cette solution est la bonne, nous tenons là un premier argument mettant en évidence l'influence du lexique attesté sur le lexique en construction, et le fait qu'un même lexème peut être pris dans plusieurs réseaux de calculs analogiques.

Par ailleurs, l'existence, hors dictionnaires, de la forme néologique /sɛkstena/ constitue un second argument en faveur de l'analogie. Une bonne façon d'expliquer le /t/ intercalaire qu'il comporte est en effet de considérer qu'il a été formé sur le modèle de *septennat* (/septena/), qui présente la double particularité d'être installé de longue date dans le lexique attesté mais surtout d'avoir un fort taux de fréquence<sup>135</sup>. À ce titre, il appartient donc probablement au lexique mental des locuteurs ayant produit *sextennat*.

Un examen des contextes d'apparition des néologismes de forme <Xennat> montre pour sa part l'effet de paradigme récurrent qu'exercent les uns sur les autres les dérivés de cette forme. Il est en effet remarquable qu'en corpus, les dérivés de cette forme absents des dictionnaires apparaissent souvent à proximité d'un autre nom de cette forme qui, lui, est tout à la fois fréquent et homologué. Par exemple, dans *LM*, chacune des deux occurrences de *biennat* répond à une occurrence de *quinquennat* (c'est moi qui souligne ; les guillemets encadrant la première occurrence de *biennat* sont d'origine) :

François Mitterrand, en recourant deux fois à la dissolution de l'Assemblée nationale au lendemain même de sa propre élection, a substitué au chiffre 7 une formule arithmétiquement insoupçonnable mais politiquement moins innocente :  $7 = 5 + 2$ . C'est-à-dire un quinquennat de plein exercice suivi d'un "biennat".

Le quinquennat voit se déployer les appareils et les mécanismes de la V<sup>e</sup> République. [...]

Le biennat est temps de pénitence pour le président.

Le même phénomène s'observe dans la citation suivante extraite de la Toile où, manifestement, c'est *quinquennat* qui entraîne à sa suite la formation de *quadriennat*, *triennat* et *biennat* (je fais ici l'hypothèse que *triennat* a été reconstruit par l'auteur de la citation) :

Pour faire un quinquennat, il faut avoir déjà fait un quadriennat, un triennat, un biennat, et... l'ENA tout court !

134. Cf. « Réfection d'apr. des mots comme *épiscopat* de *triennal* "période de trois ans" (1671, PO-MEY), subst. de l'adj. *triennal* ».

135. Comme le fait apparaître le tableau 4, *quinquennat* est plus fréquent sur la Toile que *septennat*, alors que leur fréquence respective est inversée dans les trois journaux considérés. On comprend pourquoi : il est probable que la même requête effectuée sur la Toile avant le passage du mandat présidentiel en France à cinq ans aurait donné elle aussi l'avantage à *septennat*.

ou encore dans les deux citations suivantes, toutes deux dues à Philippe de Villiers lors de la deuxième séance de l'assemblée nationale du 14 juin 2000, dans lesquelles *biennat* et *quadriennat* pour la première, *biennat* pour la seconde sont amorcés par *quinquennat* (ainsi que par *septennat* pour la seconde) :

[...] on se demande, après avoir entendu Jacques Chirac il y a quelques jours proposer le quinquennat, ce qu'a bien pu vouloir dire le Président de la République lorsqu'il a prétendu qu'il n'entendait pas modifier la Constitution. Le Président Chirac aura, en effet, présidé à six modifications de la Constitution depuis son biennat, puis son quadriennat, puis son quinquennat, soit presque autant, en cinq ans, que ses quatre prédécesseurs au cours des trente-quatre années précédentes [...]

Pour beaucoup de Français, mieux vaut finalement un quinquennat pour rien qu'un septennat pour rien et peut-être seraient-ils encore plus favorables à un "biennat"...

La citation suivante, due à Olivier Duhamel, que j'ai empruntée aux archives de l'*Humanité* en ligne, confirme cette tendance :

Prenons la proposition que vient de faire le Parti communiste pour éviter cette concomitance : organiser les législatives tous les quatre ans en cas de quinquennat présidentiel. [...]. Si on appliquait à partir de 2002 la proposition du quinquennat-quadriennat, on attribuerait le pouvoir conjointement en 2002 jusqu'aux élections législatives de 2006, et puis ensuite on l'attribuerait à la présidentielle en 2007, aux législatives en 2010, à la présidentielle en 2012, aux législatives en 2014, à la présidentielle en 2017, aux législatives en 2018, et enfin lors d'élections concomitantes en 2022.

Pour ces exemples au moins, il est à peu près certain que c'est une forme en *Xennat* à haute fréquence qui a servi de support à la construction des néologismes. Un travail systématique sur le français reste à mener sur les effets d'ancrage contextuel des lexèmes complexes, et en particulier, sur la cooccurrence, dans une fenêtre étroite, d'un néologisme et d'un mot du lexique attesté et/ou à haute fréquence mettant en jeu le même procédé de construction de lexèmes (pour une étude de ce type en néerlandais, cf. Baayen & Neijt 1997).

[16] examine d'autres données du français, plus systématiques, largement inspirées des travaux de Geert Booij pour le néerlandais<sup>136</sup>. Il s'agit des cas :

- de substitution de suffixes (par ex. *appariteur* / *apparitrice*, *travailleuse* / *travailleuse*) ou de constituants (par ex. *homme-grenouille* / *femme-grenouille*) ;
- de formation dite 'régressive' (par ex. *cohérent* / *cohérer* « rendre cohérent » (TLF), *compétition*, *compétiteur* / *compétiter* "participer à une compétition" (Toile)<sup>137</sup>, *délinquant* / *délinquer*, *saisie-rendication* / *saisir-revendiquer*), que fustige D. Corbin (1987a : 87-88) ;

---

136. Je pourrais reprendre ici les cas de lexèmes en *-et(te)* échappant au traitement constructionniste de la morphologie comme *alouette* ou *sonnette* cités au § 2.1.2.3.

137. Molière emploie un verbe *compétiter*, dans le *Dépit amoureux*, acte IV, scène 2, vers 1280 :

Ainsi, quand une femme a sa tête fantasque,  
On voit une tempête en forme de bourrasque,  
Qui veut compétiter par de certains... propos.

– de syncrétisme, *i.e.* d'utilisation d'une même forme pour des usages différents. À la suite de Booij (1997) pour le néerlandais, je développe dans [16] des cas où un adjectif supplée à un nom, de telle sorte que l'adjectif soit au moins sémantiquement apparié au nom (*Auguste / augustiniennement ; clergé / anticlérical ; naissance / prénatal*). J'aurais pu utiliser, à cet égard, le cas inverse où c'est un nom qui supplée à un adjectif relaté dans le paragraphe précédent (*portugais / portugalité*). Selon les cas, l'adjectif et le nom se servent mutuellement d'allomorphes, pour former de nouveaux lexèmes.

Je ne redévelopperai pas ces exemples ici (je renvoie à [16] et aux travaux que j'y cite). Je reprendrai en revanche les implications théoriques que je liste en conclusion de cet article, en les complétant par rapport au propos général de cette note de synthèse.

Si on abandonne le modèle dérivationnel à proprement parler et si on admet que la notion d'analogie dispense de celle d'orientation (comme en dispensent ou devraient en dispenser, comme je l'ai dit plus haut, les règles conçues comme l'expression de régularités observables), alors devient caduque une partie de la machinerie mise en place dans un modèle comme celui de Danielle Corbin permettant de passer d'un niveau de dérivation à un autre. Cela vaut tout particulièrement de la troncation, et, plus particulièrement, de la troncation réciproque dont fait l'hypothèse D. Corbin (1988), entre les suffixes *-ique*, *-iste* et *-isme* du français, dans la mesure où sa seule raison d'être est la préservation d'une morphologie syntagmatique. Dans ce cas précis, la seule régularité observable que l'on puisse enregistrer est l'existence d'un patron régulier mettant en relation, de façon non orientée, d'une part, des lexèmes en *-iste* ou *-ique* et, d'autre part, des lexèmes en *-isme*, sans qu'on ait aucune preuve qu'un suffixe tronque l'autre. Booij (sous presse) le souligne pour le néerlandais, ainsi que Plag (2003 : 187) pour l'anglais.

On peut également se passer de la notion de base possible, *i.e.* de base non attestée à l'état autonome dont on fait l'hypothèse pour expliquer certains lexèmes construits. Je prendrai ici le cas des noms en <isation> du français.

On remarque que, dans les corpus, les noms en <isation> sont tendanciellement plus fréquents que les verbes en <iser> correspondants, si bien que, quand le nom en <isation> est hapaxique ou compte peu d'occurrences, il est prévisible que le verbe en <iser> n'en compte aucune. Voici, à titre d'exemples, quelques données extraites de *Libé*, restreintes au cas où le nom en <isation> met en jeu un toponyme (tableau 5) ou un anthroponyme (tableau 6) :

toponyme	nom en <isation>	nbr d'occ.	verbe en <iser>	nbr d'occ.
Avignon	avignonnisation	1	(s') avignonniser	0
Baltimore	baltimorisation	1	(se) baltimoriser	0
Bantoustan	bantoustanisation	1	(se) bantoustaniser	0
Belgrad	belgradisation	1	(se) belgradiser	2 (1 doc.)
Burundi	burundisation	1	(se) burundiser	0
Californie	californisation	1	(se) californiser	1
Disney Land	di/ysneylandisation	2	(se) disneylandiser	3
Djibouti	djiboutisation	1	(se) djiboutiser	0
Liban	libanisation	7	(se) libaniser	1
Maastricht	maastrichtisation	1	(se) maastrichtiser	0

Tableau 5. Noms en <isation> et verbes en <iser> correspondants mettant en jeu un toponyme avec nombre d'occurrences dans *Libé*.

antroponyme	nom en <isation>	nbr d'occ.	verbe en <iser>	nbr d'occ.
Alain Jérôme	alainjéromisation	1	(se) alainjéromiser	0
Armani	armanisation	1	(s') armaniser	1
Balladur	balladurisation	7	(se) balladuriser	16
Barre	barrisation	4 (2 doc.)	(se) barriser	0
Belmondo	belmondisation	1	(se) belmondiser	0
Berlusconi	berlusconisation	1	(se) berlusconiser	1
Bill Clinton	billclintonisation	1	(se) billclintoniser	0
Bonaparte	bonapartisation	1	(se) bonapartiser	0
Brian de Palma	briandepalmisation	2 (1 aut.)	(se) briandepalmiser	0
Chaban (Delmas)	chabanisation	1	(se) chabaniser	2
Chirac	chiraquisation	1	(se) chiraquiser	10
Cresson (Edith)	cressonisation	2 (même doc.)	(se) cressoniser	4
De Niro	deNiroisation	1	(se) deN/niroiser	0
Drucker	druckerisation	2	(se) druckeriser	0
Eastwood	eastwoodisation	1	(se) eastwoodiser	0
Fletcher	fletcherisation	1	(se) fletcheriser	0
Gucci	guccisation	2 (même doc)	(se) gucciser	0
Haider	haiderisation	4 (dont 3 même doc)	(se) haideriser	0
Gloria	gloriaisation <sup>138</sup>	1	(se) gloriaiser	0
Mandela	mandelaisation	1	(se) mandelaiser	0
Zidane	zidanisation	1	(se) zidaniser	0

Tableau 6. Noms en <isation> et verbes en <iser> correspondants mettant en jeu un antroponyme avec nombre d'occurrences dans *Libé*.

Même si le verbe n'est pas toujours attesté, on constate cependant que, dès lors que le nom en <isation> est possible, le verbe l'est aussi (la réciproque est également vraie). Par exemple, si « l'avignonisation du festival » (*Libé*) fait l'objet d'une polémique, on pourra tout aussi bien répliquer qu'il peut parfois être souhaitable, pour un festival, de s'avignoniser. Pour reprendre la théorie des catégories de Williams Croft rapidement exposée au § 2.3.2., la différence entre le verbe et le nom consiste en une différence de fonction pragmatique (prédication vs référence) sans que la classe sémantique change (il s'agit dans les deux cas d'action). Ce sont donc des besoins pragmatiques, liés peut-être au style journalistique, qui expliquent que, dans un cas comme celui-là, la presse écrite choisit plutôt le nom.

Dans le cadre d'une morphologie basée sur des règles syntagmatiques orientées, on interprétera l'observation qui précède en disant que le nom en <isation> présuppose le verbe, et on reconstruira un verbe en regard de chaque nom qui en est dépourvu. On verra de la sorte dans les noms en <Xisation> les termes de la chaîne constructionnelle  $X \rightarrow {}^{\circ}Xiser \rightarrow Xisation$ , où X correspond au toponyme ou à l'antroponyme et où la rondelle marque que le verbe est reconstruit. Quand le verbe est attesté, on fera implicitement l'hypothèse que le locuteur qui produit le néologisme en <isation> le sait, ce qui est loin d'être certain.

138. La citation fait implicitement référence au film *Gloria*, de John Cassavetes, datant de 1980. La citation exacte de *Libé* est la suivante : « Une femme, un homme, follement amoureux. Elle a peur, elle le fait enfermer chez les fous. Dix ans plus tard, il sort, croyant comme elle le lui a promis qu'il ne s'est passé que trois mois. Choisira-t-elle ses enfants et son nouvel amour, adulte, ou son amour de jeunesse ? Dans cette superbe histoire vaguement actualisée par Cassavetes fils s'élabore une certaine gloriaisation du cinéma, pour le meilleur et pour le pire ».

Dans un cadre qui se contente d'enregistrer une régularité observable, puisque les contraintes sur l'input initial X sont les mêmes, qu'on construise un nom ou un verbe, on peut très bien imaginer que les noms en <isation> soient mis en relation directement, soit avec un anthroponyme, soit avec un toponyme, sans qu'il faille transiter par un verbe. La régularité observée pour le cas particulier des noms en <isation> à constituant anthroponymique ou toponymique prendrait la forme du patron à trois termes figurant sous le schéma (3), inspiré d'un traitement préconisé par Plag (2003 : 226) pour les triplets de lexèmes de forme <X> ↔ <Xate> 'provide with X' <Xation> ↔ 'action/process/result of Xating'. Rien ne s'oppose à ce que ce patron soit présent en mémoire chez le locuteur sous la forme de couples ou de triplets de lexèmes pouvant être mobilisés en tant que modèles analogiques. Dans le schéma *infra*, les flèches en pointillés marquent que la formation dite 'régressive' d'un toponyme ou d'un anthroponyme à partir d'un verbe en <iser> ou d'un nom en <isation> est peu probable. Elle n'est toutefois pas impossible. Face à un verbe en <iser> ou un nom en <isation> dans lequel on n'identifie pas le premier élément, on peut être amené à déduire qu'il doit s'agir d'un anthroponyme ou d'un toponyme, qu'on ne connaît pas forcément, et à le réemployer en tant que mot-forme :

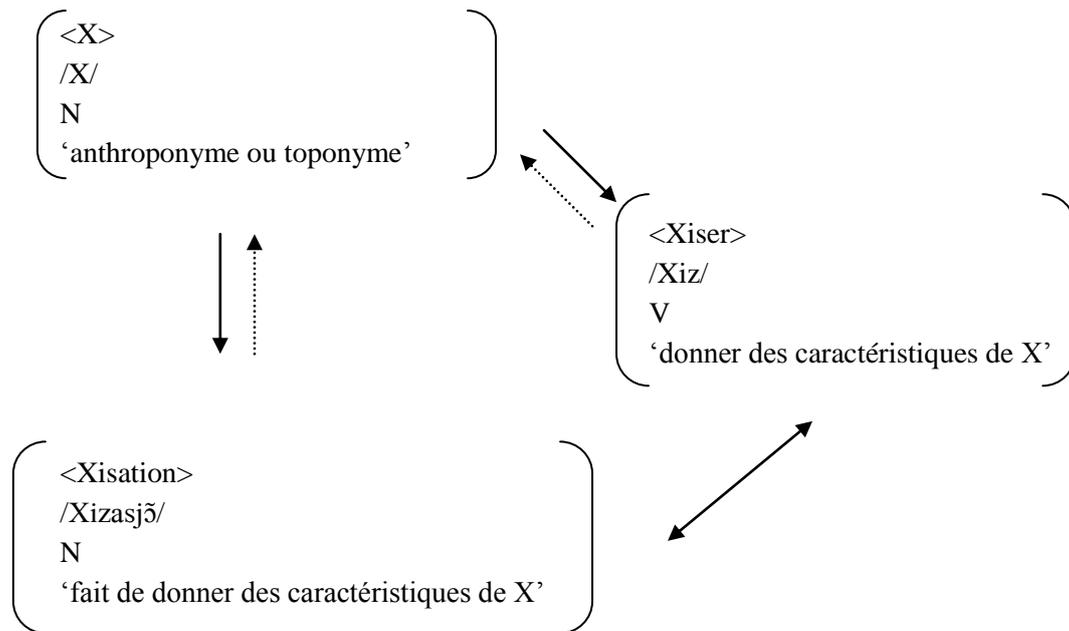


Schéma 3. Représentation des relations observables entre les noms en <isation> à constituant anthroponymique ou toponymique, les verbes en <iser> et les toponymes ou anthroponymes correspondants.

Par rapport à la dérivation en chaîne précédente, et du point de vue du linguiste averti de ses règles, ce schéma à trois termes est moins "esthétique", et plus "coûteux", puisqu'il suppose qu'il y ait deux sous patrons, l'un transitant par le verbe, l'autre évitant l'étape verbale. Cependant, du point de vue du locuteur, cette redondance en mémoire facilite le traitement, puisque en matière de lexique comme ailleurs, plus une information est redondante, plus on y accède facilement. Par ailleurs, une telle organisation prédit que, parfois, le verbe en <iser> fasse défaut en regard du nom.

Enfin, dans un modèle qui reconnaît l'analogie, il n'est plus nécessaire de mettre en place des dispositifs particuliers pour rendre compte des relations entre lexèmes autres que celles de bases à dérivés. Dans un tel cadre, il est prévisible qu'un même lexème construit puisse être pris dans plusieurs réseaux de calcul analogique. On explique dès lors les apparentes étrangetés formelles qu'on repère ici ou là dans le lexique construit (par ex. les cas de supplétion suffixale dont il s'est agi à plusieurs reprises dans ce travail, ou la récurrence d'un /y/ en français devant les suffixes *-eux* et *-el* dans des adjectifs comme *catapultueux*, *difficultueux* ou *accentuel*, *axuel*, tous attestés dans le *TLF* et construits en français <sup>139</sup>), voire les distorsions entre une forme observable et une forme attendue, étant donné un sens observable (je reprendrai ici l'adjectif exemplaire *anti-grippal*).

---

139. Les résultats de Rainer (1998) pour l'italien sont transposables au français : ces adjectifs doivent leur /P/ (/y/) soit à des adjectifs français comme *affectueux*, *impétueux*, etc., hérités du latin, soit à des lexèmes de leur propre famille constructionnelle (par ex. *sexué* pour *sexuel*). On remarque en outre que ce phonème apparaît en français dans des contextes phoniques récurrents (/Ct/ ou /Cs/ à gauche).

## **Conclusion :** **quelle morphologie aujourd'hui ?**

Pour conclure cette note de synthèse, et puisque mes certitudes d'antan se sont effondrées les unes après les autres, je me propose de réunir ici une partie des choix que j'ai été amenée à faire dans les travaux de cette dernière décennie en un cahier des charges minimal pour mes recherches futures sur le lexique construit.

### **– Compétence ou performance ?**

Comme d'autres morphologues, Danielle Corbin entendait rendre compte de la compétence constructionnelle des locuteurs. Faute de savoir ce qu'est la compétence linguistique en général et la compétence constructionnelle en particulier, et si tant est que la distinction compétence / performance doive être maintenue, je chercherai plus modestement à rendre compte de la performance des locuteurs, dans leur rapport effectif au lexique construit. Il s'ensuit :

- (i) que, comme déjà dans [3], je considère *a priori* que tous les lexèmes construits sont possibles dès lors qu'ils sont effectivement produits ;
- (ii) que je suis plus méfiante en revanche envers les lexèmes que ne produisent jamais les locuteurs : la Toile est à cet égard une ressource précieuse ;
- (iii) que, comme Fradin (2003a) notamment, par *règle*, j'entends 'expression d'une régularité observable', sans exclusive quant à la régularité observée ;
- (iv) que les régularités dégagées doivent avoir une plausibilité cognitive, sans pour autant prétendre dire quelque chose de la procédure suivie ;
- (v) qu'autant que faire se peut, la description de la régularité que fournit le linguiste doit rester à un niveau surfacique et prendre la forme de patrons à deux, voire à trois termes mettant en relation des schémas abstraits représentant des groupes de lexèmes.

Ma vision de la morphologie se situe par conséquent désormais entre la position initiale de Danielle Corbin, qui militait contre une évidence en matière de lexique construit, et une vision thomiste naïve, qui consisterait à ne croire que ce que je vois. Disons que je ne crois *a priori* pas ce que je ne vois pas.

Dans la veine des travaux menés en grammaire générative en syntaxe, et peut-être avant, dès le structuralisme, on a voulu réduire le lexique construit à une mécanique complexe de laquelle le locuteur était évincé. L'entreprise vaut peut-être pour la syntaxe, en ceci qu'il ne semble pas exister de 'syntaxe mentale' qui viendrait à l'occasion interférer avec les règles au sens génératif du terme, chaque occurrence de phrase étant réputée être une nouvelle création (*cf.* Di Sciullo & Williams 1987 : 7 *sqq.*). En ce qui

concerne le lexique construit au moins, il n'est possible que jusqu'à un certain point de faire fi du locuteur, dans la mesure où, se faisant, on oblitère l'existence du lexique mental et, plus généralement, l'ensemble des phénomènes relevant de la "gestion du lexique" (j'emprunte l'expression à Marc Plénat), pouvant influencer le choix de tel ou tel procédé, voire la forme même de l'output. On oublie également que « le langage est [...] une création naturelle et non une construction rationnelle et logique » (Darmesterer 1877 [1972] : 34), même si l'entreprise consistant à voir dans le lexique construit une mécanique était séduisante.

Il existe d'autres arguments que l'on peut invoquer en faveur d'une morphologie qui mette le locuteur au centre de ses préoccupations. J'ai déjà cité le phénomène d'« adaptation morphologique » à la suite de van Marle (1993) : cf. *supra*, § 2., n. 26. Je pourrais invoquer le phénomène de réanalyse (par ex. *forain* relié à *foire* alors qu'il résulte de l'évolution phonétique d'un *foranus* latin), ou encore la suffiximorphité de certaines finales qui, parfois, semble suffire au locuteur. Dans [16 : 232], je cite les adverbes *baudelairement* et *lamartinement*, que j'emprunte à Amiot & Flaux (sous presse), dans lesquels l'anthroponyme présente une terminaison d'adjectif, resp. <aire> et <ine>, ce qui est conforme à ce qu'attend la suffixation adverbiale en *-ment*. Je pourrais également me fonder sur des exemples comme *bibliothèque Mazarine* vs \**bibliothèque Colberte* (D. Corbin 2001b : 138), que Danielle Corbin traite au moyen d'une « contrainte d'inhibition » (*ibid.* : 140), bloquant une opération constructionnelle ultérieure <sup>140</sup>.

### – Quels outils ?

Le développement des gros volumes de données textuelles sur support informatique a introduit un changement radical dans les recherches linguistiques, en particulier en ce qui concerne la validité de l'intuition d'acceptabilité ou d'agrammaticalité (cf. Habert & Zweigenbaum 2002 cités au § 1.3.). Pour les morphologues, il n'est toutefois pas encore entièrement naturel de consulter systématiquement la Toile, ou d'autres ressources textuelles authentiques, pour valider ou, au contraire, invalider les hypothèses qu'ils formulent (j'exclus ici les morpho(phono)logues de l'ERSS, dont on a vu à plusieurs reprises l'intérêt que représente pour eux l'existence de telles ressources). Pour ma part, je suis convaincue que l'époque d'une morphologie basée sur les dictionnaires et sur l'introspection est révolue, et que, pour paraphraser le titre de l'ouvrage de Bernard Fradin, il s'agit à proprement parler d'une "nouvelle approche" de la morphologie. Si l'objectif que l'on s'assigne est une étude synchronique du lexique construit, et si l'on décide de rendre compte de la performance effective des locuteurs, les seules données réellement pertinentes sont les lexèmes construits, relevés en amont de leur entrée officielle dans le lexique conventionnel, à charge pour le morphologue d'établir des critères pour s'assurer de la fiabilité des données recueillies. L'entreprise n'est pas facile (Fiammetta Namer, Nabil Hathout, Stéphanie Lignon, Ludovic Tanguy et moi-même

140. Booij (2002 : 115) cite un cas de préfiximorphité ayant un effet inhibant : celui des verbes néerlandais dans lesquels l'initiale présente la forme d'un préfixe (*begin*, *verdwijjn*), qui se comportent comme des lexèmes complexes en ceci qu'ils n'ont pas de participe passé en *ge-*.

sommes actuellement confrontés au problème dans le cadre du projet WeSConVa cité *supra* au § 2.2.1.).

Une utilisation systématique de corpus au sens non contraint du terme permet également de recontextualiser le lexique construit et de faire apparaître, le cas échéant, des phénomènes que, par définition, une morphologie basée sur les dictionnaires et sur l'introspection ne pouvait pas voir (par ex. la récurrence de néologismes formés au moyen d'un procédé donné à proximité d'un lexème homologué et/ou fréquent présentant le même procédé).

### – Morphologie de l'input ou morphologie de l'output ?

Dans une morphologie basée sur des règles au sens génératif du terme, l'essentiel des efforts est porté sur les entrées, comme le montre le principe de la compositionnalité du sens des mots construits par rapport à leur structure tel que le concevait Danielle Corbin. Dans sa perspective, les sorties étaient des leurres, et le travail du morphologue consistait à retrouver les entrées des règles dont elles étaient le terme, ainsi que les mécanismes constructionnels et/ou proprement sémantiques permettant de passer du niveau profond au niveau de surface. Etant donné ce qui précède, la morphologie que je préconise est désormais une morphologie qui prenne en considération également les sorties au même titre, voire plus que les entrées.

Ce choix ne m'est pas propre, et correspond à une tendance qui se dessine dans le paysage morphologique actuel. On le trouve dans Fradin & al. (2003) et Fradin (2003b) pour lesquels, dans certains cas, le suffixe *-et(te)* donne une information sur les lexèmes dans lesquels il figure, indépendamment de toute procédure de calcul de sens lancée par la suffixation par rapport au lexème identifiable en position de base (on a vu au § 2.1.2.3. que je les rejoins sur ce point). On le trouve également dans Plag (1999 : 241), qui préconise une « output-oriented » morphologie, au contraire de Beard (1995), dont la morphologie, selon Ingo Plag, est orientée vers le procès (*process-oriented model of morphology*). Les travaux menés en morphophonologie à l'ERSS actuellement sont également portés par cette lame de fond, que les chercheurs toulousains ont d'ailleurs contribué à initier pour le français.

### – Morphologie constructionnelle vs syntaxe

La reconnaissance du lexique mental implique que l'on reconnaisse l'existence d'un niveau d'organisation spécifique aux lexèmes construits, qui ne soit pas celui de la syntaxe. À cet égard, même si c'est pour d'autres raisons, la position que j'ai adoptée il y a dix ans dans le sillage de Danielle Corbin n'a pas changé<sup>141</sup>. Trois de mes travaux, dont je n'ai pas parlé dans ce qui précède, portent précisément sur la question de l'autonomie de la morphologie par rapport à la syntaxe :

---

141. Pour des écrits récents sur le placement de la morphologie vis-à-vis de la syntaxe, je renvoie à Fradin (2003a : 22 *sqq.*) et à Fradin & Kerleroux (2003a). Anderson (1992) reconnaît, lui aussi, un niveau d'organisation spécifique au lexique construit. L'un des résultats qu'il dégage pour le K<sup>w</sup>ak<sup>w</sup>ala est que « [t]he rules of word-internal morphology, then, not only have different content from those of the syntax, but refer to a class of primitives (the content of  $\theta$ -roles) which are not generally considered to be available for syntactic manipulation » (*ibid.* : 36). Il serait intéressant de reprendre ce résultat pour voir s'il peut être étendu à d'autres langues.

(i) dans [13], me fondant sur un corpus d'énoncés attestés mettant en scène les mêmes noms, associés tantôt au préfixe, tantôt à la préposition (par ex. *contre-pub* vs *contre la pub* ; *contre-guérilla* vs *contre la guérilla*), je montre que le préfixe fonctionne comme marqueur de subsidiarité, tandis que la préposition déploie principalement une valeur adversative ;

(ii) dans [14], écrit avec Fiammetta Namer, nous examinons un cas de blocage présumé entre lexique construit et syntaxe. Nous montrons plus particulièrement qu'en dépit des apparences peut-être, les verbes causatifs du français construits sur des adjectifs n'équivalent pas à la séquence "rendre A", où A correspond à la base du verbe complexe. Nous nous fondons sur le fait que les contraintes pesant sur l'adjectif ne sont pas les mêmes dans les deux cas. Alors que les procédés de construction de lexèmes formant des causatifs ne sont compatibles qu'avec des bases exprimant des propriétés acquérables (cf. *supra*, § 2.2.2. pour le préfixe *-is-*), *rendre* ne fait pas peser cette contrainte sur les adjectifs qu'il sélectionne. Ainsi, un gouffre peut rendre une rivière poissonneuse (ex. adapté d'un exemple emprunté à la base de données Frantext et cité dans [14 : 193]), mais ne peut pas \*poissonnosifier une rivière. On explique cela par le fait que *X rend Y A* peut signifier "X est cause que Y soit A", sans forcément dire quelque chose du caractère inhérent ou acquis de la propriété exprimée par A, alors que les lexèmes en *-eux* expriment un certain type de propriétés inhérentes. À ce titre, ils ne constituent pas de bonnes entrées pour la suffixation par *-if-* ;

(iii) dans [28], enfin, réalisé avec Dany Amiot et présenté au *10th International Morphology Meeting* à Szentendre (Hongrie) en mai 2002, nous nous sommes focalisées sur l'absence de déterminant dans la préfixation en français pour montrer que les règles morphologiques ne peuvent pas être confondues avec les règles syntaxiques, et que, bien que des interactions soient possibles, la morphologie et la syntaxe relèvent de composants différents.

Je ne détaillerai pas plus avant quelle morphologie sera désormais la mienne, faute d'être capable de mesurer l'intégralité des implications que suppose le changement de paradigme que j'ai relaté dans cette note de synthèse et qui remet entièrement en cause ma façon de penser la morphologie par rapport à mes certitudes d'antan. Je ne renie toutefois pas la morphologie "dure" à laquelle j'ai cru et les principes forts que j'ai appliqués, dans le sillage de Danielle Corbin. Ces principes ont été nécessaires en leur temps, pour mettre de l'ordre dans un champ exploré jusque là souvent de façon aléatoire, et une partie des résultats auxquels ils ont abouti ainsi que les méthodes qui ont été utilisées doivent être conservés pour les études futures.

En tout état de cause, le changement de paradigme auquel on assiste actuellement en morphologie constructionnelle et dont témoignent à leur manière mes propres recherches va dans le sens de la théorie de Kuhn (1970), selon lequel, dans l'histoire des sciences, les changements sont des processus lents, l'effondrement d'un ancien système pouvant être synchrone à l'émergence d'un nouveau paradigme (parmi les philosophes des sciences du XX<sup>e</sup> siècle, il s'oppose à cela à Karl Popper, qui préconise un processus de réfutation permanent, et à Imre Lakatos, qui propose un tempo intermédiaire : pour ce dernier, les programmes de recherche fonctionnent par périodes, et ont une validité

dans le temps limitée, sans qu'émerge forcément un nouveau programme invalidant entièrement le premier). Pour autant que j'y voie clair dans la situation actuelle en morphologie, nous nous situons en effet actuellement à une période charnière où coexistent plusieurs paradigmes, et il est impossible de dire si les programmes émergents doivent ou non remplacer les programmes précédents, ou si ce vers quoi on tend consiste en une solution de compromis, comme le suggère Ingo Plag dans la conclusion de son ouvrage de 2003.



# Bibliographie

## 1. Études

- AITCHISON J. (1994), *Words in the mind. An introduction to the mental lexicon*, 2<sup>nd</sup> éd., Oxford UK / Cambridge USA, Blackwell ; 1<sup>st</sup> ed. : 1987.
- ALIQUOT-SUENGAS S. (1996), *Référence collective / sens collectif. La notion de collectif à travers les noms suffixés du lexique français*, Thèse de doctorat, Université de Lille III.
- AMIOT D. (1997), *L'antériorité temporelle dans la préfixation en français*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- (2002), « Re-, préfixe aspectuel ? », *Cahiers chronos* 10, pp. 1-20.
- (2003), « De l'antériorité à la postériorité : mode de repérage temporel et type de préfixes », *Cahiers chronos* 11, pp. 173-189.
- AMIOT D. & FLAUX N. (sous presse), « Pourquoi existe-t-il si peu d'adverbes en -ment dérivés de noms propres ? », *Actes du colloque franco-roumain de linguistique "L'adverbe", 26-28 mai 2003, Université d'Artois*.
- ANDERSON S.R. (1992), *A-Morphous Morphology*, Cambridge, Cambridge University Press.
- ARCHANGELI D. (1997), « An introduction to Linguistics in the 1990s », in D. Archangeli & D.T. Langendoen eds, *Optimality Theory. An overview*, Oxford, Blackwell Publishers Ltd, pp. 1-32.
- ARONOFF M. (1976), *Word Formation in Generative Grammar*, Cambridge, Mass. / London, England, The MIT Press.
- (1994), « Introduction », in M. Aronoff ed., *Morphology now*, Albany, State University of New York Press, pp. 1-3.
- ARONOFF M. ed. (1994), *Morphology now*, Albany, State University of New York Press, pp. 1-3.
- AURNAGUE M. & PLENAT M. (1996), « La préfixation en é- et la relation de partie à tout », in *Représentations et Outils pour les bases lexicales. Morphologie robuste*, Grenoble, CLIPS-IMAG, pp. 43-52.
- (1997), « Manifestations morphologiques de la relation d'attachement habituel », in D. Corbin & al. éd(s) (1997), pp. 15-24.
- BAAYEN H.R. (1992), « Quantitative aspects of morphological productivity », *Yearbook of Morphology 1991*, pp. 109-149.
- BAAYEN H.R. & LIEBER R. (1991), « Productivity and English derivation: a corpus-based study », *Linguistics* 29/5, pp. 801-843.
- BAAYEN H.R. & NEIJT A. (1997), « Productivity in context: a case study of Dutch suffix », *Linguistics* 35, pp. 565-587.
- BAKER M.C. (2003), *Lexical categories: verbs, nouns, and adjectives*, Cambridge, Cambridge University Press.
- BAUER L. (1983), *English Word-formation*, Cambridge, Cambridge University Press.
- (2001), *Morphological Productivity*, Cambridge, Cambridge University Press.

- BEARD R. (1995), *Lexeme-Morpheme Base Morphology. A general Theory of Inflection and Word Formation*, New York, State University of New York Press.
- BOERSMA P., DEKKERS J. & VAN DE WEIJER J. (2000), « Introduction », in J. Dekkers, F. Van der Leeuw & J. van de Weijer eds, *Optimality Theory. Phonology, Syntax, and Acquisition*, Oxford, Oxford University Press, pp. 1-44.
- BOOIJ G. (1977), *Dutch Morphology. A Study of Word Formation in Generative Grammar*, Lisse, The Peter de Ridder Press.
- (1996), « Inherent versus contextual inflection and the split morphology hypothesis », *Yearbook of Morphology 1995*, pp. 1-16.
- (1997), « Autonomous morphology and paradigmatic relations », in *Yearbook of Morphology 1996*, pp. 35-53.
- (2002), *The morphology of Dutch*, Oxford, Oxford University Press.
- (sous presse), « Paradigmatic morphology », in B. Fradin éd., *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, Amsterdam / Philadelphia, John Benjamins Publishing Company.
- BOOIJ G., LEHMANN C. & MUGDAN J. eds (2000), *Morphology. An international handbook on Inflection and Word-Formation*, vol. 1, Berlin / New York, Walter de Gruyter.
- BREAL M. (1890), *Essai de sémantique*, Paris ; rééd. Brionne, Gérard Montfort éditeur, 1982.
- BURZIO L. (1998), « Multiple correspondence », *Lingua* 104, pp. 79-109.
- (2000), « Cycles, non-derived-environment blocking, and correspondence », in J. Dekkers, F. Van der Leeuw & J. van de Weijer eds, *Optimality Theory. Phonology, Syntax, and Acquisition*, Oxford, Oxford University Press, pp. 47-87.
- (2001), « Zero Derivations », *Linguistic Inquiry* 32/4, pp. 658-677.
- BYBEE, J. (1985), *Morphology. A study of the relation between meaning and form*, Amsterdam / Philadelphia, John Benjamins Publishing Company.
- CADIOT P. (1994), « Représentations d'objets et sémantique lexicale : qu'est-ce qu'une boîte ? », *Journal of French Language Studies* 4, pp. 1-23.
- CHARCONNET J. (2001), « La construction de notions nouvelles par analogie », *Revue de sémantique et de pragmatique* 9-10, pp. 233-252.
- (2003), *Analogie et logique naturelle. Une étude des traces linguistiques du raisonnement analogique à travers différents discours*, Bern / Berlin / Bruxelles / Frankfurt sur M. / New York / Oxford / Wien, Peter Lang.
- CHOMSKY N. (1970), « Remarks on nominalization », in R. Jacobs & P. Rosenbaum eds, *Readings in English transformational grammar*, Waltham, Mass., Ginn, pp. 184-221.
- CLAHSEN H., SOLLENSTUHL I. & BLEVINS J.P. (sous presse), « Derivational Morphology in the German Mental Lexicon: a dual Mechanism Account ».  
[<http://www.essex.ac.uk/linguistics/papers/derivation5.pdf>.]
- CLEMENT L. (2001), *Construction et exploitation d'un corpus syntaxiquement annoté pour le français*, Thèse de doctorat, Université de Paris 7 – Denis Diderot.
- CORBIN D. (1976a), « Peut-on faire l'hypothèse d'une dérivation en morphologie ? », in J.-C. Chevalier éd., *Grammaire transformationnelle. Syntaxe et lexicale*, Lille, Presses Universitaires de Lille, pp. 47-91.
- (1986), « Qu'est-ce qu'une opération dérivationnelle ? Description et représentation des noms en –ET(TE) apparemment construits sur des bases verbales », *Cahiers de Grammaire* 11, pp. 1-65.
- (1987a), *Morphologie dérivationnelle et structuration du lexique*, 2 vol., Tübingen, Max Niemeyer Verlag ; rééd., Villeneuve d'Ascq, Presses Universitaires de Lille, 1991.

- (1987b), « Contre une transposition de la théorie X' à la morphologie dérivationnelle », Communication au VII<sup>e</sup> Colloque international de l'Université de Paris VIII, "L'hétérogénéité de la grammaire", 21-23 mai 1987, in *Acta Linguistica Academiae Scientiarum Hungaricae* 37/1-4, pp. 73-92.
- (1988), « Une hypothèse à propos des suffixes *-ISME*, *-IQUE*, *-ISTE* du français : la troncation réciproque », in R. Landheer éd., *Aspects de linguistique française. Hommage à Q.I.M. Mok*, Amsterdam, Rodopi, pp. 63-75.
- (1991a), « Introduction. La formation des mots : structures et interprétations », *Lexique* 10, pp. 7-30.
- (1991b), « La morphologie lexicale : bilan et perspectives », *Travaux de linguistique* 23, pp. 33-56.
- (1992), « Sens et définition : de la compositionnalité du sens des mots construits (Réponse à Claire Vanderhoeft) », *Linguisticae Investigationes* XVI/1, pp. 189-218.
- (1993), « Morphologie et lexicographie : la représentation du sens dans le *Dictionnaire dérivationnel du français* », in A. Hulk, F. Melka & J. Schroten éd., *Du lexique à la morphologie : du côté de chez Zwaan*, Amsterdam, Rodopi, pp. 63-86.
- (1997a), « Décrire un affixe dans un dictionnaire », in G. Kleiber & M. Riegel éd., *Les formes du sens. Etudes de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, Louvain-la Neuve / Paris, Duculot, pp. 79-94.
- (1997b), « Entre les mots possibles et les mots existants : les unités à faible probabilité d'actualisation », in D. Corbin & al. éd., (1997), pp. 78-89.
- (1997c), « La représentation d'une "famille" de mots dans le *Dictionnaire dérivationnel du français* et ses corrélats théoriques, méthodologiques et descriptifs », *Recherches Linguistiques de Vincennes* 26, Université de Paris VIII, pp. 5-38 + Supplément, pp. I-VIII.
- (2001a), « Préfixes et suffixes : du sens aux catégories », *Journal of French Language Studies* 11/1, pp. 41-69.
- (2001b), « Du nouveau sur *beurre laitier*. Note sur une fausse conversion », in C. Buridant, G. Kleiber & J.-C. Pellat éd., *Par monts et par vaux. Itinéraires linguistiques et grammaticaux*, Louvain/Paris, éditions Peeters, pp. 127-143.
- (à paraître), *Le lexique construit. Méthodologie d'analyse*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- CORBIN D. & CORBIN P. (1991), « Un traitement unifié du suffixe *-ier(e)* », *Lexique* 10, pp. 61-145.
- CORBIN D., DAL G., MELIS-PUCHULU A. & TEMPLE M. (1993), « D'où viennent les sens *a priori* figurés des mots construits ? Variations sur *lunette(s)*, *ébéniste* et les adjectifs en *-esque* », *Verbum* 1-2-3, pp. 65-100.
- CORBIN D., FRADIN B., HABERT B., KERLEROUX F. & PLENAT M. éd., (1997), *Sillexicales* 1, « Mots possibles et mots existants », Université de Lille III.
- CORBIN D. & PAUL J. (2000), « Aperçus sur la créativité morphologique dans la terminologie de la chimie », *La banque des mots* 60, pp. 51-68.
- CORBIN P. (sous presse), « Introduction : *Lexique* 16, treize ans après *Lexique* 10 », *Lexique* 16.
- CROFT W. (1991), *Syntactic Categories and Grammatical Relations: The Cognitive Organization of the Information*, Chicago / London, The University Press of Chicago.
- (2003), *Typology and Universals*, 2<sup>e</sup> éd., Cambridge, Cambridge University Press ; 1<sup>e</sup> éd., 1990.
- DAILLE B., FABRE C. & SÉBILLOT P. (2002), « Applications of computational morphology », in P. Boucher ed., *Many Morphologies*, Somerville, Mass., Cascadilla Press, pp. 210-234.

- DAL G. (1987), *Étude des suffixes -at en français moderne*, Mémoire de maîtrise, Université de Lille III.
- (1988), *Les noms d'état en français : essai de méthodologie d'analyse*, Mémoire de D.E.A., Université de Lille III.
- (1991), « Hyponymie et prototypie : les noms en -asse et -et(te) du français », *Lexique* 10, pp. 211-239.
- (1993), « Règles et exceptions : application aux noms en -et(te) du français », *Cahiers de lexicologie* 62, pp. 109-131.
- (1994), *Un exemple de traitement associatif du lexique construit : analyse unificatrice des mots suffixés par -et(te)*, Thèse de doctorat, Université de Lille III.
- DARMESTER A. (1877), *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent*, Paris, Vieweg ; Genève, Slatkine Reprints, 1972.
- DELL F. (1970), *Les règles phonologiques tardives et la morphologie dérivationnelle du français*, Ph. D. Diss., M.I.T., non publié.
- (1979), « La morphologie dérivationnelle du français et l'organisation de la composante lexicale en grammaire générative », *Revue romane* XIV/2, pp. 185-216.
- DERWING B.L. & SKOUSEN R. (1994), « Productivity and the English past tense », in S.D. Lima, R.L. Corrigan & G.K. Iverson eds, *The reality of linguistic rules*, Amsterdam / Philadelphia, John Benjamins Publishing Company, pp. 193-218.
- DIEZ F. (1874), *Grammaire des langues romanes. Tome deuxième traduit par Alfred Morel-Fatio et Gaston Paris*, Paris, Librairie A. Franck ; réimpression, Genève, Slatkine Reprints, Mareille, Lafitte Reprints, 1973.
- DI SCIULLO A.-M. & WILLIAMS E. (1987), *On the definition of Word*, Cambridge, Mass., The MIT Press.
- DOLESCHAL U. & THORNTON A. eds (2000), *Extragrammatical and marginal Morphology*, München, Lincom Europa.
- DRESSLER W.U. (2000), « Extragrammatical vs marginal morphology », in Doleschal & al. eds (2000), pp. 1-10.
- DRESSLER W. & LADÁNYI M. (2000), « Productivity in word formation (WF): a morphological approach », *Acta Linguistica Hungarica* 47-2, pp. 103-144.
- DRESSLER W.U. & MERLINI-BARBARESI L.M. (1994), *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and other Languages*, Berlin / New York, Mouton de Gruyter.
- DUBOIS J. & DUBOIS C. (1971), *Introduction à la lexicographie : le dictionnaire*, Paris, Librairie Larousse.
- DURAND J. & LAKS B. (2000), « Relire les phonologues du français : Maurice Grammont et la loi des trois consonnes », *Langue française* 126, pp. 29-38.
- FEHRINGER C. (2003), « Morphological 'gangs': constraints on paradigmatic relations in analogical change », *Yearbook of Morphology* 2003, pp. 249-272.
- FRADIN B. (1994), « L'approche à deux niveaux en morphologie computationnelle et les développements récents en morphologie », *Traitement Automatique des Langues* 35/2, pp. 9-48.
- (2001), « À propos du suffixe -ET. Analyse de Georgette DAL 1997. *Grammaire du suffixe -et(te)*. Paris. Didier Erudition », *Le français moderne* LXIX-1, pp. 86-98.
- (2003a), *Nouvelles approches en morphologie*, Paris, Presses Universitaires de France.
- (2003b), « Le traitement de la suffixation en -ET », *Langages* 152, pp. 51-77.
- (à paraître), « Problemi semantici in morfologia derivazionale », in *Atti del 37esimo convegno della SLI*, 2003, L'Aquila.

- FRADIN B., DAL G., HATHOUT N., KERLEROUX F., PLENAT M. & ROCHE M. éd(s) (2003), *Sillexicales* 3, « Les unités morphologiques », Université de Lille III.
- FRADIN B., HATHOUT N. & MEUNIER F. (2003), « La suffixation en *-ET* et la question de la productivité », *Langue française* 140, pp. 56-78.
- FRADIN B. & KERLEROUX F. (2003a), « Introduction », *Langages* 152, pp. 3-11.
- (2003b), « Troubles with lexemes », in G. Booij, J. De Cesaris, A. Ralli & S. Scalise eds, *Topics in Morphology. Selected papers from the third Mediterranean Morphology Meeting (Barcelona, September 20-22, 2001)*, Barcelona, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, pp. 177-196.
- FRADIN B. & KERLEROUX F. éd(s) (2003), *Langages* 152, « Quoi de neuf en morphologie ? ».
- FRANCKEL J.-J. (1997), « Approche de l'identité d'un préverbe à travers l'analyse des variations sémantiques des unités préverbées », *Journal of French language studies* 7, pp. 47-68.
- FRAUENFELDER U. & SCHREUDER R. (1992), « Constraining psycholinguistic models of morphological processing and representation: the role of productivity », *Yearbook of Morphology 1991*, pp. 165-183.
- FROISSART C. & LALLICH-BOIDIN G. (1996), « Morphologie robuste et analyse automatique de la langue : étude réalisée à partir des corpus de l'évaluation GRACE », in *Actes du séminaire Lexique. Représentations et Outils pour les bases lexicales. Morphologie robuste*, Grenoble, CLIPS-IMAG, les 13 et 14 nov. 1996, pp. 88-96.
- FU J., ROEPER T. & BORER H. (2001), « The VP within process nominals: evidence from adverbs and the VP anaphor *do-so* », *Natural Language & Linguistic Theory* 19, pp. 549-582.
- FUCHS C. éd. (1993), *Linguistique et traitements automatiques des langues*, Paris, Hachette supérieur.
- GERHARD-KRAIT F. (2000), *La préfixation en de(s)- : formes construites et interprétations*, Thèse de doctorat, Université Marc Bloch, Strasbourg 2.
- GINESTE M.-D. (1997), *Analogie et cognition*, Paris, Presses Universitaires de France.
- GODARD D. & JAYEZ J. (1996), « Types nominaux et anaphores : le cas des objets et des événements », *Cahiers chronos* 1, pp. 41-58.
- GODART-WENDLING B., ILDEFONSE F., PARIENTE J.-C. & ROSIER-CATACH I. (1998), « Penser le principe de compositionnalité : éléments de réflexion historiques et épistémologiques », *Traitement automatique des langues* 39/1, pp. 9-34.
- GRABAR N. & ZWEIGENBAUM P. (1999), « Acquisition automatique de connaissances morphologiques sur le vocabulaire médical », in *Actes de la VI<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN'99)*, Institut d'études scientifiques de Cargèse, Corse, 12 -17 juillet 1999, pp. 175-184.
- (2003), « Productivité à travers domaines et genres : dérivés adjectivaux et langue médicale », *Langue française* 140, pp. 102-125.
- GRIMSHAW J. (1990), *Argument Structure*, Cambridge, Mass., The MIT Press.
- GUILBERT L. (1975), *La créativité lexicale*, Paris, Larousse.
- HABERT B. & ZWEIGENBAUM P. (2002), « Régler les règles », *Traitement automatique des langues* 43/3, pp. 83-105.
- HACKEN P. TEN (2003), « Le rôle de l'analogie et des règles dans la formation des mots », *Travaux linguistiques du CERLICO* 16, pp. 11-26.
- HALLE M. (1973), « Prolegomena to a theory of word formation », *Linguistic Inquiry* 4/1, pp. 3-16.
- HALLE M. & MARANTZ A. (1993), « Distributed Morphology and the Pieces of Inflection », in K. Hale & S.J. Keyser eds, *The View From Building 20*, Cambridge / London, The MIT Press, pp. 111-176.

- HASSELROT B. (1957), *Etudes sur la formation diminutive dans les langues romanes*, Uppsala, Almqvist & Wiksells.
- HATHOUT N., PLENAT M. & TANGUY L. (2003), « Enquête sur les dérivés en *-able* », *Cahiers de grammaire* 28, pp. 49-90.
- HAY J. (2001), « Lexical frequency in morphology: is everything relative? », *Linguistics* 39/6, pp. 1041-1070.
- HOEKSEMA J. (2000), « [article 82] Compositionality in meaning », in Booij & al. eds (2000), pp. 851-857.
- HUOT H. (1997), « Des mots possibles aux mots existants : système morphologique et structuration du lexique », in D. Corbin & al. éd. (1997), pp. 141-150.  
(2001), *Morphologie. Forme et sens des mots du français*, Paris, Armand Colin.
- IACOBINI C. (2000), « [article 84] Base and direction of derivation », in Booij & al. eds (2000), pp. 865-876.
- JACKENDOFF R. (1975), « Morphological and Semantic Regularities in the Lexicon », *Language* 51/3, pp. 639-671.  
(1993), « X-bar semantics », in J. Pustejovsky ed., *Semantics and the lexicon*, Dordrecht / Boston / London, Kluwer Academic Publishers, pp. 15-26.
- JACQUEMIN C. (2000), « Présentation », *Traitement automatique des langues* 41/2, pp. 327-333.
- JACQUEMIN C. éd. (2000), *Traitement automatique des langues* 41/2, « Traitement automatique des langues pour la recherche d'information ».
- JACQUEMIN C. & TZOUKERMANN E. (1999), « NLP for term variant extraction: synergy between morphology, lexicon and syntax », in T. Strzalkowski ed., *Natural Language Processing and Information Retrieval*, Kluwer, Boston, pp. 25-74.
- JALENQUES P. (2000), *Contribution à l'étude du préfixe RE en français contemporain : pour une analyse compositionnelle du verbe regarder*, Thèse de doctorat, Université de Paris VII.  
(2002), « Etude sémantique du préfixe RE en français contemporain : à propos de quelques débats actuels en morphologie dérivationnelle », *Langue française* 133, pp. 74-90.
- JESPERSEN O. (1942), *A Modern English Grammar on Historical Principles*, vol. VI : *Morphology*, Copenhagen, Munksgaard.
- JOSEPH B. (1998), « Diachronic Morphology », in A. Spencer & A.M. Zwicky eds, *The Handbook of Morphology*, Blackwell Publishers, pp. 351-373.
- JOSEPH B. & JANDA R. eds (2003), *The Handbook of Historical Linguistics*, Oxford, Blackwell Publishers.
- KERLEROUX F. (1999), « Identification d'un procédé morphologique : la conversion », *Faits de langues* 14, pp. 89-100.  
(sous presse), « Sur quels objets portent les opérations morphologiques de construction ? », *Lexique* 16.
- KOENIG J.-P. (1999), *Lexical relations*, Stanford, CSLI.
- KUHN T.S. (1970), *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.
- KURYLOWICZ J. (1949), « La nature des procès dits 'analogiques' », *Acta Linguistica Hafniensia* 5, pp. 15-37.
- LAKS B. éd. (2000), *Langue française* 126, « Où en est la phonologie du français ? ».
- LASS R. (1997), *Historical linguistics and language change*, Cambridge, Cambridge University Press.
- LEES R.B. (1963), *The grammar of English nominalizations*, The Hague / Bloomington, Mouton & Co / Indiana University.

- LEPAGE Y. (2003), *De l'analogie rendant compte de la commutation en linguistique*, Mémoire d'habilitation à diriger des recherches, Université de Grenoble.
- LIEBER R. (1992), *Deconstructing Morphology. Word Formation in Syntactic Theory*, Chicago and London, The University of Chicago Press.
- LIEBER R. & MUGDAN J. C. (2000), « [article 42] Internal structure of words », in Booij & al. eds (2000), pp. 404-416.
- LIGNON S. (2000), *La suffixation en -ien. Aspects sémantiques et phonologiques*, Thèse de doctorat, Université de Toulouse-Le Mirail.
- LYONS J. (1977), *Semantics*, Cambridge, Cambridge University Press.
- MARLE J. VAN (1993), « Morphological adaptation », *Yearbook of Morphology 1993*, pp. 255-265.
- (2000), « [article 26] Paradigmatic and syntagmatic relations », in Booij & al. eds (2000), pp. 225-234.
- MATTHEWS P.H. (1974), *Morphology. An Introduction to the Theory of Word-Structure*, Cambridge, Cambridge University Press ; 2<sup>nd</sup> ed. : 1991.
- MC CARTHY J. & PRINCE A. (1993), « Generalized alignment », *Yearbook of Morphology 1993*, pp. 79-153.
- MOLINO J. (1985), « Où en est la morphologie ? », *Langages* 78, pp. 5-40.
- MOLINO J. éd. (1985), *Langages* 78, « Le retour de la morphologie ».
- NAMER F. (2003a), « Automatiser l'analyse morphologique non affixale : le système DériF », *Cahiers de grammaire* 28, pp. 31-48.
- (2003b) « WaliM : valider les unités morphologiquement complexes par le Web », in Fradin & al. eds (2003), pp. 142-150.
- NAZARENKO A. (1998), « Présentation », *Traitement automatique des langues* 39/1, pp. 3-7.
- NAZARENKO A. éd. (1998), *Traitement automatique des langues* 39/1, « Compositionnalité ».
- NICOLAS S. & PERRUCHET P. eds (1998), *Psychologie française* 43/1, « Mémoire et apprentissage implicites ».
- PAMIES A. (2002), « Sémantique et grammaire de la possession dans les langues d'Europe », in E. Castagne éd., *Pour une modélisation de l'apprentissage simultané de plusieurs langues apparentées ou voisines*, Publications de la Faculté de Lettres de Nice, Paris, CID Diffusion, pp. 67-98.
- [<http://ashda.ugr.es/laboratorio/possess-french-latinizado.pdf>]
- PINKER S. & PRINCE A. (1994), « Regular and irregular morphology and the psychological status of rules of grammar », in S.D. Lima, R.L. Corrigan & G.K. Iverson eds, *The reality of linguistic rules*, Amsterdam / Philadelphia, John Benjamins Publishing Company, pp. 321-351.
- PLAG I. (1999), *Morphological productivity. Structural constraints in English derivation*, Berlin / New York, Mouton de Gruyter.
- (2003), *Word-formation in English*, Cambridge, Cambridge University Press.
- PLENAT M. (1996), « De l'interaction des contraintes. Une étude de cas », in J. Durand & B. Laks eds, *Currents trends in Phonology: Models and Methods*, Salford, University of Salford, pp. 585-615.
- (2000), « Quelques thèmes de recherche actuels en morphophonologie française », *Cahiers de lexicologie* 77, pp. 27-62.
- (2002), « Jean-Louis Fossat : fossatissime. Note sur la morphologie des dérivés en -issime », in L. Rabassa éd., *Mélanges offerts à Jean-Louis Fossat. Cahiers d'études romanes (CERCLiD)* 11-12, pp. 229-248.

- PLENAT M. & ROCHE M. (2003), « Prosodic constraints on suffixation in French », in G. Booij, J. De Cesaris, A. Ralli & S. Scalise eds, *Topics in Morphology. Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, september 20-22, 2001)*, Barcelona, Institut Universitari de lingüística aplicada, Universitat Pompeu Fabra, pp. 285-299.
- POIRIER C., MATHET Y. & ENJALBERT P. (1998), « La compositionnalité à l'épreuve des faits, à travers un projet de compréhension automatique de constats d'accident », *Traitement automatique des langues* 39/1, pp. 99-129.
- PRINCE A. & SMOLENSKY P. (1993), « Optimality Theory. Constraint interaction in generative grammar », Ms., Rutgers University and University of Colorado at Boulder.
- PUSTEJOVSKY J. (1995), *The Generative Lexicon*, Cambridge, Mass., the MIT Press.
- RAINER F. (1997), « Vers une contrainte sur les stades dérivationnels virtuels », in D. Corbin & al. eds (1997), pp. 231-240.
- (1998), « Paradigmatic factors in the irradiation of allomorphy: the reanalysis of the type MANUALIS in Italian », in G. Booij, A. Ralli & S. Scalise eds, *Proceedings of the first mediterranean conference of morphology*, Patras, University of Patras, pp. 77-85.
- REY-DEBOVE J. (2001), « La morphologie lexicale du français : mots complexes et nouveaux morphèmes », in H. Kronning, C. Norén, B. Novén, G. Ransbo, L. Sundell & B. Svane, *Langage et référence. Mélanges offerts à Kerstin Jonasson à l'occasion de ses soixante ans*, Uppsala, Acta Universitatis Upsaliensis.
- RITCHIE D.G., PULMAN S.G., BLACK A.W & RUSSELL G.J. (1992), *Computational morphology*, Cambridge / London, The MIT Press.
- ROCHE M. (1997), « Briard, bougeoir et camionneur : dérivés aberrants, dérivés possibles », in D. Corbin & al. eds (1997), pp. 241-250.
- (2003), « De la "bonne formation" des dérivés en -on », *Cahiers de grammaire* 28, pp. 91-112.
- ROGER C. (2003), « Pour une individualité sémantique des affixes : rejet de la notion de paradigmes morphologiques », in Fradin & al. eds (2003), pp. 179-187.
- RUBACH J. (2000), « Glide and glottal stop insertion in Slavic languages: a DOT analysis, *Linguistic Inquiry* 31, pp. 271-317.
- SANDER E. (2000), *L'analogie, du naïf au créatif. Analogie et catégorisation*. Paris, l'Harmattan.
- SCALISE S. (1984), *Generative Morphology*, Dordrecht, Holland / Cinnaminson, U.S.A., Foris Publications.
- SCHULTINK H. (1961), « Produktiviteit als Morfologisch Fenomeen », *Forum der Letteren* 2, pp. 110-125.
- SERBAT G. (1989), « La création lexicale », *L'information grammaticale* 42, pp. 3-6.
- SKOUSEN R. (1989), *Analogical modeling of language*, Dordrecht, Kluwer Academic.
- (1992), *Analogy and structure*, Dordrecht, Kluwer Academic.
- SPERBER D. & WILSON D. (1986), *La pertinence. Communication et cognition*, trad. franç., Paris, Les Éditions de Minuit, 1989.
- SPROAT R.W. (1992), *Morphology and Computation*, Cambridge, Mass. / London, England, The MIT Press.
- TAFT M. & FORSTER K. I. (1975), « Lexical storage and retrieval of prefixed words », *Journal of Verbal Learning and Verbal Behavior* 14, pp. 638-647.
- TEMPLE M. (1996), *Pour une sémantique des mots construits*, Villeneuve d'Ascq, Presses Universitaires du Septentrion.
- (2002), « Métaphore et mots construits : éclairages réciproques », *Verbum* XXIV/3, pp. 207-227.

- TOURATIER C. (1988), « La règle et l'exception : avant-propos », *Travaux* 6, Université d'Aix-en-Provence, pp. 7-12.
- ULLMAN M.T. (2001), « A neurocognitive perspective on language: the declarative/procedural model », *Nature Reviews Neurosciences* 2, pp. 717-726.  
[[http://www.giccs.georgetown.edu/labs/ullman/PUBS/Ullman\\_NatRevNS\\_01.pdf](http://www.giccs.georgetown.edu/labs/ullman/PUBS/Ullman_NatRevNS_01.pdf)]
- VAN DE VELDE (à paraître), « Grammaire des noms d'événements ».
- ZWANENBURG W. (2000), « [article 81] Correspondence between formal and semantic relations », in Booij & al. eds (2000), pp. 840-850.
- ZWICKY A. (1992), « Some choices in the theory of Morphology », in R. Levine ed., *Formal grammar: theory and implementation*, New York / Oxford, Oxford University Press, pp. 327-371.

## 2. Ressources

Frantext. [http://www.atilf.fr/\\_ns/produits/frantext.htm](http://www.atilf.fr/_ns/produits/frantext.htm)

*L'humanité*. <http://www.humanite.presse.fr/>

*Libé* = *Libération Les archives de janvier 1995 à décembre 1999*, CEDROM-SNi inc. (Montréal), 1999.

*LM* = *Le Monde sur CD-ROM*, SA Le Monde (Paris) – CEDROM-SNi inc. (Montréal), 1999.

*LMD* = *Le Monde Diplomatique 1984-1998*, CEDROM-SNi inc. (Montréal), 1998.

*PRO1* = *Le Petit Robert. Version électronique du Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française*, CD-ROM, nouv. éd. (version 2), Paris, Dictionnaires Le Robert / VUEF, 2001.

*RE* = *Le Robert électronique DMW*, Disque optique compact CD-ROM, Paris, Dictionnaires Le Robert, 1994.

*TLF* = *Trésor de la langue française. Dictionnaire de la langue du 19<sup>e</sup> et du 20<sup>e</sup> siècle (1789-1960)*, 16 vol., Paris, Éditions du CNRS (t. 1-10) / Gallimard (depuis le t. 11), 1971-1994.

WARTBURG W. von (1922-1978), *Französisches Etymologisches Wörterbuch*, Tübingen, Mohr ; puis Basel, Helbing und Lichteahn.



**Annexe 1.**  
**Liste de mes travaux depuis ma thèse**

## 1. Travaux publiés, sous presse et à paraître

### 1.1. Ouvrage individuel

- [0] (1997), *Grammaire du suffixe -et(te)*, Paris, Didier Érudition, collection “Études de sémantique lexicale”, 269 p., ISBN : 2-86460-305-5.  
[comptes rendus par J. van Marle, in G. Booij & J. van Marle eds (1998), *Yearbook of Morphology 1997*, p. 303, et par B. Fradin, *Le français moderne* LXIX 1, pp. 86-98.]

### 1.2. Autres travaux publiés et sous presse ([A] : article ; [C] : acte de colloque sans comité de lecture ; [CC] : acte de colloque avec comité de lecture ; [CR] : compte rendu)

- [1] (1996), « Comment le sens vient-il aux unités lexicales morphologiquement construites ? », in *Représentations et Outils pour les bases lexicales. Morphologie robuste*, Grenoble, CLIPS-IMAG, pp. 193-200. [CC]
- [2] (1997a), « Une approche théorique originale de la formation des unités lexicales construites morphologiquement : le modèle “SILEX” », in *Actes du IV<sup>e</sup> Colloque International de la Lexicologie, Les fondements théoriques du lexique*, Tunis, 2-5 mai 1997. [CC]
- [3] (1997c), « Tous les mots existants sont-ils possibles ? », in D. Corbin, B. Fradin, B. Habert, F. Kerleroux & M. Plénat édés, *Mots possibles et mots existants, Silexicales 1*, pp. 91-100. [CC]
- [4] (1997b), avec Martine Temple, « Morphologie dérivationnelle et analyse sémantique des mots construits : les voies de la référence ne sont pas impénétrables », in W.U. Dressler, M. Prinzhorn & J.R. Rennison eds, *Advances in Morphology*, Berlin / New York, Mouton de Gruyter, pp. 97-110. [A]
- [5] (1997d), « Du principe d’unicité catégorielle au principe d’unicité sémantique : incidence sur la formalisation du lexique construit morphologiquement », in P.-A. Buvet, S. Cardey, P. Greenfield & H. Madec édés, *Actes du colloque international Fractal’97, « Linguistique et informatique : théories et outils pour le traitement automatique des langues »*, BULAG numéro spécial, pp. 105-115. [CC]
- [6] (1999a), « Suffixation par *-et(te)* et bases verbales », in D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché édés, *La morphologie des dérivés évaluatifs, Silexicales 2*, pp. 37-47. [CC]
- [7] (1999b), avec Nabil Hathout et Fiammetta Namer, « Construire un lexique dérivationnel : théorie et réalisations », in *Actes de la VI<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN’99), Institut d’Études scientifiques de Cargèse, Corse, 12-17 juillet 1999*, pp. 115-124. [CC]
- [8] (2000a), avec Fiammetta Namer, « Génération et analyse automatiques de ressources lexicales construites utilisables en recherche d’information », *Traitement Automatique des Langues* 41/2, pp. 423-446. [A]
- [9] (2000b), avec Fiammetta Namer, « GéDériF: automatic generation and analysis of morphologically constructed lexical resources », in *Proceedings of the Second International Conference on Language Resources and Evaluation, Athens, Greece, 31 May–2 June 2000*, pp. 1447-1454. [CC]
- [10] (2001), « Un exemple de corpus enrichi d’informations constructionnelles et sémantiques : le projet MorTAL », in actes des Journées *Corpus et formatage* organisées par l’ILF, Ivry-sur-Seine, les 13-14 juin 2001, *Programmes fédératifs de l’institut de linguistique française (ILF)*, Paris, CNRS, pp. 97-108. [C]
- [11] (2002a), « À propos d’une idée reçue, ou de la prétendue irrégularité de la dérivation », *Bulag* 27, pp. 57-73. [A]

- [12] (2002b), avec Nabil Hathout et Fiammetta Namer, « An experimental constructional database: the MorTAL project », in P. Boucher ed., *Many morphologies*, Somerville Mass., Cascadilla Press, pp. 178-209. [A]
- [13] (2003a), « Arguments pour un préfixe *contre-* », *Recherches linguistiques* 26, pp. 172-201. [A]
- [14] (2003b), avec Fiammetta Namer, « Complex words vs. phrases: the case of causative verbs in French », in G. Booij, J. De Cesaris, A. Ralli & S. Scalise eds, *Topics in Morphology. Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, september 20-22, 2001)*, Barcelona, Institut Universitari de lingüística aplicada, Universitat Pompeu Fabra, pp. 129-148. [CC]
- [15] (2003c), « Productivité morphologique : définitions et notions connexes », *Langue Française* 140, pp. 3-23. [A]
- [16] (2003d), « Analogie et lexique construit : quelles preuves ? », *Cahiers de grammaire* 28, pp. 9-30. [A]
- [17] (2004a), Compte rendu de Huot (2001), *Morphologie. Forme et sens des mots du français*, Paris, Armand Colin, coll. "Campus", in *Revue Romane* 39/2, pp. 166-169. [CR]
- [18] (2004b), Compte rendu de Fradin (2003), *Nouvelles approches en morphologie*, Paris, PUF, in *Traitement automatique des langues* 45/2. [CR]
- [19] (sous presse a), avec Nabil Hathout et Fiammetta Namer, « Morphologie constructionnelle et traitement automatique des langues : le projet MorTAL », *Lexique* 16, 2004. [A]
- [20] (sous presse b), « À propos de *Nouvelles approches en morphologie* de Bernard Fradin : à quelle dimension des catégories sont sensibles les règles de construction de lexèmes ? », *Lexique* 16, 2004, pp. 229-261. [CR/A]
- [21] (sous presse c), avec Fiammetta Namer, « L'exception infirme-t-elle la notion de règle ? ou le lexique construit et la théorie de l'optimalité », *Faits de langue* 23, 2004. [A]
- [22] (sous presse d), « L'analogie dans le domaine du lexique construit : un retour ? », in B. Fradin éd., *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, Amsterdam / Philadelphia, John Benjamins Publishing Company. [A]

### 1.3. Travaux à paraître

- [23] (en préparation), avec Bernard Fradin et Nabil Hathout, « Morphologie évaluative : la suffixation en *-ette*, *-et* », in B. Fradin, F. Kerleroux & M. Plénat édés, *Aperçus de morphologie du français*. [A]

## 2. Travaux non publiés

### 2.1. Article non publié

- [24] « *Chardonneret*, *chardonnette*, *écharдонnet(te)* : la grammaire du suffixe *-et(te)* en (presque) trois mots ».

### 2.2. Communications ne donnant pas lieu à publication

- [25] (2000a), « Le projet MorTAL », poster présenté à l'occasion d'une conférence de presse donnée par monsieur Claude Allègre, Paris, Cité des sciences et de l'industrie, 3 mars 2000.
- [26] (2000b), « Le projet MorTAL », communication à la journée d'étude « Morphologie constructionnelle et TAL » organisée par G. Dal, N. Hathout et F. Namer sous l'égide de l'ATALA (« Association pour le Traitement Automatique des Langues »), CHU-Pitié-Salpêtrière, 25 mars 2000.

- [27] (2002), avec Fiammetta Namer, « Opérations constructionnelles et bases verbales : problèmes de représentation », communication à la journée d'étude sur les noms déverbaux, organisées par l'UMR SILEX, sous le patronage de l'institut de linguistique française (ILF), 15-16 mars 2002.
- [28] (2002), avec Dany Amiot, « The absence of determiners on nominal bases in prefixation and composition in French and autonomy of morphology », *10th International Morphology Meeting*, Szentendre (Hungary), 9-12 May 2002.
- [29] (2003a), « La productivité : quelques préalables », communication à la journée d'étude « La productivité : regards actuels » organisée par l'action 5 du GDR 2220 « Description et modélisation en morphologie », Paris, CHU Pitié-Salpêtrière, 12 mai 2003.
- [30] (2003b), avec Bernard Fradin, Nabil Hathout et Fanny Meunier, « Productivités comparées : le cas de la suffixation en *-ET* », communication à la journée d'étude « La productivité : regards actuels » organisée par l'action 5 du GDR 2220 « Description et modélisation en morphologie », Paris, CHU Pitié-Salpêtrière, 12 mai 2003.
- [31] (en préparation), avec Stéphanie Lignon, Fiammetta Namer et Ludovic Tanguy, « Toile contre dictionnaires : analyse morphologique en corpus de noms déverbaux concurrents », communication au colloque « Les noms déverbaux », UMR SILEX & Université de Lille 3, 23-25 septembre 2004.

### 2.3. Conférences

- [32] (1995a), « Des impasses de l'analyse concaténaire en morphologie dérivationnelle : application au suffixe *-et(te)* », Paris X-Nanterre, mars 1995.
- [33] (1995b), « Le suffixe *-et(te)* : un suffixe mono-instructionnel », conférence au séminaire d'équipe de l'UMR SILEX, 15 novembre 1995.
- [34] (1997a), « Un point de vue sur la morphologie dérivationnelle du français », série de cours donnés dans le cadre de l'école d'été en morphologie organisé par COLEX, Université de Nantes, 1-8 septembre 1997.
- [35] (1997b), « Le modèle SILEX : un modèle dé-génératif », série de conférences données dans le cadre de la coopération scientifique et technique luso-française, faculté de lettres de l'université de Lisbonne (Portugal), 29-30 octobre 1997.
- [36] (2001a), « Construire une unité lexicale : plus qu'une simple question de forme(s) », conférence présentée dans le cadre du séminaire ILD-ISTC du Loria, Nancy, 22 mars 2001.
- [37] (2001b), « Arguments pour un préfixe *contre-* », conférence au séminaire d'équipe de l'UMR SILEX, 4 mai 2001.
- [38] (2004), « À quelle dimension des catégories les règles de construction de lexèmes sont-elles sensibles ? », conférence au séminaire d'équipe de l'UMR SILEX, 3 avril 2004.

### 3. Édition d'ouvrages collectifs

- [39] (1999), avec D. Corbin, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché, *La morphologie des dérivés évaluatifs*, Actes du colloque organisé à l'Université de Toulouse-le Mirail, 29-30 avril 1999, Toulouse, *Sillexicales 2*, publication de l'UMR SILEX.
- [40] (2003), avec B. Fradin, N. Hathout, F. Kerleroux, M. Plénat & M. Roché, *Les unités morphologiques*, Actes du colloque organisé à l'Université de Lille 3, 19-23 septembre 2002, Villeneuve d'Ascq, *Sillexicales 3*, publication de l'UMR SILEX.
- [41] (2003), *Langue Française* 140, « La productivité en questions et en expérimentations ».
- [42] (2004), avec A.-M. Berthonneau, *La linguistique de la LSF : recherches actuelles*, Actes du colloque organisé à l'Université de Lille 3, 23-24 septembre 2003, Villeneuve d'Ascq, *Sillexicales 4*, publication de l'UMR SILEX.



## Annexe 2.

# Liste des travaux générés par le projet MorTAL

[Les travaux sont cités par ordre chronologique]

### 1. Travaux immédiatement liés au projet MorTAL

#### 1.1. Travaux publiés

- DAL G, HATHOUT N. & NAMER F. (1999), « Construire un lexique dérivationnel : théorie et réalisations », in *Actes de la VI<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN'99)*, Institut d'Etudes Scientifiques de Cargèse, Corse, 12 -17 juillet 1999, pp. 115-124.
- HATHOUT N. (2000), « Morphological Pairing based on the Network Model », in *Proceedings of the Workshop on Computational Lexicography and Multimedia Dictionaries (COMLEX 2000)*, Pygos, Grèce, pp. 35-38.
- DAL G. (2001), « Un exemple de corpus enrichi d'informations constructionnelles et sémantiques : le projet MorTAL », in actes des Journées *Corpus et formatage* organisées par l'ILF, Ivry-sur-Seine, les 13-14 juin 2001, *Programmes fédératifs de l'institut de linguistique française (ILF)*, Paris, CNRS., pp. 97-108.
- HATHOUT N., NAMER F. & DAL G. (2002), « An experimental constructional database: the MorTAL project », in P. Boucher ed., *Many morphologies*, Somerville, Mass., Cascadilla Press, pp. 178-209.
- DAL G, HATHOUT N. & NAMER F. (sous presse), « Morphologie constructionnelle et traitement automatique des langues : le projet MorTAL », *Lexique* 16, 2004.

#### 1.2. Communications, conférences

- HATHOUT N. (1999), « DéCor : une méthode basée sur findaffix pour la construction d'une base de données morphologique », séminaire de l'ERSS, Université de Toulouse-le Mirail, mars 1999.
- DAL G. (2000a), « Le projet MorTAL », communication à la journée d'étude *Morphologie constructionnelle et TAL* organisée par G. Dal, N. Hathout & F. Namer, sous l'égide de l'ATALA (Association pour le Traitement Automatique des Langues), CHU-Pitié-Salpêtrière, 25 mars 2000.
- DAL G. (2000b), « Présentation du projet MorTAL », poster présenté à l'occasion d'une conférence de presse donnée par monsieur Claude Allègre, Paris, Cité des sciences et de l'industrie, 3 mars 2000.
- NAMER F. (2002), « Analyse Automatique du lexique construit en français : le système DériF », Séminaire TALANA, Paris, Université de Paris 7, 10 juin 2002.
- (2003), « Acquisition de sens construit. L'analyseur morpho-sémantique du Français DériF », communication à la journée d'étude du pôle ILD-ISTC, INIST-CNRS, Nancy, 30 janvier 2003.

#### 1.3. Journée d'étude

- (2000), *Morphologie constructionnelle et TAL*, Journée d'études organisée par G. Dal, N. Hathout et F. Namer, sous l'égide de l'ATALA (Association pour le Traitement Automatique des Langues), CHU-Pitié-Salpêtrière, le 25 mars 2000.

## 2. Travaux dérivés du projet

### 2.1. Système DériF

- DAL G & NAMER F. (2000), « Génération et analyse automatiques de ressources lexicales construites utilisables en recherche d'informations », *Traitement Automatique des Langues* 41-2, pp. 423-446.
- NAMER F. & DAL G. (2000), « GéDériF: automatic generation and analysis of morphologically constructed lexical resources », *Second International conference on language resources and evaluation (LREC), Athens, Greece, 31 May–2 June 2000*, pp. 1447-1454.
- NAMER F. (2001), « Génération automatique de néologismes bilingues morphologiquement construits en français et italien », in *Actes de la VIII<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN'01), tome 2, Université de Tours*, pp. 287-296.
- (2002), « Acquisition automatique de sens à partir d'opérations morphologiques en français : études de cas », in *Actes de la IX<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN'02), Nancy*, pp. 235-244.
- (2003a), « Productivité morphologique, représentativité et complexité de la base : le système MOQUETE », *Langue Française* 140, pp. 79-101.
- (2003b), « Automatiser l'analyse morpho-sémantique non affixale : le système DériF », *Cahiers de Grammaire* 28, pp. 31-48.
- (2004), « Acquisizione automatica di semantica lessicale in francese: il sistema di trattamento computazionale della formazione delle parole DériF », in *Atti del 27<sup>e</sup> Congresso Internazionale della Società Italiana di Linguistica*.
- NAMER F. & ZWEIGENBAUM P. (à paraître), « Acquiring meaning for French medical terminology: contribution of morphosemantics », in *Actes de MEDINFO 2004, San Francisco*.

### 2.2. Système DéCor

- HATHOUT N. (2001), « Analogies morpho-synonymiques. Une méthode d'acquisition automatique de liens morphologiques à partir d'un dictionnaire de synonymes », in *Actes de la VIII<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN'01), tome 1, Université de Tours*, pp. 223-232.
- (2002), « From WordNet to CELEX: acquiring morphological links from dictionaries of synonyms », in *Proceedings of the Third International Conference on Language Resources and Evaluation, Las Palmas de Gran Canaria, Espagne. ELRA*, pp. 1478-1484.
- HATHOUT N. & TANGUY L. (2002), « Webaffix: finding and validating morphological links on the WWW », in *Proceedings of the Third International Conference on Language Resources and Evaluation, Las Palmas de Gran Canaria, Espagne. ELRA*, pp. 1799-1804.
- TANGUY L. & HATHOUT N. (2002), « Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du Web », in *Actes de la IX<sup>e</sup> conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'02), Nancy*, pp. 245-254.
- HATHOUT N. (2003), « L'analogie, un moyen de croiser les contraintes et les paradigmes », in Duvigneau, K., Gasquet, O. et Gaume, B. éd., *Revue d'Intelligence Artificielle* 17-5/6.

# Index des auteurs

## A

Aitchison, 71  
Aliquot-Suengas, 61  
Amiot, 34, 61, 93  
Anderson, 21, 27, 49, 67, 83  
Archangeli, 75  
Aronoff, 1, 6, 7, 8, 59  
Aurnague, 54

## B

Baayen, 17, 18, 87  
Baker, 65  
Bauer, 80, 84  
Beard, 1, 63  
Boersma, 75  
Boersma & *al.*, 75  
Booij, 7, 22, 24, 50, 51, 57, 80, 88, 93  
Booij & *al.* eds, 22  
Bréal, 81  
Burzio, 2, 73  
Bybee, 72

## C

Cadiot, 18  
Charconnet, 83  
Chomsky, 59  
Clahsen, 82  
Clahsen & *al.*, 82  
Clément, 62  
Corbin & Corbin, 10, 16  
Corbin & Paul, 1  
Croft, 57, 58, 65, 66, 69, 89

## D

D. Corbin, 1, 2, 5, 6, 7, 8, 9, 10, 11, 21, 22, 24, 29, 43,  
49, 50, 55, 58, 60, 61, 63, 64, 71, 75, 80, 88, 93  
D. Corbin & *al.*, 5, 8  
Daille, 27, 31  
Daille & *al.*, 27, 31  
Dal, 5, 8, 61, 85  
Darmesterer, 93  
Dell, 6  
Derwing, 72  
Di Sciullo, 80, 92  
Diez, 84  
Doleschal, 14  
Dressler, 14, 43, 80, 84  
Dubois, 28, 45  
Durand, 35

## F

Fehringer, 72  
Flaux, 93  
Forster, 71  
Fradin, 1, 2, 5, 6, 9, 10, 13, 16, 17, 18, 19, 21, 24, 25,  
26, 30, 35, 39, 47, 49, 63, 65, 66, 68, 69, 70, 73,  
80, 82, 92  
Fradin & *al.*, 17, 19, 24, 25  
Franckel, 31, 34  
Frauenfelder, 71  
Froissart, 28  
Fu & *al.*, 6  
Fuchs, 28

## G

Gerhard-Krait, 51, 54  
Gineste, 83  
Godard, 17  
Godart-Wendling, 21  
Grabar, 27  
Grimshaw, 17  
Guilbert, 76

## H

Habert, 10, 93  
Halle, 6  
Hasselrot, 35  
Hathout, 10, 16, 18, 32, 39  
Hathout & *al.*, 10, 32, 39  
Hay, 72  
Hoeksema, 22  
Huot, 45, 46, 49, 84

## I

Iacobini, 22

## J

Jackendoff, 9, 70  
Jacquemin, 28  
Jalenques, 34  
Janda, 80  
Jayez, 17  
Jespersen, 59  
Joseph, 80

## K

Kerleroux, 1, 16, 17, 39, 68  
Koenig, 81  
Kuhn, 95

Kurylowicz, 83

## L

Ladányi, 80, 84  
Laks, 1, 35  
Lallich-Boidin, 28  
Lass, 80  
Lees, 1  
Lepage, 83  
Lieber, 6, 18, 22, 80  
Lignon, 64  
Lyons, 47

## M

Matthews, 1  
McCarthy, 74  
Merlini-Barbaresi, 14  
Meunier, 16  
Molino, 1  
Mugdan, 22

## N

Namer, 33, 39, 40, 42, 77  
Nazarenko, 22  
Neijt, 87  
Nicolas, 72

## P

P. Corbin, 1  
Pamies, 17  
Perruchet, 72  
Pinker, 72  
Plag, 2, 18, 55, 63, 67, 75, 82, 88, 90, 96  
Plénat, 10, 11, 35, 36, 43, 46, 54, 64, 73, 74  
Poirier, 22  
Prince, 72, 73, 74  
Pustejovsky, 19

## R

Rainer, 61, 91  
Rey-Debove, 49  
Richie, 35

Richie & *al.*, 35  
Roché, 35, 36, 43, 64, 75  
Roger, 56, 63  
Rubach, 2

## S

Sander, 83  
Scalise, 59, 60, 61, 80  
Schreuder, 71  
Schultink, 81  
Serbat, 67  
Skousen, 72, 83  
Smolensky, 73  
Sperber, 14  
Sproat, 28, 36

## T

Taft, 71  
Temple, 18, 39, 84  
Ten Hacken, 82  
Thornton, 14  
Touratier, 73  
Tzoukerman, 28

## U

Ullman, 72

## V

Van de Velde, 17  
van Marle, 81  
Van Marle, 24, 93

## W

Williams, 80, 92  
Wilson, 14

## Z

Zwanenburg, 22  
Zweigenbaum, 10, 27, 93  
Zwicky, 5, 70