

Université de Toulouse

Mémoire présenté pour l'obtention
d'une habilitation à diriger des recherches

Spécialité : linguistique

**La morphophonologie de l'italien :
une approche lexicaliste**

Fabio Montermini

2010

La morphophonologie de l'italien

Table des matières

Introduction	7
Partie 1. Parcours scientifique	13
1.1 Introduction	13
1.2 La recherche et l'exploitation des données en morphologie	15
1.3 La classification des unités morphologiques	17
1.4 La représentation des unités lexicales	23
1.5 Processus cognitifs et grammaticalisation	27
1.6 Conclusion	28
Partie 2. Pour un modèle lexical de la morphologie	31
2.1 La morphologie lexématique: aperçu	31
2.2 Une morphologie basée sur les mots	34
2.2.1 Définition du mot: problèmes empiriques	34
2.2.2 Définition du mot: problèmes théoriques	36
2.2.3 Définition du mot: conclusion	38
2.3 Caractérisation du lexème	39
2.3.1 Représentation lexicale des lexèmes	39
2.3.2 Variation des lexèmes: premier aperçu	40
2.4 Que reste-t-il des morphèmes ?	43
2.4.1 Problèmes posés par la morphologie incrémentale	43

2.4.2 Une morphologie inférentielle	44
2.4.3 Conséquences de la dérivation des morphèmes	48
2.4.4 Statut des unités morphologiques: racines, thèmes, affixes	50
2.4.5 Qu'est-ce qui est stocké dans le lexique ?	51
2.5 Une morphologie réalisationnelle	53
2.5.1 Caractérisation	53
2.5.2 Conséquences de l'approche réalisationnelle: des simples formes	54
2.6 Une morphologie émergente	59
2.6.1 Caractérisation	59
2.6.2 Règles vs. analogie	59
2.6.3 Une morphologie à contraintes	65
2.6.4 La directionnalité des procédés morphologiques	69
2.7 Lexique vs. compétence morphologique	75
2.7.1 Théories du lexique	75
2.7.2 Qu'est-ce que le lexique ?	77
2.7.3 Qu'est-ce qu'un mot régulier ?	78
2.7.4 L'illusion du lexique parfait	83
2.7.5 Lexique et compétence morphologique	86
2.7.6 Qu'est-ce qu'il y a dans le lexique ?	88
2.7.7 Listèmes et lexèmes	97
2.7.8 Mots possibles et mots existants	99
2.8 Une morphologie thématique	100
2.8.1 La morphologie thématique: caractérisation	100
2.8.2 Allomorphie et supplétion	101

2.8.3 Allomorphie phonologique et morphologique	105
2.8.4 Défectivité, syncrétisme et identité lexémique	109
2.8.5 La morphologie thématique et la structure des lexèmes	119
2.9 Conclusion: quelques problèmes résidus	136
2.9.1 La question des données	136
2.9.2 Descriptif vs. explicatif	138
Partie 3. La morphophonologie de l'italien dans une perspective lexicaliste: Etudes de cas	149
3.1 Introduction	149
3.2 L'accent primaire de mot en italien	151
3.2.1 Présentation	151
3.2.2 L'algorithme accentuel de l'italien	153
3.2.3 Les traitements de l'accentuation en italien	157
3.2.4 L'accentuation par défaut en italien	158
3.2.5 L'accentuation en italien: un test	163
Appendice : corpus de mots qui présentent des variations d'ac- cent	171
3.3 La représentation phonologique des affixes: une analyse du suffixe <i>-(i)ano</i>	172
3.3.1 Présentation: <i>-(i)ano</i> en italien	172
3.3.2 Un suffixe ou deux ?	173
3.3.3 Contraintes phonologiques	174
3.3.4 <i>-(i)ano</i> dans les dictionnaires de l'italien	176
3.3.5 La disponibilité de <i>-(i)ano</i>	179

La morphophonologie de l'italien

3.3.6 <i>-(i)ano</i> déonomastique	180
3.3.7 <i>-(i)ano</i> dans les ethniques	182
3.3.8 La sémantique de <i>-(i)ano</i>	184
3.3.9 La phonologie de <i>-(i)ano</i>	185
Bibliographie	192

Fabio Montermini

Introduction

i segnali spesso non significano mai

[Baustelle, *Il nulla*]

Ce mémoire, que je présente pour mon Habilitation à diriger des recherches, est le fruit des presque quinze années que j'ai passées à m'occuper, à temps plein, de linguistique. Il s'agit d'une étape importante dans ma carrière de chercheur (et d'enseignant) de cette discipline. Il est rare, en effet, qu'on se pose pour regarder derrière soi, pour tracer un bilan de ce qu'on a fait, et essayer de tirer les fils d'une manière qui soit cohérente. Le travail de chercheur, en particulier aujourd'hui, même s'il ne devrait pas l'être, est surtout orienté vers le présent et le futur immédiats. Le prochain article qu'on doit écrire, la prochaine communication qu'on doit présenter, le prochain projet qu'on doit soumettre – sans oublier les cours à préparer et les tâches administratives à assurer – focalisent généralement notre attention pour un temps assez court, et on a rarement l'occasion de lever la tête du guidon et de regarder en prenant un peu de hauteur ce qu'on est en train de faire. On peut considérer cela une chance, presque unique, que le système universitaire français nous donne. Si j'étais resté travailler dans mon pays d'origine, l'Italie, jamais je n'aurais eu besoin, au cours de ma carrière, de présenter une habilitation. J'ai donc essayé de saisir cette opportunité et de me donner un peu de recul pour reconsidérer les travaux que j'ai réalisés depuis que je m'intéresse à la linguistique, et en particulier depuis ma thèse de doctorat, que j'ai commencée en 1998 et soutenue en 2002. Il s'agit d'une activité qui peut être dé-

plaisante ; j'ai tout de même eu quelques bonnes surprises : des articles ou des communications que je considérais comme moins réussis que d'autres m'ont finalement paru meilleurs que le souvenir qu'ils m'avaient laissé. Dans ce mémoire, et notamment dans sa partie 2, j'essaie donc de résumer quelle est ma vision actuelle de la linguistique, et en particulier de la morphologie, la discipline dont je m'occupe plus directement. Certes, il resterait encore plusieurs points à éclaircir, plusieurs sujets de recherche qui sont seulement évoqués ou ébauchés dans les pages qui suivent, mais je propose tout de même un modèle qui présente une certaine cohérence dans les méthodes et dans les questions qu'il se pose.

S'il m'est permis de m'auto-citer, dans l'introduction de ma thèse de doctorat (2002), je disais que « [u]n des enjeux de tout travail de linguistique est celui d'être capable de situer son objet d'étude dans un modèle plus ample du fonctionnement du langage. En d'autres termes, un bon travail est celui qui, observant un objet particulier, même minuscule, permet de formuler des hypothèses sur le fonctionnement du langage en général ». Ce sont des affirmations que, huit ans après, je souscris encore entièrement. Trop souvent il m'est arrivé de lire ou écouter des travaux de linguistes en me demandant « et alors ? ». Très souvent les linguistes se limitent à observer un petit phénomène dans une langue, sans même nous laisser apercevoir quel est l'intérêt plus global de leur démarche. Au contraire, je considère que, quel que soit le phénomène qu'on observe, il ne faudrait pas perdre de vue le fait que ce qui nous intéresse se trouve derrière, c'est la capacité que les êtres humains ont de reproduire la réalité tels qu'ils la perçoivent (et même l'irréalité) par des mouvements de quelques-uns de leurs organes. Par le biais de la compétence morphologique et lexicale, c'est donc d'une capacité plus générale que j'ai voulu parler dans les pages qui suivent. Si je n'ai pas réussi, ce n'est pas faute d'avoir essayé.

Un moment crucial dans ma carrière de chercheur a été, sans surprise, mon arrivée au sein de l'équipe (à l'époque) ERSS de Toulouse. C'est à partir de mon intégration au groupe de morphologie de cette équipe que la plupart des idées et des réflexions contenues dans ce mémoire sont nées et ont mûri. Quelques-unes étaient déjà présentes, à l'état d'embryons, dans mon esprit ; pour d'autres, elles n'existeraient simplement pas. Une réflexion pour laquelle mon arrivée à l'ERSS a été déterminante est celle qui concerne les données. Depuis que je m'intéresse à la linguis-

tique, j'ai toujours été intéressé, entre autres, par des données 'marginales', quelqu'un dirait périphérique, sous-entendant qu'elles ne sont pas intéressantes. Au contraire, j'ai toujours été convaincu que toutes les données, y compris les périphériques méritent d'être prises en compte, parce qu'elles peuvent contribuer à jeter un éclairage sur des aspects que les données centrales peuvent cacher.

La première partie du mémoire offre un compte-rendu de mon parcours scientifique, qui s'est déroulé selon trois lignes directrices principales : la classification des unités morphologiques ; la représentation formelle des unités lexicales ; l'interaction entre processus cognitifs et grammaticalisation. La deuxième partie présente un modèle global de la morphologie, qui peut être caractérisé de réalisationnel, émergent et thématique. Tout d'abord, deux modèles concurrents de la morphologie, un basé sur les morphèmes et un basé sur les lexèmes, y sont présentés, et il est démontré que le second est plus adéquat, aussi bien du point de vue empirique que théorique. Une morphologie fondée sur les mots est plus compatible avec une approche non additive, comme l'approche réalisationnelle qui est défendue ici, dans laquelle les propriétés morphosyntaxiques ne sont pas nécessairement construites par une adjonction successive de morphèmes. Les processus morphologiques sont donc analysés non pas comme des règles orientées, mais comme des schémas non orientés, qui se construisent comme des généralisations à partir du lexique existant, d'où la caractérisation de la morphologie comme émergente. Cette vision permet de modéliser la notion d'analogie, en proposant que la compétence morphologique agit de manière probabiliste et en donnant un poids à des paramètres comme la proximité entre des unités lexicales ou la densité de leur environnement. Une thèse supplémentaire défendue ici est que cette conception du lexique est plutôt compatible avec une théorie enrichie du lexique qui prévoit que les unités puissent y être stockées de manière redondante. On peut qualifier une telle approche de « thématique » dans la mesure où les différentes formes qui peuvent instancier un lexème ne sont pas réduites à une seule forme de base, y compris dans les cas les plus réguliers. Un traitement plus satisfaisant de la morphologie à nos yeux consiste à prendre en compte l'espace de variation des lexèmes dans les différentes langues. Les notions d'allomorphie et de supplétion, et leurs conditionnements (phonologique ou morphologique) font l'objet d'une discussion approfondie. Enfin, quelques points qui restent problématiques, comme la

question des données sur lesquelles doit se fonder une étude morphologique et la dichotomie entre description et explication dans les études linguistiques sont abordés. La troisième partie propose deux études de cas qui montrent l'efficacité du modèle présenté pour rendre compte de l'interaction entre la compétence linguistique et le lexique : l'assignation de l'accent primaire en italien et la dérivation d'adjectifs en *(i)ano* à partir, principalement, de noms propres et de toponymes.

En conclusion de cette introduction, je souhaiterais remercier toutes les personnes qui ont contribué à la réalisation de ce travail, et plus en général qui sont à la base de ma réflexion sur la morphologie et sur la langue. Je suis reconnaissant en particulier à Jacques Durand, qui a accepté de diriger mon habilitation, pour sa disponibilité et ses conseils vraiment précieux. Je suis également reconnaissant à Georgette Dal, Giovanna Marotta, Fiammetta Namer, Patrick Sauzet et Anna M. Thornton, qui ont accepté de faire partie du jury, et pour leurs observations et conseils qui m'ont permis d'améliorer ce mémoire. Je souhaite aussi remercier de façon globale les collègues de CLLE-ERSS, pour leur soutien, pour les discussions intéressantes que j'ai eues avec la plupart d'entre eux et pour contribuer à créer un environnement de travail vraiment agréable. Plus particulièrement, je remercie les membres de l'ex axe « Des unités morphologiques au lexique », Gilles Boyé, Hélène Giraud, Aurélie Guerrero, Nabil Hathout, Michel Roché et Jesse Tseng, sans lesquels la plupart des idées actuelles que j'ai sur la morphologie n'existeraient simplement pas, ainsi que Josette Rebeyrolle et Ludovic Tanguy. Parmi les coresponsables de ma vision actuelle de la morphologie il faut citer aussi les membres de l'ex GDR 2220, qui participent actuellement au projet de manuel de morphologie, qui sont (en plus de quelques-uns que j'ai déjà cités) Dany Amiot, Bernard Fradin, Françoise Kerleroux, Stéphanie Lignon et Florence Villoing. Je remercie aussi Jean-François Bonnefon, Laurence Lamy et Nathalie Moullic pour leur compréhension et leur aide, qui m'ont permis de me concentrer au maximum sur ce travail, ainsi que Corinne Prunier pour la compétence et la gentillesse avec lesquelles elle a satisfait toutes mes demandes bibliographiques. Parmi les autres personnes qui m'ont, pour une raison ou pour une autre, aidé dans la réalisation de ce mémoire je n'oublie pas Giorgio Francesco Arcodia, Olivier Bonami, Basilio Calderone, Nicola Grandi, Katrin Landmann, Daniela Müller, Martina Nied, Diana Passino, Vito Pirrelli, Carmen Scherer, Fabio Tamburini, Dephine Tribout. Enfin,

Fabio Montermini

je ne peux pas ne pas remercier Giovanna pour sa patience, son soutien continu, et pour avoir même lu des parties de ce mémoire !

La morphophonologie de l'italien

Partie 1

Parcours scientifique

1.1 Introduction

La première partie de ce mémoire est consacrée à une description de mon parcours scientifique. Le but est de montrer comment des idées qui étaient déjà présentes dans mes premiers travaux se sont enrichies de nouveaux thèmes et se sont développées dans une vision globale et organisée de la compétence morphologique qui sera décrite dans la deuxième partie, et que j'illustrerai par des études de cas dans la troisième. Les travaux que j'ai réalisés au cours de ma carrière scientifique ont été répartis autour de trois thématiques principales : (i) la classification des unités morphologiques ; (ii) la représentation des unités lexicales, notamment du point de vue morphophonologique ; (iii) l'interaction entre processus cognitifs généraux et procédés morphologiques qui sont grammaticalisés dans une langue donnée. Ces trois thématiques seront développées dans les sections 1.3 à 1.5. Au préalable, je présente en 1.2 mes tout premiers travaux en linguistique, dans lesquels a émergé principalement la question des données sur lesquelles doit s'appuyer une étude linguistique, et plus particulièrement morphologique et lexicale. Au fil du temps, il m'est apparu de plus en plus clairement que cette question est loin d'être anecdotique. Au contraire, elle est centrale, et sous-tend la plupart de mes réflexions actuelles (cf. aussi la section 2.9.1).

Naturellement, les thématiques que j'ai identifiées ne sont pas des ensembles étanches ; pour quelques-uns de mes travaux l'inclusion dans une catégorie plutôt que dans une autre a été franchement problématique et, dans certains cas, reste quelque peu arbitraire. Les trois thématiques en question sont en effet strictement liées, et la segmentation que je propose a forcément été faite *a posteriori*, sur la base de l'observation rétrospective de mon propre travail. Il y a, tout de même, une notion qui est centrale dans l'ensemble de mes recherches, et qui constitue, d'un certain point de vue, le fil rouge de mon parcours scientifique : c'est la notion de mot. Dès lors que j'ai commencé à m'intéresser à la linguistique j'ai été préoccupé par des questions d'ordre lexical, et quand je me suis orienté vers la morphologie, lors de mon inscription en D.E.A. en 1997, le mot m'est apparu comme un bon candidat pour être l'unité de base de l'organisation morphologique et lexicale. Cette orientation théorique est en partie due, bien entendu, à des circonstances personnelles et liées au contexte dans lequel je me suis trouvé. Il s'avère qu'entre la fin des années 1990 et le début des années 2000 la morphologie basée sur les mots était dans une période de fort développement, surtout en Europe. En particulier, dans les deux pays dans lesquels j'ai effectué l'ensemble de mon activité les communautés de morphologues les plus actives se sont créées, à peu près à la même époque, autour du travail de deux chercheurs, Danielle Corbin en France et Sergio Scalise (qui a été le co-directeur de ma thèse de doctorat) en Italie, qui ont développé, tous les deux, des modèles de morphologie basés sur les mots.

Les réflexions que je présente dans les parties 2 et 3 représentent une tentative d'intégration des trois thématiques que j'ai identifiées comme structurantes pour mon travail dans une vision globale visant à expliquer comment la compétence morphologique se construit, comment elle est organisée et comment elle interagit avec la connaissance lexicale. Bien qu'il soit difficile de prévoir avec certitude les directions que l'activité scientifique d'un individu prendra dans l'avenir, ces deux parties constituent aussi un exposé programmatique de l'orientation de mon travail de recherche à court et à moyen terme.

1.2 La recherche et l'exploitation des données en morphologie

Le premier véritable travail de linguistique que j'ai réalisé a été mon mémoire de maîtrise (tesi di laurea) en langues étrangères, que j'ai soutenu à l'Université de Parme (Italie) en novembre 1995 sous la direction du prof. Gianpiero Piretto – mémoire consacré aux mots d'origine anglaise dans le langage jeune russe (partiellement repris dans Montermini 1997). Ce travail contenait déjà quelques-uns des thèmes que j'allais aborder dans mes travaux ultérieurs, en particulier l'importance des faits lexicaux, et l'idée que l'observation de phénomènes apparemment marginaux peut fournir un éclairage sur le fonctionnement plus global de la compétence linguistique.

En 1998 j'ai obtenu un D.E.A. en Sciences du langage à l'université de Paris X – Nanterre sous la direction de Françoise Kerleroux. Mon mémoire s'intitulait *Raccourcissements et autres phénomènes de morphologie 'mineure' en italien contemporain*. Dans ce travail, j'ai analysé les phénomènes d'apocope en italien, présents surtout dans les variétés non standard (cf. *automobile* → *auto* 'voiture'), et je les ai mis en relation avec d'autres phénomènes morphologiques, comme la composition par réduction ou la formation d'hypocoristiques. Le focus de cette recherche, qui s'inspirait de travaux déjà réalisés sur le français, l'italien et d'autres langues, était principalement morpho-phonologique, dans le sens que j'ai essayé d'y dégager des principes prosodiques pour la construction d'apocopes expressives en italien, et morphologique, puisque j'y ai traité ce type de constructions comme un procédé de construction de lexèmes. Pour la première fois je me suis servi de la notion de lexème, conçu comme une unité qui combine des informations d'ordre formel, catégoriel et sémantique, et j'ai commencé à réfléchir au mot comme unité de base de l'organisation morphologique. Encore une fois, l'observation de formes sociolinguistiquement marquées et qui, de ce fait, sont moins susceptibles de faire partie du lexique établi d'une langue, a permis – c'est du moins l'hypothèse que j'avais formulée – d'observer plus directement les processus mentaux réellement à l'œuvre dans ce type de constructions. Dans tous mes travaux successifs, je ne me suis jamais privé de me servir, pour mon argumentation, de données 'bizarres', provenant de différentes sources : dictionnaires de néologismes, de langues de spécialité ou de registres particuliers (argots ou autres), presse écrite ou audiovisuelle et plus récemment le Web. Dans d'autres cas, j'ai réalisé des enquêtes

linguistiques informelles pour tester les intuitions des locuteurs sur certains points spécifiques. Au delà du fait qu'elles sont en général plus amusantes, ces données, qui seraient probablement considérées comme marginales ou déviantes, y compris par des linguistes théoriciens, en principe opposés à une approche prescriptive, constituent, à mon sens, une source précieuse pour observer le fonctionnement réel de la compétence linguistique, au moins dans le domaine de la formation de mots. Contrairement à ce qui est généralement admis, du moins implicitement, la 'centralité' sociolinguistique et la centralité linguistique ne sont pas toujours parallèles, et on ne saurait expliquer comment elles pourraient l'être, vu que les deux critères de classification en question font référence à des domaines séparés et incomparables entre eux.

Une étape importante pour la mise en perspective et l'explicitation de ces réflexions sur les données a été mon arrivée dans l'ex axe « Des unités morphologiques au lexique » de l'ERSS, en 2003. Mon installation à Toulouse a coïncidé avec le moment où le Web a commencé à s'imposer comme source importante de données pour les recherches linguistiques. Dans ce domaine, l'équipe toulousaine de morphologie a été pionnière, notamment grâce à l'impulsion donnée par Marc Plénat (cf. Plénat 2000 pour des observations sur l'emploi des ressources électroniques pré-Web, et, entre autres, Plénat 2005 ; Plénat *et al.* 2002 ; Hathout *et al.* 2002 ; 2009, pour des analyses morphologiques qui s'appuient sur le Web). Aujourd'hui, il est devenu courant de trouver des données issues, par exemple, de requêtes effectuées sur des moteurs de recherche utilisées pour illustrer des analyses, y compris en linguistique théorique. Dans les dernières années, la collecte de données sur le Web n'a pas constitué un simple complément à la réflexion théorique, au contraire, elle a apporté un changement qualitatif important et a permis de modifier, parfois de manière substantielle, la théorie. Le Web donne accès à un très grand nombre de données, pour la plupart spontanées, ce qui était inespéré il y a encore quelques années. Si son utilisation pose des problèmes (par exemple liés au non contrôle des données rendues disponibles par cette ressource et au non équilibrage intrinsèque des corpus ainsi recueillis), il rend néanmoins possible l'accès à une quantité de données jamais atteinte auparavant, des données d'autant plus intéressantes qu'elles ne proviennent pas d'une ressource conçue pour d'autres finalités (par exemple un dictionnaire). L'apport des données spontanées et non 'institutionnalisées', dont j'avais seulement pressenti l'im-

portance lors de mes premiers travaux cités ci-dessus, s'est révélé crucial, aussi bien pour tester des idées qui étaient restées à l'état d'hypothèses, que pour ré-orienter la théorie. Parmi les options théoriques que l'accès à une grande masse de données a renforcées en morphologie, il y a par exemple l'idée que les processus morphologiques sont orientés vers l'output, et que la forme finale d'un mot complexe est le fruit de l'interaction entre différents facteurs, qui rentrent partiellement en concurrence. Dans mon parcours je me suis moi-même servi, dans plusieurs travaux, de données extensives, principalement issues du Web. J'ai utilisé cette méthode en particulier pour étudier l'effacement des voyelles finales à la frontière base / affixe (Montermini 2003a ; 2003b ; 2006b), l'insularité anaphorique (Montermini 2006a, cf. aussi 2.7.6 ci-dessous), la flexion à l'intérieur de la composition (Montermini 2008a) ou l'allomorphie suffixale en italien (Lignon & Montermini 2008). J'ai aussi mené une réflexion, avec Nabil Hathout et Ludovic Tanguy, sur les aspects méthodologiques et épistémologiques de l'emploi du Web pour des recherches morphologiques (Hathout *et al.* 2008). Dans ce travail, nous avons discuté des problèmes techniques liés à l'utilisation de cette ressource pour des recherches en morphologie dérivationnelle, et nous avons montré comment la quantité de données auxquelles le Web permet d'avoir accès en un temps limité peut avoir une incidence y compris du point de vue théorique. Pour cela, nous nous sommes appuyés sur plusieurs travaux réalisés au sein de l'équipe de morphologie de l'ERSS (en particulier par M. Plénat et collaborateurs, cf. les références ci-dessus), notamment sur les suffixations en *-esque/-este, -able, -ment*.

Dans ce domaine, parmi les projets actuellement en cours, citons la construction d'un dictionnaire des affixes dérivationnels de l'italien, en collaboration avec le Département de linguistique de l'Université de Bologne, qui s'appuiera sur un corpus de référence de l'italien écrit (le CoRIS / CoDIS¹), et qui permettra d'effectuer des recherches croisées entre le dictionnaire et le corpus en question.

1.3 La classification des unités morphologiques

Il s'agit d'une question à laquelle j'ai été surtout confronté lors de la préparation de ma thèse de doctorat, soutenue en janvier 2002, et qui avait pour sujet *La préfixa-*

1 http://corpora.dslo.unibo.it/coris_ita.html.

tion en italien contemporain (une version remaniée a été publiée sous forme d'ouvrage dans Montermini 2008b). Dans ce travail, je me suis heurté à plusieurs problèmes relevant, d'une part, de la délimitation d'une classe de préfixes en italien et de l'identification des éléments qui en font partie, et de l'autre de l'articulation interne de cette classe, qui comprend des éléments qui présentent des comportements fortement hétérogènes. Plus précisément, je me suis intéressé à la distinction entre les préfixes en tant qu'affixes, et donc exposants de schémas morphologiques dérivationnels, et d'autres types d'unités linguistiques avec lesquelles ils présentent des ressemblances fortes, allant jusqu'à l'homophonie, par exemple avec des prépositions ou des éléments de composition non autonomes d'origine néoclassique.

Pour la classification, je me suis appuyé sur une trentaine de critères relevant des différents niveaux de l'analyse linguistique (phonologiques, morphologiques, syntaxiques, sémantiques). Cela m'a permis, non pas d'établir une liste précise des préfixes de l'italien, ni de tracer des frontières nettes entre ce qui est un préfixe et ce qui ne l'est pas, mais plutôt de délimiter un espace à l'intérieur duquel placer les unités réellement observées, et de les définir, par conséquent, sur la base de leur degré d'éloignement d'un idéal de canonicité. Dans les termes que j'avais employés dans ma thèse, les unités susceptibles d'être définies comme des préfixes de l'italien ont été classées selon leur degré de prototypicité affixale. Cette classification vise à donner une image réaliste de la situation observée en synchronie. L'hypothèse de base est que les affixes d'une langue, et en particulier ses préfixes, proviennent pour la plupart de la grammaticalisation d'unités linguistiques autonomes (dans le cas des préfixes essentiellement des prépositions et des adverbes). Si cela est vrai, il est normal qu'à un stade synchronique donné coexistent dans une langue des éléments qui se trouvent à des niveaux de grammaticalisation distincts, et qui présentent donc des caractéristiques différentes, certaines plus typiquement affixales, et certaines qui restent plus proches de celles des unités linguistiques autonomes. C'est exactement la situation que l'on observe avec un certain nombre de préfixes de l'italien (notamment les préfixes spatio-temporels et les préfixes évaluatifs), qui ont un comportement plus proche des unités linguistiques autonomes, comme par exemple la possibilité d'apparaître isolément en syntaxe, la possibilité de précéder des structures plus grandes qu'un mot (syntagmes ou phrases), ou la possibilité de violer le principe de l'insulari-

té anaphorique' (cf. ci-dessous). Ces considérations étaient restées largement implicites dans la première version (2002) de ma thèse, et j'ai essayé de les rendre plus explicites dans la nouvelle version de 2008 (cf. en particulier Montermini 2008b : 11).

Dans ma thèse, je me suis aussi intéressé à la distinction entre préfixes et suffixes et à la généralisation connue sous le nom de 'suffixing preference', qui remonte au moins à Greenberg (1963). Le fait que les suffixes soient les affixes les plus répandus dans les langues du monde, et souvent à l'intérieur de la même langue, lorsqu'elle comporte les deux types d'affixes, est aussi lié au concept de prototypicité : les suffixes peuvent être considérés comme des unités plus prototypiquement affixales que les préfixes. Cette plus grande prototypicité des suffixes provient, d'une part, de leur plus grande disponibilité, et d'autre part d'un principe cognitif général selon lequel le début d'un mot est plus saillant que sa fin pour sa reconnaissance et son interprétation (cf. Montermini 2008a : 49-59 pour des détails).

Parmi les critères qui ont été retenus pour la définition de la classe des préfixes en italien signalons notamment la possibilité, pour un préfixe, de modifier la catégorie des bases auxquelles il se lie, ou, dans d'autres termes, de déterminer la catégorie de ses dérivés. Ce critère, en réalité, ne s'est pas révélé déterminant dans la mesure où cette capacité est commune à des préfixes de différentes classes, et qui présentent des degrés de prototypicité différents. Aussi bien les préfixes de changement d'état que les préfixes de repérage spatio-temporel, par exemple, qui se trouvent aux deux pôles dans l'échelle de prototypicité affixale, possèdent cette propriété : cf. *barca*_N → *imbarcare*_V ('bateau' / 'embarquer') vs. *furto*_N → *antifurto*_A ('vol' / 'antivol').

Il s'agit d'une question qui a été longuement débattue dans les études de morphologie, et une opinion commune consiste à considérer que, au moins pour les langues indo-européennes, les préfixes seraient des éléments dépourvus de capacité catégorisatrice. Cette position est à la fois fautive et clairement incompatible avec le modèle de morphologie que je défends dans ce travail. Les règles, ou schémas, de construction de nouveaux lexèmes doivent être considérées comme partiellement disjointes de leurs exposants. Posséder un pouvoir catégorisateur n'est pas une propriété des affixes en tant qu'unités, mais des règles ; cela revient à dire que, dans le schéma qui explicite cette règle, les lexèmes de sortie ont une catégorie potentiellement différente des lexèmes d'entrée. Des exemples de procédés préfixaux qui s'accompagnent d'un

changement de la catégorie de la base s'observent dans plusieurs langues, y compris des langues indo-européennes. La croyance selon laquelle les préfixes seraient dépourvus de capacité catégorisatrice n'apporte rien à la compréhension de la compétence morphologique des locuteurs, et n'est pas non plus prédictive sur le plan typologique.

J'ai abordé cette question dans quelques autres travaux. Dans Montermini (2001) j'ai pris en considération le principe de la base unique (« Unitary Base Hypothesis », UBH, cf. Aronoff 1976), selon lequel les règles de construction de lexèmes sélectionnent une seule catégorie de bases. Par conséquent, des constructions qui violent cette généralisation ont souvent été considérées comme marginales ou déviantes. Or, justement l'élargissement du nombre de données observables a montré qu'elles sont loin de l'être, du moins du point de vue quantitatif. En particulier, j'ai considéré les suffixations en *-bile* et en *-oso* en italien, ainsi que quelques préfixes. En (1) je donne la relation catégorielle supposée que ces suffixes donneraient, et des exemples de mots qui la violent :

- (1) a. V+bile → A (ex. *tollerare* → *tollerabile* 'tolérer' / 'tolérable')
papa_N → papabile ('pape' / 'papable')
azzurro_A → azzurrabile ('bleu' / 'susceptible d'être convoqué dans la sélection nationale italienne')
- b. N+oso → A (ex. *noia* → *noioso* 'ennui' / 'ennuyeux')
scivolare_V → scivoloso ('glisser' / 'glissant')
morbido_A → morbidoso ('mou' / 'mou')

En m'appuyant sur des travaux qui proposent une vision de la morphologie constructionnelle orientée vers l'output (par exemple Plag 1997 ; Corbin 1999 ; 2004), j'ai proposé que les généralisations indiquées en (1) n'expriment que le prototype des constructions en question et que les relations qui s'établissent lors de la construction d'un mot complexe peuvent agir de façon partiellement indépendante ; la priorité donnée à une de ces relations (dans ce cas les relations sémantiques) peut entraîner le fait que les autres (dans ce cas les relations catégorielles) s'éloignent du

cas canonique. De même, certains mots ambigus quant à leur construction ont été réinterprétés sur une base sémantique. Par exemple, la base sémantiquement plus plausible pour l'adjectif *studioso* ('qui étudie beaucoup') est le verbe *studiare* ('étudier'), et non pas le nom *studio* ('étude'), qui serait la base préférée selon la UBH. Les violations de celle-ci sont donc probablement plus nombreuses que ce que les données peuvent laisser penser à première vue.

Dans Montermini (2005), j'ai effectué une étude spécifique sur la capacité catégorisatrice du préfixe *anti-* dans trois langues indo-européennes, anglais, italien et russe. Dans chacune de ces trois langues, le préfixe en question peut construire un adjectif à partir d'un nom, soit par simple juxtaposition (2a), soit en prenant comme base l'adjectif de relation construit sur le même nom (2b) :

- (2) a. febbre un farmaco antifebbre ('fièvre' / 'un médicament anti-fièvre')
b. febbre un farmaco antifebbrile ('fièvre' / 'un médicament anti-fiévreux')

La fréquence avec laquelle chacune de ces stratégies est employée dans une langue a été corrélée avec la présence ou pas d'une flexion explicite sur les adjectifs et avec la possibilité d'avoir des adjectifs invariables. J'ai montré que le type (2a) est largement le plus représenté en anglais, où les adjectifs ne portent pas de marques de flexion explicites, et presque absent en russe, où les adjectifs possèdent une flexion très riche et il n'existe pas d'adjectifs invariables. En italien, où les adjectifs sont généralement fléchis, mais où il existe des adjectifs invariables, les deux types de construction sont distribuées presque également – du moins à partir des données lexicographiques que j'ai recueillies : 56,3% pour le type (2a) et 43,7% pour le type (2b).

Dans d'autres travaux, j'ai développé des thèmes que j'avais abordés ou juste ébauchés dans ma thèse. Avec Nicola Grandi (Grandi & Montermini 2004 ; 2005) j'ai analysé, sur un corpus typologiquement varié, les différences entre préfixes et suffixes évaluatifs. Dans Montermini (2003c) et Amiot & Montermini (2009) j'ai considéré la distinction entre préfixes et prépositions, respectivement en italien et en français. Dans ces travaux j'ai considéré des exemples de mots construits exocentriques comme *il sottoscala* ('le dessous d'escalier') ou *l'avant-guerre*, que certains considèrent comme des composés préposition+nom (cf. Scalise 1983 : 146-147 ; Iacobini 1999 ;

2004). Partant de l'idée que ce sont principalement les catégories lexicales majeures (noms, verbes, adjectifs) qui rentrent dans la construction de mots complexes, j'ai considéré que les prépositions sont exclues de la composition. Sur la base de critères à la fois formels et sémantiques j'ai montré que les préfixes homophones de prépositions et qui ont un sens proche (comme *sotto-* ou *avant-*) s'en distinguent en effet, alors qu'ils rentrent dans un réseau de relations communes avec d'autres éléments qui sont clairement préfixaux (par exemple, en français *avant-*, *après-*, *pré-*, *post-*). Naturellement, aussi bien pour l'italien que pour le français, il existe des éléments qui se trouvent à des stades différents de grammaticalisation, et donc qui présentent des comportements plus ou moins proches des unités autonomes desquelles ils sont issus. Nous avons mis en évidence ce fait en particulier pour ce qui concerne le français dans Amiot & Montermini (2009).

Dans Montermini (2006a) je me suis intéressé à la violation, par des mots construits, du principe de l'insularité anaphorique². Selon ce principe, la base d'un mot morphologiquement complexe ne peut pas constituer l'antécédent d'une expression anaphorique, et des phrases comme celles en (3), pourtant attestées, seraient donc agrammaticales :

(3)

Mme Verdurin était assise sur un haut siège **suéd**ois en sapin ciré, qu'un violoniste de **ce pays** lui avait donné.²

L'Unione Sovietica e l'Inghilterra [...] porterebbero avanti un'azione anti-**Opec**, per indebolire **questa organizzazione**. [*la Repubblica*, 7 mai 1985]

'L'Union Soviétique et l'Angleterre mèneraient une action anti-Opec, pour affaiblir cette organisation'

Le principe de l'insularité anaphorique, proposé pour la première fois par Postal (1969), s'est révélé être un bon exemple de ces principes abstraits formulés par les linguistes pour conforter leurs théories, sans forcément se préoccuper de leur vraisemblance. Dans le travail en question j'ai adhéré à l'idée que ce que j'ai appelé l'ana-

2 M. Proust, *Du côté de chez Swann*.

phore intralexicale est principalement réglée par des facteurs pragmatiques (cf. Ward *et al.* 1991). Néanmoins, j'ai montré que certains principes purement grammaticaux, comme le type de construction impliqué, et notamment la proximité de l'affixe avec une unité linguistiquement autonome, peuvent également influencer l'apparition des anaphores de ce type.

1.4 La représentation des unités lexicales

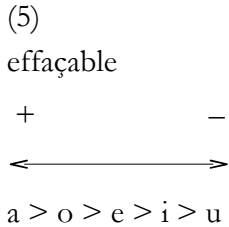
Dans un certain nombre d'autres travaux, je me suis intéressé à la forme sous laquelle les unités lexicales sont stockées, et en particulier à leur représentation phonologique. Je me suis penché, en particulier, sur les lexèmes de l'italien, une langue qui présente des propriétés particulières, même par rapport aux autres langues romanes. En italien, en effet, les mots de toutes les catégories portent typiquement une marque de flexion explicite. En d'autres termes, il n'existe pas de formes fléchies de mots non marquées, et donc monomorphémiques. Les noms et les adjectifs typiques comportent systématiquement une voyelle finale qui marque au moins le nombre (et pour les adjectifs aussi le genre), alors que les formes verbales présentent des terminaisons plus complexes. Pour les catégories nominales, la question de la nature de cette voyelle finale se pose. Dans la littérature sur la morphologie de l'italien on retrouve essentiellement deux positions : les chercheurs sont partagés entre ceux qui considèrent que la morphologie de l'italien est basée sur les morphèmes et ceux qui considèrent qu'elle est basée sur les mots. Selon les premiers (cf. Peperkamp 1995 ; Crocco Galèas 1998), ce sont des morphèmes lexicaux dépourvus de toute marque flexionnelle qui sont mémorisés dans le lexique ; selon les seconds, les mots sont mémorisés dans le lexique avec leur voyelle finale, éventuellement sous forme de thèmes³. Un des arguments en faveur de la première hypothèse est que la voyelle finale d'un nom ou d'un adjectif, correspondant à une marque flexionnelle, n'apparaît jamais lorsque le mot en question est dérivé par suffixation (cf. *tavolo* 'table' → *tavolino* 'petite table'). Il existe, cependant, un contre-argument très sérieux à cette interprétation, à savoir le fait que la voyelle finale d'un nom ou d'un adjectif n'apparaît jamais dans ses dérivés,

3 D'autres auteurs (Dressler & Thornton 1991) proposent des modèles mixtes, dans lesquels la morphologie de l'italien est basée sur les morphèmes dans certains cas et sur les thèmes dans d'autres.

même lorsqu'elle ne correspond clairement pas une marque de flexion, comme dans le cas des noms propres (4a), des emprunts invariables (4b), des adverbes (4c), etc.

- (4) a. Milano → milanese ('Milan' / 'milanais')
b. koala → koalino ('koala' / 'petit koala')
c. bene → benissimo ('bien' / 'très bien')

Pour traiter au moins ces cas, il faut imaginer qu'il existe une règle phonotactique qui efface la voyelle finale d'un mot avant un suffixe qui commence par voyelle. Dans le cadre d'un modèle de morphologie basé sur les mots, Scalise (1983 ; 1984) a proposé qu'une règle d'effacement de voyelle opère en italien en dérivation. Cette hypothèse a le mérite d'unifier le traitement des mots non analysables morphologiquement, comme ceux de (4), avec celui des mots dont la voyelle finale est une marque de flexion. En particulier, selon Scalise, un nom ou un adjectif est mémorisé dans le lexique sous sa forme de citation (respectivement le singulier et le masculin singulier). La voyelle finale de cette forme de citation est une voyelle thématique qui spécifie la classe de flexion à laquelle appartient le mot. Dans plusieurs de mes travaux (Montermini 2003a ; 2003b ; 2003d) j'ai proposé une évaluation à large échelle de l'hypothèse scalisienne de l'effacement de voyelle. J'ai notamment observé que le même phénomène s'observe, occasionnellement, aussi en préfixation (cf. *sopra+elevato* [soprele'vato] 'surélevé', *auto+ipnosi* [awtip'nɔ:zi] 'auto-hypnose'). Puisque l'effacement de voyelle est un phénomène phonologique, bien que conditionné morphologiquement, on s'attend à ce qu'il dépende aussi de facteurs phonologiques. Parmi les facteurs que j'ai identifiés comme les plus importants il y a la ressemblance entre les voyelles en contact et un principe sémiotique de reconnaissabilité de la base qui bloque l'effacement lorsque celle-ci serait compromise (cf. *Mao* → *maoista* 'maoïste'). Un autre facteur concerne le timbre de la voyelle candidate à l'effacement, avec des voyelles qui sont plus facilement effaçables que d'autres. J'ai ainsi proposé une échelle d'effaçabilité :



Cette échelle trouve probablement sa justification dans une convergence de facteurs : d'une part, elle est parallèle à la hiérarchie de sonorité pour les voyelles (cf. De Lacy 2006 : 286) ; d'autre part, les voyelles les plus facilement effaçables sont celles qui correspondent à des marques de flexion pour les noms et les adjectifs de l'italien.

Je me suis aussi intéressé à l'allomorphie affixale. Dans Montermini (2006b) j'ai étudié quelques affixes déverbaux de l'italien qui sont problématiques quant à leurs bases. Je les exemplifie ici avec le suffixe *-ione*, mais *-ivo*, *-tore* et *-tura* font aussi partie de cette catégorie :

(6) <u>verbe</u>	<u>participe passé</u>	<u>derivé</u>	
interrogare	interrogato	interrogazione	'interrogation'
estinguere	estinto	estinzione	'extinction'
decidere	deciso	decisione	'décision'
riscuotere	riscosso	riscossione	'recouvrement (des impôts)'
espellere	espulso	espulsione	'expulsion'

La seule séquence qui est commune à tous ces dérivés est [jone] ; en réalité, si on prend en considération aussi les deux phonèmes qui précèdent cette séquence, le nombre de possibilités est limité. Nous pouvons informellement le représenter comme [VCjone], où C ne peut être que l'affriquée [ts] (graphiquement *z*) ou la sibilante [s(:)/z], et V peut être n'importe quelle voyelle ou une sonorante ([r],[l] ou [n]). Il existe au total 20 formes différentes que le suffixe, ainsi défini, peut prendre. Comme on peut le constater dans la liste donnée en (6), la forme du dérivé est la plupart du temps liée à celle du participe passé du verbe⁴. La forme finale des dérivés est donnée par l'interaction entre la forme du participe passé et la forme du suffixe.

4 Il existe quelques exceptions comme *attribuito* / *attribuzione* ('attribué' / 'attribution').

J'ai proposé que la spécification phonologique de la règle dont le suffixe est l'exposant est représentée sous la forme d'une série de contraintes hiérarchisées :

$$(7) \quad a > i > e > o > u > \text{sonorante} \quad \left| \begin{array}{l} \text{ts} > \text{s:/z} \\ \text{jone} \end{array} \right.$$

La forme attestée qui correspond à la combinaison qui occupe la position la plus haute dans la hiérarchie ([at:sjone]) a la même forme que le suffixe déverbal pour la classe par défaut des verbes de l'italien (ceux à infinitif en *-are*). Dans une perspective purement morphémique, nous serions obligé, soit de lister jusqu'à vingt allomorphes différents pour le même suffixe, soit d'en donner une représentation maximale sous-spécifiée (par exemple [jone]), sans être en mesure de rendre compte du fait que les séquences qui peuvent précéder [jone] sont limitées en nombre et strictement liées au participe passé. J'ai proposé une analyse similaire pour les différents allomorphes qu'on observe pour le suffixe adjectival *-iano* en italien, entre autres dans un travail comparatif avec le français (Lignon & Montermini 2008), langue dans laquelle le suffixe *-ien* présente des problèmes semblables.

Plus récemment, je me suis penché sur la représentation lexicale des verbes, en italien et dans d'autres langues, et notamment à la structure des paradigmes verbaux. Ce travail a été réalisé principalement dans le cadre du projet ANR « MorPa » que j'ai coordonné de 2007 à 2010. Ces recherches se placent dans un cadre de morphologie basée sur les mots, et en particulier dans un modèle « mot et paradigmes ». Dans un certain nombre de travaux (cf. Boyé & Montermini 2007 ; Giraud *et al.* à p.) j'ai proposé une représentation de la structure paradigmatique des verbes de l'italien dans laquelle le fait d'avoir une représentation sous-jacente unique n'est pas une priorité, ni une propriété définitoire des verbes réguliers. La représentation que j'ai proposée, avec les collègues cités, est censée rendre compte de la complexité qu'un verbe de l'italien peut maximalelement atteindre et de la manière dont la morphologie verbale de l'italien est acquise et étendue à des données inconnues. Giraud *et al.* (à p.) relate aussi une série d'expériences psycholinguistiques que j'ai réalisées en collaboration avec d'autres collègues à l'Université de Bologne en mai 2008, dont le but était de tester quelques-unes des hypothèses théoriques formulées par rapport à la structuration du paradigme des verbes et de leur organisation en classes flexionnelles.

1.5 Processus cognitifs et grammaticalisation

Comme je l'ai rappelé plus haut, un certain nombre des premiers travaux que j'ai réalisés en linguistique ont touché à des phénomènes qui ne font pas partie du 'centre' de la langue, du moins du point de vue sociolinguistique. A plusieurs reprises j'ai été amené à m'interroger sur le statut de ces phénomènes et sur l'utilité de leur étude pour une compréhension des mécanismes linguistiques plus généraux. J'avais déjà l'intuition, au moment où j'ai réalisé ces recherches, que l'observation de ce genre de faits pouvait nous donner un aperçu du fonctionnement de la capacité linguistique. Au fil du temps cette intuition est devenue de plus en plus évidente à mes yeux, et je suis désormais convaincu qu'au moins certains de ces phénomènes peuvent être analysés comme le résultat du fonctionnement de principes cognitifs généraux. Cela les différencie des processus morphologiques canoniques, qui eux sont grammaticalisés. Dans Fradin *et al.* (2009) nous avons proposé quelques réflexions sur ce sujet, en prenant deux exemples de phénomènes que – à la suite d'autres – nous avons qualifié d'extragrammaticaux', justement parce que nous considérons qu'ils ne relèvent pas de la structuration grammaticale d'une langue spécifique, mais de principes plus généraux, les langages secrets (en particulier le verlan français) et les mots-valises. On trouve aussi dans ce groupe l'apocope, que j'ai étudiée dans mon mémoire de D.E.A. et dans d'autres travaux (Montermini 1999 ; 2002) et les hypocoristiques, que j'ai étudiés principalement en russe (Montermini 2007). Dans tous ces cas, la langue impose simplement une restriction très générale aux mots qui sont construits par les moyens en question, par exemple, le fait d'être plus courts que leurs bases. La forme finale est déterminée par l'action d'une série de principes universaux, que l'on peut représenter sous la forme de contraintes. Pour ce qui concerne les apocopes, et ce dans la plupart des langues où ce phénomène a été étudié, la forme de sortie correspond au format prosodique du mot minimal non marqué de la langue. Pour les hypocoristiques, à cette contrainte s'en ajoutent d'autres ; par exemple, le nombre et le type de syllabes qu'on rencontre dans les hypocoristiques sont normalement un sous-ensemble simplifié de ce que l'on observe normalement dans une langue déterminée (beaucoup d'hypocoristiques, dans plusieurs langues, ont la forme

CVCV).

Même si cela peut paraître surprenant, il me semble possible d'inclure dans cette thématique les travaux que j'ai réalisés sur la composition, dans une perspective typologiquement large (Montermini 2008a ; à p. ; Arcodia *et al.* 2009). Je suis de plus en plus convaincu, en effet, que les procédés de composition qui sont à l'œuvre dans les langues dérivent de l'encodage dans la grammaire d'un principe cognitif général qui consiste à relier deux mots du point de vue sémantique simplement en les juxtaposant. Cela est probablement à la base de l'idée courante selon laquelle la composition serait le moyen morphologique le plus répandu, y compris dans des langues isolantes, qui ne disposent d'aucune autre stratégie morphologique (cf. Greenberg 1963 : 92 ; Dressler 2006 : 23 ; Guevara & Scalise 2008). En revanche, les types de mots composés qu'on peut rencontrer dans une langue sont fortement contraints, aussi bien du point de vue sémantique que catégoriel et formel. Dans mes travaux consacrés à cette question (en particulier Montermini 2008a ; à p.) j'ai essayé d'intégrer l'étude de la composition dans un modèle de morphologie basée sur les lexèmes. J'ai ainsi abordé la question de la nature des éléments qui apparaissent dans des mots composés, aussi bien que des composés eux-mêmes.

1.6 Conclusion

Dans cette première partie j'ai proposé un bref tour d'horizon des activités de recherche que j'ai menées jusqu'à présent. J'espère avoir pu démontrer la cohérence d'une grande partie de mes travaux. Cette cohérence se manifeste à deux niveaux : la plupart des travaux que j'ai réalisés convergent vers une vision globale de la compétence morphologique des locuteurs et de son interaction avec le lexique ; par ce biais, la contribution que l'étude de la morphologie apporte à la compréhension de la capacité de langage plus globalement nous apparaît clairement.

J'ai eu la chance jusqu'ici de bénéficier d'un environnement de travail favorable, au niveau local, puisqu'à Toulouse sont regroupés la majeure partie des morphologues de France, mais pas seulement. La morphologie est une discipline ancienne, qui a été redécouverte à une époque relativement récente. De ce fait, la communauté européenne et mondiale des morphologues n'est pas aussi étoffée que celle des spécia-

listes d'autres disciplines comme la syntaxe ou la sémantique. Comme je l'ai rappelé plus haut, le début de ma carrière a pratiquement coïncidé avec la formation de regroupements de morphologues préoccupés par les mêmes problèmes et adoptant des hypothèses et des méthodologies similaires, aussi bien en France qu'en Italie. Faire partie de ces regroupements a été pour moi extrêmement utile et stimulant. Lorsque j'ai commencé mon activité de chercheur, je considérais le métier de linguiste comme une activité essentiellement individuelle. Je ne mesurais pas à quel point le concept d'«intelligence collective» pouvait être vrai. Tout seul, je n'aurais jamais pu élaborer ce qui apparaît dans les pages qui suivent, et sûrement pas sous cette forme. J'ai une grande reconnaissance à l'égard de chacune des personnes avec lesquelles j'ai été amené à travailler. Je leur dois le meilleur de ce que je présente dans le reste de ce dossier d'HDR, même si, bien entendu, ils ne sont nullement responsables de certaines interprétations et des erreurs éventuelles que contient ce travail.

Partie 2

Pour un modèle lexical de la morphologie

words have complex meanings precisely because neither words nor meanings are entirely linguistic objects, but rather the bastard offspring of language and the real or imagined world; it is this union of sparse linguistic resources with the vastness of the nonlinguistic universe that makes all words so rich from birth. Noncompositionality, on this account, follows from the nature of words, not from their nurture. [...] Some are more predictable than others, but complex words are never entirely the product of language. [M.Aronoff, *In the beginning was the word* (2007 : 817-818)]

2.1 La morphologie lexématique : aperçu

L'étude de la morphologie des langues a connu plusieurs changements paradoxaux dans son évolution. Si jusqu'au début du XX^e siècle elle a été fondamentalement réduite à l'étude de la conjugaison et de la déclinaison dans les langues classiques, le revirement qu'a constitué la vision structuraliste, et qui s'est poursuivie pratiquement inchangée à l'intérieur du courant générativiste, a eu des effets qui se font sentir en-

core aujourd'hui. Ceci n'est pas le lieu pour proposer une étude historique exhaustive des évolutions de la morphologie dans le siècle dernier. On peut tout de même proposer quelques observations pour situer le problème et pour introduire le modèle de morphologie qui sera discuté par la suite. Premièrement, il n'est pas exclu que les objets d'étude eux-mêmes aient influencé ce changement : les grammairiens et les spécialistes de linguistique historique se sont intéressés, jusqu'au XIX^e siècle, fondamentalement aux langues classiques, latin et grec, des langues qui présentent une morphologie extrêmement complexe et typiquement fusionnelle, pour laquelle un modèle de morphologie basé sur le mots est particulièrement adapté. La perte progressive de centralité que le mot a connue dans les études de linguistique à cheval entre les deux derniers siècles a ainsi été le fruit, probablement, d'une conspiration entre plusieurs facteurs différents. A partir de la moitié du XIX^e siècle le sanskrit devint un modèle pour les comparatistes, bien plus que le latin et le grec. Le structuralisme, de son côté, eut une grande impulsion qui lui vint en particulier des Etats-Unis. Aussi bien le sanskrit que l'anglais (que les langues amérindiennes, largement étudiées par les structuralistes à l'époque) sont des langues qui se laissent traiter aisément par une décomposition en sous-parties, dans compter que , et la plupart des structuralistes américains travaillaient sur des langues amérindiennes, qui ont aussi une structure fortement agglutinante. Tous ces facteurs sont à la base, me semble-t-il, du succès des modèles (dé)compositionnels en morphologie, au moins jusqu'au dernier tiers du XX^e siècle. Pendant une longue période une notion, pourtant intuitive, comme celle de mot a eu du mal à se faire reconnaître comme utile et opératoire en linguistique. Dans ses manifestations les plus extrêmes, les analyses de la morphologie structuraliste ne faisaient aucune place à la notion de mot ; les unités de base de l'analyse étaient des morphèmes, et les structures linguistiques n'étaient que des concaténations de morphèmes⁵. De ce point de vue, la morphologie et la syntaxe n'étaient pas différentes dans la substance, elles manipulaient le même type d'objets par les mêmes opérations. Y compris jusqu'à nos jours, une partie importante de travaux de morphologie a continué à se fonder sur des analyses qui ne manipulent que des unités plus petites que le mot par des principes similaires à ceux de la syntaxe (on peut citer, par

5 Pour un panorama de la vision structuraliste du mot cf. en particulier Dixon & Aikhenvald (2002 : 2-4). Parmi les textes fondamentaux de l'analyse morphologique structuraliste il faut citer au moins Bloomfield (1933) ; Nida (1949) ; Harris (1951) ; Hockett (1954) ; (1958).

exemple, les travaux de Halle 1973 ; Selkirk 1982 ; Di Sciullo & Williams 1987 ; Lieber 1992 et ceux qui se rattachent à la Distributed Morphology, au moins dans sa version originale, cf. Halle & Marantz 1993).

La notion de mot a recommencé à être populaire en morphologie à partir de la moitié des années 1960, d'abord portée par des linguistes européens qui, encore une fois, s'intéressaient aux langues classiques fortement flexionnelles et qui prônaient une morphologie « mots et paradigmes » (cf. Lyons 1963 : 11-12 ; Matthews 1972 : 96-101 ; Matthews 1991 (1974)). C'est dans les années 1970, avec les travaux de Jackendoff (1975) et surtout d'Aronoff (1976), qui s'inspirait directement des auteurs cités, et en particulier de Matthews, que les approches 'word-based' en morphologie ont pris un véritable essor. Sous l'influence plus ou moins directe d'Aronoff plusieurs travaux seront publiés dans les années suivantes en proposant des analyses morphologiques basées sur les mots dans un certain nombre de langues (cf. par exemple Booij 1977 ; Scalise 1983, 1984 ; Corbin 1987), en donnant naissance à une sorte de courant 'lexicaliste'⁶ dans la morphologie, surtout européenne, qui, après avoir subi plusieurs révisions et réajustements, est toujours vital aujourd'hui. Juste pour prendre un exemple, la partie la plus importante de la morphologie qui se fait aujourd'hui en France, et qui s'est développée dans les vingt dernières années, est une morphologie basée sur les mots qui prend son inspiration première dans le travail de D. Corbin, déjà citée, et du groupe de chercheurs qui s'est initialement formé autour d'elle. Parallèlement, les approches basées sur les mots ont continué de se développer dans des travaux de typologie et dans l'étude de la flexion des langues (cf. par exemple Carstairs 1987 ; Anderson 1992 ; Zwicky 1992 ; Aronoff 1994 ; Beard 1995 ; la Paradigm Function Morphology de Stump 2001 ; Blevins 2006)⁷. Il s'agit donc

6 Naturellement, 'lexicaliste' peut vouloir dire plusieurs choses. Dans ce contexte j'utilise cette étiquette de façon informelle, comme le font Scalise (1994 : 7) ou Scalise & Guevara (2005), pour me référer à une morphologie basée sur les mots et non sur les morphèmes. Stump (2001 : 1-30), par exemple, utilise 'lexicaliste' dans un sens différent pour indiquer les théories de la morphologie qui considèrent que les affixes sont listés dans le lexique au même titre que les mots pleins (cf. 2.4.1).

7 Cette brève liste n'est bien évidemment pas exhaustive. Il faudrait citer, par exemple, les approches fonctionnalistes, comme celui de la morphologie naturelle, qui propose un modèle 'mixte', basé sur les thèmes et les morphèmes (cf. Dressler 1988 ; 2005b). Sans oublier que le mot a continué à être une des unités fondamentales de la phonologie, y compris dans la phonologie générative (Chomsky & Halle 1968) et dans les modèles qui l'ont suivie, comme la phonologie autosegmentale (Goldsmith 1976) et la phonologie prosodique (Nespor & Vogel 1986).

d'un domaine bien établi dans les études linguistiques actuelles. Néanmoins, la plupart des problèmes que cette notion soulève, et qui avaient été soulignés par les chercheurs qui lui ont refusé un quelconque statut théorique, ne sont pas résolus. La section suivante est consacrée à discuter certains de ces problèmes et à montrer que les mots peuvent tout de même être considérés comme les unités de base de la morphologie, aussi bien flexionnelle que constructionnelle.

2.2 Une morphologie basée sur les mots

2.2.1 Définition du mot : problèmes empiriques

Comme observé plus haut, la notion de mot n'est pas sans poser un certain nombre de problèmes pour une approche scientifique à la langue. Ces problèmes ont souvent été mis en avant par les proposants d'une morphologie non basée sur les mots, qui en faisaient autant d'arguments décisifs pour nier tout statut théorique à cette notion.

Premièrement, puisque le mot 'mot' est d'abord un terme de la langue courante, il est de par sa nature polysémique, et ce dans la plupart des langues. Cette ambiguïté de sens fait souvent surface, y compris dans les travaux des linguistes. Je ne repropose pas ici des arguments que d'autres linguistes ont exprimé mieux que je ne pourrais le faire (cf. en particulier Matthews 1991 (1974) : chap. 2 ; Zwicky 1992 : 329-333), me limitant à indiquer la distinction fondamentale que les morphologues établissent par rapport à ce terme⁸ :

- (i) un segment (phonologique ou graphique) doté d'une certaine autonomie, par exemple parce qu'il est écrit entre deux blancs, et/ou parce qu'il peut constituer un énoncé à lui seul (il s'agit d'une définition traditionnelle du mot comme « *minimum free form* », qui remonte à Bloomfield 1933 : 178).

- (ii) une unité abstraite qui peut se présenter dans le discours sous plusieurs formes,

⁸ Pour les linguistes d'autres distinctions sont aussi pertinentes, par exemple celle entre mots phonologiques, morphologiques ou syntaxiques, qui ne coïncident pas forcément (cf. Dixon & Aikhenvald 2002 pour une discussion).

selon le contexte, par exemple syntaxique, et qui regroupe l'ensemble de ces formes.

Il faut remarquer que aussi bien le mot (i) que le mot (ii) sont des unités identifiées par les linguistes, mais aussi, de façon intuitive, par les locuteurs dans leur pratique métalinguistique. Lorsque nous allons chercher un mot dans un dictionnaire c'est à une unité du type (ii) que nous nous référons, alors que quand nous comptons les mots que comporte un texte nous nous intéressons à des unités du type (i). Le sens (i) du terme 'mot' n'est pas particulièrement utile pour une approche scientifique de la langue. Etre doté d'une certaine autonomie n'est pas en soi une preuve suffisante pour déterminer si une séquence linguistique est un mot ou pas : plusieurs séquences que nous aurions pourtant envie de considérer, pour leur sens, leurs propriétés distributionnelles, etc., comme des mots ne peuvent pas constituer un énoncé à elles seules (sinon de façon très invraisemblable ou dans la pratique métalinguistique). C'est le cas par exemple de *guise*, qui n'apparaît que dans les expressions *à sa guise* ou *en guise de*, ou de *contrecoeur*, qui n'apparaît que dans *à contrecoeur*. Inversement, rentrent dans la définition (i) des objets que nous n'aurions pas envie de considérer comme des mots à part entière, comme des prépositions (cf. ang. *Up!*) ou même des affixes (cf. l'emploi de *Re !* comme formule de salutation en français). Le mot au sens (ii), en revanche, correspond à un type d'unités qui ont été identifiées dans les grammaires depuis longtemps, et que plusieurs linguistes considèrent comme pertinent du point de vue théorique. Pour limiter les risques d'ambiguïté, pour le mot au sens (ii) a été proposé le nom 'lexème', une étiquette qui est désormais unanimement adoptée⁹. C'est l'étiquette qui sera adoptée dans la suite de ce travail.

Un deuxième problème, connexe, que pose la notion de mot pour l'analyse linguistique est que, en dépit du fait que les locuteurs ont souvent une idée intuitive de qu'est-ce qu'un mot de leur langue, il n'est pas aisé, pour les linguistes, de le définir. La plus citée des observations des anti-mots est probablement celle de Garvin (1954 : 345), selon qui

⁹ L'étiquette 'lexème' dans le sens indiqué a été proposée, à ma connaissance, pour la première fois par Lyons (1966 : 11-12) et a été rendue populaire par Matthews (1991 (1974)). Le terme lexème a été parfois utilisé avec des sens légèrement différents, par exemple par Martinet (1967) ou dans les études de sémiotique (cf. aussi Lyons 1966 : 11 n. 3).

in the present state of our techniques one may assume that we know how to isolate morphemes properly – that is unequivocally and without unaccountable residue. It is not certain that we know how to isolate words, and hence to isolate morphology from syntax.

En réalité, Garvin semble bien trop généreux avec l'analyse morphémique. Sur le plan empirique, l'analyse en morphèmes ne pose pas moins de problèmes que l'identification des lexèmes. Certes, elle est adéquate pour rendre compte d'un nombre limité de phénomènes dans un nombre limité de langues, en particulier dans celles agglutinantes. Mais les difficultés que posent beaucoup de phénomènes à l'idéal évoqué par ce linguiste sont bien documentées dans un nombre important de travaux : exponence multiple ou étendue, présence de morphèmes dépourvus de sens ou, à l'inverse, absence de morphèmes explicites en présence d'un changement de sens sont autant de phénomènes, dont il existe d'innombrables exemples dans les langues du monde, qui empêchent d'imaginer une méthodologie de décomposition en morphèmes – que j'étiquetterai informellement de 'morçeaologique'¹⁰ – universelle, propre et sans résidu. Au contraire, alors que du point de vue empirique une approche basée sur les lexèmes peut rendre compte de façon convaincante de ce genre de phénomènes, elle n'est pas plus inadéquate que l'approche morçeaologique pour traiter des langues agglutinantes, et apparaît donc comme préférable¹¹.

2.2.2 Définition du mot : problèmes théoriques

Même du point de vue épistémologique, il convient peut-être de s'interroger sur la pertinence des unités qu'on identifie. À ma connaissance, tous les linguistes acceptent le principe de la compositionnalité du langage, selon lequel le sens des structures complexes se construit à partir du sens des éléments qui les composent. Aronoff (2007 : 803) fournit une citation de Frege (1980 (1914)) qui résume de manière très claire le principe en question :

¹⁰ Cette étiquette a été originellement proposée par Gilles Boyé.

¹¹ Pour des traitements détaillés des problèmes posés par l'analyse en morphèmes cf. en particulier Matthews (1972 : ch. 6) ; Stump (2001 : chap. 1) ; Fradin (2003 : chap. II). Sur la sélection des données opérée par la morphologie morphématique cf. Blevins (2006 : 541).

The possibility of our understanding sentences which we have never heard before rests evidently on this, that we can construct the sense of a sentence out of parts that correspond to words.

Ce principe présuppose donc que l'on identifie, dans le flux linguistique, des « parties qui correspondent à des mots », des unités discrètes qui structurent la réalité telle que nous la percevons. Ces unités élémentaires sont des dénominations (dans le sens proposé par exemple par Kleiber 1984) : elles servent à désigner des objets, des propriétés ou des actions et sont, dans le cas le plus typique, des signes au sens saussurien, non motivés, qui sont mémorisés en tant que tels par les locuteurs. Leur ensemble constitue le lexique d'une langue. L'existence de tels éléments peut raisonnablement être considérée comme relevant de la capacité de langage, et donc comme une propriété de toutes les langues (cf. Pinker & Jackendoff 2005 : 212-215). Je considère que, dans leur version la plus prototypique, ces éléments correspondent aux lexèmes, tels qu'ils ont été caractérisés ci-dessus. Sur la réalité psychologique du mot (dans le sens de lexème), et en particulier sur sa valeur universelle il existe un certain débat (cf. Bauer 2000 : 253-254 ; Dixon & Aikhenvald 2002 pour un résumé). Les détracteurs de son universalité affirment souvent que cette notion serait induite par la séparation des mots graphiques dans la pratique de l'écriture dans les sociétés alphabétisées. Il s'agit, cependant, d'un argument circulaire : il est probable que la capacité de lire et écrire et les conventions d'écriture des langues aient pu renforcer la conscience d'une telle unité chez les locuteurs et contribué à en augmenter l'importance. Il est connu qu'en chinois, une langue qui n'a pas une écriture alphabétique, ce qui est communément considéré comme mot coïncide avec un caractère, même si ce 'mot' correspond rarement à une unité syntaxique (cf. Dixon & Aikhenvald 2002 : 32-33). Néanmoins, dans toutes les langues les locuteurs semblent avoir une conscience du fait que la chaîne sonore peut se décomposer en unités plus petites. Apparemment, même dans les langues non écrites, les locuteurs sont capables de prononcer les phrases lentement mot par mot ou syllabe par syllabe, mais – ce qui est particulièrement significatif – jamais morphème par morphème, lorsque ceux-ci ne coïncident pas avec les mots ou les syllabes. De plus, Dixon & Aikhenvald (2002 :

34), par exemple, citent un certain nombre de faits linguistiques qui semblent confirmer la valeur du mot comme unité structurante de la langue à un niveau universel, comme l'acquisition de la langue, les emprunts ou la tabouisation (cf. aussi Ramat 1990 : 4). Ces observations sont naturellement à prendre avec prudence, mais, à la différence du mot, le morphème ne semble jouer aucun rôle particulier dans la conscience métalinguistique des locuteurs.

2.2.3 Définition du mot : conclusion

Pour contourner les problèmes que pose sa définition, la plupart des linguistes qui tiennent à conférer au mot une valeur théorique proposent de considérer cette notion en termes de prototype (cf. Ramat 1990 ; Bauer 2000 : 255), ou bien de la moduler selon différents paramètres, par exemple la langue ou la composante de la langue considérées (cf. Dixon & Aikhenvald 2002, mais aussi Haspelmath 2009). Une attitude pragmatique qui consiste à reconnaître que le mot est une catégorie qui peut et doit être relativisée semble en effet être la plus raisonnable (cf. aussi Bertinetto 2009 : 131). Si les catégories théoriques inventés par les linguistes résistent difficilement à l'épreuve des données (et il est probable que le morphème ne résisterait pas mieux que les lexèmes à une investigation rigoureuse, cf. Blevins 2006), une attitude pragmatique qui consiste à renoncer à des définitions universelles pour se doter de catégories qui ont le mérite d'avoir une valeur intuitive certaine semble préférable. Haspelmath (2009) compare la notion de mot à celle de planète, pour laquelle il est autant difficile de trouver une définition univoque, mais les astronomes ne cessent pas d'en parler pour autant. Je propose donc de considérer que prototypiquement un objet qui (i) structure notre perception du monde en une unité discrète et est une dénomination, (ii) a les propriétés syntaxiques et distributionnelles d'un nom, d'un verbe ou d'un adjectif¹², (iii) fait ou peut faire partie du lexique est un mot, dans le sens de

12 Naturellement, ce paramètre en particulier est relatif et dépend de façon cruciale du type de langue, puisque le comportement et la distribution des unités que dans les langues indo-européennes nous identifions comme noms, verbes et adjectifs peut varier largement. Cela n'empêche que dans une même langue des unités du même type aient la même distribution, si on prend 'distribution' dans un sens plus large qu'uniquement syntaxique. Par exemple, si on considère que l'incorporation est, au moins dans certains cas, un procédé morphologique de composition (cf. Mithun 2010), on peut attribuer le statut de lexèmes aux noms qui peuvent être incorporés, en considérant que cela relève de leurs propriétés distributionnelles.

lexème donné ci-dessus et qui sera précisé par la suite.

2.3 Caractérisation du lexème

2.3.1 Représentation lexicale des lexèmes

La plupart des travaux en morphologie qui assument le lexème comme unité de base concordent à le caractériser comme un signe linguistique au sens saussurien qui unit de manière arbitraire au moins trois types d'informations : une caractérisation formelle (phonologique, éventuellement graphique), des propriétés sémantiques et des propriétés morphosyntaxiques (qui spécifient, par exemple, la catégorie lexicale, la classe de flexion éventuelle, la valence, etc.). En (1) je donne une représentation des lexèmes *TABLE*, *PUR* et *LAVER*¹³ du français, à partir de la formalisation proposée par Fradin (2003 : 105)¹⁴ :

(1)	(G)	<i>table</i>	<i>pur</i>	<i>laver</i>
	(F)	tabl	pyr	lav
	(SX)	cat:N	cat:A	cat:V, str-arg<SN, SN>
	(M)	ger:f		conj:1
	(S)	'...'	'...'	'...'

(1) correspond au format dans lequel on peut considérer que les lexèmes sont stockés dans la mémoire des locuteurs à l'intérieur du lexique. Les rubriques identifiées par Fradin sont les suivantes : (G) graphémique ; (F) phonologique ; (SX) syntaxique, où sont stockées des informations concernant l'emploi du lexème par la syntaxe, comme la catégorie ou la structure argumentale ; (M) morphologique, où sont stockées des informations idiosyncratiques, comme le genre pour les noms en français ou la classe de flexion ; (S) sémantique. La dernière rubrique est laissée délibérément sous-spécifiée ici. Dans ce qui suit, seules les rubriques pertinentes pour la discussion seront indiquées. En particulier, la rubrique (G) ne sera pas prise en considération da-

¹³ Je précise, s'il le faut, que, selon l'usage courant, j'indique les lexèmes en petites majuscules.

¹⁴ Sur la caractérisation du lexème cf. en particulier Lyons (1977 : 512-520) ; Beard (1995 : 44-50) ; Fradin (2003 : 80-106).

vantage.

Naturellement, les rubriques sont liées entre elles de façon arbitraire seulement dans les cas limites. Il existe plusieurs cas dans lesquels les locuteurs peuvent inférer le contenu d'une rubrique du contenu d'une autre, ou dans lesquels les contenus de deux rubriques sont strictement liés. Par exemple, il est évident que la rubrique (M) contient des informations sensiblement différentes selon la catégorie lexicale qui est indiquée dans la rubrique (SX) (du moins dans les langues, comme le français, dans lesquelles les catégories lexicales sont bien distinctes). De même, on sait que les catégories lexicales prototypiquement expriment des types de significations et remplissent des fonctions sémantiques différentes (cf. Croft 1991, 2000 ; Baker 2003), ce qui, par exemple, rend dans certains cas le contenu de la rubrique (SX) inférable de la rubrique (S). De même, la rubrique (F) et la rubrique (SX) peuvent être étroitement liées. Par exemple, l'interrogation d'un corpus du français¹⁵ fournit 537 adjectifs qui se terminent en *-able*, mais seulement 13 noms qui ne soient pas analysables aussi comme adjectifs (par ex. *sable*) ; pour *-ible* la proportion est de 115 adjectifs et 3 noms inanalysables (par ex. *cible*). Un mot se terminant par *-able* ou *-ible*¹⁶ a donc 40 fois plus de possibilités d'être un adjectif qu'un nom, une proportion qui doit jouer dans la possibilité pour un locuteur d'inférer la catégorie à partir de la forme phonologique lorsqu'ils entendent un mot inconnu et non analysable pour la première fois. Enfin, il est clair que la relation entre (F) et (S) peut être plus ou moins transparente, et elle est transparente, notamment, dans le cas des lexèmes construits de manière productive.

2.3.2 Variation des lexèmes : premier aperçu

En (1) je n'ai présenté que des lexèmes qui constituent des cas simples. Prenons par exemple l'adjectif *PUR*, qui est invariable (en français de référence) : dans toutes les formes de son paradigme et dans tous les dérivés qu'il construit (par exemple *pureté*, *purement*, *purifier*) il est toujours [pyr]. En français le cas par défaut est qu'un nom ou un adjectif soient phonologiquement invariables en flexion. Dans d'autres

15 Où cela n'est pas spécifié autrement, les interrogations de corpus pour les français ont été effectuées sur Lexique.3 (<http://www.lexique.org/>).

16 La proportion de lexèmes se terminant par *ble* précédé d'une voyelle différente de [a] ou [i] est insignifiante (respectivement 7 noms et 8 adjectifs).

langues, comme en italien, les noms et les adjectifs varient selon la case du paradigme qui est exprimée. Ainsi, l'adjectif correspondant à PUR en italien possède les formes [ˈpuro], [ˈpuri], [ˈpura], [ˈpure] respectivement au masculin singulier, masculin pluriel, féminin singulier et féminin pluriel. Si l'entrée lexicale de PUR, exemplifiée en (1), ne comporte qu'une forme, cependant, il n'est pas nécessaire que celle de PURO en comporte quatre. Puisque l'alternance qu'il met en jeu se retrouve dans un nombre important d'adjectifs en italien, elle peut être exprimée par moyen d'un schéma qui s'applique à tous les adjectifs qui se comportent de la même manière. Les mêmes observations valent pour les verbes du français : les cinquante et quelques formes que comporte le paradigme du verbe LAVER ne doivent pas obligatoirement être incluses dans sa représentation lexicale¹⁷, dans la mesure où il met en jeu des alternances qui sont également mises en jeu par de nombreux autres verbes. Les adjectifs comme PUR représentent le cas le plus simple, et ceux comme PURO ou les verbes comme LAVER sont un cas légèrement moins simple, mais commun dans toutes les langues (du moins celles qui possèdent une flexion). Le cas opposé est celui des lexèmes pour lesquels la rubrique (F) doit nécessairement spécifier toutes les formes, parce qu'elles ne peuvent pas être reconstruites les unes à partir des autres. Pour le nom ŒIL en français, par exemple, les deux formes du paradigme ([œj] pour le singulier et [jø] pour le pluriel¹⁸) doivent certainement être spécifiées dans l'entrée lexicale, puisqu'il s'agit d'une alternance qui n'est mise en jeu que par ce lexème (cf. 2.8.2). Entre ces deux extrêmes, il existe une série de cas pour lesquels la décision est plus difficile. Par exemple, l'adjectif BON possède, en flexion, deux formes, [bɔ̃] et [bɔn]. Selon le critère qu'on a utilisé pour PURO, on pourrait avoir envie d'explicitier une règle pour les adjectifs de ce type, puisque l'alternance [ɔ̃] ~ [ɔn] se retrouve dans un certain nombre d'autres adjectifs du français (par exemple BRETON, MIGNON, etc.). Mais pour d'autres adjectifs la séquence [ɔ̃] alterne avec d'autres séquences, par exemple [ɔ̃d] dans BLOND, ROND, [ɔ̃g] dans LONG, ou n'alterne pas, comme dans MARRON. Une grande partie de ce travail sera consacrée à ce genre de questions. L'enjeu n'est pas uniquement celui de rendre compte du lexique existant et stabilisé, pour autant que l'on puisse le définir

17 Le fait que les formes fléchies de laver ne *doivent* pas nécessairement être incluses dans sa représentation lexicale ne signifie pas qu'elle ne puissent pas l'être. Comme on verra par la suite (2.7.1), il n'est pas nécessaire, ni probablement réaliste, de penser que le lexique ne contient que l'information non redondante.

18 Ainsi que, peut-être, [zjø] pour le dérivé *zyeuter*.

avec précision, tel que l'exemplifient les lexèmes cités, mais aussi, et peut-être surtout, pour les cas dans lesquels il 'se construit' ; lorsque, par exemple, des restructurations ont lieu ou de nouveaux lexèmes sont créés. Les exemples de (2) montrent que l'alternance [ʃ] ~ [ɔn] est active dans la compétence morphologique des locuteurs, y compris pour des adjectifs généralement considérés comme invariables (par exemple *souillonne*, dont le TLFi ne donne qu'un exemple qu'il indique comme « hapax », mais dont on retrouve de très nombreux exemples sur le Web) ou des adjectif dérivés de noms (en l'occurrence ici du nom propre d'Eric Besson) :

- (2) Celles où ça traîne sont des **souillonnes** qui se fichent que leurs enfants attrapent des microbes.¹⁹

Cette brusque générosité de Besson a fait trembler les cœurs, jusqu'à Port-au-Prince. [...] Petite précision : la clémence **bessonne** ne concerne que les Haïtiens.²⁰

Plus généralement, ce travail s'occupe de la rubrique (F) des lexèmes, tels qu'ils ont été caractérisés en (1), de la manière dont elle est remplie et de sa relation avec les autres rubriques, que ce soit pour les lexèmes présents de façon stable dans le lexique ou pour les lexèmes nouvellement construits, qui présentent, comme on l'a dit, des problèmes spécifiques et qui permettent, c'est l'une des hypothèses que j'avance, d'observer de manière plus directe la compétence morphologique en action. Puisque dans les deux dernières sections je me suis attaché à montrer les avantages, à la fois empiriques et théoriques, que la morphologie peut tirer en adoptant le lexème et non pas le morphème comme unité de base et à éliminer, par conséquent, les morphèmes lexicaux de l'inventaire de ses unités, celle qui suit sera consacrée à l'autre type de morphèmes identifiés par les tenants des approches morceaologiques, à savoir les morphèmes grammaticaux.

¹⁹ <http://www.momentdefolie.com/?p=328>.

²⁰ <http://achillengoma.over-blog.com/article-eric-besson-offre-a-la-france-sa-coupe-du-monde-de-l-humour-politique-43470579.html>.

2.4 Que reste-t-il des morphèmes ?

2.4.1 Problèmes posés par la morphologie incrémentale

Stump (2001 : ch. 1) propose une classification des approches de la morphologie fondée sur deux paramètres croisés : il distingue celles qu'il appelle les approches « lexicales » des approches « inférentielles » et les approches « incrémentales » des approches « réalisationnelles ». Les deux premières dimensions distinguent les théories selon lesquelles les éléments qui servent à construire, par exemple, les formes fléchies d'un lexème sont listés dans le lexique, au même titre que les unités lexicales, de celles qui considèrent les relations entre les différentes formes fléchies d'un lexème en termes de règles ou de fonctions. Les deux autres distinguent les théories qui considèrent qu'en morphologie, typiquement, la quantité d'information véhiculée par une forme augmente en concomitance avec l'adjonction de matériel segmental de celles pour lesquelles ce sont les propriétés morphosyntaxiques d'un mot qui demandent que celui-ci ait une certaine forme. Celle que j'ai défini l'approche morcealogique typique est donc une théorie qui est à la fois lexicale (selon la terminologie de Stump) et incrémentale. Si on reprend l'exemple de Stump et on l'adapte au français, dans une forme verbale comme *mangeons*, pour la morphologie lexicale-incrémentale la partie [mãʒ] n'apporte qu'une information de type lexical, tandis que la partie [ɔ̃] apporte une information de type morphosyntaxique, à savoir [1^e personne] et [pluriel]. Parallèlement, dans la forme *mangerons*, une troisième partie, [əʁ] intercalée entre les deux premières, apporte l'information, toujours de type morphosyntaxique, [futur]. Comme je l'ai observé en 2.2.1, une approche de ce genre est efficace uniquement pour rendre compte d'un certain nombre de faits dans un nombre limité de langues, et notamment des phénomènes morphologiques agglutinants. Prenons un exemple classique comme celui du turc :

- (3) turc
- | | |
|----------|-----------------|
| ev | 'maison' |
| evler | 'maisons' |
| evlerim | 'mes maisons' |
| evlerime | 'à mes maisons' |

Des données comme celles de (3) s'expliquent très bien dans une perspective additive (ou incrémentale, selon la terminologie de Stump) : chaque portion de matériel phonologique qui est rajoutée apporte une portion de sens supplémentaire. Ainsi, on peut identifier un morphème *-ler* qui 'signifie' [pluriel], un morphème *-im* qui 'signifie' [possesseur de 1^e personne singulier], etc. Cependant, il est bien connu que des données comme celles de (3) constituent plutôt l'exception que la norme dans les langues du monde. Plus fréquents sont les cas dans lesquels il n'est pas possible d'établir une correspondance directe entre un segment phonologique et une valeur morphosyntaxique sans restes, pour réutiliser la métaphore de Garvin. En revanche, même pour ces données on peut envisager une analyse différente de celle ébauchée, qui a le mérite de pouvoir s'appliquer également à d'autres types de phénomènes non agglutinants. Si, en effet, le singulier est plus souvent la forme non marquée par rapport au pluriel (ou aux autres nombres) (cf. Greenberg 1963 ; Corbett 2000 : 151), il existe aussi dans les langues de nombreux cas dans lesquels le singulier et le pluriel sont marqués (4b), aucun de deux n'est marqué (4c), et même dans lesquels le pluriel ne porte aucune marque, contrairement au singulier (4d) :

(4)		singulier	pluriel	
a.	turc	ev	evler	'maison'
b.	italien	casa	case	'maison'
c.	français	maison [mezɔ̃]	maisons [mezɔ̃]	
d.	murle	nyoon	nyoo	'agneau'

Naturellement, tous ces cas peuvent être pris en compte dans une approche incrémentale, par exemple en ayant recours à des morphèmes vides dans les cas non marqués, mais justement le recours à des expédients de ce genre que rien ne justifie sur le plan empirique est emblématique des difficultés que des cas de ce genre posent aux modèles incrémentaux. Les exemples de (4) montrent que dans la relation qui lie une forme du singulier à une forme du pluriel l'ajout de matériel segmental (ou la suppression) n'est qu'une des possibilités mises en œuvre par les langues, d'autres stratégies pouvant être la substitution (italien) ou l'identité (français).

2.4.2 Une morphologie inférentielle

Nous pouvons avoir un point de vue différent sur les exemples de (4) et prendre en considération non pas les formes individuellement, mais les lexèmes entiers dont elles relèvent. De ce point de vue, dans les langues exemplifiées, un nom est une unité qui est faite pour avoir (au moins) deux nombres, singulier et pluriel ; ces deux nombres sont réalisés par des formes qui, selon la langue, sont reliés par des fonctions différentes, que nous pouvons symboliser comme il suit :

(5)	sg.	pl.
a. turc	X	— Xler
b. italien	Xa	— Xe
c. français	X	— X
i. russe	Xa	— X

Un point de vue comme celui-ci, à la différence de l'approche incrémentale, est non orienté : ce qui compte n'est pas le fait de 'rajouter' quelque chose du point de vue phonologique, mais la fonction qui exprime la relation entre deux formes. Cette vision a des conséquences théoriques intéressantes, et principalement le fait qu'un lexème ne possède pas une forme de base unique, à partir de laquelle toutes les autres peuvent être construites (cf. 2.8.5). Nous reviendrons sur cette vision du lexème plus bas. Ici je remarquerai simplement que cette non orientation possède également une certaine plausibilité aussi bien sur le plan cognitif que sur le plan acquisitionnel. Plus haut on a rappelé que le singulier est souvent la forme morphologiquement non marquée ; on peut effectivement faire l'hypothèse que cela correspond au fait qu'il s'agit aussi de la forme la moins marquée sémantiquement (il s'agit, par exemple, d'un des principes souvent évoqués par les tenants de la morphologie naturelle, cf. Dressler 1985b : 328 ; 2005a ; Dressler *et al.* 1987). Cependant, le fait d'être non marqué morphologiquement n'est pas forcément un indice du fait qu'une forme est plus simple. Il serait par exemple difficile de soutenir que la 3^e personne du singulier des verbes en anglais, qui comporte un suffixe *-s*, est sémantiquement marquée par rapport à toutes

les autres qui ne comportent pas de suffixe. Sur le plan empirique, donc, une approche basée sur les lexèmes n'est pas plus inadéquate qu'une approche incrémentale pour traiter des phénomènes morphologiques de type agglutinant, comme ceux du turc en (3) (malgré ce que semble penser Jackendoff 2002 : 161-162). Là où elles ne s'équivalent pas c'est pour traiter des données telles que celles de (4b-d), dont il est plus difficile de rendre compte dans une approche de type incrémental, sinon avec une série d'artifices théoriques. Tenons compte, par ailleurs, du fait que les exemples de (4) représentent des cas qui sont encore 'simples' par rapport à la diversité de variations morphologiques qu'on retrouve dans les langues du monde, puisqu'il s'agit, dans tous les cas de variations segmentales et concaténatives, alors que l'existence de phénomènes non concaténatifs et même non segmentaux est bien documentée, comme dans le cas de l'apophonie (6a), de la reduplication (6b), de la morphologie gabaritique (6c), etc. :

(6)

a. allemand	Garten	Gärten	'jardin / jardins'
b. washo ²¹	sukuʔ	sukukuʔ	'chien / chiens'
c. arabe	kitāb	kutub	'livre' / 'livres'

Naturellement, les approches de type morphématique incrémental ont essayé de traiter ces phénomènes, la plupart des fois en proposant des morphèmes spéciaux, comme des « replacive morphemes », dans le cas de l'apophonie (cf. Nida 1949), ou comme dans la « morphemic tier hypothesis » de McCarthy (1982) pour expliquer les procédés de la morphologie gabaritique des langues sémitiques.

Le principal problème de toutes ces approches est qu'elles ne renoncent pas à considérer le morphème comme un objet, c'est-à-dire comme un signe qui, au moins dans les cas typiques, se doit d'avoir un corrélat segmental. La version extrême de cette vision est celle qui considère que les morphèmes, par exemple les affixes, sont des unités lexicales listées dans le lexique au même titre que les lexèmes (ou les morphèmes lexicaux), qui s'exprime dans une certaine tradition de la morphologie générative, celle que Stump (2001 : 1-2) étiquette d'approche lexicale (cf. Halle 1973 ; Lie-

21 Yu (2005).

ber 1981, 1992 ; Selkirk 1982 ; Di Sciullo & Williams 1987 ; Jackendoff 1997 ; 2002). Selon Lieber (1981 : 36-38 ; 1992), par exemple, les thèmes lexicaux et les affixes ont des entrées lexicales qui sont identiques, à la différence près que les derniers, mais pas les premiers, sont sous-catégorisés selon la catégorie lexicale des autres morphèmes auxquels ils peuvent s'attacher. Les approches de type inférentiel, au contraire, (cf. Zwicky 1985 : 25 ; 1988 ; 1992 ; Anderson 1992 ; Beard 1995 ; Stump 2001) considèrent les morphèmes non pas comme des objets, mais comme des opérations (cf. aussi Aronoff 1976 : 65-66, 78). A mes yeux, une des meilleures définitions du 'morphème' dans ce sens est celle qui est donnée par Janda (1983 : 79) :

the only generally-valid (certainly the only cross-linguistically valid) notion of the "morpheme" is that of a rule, a formal operation which may apply to a stem on the basis of its associated morphosyntactic features, rather than that of a thing, or unit-representation.

Dans une vision comme celle-ci, les affixes ne sont plus vus comme des unités, mais comme des exposants de règles ou des fonctions morphologiques, et ils ne sont plus différents dans la substance des autres types d'opérations non affixales, exemplifiés en (6). Du point de vue empirique il s'agit d'un effet désirable, puisque il est bien connu que, par exemple en flexion, les mêmes valeurs morphosyntaxiques peuvent être exprimées aussi bien par des opérations affixales que par des opérations non affixales, y compris dans la même langue et y compris pour les mêmes valeurs morphosyntaxiques dans la même langue :

(7)	sg.	pl.	
allemand	Hund	Hunde	'chien' / 'chiens'
	Garten	Gärten	'jardin' / 'jardins'
	Kran	Kräne	'grue' / 'grues'

un fait qui est exprimé par Stump (2001 : 9) dans le principe selon lequel « [t]here is no theoretically significant difference between concatenative and nonconcatenative inflection ». Celui-ci est un des principes-clés des approches inférentielles de la morphologie. Comme le résume bien Zwicky (1992 : 346), alors que la morphologie

de type incrémental, celle que j'ai appelée morceaulogique, essaie d'assimiler les opérations non concaténatives et non segmentales, comme celles exemplifiées en (7), à l'affixation, la morphologie inférentielle traite l'affixation comme une opération parmi les autres.

2.4.3 Conséquences de la dérivation du morphème

Celle que j'appellerai la 'dérivation' du morphème a d'autres conséquences intéressantes. En premier lieu, elle nous permet de voir l'allomorphie affixale d'une manière beaucoup moins dramatique. En effet, si les morphèmes perdent leur substance segmentale et se caractérisent comme des opérations phonologiques (au sens large d'opérations concaténatives, segmentales ou prosodiques), ils peuvent avoir des réalisations de surface différentes selon la base sur laquelle ils opèrent. Ainsi, par exemple, pour les lexèmes français se terminant en *-ien* ou *-éen* (deux séquences pour lesquelles il est possible de démontrer qu'elles sont les exposants de la même règle morphologique), on peut proposer un traitement qui ne doit pas avoir recours à des règles phonologiques douteuses, puisqu'il arrive qu'à une même séquence dans la base correspondent des dérivés où le suffixe a une forme différente (la plupart des données proviennent de Lignon 2000) :

- (8) Zola → zolien Ghana → ghanéen Kafka → kafkaïen
Mussolini → mussolinien Nancy → nancéen
Téchiné → téchinien Pyrénées → pyrénéen Juppé → juppéen

La section 3.3 sera consacrée au(x) suffixe(s) correspondant(s) en italien, et on y proposera un traitement de ces opérations morphologiques en termes de contraintes. Une autre tâche dont la dérivation des affixes nous libère est l'obligation de définir avec précision la frontière entre une base et un affixe, un effort qui est souvent destiné à être frustré, puisque dans beaucoup de cas il n'y a pas de réponse univoque possible. Pour reprendre les exemples ci-dessus, le [e] qui apparaît dans *pyrénéen* appartient-il à la base ou au suffixe ? Dans une perspective morphématique qui voudrait tracer des frontières nettes, on ne s'en sort pas : si on l'attribue à la base on ne s'ex-

plique pas pourquoi on retrouve la forme [eẽ] avec des lexèmes qui ne comportent pas de [e] final, comme dans *nancéen* ; si on l'attribue au suffixe il faut encore expliquer la prépondérance du type *Pyrénées* → *pyrénéen* pour les bases se terminant en [e] et du type *Mussolini* → *mussolinien* pour les bases se terminant en [i]²².

La dérèification des morphèmes permet aussi de jeter un nouveau regard sur les procédés de grammaticalisation, plus ou moins innovateurs, qui sosu-tendent la création de nouveaux schémas en morphologie dérivationnelle, par exemple la naissance de nouveaux affixes qui a souvent lieu à partir de la segmentation artificielle et de la remotivation des parties d'un lexème originaire, comme dans le célèbre exemple du suffixe *-gate* qui dérive du prototype (anglais) *Watergate* et qui s'est répandu dans plusieurs langues européennes. Dans le cas de procédés qui ne sont pas encore totalement stabilisés, le lien avec le prototype peut être marqué au niveau formel de différentes manières, et non seulement par la simple juxtaposition d'un affixe et d'une base, comme dans les exemples, cités par Fradin (2000 : 39-40), *Castroïka* (← *Castro*) ou *Béréstroïka* (← *Béré* <*govoy*>). Dans ces cas, on observe une dynamique dans laquelle rentrent trois éléments : la base (par exemple *Castro*), le prototype (*pérestroïka*) et l'affixe présumé (*-stroïka*, dans le sens de 'renouvellement politique et social lié à'). L'output qu'on observe est donc le résultat optimal de la combinaison des trois. Si l'affixe a du succès (cela n'a visiblement pas été le cas pour *-stroïka*, pour des raisons qui n'ont clairement rien de linguistique), le lien avec le prototype est destiné à s'estomper et à devenir de plus en plus ténu, et la construction à fonctionner plutôt comme une suffixation canonique. Naturellement, les phénomènes de ce type – qui sont peut-être sociolinguistiquement marqués, mais néanmoins sont loin d'être marginaux dans le fonctionnement actuel de la langue – restent largement inexpliqués dans une perspective morceaologique.

Plus globalement, en dérèifiant les morphèmes, la morphologie inférentielle rend possible une approche à la morphologie « output-based », dans laquelle les formes réellement observées sont conçues comme le résultat optimal de l'application de l'opération à une base sous l'influence de contraintes diverses. Cette approche fondée sur les contraintes, inspirée de la Théorie de l'Optimalité, a déjà été appliquée à la morphologie flexionnelle et constructionnelle dans bon nombre de travaux, en parti-

22 Respectivement 55,35% et 73,84% selon les calculs effectués par Lignon (2000 : 175) sur son corpus.

culier issus de l'approche élaborée dans notre laboratoire, CLLE-ERSS (cf. Plénat & Roché 2003 ; Plénat 2009b ; Roché 2003 ; 2007 ; 2008 ; 2009 ; Hathout 2009 ; Lignon & Plénat 2009 ; Boyé à paraître). Elle sera illustrée dans le détail dans la section 2.6.3.

2.4.4 Statut des unités morphologiques : racines, thèmes, affixes

Un autre effet désirable de la dérègulation des morphèmes est qu'elle nous permet de nous libérer d'une panoplie d'objets morphologiques, racines, thèmes, affixes, etc., pour lesquels quasiment chaque morphologue a sa propre définition et qui, d'un travail à l'autre, correspondent à des objets différents, dont l'existence est plus ou moins plausible. Tout ce dont le morphologue a besoin dans l'approche développée ici sont des formes de lexèmes et des opérations (une option adoptée aussi par Blevins 2006 : 536). Celle-ci est une option radicale, mais à mon sens salutaire. Je maintiens que toutes les étiquettes en question renvoient à une pratique qui aime la segmentation et non pas la mise en relation de différentes formes, qui est la base de la vision de la morphologie que je défends ici. A mon sens, la segmentation, l'identification et l'étiquetage de sous-parties dans les formes des lexèmes sont des opérations qui peuvent avoir une certaine utilité pratique et descriptive dans certains cas, mais les éléments identifiés ne doivent pas être confondus avec des objets qui ont une valeur théorique précise (cf. aussi la section 2.9.2 dans la conclusion de cette partie). De cela découle un autre fait important, à savoir qu'entre les formes abstraites, telles que les thèmes, et les formes fléchies concrètes il n'y a plus de différence de substance. Cela est compatible avec une observation récurrente en morphologie (cf. Bonami & Boyé 2003 ; Boyé 2006 : 297), que dans un lexème la supplétion peut concerner aussi bien un thème (c'est-à-dire, typiquement plusieurs cases du paradigme) qu'une forme fléchie individuelle, comme dans les verbes très fréquents du français ÊTRE ou FAIRE qui font respectivement *êtes* et *faites* à la 2^e personne du pluriel du présent de l'indicatif. On verra plus loin qu'une explication possible de ces supplétions est de considérer qu'une forme remplit une portion du paradigme (une case ou une série de cases) de façon idiosyncrasique, en conséquence de quoi le lien entre deux formes est brisé.

Par commodité je vais appeler les formes de lexèmes sur lesquelles opère la morphologie tantôt 'bases' tantôt 'thèmes', sans que ces étiquettes n'aient un statut théo-

rique précis ; il s'agit simplement d'étiquettes descriptives qui renvoient à des objets linguistiques abstraits qui peuvent ou ne peuvent pas apparaître en surface. Ainsi, si je dis que le thème du verbe *LAVARE* en italien est [lava], je ne veux nullement suggérer que cette séquence phonologique constitue une sorte de forme sous-jacente de base du lexème, ou qu'elle a une quelconque priorité par rapport aux autres. En réalité, que la représentation phonologique d'un lexème (ce qui remplit sa rubrique (F)) puisse être réduite à une forme unique sous-jacente est illusoire. Bien entendu, il existe dans les langues des lexèmes qui n'ont qu'une seule représentation phonologique, c'est le cas par exemple de l'adjectif français *PUR* cité plus haut ; il existe même des systèmes linguistiques entiers (les langues isolantes et en partie les langues agglutinantes) dans lesquels les lexèmes ont tous, ou presque, cette caractéristique. Mais le fait même qu'il puisse y avoir des lexèmes qui peuvent varier, et que cela est loin de représenter une exception dans un nombre considérable de langues indique que la non variabilité des lexèmes n'est qu'une des options que les langues ont à disposition. Je ne pourrais pas exprimer cela mieux qu'Aronoff (1994 : 9 ; cf. aussi Zwicky 1992 : 333-336) :

It follows that the search for invariance is only a methodological or aesthetic desideratum (albeit not an important one) within such a framework [la morphologie basée sur les lexèmes, FM] and not of particular theoretical significance.

2.4.5 Qu'est-ce qui est stocké dans le lexique ?

Sous quel format est donc stockée l'information phonologique des lexèmes dans le lexique ? Je me réfère ici aux formes régulières, puisqu'il est évident, comme observé ci-dessus, que chaque forme irrégulière doit être mémorisée séparément (cf. tout de même 2.7.3 et 2.8.5 pour une discussion du concept de régularité). Les approches incrémentales choisissent l'option minimale : en essayant de réduire le plus possible la variation réellement observée à une forme sous-jacente unique, elle postulent que le format canonique est qu'à chaque unité mémorisée dans le lexique corresponde une forme unique. A l'inverse, les défenseurs d'un lexique « full-entry », en particulier dans les approches connexionistes (cf. Bybee 1985 ; 1988 ; Blevins 2006) font l'hypo-

thèse d'un lexique riche et largement redondant. En général, pour ces approches la réduction de la charge mnémotique se fait plutôt sur la base de la fréquence que de la régularité. Il semble avéré que les locuteurs stockent dans leur lexique mental aussi des formes régulières (cf. Stemmerger & MacWhinney 1986 ; 1988 ; Baayen *et al.* 1997 ; 2003 ; Bybee 2007 : 207-208). Par ailleurs, une représentation exacte de tout ce qui est stocké dans le lexique de tous les locuteurs est évidemment impossible ; il est probable que la réponse est variable pour chaque item mémorisé par chaque locuteur. Par exemple, lorsque des enfants italiens disent *una caria* ('une carie') ou *un ceci* ('un pois chiche') au lieu des attendus *una carie* et *un ceci* (en l'occurrence, mes enfants), ils ont évidemment stocké dans leur lexique uniquement la forme plurielle de ces lexèmes (*carie* et *ceci*), dont ils reconstruisent le singulier selon les schémas les plus communs en italien²³. Tout ce que nous pouvons proposer est un modèle idéalisé du lexique tel qu'il peut être mémorisé par les locuteurs. La morphologie incrémentale se fonde, pour élaborer ce modèle, sur une série de postulats, qui ne sont jamais discutés, mais dont l'intérêt théorique est relatif et qui ne sont pas prouvés empiriquement, à savoir que (i) la représentation idéale est celle 'une unité = une forme' et (ii) cette forme doit être la même pour toutes les unités. En s'imposant de telles contraintes, ces approches se condamnent à opérer des choix arbitraires, et font le pari qu'elles sont capables de deviner quelle est 'la' forme qui doit être mémorisée pour chaque lexème. Les théories « full-entry », en revanche, nous permettent de poser notre regard ailleurs, et en particulier de distinguer ce qui *doit* être mémorisé et ce qui ne l'exige pas (mais *peut* néanmoins) l'être. Puisqu'elles partent du présupposé que tout le lexique peut être mémorisé, ce qui est effectivement mémorisé par des locuteurs réels est forcément un sous-ensemble de cela, et le deviner cesse d'être un enjeu. De plus, ces théories nous permettent de prédire comment est (ou pourrait être) ce qui n'est pas mémorisé, le lexique tel qu'il se construit (par exemple, le singulier de *CARIE* et *CECE* dans le lexique de mes enfants).

Une autre question sur laquelle il convient de se pencher brièvement est s'il faut considérer que tout ce qui dans la langue est mémorisé est dans le lexique, ou s'il peut y avoir des séquences phonologiques reliées à une fonction linguistique qui ne le sont

²³ Plus systématiquement, il a été montré, par exemple, pour les verbes du français qu'il existe pour tous les réguliers des formes fléchies qui permettent de reconstruire tout le paradigme, mais que ce ne sont pas les mêmes pour tous les verbes (cf. Bonami & Boyé 2007).

pas. Clairement, si nous considérons que les morphèmes de la tradition ne sont que des exposants d'opérations morphologiques, il en découle qu'il existe des objets linguistiques dont la forme phonologique est mémorisée, mais elle ne l'est pas dans le lexique ; plutôt, cette mémorisation a lieu directement dans la compétence morphologique. De la même manière, il est peut-être cohérent de penser que les noms, les adjectifs et les verbes constituent les lexèmes prototypiques, et donc typiquement mémorisés dans le lexique, et que les unités relevant des autres catégories, notamment les unités fonctionnelles qui servent à la syntaxe (comme les prépositions, les pronoms, les déterminants, etc.) ne sont pas mémorisées dans le lexique, mais coïncident avec les règles syntaxiques qui les insèrent et sont donc mémorisées plutôt dans la compétence syntaxique des locuteurs. La tradition qui veut que ces unités soient des parties du discours au même titre que les autres doit donc être revue (cf. aussi Baker 2003).

Pour revenir au lexème *LAVARE*, dire que son thème principal est [lava] revient à dire que cette forme est celle qui nous permet d'obtenir le plus de formes de ce lexème avec le moins de règles. C'est donc un raccourci descriptif, partiellement arbitraire, que j'adopte, parce que, malheureusement, nous sommes obligés, dans la pratique métalinguistique, d'indiquer la forme abstraite d'un lexème, c'est-à-dire l'ensemble de ses formes et le réseau par lequel elles sont liées, par une forme linguistique concrète. Confondre une forme conventionnelle de citation avec 'la' forme du lexème signifierait faire rentrer les morceaux par la fenêtre une fois que nous les avons chassés par la porte.

2.5 Une morphologie réalisationnelle

2.5.1 Caractérisation

Comme je l'ai rappelé plus haut, dans les approches réalisationnelles c'est l'association d'un lexème avec certaines propriétés morphosyntaxiques qui détermine le fait qu'il ait une certaine forme, par exemple via l'adjonction d'un suffixe. En cela, elles se distinguent des approches incrémentales, pour lesquelles le suffixe lui-même porte en lui certaines propriétés morphosyntaxiques. Par exemple, pour une ap-

proche de type incrémental, dans la forme *mangeons* c'est le suffixe *-ons* qui ajoute les valeurs [1^e personne] [pluriel] [présent] [indicatif] au verbe MANGER. Pour une approche de type réalisationnel c'est l'association du verbe MANGER avec les valeurs mentionnées qui demande l'adjonction du suffixe *-ons* (ou, pour être encore plus précis, qu'il y ait une forme de ce lexème qui se termine par [ɔ]). Selon cette approche, les propriétés morphosyntaxiques des mots fléchis sont donc dissociées des formes qui les réalisent. Dans cette perspective nous pouvons considérer que le paradigme d'un lexème peut être présent dans la compétence morphologique des locuteurs indépendamment des formes concrètes qui le remplissent, ce qui rend compte, par exemple, du fait que les locuteurs sont capables, en général, de générer les formes fléchies de lexèmes inconnus, ou du moins de formuler des hypothèses sur leur forme. Si c'est le cas, on peut penser qu'un paradigme est acquis indépendamment et avant les formes qui réalisent concrètement les cases qu'il comporte. En ce sens, la morphologie naturelle (cf. les références ci-dessus) a raison quand elle soutient que l'idéal morphologique est la différenciation maximale entre des formes qui réalisent des significations différentes (cf. aussi l'approche canonique de Corbett 2007a ; 2007b). Plusieurs travaux ont mis en évidence une particularité apparemment paradoxale de l'apprentissage morphologique, à savoir qu'un système plus complexe peut être acquis plus rapidement qu'un système plus simple. Par exemple, il a été montré que les enfants italo-phones maîtrisent l'indicatif présent de leur langue plus tôt que les enfants anglo-phones, bien que ce dernier ne comporte que deux formes (par exemple *loves* pour la 3^e personne du singulier et *love* pour toutes les autres), alors que celui de l'italien en comporte six pour les verbes réguliers, c'est-à-dire une forme différente pour chaque personne (*amo, ami, ama, amiamo, amate, amano*) (cf. Calderone *et al.* 2007 pour plus de détails). Comme l'observent Calderone *et al.* (2007 : 175-176) ce fait ne s'explique pas dans un modèle de l'apprentissage qui considérerait que ce dernier se fait par une décomposition systématique des formes.

Pour renverser le point de vue des morphologues naturalistes, toutefois, ce qui est intéressant dans les langues ce n'est pas le fait que la morphologie peu ou non iconique (par exemple la non différenciation des formes dans le paradigme du présent de l'anglais, mais aussi l'allomorphie thématique) est rare, mais le fait qu'elle est possible tout court. Cela signifie que dans la langue coexistent plusieurs tendances

contradictoires ; en l'occurrence, la tendance à l'iconicité se heurte à une autre tendance qui est celle à l'économie des moyens (cf. la discussion sur le recyclage en morphologie constructionnelle dans la section suivante), qui fait qu'une même forme peut être utilisée pour plusieurs fonctions. Nous avons vu en (5) ci-dessus que l'identité doit être considérée comme une fonction reliant deux formes morphologiques qui a le même statut que les autres.

2.5.2 Conséquences de l'approche réalisationnelle : des simples formes

La dissociation entre une forme et la fonction qu'elle remplit qui est rendue possible par les approches réalisationnelles permet aussi de donner une réponse satisfaisante à un certain nombre de problèmes comme la question de la 'flexion dans la dérivation', c'est-à-dire des lexèmes qui sont apparemment construits sur des formes fléchies (cf. Booij 1994 : 36-41 ; 1996 : 6-11 ; Rainer 1996b). Mis à part les cas évidents, comme les dérivés du français apparemment construits sur la forme du féminin (cf. *grand* → *grandeur*, *fort* → *fortement*), des cas souvent discutés en littérature, comme les composés verbe+nom (VN) des langues romanes ou les composés dans les langues germaniques peuvent être expliqués si on considère qu'ils emploient une forme d'un lexème qui est disponible parce qu'elle est aussi employée par la flexion, mais qui, dans les lexèmes construits en question, ne véhicule aucune instruction de type flexionnel. Dans le cas des composés VN, que j'exemplifie ici sur l'italien, pour expliquer la forme du premier membre on a fait l'hypothèse qu'il s'agit (i) d'un impératif, (ii) d'une 3^e personne du singulier du présent de l'indicatif, (iii) du thème verbal 'nu' (Vogel & Napoli 1995), (iv) d'un thème spécial du verbe employé par la morphologie constructionnelle (Bisetto 1999)²⁴. Les hypothèses (ii), (iii) et (iv) rencontrent des problèmes formels : elles sont obligées d'expliquer le décalage entre la forme identifiée comme base et celle effectivement observée dans les composés ; l'hypothèse (i) ne rencontre aucun obstacle du point de vue de la forme (la forme du V dans les VN est systématiquement homophone à l'impératif). Le Tableau 1 (que j'adapte de Thornton 2005 : 158) montre des exemples de composés VN en italien, en mettant en évidence l'identité ou la non identité entre leur premier membre et les

24 Cf. Floricic (2008) pour un compte-rendu détaillé des différentes hypothèses.

formes fléchies des verbes correspondants avec lesquelles il devrait coïncider selon les hypothèses mentionnées ci-dessus (pour l'hypothèse (iv) – thème du verbe – je prends comme exemple l'imparfait, dont le thème est le plus stable entre les classes verbales de l'italien)²⁵

lexème	impératif (i)	3 sg. (ii)	imparfait (iii)	dérivés (iv)	VN
PORTARE 'porter'	porta	porta	portavo	portatore 'porteur'	portabagagli 'porte-bagages'
TENDERE 'tendre'	tendi	tende	tendevo	tenditore 'tenseur'	tendinghia 'tenseur pour courroies'
COPRIRE 'couvrir'	copri	copre	coprivo	copritore 'couvreur'	copricapo 'couvrechef'
PULIRE 'nettoyer'	pulisci	pulisce	pulivo	pulitore 'nettoyeur'	puliscipenne 'nettoyeur pour stylos'
CONTENERE 'contenir'	contieni	contiene	contenevo	contenitore 'conteneur'	contienipancia 'ceinture qui contient le ventre'

Tableau 1 : thèmes verbaux et composés VN en italien

L'hypothèse de l'impératif, qui a souvent été écartée sur une base sémantico-pragmatique, peut être réhabilitée, si on considère que ce que l'on retrouve dans les composés VN n'est pas un impératif, mais la forme qui sert en flexion à réaliser l'impératif et en dérivation à construire des composés verbo-nominaux (cf. aussi Rainer 2001 : 389-391 ; Floricic 2008). Dans d'autres langues, par exemple dans les langues germaniques, il existe plusieurs exemples de composés dans lesquels un des éléments est formellement identique à une forme fléchie du lexème, sans que l'on puisse identifier dans la sémantique du composé les valeurs morphosyntaxiques correspondantes. Il peut même arriver que plusieurs formes soient disponibles pour le même lexème

²⁵ Par rapport à Thornton (2005) je rajoute le verbe *CONTENERE*, qui présente une diphtongaison à l'impératif. Une recherche Google (janvier 2010) a donné huit attestations utiles de *contienipancia*.

dans différents composés (9a)²⁶. Dans une optique incrémentale ces composés ont souvent été interprétés comme comportant un interfixe vide ou un élément de liaison (« linking element », cf. Dressler & Merlini Barbaresi 1991 ; Dressler *et al.* 2001 ; Libben *et al.* 2009). Dans Montermini (2010) j’ai soutenu que des cas semblables s’observent, plus rarement, dans les langues romanes, comme en italien : dans les exemples de (9b), par exemple, le deuxième élément peut correspondre à un singulier ou à un pluriel mais les composés peuvent être synonymes²⁷ :

(9)

a. allemand

Kindbett ‘lit pour enfants (enfant_{NOM}+lit)’

Kindergarten ‘crèche (enfant_{PL}+jardin)’

Kindesannahme ‘adoption (enfant_{GEN}+adoption)’

b. italien

il portafoglio / il portafogli

il marciapiede / il marciapiedi

Du point de vue sémantique, il ne semble pas justifié de considérer que les exemples de (9a) *Kindergarten* et *Kindesannahme* sont construits, respectivement, sur un pluriel et sur un génitif, et les triplets cités dans la note 22 confirment cette idée. L’hypothèse qu’en (9a) *-er-* et *-es-* constituent des linking elements, c’est-à-dire des morphèmes dépourvus de sens, semble être démentie par le fait que dans les composés allemands où apparaissent des lexèmes qui ont une flexion non concaténative (par exemple un pluriel à umlaut, avec ou sans suffixe) c’est toujours la forme qui est aussi utilisée en flexion qu’on retrouve, encore une fois avec la possibilité d’avoir des doublons :

26 Sur Google on trouve un nombre non négligeable de triplets comme *Kindbett*, *Kinderbett*, *Kindesbett* ou *Kindannahme*, *Kinderannahme*, *Kindesannahme*.

27 *Portafogli* e *marciapiedi* peuvent fonctionner aussi comme pluriels de *PORTAFOGLIO* et *MARCIAPIEDE*.

(10) allemand²⁸

Hausmarkt	Häusermarkt	‘marché immobilier (maison+marché)’
Buchkunde	Bücherkunde	‘guide bibliographique (livre+information)’
Stadtbau	Städtebau	‘construction urbaine (ville+construction)’
Mantelhaus	Mäntelhaus	‘maison des manteaux (manteau+maison)’
Muttergemeinschaft	Müttergemeinschaft	‘société des mères (mère+société)’

On pourrait objecter que dans les exemples cités l'apparition du pluriel peut avoir une justification sémantique : par exemple, un guide bibliographique concerne forcément plusieurs livres. La présence de formes fléchies à l'intérieur des composés est un fait très commun (cf. Booij 1994 ; 1996 ; Rainer 1996b ; Bauer 2006, entre autres). En réalité, si la sémantique était le critère décisif, on ne devrait pas s'attendre, parallèlement, à avoir *Hausmarkt*, *Buchkunde* ou *Muttergemeinschaft*. On trouve même des composés dans lesquels le premier élément est formellement identique au pluriel, alors que la seule interprétation sémantique possible, ou du moins la plus plausible, est au singulier, comme *Spinnennetz* ('toile d'araignée') *Männerstimme* ('voix d'homme, les deux cités par Blevins 2001) ou *Hühnerei* ('œuf de poule', Krott 2009 : 123). On peut considérer qu'il y a là une convergence de facteurs : les composés 'optimaux' sont ceux où la forme du pluriel correspond effectivement à une interprétation sémantique plurielle possible, ce qui n'exclut pas la possibilité d'utiliser une des formes du lexème indépendamment des valeurs morphosyntaxiques qui leur sont associées en flexion (cf. aussi Blevins 2001)²⁹.

Il est clair que des données comme celles de (10) sont difficilement explicables dans une approche à la morphologie qui considérerait qu'il y a une différence de nature entre les procédés morphologiques affixaux et concaténatifs et les procédés non concaténatifs. La dissociation entre les formes et les valeurs sémantiques qu'elles réa-

28 Toutes les formes citées en (10) sont attestées sur le Web (recherche Google janvier 2010). *Häusermarkt* et *Städtebau* sont pris de Libben *et al.* (2009). *Mantelhaus* et *Mäntelhaus* sont les noms de plusieurs magasins repérés sur Google.

29 Libben *et al.* (2009 : 151) citent aussi l'exemple de *Zukunftsangst* ('peur du futur, futur+peur') qui semblerait militer plutôt en faveur de l'hypothèse que -s- est un interfixe, puisque ZUKUNFT (un nom féminin) n'a aucune forme fléchie dans son paradigme se terminant par [s]. Toutefois, imaginer qu'un lexème possède une forme qui n'est utilisée qu'en dérivation ou en composition n'a rien d'aberrant. En revanche, lorsqu'un mot apparaît avec un umlaut dans un composé, il possède toujours une forme avec umlaut aussi dans son paradigme flexionnel.

lisent est une des avancées principales qui sont rendues possibles par les approches inférentielles-réalisationnelles de la morphologie. Dans la section suivante j'esquisserai les avantages que ce type de morphologie présente du point de vue cognitif et acquisitionnel.

2.6 Une morphologie émergente

2.6.1 Caractérisation

Dans les sections précédentes j'ai déjà caractérisé l'approche à la morphologie que je développe comme inférentielle et réalisationnelle. A celles-ci, je voudrais ajouter une troisième dimension, et définir cette morphologie comme 'émergente' (ou « abstraactive », selon la définition de Blevins 2006). Cette étiquette renvoie à l'idée que celles que j'ai jusqu'à présent informellement défini 'règles' ou 'opérations' morphologiques émergent dans la compétence morphologique des locuteurs en tant que généralisations à partir du lexique, et que le lexique émerge lui-même comme une généralisation à partir des manifestations concrètes de la langue (énoncés) auxquels un locuteur est exposé. Cette conception 'émergente' de la morphologie se situe dans la lignée d'une série de travaux qui considèrent le lexique comme un réseau dans lequel les constructions morphologiques émergent comme des schémas à partir des associations, formelles et sémantiques, entre des formes ou des portions de formes, tels les travaux de Bybee (1985 ; 1988 ; 1995) ; Ford *et al.* (1997) ; Burzio (2002, ms.) ; Blevins (2006) ; Pirrelli (2007), et plus en général aux approches connexionistes (cf. Westermann & Plunkett 2007 pour un aperçu). Elle est aussi compatible avec cette évolution récente de la morphologie qui va sous le nom de Morphologie Constructionnelle (« Construction Morphology »), proposée notamment par Booiij (2005 ; 2007 ; 2009a).

2.6.2 Règles vs. analogie

Il est clair qu'une conception non-symbolique de la morphologie est plus compatible avec cette vision émergente qu'une conception qui adopte comme primitifs des

objets de la même nature qui se combinent entre eux³⁰. Cette conception de la morphologie reprend le concept classique d'analogie pour essayer de lui donner une modélisation (cf. Derwing & Skousen 1988 ; Skousen 1989 et, pour un aperçu des approches les plus récentes à l'analogie, Blevins & Blevins 2009). Elle s'oppose à l'idée qu'il existe une césure nette entre une grammaire procédurale et un lexique d'items mémorisés ; au contraire, ce sont les liens existants entre ces items qui font émerger des régularités et des sous-régularités qui structurent le premier (cf. Bates & Goodman 1999). Comme l'observent Blevins & Blevins (2009 : 10), rien n'empêche de considérer que celles que l'on appelle 'règles' soient, en effet, des analogies très générales ; si cela est vrai, la différence entre règles et analogies est plutôt quantitative que qualitative (cf. la discussion sur la naissance de nouveaux schémas dérivationnels, comme celui des noms en *-stroika* en 2.4.3, cf. aussi Bertinetto 2009 : 117-118). Enfin, cette vision contribue aussi à projeter une nouvelle lumière sur des notions comme celle de régularité ou d'irrégularité : des travaux récents (par exemple Baayen *et al.* 2003 ; Albright 2002) ont montré que l'analogie n'opère pas uniquement sur les procédés irréguliers, mais que ceux réguliers sont sensibles aussi à des effets de fréquence ou de voisinage (par exemple phonologique). De tels travaux nient l'idée qu'il existe une différence dans le traitement de la morphologie régulière et de la morphologie irrégulière de la part des locuteurs, en ouvrant la voie un modèle probabiliste de l'organisation morphologique (cf. Baayen 2003 ; Albright 2009 ; Plag 2009).

Plusieurs modèles récents utilisent, pour modéliser l'analogie, le paramètre du voisinage entre les expressions linguistiques. Je reprendrai ici l'idée de « proximité morphologique » et en particulier de « proximité lexicale » (qui constitue une extension de la première), telles qu'elles sont définies par Hathout (2009 : 32-35)³¹ :

- a. Deux formes qui partagent des propriétés sémantiques, formelles ou catégorielles sont lexicalement proches.
- b. La proximité lexicale de deux lexèmes est d'autant plus grande que le nombre des propriétés sémantiques, formelles et catégorielles qu'ils partagent est grand.

30 Mais cf. Baayen (2003 : 232) sur l'introduction progressive de l'idée d'émergence et des explications de type probabiliste dans des approches symboliques, comme par exemple la Théorie de l'Optimalité (cf. aussi Burzio 2002 ; Bermúdez-Otero 2008).

31 Hathout reconnaît la proximité de sa proposition avec celle de Jackendoff (1975). Elle est proche aussi de la « Representational Entailment Hypothesis » de Burzio (2002 ; ms.).

- c. La proximité lexicale de deux lexèmes est d'autant plus grande que les propriétés qu'ils partagent leur sont spécifiques, c'est-à-dire que peu d'autres lexèmes en sont munis.

L'idée de base est que toutes les formes présentes dans le lexique sont connectées, que de ce fait chaque forme exerce une influence sur chaque autre forme et que cette influence peut être mesurée sur la base de la proximité formelle (par exemple en termes de phonèmes partagés) et sémantique (par exemple en termes de traits sémantiques partagés).

Pour illustrer ce point, je prends l'exemple de la palatalisation du segment [k] devant une voyelle antérieure en italien. Cette alternance s'observe principalement devant le suffixe pluriel *-i* et devant des suffixes dérivationnels commençant par la même voyelle (par exemple *-ità*, *-ista*, *-ismo*, *izzare*, *-issimo*, etc.), mais pas devant d'autres ; de plus, tous les lexèmes dont la dernière consonne est [k] ne présentent pas de palatalisation dans les mêmes contextes (cf. les données en (11)) :

(11)	elasti[k]o 'élastique'	elasti[tʃ]i 'élastiques'	elasti[tʃ]ità 'élasticité'	elasti[k]ino 'petit élastique'
	chimi[k]o 'chimique'	chimi[tʃ]i 'chimiques'		chimi[k]iera 'chimiquier'
	anti[k]o 'ancien'	anti[k]i 'anciens'	anti[k]ità 'antiquité'	anti[k]issimo 'très ancien'
	opa[k]o 'opaque'	opa[k]i 'opaque'	opa[tʃ]ità 'opacité'	

Plusieurs auteurs (cf. Dressler 1985a : 168-180 ; Celata & Bertinetto 2005 ; van de Veer 2006) ont plus ou moins explicitement reconnu que la palatalisation en question est un phénomène lexical en synchronie. Il s'agit du reflet d'une règle phonologique qui était active anciennement dans les langues romanes, mais qui ne l'est plus en ita-

lien contemporain (les séquences [tʃo] et [ki] sont possibles et fréquentes). Cependant, la relation entre [k] et [tʃ] peut encore être perçue dans certains contextes par les locuteurs et avoir un effet sur la construction de nouveaux lexèmes. Considérons les exemples suivants, qui montrent des lexèmes semblables du point de vue segmental qui présentent et ne présentent pas la palatalisation :

(12) avec palatalisation	sans palatalisation
cúbi[tʃ]i	cári[k]i
‘cubiques’	‘chargés’
amí[tʃ]i	antí[k]i
‘amis’	‘anciens’
austría[tʃ]i	ubriá[k]i
‘autrichiens’	‘ivres’
sínda[tʃ]i	ába[k]i
‘maires’	‘abaques’
gré[tʃ]i	spré[k]i
‘grecs’	‘gaspillages’

En observant les formes attestées, il est possible de tirer quelques généralisations, et en particulier d’identifier deux paramètres qui favorisent la palatalisation : (i) la structure accentuelle du lexème (paroxyton ou proparoxyton) et (ii) le timbre de la voyelle qui précède la consonne candidate à la palatalisation ([i] vs. une autre voyelle ou une sonante). En (13a) j’indique la structure typique pour les lexèmes qui palatalisent et pour ceux qui ne palatalisent pas. (13b) exprime la même généralisation sous la forme d’une hiérarchie, les formes les plus à gauche étant les plus palatalisantes³² (cf. aussi,

32 Il existe d’autres cas de palatalisation qui impliquent des noms en italien, par exemple la palatalisation de [g] dans *bel[g]a* – *bel[dʒ]i* (‘belge(s)'), *psicologo* – *psicologo[dʒ]i* (‘psychologue(s)'). Il s’agit cependant de cas plus marginaux et qui peuvent difficilement être étendus à des mots inconnus.

pour des résultats similaires dans une approche expérimentale, Krämer 2009a) :

(13) a.

	iko	Vko
paroxyton	☞ anti[k]i	☞ ubriá[k]i
	amí[tʃ]i	gré[tʃ]i
proparoxyton	☞ elásti[tʃ]i	☞ sínda[tʃ]i
	cári[k]i	ába[k]i

b.

proparoxyton		paroxyton	
iko	Vko	iko	Vko
elásti[tʃ]i	sínda[tʃ]i	anti[k]i	ubriá[k]i

Le croisement des deux paramètres nous permet de délimiter un objet morphologique, le suffixe *-ico*, qui palatalise toujours et impose une accentuation proparoxytone à ses dérivés, bien que, naturellement, tous les lexèmes qui se terminent par cette séquence et palatalisent ne comportent pas forcément ce suffixe, ou ne sont pas analysables en synchronie (comme *elastico* ou *chimico*). Observons maintenant plus en détail les lexèmes du type *sindaco* / *abaco*. Un dictionnaire de l'italien de taille moyenne (Zingarelli 1997) contient 61 noms ou adjectifs proparoxytons se terminant par la séquence [ako], candidats donc à la palatalisation. La proportion est la suivante :

avec palatalisation	sans palatalisation	avec ou sans palatalisation
50	8	3

Tableau 2 : nombre de lexèmes proparoxytons en [ako] avec ou sans palatalisation (dans Zingarelli 1997)

Parmi les 53 lexèmes qui peuvent palataliser, 35 contiennent le suffixe *-iaco*, qui dans beaucoup de cas est toujours transparent, bien qu'il ne soit plus productif (par

exemple *austriaco*, *elefantiaco* ‘éléphantiaque’, *paradisiaco* ‘paradisique’...). Des 18 lexèmes qui restent, 9 autres se terminent par la séquence [iako], et la plupart sont des lexèmes très rares. Dans le même dictionnaire, il y a 134 lexèmes en [iko] (sur 4 288) qui sont proparoxytons mais ne palatalisent pas, comme *carico*. Parmi ceux-ci, le seul qui est peut-être transparent est *francico* (‘francique’), pour lequel la palatalisation entraînerait la présence de l’affriquée [tʃ] dans deux syllabes contiguës (cf. *['frantʃitʃi]³³).

Tous ces faits suggèrent que la phonologie à elle seule ne peut pas rendre compte des cas dans lesquels la palatalisation se réalise. La hiérarchie de (13b) peut être affinée en y ajoutant le paramètre ‘morphologiquement transparent’ :

(14)

proparoxyton				paroxyton	
iko		Vko		iko	Vko
transp.	non transp.	transp.	non transp.		
cúbi[tʃ]i	elásti[tʃ]i	austriá[tʃ]i	sínda[tʃ]i	antí[k]i	ubriá[k]i

Dans les exemples présentés en (11) il y avait l’adjectif *opaco* qui fait partie d’un petit groupe de lexèmes (qui comprend aussi *cieco* ‘aveugle’ et *mendico* ‘mendiant’) qui ont un comportement quelque peu différent de ceux qu’on a vus jusqu’ici, en ce qu’ils ne palatalisent pas au pluriel, mais uniquement dans certains dérivés. En principe, lorsque les locuteurs utilisent un lexème comme ceux de (12) avec un suffixe qui peut entraîner une palatalisation, il ne devrait pas y avoir d’hésitations : il s’agit de lexèmes qui soit palatalisent dans tous les contextes appropriés, soit ne palatalisent en aucun contexte. En revanche, lorsqu’ils utilisent un lexème du type de *opaco*, ils doivent se

33 L’impossibilité ou la difficulté à avoir deux segments identiques dans deux constituants (par exemple deux syllabes) contigus est un principe largement documenté dans les langues. Il est par exemple exprimé dans le Obligatory Contour Principle, dont on verra une application en 3.3.3). Des restrictions semblables opèrent dans d’autres cas dans la morphologie de l’italien. Thornton (1988 : 356), par exemple, a observé que le suffixe déverbal *-zione* (phonologiquement [tsjone]) ne s’applique pas à des verbes qui comportent le phonème [ts] dans la dernière syllabe du thème (comme *disprezzare* ‘mépriser’, *indirizzare* ‘adresser’), mais que la même restriction ne s’applique pas aux verbes qui comportent, à la même position, le phonème [dz] (par exemple, les verbes contenant le suffixe *-izzare*).

baser sur d'autres paramètres que sur son appartenance ou non à la classe des lexèmes palatalisants. Le suffixe *-issimo* normalement entraîne la palatalisation de la base à laquelle il s'applique (s'il s'agit d'un lexème palatalisant, cf. *amicissimo* 'très ami', *comiccissimo* 'très comique') ; ni *opaco* ni *cieco* ne possèdent un dérivé superlatif en *-issimo* lexicalisé (qui soit, par exemple, enregistré dans les dictionnaires), même si leur sémantique est tout à fait compatible avec ce suffixe. Dans le Tableau 3 je donne le nombre d'occurrences des dérivés de *opaco* et *cieco* + *-issimo*, avec ou sans palatalisation, sur Google (données brutes, recherche de mars 2010), comparé avec le nombre d'occurrences du lexème non dérivé et du dérivé en *-ità* (qui est un lexème d'usage commun pour les deux adjectifs)³⁴ :

opa[k]o 1 130k	opa[tʃ]ità 176k	opa[k]issimo 518	opa[tʃ]issimo 3 922
cie[k]o 4 750k	ce[tʃ]ità 298k	c(i)e[k]issimo 588	c(i)e[tʃ]issimo –

Tableau 3 : dérivés de *opaco* et *cieco* en *-issimo* avec ou sans palatalisation

La donnée la plus frappante est sûrement l'absence absolue de dérivés en *-issimo* de *cieco* avec palatalisation, face à une très forte prévalence de dérivés de ce type pour *opaco* (7,5 fois de plus que les dérivés sans palatalisation). Cette donnée peut être mise en relation avec la proportion d'occurrences du dérivé correspondant en *-ità* : *opacità* est presque deux fois plus fréquent par rapport à sa base que *cecità*. Sans compter que *opacità* appartient plutôt à la langue courante, alors que *cecità* est plutôt typique du langage technique et médical, et est probablement surreprésenté sur le Web. De plus, la base [tʃetʃ] n'apparaît que dans le dérivé *cecità*, alors que la base [opatʃ] apparaît dans un certain nombre d'autres dérivés (*opacizzare*, *opacificare*) et peut donc être perçue comme globalement plus disponible par les locuteurs. Cet exemple montre que les choix des locuteurs pour la construction de nouvelles unités ne sont pas univoques, et se basent sur le voisinage avec des éléments existants dans le lexique.

³⁴ L'apparition du signe *i* dans *cieco* mais non dans *cecità* est un fait purement orthographique.

2.6.3 Une morphologie à contraintes

Comme je l'ai observé plus haut, l'application des contraintes à l'analyse morphologique est un des traits communs aux travaux de morphologie réalisés dans l'équipe ERSS de Toulouse depuis une dizaine d'années. Ce modèle est clairement inspiré de la Théorie de l'Optimalité (OT) dans l'idée que les formes linguistiques effectivement observées sont le fruit de l'interaction entre plusieurs contraintes différentes, partiellement contradictoires et hiérarchisées. Comme en OT, il existe des contraintes de fidélité et des contraintes anti-marque. Cependant, OT est née pour traiter des phénomènes phonologiques, et la plupart des analyses de faits morphologiques proposés par des optimalistes mettent en jeu principalement, ou exclusivement, des contraintes d'ordre phonologique, bien que plusieurs travaux tendent de plus en plus à intégrer des contraintes purement morphologiques³⁵. La phonologie seule, cependant, n'est clairement pas suffisante pour rendre compte de la forme réellement observable de tous les lexèmes morphologiquement complexes. En traitant des dérivés comme le français *iotacisme*, « construits sur le nom d'une lettre grecque pour désigner un phénomène phonétique, phonologique ou stylistique », Roché (2007 : 50-51) donne un exemple éclairant de l'utilité d'intégrer des contraintes phonologiques et lexicales pour rendre compte de la forme attestée des lexèmes dérivés :

- (15)
- | | | |
|-------|---|------------|
| iota | → | iotacisme |
| bêta | → | bétacisme |
| gamma | → | gammacisme |
| rhô | → | rhotacisme |
| sigma | → | sigmatisme |

La présence d'un [s] dans *iotacisme* (et dans *lambdacisme*) est l'effet d'un héritage historique, puisque ce lexème est entré en français déjà construit du grec ancien en passant par le latin. Il a représenté le prototype pour les autres lexèmes de la série construits

³⁵ Sans souci d'exhaustivité, pour des analyses OT qui intègrent des contraintes d'ordre morphologique cf. par exemple Russell (1997) ; Yip (1998) ; McCarthy (2001) ; Xu (2007) ; Ohannesian & Pons (2009).

en français. La présence d'un [s] dans *bétacisme* et *gammacisme* ne s'explique que par la nécessité de faire rentrer ces dérivés dans la même série que *iotacisme* et *lambdacisme* : la consonne épenthétique par défaut du français est [t], et le [s], de plus, viole une contrainte qui vise à éviter la répétition de deux consonnes identiques dans deux syllabes contiguës ou proches. Dans *rhotacisme* c'est un segment plus long qui est ajouté à la base ; puisque la base idéale pour ces dérivés est dissyllabique (ce qui est exprimé par une contrainte phonologique anti-marque) et *rhô* ne comporte qu'une seule syllabe, l'insertion de [tas] permet à la fois de respecter cette contrainte et d'insérer le dérivé dans la série de *bétacisme*, *deltacisme*, *zétacisme*. Pour la base *sigma*, enfin, la contrainte 'pas de répétition du même segment dans deux syllabes proches' serait violée deux fois par la forme ^o*sigmacisme* ; parallèlement cette forme serait favorite par la contrainte d'intégration dans la série lexicale, mais il existe une autre série lexicale qui contient des lexèmes d'origine grecque se terminant en [gma] et qui ont leurs dérivés en [gmat] : *dogmatisme*, *pragmatisme*, *magmatisme*, etc. Sémantiquement ils sont plus éloignés des autres dérivés de noms de lettres grecques, mais l'alliance entre la contrainte anti-marque et la possibilité d'intégrer le lexème dans la série en [gmat] prime sur la nécessité de l'insérer dans la série en [asisme] (sur l'alliance entre contraintes cf. Burzio 2002 : 152 ; Hathout 2009 : 42). Je réélabore légèrement la formalisation proposée par Roché pour la rendre plus semblable à une notation OT classique et je donne en (16), pour comparaison, les tableaux relatifs à *bétacisme*, *rhotacisme* et *sigmatisme*³⁶ :

(16)

a. bétacisme

	$\sigma.\sigma$.isme	*C _i .C _i .C _i	[gmatism]	[tasism]	[sism]
bétatisme		*		*	*
☞ bétacisme		*			

36 Les contraintes $\sigma.\sigma$.isme et *C_i.C_i.C_i indiquent, respectivement, la tendance à préférer une base dissyllabique et celle à éviter des consonnes identiques dans l'attaque des syllabes proches. Les formes *bétaïsme, rhoïsme et *sigmaïsme sont exclues par une contrainte, superordonnée aux autres, *hyatus ; *bétisme, *rhisme et *sigmisme à la fois par $\sigma.\sigma$.isme et par la contrainte de fidélité base – dérivé (la dernière aussi par *C_i.C_i.C_i).

b. rhotacisme

	$\sigma.\sigma$.isme	*C _i .C _i .C _i	[gmatism]	[tasism]	[sism]
rhotisme	*			*	*
rhocisme	*	*		*	
rhotatisme		*		*	*
☞ rhotacisme		*			

c. sigmatisme

	$\sigma.\sigma$.isme	*C _i .C _i .C _i	[gmatism]	[tasism]	[sism]
☞ sigmatisme				*	*
sigmacisme		*	*	*	

Les contraintes qui se trouvent à la droite des tableaux sont censées représenter de manière informelle le poids du lexique sur la forme qu'ont les lexèmes dérivés. Malgré leur forme, elles n'expriment pas uniquement des spécifications formelles. Par exemple, la contrainte [tasism] doit être interprétée comme une préférence pour les lexèmes construits sur la base d'un nom de lettre grecque à se terminer par cette séquence. Elle domine une autre contrainte qui exprime la préférence à se terminer, à défaut, par la séquence [sism].

Il est évident que ce type de contraintes est plus compatible avec une conception des affixes comme des exposants de règles morphologiques que comme des signes. Dans une perspective morphématique le statut à attribuer aux séquences [sism], [tasism], [gmatism], ou à des sous-parties de celles-ci, serait problématique : [s] et [tas], par exemple, sont-ils des interfixes ? *-isme* et *-issime* sont-ils des allomorphes de *-isme* ? Au contraire, dans une perspective « output-based », comme celle défendue ici, qui ne s'oblige pas à identifier un statut morphémique pour chacun des segments, le suffixe n'a pas une forme de base et des variantes ; plutôt, la règle morphologique *-ISME*³⁷ liste plusieurs formats phonologiques pour les lexèmes qu'elle construit, de plus en plus spécifiés (par exemple la suffixation de [sism], [tasism], [gmatism]) et dont le

³⁷ Pour distinguer les règles morphologiques de leurs exposants (par exemple le suffixe *-isme*), j'adopte la convention de les marquer en petites majuscules, en m'inspirant de celle utilisée pour les lexèmes.

choix dépend de l'interaction de facteurs purement phonologiques avec d'autres facteurs, et notamment la pression du lexique existant.

Dans les tableaux de (16) les contraintes sont indiquées de manière informelle. Jusqu'à présent, Hathout (2009 : 42-53) représente la tentative la plus élaborée de formaliser le type de morphologie à contraintes défendu ici. Je renvoie le lecteur à ce texte pour un traitement détaillé. Pour résumer, dans le système proposé par Hathout les contraintes tiennent compte à la fois de la proximité formelle et sémantique entre deux lexèmes dans le lexique, de la fréquence d'un lexème et de la densité de son environnement, c'est-à-dire du nombre de voisins morphologiques que chaque lexème possède. Ces contraintes opèrent parallèlement aux contraintes phonologiques classiques (pour quelques contraintes spécifiques au français cf. notamment Plénat & Roché 2003 ; Plénat 2009b). Plusieurs exemples de contraintes et de leur emplois seront illustrés dans les sections qui suivent.

En OT, au moins dans sa version originale (McCarthy & Prince 2001 (1993) ; Prince & Smolensky 2002 (1993)), les contraintes découlent de principes universaux d'organisation de la grammaire ; autrement dit, elles sont un reflet de la Grammaire Universelle, et les différences entre les langues sont dues uniquement à un agencement différent des mêmes contraintes. Nous pouvons aussi interpréter les contraintes proposées par Hathout comme des contraintes sur l'organisation du lexique valables à un niveau général. Des contraintes 'descriptives' et propres à chaque langue (language-specific) comme [tasism] >> [sism] ci-dessus seront employées comme des raccourcis pour des contraintes lexicales plus spécifiques, telles celles décrites en Hathout (2009 : 42-53)³⁸.

2.6.4 La directionnalité des procédés morphologiques

De la morphologie incrémentale les approches lexématiques ont hérité l'idée que les procédés de formation de lexèmes sont orientés. Aronoff (1976 : 21), par exemple, définit une règle de construction de lexèmes comme une opération dans laquelle (le gras est mien) : « [a] **new** word is formed by applying a regular rule to a

38 Sur l'emploi de contraintes language-specific cf. cependant Mohanan & Mohanan (2003) ; Xu (2007 : 3).

single **already existing** word »³⁹. Malheureusement, les auteurs n'ont pas toujours distingué entre une directionnalité qui peut être définie sur la base d'une plus grande complexité sémantique ou cognitive et une simple priorité chronologique ou étymologique (sur le principe de directionnalité en morphologie constructionnelle cf. Iacobini 2000 ; Grossmann & Rainer 2004 : 9-11). Naturellement, seule la première nous intéresse pour une étude de la morphologie synchronique d'une langue. Cela exige, par exemple, qu'on renonce à l'idée que les lexèmes sémantiquement plus complexes sont automatiquement, ou même de préférence, plus complexes du point de vue de la forme. Il est bien connu que le même type de signification, qu'il soit simple ou complexe, peut être exprimé aussi bien par des lexèmes construits que par des lexèmes simples. Cela signifie aussi que le principe de directionnalité peut être considéré comme relatif et non pas comme universellement valide. En d'autres termes, alors que dans certains cas il est clair lequel des membres d'un couple dérivationnel est primaire par rapport à l'autre, dans d'autres cette recherche d'une primauté n'est pas seulement difficile à établir, mais elle est simplement inutile. Prenons l'exemple des couples de noms et de verbes construits sur la même base sans affixe explicite en français, comme *saut* / *sauter*, *clou* / *clouer*. Dans le cas de *clou* / *clouer* on peut effectivement soutenir que le nom est primaire par rapport au verbe : en tant que nom d'objet c'est un nom prototypique, et le sens du verbe peut être décrit comme une action portant sur cet objet⁴⁰. En revanche, le nom *saut* désigne plutôt une action ; on serait donc porté à penser que *sauter* soit premier et que *saut* soit le nom d'action dérivé. Mais, comme tous les noms d'action, *saut* est ambigu et désigne aussi un objet, à savoir le résultat de l'action de sauter, et en ce sens *sauter* pourrait être interprété comme une action portant sur l'objet *saut*⁴¹. On ne sort de cette circularité qu'en considérant que *saut* et *sauter* coexistent dans le lexique et qu'ils sont liés par une rela-

39 Un modèle, contemporain de celui d'Aronoff, est celui proposé par Jackendoff (1975) dans lequel les mots sont liés par des règles de redondance. Aronoff (1976 : 31) discute brièvement ce modèle pour conclure que les seules règles qui peuvent être utilisées comme des règles de redondance pour l'analyse du lexique existant sont les règles de construction de mots productives.

40 Sur le principe de directionnalité pour les convertis cf. par exemple Marchand (1964), qui propose de considérer qu'un lexème est primaire, dans une telle relation, si sa signification est nécessaire pour décrire la signification d'un autre (comme pour *clou* par rapport à *clouer*) (cf. aussi Thornton 2004 : 501-502).

41 Pour des observations sur les conversions de noms, adjectifs et verbes en anglais et en allemand cf. Plank (ms.).

tion bidirectionnelle, ce qui revient à reconnaître que le principe de directionnalité peut valoir pour certains couples de lexèmes dans le lexique et pas pour d'autres. Relativiser le principe de directionnalité a plusieurs avantages, et en particulier permet de rendre compte d'une série de phénomènes qui, dans une optique orientée, ne peuvent être expliqués qu'en opérant des choix arbitraires. C'est, par exemple, le cas bien connu des affixes qui s'excluent mutuellement, comme *-isme* et *-iste* en français (cf. Roché 2007), ou des suffixes italiens *-ore* et *-ido* dans certains couples comme *pallore* ('pâleur') / *pallido* ('pâle'). Roché (2008 : 1573-1575) donne plusieurs exemples du principe d'économie « qui fait que la langue, plutôt que de construire une forme nouvelle, réutilise une forme déjà existante ». Parmi ses exemples il y a *éducatif* dans le sens 'qui concerne l'éducation' et non pas 'qui éduque' (cf. Roché 2008 : 1574). Ce 'recyclage' de l'adjectif *éducatif* est possible parce qu'en français le triplet constitué d'un verbe, du nom d'action correspondant en *-tion* et de l'adjectif correspondant en *-if* est courant, et on ne peut en rendre compte que si on considère que chacun de ses membres peut fonctionner comme la 'base' de l'autre. Ce genre d'observations étaient déjà implicites dans les théories du lexique fondées sur la redondance à la Jackendoff, même si depuis la recherche en morphologie s'est plutôt développée dans la direction d'un modèle à règles orientées⁴². Dans mes travaux précédents (Montermini 2005 ; 2008b : 203-205) j'ai utilisé cette idée pour expliquer le fait que les dérivés par des préfixes de repérage (y compris *anti-* qui l'est étymologiquement et se comporte comme tel) ont comme base soit un nom (17a), soit un adjectif de relation qui est systématiquement homophone à un adjectif déjà disponible dans le lexique, y compris lorsqu'il est construit sur une base allomorphique (cf. aussi Hathout 2009 : 110-111 ; Roché 2009 : 83) (17b-c), et il n'est pas rare de trouver des doublons voire des triplets :

42 En commentant les règles de redondance de Jackendoff (1975), Covington (1981 : 32, emphase dans l'original) observe que « the language provides no motivation for choosing *aggression*, *aggressor* or *aggressive* as prior to the other two members of a triple. [...] although it is clear that no member of the triple is prior in English **as a whole**, it is quite plausible that one of them is prior **in each speaker's head** ».

(17)

- | | |
|---------------------|-----------------------------|
| a. un anti-fièvre | un traitement anti-fièvre |
| b. un anti-fiévreux | un traitement anti-fiévreux |
| c. un antipyrétique | un traitement antipyrétique |

Ce traitement représente une évolution de l'explication basée sur le principe de copie proposée par Corbin (1999 ; 2004), selon laquelle le suffixe relationnel qui apparaît, par exemple dans *anti-fiévreux* serait un « suffixoïde calquant le suffixe qui sert à construire un adjectif sur la même base nominale » (Corbin 1999 : 68) : une fois qu'on s'est éloigné complètement de l'optique morceaologique on n'est plus dans l'obligation de segmenter le *fiévreux* qu'on retrouve dans *anti-fiévreux* et d'étiqueter les bouts que nous obtenons en parlant de « suffixoïdes », une explication qui, de toute façon, devrait être revue pour être adaptée aux cas comme *antipyrétique*.

Pour prendre un exemple plus détaillé on peut observer les noms italiens en *-anza* / *-enza*. Ce suffixe s'ajoute généralement à une base qui a la forme d'un participe présent (cf. Gaeta 1999) et rentre dans une famille dérivationnelle assez longue, que quelques lexèmes exemplifient de manière relativement transparente. Cette série peut être illustré sur la base du parcours qui relie *assistere* ('assister') à *assistenzialità*⁴³ :

(18)

1	2	3	4	5
assistere	→ assistente	→ assistenza	→ assistenziale	→ assistenzialità
V:'X'	pp:'qui X'	N:'le fait de X'	A:'qui a la propriété de ce qui X'	N:'la propriété de ce qui X'

Cependant, dans le lexique attesté, cette série peut comporter plusieurs lacunes, qui peuvent toucher n'importe laquelle de ses étapes, y compris celle du verbe de départ, comme le montre le tableau ci-dessous⁴⁴ :

43 Pour les nécessités de l'illustration je traite la construction du participe présent comme une relation de type dérivationnel, sans préjuger de son réel statut.

44 Je n'indique dans le tableau que les lexèmes enregistrés dans un dictionnaire moyen d'italien (Zingarelli 1997). Naturellement, le tableau ne tient pas compte de relations lexicales qui se sont maintenant complètement opacifiées en italien, comme celle entre *essere* ('être') et *essenza* ou entre *patire* ('souffrir') et *paziente*. Le tableau n'est pas exhaustif ; les noms en *-anza/-enza* par exemple peuvent construire également des dérivés qui ont un sens collectif, comme *dirigenza*, *militanza*

glose	1	2	3	4	5
‘assister’	assistere	assistente	assistenza	assistenziale	assistenzialità
‘adolescent’		adolescente	adolescenza	adolescenziale	
‘partir’	partire	partente	partenza		
‘essence’			essenza	essenziale	essenzialità
‘patient’		paziente	pazienza		

Tableau 4 : famille dérivationnelle des lexèmes en *-anza/-enza*

Il existe des lexèmes en *-anza/-enza* qui sont ambigus pour la case 3 ou la case 5, comme *aderenza* (‘adhérence’), par exemple dans *migliorare l’aderenza di un pneumatico* (‘améliorer l’adhérence d’un pneu’), qui peut vouloir dire aussi bien ‘le fait d’adhérer au terrain’ et ‘la propriété d’être adhérent’. Parallèlement, un adjectif comme *aderente* peut vouloir dire à la fois ‘qui adhère’ et ‘qui a la propriété d’être adhérent’, c’est-à-dire peut occuper les cases 2 et 4. En effet, cette propriété n’est pas fortuite mais systématique : les cases 4 et 5 peuvent être occupées, respectivement, par un lexème en *-ante/-ente*, qui occupe déjà la case 2, ou par un lexème en *-anza/-enza*, qui occupe déjà la case 3. Cette circonstance est vraie aussi bien pour des lexèmes attestés, que pour des usages ‘non canoniques’ de lexèmes qui ne sont pas enregistrés dans les dictionnaires, en donnant lieu parfois à des doublons. Des exemples des usages des lexèmes ‘recyclés’ en *-ante/-ente* et *-anza/-enza* (indiqués en gras dans le tableau) sont donnés en (19) dessous :

(‘l’ensemble des dirigeants / militants’) (cf. Gaeta 1999 : 557-560).

glose	1	2	3	4	5
'dément'		demente	demenza	demenziale demente (19a)	demenzialità demenza (19b)
'concourir'	concorrere	concorrente	concorrenza	concorrenziale concorrente (19c)	concorrenzialità ⁴⁵ concorrenza (19d)
'accueillir'	accogliere	accogliente	accoglienza		accoglienza (19e)
'pestilence'			pestilenza	pestilenziale	pestilenzialità ⁴⁶ pestilenza (19f)

Tableau 5 : recyclage des lexèmes en -ante/-ente et -anza/-enza

(19)

a. una domanda **demente**

'une question démente'

b. la **demenza** della domanda

'la démence de la question'

c. Non solo per manager ma anche per ragazzi. Il prezzo è **concorrente**, gli optional all'altezza della casa costruttrice⁴⁷.

'Elle n'est pas seulement pour les managers, mais aussi pour les jeunes. Le prix est concurrentiel, les options à la hauteur des standards de la maison'.

d. Per aiutare gli Stati membri ad accelerare l'accoglienza sul mercato dei medicinali generici e migliorare la **concorrenza** dei prezzi, la relazione riporta una rassegna delle misure nazionali e dei loro effetti sull'accoglienza dei generici⁴⁸.

45 52 400 attestations sur Google (données brutes), février 2010.

46 63 attestations sur Google (données brutes), février 2010.

47 http://auto.ciao.it/Opinionisulprodotto/Alfa_Romeo_147_1_6__88399/Start/45.

48 <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/1098&format=DOC&aged=0&language=IT&guiLanguage=en>.

‘Pour aider les Etats membres à accélérer l’accueil sur le marché des médicaments génériques et améliorer la concurrentialité des prix, la relation contient une revue des mesures nationales et de leurs effets sur l’accueil des génériques’.

- e. Lascia stare l’**accoglienza** degli italiani che è relativa (infatti magari un romano o un napoletano può essere più allegro ed accogliente di un milanese o valdostano)⁴⁹.

‘Laisse tomber la capacité d’accueil des italiens, qui est relative (en effet il se peut qu’un romain ou un napolitain soit plus joyeux et accueillant qu’un milanais ou un valdôtain)’.

- f. La monnezza dopo quasi due mesi è tutta lì intatta in tutta la sua **pestilenza**⁵⁰.

‘Les ordures après presque deux mois sont toujours là intactes dans toute leur pestilence’.

Encore une fois, une approche incrémentale, qui représenterait les données comme on l’a fait en (18) serait incapable de capter les allées-retours que les lexèmes en *ante/-ente* et *-anza/-enza* font à l’intérieur de leur série dérivationnelle.

Pour des ensembles de données comme celles illustrées dans les Tableaux 4 et 5 il est désormais plutôt courant de parler de paradigmes dérivationnels (cf. Bauer 1996 ; Roché 2007 ; Hathout 2009 : 36, qui distingue même les paradigmes morphologiques des paradigmes lexicaux), une étiquette qui montre bien que le lexique est une entité structurée et que des effets paradigmatiques existent en dérivation de la même manière qu’ils existent en flexion. C’est au lexique et à son interaction avec la compétence morphologique qu’est consacrée la section suivante.

2.7 Lexique vs. compétence morphologique

2.7.1 Théories du lexique

La relation entre des objets mémorisés et des objets construits « on-line » est une question récurrente dans les études de morphologie. Les théories sur la représentation des mots complexes (que ce soit des mots fléchis ou construits) se situent entre

49 <http://it.answers.yahoo.com/question/index?qid=20090316115831AAH2hS2>.

50 http://www.antoniodipietro.com/2008/06/ancora_leggi_bavaglio.html.

deux pôles, et peuvent globalement être réparties en trois groupes :

-les théories dites « impoverished-entry » (cf. Jackendoff 1975 : 642), selon lesquelles le lexique contient uniquement l'information idiosyncratique, c'est-à-dire non prédictible par règle. Le lexique est une entité fondamentalement non structurée qui ne contiendrait que des unités irrégulières (par exemple les pluriels de l'anglais *oxen* 'bœufs' ou *feet* 'pieds') et aucune unité régulière (par exemple *books* 'livres'). Cette vision du lexique, héritée de la linguistique structuraliste (cf. Bloomfield 1933), a été répandue principalement par la linguistique générative (cf. Chomsky 1965 : 87-88 ; 1995 : 130-131 ; Halle 1973, en particulier pour l'introduction de cette idée dans les études de morphologie) et elle a été également adoptée dans un certain nombre de travaux de psycholinguistique et de linguistique cognitive, par exemple dans le célèbre modèle à deux voies (« Dual Route Model », cf. Pinker 1991 ; Pinker & Ullman 2002).

-les théories « full-entry », selon lesquelles toutes les formes sont mémorisées, indépendamment de leur constitution (Bybee 1985 ; 1988 ; Blevins 2006). Selon ces théories le lexique est structuré et s'organise plutôt en un réseau de correspondances.

-des théories mixtes, selon lesquelles dans le lexique sont mémorisées a minima les formes irrégulières, mais il n'existe aucune nécessité d'imaginer que des formes régulières ne le soient pas aussi, ou, en suivant Jackendoff (2002 : 153) « that the brain stores information non-redundantly » (cf. aussi Hohenhaus 2005 ; 357 ; Booij 2009 : 15 ; Gaeta & Ricca 2009 : 38-39). Comme dans le cas des théories full-entry le lexique est une entité structurée, dans lequel les unités sont mises en relation, par exemple par des règles de redondance (cf. Jackendoff 1975 ; 1997 ; 2002).

Clairement, l'hypothèse du lexique « appauvri » est la plus incompatible avec le modèle de morphologie inférentiel et émergent défendu ici, et celle qui rencontre le plus de problèmes empiriques. Il semble avéré, par exemple, qu'au moins une partie des formes régulières les plus fréquentes sont mémorisées dans le lexique conjointement à celles irrégulières (cf. Stemberger & MacWhinney 1986 ; 1988 Bybee 2007 : 207-208 ; Baayen *et al.* 1997 ; 2003). Même du point de vue acquisitionnel, l'idée d'un

lexique enrichi et non limité aux formes non prévisibles semble plus compatible avec les données observées : plusieurs auteurs proposent en effet de considérer que l'acquisition de la morphologie régulière d'une langue se fait progressivement au fur et au mesure que l'on mémorise de plus en plus de mots qui présentent des propriétés communes, à partir desquels on extrait des régularités (cf. Becker 1993 : 12 ; Jackendoff 1997 : 122 ; Bauer 2001 : 114). En revanche, le choix entre l'hypothèse « full-entry » et une hypothèse intermédiaire est plutôt un faux problème, et en tout cas risque de rester une pure spéculation tant qu'on n'a pas répondu à deux questions, pour lesquelles tout le monde semble avoir une réponse, mais qui ne sont jamais traitées de front, et que je discuterai dans les sections qui suivent.

2.7.2 Qu'est-ce que le lexique ?

A cette première question on peut répondre de deux manières : soit il s'agit du lexique mental, et donc, par définition, il existe tant de lexiques qu'il existe de locuteurs d'une langue, soit il s'agit du lexique idéalisé d'un locuteur idéalisé, une sorte de moyenne de tous les lexiques mentaux⁵¹. Prétendre établir avec précision qu'est-ce qui est mémorisé dans les lexiques mentaux de tous les locuteurs est évidemment une tâche aussi déraisonnable qu'inutile, sans compter qu'il s'agit sans doute d'une entité qui n'est pas stable, ni même au cours de la vie d'un locuteur. Le problème du lexique idéalisé, en revanche, est qu'il s'agit d'une entité qui dépend de facteurs à la fois linguistiques et historico-sociaux. L'éducation scolaire, l'établissement d'un canon littéraire et d'une norme grammaticale, et leur fixation dans les dictionnaires, ou tout autre facteur culturel sont, évidemment, des phénomènes qui ne doivent pas rentrer en compte dans une étude de la compétence linguistique, mais qui contribuent au moins autant que celle-ci à créer un lexique des langues. Malheureusement, aussi pour des raisons pratiques, les chercheurs ont souvent confondu ce lexique idéalisé avec la liste des mots qu'on peut trouver dans les dictionnaires, ou avec ce qu'à l'introspection leur paraissait acceptable (sur la collecte des données en morphologie, cf. 2.9.1 ci-dessous). En réalité, il est probable qu'il faille renoncer entièrement à la prétention de définir avec précision le lexique idéalisé et ses frontières. De ce point de vue, la

51 Sur l'ambiguïté du mot 'lexique' et sur son usage de la part des linguistes cf. aussi Aronoff (1994 : 16-22).

position la moins déraisonnable est de considérer que des formes 'régulières' et 'irrégulières' *peuvent* coexister dans le lexique.

2.7.3 Qu'est-ce qu'un lexème régulier ?

Ci-dessus j'ai cité plusieurs études montrant l'action des effets analogiques de voisinage pour des constructions traditionnellement considérées comme régulières. Encore une fois, une théorie qui préconise une dichotomie rigide entre lexique et règles privilégie une vision du premier comme une entité statique, ou du moins comme un point d'arrivée : une fois qu'un lexème est accueilli dans le lexique ses propriétés (phonologiques, syntaxiques, morphologiques) sont stabilisées ou au plus varient dans le sens d'une plus grande idiosyncrasie. L'idée que l'on puisse délimiter avec précision ce qui, dans un lexème, est idiosyncratique et ce qui est régulier et prévisible est toutefois illusoire, notamment du point de vue sémantique. Dans leur version la plus élémentaire, les règles de construction de lexèmes spécifient, pour les lexèmes complexes, une lecture sémantique qui est une fonction de la lecture de la base (Aronoff 1976 : 22). Cette définition est suffisamment vague pour être difficilement falsifiée par des lexèmes construits réellement observés dans les différentes langues. Elle ne tient pas compte du fait, cependant, que l'interprétation sémantique d'un lexème construit, comme de tout autre lexème, ne se fait pas de façon abstraite, et dépend au moins de deux facteurs : le réseau lexical dans lequel un lexème est inséré (cf. Roché 2009) et la pragmatique.

L'étude de la morphologie constructionnelle ne peut pas faire abstraction de la fonction principale des nouveaux lexèmes qui sont créés, qui est celle d'être une dénomination, dans le sens où on a utilisé cette notion en 2.2.2. Une fois qu'il est construit, un lexème rentre donc dans une double relation, avec les lexèmes qui appartiennent à la même série (par exemple, celle des lexèmes en *-ien*, en *-iste*, en *-ier*, etc.), mais aussi avec ceux qui relèvent du même schème de nomination (cf. Roché 2009 : 81-83 ; à p.), par exemple les noms [+humain] qui indiquent des activités ou des métiers, qui, eux, peuvent avoir plusieurs sources différentes : être des lexèmes simples (*poète*), des dérivés, plus ou moins 'canoniques' (*physicien*, *artiste*, *poissonnier*), des composés (*porte-drapeau*), des lexicalisations de syntagmes ou de sigles (*gendarme*, *CRS*),

des emprunts (*leader*), des lexèmes issus d'un glissement sémantique (*cadre*), etc. A la différence de ce que croient ceux qui le considèrent comme une liste d'irrégularités, le lexique a sa propre organisation qui se superpose aux schémas morphologiques. Il peut arriver que des schémas morphologiques et lexicaux soient en compétition. Par exemple, Roché (à p.) explique la lecture sémantiquement déviante d'*équipier* ('membre d'une équipe', alors que le sens devrait être plutôt 'chef d'équipe', cf. *brigadier* ou *caravanier*) par le fait que la désignation d'une personne par le groupe à laquelle elle appartient correspond à un schème de nomination très répandu (cf. *académicien*, *parlementaire*, *CRS*, etc.).

De plus, plusieurs auteurs ont souligné la dépendance de la lecture sémantique des lexèmes dérivés de facteurs pragmatiques ; certains soutiennent même que, au moins pour certains dérivés, la spécification sémantique ne peut pas être plus spécifique qu'une relation générale entre la base et le dérivé, et que le reste est fait par la pragmatique (cf. Aronoff 1980 ; 1983 : 372 ; 2007 : 812-818 ; cf. aussi la discussion de Plag 1998 sur le suffixe de l'anglais *-ize*). Aronoff (2007 : 817-818), par exemple, cite le verbe anglais *friend* qui a acquis le sens de 'inviter quelqu'un à être ami dans la communauté Facebook' et qui met en jeu une instruction sémantique qui n'est compatible avec aucun des nombreux sens listés dans les taxinomies sémantiques des noms convertis de l'anglais. De la même manière, Roché (2004 : 464) commente dans ces termes l'interprétation du lexème *ardoisier* dans le sens de 'personne chargée d'inscrire des indications pour les coureurs sur une ardoise pendant les courses cyclistes' : « [l]a mémoire mobilisée [...] n'est pas celle du mot [...] mais celle des *realia*. Il y a donc bien construction du sens, mais une construction 'incarnée', dans laquelle le contexte et la connaissance du monde jouent un rôle indispensable à côté de l'opération dérivationnelle proprement dite ». Naturellement, la « connaissance du monde » dont il est question ici peut-être partagée par un nombre plus ou moins grand de locuteurs, et peut même se référer à un monde imaginaire. Dans plusieurs albums d'Astérix⁵² on utilise le mot *amphorissage* pour se référer aux encombrements de chars sur les routes. Ce mot est clairement construit sur l'exemple de *embouteillage* (qui lui-même a une sémantique non transparente), et n'a de sens que dans l'univers fictif créé par les auteurs de cette bande dessinée. Pour revenir à l'exemple d'*ardoisier* ci-dessus, il montre

52 Par exemple, dans *La serpe d'or* (1962), p. 23 ; *L'anniversaire d'Astérix et Obélix* (2009), p. 20.

un lexème qui a un sens institutionnalisé ('personne qui possède ou qui travaille dans des carrières d'ardoise', TLFi) et un sens différent dans un emploi plus spécifique, limité à un domaine spécialisé. Les études sur le lexique et la morphologie constructionnelle dans la plupart des cas font comme s'il était possible d'identifier un sens premier stable pour chaque lexème, qui est institutionnalisé et partagé par tous les locuteurs d'une langue. Or, il est déjà discutable que ce soit le cas pour n'importe quel lexème bien ancré dans le lexique ; et ce n'est certainement pas le cas pour les lexèmes qui ne sont pas institutionnalisés. Je prends comme exemple un lexème construit qui n'est pas dans les dictionnaires d'italien, *poltronista*, construit à partir du nom *poltrona* ('fauteuil') avec le suffixe *-ista*. Sur Google on trouve environ 14 440 attestations de ce lexème (données brutes, recherche effectuée en février 2010), et le lexème semble être en train de se spécialiser dans le sens de 'politique qui est accroché à son siège' (20). Dans ce cas, c'est l'emploi de *poltrona* dans le sens de 'siège dans une assemblée publique (parlement ou autre)' qui est en jeu. Cependant, en observant les contextes d'emploi on trouve plusieurs significations possibles pour ce lexème (20b-d). Chacune d'elles met en jeu un sens différent du lexème *poltrona*, partagé par une communauté plus ou moins grande de locuteurs⁵³ :

(20)

a. 'politique accroché à son siège'

Descrivere un politico 'capo dei **poltronisti**' non configura alcun reato di diffamazione⁵⁴.

'Décrire un politique comme le 'chef des *poltronisti*' ne constitue pas un crime de diffama-

53 Il est significatif que la quasi totalité des sens relevés pour *poltronista* se retrouvent aussi pour un lexème construit sur la même base mais avec un autre suffixe, *poltronaio* (qui en plus peut aussi indiquer un fabricant de fauteuils, conformément à la spécialisation du suffixe en question). Ces suffixes donnent lieu à un nombre particulièrement important de dérivés, puisqu'ils n'opèrent pas de sélection sémantique particulière sur leurs bases. Comme Fradin & Kerleroux (2009) je considère que chaque combinaison de *poltrona* avec *-ista* crée un nouveau lexème ; chacune des unités exemplifiées en (20) répond en effet de façon autonome à la définition de mot que j'ai donnée en 2.2.2 en ce qu'elle : (i) est ou peut être une dénomination, (ii) a les propriétés distributionnelles d'un nom, (iii) pourrait faire partie du lexique.

54 <http://www.comuniclab.it/21984/calderoli-poltronista-il-giudice-non-reato>.

tion’.

b. ‘amant du confort, paresseux’

In Italia più che essere amanti delle passeggiate, sono amanti [...] della comodità... Odio dirlo... ma siamo un popolo di pantofolai e **poltronisti**⁵⁵.

‘En Italie, plutôt que des amants des promenades, on est des amants [...] du confort... Je déteste dire ça... mais nous sommes un peuple d’amants des pantoufles et des fauteuils’.

c. ‘spectateur qui a loué un fauteuil dans un théâtre’

Al momento della sottoscrizione, il **poltronista** indica il suo giorno preferito di occupazione della poltrona che dovrà rispettare⁵⁶.

‘Au moment de la souscription, le locataire d’un fauteuil indique le jour préféré d’occupation du fauteuil qu’il devra respecter’.

d. ‘supporter qui aime regarder le sport à la télévision (par opposition à ceux qui vont au stade)’

Non capisco assolutamente il rapporto stadisti / **poltronisti**. Guarda che allo stadio tanta gente non ci va più, proprio per i tanto bravi ‘stadisti’ che tu citi⁵⁷.

‘Je ne comprends absolument pas la relation entre supporters de stade et supporters de fauteuil. Tu dois savoir que si tant de gens ne vont plus au stade, c’est justement à cause des bons supporters de stade que tu cites’.

Contrairement à ce qu’affirme Fradin (2009 : 98) (le gras est mien, cf. aussi Fradin & Kerleroux 2009) le sens des lexèmes construits n’est pas toujours construit « à partir des sens **lexicalisés** des lexèmes-bases ». (20d), par exemple, met en jeu une sphère d’expérience qui n’est partagée que par un nombre réduit de locuteurs. Le sens de *poltronista*, dans ce cas, s’interprète correctement seulement par opposition à *stadista*, et plutôt qu’au sens du lexème POLTRONA, il fait référence à un prédicat que nous pourrions paraphraser avec ‘regarder du sport à la télévision assis dans son fauteuil’. Ceux

55 <http://www.mauriziomaraglino.it/2008/12/trenitalia-preferisco-tornare-a-casa-a-piedi/>.

56 <http://www.teatridivita.it/italiano/qualita.html>.

57 <http://www.mad4games.it/forum/showthread.php?p=1682471>.

qui font ça méritent d'être désignés par un nom spécifique seulement dans un milieu – qui s'avère être restreint – où l'activité en question est suffisamment saillante. Le fait que *poltronisti* dans les sens (20a) et (20b) fasse référence à des significations du lexème POLTRONA qui sont partagées par la majorité des locuteurs, alors que le *poltronista* de (20d) fasse référence à une activité qui n'est saillante que pour une minorité d'entre eux est un cas fortuit qui n'a rien à voir avec la compétence morphologique. Une caractérisation de la morphologie de l'italien doit simplement nous dire qu'une construction morphologique (ici le suffixe *-ista*) est disponible pour construire des dérivés qui ont certaines propriétés sémantiques (par exemple 'celui qui effectue habituellement une activité dont X [la base] est un des arguments saillants'). Elle ne peut et ne doit pas nous dire si la base exprime une activité saillante pour la totalité ou pour un nombre réduit de locuteurs, ni si le dérivé en question est destiné à devenir la dénomination stable d'une classe d'individus dans une phase historique précise (par exemple les politiques *politronisti* dans la société italienne du XXI^e siècle) et donc à se lexicaliser, ou à rester une création occasionnelle dont l'usage peut difficilement sortir d'un milieu restreint (cf. aussi la notion de « *nameworthiness* » de Downing 1977).

Les exemples cités mettent en jeu la pragmatique en tant que connaissance partagée du contexte extralinguistique. L'interprétation sémantique d'un dérivé, cependant, peut dépendre de façon cruciale du contexte linguistique à proprement parler. Par exemple, en italien le lexème *perditempo* est enregistré dans les dictionnaires en tant que nom [+humain] qui signifie 'quelqu'un qui perd son temps, glandeur'. Cependant, on trouve plusieurs attestations de ce lexème comme [-humain] dans le sens de 'activité qui fait / permet de perdre son temps'. Curieusement, une recherche Google révèle que, dans la grande majorité des cas dans lesquels *perditempo* est utilisé avec ce sens, il a le lexème *passatempo* ('passe-temps' [-humain]) dans son environnement linguistique immédiat, comme dans les exemples suivants :

(21)

- a. In queste condizioni il dialogo non è neppure un *passatempo*, ma semplicemente un **perditempo**⁵⁸

⁵⁸ http://www.fiducia-tivoli.it/chiesa_tivoli/Il%20Vescovo/Catechesi/2009_catechesi/LECTIO%20DIVINA%20PER%20I%20GIOVANI%20DELLA%20I%20VICARIA.pdf.

‘Dans ces conditions le dialogue n’est même pas un passe-temps, mais simplement une perte de temps’

- b. Spesso per ridurre la stanchezza e lo stress un bel gioco online può aiutare, un passatempo o un **perditempo** che dir si voglia, aiuta...⁵⁹

‘Souvent pour diminuer la fatigue et le stress un beau jeu on-line peut aider, un passe-temps ou *perd-temps*, comme on veut, est utile...’

Il y a donc au moins trois facteurs qui peuvent déterminer la lecture sémantique d’un lexème construit, au-delà de l’instruction véhiculée par la règle de construction elle-même : le réseau lexical dans lequel le lexème construit s’insère, la pragmatique et le contexte linguistique immédiat.

J’ouvre une parenthèse : les facteurs cités, naturellement, ne touchent pas uniquement la sémantique, mais aussi les autres dimensions qui sont impliquées dans la construction de nouveaux lexèmes. Ainsi, un lexème peut être issu de la combinaison inattendue d’une base et d’un affixe parce que cela le fait rentrer dans une série lexicale plus large. C’est le cas, par exemple, des noms en *-eur* du français qui désignent des sportifs construits sur des noms et non pas sur des verbes (*footballeur, kilomètreur, sabreur...*), et qui rentrent dans une série avec les autres noms désignant des sportifs (*plongeur, skieur, coureur...*). De même, le contexte linguistique peut favoriser l’apparition de lexèmes construits, y compris d’occasionnalismes. L’apparition de *antiquestismo* et *quellismo* dans l’exemple qui suit ne s’explique que par le souci d’intégrer ces dérivés à un paradigme de lexèmes en *-ismo*, qui dans ce cas est strictement ‘local’ :

(22)

Italiani sarete sempre tacciati di razzismo, fascismo, nazismo, terrorismo, antisemitismo, antiquestismo e quellismo⁶⁰.

‘Italiens, vous serez toujours accusés de racisme, fascisme, nazisme, terrorisme, antisémitisme, anti-ceci-isme et cela-isme’

⁵⁹ <http://approvatepervoi.com/tag/social-games/>.

⁶⁰ <http://www.cataniaomnia.it/cronache/8307-ossessioni-ora-ce-anche-il-razzismo-gastronomico>.

2.7.4 L'illusion du lexique parfait

L'idée que l'on puisse avoir un lexique qui ne contient que les lexèmes qui présentent des idiosyncrasies, sémantiques ou autres, est donc une illusion. Halle (1973), par exemple, qui a été un des premiers à élaborer une théorie du lexique et des règles de construction de mots globale et articulée, propose que les mots construits, avant de rentrer dans le lexique, passent par un filtre qui ajoute les idiosyncrasies observées dans le lexique attesté. Cela présuppose que les règles de construction de mots formeraient des unités parfaitement compositionnelles. Cet idéal de la compositionnalité parfaite reflète une conception de la sémantique bien trop naïve. Pouvoir décrire exactement le sens compositionnel d'un lexème construit (et par conséquent le distinguer des sens idiosyncratiques, qui méritent d'être listés dans le lexique) signifierait pouvoir identifier avec certitude *le* sens de sa base. Or, comme l'a observé très lucidement Lyons (1977 : 550) :

Current attempts to formalize the semantic structure of language-systems and to generate all and only the possible interpretations of sentences are based upon the assumption that, not only the number of lexemes in any language, but also the number of senses associated with each lexeme, is finite and enumerable. Metaphor constitutes a very serious theoretical problem for any theory of semantics that is based on such assumptions.

Dans un monde idéal, il devrait être possible de prendre la liste des lexèmes simples d'une langue, la liste des affixes (ou des procédés morphologiques plus en général), les combiner et obtenir un lexique parfaitement compositionnel et transparent (une possibilité évoquée par exemple par Aronoff 1983 : 359, bien qu'il formule quelques réserves). Il est probable qu'une telle tentative, cependant, n'aboutirait à rien : que signifierait *poltronista* dans un tel lexique ? Et *perditempo*, *ardoisier*, *amphoris-sage* ? Cette illusion de pouvoir identifier *le* sens premier des lexèmes, et par conséquent *le* sens compositionnel parfait est le pendant sémantique de la chimère de pouvoir trouver *la* forme phonologique sous-jacente unique pour tous les lexèmes. La morphologie est la capacité à construire des lexèmes complexes, comme ceux cités ci-dessus, qui sont sémantiquement compatibles avec un nombre plus ou moins

grand de réalités. En d'autres termes, c'est la capacité à construire des lexèmes qui peuvent vouloir dire quelque chose. Le fait que ces lexèmes veuillent effectivement dire cette chose dans une certaine construction linguistique n'est pas du ressort de la morphologie ; elle dépend, comme on l'a vu, de facteurs qui ne sont plus purement 'grammaticaux', mais relèvent de l'articulation entre la langue et le monde réel. La possibilité d'effectuer un calcul exact du sens des lexèmes construits est surtout compromis par le fait que la polysémie est loin d'être un épiphénomène, et doit être considérée plutôt comme une propriété constitutive des mots, ou du moins de la plupart d'entre eux (cf. Cruse 1986 : 71-74 ; Chomsky 1993 : 23, ou les considérations qui sont à la base de la théorie de Pustejovsky 1991 : ch. 3, 47-54⁶¹). Comme l'ont montré plusieurs études sur la métaphore, cette partie d'indétermination dans la signification des mots est une des propriétés de la langue qui nous permettent de donner un nom à de nouvelles entités. C'est d'ailleurs une possibilité que les linguistes eux-mêmes exploitent abondamment : *tête*, (*catégorie*) *vide*, (*structure*) *profonde*, *atome (lexical)*, *arbre*⁶². De ce point de vue, les mots peuvent être considérés comme des unités au moins en partie programmées pour être polysémiques. Les glissements de sens qu'ils subissent peuvent être fortuits ou bien se faire selon des schémas systématiques. Cette double nature du changement sémantique est exactement parallèle à celui qu'on observe avec la morphologie constructionnelle : un lexème complexe peut être le seul représentant d'un type de construction (et donc s'approcher des lexèmes simples en tant qu'unité mémorisée dans le lexique) ou bien illustrer un schéma constructionnel systématique et productif. Le lexème *cadre* dans le sens de 'ensemble du personnel d'une entreprise investi d'une fonction de commandement de contrôle, de direction' (TLFi) relève probablement du premier type. En revanche, comme l'observe Roché (2009 : 74-75), le procédé qui consiste à nommer un objet selon la ressemblance qu'il a avec un objet d'une autre classe (*puce* 'insecte' → 'microprocesseur' ; *souris* 'mammifère rongeur' → 'instrument informatique'...) relève probablement d'un schéma systématique dans diverses langues. Un procédé encore plus clair, de ce point de vue, est celui qui consiste à créer un nom de couleur à

61 On citera également de l'hypothèse distributionnelle' en sémantique (basée sur l'idée de Harris 1954), selon laquelle la distance sémantique entre deux mots se mesure en fonction de leurs contextes d'apparition, et qui n'est pas sans rappeler l'hypothèse d'une morphologie émergente illustrée en 2.6.

62 La plupart de ces expressions sont prises (et traduites par moi) de l'index de Chomsky (1995).

partir du nom d'un objet qui a une couleur saillante (*rose, orange, saumon, crème...*). L'ensemble des noms de couleurs n'est clairement pas clos ; de manière intéressante, exactement comme on l'a observé pour le lexème construit *poltronista* ci-dessus, quelques-uns des noms de couleur qu'il est possible de repérer font appel à une connaissance du monde partagée par la majorité des locuteurs, alors que d'autres sont plutôt liés à la sphère d'expérience d'un nombre réduit d'entre eux, et difficilement pourront devenir la dénomination stable d'une couleur, bien que le procédé qui les a créés soit le même. Si on regarde sur le Web les unités qui peuvent être combinés en italien avec le mot *color(e)* on trouve des expressions aussi variées que *color Ferrari* (rouge), *color vampiro* (pâle) ou *color Avatar* (bleu comme les extraterrestres qui apparaissent dans le film du même nom), ainsi que des exemples, comme en (23) ci-dessous, dans lesquels le sens à attribuer au nom de la couleur (dans cet exemple, bleu clair comme le maillot de l'équipe de football de Naples) ne peut pas être déduit sans faire référence au contexte linguistique qui l'entoure (et plus généralement, en l'occurrence, à une photo qui accompagne la citation en question et au type de texte duquel elle est prise – le site Web d'un quotidien sportif – dans lequel *Napoli* doit plus souvent désigner l'équipe de football que la ville) :

(23)

Lavezzi si regala una Ferrari color Napoli. [...] Ieri [...] è andato a ritirare alla concessionaria Ferrari la sua California, modello interamente azzurro⁶³.

'Lavezzi s'offre une Ferrari de couleur Naples. [...] Hier [...] il est allé chercher au magasin Ferrari sa California, modèle entièrement bleu'.

Roché (2009 : 74-75) propose, très justement, de considérer les cas comme ceux de *puce* et *souris* parmi les procédés de construction de lexèmes, et crée à ce propos l'étiquette de « conversion sémantiques » (cf. aussi Grossmann & Rainer 2004 : 24-30).

2.7.5 Lexique et compétence morphologique

Les exemples considérés jusqu'ici montrent qu'une étude de la construction des

⁶³ <http://www.corrieredellosport.it/Notizie/Calcio/101173/Lavezzi+si+regala+una+Ferrari+color+Napoli+LA+FOTO>.

lexèmes ne peut pas faire abstraction du fait que celle-ci a le lexique comme point de départ et comme point d'arrivée. C'est là tout l'enjeu : trouver l'articulation exacte entre une activité créative abstraite, qui relève de la capacité linguistique des êtres humains, et une entité historiquement déterminée, composite, stratifiée, variable selon les locuteurs, les phases historiques, les situations comme le lexique. Une articulation qui est très bien résumée par Aronoff (2007 : 813-81) dans la citation mise en exergue de cette partie (cf. aussi Corbin 1997 ; Aronoff & Fudeman 2005 : 131-132). Un lexème lorsqu'il est créé, même par des moyens morphologiques productifs, n'est pas créé de façon abstraite, mais est inséré dans une construction linguistique concrète. Autrement dit, lorsque nous créons un néologisme (consciemment ou pas), nous ne créons jamais une unité dictionnaire ; nous créons un discours qui est constitué de structures syntaxiques (énoncés, syntagmes...), qui à leur tour sont constituées de mots. Dans la grande majorité des cas ces mots sont des mots que nous connaissions déjà et que nous avons déjà utilisés ; il arrive parfois que ce ne soit pas le cas et que nous créons de nouveaux lexèmes. Au contraire, la majorité des constructions syntaxiques que nous utilisons nous les créons sur le champ. Toutefois, il n'y a aucune raison de penser que l'activité qui préside à la création de ces lexèmes et de ces constructions syntaxiques soient qualitativement différentes. Je reviendrai sur ce point plus bas. Puisqu'un nouveau lexème est toujours créé simultanément à son usage réel dans une construction linguistique concrète, de ce fait il rentre dans un réseau avec le reste du lexique, et les relations qu'il établit peuvent être différentes à chaque fois qu'il est recréé dans des situations différentes, comme c'est sûrement le cas pour les différents *poltronista* de (20). L'objectif d'une théorie de la morphologie constructionnelle est de rendre compte de la manière dont celle-ci produit des unités qui sont susceptibles de rentrer dans cette multiplicité de relations. Dans les études sur le lexique la perte de transparence (formelle ou sémantique) d'une unité morphologiquement complexe est directement liée, pratiquement sans exception, à la variable temps : plus un lexème reste dans le lexique, plus il s'opacifie. De ce qui a été dit plus haut, cependant, découle que les deux paramètres doivent en réalité rester séparés : un lexème peut être non complètement transparent dès sa naissance (cf. aussi Carstairs-McCarthy 1992 : 47-51 ; Aronoff 2007), et il existe des constructions qui forment des lexèmes tellement non spécifiés sémantiquement qu'on peut difficile-

ment les définir comme transparents (par exemple la suffixation en *-ista* en italien, dont on a vu plusieurs exemples ci-dessus ; Fabrizio 2010 a proposé la même idée pour les verbes dénominaux convertis). Déconnecter la lexicalisation de la variable temps est une conséquence souhaitable du modèle proposé, étant donné que cette dernière est une dimension clairement externe au fonctionnement de la langue et en tout cas difficilement mesurable avec précision.

2.7.6 Qu'est-ce qu'il y a dans le lexique ?

Un problème qui a été évoqué sans le développer, et sur lequel il convient de se pencher, est celui de la mémorisation des unités lexicales. Il est clair que le lexique et l'ensemble des lexèmes d'une langue ne sont pas deux ensembles parfaitement superposables. Comme on vient d'observer, l'idée que tous les lexèmes seraient mémorisés alors que toutes les phrases seraient construites on-line est clairement fautive. Le lexique contient un nombre important d'unités qui ne sont pas des lexèmes, par exemple, toutes les phrases figées (cf. Bolinger 1974, cité dans Derwing & Skousen 1988 : 63 ; Di Sciullo & Williams 1987 : 14-15 ; Jackendoff 1997 : 109-110, ch. 7 ; 2002 : 167-176), et inversement beaucoup de lexèmes ne peuvent pas être considérés comme faisant partie du lexique. Plus haut j'ai cité plusieurs exemples de nouveaux lexèmes (ou même de nouveaux sens de lexèmes) construits on-line, qui n'ont pas pour vocation de se figer et qui ont pu être conservés seulement grâce au support technologique sur lequel ils ont été enregistrés (Internet). Gaeta & Ricca (2009 : 38-40) proposent une typologie des objets linguistiques basée sur les deux paramètres [\pm morphologique] et [\pm lexical], où la première dimension se réfère au fait d'être construit par la morphologie et la deuxième au fait d'être mémorisé dans le lexique. Le croisement des deux dimensions définit quatre catégories d'objets.

	+morphologique	-morphologique
+lexical	(a)	(b)
-lexical	(c)	(d)

Figure 1 : types d'objets linguistiques (inspiré de Gaeta & Ricca 2009)

Les cas (a) et (d) correspondent, respectivement, au cas canonique des lexèmes construits par la morphologie et qui appartiennent au lexique et aux unités non construites par la morphologie et qui ne sont pas mémorisées (typiquement des constructions syntaxiques). Les cas qui nous intéressent le plus, cependant, sont les cas (b) et (c), et en particulier ce dernier. Le cas (b), en effet, ([–morphologique] [+lexical]) correspond aux unités qui sont dans le lexique et qui ne sont pas construites par la morphologie, mais par d’autres moyens, par exemple syntaxiques⁶⁴ ; c’est le cas des locutions figées, comme *lune de miel*, *pomme de terre*, etc. L’identification d’une classe bien définie d’objets qui sont construits par la morphologie mais ne sont pas lexicaux ([+morphologique][–lexical]) constitue la vraie nouveauté de la proposition de Gaeta et Ricca. Pour illustrer ce cas, les deux auteurs citent un certain nombre d’exemples assez divers, comme des composés ou des dérivés contenant des unités syntaxiques (24a-b), des composés qu’on pourrait étiqueter d’‘occasionalismes’ (24c), des dérivés évaluatifs qui ont uniquement une valeur pragmatique (24d) ou des dérivés de noms propres (24e) :

(24)

- a. porta-rotolo delle striscie di carta che si usano per coprire la tavoletta del wc
‘porte-rouleau des rubans en papier qui servent à couvrir la cuvette du wc’
- b. marcia anti-moschea di Lodi⁶⁵
‘marche anti mosquée de Lodi’
- c. È abbastanza calda l’acquetta per il bagnetto?
‘Elle est assez chaude l’eau-DIM pour le bain-DIM ?’
- d. terzo posto acchiappa-Uefa
‘troisième place qui permet d’accéder à la coupe Uefa (lit. attrape-Uefa)’

64 A rigueur, cette catégorie devrait contenir aussi les lexèmes simples. Comme Gaeta e Ricca je ne m’intéresse ici qu’aux unités construites.

65 Déjà dans Montermini (2008b : 116).

e. *rappresaglia gheddafiana*
'représaille khadafienne'

C'est sûr que, à un certain niveau, toutes les unités citées par Gaeta e Ricca peuvent être considérées comme [-lexical], c'est-à-dire non mémorisées. Toutefois, quasiment chacune d'entre elles représente un cas différent, et a des raisons spécifiques pour ne pas être mémorisée. Comme le reconnaissent les deux auteurs eux-mêmes (Gaeta & Ricca 2009 : 65 n. 6), (24d) et (24e) ne sont peut-être pas mémorisés (du moins pour tous les locuteurs), mais sont des unités qui ont le format canonique des lexèmes. Si *GHEDDAFIANO*, en particulier, n'est pas mémorisé c'est certainement pour des raisons pragmatiques, mais c'est un lexème tout à fait régulier du point de vue de son mode de formation, de sa sémantique, et qui possède un paradigme que les locuteurs sont en mesure de reconstruire sans problèmes. Justement le suffixe *-iano* (sur lequel cf. la section 3.3) est très éclairant à cet égard : puisque virtuellement il peut se lier à toute base ayant un référent unique, il n'est pas rare qu'il soit attaché de façon occasionnelle à des noms propres de personne, de groupes, de communautés, etc., sans que ces dérivés dépassent la sphère immédiate du locuteur et de son entourage. Il se peut, par exemple, que les membres d'une même équipe sportive ou que les habitués d'une communauté virtuelle (site Web, forums, etc.) se définissent entre eux les *Xiani*, ou qu'un nom propre cité dans l'environnement linguistique immédiat fonctionne comme base d'un dérivé en *-iano* :

(25)

Io propongo a Beppe e tutti i seguaci della filosofia **beppiana** di andare in strada, fermare il primo che passa e dirgli 'Non sei un pirla, ma sei l'equivalente di un pirla'⁶⁶.
'Je propose à Beppe et à tous les adeptes de la philosophie *beppienne* d'aller dans la rue, arrêter le premier qui passe et de lui dire "Tu n'es pas un idiot, mais tu es l'équivalent d'un idiot".

Pensons aussi, par exemple, au cas des ethniques : il existe naturellement des ethniques qui sont lexicalisés pour tous les locuteurs, cependant, virtuellement le nom de chaque région, ville ou village peut avoir un ethnique dans une langue qui possède

⁶⁶ http://forum.corriere.it/il_campionato_dei_lettori/20-09-2007/chiosa_alla_domanda_di_beppe-923775.html.

des moyens de construire des ethniques à partir de noms géographiques. Dans la plupart des cas, l'usage de ces ethniques restera limité aux habitants du lieu en question ou aux personnes qui pour différentes raisons le fréquentent ; il n'empêche qu'il s'agit dans tous les cas de lexèmes absolument réguliers du point de vue formel et sémantique. Ces exemples montrent aussi que le trait [-lexical] doit être nuancé, selon la signification qu'on donne au mot 'lexique' (cf. 2.7.2 ci-dessus). Peu importe laquelle des conceptions du lexique on prend comme référence (les lexiques mentaux des locuteurs ou le lexique idéalisé) : ce qui est important c'est que, que l'on prenne le lexique dans l'un ou dans l'autre sens, l'articulation entre le trait [\pm morphologique] et le trait [\pm lexical], c'est-à-dire entre la morphologie productive et les unités mémorisées, est la même. Le cas de *acchiappa-Uefa* est similaire : ce lexème n'est vraisemblablement pas dans le lexique simplement parce qu'il désigne une réalité qui n'a pas beaucoup de chances de devenir un concept suffisamment saillant pour mériter une dénomination spécifique (n'est pas « *nameworthy* », selon les termes de Downing 1977), mais du point de vue de sa construction il ne se distingue guère des autres composés VN de l'italien. En fait, il fait partie du grand nombre de composés VN employés en tant qu'adjectifs, très fréquents dans le langage journalistique italien qui sont le plus souvent des occasionnalismes (cf. Ricca 2005 ; 2009 ; Gaeta & Ricca 2009 pour plus d'exemples). La situation est légèrement différente pour l'exemple (24a), qui est toujours un composé verbal, mais pas un VN. Je reviens sur ces cas ci-dessous.

Pour ce qui concerne l'exemple (24c), il s'agit effectivement d'un cas de dérivation évaluative dont la portée sémantique est, plutôt que sur le lexème de base, sur l'acte de langage lui-même, comme démontre le fait que le diminutif est répété deux fois dans l'exemple en question (cf. Dressler & Merlini Barbaresi 1994 ; Dressler 2009). De ce fait, la dérivation en question n'est pas à proprement parler 'lexicale', du moins du point de vue sémantique. Du point de vue phonologique, et même morphologique, les suffixes en question sont bien attachés à des lexèmes, mais leur portée sémantique est plutôt sur la phrase. Cela est démontré par le fait que lorsque dans ce type d'énoncés l'évaluation se fait par préfixation, un mode de construction de lexèmes beaucoup plus libre quant aux unités sur lesquelles il opère (cf. Montermini 2008b), l'attachement peut se faire aussi bien sur le lexème que sur une construction

syntactique plus large (cf. *en super retard* vs. *super en retard*, qui sont tous les deux attestés). Dans ces cas on est face à un décalage entre la structure formelle et la structure sémantique des dérivés. Du point de vue formel, le procédé de suffixation opère sur le lexème qui dans la phrase de base est le plus approprié à son application (dans l'exemple en question des noms, la catégorie à laquelle des diminutifs s'appliquent le plus facilement). Du point de vue sémantique elle porte sur la phrase, voire sur l'acte illocutoire lui-même. De ce fait, les lexèmes auxquels sont attachées les marques de la diminution dans ce type de constructions n'ont pas pour vocation à devenir des dénominations, et donc à rentrer dans le lexique⁶⁷.

Les cas (24a) et (24b), enfin, sont des constructions qui prennent comme bases⁶⁸ des structures syntaxiques. Les dérivés de ce genre ont été identifiés depuis longtemps (cf. Lieber 1992 : 12-14 ; Lieber & Scalise 2006, et les références citées dans ces travaux), même si les traitements qui ont été proposés sont différents. Clairement, la restriction selon laquelle la base de ces dérivées devrait être constituée d'une construction syntaxique lexicalisée (cf. Scalise 1994 : 208-209) est trop restrictive, comme le démontrent les exemples mêmes cités en (24a-b) (cf. aussi plusieurs exemples dans Gaeta 2006 ; Montermini 2008b : 116-119). Dans mes travaux précédents (Montermini 2008b : 113-118 ; 2010 : 89-91) j'ai soutenu que les structures qui apparaissent dans ces lexèmes complexes sont effectivement des structures construites par la syntaxe qui sont utilisées comme base par la morphologie. Ce genre de données est particulièrement gênant pour les modèles linguistiques parfaitement modulaires et sériels ou cycliques, dans lesquels toutes les opérations morphologiques ont lieu avant toutes les opérations syntaxiques (cf. Halle 1973 : 8 ; Kiparsky 1982 : 133 ; Selkirk 1982) et pour le principe de l'intégrité lexicale qui en découle (cf. Lieber & Scalise 2006 pour une discussion). De plus en plus de travaux, cependant, ont montré qu'il existe des constructions syntaxiques qui possèdent des propriétés typiques d'unités mémorisées, ce qui fait penser que la césure entre la construction de lexèmes et la construction de structures syntaxiques est moins nette que ce qu'on

67 Naturellement, cela n'implique pas que toutes les constructions évaluatives aient ces propriétés : un lexème issu d'une construction évaluative peut très bien constituer une dénomination et de ce fait entrer dans le lexique : *table* et *tablette* désignent bien deux référents différents.

68 Puisque je ne le considère pas un concept théorique, mais descriptif, je prends la liberté d'employer 'base' ici dans le sens large d'unité sur laquelle porte une construction morphologique (cf. Roché 2009 : 71).

pensait auparavant (cf. Sadler & Arnold 1994 ; Jackendoff 2002 : 178-182 ; 2008 : 27 ; Booij 2009b). En particulier, dans le cadre de la grammaire des constructions, à laquelle se réfèrent la plupart des auteurs cités, il n’y a pas de différence de nature entre les constructions morphologiques et les constructions syntaxiques, et donc il n’y a aucun obstacle à considérer que les premières peuvent être incluses dans les deuxièmes, et vice-versa (cf. Booij 2005 ; 2007 ; 2009a pour une présentation de la morphologie constructionnelle et Booij 2008 ; 2009b pour une application à des données semblables à celles dont il est question ici). Naturellement, cela présuppose que la morphologie et la syntaxe n’opèrent pas sériellement, mais simultanément. Il s’agit d’une position qui empiriquement se défend, au moins dans les cas dont il est question ici. Je suis raisonnablement sûr de n’avoir jamais entendu, avant la construction exemplifiée en (24b), le syntagme *moschea di Lodi*, mais cela ne m’a pas empêché de comprendre immédiatement sa signification et celle du lexème dérivé en entier.

Pour revenir à nos exemples de (24a-b), il n’y a pas de doute que dans ces cas on est face à des unités construites par la syntaxe. Du point de vue sémantique, on ne peut certainement pas soutenir que les phrases incluses dans le composé de (24a) et dans le dérivé de (24b) ne sont pas compositionnelles. De plus, comme je l’ai remarqué dans Montermini (2006a ; 2008b), ce qui est déterminant pour en faire des constructions syntaxiques est le fait que ces unités maintiennent leur capacité référentielle, alors qu’habituellement les bases des lexèmes construits n’ont pas de valeur référentielle autonome. Par exemple, elles peuvent servir d’antécédents dans une relation anaphorique :

(26)

Anche in Italia ben vengano le proposte come quella del Ministro Ferrero per un DDL anti **privatizzazione delle risorse idriche**, **che** sono un bene pubblico⁶⁹.

‘Même en Italie soient les bienvenues les propositions comme celle du ministre Ferrero pour un décret de loi anti privatisation des ressources idriques qui sont un bien public’.

Il est significatif que la même propriété s’observe avec certains dérivés (27a) ou certains composés (27b) (notamment des composés subordonnés, cf. Bisetto & Scalise

69 <http://www.sardegna.italiadeivalori.it/home/content/view/501/115/>.

2005 pour la classification des composés) qui sont apparemment construits sur des lexèmes, et non pas sur des unités syntaxiques (les exemples sont tirés de Montermini 2006a, cf. aussi Montermini 2008b pour des exemples dans d'autres langues)⁷⁰ :

(27)

a. L'Unione Sovietica e l'Inghilterra [...] porterebbero avanti un'azione anti-**Opec**, per indebolire **questa organizzazione**. [*la Repubblica*, 7 mai 1985]

'L'Union Soviétique et l'Angleterre mèneraient une action anti-Opec, pour affaiblir cette organisation'

b. Era disponibile a essere **capogruppo**, di **quello**, da noi appena costituito.

'Il était disponible pour être le chef du groupe que nous venions de constituer'

Dans des travaux précédents (Montermini 2006a ; 2010) j'ai utilisé cet argument, conjointement à d'autres⁷¹, pour soutenir que l'élément qu'on retrouve dans tous ces lexèmes complexes, affixés ou composés, est une construction syntaxique, et notamment que la position en question peut être occupée par un syntagme nominal sans déterminant⁷². En d'autres termes, il n'y aurait pas de différence de nature entre le composé *raccolta rifiuti* ('ramassage des ordures') et *raccolta rifiuti di provenienza agricola* ('ramassage des ordures d'origine agricole'), ainsi qu'il n'y a pas de différence entre *i rifiuti* et *i rifiuti di provenienza agricola*, si ce n'est la présence / absence d'un complément. L'élément nominal dans les composés VN (qui comme le montre l'exemple (24a) peut aussi être un syntagme sans déterminant) présente exactement les mêmes caractéristiques, comme le montrent les exemples suivants :

70 Le débat sur l'insularité anaphorique (« anaphoric islandhood ») des lexèmes construits, lancé par l'article de Postal (1969) a donné lieu à un grand nombre de travaux qui ont montré que la question ne peut pas être réduite à des conditions syntaxiques purement locales, comme c'était le cas dans les premiers traitements, réalisés dans le cadre de la grammaire générative (cf. Montermini 2006a pour un historique). Je ne prends en considération que des exemples de lexèmes composés ou préfixés ; le phénomène en question, cependant, se réalise plus rarement aussi avec des lexèmes suffixés (cf. Montermini 2006a).

71 Par exemple, le fait que la lecture spécifique / générique admise au pluriel est la même avec ces lexèmes complexes et des syntagmes (cf. Montermini 2008b : 111-113),

72 On peut faire l'hypothèse que, en ce qui concerne les lexèmes affixés de ce genre, on est face à des affixes de phrase, tels ceux qu'Anderson (2005) appelle « clitiques spéciaux » (par exemple le 's génitif anglais.

(28)

a. collanina ‘trasgressiva’ acchiappatalenti_i, tra i quali_i, ha figurato anche il primo Ammanniti [Ricca (2005 : 479)]

‘petite série ‘transgressive’ attrape-talents parmi lesquels a figuré aussi le premier Ammanniti’

b. porta biberon_i, termico che ne_i, mantiene la temperatura⁷³

‘porte-biberon thermique qui en maintient la température’

Que l’élément nominal des VN soit une construction syntaxique n’a rien d’étonnant, puisqu’il s’agit d’un des procédés morphologiques des langues romanes qui présentent le plus de traits communs avec la syntaxe (y compris du point de vue étymologique, cf. Floricic 2008). Naturellement, cela vaut en particulier pour les VN construits productivement : l’élément nominal de composés comme *marciapiede* (‘marcher+pied, trottoir’), *portafoglio* (‘portefeuille’) ou *guastafeste* (‘gâcher+fêtes, rabat-joie’) peut difficilement être considéré comme ayant une référence propre. Mais justement ce positionnement incertain entre syntaxe et morphologie rend possible le grand nombre d’occasionnalismes qu’on retrouve avec ce type de constructions (cf. Ricca 2005 ; 2010 pour plusieurs exemples). S’il est vrai qu’il s’agit d’unités construites en partie par la morphologie et en partie par la syntaxe, cela explique pourquoi elles sont moins facilement lexicalisées que d’autres unités morphologiquement complexes.

Les exemples fournis par Gaeta et Ricca ne sont pas exhaustifs de toutes les unités qui peuvent être construites mais ne pas être mémorisées. Les travaux spécialisés font état d’autres cas où l’interprétation d’unités données dépend entièrement du contexte – linguistique et extralinguistique – et qui n’ont aucune raison d’exister en dehors de celui-ci. Un exemple est celui de *beppiano*, cité en (25) ; d’autres exemples sont ceux bien connus des composés anglais comme *apple juice seat* (Downing 1977) qui peut désigner un siège qui se trouve à côté du jus de pomme dans une situation extralinguistique particulière, ou *London bus*, dont l’interprétation peut varier selon la phrase dans laquelle il est inséré (*Has the London bus left / arrived yet?*, exemple cité par Lyons 1977 :

73 <http://www.tuttomamma.com/hello-kitty-linea-prenatal/5736/>.

538-539).

Les procédés de composition peuvent relever de schémas grammaticalisés dans une langue individuelle, mais peuvent aussi relever d'une capacité cognitive plus générale chez les humains à relier deux lexèmes sémantiquement par une simple juxtaposition (cf. 2.9.2). Il existe d'autres procédés qui font appel non pas à des processus grammaticalisés, mais à des capacités cognitives plus générales ou à des contraintes de bonne formation universelles ; pour cette raison j'applique à ces constructions l'étiquette d'extragrammaticales'. Des cas de ce genre existent probablement dans toutes les langues. Des exemples en français sont les raccourcissements, les reduplications et en général les langages secrets comme le verlan ou autres (cf. Fradin *et al.* 2009 : 28-32 ; Plénat 2009b : 130-131). Tous ces procédés donnent lieu à des unités qui, pour des raisons diverses, peuvent se lexicaliser (par exemple *beau-frère* → *beauf*, *guerre* → *guéguerre* ou *pourri* → *ripou*), mais qui n'ont pas de vocation de le faire et en général ne le font pas (cf. Fradin 2003 : 206-218).

Pour résumer, la classification de Ricca & Gaeta (2009) capte bien le fait que la morphologie peut construire des objets qui ne sont pas mémorisés ou ne sont pas faits pour l'être. Nous avons vu que les raisons pour cette inadéquation à la lexicalisation peuvent être diverses, et en particulier on a distingué :

- des unités qui ont les propriétés de compositionnalité, de capacité de référence et d'ancrage au contexte typiques des constructions syntaxiques, comme *anti-moschea di Lodi* (24b) ;
- des unités qui sont complexes uniquement du point de vue de la forme, mais qui ont plutôt une fonction pragmatique, comme *acquetta* ou *bagnetto* (24c) ;
- des unités qui ne sont pas lexicalisées pour des raisons pragmatiques, tout en ayant les propriétés typiques des lexèmes qui normalement sont lexicalisés, par exemples parce qu'elles ne désignent pas une entité suffisamment saillante pour mériter une dénomination, comme *acchiappa-Uefa* (24d) ou *gheddafiano* (24e) ;
- des unités dont l'interprétation dépend entièrement du contexte, comme *beppiano*, cité en (25), ou *apple juive seat* ;
- des unités construites par des moyens extragrammaticaux, comme *beauf*, *guéguerre* ou *ripou*.

graduelle, il peut y avoir des différents degrés de prototypicité pour un lexème. Les unités exemplifiées en (24) et tout au long de cette section (et reprises dans la Figure 2) sont des lexèmes à différents degrés : syntaxiquement, ils ont tous la même distribution des lexèmes typiques (noms, verbes ou adjectifs) ; ils ont une capacité référentielle, mais tous ne sont pas des dénominations (au sens indiqué en 2.2.2) ; certains d'entre eux, mais pas tous, ont un paradigme flexionnel qui est identique à celui des autres lexèmes de la même catégorie⁷⁴.

La Figure 3 permet aussi de délimiter une portion plus homogène que les autres, celle des unités [+lexical]/[+morphologique], qui sont à la fois des lexèmes et des listèmes⁷⁵. Je considère que ces unités sont prototypiquement les unités basiques de la langue, c'est-à-dire les lexèmes typiques. Il est probablement vrai que le nombre de constructions syntaxiques mémorisées est très grand et a longtemps été sous-estimé (cf. Jackendoff 1997 : 154-157), mais cela ne signifie pas que les lexèmes et les constructions syntaxiques sont équivalents quant à leurs possibilités d'être mémorisés. En réalité, j'estime que dans les langues les lexèmes sont des unités qui tendent plutôt à être mémorisées et que les unités mémorisées tendent plutôt à être des lexèmes (cf. Bauer 2001 : 122). Cela va à l'encontre des positions quelque peu relativistes exprimées par exemple par Jackendoff, selon qui l'idée que les unités mémorisées sont surtout des lexèmes serait « a widespread stereotype in linguistic theory (and philosophy and the popular conception of language) » (Jackendoff 2002 : 153). Au contraire, les unités qui sont mémorisées et qui ne sont pas des lexèmes tendent souvent à devenir plus synthétiques, et donc plus semblables à un mot, comme on peut voir, par exemple, dans la prononciation avec liaison des syntagmes figées du français *pied-à-terre* ou *pot-au-feu*. Gaeta (2006 : 75-76) propose aussi plusieurs exemples de lexèmes dérivés par suffixation d'expressions idiomatiques, dans lesquels n'est retenu que le matériel lexical, alors que le matériel grammatical (déterminants, prépositions, etc.) est éliminé : *cerchiobottista* (← *dare un colpo*) *al cerchio e uno alla botte* 'partager tous les avis, ne pas prendre position'), *saccopelista* (← *sacco a pelo* 'sac de couchage'), et des exemples semblables existent en français, surtout comportant des eth-

74 A cela il faut ajouter que certains d'entre eux (*gheddafiano*, *acquetta*...) constituent un seul mot phonologique, d'autres (*anti-moschea di Lodi*) non, ce qui certainement contribue à rendre les premiers encore plus typiques comme lexèmes.

75 Naturellement, les lexèmes simples sont des unités lexicales encore plus prototypiques.

niques (*francilien* ← *Île de France*). Plutôt, le fait que la mémorisation ne soit pas un critère pour distinguer les lexèmes des constructions syntaxiques est une invitation à considérer la frontière entre la morphologie et la syntaxe comme moins étanche. En effet, récemment plusieurs travaux, en particulier en s’inspirant à la grammaire des constructions, ont proposé d’analyser de manière similaire des phénomènes morphologiques et des phénomènes que les morphologues auraient eu tendance à ignorer il y a quelques années en raison du fait qu’ils étaient « construits par la syntaxe », comme les constructions du français N-Prép-N (*moulin à poivre*, *moteur à essence*, cf. Booij 2009b : 221-222), les constructions comportant une conjonction, comme en italien *alti e bassi* (‘des hauts et des bas’, Masini 2006) ou *usa e getta* (‘utilise et jette, jetable’, Masini & Thornton 2008), etc. On pourrait faire l’hypothèse, alors, que la distinction entre les règles de la morphologie et celles de la syntaxe a son fondement dans la différence de nature entre les unités qu’elles manipulent, construisent et des données desquelles elles émergent, typiquement mémorisées les unes, typiquement non mémorisées les autres (sur la distinction entre morphologie et syntaxe cf. aussi Haspelmath 2007 ; 2009).

2.7.8 Mots possibles et mots existants

On pourrait se demander si le schéma proposé à la Figure 3 ne permet pas de délimiter un espace supplémentaire, notamment celui des unités qui sont [-lexical], mais [+morphologique], c’est-à-dire l’ensemble des lexèmes construits par la morphologie mais qui ne sont pas lexicalisés. Plusieurs auteurs se sont effectivement penchés sur l’existence d’un tel « lexique potentiel »⁷⁶ et sur l’existence de « mots possibles mais non existants ». Je n’utiliserai ni l’une ni l’autre des deux expressions. D’une part, je considère que l’ensemble (potentiellement infini) des lexèmes construits par la morphologie n’est pas un lexique. Ce mot, comme je l’ai déjà observé, renvoie en effet à une entité qui est à l’interface entre la compétence linguistique et la connaissance de la réalité, et qui est socialement et historiquement déterminée. Puisque potentiellement tout ce qui est linguistique peut être lexicalisé (cf. Jackendoff 2002 : 152-153),

76 C’est l’étiquette la plus employée en anglais (par exemple par Aronoff 1994 ; 2000). En français on parle plus volontiers de « lexique virtuel » (cf. Corbin 1997). Zwicky (1992 : 332) créé pour un tel ensemble l’étiquette de « morphicon » (calqué sur l’anglais « lexicon »).

parler de lexique potentiel perd de toute signification. D'autre part, les adjectifs « possible » et « existant » renvoient à deux dimensions qui restent séparées et incommensurables. Le premier se réfère à la capacité d'exploiter la compétence linguistique (en l'occurrence morphologique) pour générer de nouvelles unités. C'est donc un concept purement linguistique. Le deuxième se réfère à une propriété qui est étrangère à la compétence linguistique au sens strict. Puisque les limitations techniques nous empêchent (et vraisemblablement nous empêcheront toujours) d'avoir accès et de lister tous les mots qui sont et ont été produits dans tous les échanges effectués en une langue déterminée, l'existence (ou la non existence) d'un lexème ne peut-être calculée qu'à l'intérieur d'un système de référence donné (et donc il est plus prudent, en tout cas, de parler de mot « attesté »). Dans les études de morphologie ce système de référence est pour la plupart constitué soit de l'ensemble des mots inclus dans un ou plusieurs dictionnaires ou dans un corpus, soit – regrettablement – de ceux que les chercheurs 'sentaient' comme de bons mots de leur langue. Naturellement, effectuer des études et des calculs sur un système de référence clos et déterminé, et considérer qu'il peut être plus ou moins représentatif du lexique idéalisé d'une langue n'est pas en soi une erreur. L'erreur est le raccourci conceptuel qui consiste à considérer que ce système coïncide avec *le* lexique, c'est-à-dire avec l'ensemble des lexèmes existants d'une langue. Les critères qui permettent de décider si un mot 'existe' ou pas ne sont d'ailleurs pas clairs. Quelques auteurs identifient différents étapes dans la naissance et la vie d'un mot. Par exemple, Hohenhaus (2005) distingue plusieurs types de mots : occasionnalismes (« nonce formations »), néologismes, mots institutionnalisés, mots lexicalisés. Le problème de ces classifications est qu'elles restent toujours quelque peu arbitraires. Il convient peut-être renoncer du tout à utiliser des concepts comme 'mot existant' ou 'mot attesté' comme des concepts théoriquement utiles, et se limiter à établir si, et à quelles conditions, un mot est possible.

2.8 Une morphologie thématique

2.8.1 La morphologie thématique : caractérisation

Comme je l'ai rappelé à plusieurs reprises, ce travail s'intéresse principalement à la

rubrique (F) des lexèmes, c'est-à-dire à leur représentation phonologique et aux relations qui s'instaurent entre les différentes formes qu'un lexème peut prendre (en flexion) et entre les forme de différents lexèmes (en dérivation). En particulier, il s'inscrit dans le modèle de morphologie qui a été étiqueté, notamment en France, de 'thématique' (cf. Plénat 2008 ; 2009a). Il s'agit d'un modèle qui s'est développé partiellement indépendamment à partir de travaux comme ceux de Bonami & Boyé (2003 ; 2005 ; 2007) (cf. aussi Bonami *et al.* 2009), ou de Pirrelli (2000) ; Pirrelli & Battista (2000), mais qui avait déjà été anticipé par une série de travaux, aussi bien dans la morphologie « mots et paradigmes » que dans le courant générativiste (Matthews 1981 ; Morin 1987 ; Plénat 1987 ; Perlmutter 1988 ; Maiden 1992 ; Aronoff 1994, entre autres). Il s'agit d'une approche qui découle tout naturellement du modèle de morphologie qui est défendu ici. En particulier, pour la morphologie dite thématique, que par exemple les formes fléchies d'un lexème puissent être reconduite à une forme unique sous-jacente n'est pas une priorité, ni même une propriété définitoire des lexèmes réguliers par opposition aux lexèmes irréguliers. L'attention est plutôt tournée vers la forme globale du paradigme que vers les relations locales entre les formes. Comme on le voit, cette vision est particulièrement compatible d'une part avec une vision du lexique enrichie, dans laquelle avoir une seule représentation phonologique ne constitue pas, même idéalement, un avantage, et d'autre part avec une morphologie non symbolique et non incrémentale, dans laquelle les significations ne doivent pas être nécessairement associées à un morphème, et vice-versa.

2.8.2 Allomorphie et supplétion

La préoccupation principale de la morphologie thématique est de rendre compte de la variation allomorphique que subissent les lexèmes, que ce soit en flexion ou en dérivation. Naturellement, il existe une multitude de travaux sur l'allomorphie, qui tentent de systématiser ce concept et d'établir une distinction entre celle-ci et la supplétion. De manière souvent implicite, l'allomorphie est généralement considérée comme un phénomène qui a une certaine connexion avec la phonologie (par exemple parce qu'elle en subit le conditionnement, cf. Kiparsky 1996 ; Paster 2006 ; Nevins à p.), alors que la deuxième est simplement synchroniquement opaque (Boyé 2006 :

297). Dans ce travail je considérerai que, la supplétion étant un concept gradable (Dressler 1985c : 98-99 ; Mel'čuk 2006 : 418-419), il n'y a pas de distinction de substance entre ces deux notions, et que le passage de l'allomorphie typique à la supplétion typique se fait de manière graduelle, en particulier en prenant en compte deux paramètres : (i) l'alternance correspond ou pas à une opération phonologique unique et (ii) l'alternance concerne un seul ou plusieurs lexèmes (cf. aussi Kiparsky 1996 : 17 et, en particulier pour (ii), Mel'čuk 2006 : 419). La Figure 4 illustre la situation ; des exemples de chaque type, pris au français, sont donnés en (29) ; les cas (a) et (d) illustrent, respectivement, l'allomorphie typique et la supplétion typique :

	opération phonologique	
	+	-
plusieurs lexèmes	(a)	(b)
un lexème	(c)	(d)

Figure 4 : allomorphie vs. supplétion

- (29) a. lent – lente
b. cheval – chevaux
c. sec – sèche
d. allons – va

La classification de la Figure 4 essaie de systématiser les paramètres, parfois invoqués pour rendre compte de la ressemblance (ou de la différence) entre deux formes (cf. par exemple Carstairs 1988 et le concept de supplétivisme fort vs. supplétivisme faible, cf. Dressler 1985c). Cependant, la ressemblance ou la différence entre deux formes phonologiques est un concept difficilement formalisable et calculable, et surtout, il n'est pas clair en quoi elle pourrait servir de critère pour placer la frontière entre deux notions différentes dans la substance. En effet, la classification proposée n'a pas pour but de distinguer des objets de nature différente, mais de rendre compte du passage non discret entre des formes tendanciuellement transparentes et prédictibles (les cas (a), qui comprend aussi les allomorphies entièrement phonologiques,

exemplifiées par l’alternance [e] ~ [ɛ] dans *sécher* – *sèche*) et des formes maximale-
 imprédictibles (les cas (d)). C’est pourquoi dans ce qui suit le terme ‘allomorphie’ sera
 utilisé pour parler de tous les phénomènes inclus dans la Figure 4. Il faut cependant
 remarquer que les paramètres de la transparence et de la prédictibilité coupent trans-
 versalement le classement proposé. La transparence, d’une part, dépend de plusieurs
 facteurs, et ne peut pas être calculée dans l’absolu, mais seulement en relation aux
 couples de formes qui sont en relation entre elles ; on peut difficilement considérer,
 par exemple, que la relation entre *maux* et *mal* soit aussi transparente que celle entre
cardinaux et *cardinal*, bien que les deux couples exemplifient la même alternance. Une
 mesure purement arithmétique, par exemple la distance de Levenshtein, ne peut pas
 tout nous dire. Si nous prenons, par exemple, deux lexèmes comme LENT ou
 IMPORTANT, qui exemplifient tous les deux l’alternance $\emptyset \sim t$, la distance, ainsi calcu-
 lée, entre leur forme du masculin et celle du féminin est toujours de 1. Cependant, si
 on calcule le nombre total de phonèmes que les formes contiennent respectivement,
 cette distance est de 1/3 (33,3%) pour LENT et seulement de 1/7 (14,2%) pour
 IMPORTANT. Sur la base de ces observations nous pouvons établir, par exemple, une
 échelle de transparence des formes suivantes (j’indique l’alternance mise en jeu par
 chacun des couples) :

(30) lent – lente	$\emptyset \sim t$
suspect – suspecte	$\emptyset \sim kt$
fou – folle	$u \sim \text{ɔl}$
directeur – directrice	$\text{œr} \sim \text{ris}$
allons – va	$\text{al}\text{ɔ} \sim \text{va}$

Par ailleurs, transparence et prédictibilité ne sont pas deux concepts parfaitement su-
 perposables. Comme l’observe Boyé (2006 : 297), en morphologie flexionnelle l’iden-
 tité aussi (qui est pourtant la relation la plus transparente qui soit) peut paradoxa-
 lement être considérée comme une supplétion, dans les cas où il n’y a pas d’alternance
 là où on en attendrait une (cf. le verbe anglais *cut*_{PRES} – *cut*_{PRET} ‘couper’ vs. *love*_{PRES} – *loved*_{PRET}
 ‘aimer’). La prédictibilité dépend en effet aussi de façon cruciale du nombre de
 formes qui exemplifient une certaine alternance, et elle est donc strictement liée au

paramètre (ii) cité ci-dessus. Prenons par exemple (29a) ci-dessus : à un masculin en [ã] peut correspondre un féminin en [ãt], comme pour LENT, mais aussi un féminin en [ãd] (GRAND), en [ãʃ] (BLANC) ou en [an] (OCCITAN) (cf. le Tableau 8 ci-dessus) ; un féminin en [an], cependant, peut avoir aussi un masculin homophone (PROFANE). En (31) sont indiqués les différents cas avec le nombre de lemmes qui illustrent chacun d'eux dans la base lexique.org :

(31)	alternance	exemple	lemmes
	[ã] ~ [ãt]	lent – lente	1 141
	[ã] ~ [ãd]	grand – grande	14
	[ã] ~ [ãʃ]	blanc – blanche	2
	[ã] ~ [an]	occitan – occitane	53
	[an] ~ [an]	profane – profane	21

Le croisement des deux paramètres relatifs à la transparence et à la prédictibilité de la relation nous donne une autre classification des cas de variation allomorphique, qui coïncide seulement partiellement avec celle ci-dessus :

			prédictible	
			+	-
	transparent	+	(a)	(b)
		-	(c)	(d)

Figure 5 : types d'allomorphies transparentes et prédictibles

- (32) a. lent – lente
 b. suspect – suspecte
 c. directeur – directrice
 d. allons – va

L'intérêt principal des classements proposés dans les Figures 4 et 5 est qu'ils nous permettent d'identifier clairement deux pôles, correspondant respectivement à la va-

riation allomorphique entièrement transparente et prédictible et à la supplétion complètement non transparente et non prédictible. Ils ne nous disent rien, cependant, sur la nature, phonologique ou morphologique, de la variation, qui constitue le sujet de la section suivante.

2.8.3 Allomorphie phonologique et morphologique

L'émergence de l'allomorphie est clairement liée à la phonologie. La grande majorité des travaux de phonologie, aussi bien dans la phonologie bâtie sur des règles à la SPE que dans le cadre de OT, a essayé de rendre compte par des moyens phonologiques de la variation de forme des mots. Cependant, la variation allomorphique ne peut clairement pas être réduite à l'effet du fonctionnement des règles phonologiques. Les alternances exemplifiées en (30), par exemple, ne peuvent pas être interprétées comme des alternances purement phonologiques, puisque (i) elles n'ont pas le caractère automatique et entièrement prédictible de celles-ci ; (ii) leur conditionnement n'est pas phonologique, dans la mesure où l'alternance ne sert pas, par exemple, à résoudre une séquence phonotactiquement inacceptable. Dans les travaux de phonologie, il est en effet commun de distinguer au moins trois types d'alternances :

- i. des alternances purement phonologiques, dans lesquelles se manifeste simplement l'action des règles phonologiques à l'œuvre dans la langue ; par exemple, l'alternance [s] ~ [z] pour le suffixe de pluriel anglais (*back[s]* 'dos' vs. *bag[z]* 'sacs'), ou l'alternance [e] ~ [ɛ], rappelée ci-dessus, dans le paradigme des verbes français comme SÉCHER ; dans ces cas aussi bien le conditionnement (le trait [\pm sonore] de la consonne qui précède le suffixe de pluriel ou le type de syllabe – ouverte vs. fermée – dans laquelle apparaît la voyelle) que l'effet sont purement phonologiques ;
- ii. des alternances purement morphologiques (qui dans ce cas est synonyme de lexical), comme le pluriel anglais *ox* – *oxen* ('bœuf(s)') ou l'alternance *allons* – *va* en français ;
- iii. des alternances de type morphologique (c'est-à-dire, non prédictibles par une règle

phonologique), mais dont le conditionnement est purement phonologique (cf. Kiparsky 1996 ; Paster 2006 ; Nevins à p., entre autres). Paster (2006) fournit plusieurs exemples très clairs de ce type de phénomène ; par exemple, en dia:bugay (Paster 2006 : 1) la forme du suffixe de génitif ([n] ou [ɲun]) varie selon que le phonème immédiatement précédent est une voyelle ou une consonne :

- (33) dia:bugay (langue pama-nyungan, Australie)
- | | |
|-------------------------------------|--------------------------------------|
| guludu-n ('colomb _{GEN} ') | gənal-ɲun ('varan _{GEN} ') |
| djama-n ('serpent _{GEN} ') | bibuy-ɲun ('enfant _{GEN} ') |

En réalité, comme c'est fréquemment le cas, alors que les cas extrêmes ((i) et (ii)) sont assez clairs, les cas intermédiaires sont plus difficiles à classer. Kiparsky (1996) propose qu'à côté de la variation morphologique phonologiquement conditionnée (le cas (iii)), il existe une variation qui est phonologique, mais qui est morphologiquement conditionnée. Cela nous donne une classification quadripartite basée sur les deux paramètres du conditionnement et de l'effet, qui peuvent tous les deux être phonologique ou morphologique (lexical) :

		conditionnement	
		phonologique	morphologique
effet	phonologique	(a)	(b)
	morphologique	(c)	(d)

Figure 6 : types de conditionnements et d'effets de l'allomorphie

(a) et (d) correspondent respectivement aux cas (i) et (ii) cités ci-dessus. (c) correspond au cas (iii). Le français nous fournit aussi plusieurs exemples de ce cas, en particulier en ce qui concerne la morphologie dérivationnelle. La plupart des suffixes dénominaux et désadjectivaux du français commencent par voyelle. L'ajout d'un tel suffixe à une base se terminant par voyelle crée un contexte de hiatus, défavorisé dans les langues du monde, qui incluent souvent une ou plusieurs contraintes anti-hiatus.

Cette contrainte anti-hiatus peut entrer en concurrence avec d'autres contraintes⁷⁷ ; ce qu'il est important de retenir ici c'est que le cas le plus fréquent est celui dans lequel le hiatus est évité par l'emploi d'une forme du nom ou de l'adjectif qui se termine par consonne, sans qu'elle ne puisse être dérivée phonologiquement de l'autre ou d'une forme sous-jacente commune (pour un traitement des alternances $\emptyset \sim C$ en français cf. en particulier Tranel 1981 : 222-276 ; Bonami & Boyé 2005, et les références qui y sont citées) :

- | | | |
|------|-------------------------------------|---------------------|
| (34) | roux / rousse ([ru] / [rus]) | rousseur ([rusœr]) |
| | jaloux / jalouse ([ʒalu] / [ʒaluz]) | jalousie ([ʒaluzi]) |
| | fou / folle ([fu] / [fɔl]) | folie ([fɔli]) |

Dans le cas des bases se terminant par une voyelle nasale, la contrainte phonologique, au lieu d'être une préférence, est catégorique : les séquences $\tilde{V}V$ sont absolument exclues en français. Si nous reprenons les exemples de (31), nous retrouvons en dérivation toutes les alternances représentées en flexion :

- | | | |
|------|--|------------------------------|
| (35) | lent / lente ([lɑ̃] / [lɑ̃t]) | lenteur ([lɑ̃tœr]) |
| | grand / grande ([grɑ̃] / [grɑ̃d]) | grandeur ([grɑ̃dœr]) |
| | blanc / blanche ([blɑ̃] / [blɑ̃ʃ]) | blancheur ([blɑ̃ʃœr]) |
| | occitan / occitane ([ɔksitɑ̃] / [ɔksitan]) | occitaniste ([ɔksitanist]) |
| | profane / profane ([prɔfɑ̃] / [prɔfɑ̃]) | profanation ([prɔfɑ̃nasjɔ̃]) |

Aucune des alternances exemplifiées ci-dessus n'est entièrement phonologique (ni même la (dé)nasalisation qu'on observe dans *occitan* – *occitane*, cf. Tranel 1981 pour une discussion du traitement génératif classique) : comme on l'a vu plus haut, aucune des alternances en question n'étant complètement régulière, elles doivent toutes être mémorisées dans le lexique.

Pour ce qui est du cas (b), je soupçonne que le plus souvent les cas de ce type sont

⁷⁷ Par exemple des contraintes prosodiques relatives à la taille de la base dans les dérivés : un dérivé optimal est celui dans lequel la séquence qui précède le suffixe est dissyllabique, ce qui peut provoquer la chute de la voyelle finale d'une base plus que dissyllabique (*Poincaré* → *poincarisme*) ou rendre le hiatus acceptable pour préserver la taille dissyllabique de la base (*Kafka* → *kafkaïen*) (cf. Plénat 2009a ; 2009b).

considérés par les chercheurs comme étant conditionnés phonologiquement. Pour prendre un exemple simple de ce type de variations, on peut citer les verbes du français dérivés par conversion de noms qui se terminent par voyelle (en particulier [o], [a] ou [ø]⁷⁸), dont je donne quelques exemples en (36) :

(36) numéro	→	numéroter / (il) numérote ([nymerote] / [nymerot])
réseau	→	réseauter / (il) réseaute ([rezote] / [rezot])
Stabilo	→	stabiloter / (il) stabilote ([stabilote] / [stabilot]) ⁷⁹
blabla	→	blablater / (il) blablate ([blablate] / [blablat])
mafia	→	mafiater / (il) mafiate ([mafjate] / [mafjat]) ⁸⁰
sépia	→	sépiater / (il) sépiate ([sepjate] / [sepjat]) ⁸¹
bleu	→	bleuter / (il) bleute ([bløte] / [bløt])
milieu	→	milieuter / (il) milieute ([miljøte] / [miljøt]) ⁸²
queue	→	queu(e)ter / (il) queute ([køte] / [køt])

Dans tous les cas exemplifiés, le verbe est dérivé via l'adjonction d'une consonne [t] (qui est la consonne épenthétique par défaut en français). Cette adjonction n'a pas de vraie justification phonologique : aussi bien à l'infinitif qu'au présent de l'indicatif (qu'à toutes les autres formes du paradigme) la simple juxtaposition des suffixes verbaux ne donnerait pas de séquences phonologiquement impossibles en français. En particulier, au présent de l'indicatif (représenté en (36) par la 3^e personne du singulier), les voyelles [o], [a] et [ø] sont tout à fait courantes en français en finale de mot. Cependant, il n'existe aucun verbe du français dont le thème se termine en une voyelle pleine [o] ou [a], et par conséquent dont les formes du singulier du présent de l'indicatif auraient cette voyelle comme finale absolue. Il existe des verbes dont les formes du présent de l'indicatif peuvent se terminer par [ø] (*veux, peux*), mais aucun

78 Parfois [u] (cf. *chouchou* → *chouchouter*), mais les verbes en *-ouer* sont possibles, et il existe même des doublons (*clou* → *clouer* / *clouter*).

79 Dans les transcriptions je suis la tradition prescriptive qui prévoit [ø] lorsque la graphie est *o* et [o] lorsque la graphie est *(e)au*.

80 Cité dans Morin (1972 : 104).

81 « perso je **sépiate** un peu, ou je retouche au plus soft »
[http://www.zphoto.fr/venelle_photo21146.html].

82 « un blog qui débute au début, qui finit à la fin, et qui **milieute** au milieu... »
[<http://bertran.blogtrafic.com/>].

d'eux n'appartient à la première classe (infinitif en *-er*). Le conditionnement est donc morphologique (l'impossibilité absolue d'avoir des formes comme **il* [rezo] ou **il* [blabla], et des formes comme **il* [blø] pour un verbe de la première classe), mais l'effet est phonologique (l'ajout d'une consonne épenthétique).

Pour conclure cette section, les paramètres vus jusqu'ici permettent d'établir une échelle, dont les deux pôles sont constitués, respectivement, par l'allomorphie (phonologique) prototypique et par la supplétion (lexicale) prototypique :

	allomorphie	supplétion
opération phonologique	unique	non unique
lexèmes concernés	plusieurs	un
transparent	oui	non
prédictible	oui	non
conditionnement	phonologique	morphologique
effet	phonologique	morphologique

Tableau 6 : allomorphie vs. supplétion

On peut, naturellement, identifier des critères pour classer tous les cas intermédiaires. Cependant, l'existence justement d'un continuum entre deux pôles prototypiques met en doute le fait que ces types de variations doivent être traités de manière distincte.

2.8.4 Défectivité, syncrétisme et identité lexémique

Dans les sections précédentes je n'ai pris en considération, parmi les types de déviations qu'on peut observer en morphologie flexionnelle ou dérivationnelle, que la variation de forme, des bases ou des exposants. Naturellement, ce n'est pas le seul type de déviation qu'il est possible d'observer dans les langues du monde. Corbett (2007b : 30) propose une classification des types de déviations qui peuvent se manifester dans les systèmes morphologiques des langues (notamment dans les paradigmes flexionnels), et en identifie treize types différents (je renvoie à sa classification

pour plus de détails). En particulier, parmi les types de déviations identifiées je m'intéresserai ici à la défektivité, c'est-à-dire à l'absence d'une ou plusieurs formes dans le paradigme d'un lexème. La défektivité peut en effet être considérée comme un cas particulier d'allomorphie. Il a été montré, par exemple, que dans certains cas la défektivité et l'allomorphie suivent la même distribution à l'intérieur des paradigmes (par exemple dans les paradigmes verbaux des langues romanes, cf. Boyé & Cabredo Hoffherr 2008 ; Maiden & O'Neill 2008, ou slaves, cf. Baerman 2008). Comme pour les différents types d'allomorphie présentés à la section précédente, la défektivité peut être due à plusieurs types de conditionnement différents, et notamment à des conditionnements purement phonologiques ou à des conditionnements morphologiques. Rice (2007) présente plusieurs exemples du premier type, pour la plupart des 'trous' en correspondance de cases d'un paradigme qui, à la suite d'une application automatique des règles morphologiques, seraient remplies par une forme phonologiquement inacceptable voire « ineffable ». Un de ses exemple est constitué de l'impératif des verbes en norvégien, qui n'est pas possible lorsque la base contient une séquence de consonnes de sonorité croissante (37b) (Rice 2007 : 202-203) :

- (37) a. å spise / spis 'manger / mange_{IMPER}'
 å løfte / løft 'soulever / soulève_{IMPER}'
 b. å åpne / *åpn 'ouvrir / ouvre_{IMPER}'
 å sykle / *sykl 'faire du vélo / fais du vélo_{IMPER}'

Un autre exemple – dérivationnel – est celui des noms d'agent déverbaux féminins en italien : seulement les noms en *-ore* où ce suffixe est précédé d'un segment [t] peuvent avoir un correspondant en *-rice* (cf. Passino 2007 pour un traitement) :

- (38) a. vincitore vincitrice 'vainqueur'
 indossatore indossatrice 'mannequin'
 b. uccisore *uccisrice 'tueur'
 difensore *difensrice 'défenseur'

Un exemple classique de défektivité conditionnée morphologiquement est celui de

certaines verbes espagnols de la classe à infinitif en *-ir*. Il existe deux sous-classes de verbes de ce type (je ne donne ici que le présent de l'indicatif et le présent et l'imparfait de l'indicatif, d'autres formes du paradigme sont concernées, cf. Albright 2003 ; à p. ; Maiden & O'Neill 2008 et l'entrée « Spanish » du *Typology of Defectiveness Project - TDP*⁸³ pour des détails) :

(39) espagnol

a.

DORMIR ('dormir')

	1	2	3	4	5	6
PRES IND	duermo	duermes	duerme	dormimos	dormís	duermen
PRES SBJ	duerma	duermas	duerma	durmamos	durmáis	duerman
IMPF SBJ	durmiera	durmiera	durmiera	durmiéramos	durmierais	durmieran

ABOLIR ('abolir')

	1	2	3	4	5	6
PRES IND	–	–	–	abolimos	abolís	–
PRES SBJ	–	–	–	–	–	–
IMPF SBJ	aboliera	aboliera	aboliera	aboliéramos	abolierais	abolieran

b.

CONDUCCIR ('conduire')

	1	2	3	4	5	6
PRES IND	conduzco	conduces	conduce	conducimos	conducís	conducen
PRES SBJ	conduzca	conduzcas	conduzca	conduzcamos	conduzcáis	conduzcan
IMPF SBJ	condujera	condujeras	condujera	conduciéramos	condujeráis	condujeran

⁸³ <http://www.defectiveness.surrey.ac.uk/WALS/Spanish.htm>.

BALBUCIR ('bégayer')						
	1	2	3	4	5	6
PRES IND	–	balbucēs	balbuce	balbucimos	balbucís	balbucen
PRES SBJ	–	–	–	–	–	–
IMPF SBJ	balbuciera	balbucieras	balbuciera	balbucieramos	balbucierais	balbucieran

Pour résumer, la défektivité des verbes comme *ABOLIR* et *BALBUCIR* est due au fait que ces verbes ont les mêmes propriétés segmentales que d'autres verbes qui présentent une double alternance dans leur paradigme, notamment l'alternance [o] ~ [we] ~ [u] de la voyelle du thème pour les verbes comme (39a), et l'alternance [θ] ~ [θk] ~ [x] de la consonne qui précède la terminaison pour les verbes comme (39b) (les cases du paradigme qui alternent sont indiquées en gris). Le fait que *ABOLIR* et *BALBUCIR* aient la structure de verbes alternants, mais qu'ils ne réalisent pas une alternance qui serait attendue, ne permet pas d'établir de manière univoque les formes qui devraient réaliser l'autre alternance (cf. aussi Boyé & Cabredo Hofherr 2008 pour une analyse). L'explication donnée par Albright (à p.) pour les trous paradigmatiques comme ceux de l'espagnol, à savoir que « gaps occur when speakers know that an inflected form must stand in a certain relation to another inflected form, but the language does not provide enough data to be certain of what that relation should be », est tout à fait compatible avec le modèle de morphologie émergente défendu ici.

Un dernier type de défektivité s'observe lorsqu'un paradigme présente des trous qui ne sont justifiés ni phonologiquement ni morphologiquement. Corbett (2007b : 27-28), par exemple, discute le cas du nom russe *MEČTA* ('rêve') qui pour beaucoup de locuteurs ne possède pas de génitif pluriel (en (40) ce nom est présenté en parallèle avec le nom *MAČTA* 'mât' qui a un paradigme complet, cf. aussi l'entrée « Russian » du *TDP*⁸⁴) :

84 <http://www.defectiveness.surrey.ac.uk/WALS/Russian.htm>.

(40) russe

	MEČTA		MAČTA	
	SG	PL	SG	PL
NOM	mečta	mečty	mačta	mačty
GEN	mečty	–	mačty	mačt

En italien, plusieurs verbes à infinitif en *-ere* ne possèdent pas de participe passé. Pour certains d'entre eux on peut invoquer des incertitudes d'ordre morphologique, comme celles illustrées pour l'espagnol ci-dessus (PRUDERE 'démanger', SPLENDERE 'resplendir', URGERE 'être urgent'), pour d'autres, comme SOCCOMBERE ('succomber'), l'absence en synchronie de participes passés réguliers (qui sont parfois attestés en diachronie, comme **soccombuto*), il n'y a apparemment pas d'explication. Pour tous ces cas, qui résistent à toute explication phonologique et morphologique, on doit considérer qu'il s'agit de trous arbitraires, donc lexicaux, dans les paradigmes de certains lexèmes, exactement comme certains lexèmes spécifient des formes fléchies, ou des ensembles de formes fléchies, qui présentent une variation allomorphique de façon arbitraire.

La question de la défektivité des lexèmes est strictement liée à une autre question, à savoir celle de l'identité lexémique. Albright (2003 : 4) observe que pour plusieurs des verbes défectifs de l'espagnol du type de ABOLIR ou BALBUCIR les locuteurs remplacent les formes absentes par les formes correspondantes prises à d'autres verbes qui ont la même signification et qui appartiennent à la classe de verbes régulière, par exemple BALBUCIR / BALBUCEAR, COLORIR / COLOREAR, GARANTIR / GARANTIZAR, etc. La question qui peut se poser est jusqu'à quel point il est légitime de placer des frontières entre les lexèmes dans ces cas. Pour reprendre un des exemples ci-dessus, la 3^e personne du présent de l'indicatif *balbucea* est clairement une forme du verbe BALBUCEAR ; on pourrait, cependant, considérer que c'est aussi une forme du verbe BALBUCIR, si elle remplace systématiquement la forme correspondante de ce verbe. Des problèmes de paradigme se poseraient, puisque l'alternance entre *-ir* à l'infinitif et *-ea* à la 3^e personne du présent de l'indicatif n'est pas typique des verbes de cette classe, mais cela est plutôt une question de choix théoriques et descriptifs. Très souvent les grammaires de l'espagnol indiquent, pour le prétérit du verbe IR ('aller'), qui

est par ailleurs hautement irrégulier, les formes *fui, fuiste, fue, fuimos, fuisteis, fueron*, qui sont issues et coïncident avec les formes du prétérit du verbe SER ('être') et coïncident avec elles/ Elles suggèrent donc, implicitement, que les deux lexèmes possèdent une série de formes homophones, et que les locuteurs peuvent ne pas faire le lien entre les deux ensembles. Des cas de ce genre, bien que moins évidents que celui du verbe IR de l'espagnol, existent probablement dans toutes les langues. En italien il existe une forme *sepolto* ('enseveli'), qui est clairement un participe passé doté en synchronie d'un fonctionnement verbal (cf. *è stato sepolto dalla neve* 'il a été enseveli par la neige'). Etymologiquement, il s'agit du participe passé du verbe SEPPELLIRE ('ensevelir') (*sepultus* était le participe passé du verbe SEPELIO en latin). Cependant, ce verbe possède aussi un deuxième participe passé, *seppellito* qui, lui, est régulier. En fait, SEPPELLIRE appartient à la classe des verbes à infinitif en *-ire* qui présentent une extension *-isc-* à certaines personnes du présent, du subjonctif et de l'impératif. Il s'agit de la deuxième classe verbale la plus nombreuse en italien, et qui, mis à part l'extension en question, est fortement régulière (cf. Dressler *et al.* 2003). De ce fait, SEPPELLIRE serait le seul verbe de cette classe à avoir un participe passé – et une quelconque forme fléchie – irrégulier⁸⁵, ce qui a probablement conduit à la reconstruction d'un participe passé régulier *seppellito*. Thornton (à p.) a effectué des calculs à partir du corpus du quotidien *la Repubblica* et a observé que *sepolto* est préféré comme adjectif et avec des thèmes [+humain], alors que *seppellito* est préféré comme participe passé et avec des thèmes [-animé], mais les deux formes sont possibles dans toutes les configurations, sans que l'on puisse, en réalité, dégager des régularités nettes. On peut se demander quel est le lien qui existe entre *sepolto* et le lexème SEPPELLIRE dans la conscience des locuteurs, tenu compte de l'existence de la forme *seppellito* et du fait qu'un verbe (notamment de la classe *-ire/-isc-*) ne possède typiquement qu'un seul participe passé. Il n'est pas exclu qu'un certain nombre d'entre eux (dont moi-même avant de m'intéresser à la question) ne relient pas les deux de manière automatique, et que *sepolto* puisse être considéré, au moins pour ceux-ci, comme un verbe défectif de toutes les formes sauf le participe passé (et les temps composés, ou que le verbe SEPPELLIRE soit considéré comme 'surabondant' au participe passé, comportant une forme exceptionnelle et

85 D'autres possibles candidats sont PROFFERIRE ('proférer'), INFERIRE ('porter un coup'), INERIRE ('concerner'), tous des verbes très rares, et qui présentent des oscillations dans le paradigme et/ou sont défectifs à plusieurs endroits (cf. Spina 2007 : 52).

une régulière, comme semble témoigner l'exemple suivant, qui est une phrase qui a été réellement prononcée lors d'une émission télévisée (cité par Cappellaro 2010) :

(41)

Ottavia, dodici anni, scomparsa, mai più trovata; una lettera che dice che è stata violentata, uccisa e... **sepolta, seppellita**.

'Ottavia, douze ans, disparue, jamais retrouvée ; une lettre qui dit qu'elle a été violée, tuée et... *sepolta, seppellita* [ensevelie]'

De plus, sur Google on trouve plusieurs exemple de *sepolgere*, un infinitif hypothétique plausible pour ce participe passé (sur le modèle de *volgere* – *volto* 'adresser, tourner'). Bien que dans la plupart des cas cette forme soit employée dans le contexte d'un discours métalinguistique (du type « X n'est pas le participe passé de Y »), cela indique qu'il existe plusieurs locuteurs qui se sont au moins posé la question. Un autre exemple est celui du verbe *radere* ('raser') en italien. Ce verbe a en italien fondamentalement deux emplois : (i) avec le sens de 'raser' (par exemple une ville) quasi exclusivement dans la locution figée *radere al suolo* (littéralement, 'raser au sol') et (ii) avec le sens de 'couper les poils', dans ce cas il peut être employé à la forme réflexive et absolue (sans complément d'objet direct) dans le sens plus spécifique de 'couper sa barbe'. Une recherche Google (données brutes, recueillies en mars 2010) sur le participe passé de ce verbe donne les résultats suivants⁸⁶ :

(42)	i.	raso/a al suolo	433k
		rasato/a al suolo	25k
	ii.	mi sono raso/a	11k
		si è raso/a	17k
		mi sono rasato/a	114k
		si è rasato/a	154k

⁸⁶ Je n'ai effectué la recherche Google que sur les singuliers, masculin et féminin ; les formes plurielles de *raso* (*rasì, rase*) étant homophones de deux formes du prétérit, je n'avais pas les moyens de distinguer *rase al suolo* participe féminin pluriel de *rase al suolo* 3^e personne du singulier du prétérit.

Assez clairement, dans chacun des deux emplois le verbe a, tendanciellement, un participe passé différent : *raso* pour la signification (i) et *rasato* pour la signification (ii)⁸⁷. On pourrait penser qu'il existe en italien deux lexèmes *RADERE*, homophones à toutes les formes sauf au participe passé⁸⁸ ; cependant, la situation est compliquée par le fait que *rasato* est aussi le participe passé du verbe régulier *RASARE*. La relation entre ce dernier et les deux *RADERE* n'est pas si claire, et surtout rien n'indique qu'elle soit la même pour tous les locuteurs. Le but d'une théorie de la morphologie ne doit pas être celui d'établir quel est le participe passé de *SEPPELLIRE* ou de *RADERE*, ou à quel lexème appartiennent *sepolto* ou *raso* – un travail plutôt de lexicographe – mais de rendre compte du fait que, par exemple, *rasato*, au moins pour certains locuteurs et/ou dans certains cas, peut être le participe passé de *RADERE* (plus en général sur la surabondance dans les paradigmes, en particulier en italien, cf. Thornton à p.). Cela est possible dans un modèle qui considère que les lexèmes eux-mêmes, en tant qu'objets dotés de propriétés typiques, sont des généralisations qui émergent de l'ensemble des formes avec lesquelles les locuteurs entrent en contact, et que ces généralisations ne sont pas nécessairement les mêmes pour tous les locuteurs. Encore une fois, sur la base de principes d'organisation purement linguistiques nous sommes en mesure d'émettre des hypothèses sur la forme qu'un lexème peut avoir ; pour établir quelle forme il a concrètement nous devons prendre en compte d'autres critères, et renoncer à donner une réponse univoque.

Un domaine qui pose plus d'un problème de ce point de vue est celui des bases d'origine néoclassique qu'on observe dans plusieurs langues indo-européennes d'Europe. Une certaine tradition au sein de la morphologie lexématique (cf. Corbin 1992 ; 2001 ; Scalise 1984 ; 1994) a souvent défini ces bases comme des lexèmes qui ne peuvent pas être employés de façon autonome en syntaxe, et qui n'apparaissent que

87 Des doublons pour une même case du paradigme avec des spécialisations sémantico-fonctionnelles différentes ont été observées aussi dans d'autres langues par exemple en anglais où seulement à la forme écrite, selon Quirk (1970), des prétérits en *-t* (cf. *spoilt* 'dépouiller_{PRÉT}', *dreamt* 'rêver_{PRÉT}') sont de préférence utilisés dans des contextes perfectifs, alors que des variantes des prétérits des mêmes verbes en *-ed* (cf. *spoiled*, *dreamed*) sont plutôt employées dans des contextes duratifs (cf. aussi Kempson & Quirk 1971, je remercie Anna M. Thornton de m'avoir signalé ces exemples).

88 Et peut-être au prétérit ; ce temps verbal est très rarement employé, notamment à l'oral, au moins dans ma variété, et je n'ai pas de véritable intuition sur ce point.

dans des lexèmes composés ou dérivés. Cette solution est quelque peu insatisfaisante pour au moins deux raisons : d'une part, elle ne permet pas de voir la relation qui existe entre ces lexèmes et d'autres lexèmes autonomes de la langue (la relation qui lie *jeu* à *ludique* n'est pas différente du point de vue sémantique de celle qui lie *alcool* à *alcoolique*) ; d'autre part, encore une fois, les frontières entre ce qui est clairement une forme alternative d'un lexème et un lexème potentiellement différent ne sont pas nettes. Par exemple, il existe, en français, au moins trois types d'adjectifs en *-ique* : un premier groupe (43a) est constitué de lexèmes reliés de façon transparente (parfois modulo une opération phonologique générale) à un lexème existant ; un deuxième groupe (43b) est constitué de lexèmes qui ont un lien formel, plus ou moins évident, avec un lexème existant, mais pour lesquels la relation ne peut pas être décrite comme une opération phonologique simple et généralisable à d'autres dérivés ; un dernier groupe (43c) comprend des adjectifs en *-ique* qui sont reliés sémantiquement à des lexèmes existants, mais qui n'ont aucun lien formel avec ceux-ci :

- | | | |
|---------|-------------------|-------------|
| (43) a. | alcool | alcoolique |
| | Babel | babélique |
| | volcan | volcanique |
| b. | (Grande) Bretagne | britannique |
| | moine | monastique |
| | poumon | pulmonique |
| c. | cheval | hippique |
| | homme | anthropique |
| | jeu | ludique |

Plus récemment, il a été proposé (cf. Amiot & Dal 2007) de reconnaître que les bases en question ne constituent pas une classe homogène, et que certaines des formes exemplifiées en (43) doivent être considérées comme des thèmes alternatifs de lexèmes existants spécialisés dans certaines dérivations (par exemple la suffixation en *-ique*). En particulier, les deux auteures considèrent, sur la base de critères principale-

ment distributionnels, que [lyd] et [ãtrɔp] constituent, respectivement, le thème des lexèmes JEU et HOMME spécialisé, entre autres, dans la dérivation en *-ique*. Encore une fois, l'enjeu est plus de définir les relations que ces bases non autonomes peuvent avoir avec des lexèmes de la langue que d'établir avec exactitude si une de ces formes doit être considérée comme le thème alternatif d'un lexème dans l'absolu. La réponse n'est sans doute pas univoque pour toutes les formes et pour tous les locuteurs. Alors, par exemple, qu'il est raisonnable de penser que des bases non autonomes qui apparaissent dans des lexèmes très communs comme *hydr(o)* ou *hipp(o)*- constituent un thème alternatif de lexèmes existants par ailleurs (en l'occurrence EAU et CHEVAL) pour un nombre important, voire pour la majorité des locuteurs, d'autres lexèmes très employés dans des langages de spécialité (par exemple FOIE ou RATE) ne possèdent de thème alternatif que dans le lexique des spécialistes d'une discipline donnée (*hépat(o)*-, *splén(o)*-...).

La question de l'identité lexémique est strictement connectée, il est clair, à la question de l'organisation du lexique. Un domaine où ce genre de variation se manifeste de manière très nette est celui des (noms et adjectifs) ethniques, au moins dans les langues romanes. Comme il est bien connu, il existe de très nombreux ethniques pour lesquels, pour des raisons historiques, la relation avec le nom de la ville ou du pays auxquels ils se réfèrent n'est pas transparente, bien que, dans la plupart des cas, ils comportent un suffixe typique de leur catégorie (cf. *Haute-Loire* → *altiligérien*, *Saint-Etienne* → *stéphanois*). Les ethniques constituent un domaine du lexique fortement structuré : le cas canonique est celui dans lequel à un nom géographique (de ville, de région ou de pays) correspond un ethnique, et la relation entre les deux est donc globalement moins aléatoire et plus prévisible que celle entre un nom et un adjectif de relation en *-ique* illustrée en (43). Un ethnique, cependant, n'est pas forcément un lexème construit (cf. *belge*, *corse*, *suisse*...), n'a pas forcément de lien phonologique ni étymologique avec le nom géographique auquel il est relié (cf. *magyare*, *néerlandais*) et les doublons ne sont pas rares (*bongrois* / *magyare*, *néerlandais* / *hollandais*). Quelques données sont présentées en (44) (j'indique comme « ethnique 1 » ceux qui sont parfaitement transparents, comme « ethnique 2 » ceux qui présentent des variations de forme et comme « ethnique 3 » ceux qui sont totalement déconnectés du nom du pays) (sur les ethniques cf. en particulier Crocco Galès 1991 ; Roché 2008) :

(44)

<u>pays</u>	<u>ethnique 1</u>	<u>ethnique 2</u>	<u>ethnique 3</u>
Italie	italien		
Grande Bretagne		britannique	
Nouvelle-Zélande		néozélandais	
Espagne	espagnol	hispanique	
Suisse	suisse		helvétique
Japon	japonais		nippon/nipponique
Pays-Bas			néerlandais hollandais

La relation entre l'ethnique 1 et le nom du pays est clairement linguistique ; celle entre l'ethnique 2 et le nom du pays probablement aussi, si nous considérons que le lexème GRANDE BRETAGNE possède un thème alternatif [britan] pour la construction d'un lexème en *-ique* qui lui sert d'ethnique. La relation entre l'ethnique 3 et le nom du pays est plus discutable : du point de vue sémantique la relation est clairement la même (peut-être avec des différences d'ordre stylistique) ; cependant, on aurait probablement plus de mal à admettre que [elvet] et [nipɔn] sont des thèmes alternatifs de SUISSE et JAPON. Et que faire de *néerlandais* et de *hollandais*, qui sont les seuls ethniques disponibles pour *Pays-Bas* et dont le deuxième est morphologiquement relié à un autre nom (*Hollande*) ? Encore une fois, nous nous trouvons là à la frontière entre la compétence purement linguistique et la connaissance du monde : la morphologie peut simplement nous dire que tous les lexèmes de (44) sont des candidats plausibles pour les noms d'habitants des pays identifiés, elle n'est pas censé rendre compte les liens qui se créent entre ces différents lexèmes dans les lexiques individuels des locuteurs.

2.8.5 La morphologie thématique et la structure des lexèmes

Dans cette section je présente les principes de base de celle que j'ai appelée la morphologie thématique. Comme je l'ai déjà observé, ce modèle de morphologie ne

considère pas le fait de pouvoir identifier une forme sous-jacente unique pour les lexèmes comme une exigence théorique prioritaire, y compris pour les lexèmes réguliers⁸⁹. Plutôt, il déplace l'attention vers la structure globale des lexèmes, réguliers ou irréguliers, et vers les relations possibles entre leurs formes. Ce positionnement provient, d'une part, du constat que des faits tels ceux présentés ci-dessus sont tout sauf que rares dans les langues, et d'autre part que toutes les théories sont de toute façon obligées de reconnaître, au moins dans les cas les plus extrêmes, la possibilité que pour un même lexème soient mémorisées plus d'une forme. A la base de la morphologie thématique il y a aussi l'observation que pour les lexèmes appartenant à la même catégorie dans une langue l'allomorphie ne se distribue pas de manière aléatoire, mais suit des schémas récurrents (cette observation est évidemment liée à la notion de morphème d'Aronoff 1994 ; cf. aussi Maiden 1992 ; Pirrelli & Battista 2000 ; Bonami & Boyé 2003 ; 2005 ; 2007). Le positionnement relatif des formes dans la structure du lexème est plus importante que la nature de la relation qui lie une forme à une autre (par exemple, si elle est phonologiquement prédictible ou arbitraire). Prenons un exemple relativement simple, mais qui illustre bien le type de représentations qui est proposée par la morphologie thématique, celui des adjectifs du français (la discussion s'appuie principalement sur Bonami & Boyé 2005, auquel je renvoie pour plus de détails). Les adjectifs du français possèdent un paradigme flexionnel à quatre cases, correspondant au croisement des deux traits de genre et de nombre qui peuvent prendre, chacun, deux valeurs. En général, il est bien connu que pour un adjectif du français soit les quatre formes sont homophones, soit les deux formes du masculin se distinguent des deux formes du féminin. J'illustre cela avec les deux adjectifs PUR et BEAU :

⁸⁹ L'idée qu'il n'y a pas une forme sous-jacente unique du point de vue morphologique n'est pas en soi incompatible avec l'idée d'avoir des formes sous-jacentes en phonologie, bien qu'elle plaide plutôt en faveur d'une phonologie où il n'y aurait que des relations output-output (cf. par exemple Burzio 1996 ; 2002). Dans ce travail j'ai opté systématiquement pour des représentations phonologiques surfacistes de ce type.

(45)

a. PUR

	MASC	FEM
SG	pyʀ	pyʀ
PL	pyʀ	pyʀ

b. BEAU

	MASC	FEM
SG	bo	bɛl
PL	bo	bɛl

La relation entre [bo] et [bɛl] ne peut pas être décrite comme une relation phonologique simple en synchronie ; il faut donc admettre que les deux formes sont mémorisés dans l'entrée lexicale de ce lexème. A partir des données de (45) nous pouvons tirer la généralisation suivante :

(46)

	MASC	FEM
SG	A	B
PL	A	B

où A correspond au thème utilisé pour le masculin et B à celui utilisé pour le féminin⁹⁰. La distribution des formes dans le paradigme, illustrée en (46), a été appelé « espace thématique » (une étiquette proposée par Bonami & Boyé 2003 ; 2005, mais déjà présente dans la proposition élaborée indépendamment par Pirrelli & Battista 2000). La représentation lexicale du lexème BEAU comprend donc une spécification de la forme du thème A ([bo]) et du thème B ([bɛl]), celle du lexème PUR ne comprend qu'une spécification ([pyʀ]) et une 'règle de référence' qui indique que le thème A et le thème B sont identiques (B=A). Pour d'autres adjectifs, la relation entre le thème A

⁹⁰ Je ne prends pas en considération ici les adjectifs qui ont un masculin pluriel différent du singulier (comme *moral – moraux*), pour lesquels il est possible qu'il faille postuler l'existence d'un troisième thème.

et le thème B peut être exprimée comme une relation phonologique simple, par exemple effacement/ajout d'une consonne (*lent – lente, froid – froide*), (dé)nasalisation de la voyelle (*bon – bonne, serein – sereine*), (dé)voisement (*vif – vive*). Cependant, on a déjà observé que dans aucun de ces cas la relation n'est parfaitement prédictible :

- (47) *lent – lente* vs. *grand – grande*
serein – sereine vs. *pérenne – pérenne*
vif – vive vs. *brave – brave*

Tous les traitements qui essaient d'identifier une forme sous-jacente unique pour tous ces adjectifs (principalement celle du féminin de laquelle celle du masculin est dérivée par effacement de la consonne finale, cf. par exemple Dell 1973 et, pour une discussion détaillée, Tranel 1981) sont insatisfaisants, dans la mesure où ils ne prennent en compte qu'une partie des données. Plus satisfaisants sont les traitements qui reconnaissent que le thème du masculin et celui du féminin sont mémorisés simultanément dans le lexique (avec quelques variantes, cf. Tranel 1981 ; Bonami & Boyé 2005), éventuellement en faisant la généralisation que seulement certaines relations sont possibles (par exemple [ã] ~ [ãt], [ãd], [an], cf. 2.8.2 ci-dessus). La relation existante entre le thème A et le thème B (identité, phonologiquement prédictible, arbitraire) est donc moins pertinente, dans ce modèle que le positionnement relatif des deux thèmes dans le paradigme des adjectifs du français. Comme on le voit, à la différence d'autres approches de la morphologie (par exemple la Morphologie Naturelle), la morphologie thématique considère l'identité comme une relation qui n'est pas substantiellement différente des autres types de relations possibles entre les formes d'un paradigme. Comme on l'a vu ci-dessus à propos du verbe *cut* en anglais, l'identité peut être considérée comme un cas particulier d'allomorphie, lorsqu'une forme pour laquelle on s'attendrait qu'elle varie ne varie pas. Si nous nous tournons maintenant vers la dérivation⁹¹, nous nous apercevons que l'espace thématique des adjectifs nécessite au moins un autre thème pour certains types de suffixations. Plénat (2008 :

91 En réalité, pour la flexion des adjectifs du français on a besoin d'un autre thème, celui de la forme de liaison du masculin singulier, ce qui rend compte de la variation entre *gros* [gro], *grosse* [gros] et *gros* [groz] dans *gros ennui*. Je ne rentre pas dans le détail de cette question, pour laquelle je renvoie à Bonami & Boyé (2005).

1615-1616) analyse les dérivés désadjectivaux en *-ité* et propose que ce suffixe (ainsi que d'autres, comme *-isme*, *-itude* ou *-iser*) s'attache à un thème spécifique. Encore une fois, la relation qui peut exister entre ce thème et les autres thèmes du paradigme flexionnel peut être l'identité (48a-b), une relation phonologique simple (48c-d) ou complètement arbitraire (48e) :

(48)⁹²

		A (masculin)	B (féminin)		L (dérivé en <i>-ité</i>)
a.	DOCILE	dɔsil	dɔsil	=	dɔsil
	PÉRENNE	perɛn	perɛn	=	perɛn
b.	FÉCOND	fekɔ̃	fekɔ̃d	=	fekɔ̃d
	NOCIF	nɔsif	nɔsiv	=	nɔsiv
c.	ADAPTABLE	adaptabl	adaptabl	∅ ~ i	adaptabil
	INFÉRIEUR	ɛ̃ferjœr	ɛ̃ferjœr	P.S.	ɛ̃ferjœr
d.	FAMILIER	familje	familjer	P.S.	familjar
	HUMAIN	ymɛ̃	ymɛn	P.S.	yman
e.	AIGU	ɛgy	ɛgy	≠	aky
	ROND	rɔ̃	rɔ̃d	≠	rɔ̃tɔ̃d

A plusieurs endroits dans cette partie, j'ai avancé l'idée que la relation entre un thème du masculin et un thème du féminin pour les adjectifs du français, lorsqu'ils ne sont pas identiques, doit être spécifiée dans le lexique et ne peut pas être déduite par règle. En effet, la relation par défaut entre les deux thèmes de flexion d'un adjectif français est l'identité, et les adjectifs les plus réguliers sont donc ceux qui se fléchissent comme PUR (45a) (cf. le Tableau 7 ci-dessous pour des données numériques des types d'alternances observées dans les adjectifs du français). Les données de (48) montrent clairement un autre fait intéressant : lorsque le thème L n'est pas supplétif, il est tou-

92 L est l'étiquette donnée par Plénat (2008) au thème qui construit les dérivés en *-ité* (de « learned »). P.S. indique une règle de postériorisation de voyelle dite 'savante', puisqu'elle s'observe principalement dans des dérivés d'origine grecque ou latine.

jours relié au thème B par une relation prédictible, alors que ce dernier peut ne pas être relié au thème A. En d'autres termes, il n'existe pas d'adjectifs qui auraient un thème L identique au thème A avec un thème B différent. A partir de cette observation nous pouvons proposer une représentation spatiale (en s'inspirant des schémas de dépendance entre les thèmes qui ont été proposés indépendamment par Pirrelli & Battista 2000 pour la conjugaison de l'italien et par Boyé 2000 pour celle du français, de l'italien et de l'espagnol) de l'espace thématique des adjectifs du français (illustrés ici par *DOCILE*), dans lequel le thème B est directement relié aussi bien au thème A qu'au thème L, mais ces derniers ne sont pas reliés directement entre eux (j'indique l'identité comme la relation par défaut) :

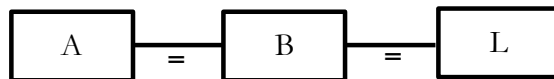


Figure 7 : schéma de dépendance des adjectifs du français (portion)

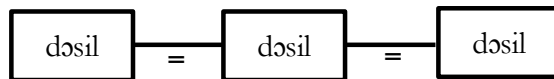


Figure 7' : schéma de dépendance de l'adjectif *DOCILE*

Comme on le voit, pour un adjectif comme *DOCILE*, il est suffisant que les locuteurs aient mémorisé un thème, et ils peuvent reconstruire la totalité du paradigme en se basant simplement sur les relations par défaut. La supplétion correspond aux cas dans lesquels le lien entre deux portions du paradigme est interrompu. Le contenu de chacune des cases doit alors être mémorisé. Les portions du paradigme qui sont reliées entre elles peuvent covarier indépendamment du reste, ce qui rend compte du fait que généralement l'allomorphie n'est pas distribuée de manière aléatoire dans les lexèmes de la même catégorie dans une langue, mais suit des schémas récurrents. A chaque lexème est associé un schéma paradigmatique ; en particulier, on fait l'hypothèse qu'un lexème régulier est un lexème pour lequel toutes les portions sont

connectées par des relations par défaut, alors qu'un lexème irrégulier présente un schéma paradigmatique particulier, mais qui est toujours une variation du schéma par défaut dans laquelle certaines connexions sont brisées (cf. Boyé 2000). Dans la Figure 8 je montre les schémas de dépendance de trois des adjectifs listés en (48) pour lesquels il est nécessaire que plus d'une forme soit mémorisée ; les relations qui ne sont pas actives pour les lexèmes en question ont été brisées :

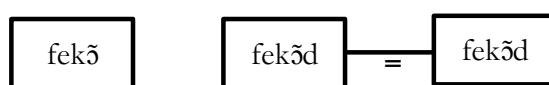


Figure 8a : schéma de dépendance de l'adjectif FÉCOND

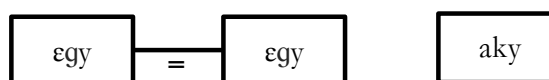


Figure 8b : schéma de dépendance de l'adjectif AIGU

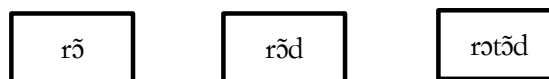


Figure 8c : schéma de dépendance de l'adjectif ROND⁹³

La représentation des lexèmes proposée dans les Figures 7 et 8 correspond à des positionnements théoriques qui ont été discutés dans les sections précédentes, et en particulier à l'idée que (i) un lexème peut avoir une ou plusieurs représentations phonologiques qui sont mémorisées ; (ii) la flexion et la dérivation interagissent et manipulent le même type d'unités ; (iii) la capacité à employer un lexème morphologique-

⁹³ Lorsque j'ai inséré le même exemple dans un article, un relecteur anonyme m'a fait remarquer que la 'bonne' base de *rotondité* est *rotond*. Certes, il existe en français un adjectif *rotond*, mais on revient là à la question de l'identité lexémique : est-on sûr qu'il existe, de manière absolue, une seule bonne base pour les mots complexes d'une langue ? *Rotond* est étiqueté par le *TLFi* de « vieilli ou littér[aire] ». On peut donc faire l'hypothèse qu'il n'est pas disponible pour un nombre considérable de locuteurs, certains desquels ont *rotondité* dans leur lexique, qu'ils peuvent lier à *rond* en vertu de son lien sémantique et de sa ressemblance phonologique avec ce dernier.

ment (le fléchir ou le dériver) est la capacité de combiner une compétence de type grammatical (connaître le schéma paradigmatique qui est associé aux lexèmes d'une catégorie déterminée) et une compétence de type lexical (connaître la ou les formes qui sont associées à l'entrée d'un lexème déterminé et le schéma paradigmatique particulier associé à ce lexème).

Celles que j'ai appelées 'portions' du paradigme correspondent à des groupes de formes fléchies qui covarient systématiquement pour tous les lexèmes. Par exemple, pour les adjectifs du français, celui que j'ai appelé le thème B correspond à la forme qui sert à construire (en ce cas par identité) les formes du féminin singulier et du féminin pluriel. Ces portions dérivent des généralisations que les locuteurs peuvent faire à partir de l'ensemble des formes (fléchies ou qui entrent en dérivation) d'un lexème. Elles correspondent à celles que Pirrelli & Battista (2000) ont appelé « classes de partition » (« partition classes ») pour la conjugaison de l'italien. Ces classes de partition ne constituent pas une classe naturelle, ni du point de vue sémantique, ni du point de vue phonologique : les formes y sont regroupées uniquement sur la base du fait de leur covariation systématiquement, donc sur la base d'un principe d'organisation purement morphologique. L'existence de tels sous-paradigmes qui ont un comportement homogène, et dont la justification est purement morphologique, est un fait désormais bien reconnu et étudié dans un grand nombre de travaux : ce sont ces sous-paradigmes, par exemple, qui correspondent à ce qu'Aronoff (1994) a appelée 'morphomes'. L'existence de tels sous-paradigmes est de plus en plus prise en compte par des approches plutôt incrémentales, comme OT (cf. McCarthy 2001 ; Ohannesian & Pons 2009, entre autres). L'utilité de ce niveau d'organisation a été montrée, par exemple, dans l'étude de phénomènes morphologiques particuliers comme le syncrétisme ou la déponence (cf. Corbett 2007b pour un aperçu), la défec-tivité (cf. Boyé & Cabredo Hofherr 2008 ; Maiden & O'Neill 2008), et aussi dans l'évolution diachronique des langues (cf. Maiden 1992 ; 2003).

Je fais l'hypothèse que les relations qui relient les formes fléchies entre elles et celles qui relient les classes de partition sont du même type, et en particulier qu'elles peuvent correspondre à l'identité, à une opération phonologiquement motivée, à une relation phonologique arbitraire ou être entièrement supplétives. Ainsi, par exemple, pour les adjectifs exemplifiés dans les Figures 7 et 8, chacun des thèmes est relié aux

formes fléchies qu'il sert à construire par une relation d'identité :

- (49) MASC SG = Th A
 MASC PL = Th A

Contrairement à une approche purement connexionniste, dans laquelle toutes les formes fléchies d'un lexème seraient reliées entre elles, je défends donc l'hypothèse qu'il existe un niveau d'organisation morphologique, qui correspond au niveau thématique.

Les exemples présentés jusqu'ici ne mettent en jeu que l'identité comme relation par défaut entre les thèmes ou entre les thèmes et les formes. Comme je l'ai indiqué, cependant, cette relation peut être de différents types, et correspondre à une opération phonologique, plus ou moins naturelle et motivée. Les adjectifs du catalan ont une structure paradigmatique sensiblement semblable à ceux du français, avec un thème qui est utilisé pour les formes du masculin, et un qui est utilisé pour les formes du féminin, et qui peuvent être reliés, comme en français, par une relation d'identité ou par une relation qui est phonologique, mais pas entièrement prédictible. Je présente deux exemples en (50)⁹⁴ :

(50)

a. PRIM 'maigre'

	MASC	FEM
SG	prim	primə
PL	primɪs	priməs

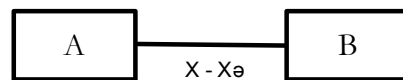
94 Je simplifie un peu les termes de la question : pour certains adjectifs le masculin pluriel est construit sur un thème qui est identique à celui du féminin, mais différent de celui du masculin singulier (cf. *bufó* – *bufons* – *bufona* 'bouffon'). Les adjectifs du catalan peuvent de plus être répartis en plusieurs classes (cf. ci-dessous). Je renvoie à Guerrero (2008) pour une analyse thématique complète des adjectifs du catalan.

b. CLAR 'claire'

	MASC	FEM
SG	kla	klarə
PL	klas	klarəs

Comme on le voit, le thème B (celui du féminin) se distingue du thème A minimale-ment par le fait qu'il se termine par un [ə], et les formes du pluriel se distinguent de celles du singulier parce qu'elles se terminent par un [s]. On peut donc proposer un schéma de la structure paradigmatique des adjectifs du catalan comme PRIM (51a) et une série de fonction qui permettent de relier les thèmes avec les formes fléchies (51b) :

(51) a.



- b.
- | | | |
|---------|---|----------|
| MASC SG | = | Th A |
| FEM SG | = | Th B |
| MASC PL | = | Th A ⊕ s |
| FEM PL | = | Th B ⊕ s |

Assimiler les relations qui peuvent être partiellement motivées phonologiquement et celles qui sont arbitraires ne signifie pas nier en entier qu'il existe des règles purement phonologiques. Par exemple, il est clair que l'abaissement de la voyelle qu'on observe dans des verbes comme *sécher* ([seʃe]) – *il sèche* ([seʃ]) est un effet purement phonologique, mais le rôle de la phonologie est limité à ces cas d'application automatique de règles phonotactiques générales.

Les relations entre thèmes et entre thèmes et formes exemplifiées dans ce travail ne sont pas orientées. Cela correspond à l'idée qu'un locuteur est généralement en mesure, à partir d'une forme fléchie d'un lexème inconnu, de reconstruire, totalement ou en partie, son paradigme ; cette capacité est maximale dans le cas des lexèmes réguliers. La non orientation des relations découle naturellement de l'idée que, même lorsque ces relations correspondent à une fonction phonologique, celle-ci

peut être partiellement arbitraire, et donc lexicale. Prenons l'exemple des verbes de l'italien en *-ire* avec l'extension *-isc-*, déjà cités en 2.8.4. Pour clarté, je reproduis en (52) le présent de l'indicatif d'un verbe de ce type (FINIRE 'finir'), avec la transcription phonologique, et, pour permettre une comparaison, le présent de l'indicatif d'un verbe en *-ire* qui ne présente pas d'extension (PARTIRE 'partir') :

(52)

1	2	3	4	5	6
finisco	finisci	finisce	finiamo	finite	finiscono
f'nisko	f'niʃi	f'niʃe	f'njamo	f'nite	f'niskono
parto	parti	parte	partiamo	partite	partono
'parto	'parti	'parte	par'tjamo	par'tite	'partono

L'extension – qui phonologiquement correspond tantôt à [isk] tantôt à [iʃ:] – a souvent été analysée comme un objet unique, qui est rajouté à une forme sous-jacente (par exemple [fini], cf. DiFabio 1990 ; Schwarze 1999 ; Paster 2006 : 121-123, entre autres) ou bien qui est supprimé à partir d'une forme sous-jacente qui l'inclut (cf. Vogel 1993 : 223-225, qui, sauf erreur de ma part, n'indique pas quelle serait la forme sous-jacente en question, mais vraisemblablement [fniʃ:i] ; cf. aussi Vogel & Napoli 1995). L'hypothèse que l'extension est ajoutée à une forme sous-jacente a le désavantage qu'une base verbale en [i] doit être lexicalement spécifiée comme étant [\pm extension]. C'est la raison principale qui pousse Vogel (1993) à considérer que l'extension en question fait partie du thème principal du verbe et qu'elle est effacée à certains endroits du paradigme. Cette hypothèse, cependant, n'est ni plus économique, ni plus prédictive : elle est entièrement fondée sur l'assomption que l'entrée lexicale d'un verbe comprend toujours un seul thème ; mais dans la mesure où il existe en italien un certain nombre de verbes, notamment de la première classe, se terminant en *-iscare* ou en *-isciare*, il y aura toujours des formes fléchies qui demeureront ambiguës, par exemple, une première personne en [isko] ou une deuxième personne en [iʃ:i] (*confisco* 'je confisque' ou *scudisci* 'tu fouettes'). En réalité, le fait même que les linguistes hésitent entre les deux options montre qu'identifier laquelle des deux options (ajout ou

effacement) est la plus naturelle pour les locuteurs ne va pas de soi. Dans l'approche défendue ici, où les lexèmes peuvent avoir une représentation phonologique complexe, l'hypothèse la plus raisonnable semble être celle de considérer qu'il existe une classe de verbes en italien pour lesquels l'alternance $\emptyset \sim [\text{isk}] \sim [\text{iʃ}]$ exprime une relation (phonologiquement arbitraire) symétrique entre trois thèmes qui occupent des places spécifiques dans le paradigme des verbes. .

Dans les Figures 7 et 8 j'ai proposé une représentation des paradigmes adjectivaux du français qui présupposent que la fonction par défaut (en l'occurrence l'identité) est la seule possible pour relier les thèmes de l'espace thématique. En réalité, dans la morphologie thématique, rien n'empêche que pour les lexèmes de la même classe plusieurs fonctions, éventuellement en concurrence entre elles, soient actives. Cela est même le principe qui est à la base de la répartition des lexèmes en classes de flexion. Reprenons l'exemple des adjectifs du catalan illustrés en (50). En plus des modèles flexionnels illustrés par PRIM et CLAR, il en existe au moins trois autres :

(53)

a. PRIM 'maigre'

	MASC	FEM
SG	prim	primə
PL	primɪs	priməs

b. JOVE 'jeune'

	MASC	FEM
SG	ʒəβə	ʒəβə
PL	ʒəβəs	ʒəβəs

c. GRAN 'grand'

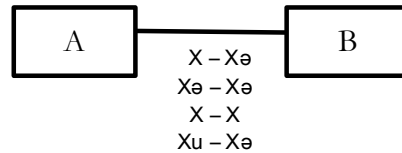
	MASC	FEM
SG	gran	gran
PL	grans	grans

d. FLONJO ‘mou’

	MASC	FEM
SG	flon <u>ʒ</u> u	flon <u>ʒ</u> ə
PL	flon <u>ʒ</u> us	flon <u>ʒ</u> əs

Si on prend en compte les données ci-dessus, le schéma proposé en (51a) peut être complété de la manière suivante (les fonctions qui relient les thèmes aux formes fléchies restent les mêmes pour tous les adjectifs) :

(54)



En réalité, la relation $Xə - Xə$ est déjà incluse dans la relation $X - X$ et donc redondante. Le thème A et le thème B des adjectifs du catalan peuvent donc être reliés par une des trois relations $X - X$, $X - Xə$ ou $Xu - Xə$. Ces relations sont en concurrence mutuelle, dans la mesure où il existe des formes systématiquement ambiguës : un féminin en [ə], par exemple, peut correspondre à un masculin en [ə], en [u] ou se terminant par consonne.

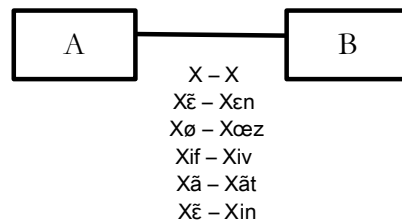
Les trois relations identifiées pour le catalan sont distribuées de manière relativement homogène parmi les adjectifs de cette langue. En revanche, pour les thèmes du masculin et du féminin des adjectifs du français, il existe une relation (l'identité) qui est de loin la plus représentée par rapport aux autres, et que j'ai identifiée ci-dessus comme étant le défaut. Dans le Tableau 7 je présente un calcul des relations entre les thèmes des adjectifs du français que j'ai effectué à partir de la base lexique.org :

invariables	<i>pur – pure</i>	80,2%
ã – at	<i>lent – lente</i>	7,1%
ø – œz	<i>joyeux – joyeuse</i>	2,8%
ẽ – en	<i>romain – romaine</i>	2,3%
if – iv	<i>vif – vive</i>	1,6%
e – er	<i>léger – légère</i>	0,9%
ar – ard	<i>bâtard – bâtarde</i>	0,7%
ẽ – in	<i>fin – fine</i>	0,7%
autres		3,8%

Tableau 7 : proportion des différents schémas flexionnels pour les adjectifs du français

La différence entre les schémas présentés pour le catalan en (54) et pour le français dans le Tableau 7 n'est pas une différence de nature, mais uniquement numérique. Cette conception est cohérente avec l'idée d'une morphologie dans laquelle les 'règles' émergent comme des généralisations à partir d'analogies sur le lexicque observé. La représentation qui a été proposée en (54) pour le catalan peut donc être appliquée aux adjectifs du français :

(55)



Cela explique les hésitations des locuteurs, en particulier, mais pas seulement, lorsqu'ils doivent fléchir des adjectifs nouveaux et/ou inconnus et l'apparition de formes telles que *pécunier – pécuniaire*, *avare – avarde* (cités dans Tranel 1981 : 255, 264) *pérein –*

pérenne (cité dans Plénat 2008 : 1615), *rabat-joie* – *rabat-joise*, toutes basées sur des schémas bien attestés pour les adjectifs du français. Comme on le voit, ces exemples militent en faveur d’une vision non orientée des relations paradigmatisques, puisque l’application de schémas récurrents va parfois dans le sens masculin → féminin (*avarde*, *rabat-joise*), parfois dans le sens inverse (*pécunier*, *pérein*). Pour résoudre la concurrence entre les différents schémas les locuteurs peuvent se baser sur différents paramètres, qui peuvent interagir entre eux dans un système de contraintes, comme il a été discuté en 2.6.3 (cf. aussi Boyé à p. pour une vision des fonctions qui relient les thèmes dans l’espace thématique en termes de contraintes), et en particulier sur la fréquence des différents schémas. Par exemple, alors que l’identité est globalement la relation la plus représentée, lorsqu’un adjectif se termine au masculin par une voyelle nasale, elle est loin de représenter le cas par défaut. Le Tableau 8 montre un calcul des différentes alternances observées avec un masculin en voyelle nasale effectué à partir de lexique.org⁹⁵ :

95 Pour les adjectifs qui présentent une alternance, comme *marron* – *marron/marronne* ou *malin* – *maline/maligne*, je n’ai tenu compte que de la forme la plus commune.

masculin	féminin		exemple
ã	ã	4	SAFRAN
	ãt	1 141	LENT
	an	36	OCCITAN
	ãd	14	GRAND
	ãʃ	2	BLANC
ẽ	ẽ	4	ZINZIN
	en	378	SAIN
	in	107	FIN
	ẽt	26	SAINT
	yn	7	BRUN
	ijn	1	BÉNIN
	õ	õ	7
on		46	BON
õd		12	ROND
õg		2	LONG
õt		1	PROMPT
õs		1	ABSCONS

Tableau 8 : alternance des adjectifs à masculin en voyelle nasale en français

Je propose donc un modèle dans lequel les différentes relations que les formes d'un lexème peuvent entretenir entre elles (formes fléchies ou qui sont utilisées en dérivation) ne sont pas distinguées sur la base de leur nature (par exemple, celles qui sont régulières de celles qui ne le sont pas), mais sont au plus hiérarchisées, par exemple sur la base de leur fréquence dans le lexique, la hiérarchie elle-même pouvant varier sous l'effet de différentes contraintes (cf. la discussion en 2.6.2 sur le rôle de la fréquence et du raisonnement de type probabiliste dans la structuration de la compétence morphologique). Cela nous permet, d'une part, de relier la régularité à la fré-

quence, en en faisant une notion gradiente, et d'autre part d'en proposer une définition explicite : un lexème régulier est un lexème pour lequel toutes les zones de l'espace thématique, et donc toutes les formes, sont reliées par des fonctions prédictibles et qui peut entièrement servir de modèle pour de nouveaux lexèmes.

Pour résumer, dans cette section j'ai essayé de donner une caractérisation de la morphologie dite thématique qui se base, en particulier, sur les présupposés suivants :

- un lexème peut avoir une représentation phonologique multiple, c'est-à-dire que plusieurs thèmes peuvent lui être associés ;
- dans une langue, pour les lexèmes de la même catégorie, l'allomorphie ne se distribue pas de manière aléatoire, mais suit des schémas récurrents ;
- ces schémas permettent de définir l'espace thématique des lexèmes d'une catégorie déterminée ; on peut prévoir que l'espace thématique de chaque lexème soit par défaut une version simplifiée de l'espace thématique de sa catégorie ;
- les différentes zones de l'espace thématique (thèmes) sont reliés entre elles par des relations qui peuvent correspondre à une relation phonologique motivée ou à une relation phonologique arbitraire ; deux zones peuvent aussi ne pas être reliées ; dans ce cas, plusieurs thèmes doivent être mémorisés pour le même lexème (cas de supplétion) ;
- il peut y avoir plusieurs schémas en concurrence (plusieurs relations différentes reliant les mêmes zones de l'espace thématique) pour les lexèmes de la même catégorie ;
- un lexème régulier est un lexème pour lequel toutes les zones de l'espace thématique sont reliées par des relations prédictibles et qui peut servir de modèle pour d'autres lexèmes.

2.9 Conclusion : quelques problèmes résidus

2.9.1 La question des données

La question des données se pose pour toute étude linguistique, et tout particulièrement pour les études de morphologie. Dans le paragraphe 2.7 j'ai longuement discuté la question de la mémorisation des unités linguistiques et l'interaction entre le lexique et la compétence morphologique des locuteurs. Le but d'une théorie morphologique est de donner une description la plus précise possible de la deuxième, tout en tenant compte du fait que le premier fait partie de la compétence linguistique plus globale et qu'il est à la fois le point d'arrivée et le point de départ de cette compétence morphologique. La morphologie s'occupe principalement de lexèmes, des unités qui sont, plus typiquement que d'autres, mémorisées, bien que, comme on l'a vu, tous les lexèmes ne le soient pas forcément. Un morphologue est quelqu'un dont le premier souci, pour entreprendre son étude, est celui de constituer une liste de mots. Tout naturellement, les morphologues se sont principalement servis des listes de mots qui existaient déjà, des dictionnaires, éventuellement des corpus, notamment pour les langues avec une riche tradition grammaticale et lexicographique. Cependant, les dictionnaires sont censés refléter le lexique d'une langue, pas le produit de la compétence morphologique de ses locuteurs. Les plus gros dictionnaires contiennent effectivement aussi un grand nombre de mots morphologiquement construits plus ou moins transparents, mais la fonction principale d'un dictionnaire est celle de permettre d'avoir accès à une information qu'on ne peut pas déduire autrement : justement toutes les idiosyncrasies des mots qui, en principe, proviennent plus du lexique que de la compétence morphologique. En d'autres termes, on aura plus de chances de devoir aller chercher dans un dictionnaire *remonter* que *renationaliser*.

Le lexique, je l'ai déjà observé, est une entité composite, stratifiée, historiquement et socialement déterminée, des caractéristiques qui le rendent, évidemment, inapte à fournir des indications, sinon très vagues, sur le fonctionnement concret de la langue en synchronie. Naturellement, cela ne veut pas dire que les données lexicographiques doivent être complètement délaissées dans les études de morphologie. Ce qui est illusoire est de pouvoir écrire, par exemple, la 'grammaire' d'un certain suffixe enregar-

dant dans un dictionnaire tous les mots qui se terminent par ce suffixe, et uniquement ceux-là. Une telle entreprise équivaldrait à essayer de décrire les principes de l'architecture moderne en se promenant dans les rues de Rome. Probablement, on arriverait à donner une description précise, même cohérente, de ce qu'on aurait observé, peut-être en excluant comme des exceptions quelques bâtiments qui ne sont pas tout à fait conformes à cette description. Cependant, ce ne serait vraisemblablement pas une représentation très réaliste de l'objet qu'on voulait décrire.

Dans le passé, les morphologues ont eu recours aux données lexicographiques principalement pour des raisons pratiques. Cependant, j'ai l'impression que cette activité correspond aussi à une prise de position théorique, plus ou moins consciente, qui s'enracine dans le prestige que nous tous – scientifiques ou pas – attribuons à l'autorité en matière de langue, principalement les grammaires et les dictionnaires. En particulier à l'intérieur d'une mouvance théorique sélective comme la grammaire générative, on a toujours tenu à distinguer ce que, selon une métaphore spatiale, on étiquetait comme le 'centre' de la grammaire de sa 'périphérie', la compétence de la performance. Plusieurs phénomènes, notamment relevant de la créativité lexicale, qui ne rentraient pas forcément dans les catégories établies, ont été ainsi dédaigneusement exclus du domaine d'intérêt de la linguistique en raison du fait qu'ils appartenaient à la périphérie. Curieusement, ce type de jugements concerne le plus souvent des données qui sont marginales aussi du point de vue sociolinguistique, par exemple des données néologiques, relevant de la créativité lexicale, voire appartenant à des registres non standard. Les données néologiques et les occasionnalismes, au contraire, lorsqu'on peut y avoir accès, constituent une ressource précieuse, justement pour observer le fonctionnement « on-line » de la compétence morphologique⁹⁶. Les moyens de communication modernes, et notamment Internet, ont rendu possible, au cours des quinze dernières années, l'accès à une masse de données sans précédent, une grande partie desquelles n'auraient jamais été enregistrées autrement, et qui ont en partie bousculé des concepts comme ceux de 'mot existant' ou de 'mot attesté'. De

⁹⁶ Dans la même catégorie il faudrait probablement inclure, comme exemple extrême, les formations qui relèvent des 'langages secrets', observés dans plusieurs langues, où les unités sont modifiées de manière automatique selon des principes qui, en général, relèvent plutôt de principes cognitifs généraux que de procédés grammaticaux spécifiques aux langues (cf. déjà Yip 1982 : 637 pour des observations sur l'utilité d'étudier de telles données dans ce but, et aussi Fradin *et al.* 2009 : 28-32 ; Plénat 2009).

plus en plus de travaux de morphologie, surtout dérivationnelle, se servent des données tirées du Web, ce qui, qualitativement et quantitativement, représente un changement important par rapport au passé⁹⁷. Les mêmes préjugés sociolinguistiques que j'ai cités – encouragés par les genres des textes qu'on trouve typiquement sur le Web – se manifestent d'ailleurs souvent, aujourd'hui, vis-à-vis de ce moyen. Dans les pages précédentes j'ai abondamment, et sans aucun état d'âme, utilisé des données tirées du Web, un grand nombre desquelles relevaient clairement d'un usage créatif de la langue. Ces données sont à mon avis utiles et intéressantes, justement dans la perspective d'une théorie de la langue qui ne repose pas sur une vision binaire des faits linguistiques, 'bons' ou 'mauvais', 'recevables' ou 'irrecevables', 'grammaticaux' ou 'non grammaticaux', mais qui est capable de décrire les mécanismes à travers lesquels toutes les données réellement observées ont été construites et éventuellement de les situer par rapport au cas qui peuvent être considérés comme canoniques.

2.9.2 Descriptif vs. explicatif

Un leitmotif de la linguistique moderne, en particulier de la linguistique formelle, est la recherche de l'« explicativité », et par conséquent la relégation d'un certain nombre d'approches et de travaux parmi les études purement descriptives, donc non scientifiques et inintéressantes. Encore, faut-il avoir clairement à l'esprit ce qu'on veut expliquer. En matière de morphologie notamment, une approche qui se veut explicative peut viser à rendre compte des mécanismes cognitifs qui sont à la base de la segmentation de la réalité en unités discrètes, de la projection de ces unités dans du matériel linguistique et de sa mémorisation, du fait que certains de ces processus cognitifs peuvent se grammaticaliser⁹⁸ dans telle ou telle langue, et de comment la grammaire

97 Pour un aperçu des enjeux théoriques et pratiques de l'usage du Web comme corpus pour les études de morphologie, cf. Hathout *et al.* (2008 ; 2009) et les observations de Rainer (2005 : 350) ; Ricca (à p. : 254). Plus en général sur la relation théorie / données et sur les développements récents, notamment de la phonologie et de la morphologie, dans ce domaine, cf. Durand (2009).

98 J'utilise ici l'étiquette de 'grammaticalisation' de façon pas tout à fait canonique, parce que en général par ce terme on définit la transformation d'un élément lexical en un élément grammatical. Il me semble, cependant, que le fait d'utiliser grammaticalisation pour se référer à la conventionnalisation d'une construction est cohérent avec le modèle de morphologie émergent et 'déréifié' défendu ici. (Pour des approches récentes qui essaient d'étendre le concept de grammaticalisation aux constructions cf. par exemple Bybee 2003 ; Traugott 2003).

ainsi créée interagit avec d'autres processus cognitifs plus généraux. Il n'est pas sûr, en revanche, que chaque phénomène morphologique particulier puisse être étudié dans des termes qu'on puisse qualifier de véritablement explicatifs. Beaucoup de travaux en linguistique formelle, notamment en morphologie, se focalisent sur un phénomène particulier en supposant qu'il peut servir de clé pour donner une explication du fonctionnement du langage. Ce faisant, ils manquent du recul nécessaire, me semble-t-il, pour voir à quel niveau doivent se placer les explications et où, au contraire, le linguiste – ainsi que le locuteur qui apprend une langue – doit se limiter à constater des faits. L'approche thématique à la morphologie que j'ai caractérisée dans la section précédente se prête bien à être accusée de 'descriptivisme' : pour beaucoup des faits observés, en effet, elle se limite à constater l'existant ; par exemple, j'ai déjà affirmé à plusieurs reprises que la priorité, dans ce modèle, est donnée à la structure des relations entre différentes zones du paradigme d'un lexème, plutôt qu'à la nature de ces relations, et qu'aucune différence de substance n'est faite entre des relations phonologiquement motivées et des relations phonologiquement arbitraires. Dans ce qui suit, je présente quelques arguments pour soutenir que dans l'analyse des phénomènes linguistiques le niveau de l'explication ne coïncide pas nécessairement avec la prise en compte des données immédiatement observables.

En premier lieu, on peut se demander quelle est la plus-value empirique qu'une analyse abstraite rapporte en comparaison avec une approche à l'apparence descriptive. Je prends l'exemple de l'article défini masculin en italien, dont la distribution de base est fournie en (56) :

(56)	l'/gli ⁹⁹	_V	l'albero / gli alberi ('le(s) arbre(s)')
	il/i	_C	il cane / i cani ('le(s) chien(s)')
	sauf		
	lo/gli	_sC	lo scorpione / gli scorpioni ('le(s) scorpion(s)')
		_ts, _dz, _ʃ, _ɲ	lo zucchero ['ts:ukɛro]/ gli zuccheri ('le(s) sucre(s)')

Pour les données de (56) on peut imaginer que la description qu'on en a donnée correspond de manière approximative à la représentation mentale qu'en ont les

99 Phonologiquement *gli* correspond à [ki].

locuteurs : l'apparition des différents allomorphes pour le grammème qui réalise le déterminant défini singulier et pluriel (qui sont mémorisés en une relation de correspondance réciproque) est définie par la séquence phonologique qui le suit immédiatement. Les analyses formelles qui ont été données de l'alternance en question (cf. Davis 1990 ; 1999 ; Larsen 1998) ont en général essayé de relier l'apparition des différents allomorphes à la structure syllabique du mot qui suit et éventuellement au domaine prosodique d'application (qui est généralement considéré comme le groupe clitique, cf. Renzi & Vanelli 1993). Apparemment, ces traitements sont préférables, en ce qu'ils renoncent à des conditionnements purement locaux, et relient le choix de l'allomorphe à des principes phonologiques plus globaux qui servent à traiter d'autres phénomènes en italien (par exemple le « raddoppiamento sintattico »). Si cela est vrai, ils devraient être au moins aussi prédictifs qu'une approche purement stipulative comme celle qu'on a esquissée ci-dessus. Par contre, les traitements auxquels je me réfère ont comme premier et à mon sens plus sérieux défaut de s'intéresser exclusivement à la forme du singulier du déterminant, alors qu'il est très clair que le déterminant en question (singulier et pluriel) entre dans une série de relations qui comprennent aussi le déterminant indéfini singulier UN et ses dérivés (NESSUNO, ALCUNO...), le démonstratif QUELLO et les adjectifs BUONO et BELLO lorsqu'ils sont antéposés à un nom. Or, l'hypothèse que *lo* est préféré à *il* dans les contextes indiqués parce qu'il permet de fournir un noyau vocalique pour des séquences qui ne sont pas de 'bonnes' attaques syllabiques (comme [sC] et [ts], [dz], [ʃ], [ɲ], qui sont intrinsèquement géminées)¹⁰⁰ perd de sa valeur, si on considère que la forme du singulier et celle du pluriel covariant systématiquement, et que le noyau vocalique en question serait fourni aussi bien par la forme du pluriel qui correspond à *lo* (*gli* [ʎi]) que par celle qui correspond à *il* (*i* [i]) (cf. aussi Nevins à p.). De plus, alors que [ts], [dz], [ʃ], [ɲ] sont effectivement intrinsèquement géminées dans un grand nombre de variétés de l'italien, il existe des variétés (par exemple celle, septentrionale, de l'auteur de ce texte) dans lesquelles au moins le phonème [ʃ] ne l'est pas en début de mot. Dans ces variétés la structure de l'attaque de *scivolo* ([ʃivolo], 'toboggan') n'est pas différente de celle de *simbolo* ([simbolo] 'symbole') ; au moins dans la grammaire qui décrit ces variétés,

100 Avec des variantes : pour Larsen (1998) la forme sous-jacente de l'article est /ilo/, Davis (1990 ; 1999) n'indique pas quelle est la représentation de l'article, mais vraisemblablement pour lui *lo* et *il* sont pour lui deux allomorphes dont l'apparition est phonologiquement déterminée.

il semble difficile de pouvoir caractériser la sélection de l'allomorphe sur une base purement phonologique. Même en se limitant à l'italien standard, les correspondances de (55) fonctionnent systématiquement pour des attaques plutôt communes, comme [ts], [dz], [ʃ], mais pour des attaques moins communes, comme [ɲ], [ps], [pn], etc. les choix des locuteurs sont beaucoup moins tranchés et on observe des hésitations comme *il/lo gnocco* ('le gnocchi'), *il/lo PSI* ('le PSI, Parti Socialiste Italien', exemple de Marotta 1993), *il/lo pneumatico* ('le pneu') (cf. Marotta 1993, mais aussi Bertinetto 1999 : 86-98 ; McCrary 2002) ; de plus, Marotta (1993) a montré que le comportement de l'article défini *il/lo* et celui de l'article indéfini *un/uno* ne sont pas uniformes. Tous ces faits sont compatibles avec l'idée que dans le choix de l'allomorphe pour les cas nouveaux les locuteurs se basent sur l'analogie avec les formes attestées et sur leur fréquence. Enfin, du point de vue diachronique, le contexte d'apparition des différentes formes singulières et plurielles du déterminant a été fortement variable à différentes époques et ne s'est stabilisé qu'à partir de la deuxième moitié du XIX^e siècle, sur la base des recommandations des grammairiens du XVI^e (Rohlf's 1968 : 99-114 ; Migliorini 1988 (1958) ; Renzi & Vanelli 1993). Il s'agit donc, en partie, du fruit d'une codification explicite et consciente, et inversement pour les locuteurs il s'agit souvent d'un apprentissage de l'emploi de l'article en tant que 'règle' codifiée¹⁰¹. Sans vouloir nier qu'entre le choix de l'article en italien et la structure phonologique il y a un lien, on peut s'interroger sur l'opportunité d'essayer d'expliquer les faits observés sur une base purement phonologique. Cela implique l'idée, en effet, que le système que les linguistes peuvent observer en synchronie est phonologiquement optimal. Or, il me semble qu'il y a deux présupposés qui peuvent être à la base de cette idée : (i) que la grammaire constitue toujours un système optimal (ce qui est bizarre à soutenir, puisque, comme on l'a vu plus haut à propos des mots qui commencent par [ʃ], il existe dans les différentes variétés de l'italien, des incohérences) ; (ii) que les mécanismes de sélection de l'article en italien ont trouvé un point d'équilibre optimal dans le système actuel. Dans les deux cas, il s'agit de présupposés non vérifiés, et l'idée implicite (également non vérifiée) sur laquelle on se base est que les données observables par les linguistes à un moment déterminé sont toujours un reflet du système

101 Cf. Krämer (2009c : 147-148 n. 14) qui observe que « either we are still far from an understanding of Italian article selection or it is indeed a quirky offshoot of prescriptivism, and therefore outside the realm of grammatical analysis ».

de règles inconscientes que les locuteurs possèdent.

La relation entre les données observables en synchronie et l'histoire de la langue, et plus précisément entre des règles qui étaient actives à des étapes anciennes de l'évolution des langues, et la conscience linguistique des locuteurs en synchronie est un problème récurrent dans les études de morphologie. Un grand nombre des faits que j'ai discutés dans les sections précédentes, et en particulier en 2.8, et que j'ai traités comme des alternances spécifiées lexicalement ont en effet une explication qui est diachronique. Les alternances dans les terminaisons des formes du féminin et du masculin des adjectifs du français, par exemple, correspondent pour la plupart au résultat des processus phonologiques qui ont eu lieu dans le passage du latin au français et au cours de l'évolution de ce dernier. Comme l'observe Bertinetto (2009 : 117-118)¹⁰², à propos de la phonologie générative :

Molti dei più ardui problemi (non di rado paradossali, veri e propri rompicapo) tradizionalmente discussi in fonologia generativa non sono altro, a mio avviso [...], che il prodotto di un errore ottico, dovuto alla pretesa di comprimere in sincronia ciò che, in realtà, si è disteso su un ampio arco di tempo. Molte presunte regole sincroniche non sono altro – nella prospettiva del non generativista – che il riflesso attuale di eventi precedenti, definitivamente incapsulati nei meccanismi morfofonologici [...]. Il parlante li interiorizza apprendendo il lessico della propria lingua, e su questa base sviluppa la capacità di – e la propensione a – estenderli all'occorrenza anche a nuovi materiali (prestiti o neologismi).

Le principe explicatif ne réside pas, dans ce cas, dans le fait de pouvoir réduire la variété des données observées à une unité, mais plutôt dans la reconnaissance du fait que la grammaticalisation (dans ce cas la morphologisation) est un phénomène généralisé et une source importante pour la structuration des langues. Le même problème se pose pour la classification des unités linguistiques. Les catégories identifiées par les linguistes, qui sont des objets éminemment descriptifs et classificatoires, et ont été pour la plupart élaborées pour la description de langues particulières (principalement indo-européennes) sont souvent confondues avec des principes explicatifs (cf. Haspelmath 2007). J'ai été confronté avec ce problème lors de mon travail de thèse (cf.

102 Cf. aussi Plénat (1987 : 138).

Montermini 2008b), quand j'ai essayé d'établir une liste des préfixes de l'italien. Beaucoup des unités observées ne rentraient pas dans les définitions qu'il était possible de donner pour la catégorie préfixe, mais pouvaient difficilement être classées autrement. De plus, la catégorie même des préfixes était très différenciée à son intérieur, avec plusieurs sous-catégories présentant des propriétés différentes. Puisque la source première pour la naissance des affixes est la grammaticalisation à partir d'unités lexicales autonomes (pour la plupart des prépositions dans le cas des préfixes), je me suis rendu compte que ce que j'étais en train d'observer, en regardant le fonctionnement des préfixes en synchronie, était une 'photographie' d'une situation dans laquelle coexistent plusieurs unités qui se trouvent à des stades différents de grammaticalisation, et que les comportements différents qu'ils montraient étaient principalement dus à cela. Une photographie nécessairement écrase la profondeur et montre sur le même plan des objets qui se trouvent à des distances différentes. Encore une fois, Bertinetto (2009 : 77) a exprimé cela avant et mieux que moi :

Nella loro *libido classificandi*, gli strutturalisti non hanno prestato la debita attenzione al fatto che gli elementi che convivono in un certo stadio sincronico possono appartenere, in realtà, a stratificazioni diverse della struttura linguistica, e risultano pertanto dotati di gradi diversi di vitalità. Qui entra in gioco il concetto di 'grammaticalizzazione' [...]. Ogni lingua presenta – in qualsiasi momento la si osservi – elementi vitali ed in fase di crescita accanto ad altri in graduale raffreddamento [...] la proiezione di tutte queste entità su un medesimo piano costituisce un'illusione ottica, che non rende affatto conto della variegazione dei dati.

Les classes établies par les linguistes ne constituent pas un principe explicatif en soi. Plutôt que remodeler leur classes pour y faire rentrer plus de données, ou en créer de nouvelles, ou encore exclure de leur analyse les données qui ne rentrent pas dans les classes préétablies, les linguistes devraient se dire que, dans ce cas, le niveau réellement explicatif est celui dans lequel on s'intéresse à la manière dont l'esprit humain segmente et classe les données linguistiques auxquelles les locuteurs sont confrontés, et à la manière dont les langues organisent à chaque étape de leur évolution le matériel hétérogène issu, entre autres, des processus de grammaticalisation. De ce point de vue, on devrait sans doute renoncer à l'idée qu'il puisse y avoir une iso-

morphisme parfait entre les constructions mentales des locuteurs et une réalité nécessairement multiforme, et que le modèle le plus efficace est sans doute celui qui prévoit que la classification des objets linguistiques se fait de manière non discrète autour d'un centre prototypique, ce qui n'implique pas qu'on doive exclure de l'analyse les éléments périphériques. Je prends l'exemple des éléments de composition d'origine grecque ou latine qui apparaissent dans beaucoup de langues européennes, et en particulier des éléments finaux. Quelques exemples pris à l'italien sont donnés en (57) :

- (57) -ficio
calzaturificio 'usine à chaussures'
zuccherificio 'usine à sucre'
- fono
germanofono 'germanophone'
italofono 'italophone'
ispanofono 'hispanophone'
- geno
cancerogeno 'cancérogène'
schiumogeno 'moussant, qui produit de la mousse'
- poli
calciopoli 'scandale lié au football'
sanitopoli 'scandale lié à la santé'

Ces éléments font partie d'un ensemble bien connu pour lequel ont été proposées plusieurs étiquettes (« affixoïdes », cf. Migliorini 1963 ; « semi-mots », cf. Scalise 1990 ; 1994 ; « confixes », cf. Sgroi 2003, etc.). Il s'agit d'un ensemble très hétérogène, dont apparemment les seuls traits communs sont le fait d'apparaître dans des lexèmes complexes, d'être typiquement (mais pas nécessairement) d'origine latine ou grecque et d'être communs dans la langue technique et scientifique. Typiquement, il s'agit

d'un domaine dans lequel on retrouve des éléments à différents degrés de grammaticalisation. En particulier, pour les éléments de (57) – ainsi que pour certains autres, comme *-cida*, *-logo*, *-voro*, etc. – je ne vois pas de véritables arguments, si ce n'est des arguments étymologiques, pour distinguer ces éléments des autres affixes de l'italien. Du point de vue formel, cela est même préférable, puisque, de cette manière, nous pouvons nous libérer du problème de la voyelle de liaison ([i] pour *-ficio*, [o] pour *-fono*, *-geno* et *-poli*), dont le statut est peu clair, et nous pouvons simplement considérer qu'elle fait partie de la représentation de la règle : nous avons donc quatre suffixes [ifitʃo], [ɔdʒeno], [ɔfono] et [ɔpoli], qui, comme beaucoup d'autres suffixes de l'italien, commencent par voyelle, causent l'effacement de la voyelle finale du mot auquel ils se lient et portent l'accent¹⁰³. Certains (cf. Corbin 2001) considèrent que la distinction entre les affixes et les éléments en question est sémantique : un affixe véhicule typiquement une signification grammaticale (relationnelle), alors qu'un mot véhicule une signification de type lexical, ce qui serait le cas de *-ficio*, etc. Cependant, la distinction entre signification grammaticale et signification lexicale n'est pas facile à tracer, et il est bien connu que le même sens peut être exprimé par des affixes dans une langue et par des unités lexicales autonomes dans une autre, ou même par des éléments différents dans la même langue. Pensons, par exemple, aux sens liés à la dimension ou à l'évaluation qui sont exprimés dans beaucoup de langues par les affixes évaluatifs (Grandi 2002). Enfin, du point de vue morphologique, il existe en effet quelques cas dans lesquels les éléments en question n'ont pas la forme que j'ai indiquée (cf. *beneficio* 'bénéfice', *megafono* 'porte-voix'), mais c'est un problème qu'on retrouve avec d'autres suffixes (cf. *solubile* 'soluble', *ribellione* 'rébellion'), simplement parce que d'autres suffixes aussi ont eu différents étapes de grammaticalisation, et que tous les lexèmes ne sont pas rentrés dans le lexique d'une langue simultanément. Il existe, en revanche, de bons arguments morphologiques pour considérer que ces éléments sont des suffixes. D'une part, les lexèmes qui les comportent rentrent souvent dans les mêmes relations que des lexèmes suffixés 'canoniques'. Certains d'entre eux, par exemple, apparaissent dans des lexèmes préfixés en *anti-* en alternance avec des noms non suffixés

103 Une autre hypothèse serait de considérer que les suffixes en question sélectionnent systématiquement un thème de leur base qui se termine par [i] ou par [o]. Du point de vue théorique, je ne vois pas une vraie différence de nature entre cette explication et celle que je propose. Du point de vue empirique, il est préférable d'attribuer la voyelle aux suffixes, qui sont en nombre limité, plutôt qu'aux bases, qui sont potentiellement infinies.

ou des adjectifs de relation typiques (cf. la discussion en 2.5.2 et les exemples présentés en (17) ci-dessus :

(58)	anti-schiuma		antischiumogeno	‘anti-mousse’
	anti-ansia	anti-ansioso	antiansiogeno	‘anti-anxiété’

Le fait que certains d’entre eux (par exemple, ceux en *-fono* en (57)) soient liés à des bases qui ne sont pas des formes autonomes n’est pas non plus un problème, puisque ces mêmes bases se retrouvent dans des dérivés typiques (cf. *germanofono* – *germanista* ‘germaniste’, *ispanofono* – *ispanico* ‘hispanique’), et que ces éléments peuvent aussi être liés à des formes autonomes (cf. *dialettofono* ‘dialectophone’, *slovenofono* ‘slovénophone’). Il faut donc considérer ces éléments comme des suffixes à part entière, c’est-à-dire des éléments qui ont accompli leur parcours de grammaticalisation d’éléments lexicaux autonomes en grec ou latin à exposants de règles de construction de lexèmes, et pour certains desquels on a pu voir suivre ‘en direct’ l’évolution (par exemple, l’évolution sémantique de *-poli* de ‘ville’ à ‘scandale’ au début des années 1990).

La grammaticalisation peut ne pas être interprétée uniquement comme un concept diachronique. Il n’est pas aberrant d’imaginer que plusieurs des phénomènes qui apparaissent dans nos langues dérivent de la grammaticalisation de capacités cognitives plus générales (c’est d’ailleurs une des idées de base de plusieurs théories linguistiques formelles, comme la Théorie de l’Optimalité). Dans Fradin *et al.* (2009), nous avons montré que certains procédés morphologiques à l’œuvre dans les langues peuvent être considérés comme ‘extragrammaticaux’, dans la mesure où ils ne recourent pas aux moyens qui sont grammaticalisés dans une langue, mais font appel à des capacités cognitives plus globales (cf. 2.7.6). Font partie de ces procédés, par exemple, les différents langages secrets, le verlan, etc. Un autre exemple, me semble-t-il, est celui des composés. Un leitmotiv des études sur la composition est que ce procédé est le moyen de construction de lexèmes le plus répandu dans les langues et qu’il existe dans toutes les langues naturelles (cf. Greenberg 1963 : 92 ; Dressler 2006 : 23 ; Guevara & Scalise 2008). En réalité, ce qui est peut-être universel est la capacité des êtres

humains à relier deux lexèmes sémantiquement par simple juxtaposition¹⁰⁴. La composition, en revanche, est un concept principalement grammatical car tous les types de composition ne sont pas représentés dans les langues, ni toutes les combinaisons catégorielles, la manière dont les éléments se disposent dans un composé varie de langue à langue, etc. Si dans toutes les langues on observe de vrais procédés de composition, ou si dans certains cas c'est cette capacité cognitive qui est à l'œuvre n'est pas facile à établir (cf. Bauer 2006 : 721 ; 2009). Distinguer différents types de composés et les classer n'est pas en soi explicatif ; essayer d'établir comment des capacités cognitives propres aux êtres humains se grammaticalisent dans les systèmes linguistiques peut l'être.

104 Et peut-être pas seulement des êtres humains. Un exemple très connu dans la littérature sur l'apprentissage linguistique par les primates est celui d'un singe qui, pour indiquer un canard, était capable d'unir les deux signes qui pour lui signifiaient 'eau' et 'oiseau' (cité par exemple dans Kennally 2007 : 42-43).

Fabio Montermini

Partie 3

La morphophonologie de l'italien dans une perspective lexicaliste : Etudes de cas

if the trappings of the grammar are indeed as varied and complex as contemporary theories suggest, why is it that no one but a few professional linguists seem to have even the slightest inkling of any of it?

[B.L. Derwing, R. Skousen, *Morphology in the mental lexicon: A new look at analogy* (1989 : 58)]

3.1 Introduction

Dans la partie précédente j'ai esquissé un modèle de la morphologie et de son interaction avec le lexique fondé sur l'influence réciproque que les éléments qui le constituent exercent les uns sur les autres et dans lequel j'ai suggéré que la compétence morphologique, et peut-être linguistique plus en général, émerge à partir des généralisations que les locuteurs sont capables d'opérer à partir du lexique mémorisé. Dans cette dernière partie, je vais illustrer le modèle en question par deux études de cas concernant l'italien. En premier lieu, je m'intéresse à l'assignation de l'accent primaire en italien, une question récurrente dans les études de phonologie de l'italien.

Alors que quelques auteurs sont persuadés qu'il est possible d'identifier des principes purement phonologiques pour l'assignation de l'accent en italien, pour d'autres il s'agit d'un phénomène principalement lexical. Naturellement, ce travail s'inscrit dans cette deuxième lignée, et essaie de donner une réponse à la question de comment l'accent est assigné à des mots nouveaux ou inconnus, et donc de dégager des principes qui rendent compte du fonctionnement réel de la compétence linguistique. La réponse qui sera proposée est que la référence à des unités phonologiques et prosodiques est insuffisante, au moins dans certains cas, pour établir la structure accentuelle d'un mot, et que l'identification d'une unité plus abstraite, un lexème, à laquelle le mot candidat à l'assignation de l'accent est relié. Dans cette perspective, parmi les principes qui déterminent l'assignation de l'accent il y a le voisinage lexical d'un mot, ce qui relie le traitement proposé à l'idée plus générale du poids lexical comme moteur pour la construction de la compétence linguistique.

Le deuxième phénomène que je prends en considération est plus proprement morphologique et relève de la morphologie constructionnelle. En particulier, je me penche sur la segmentation des mots morphologiquement complexes, et sur l'identification des unités qui les composent, les affixes, d'un côté, et les unités qui correspondent aux lexèmes, thèmes ou racines, de l'autre, et de la variation de forme de celles-ci. L'idée fondamentale est que les lexèmes et les affixes sont des unités de nature différente, des unités lexicales pour les premiers, et des exposants de constructions morphologiques pour les seconds, et que, par conséquent, leur représentation phonologique et leur variation éventuelle peuvent être traitées de manière différente. En particulier, je suggère que la forme phonologique des affixes peut être traitée en termes de contraintes pesant sur la forme de surface que les dérivés doivent avoir, et que cette forme est le fruit d'un calcul qui tient compte de l'interaction entre ces contraintes et la forme phonologique de la base sur laquelle la construction morphologique opère. Dans cette perspective, l'attribution précise des segments qui apparaissent dans l'output d'une construction morphologique à la base ou à l'afixe cesse d'être un enjeu, un effet positif, qui nous permet de sortir d'une optique strictement segmentale dans l'analyse de la morphologie constructionnelle. Plus précisément, je m'intéresse, dans cette deuxième section à la construction d'adjectifs dénominaux, en particulier à partir de noms propres et toponymes, par moyen du suffixe *-(i)ano*, et à la

présence / absence d'un segment semivocalique [j] dans ces dérivés. Les données sur lesquelles s'appuient les deux traitements proposés dans cette section sont tirées, respectivement, d'une série de tests que j'ai effectués sur des locuteurs natifs italiens afin de mesurer leurs intuitions sur un certain nombre de cas, et sur un corpus de mots dérivés en *-(i)ano* non attestés dans les dictionnaires que j'ai recueillis. Dans les deux cas, les données présentées ont pour but de démontrer que, pour établir des principes de fonctionnement de la compétence linguistique, le recours à des données dont on sait qu'elles ne sont pas mémorisées dans le lexique est plus efficace, et que, au contraire, le lexique existant n'est pas le meilleur point d'observation pour ces mêmes principes.

3.2 L'accent primaire de mot en italien

3.2.1 Présentation

Une question récurrente dans la littérature sur la phonologie de l'italien concerne l'assignation de l'accent primaire. Alors qu'un certain nombre de chercheurs défendent l'idée qu'il est possible d'identifier des règles pour l'assignation de l'accent primaire en italien (cf. entre autres Den Os & Kager 1986 ; Sluyters 1990 ; Jacobs 1994 ; D'Imperio & Rosenthal 1999 ; Roca 1999 ; Guertzoni 2000 ; Morén 2001), d'autres considèrent, plus ou moins explicitement, que l'accent est marqué lexicalement pour les mots de l'italien (cf. Muljačić 1969 ; Burzio 1994 ; Bafle 1999 ; Marotta 1999 ; Krämer 2009b ; 2009c). En faveur de l'hypothèse de la spécification lexicale on peut citer le fait que l'accent principal de mot a une valeur distinctive en italien, et donc que, pour au moins une partie des mots de cette langue, il n'est pas assigné par règle :

- | | | |
|-------------------|---------------|---------------|
| (1) <i>cápito</i> | <i>capító</i> | <i>capitó</i> |
| 'j'arrive' | 'compris' | 'il arriva' |

En faveur de la deuxième hypothèse on peut citer le fait qu'un locuteur d'italien, lorsqu'il se trouve face à un mot nouveau et / ou inconnu, est généralement en mesure

d'en prévoir la position de l'accent primaire, ou au moins d'émettre des hypothèses à cet égard. En réalité, les jugements des locuteurs ne sont pas univoques à cet égard. Dans un test de lecture de pseudo-mots par des locuteurs italiens, pour des mots trisyllabiques comportant trois syllabes ouvertes (par exemple, *frunaco*, *fiesova*, *frudalo*), Krämer (2009b ; 2009c : 167-177) a observé une répartition de 55,1% vs. 44,9% entre accent sur la pénultième et sur l'antépénultième syllabe. Même si on peut s'interroger sur la valeur réelle des données recueillies par Krämer de manière assez informelle et sur un échantillon réduit de formes et de locuteurs, cela montre qu'au moins pour un certain nombre de mots les locuteurs hésitent et ne sont pas en mesure de donner des réponses univoques. Krämer attribue cette incertitude à la nature lexicale de l'accent de l'italien et observe correctement, dans les termes optimalistes qui sont les siens (Krämer 2009b) :

the lexical nature of the system causes the same problem for the learner as for the linguist: the learner has no data that support unambiguous ranking arguments, hence, many of the markedness constraints that determine stress remain unranked.

En effet, l'incertitude que les locuteurs montrent dans un certain nombre de cas représente un argument fort contre l'idée qu'il existe des principes d'ordre phonologique pour l'assignation de l'accent en italien, ou du moins que la phonologie seule en soit responsable. Comme nous le verrons ci-dessous, pour le lexique attesté l'identification d'un schéma accentuel comme celui par défaut n'offre aucun avantage empirique et en même temps pose plusieurs problèmes d'ordre théorique. Si l'accent est principalement un fait lexical, l'assignation d'un schéma accentuel aux mots nouveaux et / ou inconnus peut être soumise au même type de contraintes qu'on a identifiées précédemment pour la construction et l'interprétation de nouveaux mots, et en particulier à l'action conjointe de contraintes purement phonologiques et de contraintes d'ordre lexical. Parmi ces dernières le poids du lexique existant, et la similarité entre le mot cible et d'autres mots dans le lexique et la densité de son environnement jouent un rôle essentiel¹⁰⁵. Observons que dans le cas où on faisait l'hypothèse que

105 Cf. Burani & Arduino (2004) pour une expérience psycholinguistique qui confirme le rôle de l'environnement linguistique sur l'assignation de l'accent, et Eddington (2000) pour un modèle de l'assignation de l'accent fondé sur l'analogie appliqué à l'espagnol.

l'accent est assigné par règle pour un certain nombre de mots, pour les mots qui seraient réguliers de ce point de vue l'effet serait exactement le même : des mots semblables tendent à être encore plus semblables, y compris en ce qui concerne l'accent.

3.2.2 L'algorithme accentuel de l'italien

En général, pour l'identification d'une règle d'assignation de l'accent primaire en italien on prend comme base l'algorithme accentuel latin (mais cf. Vincent 1988 pour quelques observations sur l'assignation de l'accent dans cette langue) :

(2) a.

- accent sur l'antépénultième, s'il y en a une et si la pénultième est légère
- accent sur la pénultième, s'il y en a une

b. réx	'roi'	
rósa	'rose'	
romānus	'romain'	* rómānus
cucúrri	'je courus'	* cúcurri
pópŭlus	'peuple'	*popŭlus

La principale différence entre l'italien et le latin, dans le domaine syllabique, est la perte de la valeur distinctive de la longueur vocalique. En italien, une syllabe lourde coïncide dans tous les cas avec une syllabe fermée. Un algorithme accentuel qui peut être proposé pour l'italien est donc une version réduite de celui proposé en (2) (cf. Mioni 1993 ; Bertinetto & Loporcaro 2002) :

(3) a.

- accent sur l'antépénultième, s'il y en a une et si la pénultième est ouverte
- accent sur la pénultième, s'il y en a une

b. ré 'roi'

rósa	'rose'
áncora	'ancre'
conténto	'heureux'

L'attraction de l'accent par une syllabe antépénultième lourde en italien peut difficilement être remise en doute. Premièrement, elle est amplement confirmée par l'observation statistique (cf. par exemple Kenstowicz & Zuraw 2002). Parmi les mots qui peuvent être identifiés comme authentiquement italiens du point de vue phonotactique (en gros, ceux qui se terminent par une voyelle, cf. 3.3.4), ceux qui présentent une structure accentuelle de type $'\sigma_H\sigma$ ¹⁰⁶ sont une petite minorité, et pour la plupart il s'agit d'une accentuation de type (pseudo)savant qui mime une structure prosodique alloglotte (par exemple grecque, cf. Rohlfs 1966 : 445 ; Loporcaro 1996 : 162, n. 16)¹⁰⁷. En deuxième lieu, le schéma $\sigma'_H\sigma$ est encore largement prépondérant dans l'assignation de l'accent à des mots nouveaux et / ou inconnus (par exemple des sigles) (cf. *Anúscá* 'Association nationale des employés de l'état civil' ; *Unéscó*).

Statistiquement, cependant, l'italien montre aussi une très forte tendance à avoir une accentuation paroxytonique dans tous les cas, indépendamment du poids de la pénultième. Plusieurs auteurs ont proposé des estimations de la distribution des schémas accentuels en italien (cf. Mancini & Voghera 1993 et Krämer 2009c : 161 pour un panorama), et la proportion de paroxytons semble varier entre 70% et 80%. Je donne des chiffres plus complets en (4)¹⁰⁸ :

106 J'indique avec L une syllabe ouverte, avec H une syllabe fermée et avec σ une syllabe indépendamment de son degré d'ouverture.

107 Au total, j'ai repéré treize mots qui présentent cette structure en italien : les toponymes *Ágosta*, *Lépanto*, *Lévanto*, *Lévanzo*, *LépiZZa*, *Ótranto*, *Tárantó*, un nom de fleuve, *Ófanto*, plus les noms *árista* 'échine de porc', *canéderlo* ('knödel'), *finferlo* 'chanterelle', *mándorla* 'amande' (et *mándorlo* 'amandier'), *póliZZa* 'contrat d'assurance', ainsi que *ácanto* ('acanthé'), pour lequel les dictionnaires recommandent la prononciation paroxytone. Parmi ces mots, dans un grand corpus d'italien écrit (CORIS/CODIS, 120M de mots, http://corpora.dslo.unibo.it/coris_ita.html) *póliZZa* occupe le rang le plus élevé (le 7 196^e) avec 1 281 occurrences.

108 Le décalage entre les chiffres donnés dérive, outre que des différentes sources considérées, du fait que certaines sources ne prennent en compte que des types (exposants de lexèmes) alors que d'autres prennent en compte des tokens (formes fléchies). Ce dernières sont, bien entendu, plus précises, dans la mesure où les différentes formes fléchies d'un mot peuvent présenter des schémas accentuels différents.

(4) proparoxytons	16/30%
paroxytons	70/80%
oxytons	2/3%

En réalité, dans le lexique attesté il existe plusieurs facteurs externes, non phonologiques, qui peuvent influencer la position de l'accent d'un mot. Il est connu, par exemple, que la plupart des suffixes de l'italien construisent des mots paroxytons ; de même, on peut considérer que beaucoup d'emprunts sont paroxytons parce qu'ils reproduisent le schéma accentuel dans la langue d'origine (cf. *computer* [kom'pjuter] 'ordinateur', *corrida* [kor'ri:da]¹⁰⁹) ; enfin, plusieurs mots conservent simplement le schéma accentuel latin, même indépendamment de la longueur vocalique :

(5) amico	< amicus	'ami'
fedele	< fidelis	'fidèle'

Si on prend en considération tous ces facteurs, il faut considérer que l'algorithme exprimé en (3) ne vaut que pour les mots non dérivés et pour le lexique autochtone. Malheureusement, comme on l'a vu dans la partie précédente, aucune des deux notions n'est claire pour ce qui concerne l'italien (et vraisemblablement toutes les autres langues). En effet, la plupart des tentatives d'identifier un algorithme d'assignation de l'accent en italien partent de l'observation du lexique attesté. Or, comme on l'a vu, essayer de dégager des principes grammaticaux à partir d'une entité composite, stratifiée et historiquement déterminée comme le lexique n'est globalement pas une bonne idée. Comme dans les autres cas qu'on a étudiés, la forme finale des mots est en réalité le fruit d'un ensemble de facteurs, éventuellement en compétition entre eux.

Pour reprendre de manière plus détaillée les données, en italien chacune des trois dernières syllabes d'un mot peut porter l'accent :

109 La transcription des mots italiens, en particulier des emprunts, s'appuie sur ma propre prononciation, et a été vérifiée, en cas de doute, dans Canepari (2008).

	3 syllabes ou plus		2 syllabes		1 syllabe
oxyton	virilitá 'virilité'	povertá 'pauvreté'	paté 'paté'	virtú 'vertu'	ré 'roi'
paroxyton	amíco 'ami'	conténto 'heureux'	páne 'pain'	ténda 'tente'	
proparoxyton	sólido 'solide'	mándorla 'amande'			

Tableau 1 : schémas accentuels de l'italien

La seule généralisation qu'il est possible de tirer du Tableau 1, qui s'appuie sur le lexique attesté, est qu'en italien l'accent ne peut pas remonter plus loin que la syllabe antépénultième¹¹⁰. Certaines des configurations exemplifiées, cependant, sont marquées. J'ai déjà parlé du type *mándorla* (note 3), qui est représenté par un nombre réduit de mots, la plupart desquels ne sont pas particulièrement fréquents, et qui est peu (ou pas) disponible pour être étendu à des mots nouveaux. Pour ce qui concerne les mots oxytons, lorsqu'il s'agit de noms ou d'adjectifs ils sont tous invariables et contiennent un suffixe, plus ou moins transparent (le seul véritablement productif en italien contemporain est *-ità*), ou bien sont des emprunts (cf. *bordeaux* [bor'dɔ]). On peut considérer donc que le schéma oxyton n'est pas non plus disponible : lorsqu'ils sont face à un mot nouveau et / ou inconnu les locuteurs ne choisissent jamais ce type d'accentuation. Les schémas accentuels qui sont actuellement disponibles en italien sont donc ceux qui sont grisés dans le Tableau 1, et l'oscillation concerne en particulier les mots plus que dissyllabiques qui ont une syllabe pénultième ouverte. En d'autres termes, si une règle d'accentuation en italien existe, elle doit être capable de nous dire lequel entre *amíco* et *sólido* est construit par règle et lequel a un accent marqué lexicalement.

110 Les seuls mots qui présentent un accent plus loin que l'antépénultième sont certains groupes clitiques (cf. *telefonagli* 'téléphone-lui') et quelques 3^e personnes du pluriel du présent de l'indicatif (cf. *telefonano* 'ils téléphonent'), pour lesquelles on a de bonnes raisons de considérer le suffixe *-no* comme un clitique, du moins du point de vue phonologique (cf. Spagnoletti & Dominicy 1992).

3.2.3 Les traitements de l'accentuation en italien

Les positions rencontrées dans la littérature spécialisée à propos de l'assignation de l'accent en italien peuvent, en gros, être réparties en deux groupes :

- i. il existe un schéma accentuel non marqué unique, et tous les mots qui ne le respectent pas sont marqués comme exceptionnels dans le lexique. Le schéma non marqué est tantôt identifié avec le paroxyton (Den Os & Kager 1986 ; Sluyters 1990 ; Jacobs 1994 ; D'Imperio & Rosenthal 1999), tantôt avec le proparoxyton (Guerzoni 2000 ; Morén 2001).
- ii. il n'existe pas de schéma accentuel non marqué ; le type accentuel auquel un mot appartient est spécifié dans son entrée lexicale (Burzio 1994 ; Bafle 1999 ; Marotta 1999 ; Krämer 2009c).

L'intérêt principal des explications de type (i) est qu'elles permettent de respecter, à la fois, le principe d'uniformité de l'algorithme, selon lequel une langue possède un seul type de pied (en italien typiquement un trochée dissyllabique) et le principe de binarité, selon lequel un pied est toujours binaire (c'est-à-dire comporte deux syllabes ou deux mores). La plupart des chercheurs qui adhèrent à une explication de type (ii), au contraire, proposent d'élargir l'inventaire métrique de l'italien, en incluant un pied ternaire, à savoir le dactyle (x . .), qui serait associé aux mots comme *sólido* (mais cf. Krämer 2009c pour une explication de type lexical fondée sur l'existence d'un seul pied binaire). Les explications de type (ii), cependant, ne peuvent pas faire l'impasse d'expliquer comment l'accent est assigné pour les mots nouveaux et / ou inconnus, un domaine, on l'a vu, dans lequel les locuteurs ont souvent des intuitions uniformes et précises. Clairement, pour les mots morphologiquement complexes, et en particulier suffixés, le schéma accentuel est assigné par la règle de suffixation elle-même :

(6) <i>távolo</i>	→	<i>tavolíno</i>	'table' / 'petite table'
<i>átomo</i>	→	<i>atómico</i>	'atome' / 'atomique'

Dans la plupart des cas, cela est vrai même pour les séquences qui sont homophones d'un suffixe (cf. *limóne* 'citron', ou la prononciation, fréquente, de mon nom de famille comme *Montermíni*).

3.2.4 L'accentuation par défaut en italien

La question de l'accentuation en italien reste donc ouverte pour les mots simples. Les mots avec une penultième fermée ne posent pas vraiment de problème, dans la mesure où les exceptions à la généralisation $\sigma'\sigma_H\sigma$, comme on l'a vu, sont rares, et pour la plupart anecdotiques. Pour ce qui concerne les mots ayant une structure $\sigma\sigma_L\sigma$, le lexique existant ne nous permet pas vraiment de proposer une généralisation. Quelques auteurs pensent pouvoir justifier le fait que la structure $\sigma'\sigma_L\sigma$ serait celle non marquée en italien sur la base de la diachronie (par exemple Repetti 1993 : 183, sur la base, pratiquement, du seul exemple *varíce* < lat. *várīce(m)*, 'varice'). En effet, la tendance à créer des paroxytons à partir de mots proparoxytons a probablement été plus forte dans des étapes antérieures de l'évolution de l'italien (cf. les formes de l'italien ancien, toujours citées par Repetti 1993 *umíle* 'humble', *oceáno* 'océan' vs. italien actuel *úmíle*, *océano*). Dans l'italien contemporain, cependant, la situation est beaucoup moins claire. À partir d'un certain nombre de ressources qui pointent les « erreurs » d'accentuation les plus communes en italien¹¹¹, j'ai pu établir un petit corpus de 148 mots pour lesquels les deux types de prononciations sont communs, ou du moins attestés (la liste est donnée en appendice à cette section). De ces mots, 89 (60%) ont un accent étymologique paroxyton¹¹². De manière encore plus intéressante, si nous regardons l'accentuation prépondérante¹¹³ pour les 148 mots recensés, nous observons la distribution suivante (dans la suite de cette section, S=accentuation proparoxytone, P=accentuation paroxytone) :

111 En particulier, une liste d'erreurs incluse dans Zingarelli (1997), et des sites Web (<http://www.achyra.org/cruscate/index.php>, <http://www.manuscritto.it/pronuncia.html>).

112 Par 'accent étymologique' je veux indiquer l'accent du mot latin (ou éventuellement d'une autre langue) duquel le mot italien est dérivé, ou l'accent de la base, dans le cas de mots dérivés morphologiquement (comme dans *vapóre* → *evápora*_{1SG PRES IND}).

113 L'accentuation prépondérante a été établie sur la base de ma propre intuition et de dictionnaires de prononciation (en particulier, *DOP* et Canepari 2008).

accentuation		total
prépondérante	étymologique	
P	P	52
P	S	17
S	P	37
S	S	42

Tableau 2 : types d'accentuation dans un corpus de mots italiens

Les données observables montrent donc qu'imaginer que dans l'histoire de l'italien il y ait eu une tendance globale de glissement d'accent d'un schéma proparoxyton vers un schéma paroxyton n'est pas justifié. Des glissements s'observent dans les deux sens, avec un léger avantage pour le passage de paroxyton vers proparoxyton. Rohlfs (1966 : 441) observe que cette tendance trouve probablement son origine dans la « *lingua delle persone istruite, in seguito a una specie di pronuncia ipercorretta* ». Quelles qu'en soient l'origine et la motivation sociolinguistique, nous pouvons considérer (comme le font par exemple Marotta 1999 et Bertinetto 1999 : 83) que l'italien se trouve dans une phase dans laquelle le schéma proparoxyton est en train de gagner du terrain en tant que patron accentuel non marqué. D'autres indices semblent confirmer cette observation. Nous pouvons, par exemple, observer le paradigme verbal. Il existe un certain débat dans les études sur l'italien si les verbes doivent être considérés comme partageant les mêmes principes d'accentuation que les noms et les adjectifs (cf. Roca 1999 ; Krämer 2009b). Il y a de bonnes raisons, en réalité, de considérer qu'il ne faut pas faire de distinction entre les deux. La plupart des formes verbales possèdent un schéma accentuel qui est spécifié par la règle de flexion, mais dans le paradigme des verbes de la classe à infinitif en *-are* (la plus nombreuse et productive en italien), il existe neuf cases dans lesquelles l'assignation de l'accent se fait de manière relativement libre (toujours à l'intérieur de la fenêtre comprenant l'antépénultième et la pénultième). Le détail est donné dans le Tableau 2 pour les verbes *COMPLICARE* ('compliquer') et *FATICARE* ('se fatiguer') :

	1	2	3	4	5	6
PRÉS IND	cómplico fático	cómplichí fatíchi	cómplica fatíca	complicíamo fatíchiamo	complicáte faticáte	cómplicano fatícano
PRÉS SUBJ	cómplichí fatíchi	cómplichí fatíchi	cómplichí fatíchi	complicíamo fatíchiamo	complichíate fatichiáte	cómplichino fatíchino
IMPÉR			cómplica fatíca			

Tableau 3 : schémas accentuels des verbes italiens en *-are*

Les principes d'accentuation des mots italiens indiqués ci-dessus semblent en effet valoir aussi bien pour les verbes que pour les noms et les adjectifs. Par exemple, Kenstowicz & Zuraw (2002), sur un corpus d'environ 1.500 verbes du type illustré dans le Tableau 3, ont relevé les chiffres suivants :

	penultième	
	fermée	ouverte
paroxytons	963	307
proparoxytons	–	251

**Tableau 4 : accentuation des verbes italiens
(d'après Kenstowicz & Zuraw 2002)**

Un dernier argument en faveur du fait qu'il existe un même principe d'accentuation pour les verbes et pour les mots des autres catégories est le fait que les verbes dérivés par préfixation ou par conversion d'un nom ou d'un adjectifs, et qui appartiennent dans la grande majorité des cas à la classe à infinitif en *-are*, aux neuf cases du paradigme illustrées au Tableau 3 maintiennent pour la plupart le même schéma accentuel que leur base :

(7) rótolo	→	arrótolo _{1SG PRÉS IND}	'rouleau' / 'enrouler'
caténa	→	incaténo _{1SG PRÉS IND}	'chaîne' / 'enchaîner'
spázzola	→	spázzolo _{1SG PRÉS IND}	'brosse' / 'brosser'

Justement, l'analyse de quelques formes verbales ainsi que d'autres formes nous permet, dans certains cas, d'observer ce que nous pouvons considérer comme de 'purs' principes phonologiques à l'œuvre. D'Imperio & Rosenthal (1999), par exemple, affirment, quelque peu arbitrairement, qu'un mot hypothétique avec un accent sous-jacent sur la syllabe pré-antépénultième apparaîtrait sûrement à la surface avec un accent paroxyton. Or, de rares mots de ce type en réalité existent et le patron accentuel que les deux auteurs préconisent est le moins acceptable selon le jugement des locuteurs natifs. Il s'agit, par exemple, de verbes dérivés de mots étrangers plus que dissyllabiques qui, en italien, ont un accent proparoxyton. En général, comme dans les cas exemplifiés en (7), aux neuf formes mentionnées plus haut les verbes dérivés de mots étrangers préservent l'accent de leur base :

- (8) *clacson* ['klakson] → *clacsono*_{1SG PRES IND} ['klaksono] 'klaxon' / 'klaxonner'
shaker ['ʃeker] → *shakero*_{1SG PRES IND} ['ʃekero] 'shaker' / 'mixer avec un shaker'

Lorsqu'ils dérivent de mots se terminant en consonne, ces formes comportent l'adjonction systématique d'une syllabe (cf. *clac.son* → *clac.so.no*). Par conséquent, dans le cas de verbes dérivés mots étrangers se terminant en consonne et prononcés avec une accentuation proparoxytone, comme *cellophane* ([tʃellofan]) ou *monitor* ([mɔnitor]), l'accent des formes en questions tomberait hors de la fenêtre accentuelle admise en italien. Lors d'un test informel, j'ai demandé à huit locuteurs natifs d'italien de conjuguer les verbes *INCELLOFANARE* ('envelopper dans du cellophane') et *MONITORARE* ('surveiller à travers un écran') à la 1^e personne du singulier du présent de l'indicatif. En même temps, je leur ai demandé de lire une phrase dans laquelle apparaissait l'adjectif composé *senegalo-guineano* ('sénégal-guinéen'). L'élément *senegalo*, dérivé de l'adjectif *senegalese*, lui-même dérivé de *Senegal* ([senegal]), présente en effet le même genre de problème. On avait donc affaire potentiellement à trois formes de surface, dont deux qui violaient la fidélité base-dérivé (*incellofano*, *incellofáno*, etc.) et une qui violait la fenêtre d'assignation de l'accent (*incellofano*, etc.). Les résultats ont été assez clairs et ont confirmé celle qui était mon intuition : quand on leur a demandé de classer les formes de surface possibles en ordre de préférence, les locuteurs ont indiqué la

forme paroxytone comme la moins bonne, et accordé une légère préférence à la forme proparoxytone contre celle qui violait la fenêtre accentuelle. Les résultats du test sont résumés dans le Tableau 5 (les différents visages indiquent les préférences des locuteurs) :

	fenêtre accentuelle	fidélité base-dérivé
☺ incellófano monítoro senégalo		*
☹ incéllofano mónitoro sénegalo	*	
☹ incellofáno monitóro senegálo		*

Tableau 5 : accentuation de quelques dérivés de mots étrangers en italien

Dans le cas des mots illustrés ci-dessus, le voisinage lexical plus proche suggère une accentuation phonologiquement non acceptable. Nous pouvons donc considérer qu'ici ce sont des principes purement phonologiques qui sont à l'œuvre, et, comme le montrent les résultats présentés, dans ces cas l'accentuation de type proparoxyton semble être préférée.

Le cas, que j'ai évoqué, de *Sénégal*, mérite qu'on s'y intéresse brièvement. La tendance à avoir un accent proparoxyton est plus forte avec des mots qui ont un aspect 'étranger', c'est-à-dire, en gros, des mots qui se terminent en consonne. Cela vaut pour les « English-sounding words » (cf. Bertinetto & Loporcaro 2002), mais aussi pour des mots qui ont des origines différentes (9a), y compris des mots qui étymologiquement ne sont pas étrangers, comme des sigles ou des noms de famille italiens (9b) :

- (9) a. Ásterix ['asteriks]
Écuador ['ekwador]
mátador ['matador]

Górbaciov ['gɔrbatʃov]

Pásternak ['pasternak]

b. Bénetton ['benet:ɔn]

Bérlinguer ['bɛrlingwer]

Fíninvest ['fininvest]

Comme le montrent la plupart de ces exemples, et notamment ceux en (9a), dans ces cas la tendance à avoir un accent proparoxyton prime sur la fidélité avec l'accent dans la langue d'origine (qui, d'ailleurs, n'est pas nécessairement connu par les locuteurs). Cela peut être vrai aussi pour des mots anglais, comme *pérformance* ou *cómponent*. De plus, quelques-uns des exemples ci-dessus (*Pas.ter.nak*, *Be.net.ton*, *per.for.mans*, etc.) montrent que celui-ci est le seul cas d'accentuation, pour ainsi dire, 'productive' qui peut violer la généralisation $\sigma'\sigma_H\sigma$. Pour ce qui concerne le statut des mots que j'ai définis d'aspect étranger, il semblerait que la seule généralisation possible soit que le mot se termine par consonne, ce qui démontre, à mon sens, la supériorité de contraintes de type lexical : à ma connaissance aucune théorie phonologique ne saurait traiter une 'contrainte' d'assignation de l'accent qui ferait référence à la structure de la dernière syllabe d'un mot, indépendamment du fait qu'elle soit accentuée ou pas.

3.2.5 L'accentuation en italien : un test

Je présente ici les résultats d'une étude que j'ai effectuée sur un groupe particulier de mots en italien, les mots se terminant avec la séquence [inV], qui démontre, à mon avis, qu'il n'est pas possible de décrire les mécanismes d'assignation de l'accent en italien sur la base de principes uniquement phonologiques. En italien il existe fondamentalement trois groupes de mots qui présentent la séquence en question en finale :

1. les lexèmes construits au moyen de deux suffixes paroxytons homophones *-ino*, un diminutif (cf. *tavolo* 'table' → *tavolino* 'petite table') et un relationnel (cf. *Alpi* 'Alpes' → *alpino* 'alpin') ;

2. des lexèmes contenant quelques suffixes féminins proparoxytons, comme *-aggine* (cf. *stupido* 'bête' → *stupidággine* 'bêtise') ou *-itudine* (cf. *solo* 'seul' → *solitúdine* 'solitude') ;
3. des noms inanalysables dérivant de noms latins de la troisième déclinaison en *-men/-minis* (cf. *crímine* 'crime', *términe* 'terme').

De plus, dans l'histoire de l'italien, la plupart des mots empruntés se terminant par la séquence [inV] se sont vu assigner une accentuation de type paroxyton :

- (10) giardíno < fr. jardin
 'jardin'
- mandaríno < port. mandarim
 'mandarin'
- Pechíno < chin. Beijīng
 'Pékin'

Il semblerait, donc, que le schéma paroxyton fonctionne comme défaut pour les mots de ce type. Pourtant, comme on va le voir, les jugements des locuteurs sont parfois partagés, et cela est probablement dû à des contraintes qui ne sont pas d'ordre purement phonologique.

Pour vérifier cette hypothèse j'ai pris les noms et les adjectifs se terminant par [inV] dans un corpus d'italien parlé, le *LIP*¹¹⁴. Le choix d'un corpus oral a été dicté par le désir d'avoir une liste de lexèmes suffisamment fréquents, et que l'échantillon considéré ne contienne pas un nombre excessif de lexèmes rares ou d'usage limité. Au total, il y a dans le *LIP* 356 lexèmes se terminant par cette séquence. Dans le Tableau 6 je donne le total des mots pour chacune des classes de flexion nominales ou

114 Le *LIP* contient environ 15 000 tokens et 500 000 types. Pour homogénéité, je n'ai pris en compte que les noms et les adjectifs se terminant en la séquence en question ; je n'ai pas pris en compte les verbes se terminant en *-inare* (qui font, par exemple, [ino] à la première personne du singulier du présent de l'indicatif). Je n'ai pas pris en compte non plus les mots invariables en [ini] (uniquement des noms de famille, tous paroxytons), ni la forme fléchie *uómìni*, pluriel irrégulier de *UOMO* ('homme').

adjectivales selon le type d'accent ; dans le 'Tableau 6' je donne le total de mots pour chaque séquence finale possible et, pour des raisons qui seront claires immédiatement après, je partage les mots se terminant en [ine] selon que cette séquence apparaît dans une forme au singulier ou au pluriel (les mêmes données sont reproduites dans les Figures 1 et 1' :)

	P	S
invariables en [ino]	29	–
invariables en [ina]	40 ¹¹⁵	2
invariables en [ine]	3	2
noms (masc.) en [ino]~[ini]	119	5
noms (fém.) en [ina]~[ine]	77	8
noms (masc./fém.) en [ine]~[ini]	1	25
noms et adj. en [ino]~[ini]~[ina]~[ine]	45	–
Total	315	41

**Tableau 6 : distribution de l'accent des mots en [inV]
dans le LIP selon la classe**

	P	S
[ino]	193	5
[ina]	162	10
[ini]	165	30
[ine]	126	35
[ine] _{SG}	1	25
[ine] _{PL}	122	8

**Tableau 6' : distribution de l'accent des mots en [inV]
dans le LIP selon la séquence finale**

115 Parmi les invariables en [ina] apparaît *Ucraina* ('Ukraine'), qui peut recevoir les deux accentuations *Ucrâina* et *Ucrâina*, et a été comptabilisé parmi les paroxytons, le type d'accent indiqué comme le plus fréquent dans Canepari (2008), et qui me semble être tel, au moins dans mon entourage.

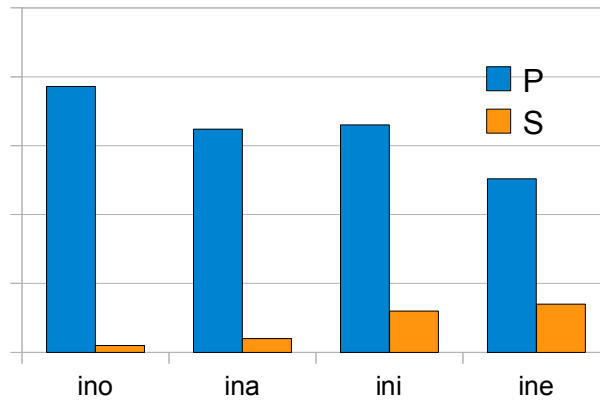


Figure 1

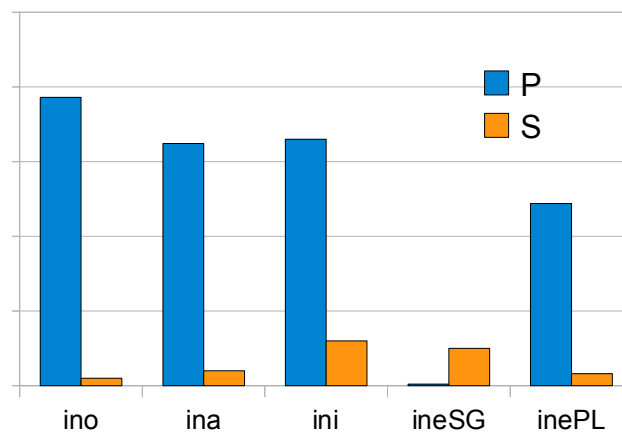


Figure 1'

Comme on le voit des données ci-dessus, la prévisibilité du schéma accentuel des mots se terminant en [ine] dépend de façon cruciale de facteurs morphologiques, comme le nombre exprimé par la forme en question, ce qui milite en faveur d'une

analyse lexicale de l'assignation de l'accent en italien. Pour aller plus loin, et pour raccrocher l'analyse au modèle émergent du lexique esquissé dans la partie 2, on peut proposer des considérations ultérieures. Si nous regardons, en effet, la consonne qui précède immédiatement la séquence [inV], le schéma paroxyton est minoritaire dans deux cas (avec les consonnes [m] et [dʒ]) et les deux schémas sont grosso modo équivalents avec la consonne [d] ; dans tous les autres cas le schéma paroxyton est majoritaire, voire exclusif. La totalité des données est résumée dans le Tableau 7 ; dans la Figure 2 sont mis en évidence les trois cas dans lesquels chacun des deux schémas est le plus représenté :

	t	l	r	tʃ	d	k	z	n	b	p	ts	m	dʒ	f	s	ʎ	v	g	ɲ
P	71	50	47	30	16	13	9	9	8	8	8	7	6	5	5	3	3	2	1
S	1	1	–	–	13	1	2	–	–	–	–	10	10	–	–	–	–	–	–

Tableau 7 : distribution des schémas accentuels selon le segment précédant la séquence [inV]

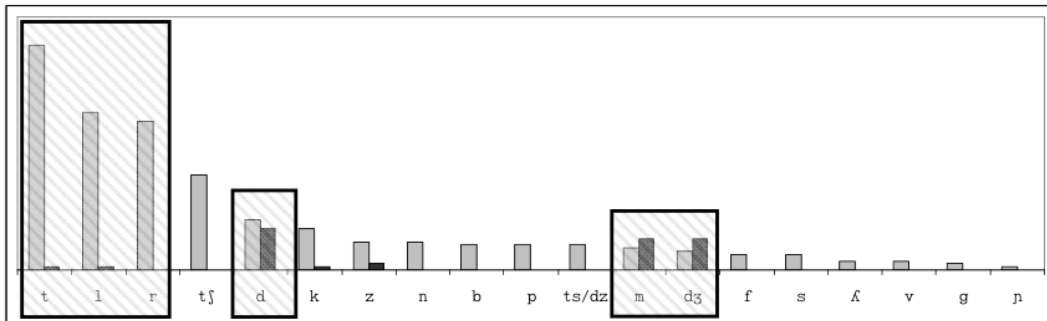


Figure 2

Quelques-uns des chiffres observés sont clairement influencés par la morphologie. Le grand nombre de formes en [tinV], [linV] et [tʃinV], par exemple, est clairement une conséquence des contextes les plus fréquents dans lesquels on rencontre le suffixe diminutif *-ino*, par exemple en combinaison avec d'autres suffixes évaluatifs (11a), ou dans la variante *-ino* lorsqu'il est attaché à une base se terminant par la sé-

quence [one] (11b) :

- (11) a. polpetta → polpettina
 'boulette' 'boulette-DIM'
- cartello → cartellino
 'panneau' 'étiquette'
- b. balcone → balconcino
 'balcon' 'balcon-DIM'

Pour vérifier l'incidence réelle de tous les facteurs identifiés, et de leur interaction, j'ai réalisé un test informel en demandant à huit locuteurs natifs (tous d'un âge compris entre 30 et 65 ans, de niveau d'instruction moyen ou supérieur et de provenance géographique variée, mais tendancielle du nord de l'Italie) de lire des mots inventés se terminant par les séquences *ino* et *ine* précédés des consonnes *t, l, r, d, m* et *g* (phonologiquement d₃). Aucun des sujets n'a eu à lire un couple de stimuli (chaque stimulus était présenté soit se terminant par *o*, soit se terminant par *e*), et aucune indication n'a été donnée quant au genre ou au nombre des mots présentés. Les résultats obtenus sont donnés dans le Tableau 8, dans lequel je mets en évidence les cas dans lesquels les sujets ont préféré l'accentuation paroxytones (cases en blanc), l'accentuation proparoxytone (cases en gris foncé), ou ont montré des incertitudes importantes (cases en gris clair) (les cases vides correspondent à des séquences redondantes ou phonotactiquement impossibles) :

prutino	prulino	prurino	prudino	prumino	prugino
prutine	pruline	prurine	prudine	prumine	prugine
croltino			croldino	crolmino	crolgino
croltine			croldine	crolmine	crolgine
afetino	afelino	aferino	afedino	afemino	afegino
afetine	afeline	aferine	afedine	afemine	afegine

Tableau 8 : résultats d'un test d'accentuation de pseudo-mots

Les résultats du test sont intéressants, notamment parce qu'ils mettent en évidence que l'assignation de l'accent ne peut pas se faire, ou pas toujours, sur une base purement phonologique. En particulier, me semble-t-il, ces résultats montrent que, dans les cas les moins immédiatement interprétables, l'assignation de l'accent passe par un calcul qui prend en compte des facteurs morphologiques et lexicaux, comme le genre et / ou le nombre, le lexème potentiel auquel une forme nouvelle peut appartenir, et vraisemblablement le nombre de voisins lexicaux de ce lexème. Pour exemplifier cette conclusion, je prends trois des pseudo-mots utilisés dans le test qui me semblent particulièrement éloquents : *prugine*, *croimine* et *prurine*. Pour le premier les locuteurs ont montré un certain degré d'incertitude ; pour le second ils ont massivement penché pour l'accentuation proparoxytone, et pour le dernier pour l'accentuation paroxytone. Dans les Tableaux 9'-9''' je mets en relation ces tendances avec les lexèmes desquels les pseudo-mots en question peuvent être des formes et avec le nombre de voisins (c'est-à-dire, de lexèmes présentant la même structure accentuelle et les mêmes quatre phonèmes en finale) que chacun de ces lexèmes a dans le *LIP*. Comme on le voit, dans les trois cas la corrélation entre cette donnée et le type d'accentuation choisi apparaît comme significative :

forme	lexème potentiel	valeurs gramm.	voisins
prugine	PRUGINA	N FEM PL	3 (REGINA 'reine')
	PRUGINO	A ¹¹⁶ FEM PL	1 (CUGINO 'cousin')
prúgine	PRUGINE	N FEM SG	7 (IMMAGINE 'image')
	PRUGINA	N FEM PL	1 (PAGINA 'page')
	PRUGINE	N MASC SG	1 (MARGINE 'marge')

Tableau 9' : lexèmes potentiels et leurs voisins pour le pseudo-mot *prugine*

116 Avec A j'indique aussi bien les adjectifs que les noms variables en genre à quatre formes.

forme	lexème potentiel	valeurs gramm.	voisins
crolmíne	CROLMINA	N FEM PL	1 (VITAMINA 'vitamine')
	CROLMINO	A FEM PL	–
crólmíne	CROLMINE	N FEM SG	–
	CROLMINA	N FEM PL	4 (FEMMINA 'femme')
	CROLMINE	N MASC SG	4 (FULMINE 'foudre')

Tableau 9'' : lexèmes potentiels et leurs voisins pour le pseudo-mot *ardmine*

forme	lexème potentiel	valeurs gramm.	voisins
pruríne	PRURINA	N FEM PL	11 (FARINA 'farine')
	PRURINO	A FEM PL	8 (CARINO 'mignon')
prúgíne	PRURINE	N FEM SG	–
	PRURINA	N FEM PL	–
	PRURINE	N MASC SG	–

Tableau 9''' : lexèmes potentiels et leurs voisins pour le pseudo-mot *prurine*

Tous les travaux qui se sont occupés de l'assignation de l'accent en italien concordent à dire qu'il est assigné à des « mots ». Cependant, comme on l'a vu tout au long de ce travail, il est nécessaire de définir clairement les objets qui correspondent à cette étiquette. On peut imaginer que les phonologues qui se sont penchés sur la question considèrent que le 'mot' pertinent dans ce contexte est le mot phonologique ; cependant, on a vu que cette notion n'est pas suffisante pour rendre compte de l'assignation de l'accent dans tous les cas. Les données illustrées ci-dessus montrent que d'autres facteurs, notamment des facteurs morphologiques et lexicaux doivent être pris en compte, par exemple des informations concernant le genre, le nombre ou la classe de flexion, et par conséquent le lexème dont le mot en question est potentiellement une forme. On peut donc considérer que les locuteurs reconstruisent un mot lexical, un lexème, pour procéder à l'assignation de l'accent ; l'assignation de l'accent ne peut par conséquent pas être considérée comme un phénomène purement phonologique, mais lexical, et le voisinage qui influence cette assignation n'est pas purement

segmental ou prosodique, mais lexical plus globalement.

Appendice : corpus de mots qui présentent des variations d'accent

(P= paroxyton, S = proparoxyton ; l'abréviation 1SG indique que le mot en question est un verbe et présente des variations d'accent à neuf personnes du paradigme, cf. 3.2.4 ci-dessus)

Prép. = P	episteme	rubrica	serotino
Etym. = P	epodo	scorpena	termite
	esecrare _{1SG}	surrogare _{1SG}	titubare _{1SG}
abbaino	Esopo	sutura	varice
abnegare _{1SG}	fustigare _{1SG}	tracimare _{1SG}	
abrogare _{1SG}	giacere	tralice	Prép. = S
adulare _{1SG}	imperare _{1SG}	voivoda	Etym. = P
aprico	impudico		
arrogio	inane	Prép. = P	abbacinare _{1SG}
bismuto	ippocastano	Etym. = S	asseverare _{1SG}
bolscevico	irrigare _{1SG}		Belice
bramito	isocolo	alacre	collaborare _{1SG}
caduco	lorica	anodino	colubro
centellinare _{1SG}	mollica	comparare _{1SG}	confutare _{1SG}
cervice	onagro	elevare _{1SG}	dissuadere
comminare _{1SG}	pedagogo	esegesi	edema
congrega	perequare _{1SG}	flogosi	educare _{1SG}
constatare _{1SG}	peritare _{1SG}	harakiri	efebo
cosmopolita	perorare _{1SG}	iconostasi	elaborare _{1SG}
demagogo	podere	lubrico	espletare _{1SG}
diegesi	poliptoto	matrice	evaporare _{1SG}
duodeno	pudico	pignorare _{1SG}	evitare _{1SG}
edile	quirite	scorbuto	fagocitare _{1SG}
edule	regime	separare _{1SG}	gratuito

La morphophonologie de l'italien

incavo	svalutare _{1SG}	collocare _{1SG}	mangano
indicare _{1SG}	tenebra	cuculo	melico
infervorare _{1SG}	valutare _{1SG}	dagherrotipo	metabasi
infido	zaffiro	diospero	metopa
irritare _{1SG}		egloga	neofita
isteresi	Prép. = S	elleboro	olibano
manticora	Etym. = S	enclisi	ossimoro
necrosi		eureka	patera
palpebra	allobrogo	Florida	ragano
permuta	apodosi	folaga	retore
permutare _{1SG}	archetipo	ilare	satrapo
perseverare _{1SG}	baggiolo	imitare _{1SG}	sinchisi
persuadere	baratro	impari	ubere
Salgari	bevuta	implicare _{1SG}	upupa
salubre	buccina	intersecare _{1SG}	utensile
sbellicare _{1SG}	callifugo	intimare _{1SG}	velodromo
sgattaiolare _{1SG}	circuito	ippodromo	

3.3 La représentation phonologique des affixes : une analyse du suffixe *-(i)ano*

3.3.1 Présentation : *-(i)ano* en italien

Alors que dans la section précédente j'ai pris en considération un phénomène qui concerne la représentation phonologique des lexèmes, dans celle-ci je m'occuperai plus directement de la représentation des exposants des règles de construction des mots, en m'intéressant en particulier au(x) suffixe(s) italien(s) issu(s) du latin *-anus*, qui présente(nt) de nombreux problèmes d'analyse, notamment en relation à la présence / absence apparente d'un segment semivocalique [j] entre la base et l'affixe.

Du point de vue diachronique la présence de ce segment dans certains dérivés ne constitue pas un mystère, puisqu'il dérive de la réanalyse d'adjectifs dérivés par *-anus* à partir de bases qui comportent un segment final [j]/[i], comme dans *Octavius* → *octa-*

vianus (cf. Tekavčić 1972 : 68-69 ; Rainer 2009). Déjà en latin, d'ailleurs un allomorphe *-ianus* s'était développé, qui pouvait s'attacher à des bases qui ne comportaient pas de segment [j] (cf. *Caesar* → *caesarianus*). Du point de vue interlinguistique, un suffixe dérivé de *-anus* existe dans toutes les langues romanes principales (et en anglais), et dans tous les cas une variante avec la semivoyelle et une sans la semivoyelle coexistent. De plus, dans toutes les langues en question, à côté des deux variantes citées il en existe une troisième, dans laquelle à la place de la semivoyelle apparaît un segment vocalique plein, rendu graphiquement avec *e*. Dans ces cas le suffixe est donc systématiquement dissyllabique. Quelques exemples sont données en (12) :

- (12) fr. : Kafka → kafk(a)ien / kafkéen
esp. : Borges → borgiano / borgeano
ang. : Shakespeare → Shakespearian / Shakespearean

L'existence de la variante dissyllabique peut aussi être expliquée facilement du point de vue diachronique comme une contamination entre les deux suffixes latins *-(a)eus* et *-anus* (cf. Schweickard 1992 : 184), comme le démontre la comparaison entre les formes de l'italien *uropeo* et du français *européen*. Du point de vue synchronique, toutefois, les deux variantes continuent de fonctionner de façon productive dans les langues considérées, et il faut donc identifier d'autres facteurs qui sont responsables de leur distribution.

3.3.2 Un suffixe ou deux ?

Pour ce qui concerne l'italien, il est légitime d'abord de se demander si les deux variantes *-ano* et *-iano* (pour *-eano* cf. 3.3.9) doivent être considérées comme deux suffixes distincts, malgré leur parenté étymologique et sémantique, ou s'ils doivent être considérés comme deux variantes du même suffixe. La première hypothèse est celle qui est adoptée, souvent implicitement, par la plupart des dictionnaires de l'italien. Selon Zingarelli (1997), par exemple, *-ano* sert à construire des adjectifs qui indiquent « mestiere, categoria, classe, dignità e sim. » et des ethniques, alors que *-iano* sert plus généralement à construire des adjectifs de relation à partir de noms propres. En ita-

lien il existerait donc deux suffixes, qui servent tous les deux à construire des adjectifs de relation : *-iano*, spécialisé dans la dérivation déonomastique, et *-ano* qui, par exclusion, s'appliquerait aux autres types de bases, et principalement à des toponymes pour construire des ethniques. Une observation sommaire des lemmes inclus dans le même dictionnaire, cependant, suggère une certaine prudence vis à vis d'une distinction si nette entre les deux variantes, uniquement fondée sur le type de bases auxquelles ils se lient : à côté d'adjectifs apparemment dérivés de noms propres dans lesquels le segment [j] n'apparaît pas (cf. *Copernico* 'Copernic' → *Copernicano*, *Maometto* 'Mahomet' → *maomettano*¹¹⁷), il existe de nombreux exemples dans lesquels la semi-voyelle apparaît lorsqu'on ne l'attendrait pas, comme avec des bases qui sont des toponymes (cf. *Brasile* 'Brésil' → *brasiliano*) ou des noms communs (cf. *rettile* 'reptile' → *rettiliano*). En réalité, cependant, la distinction proposée par Zingarelli (1997) reflète assez bien une intuition qui est globalement confirmée par les lemmes contenus dans le dictionnaire.

3.3.3 Contraintes phonologiques

Avant d'avancer dans l'analyse du suffixe *-(i)ano*, il convient de se pencher sur les quelques contraintes phonologiques auxquelles ses dérivés sont soumis. En premier lieu, lorsque la dernière consonne de la base est une palatale ([ʃ], [tʃ], [dʒ], [ʎ] ou [ɲ]), le segment [j], suivant une règle phonotactique systématique en italien, n'apparaît jamais :

- (13) Sciascia → sciasciano [ʃaʃʃaːno]
Croce → crociano [kroˈtʃaːno]
Badoglio → badogliano [badoˈʎaːno]

Deuxièmement, lorsque la base se termine par une diphtongue ascendante ou descendante dont l'élément semivocalique est un [j], cet élément apparaît toujours (plus

117 Dans la suite de cette section je ne donne la glose que pour les noms propres (de personne ou autre) qui ne sont pas identiques en italien et en français, ou je donne des explications pour des noms propres qui ne sont pas immédiatement compréhensibles en français. Puisque la glose des dérivés transparents en *-(i)ano* est systématiquement 'relatif à', j'évite aussi de gloser tous les dérivés.

rarement, il apparaît aussi dans des cas de hiatus comportant un [i] :

- (14) Italia ‘Italie’ → italiano
 Tolstoj → tolstoiano / tolstoiano
 Maria ‘Marie’ → mariano

Dans une étude récente consacrée à la dérivation déonomastique, Seidl (2004 : 412) s’appuie sur des exemples comme ceux de (13) et (14) pour affirmer que « [d]a un punto di vista sincronico l’esistenza di due suffissi differenti *-iano* e *-ano* e fuori di dubbio » : *-ano* serait l’allomorphe choisi dans les cas, comme ceux qu’on a cités, dans lesquels l’émergence de la semivoyelle est bloquée pour des raisons phonologiques. Mais, puisqu’en italien l’effacement de la semivoyelle [j] dans les contextes cités est une règle phonologique systématique¹¹⁸, se demander si le suffixe a la forme [ano] ou [jano] dans les cas en question est vain, et ne peut qu’aboutir à une réponse arbitraire.

Seidl (2004) soutient aussi que l’apparition de *-iano* serait impossible après des bases qui se terminent elles-mêmes par la séquence [janV] et rare après une « racine » se terminant en voyelle (c’est-à-dire avec une base qui, à la forme autonome, se termine par un hiatus). La deuxième question sera traitée dans le détail en 3.3.9. Pour ce qui concerne la première, la séquence [janjano] est effectivement rare. Toutefois, la présence de dérivés comme *Diana* → *dianiano* ou *Galliani* → *gallianiano* montre qu’il s’agit, tout au plus, d’une contrainte ‘soft’, qui peut être violée à l’occurrence. Quant à la préférence présumée des bases qui se terminent par la vélaire [k] pour la variante *-ano*, que Seidl (2004) établit à partir d’exemples comme *Seneca* ‘Sénèque’ → *senecano*, elle semble être peu active en italien, et plutôt liée au timbre de la voyelle finale de la base (cf. 3.3.9).

D’autres restrictions sur les bases potentielles de *-iano* sont discutées par Thornton (2001), et concernent principalement la présence de diphtongues et / ou d’attaques complexes dans la base. En particulier, Thornton observe une certaine tendance à éviter, dans les dérivés, les séquences diphtongue + attaque complexe + diphtongue ([jV.Crja.no]), et cela serait démontré par le fait que *dipietrano* est attesté plus fréquem-

118 Comme le démontrent, par exemple, les formes verbales fléchies à la première personne du pluriel du présent de l’indicatif ou du subjonctif (dont l’exposant normalement a la forme [jam]) : *mangiamo* ([man'dʒa:mo], ‘nous mangeons’), *cambiamo* ([kam'bjɑ:mo], ‘nous changeons’).

ment que *dipietriano* en tant que dérivé de *Di Pietro*. En réalité, encore une fois, on est probablement face à une contrainte qui peut être violée, et les locuteurs ont le choix entre la transparence maximale (l'emploi de la variante *-iano*, qui comme on la vu est majoritaire avec les noms propres) et l'optimalité phonologique, même si la première semble l'emporter : sur Google (données brutes, recherche de mai 2010) on trouve environ 2 000 exemples de *dipietriano* et 500 de *dipietrano*¹¹⁹. La restriction en question, d'ailleurs doit être liée à une contrainte plus générale qui défavorise l'apparition de semivoyelles dans deux syllabes contiguës (à son tour liée à des contraintes vraisemblablement universelles, comme l'Obligatory Contour Principle, cf. McCarthy 1986 ; cf. aussi Lignon 2000 ; Lignon & Plénat 2009 pour l'application de la même contrainte à la suffixation en *-ien* en français). En (15) je liste quelques exemples de dérivés dont la base comporte une semivoyelle [j] dans la syllabe qui précède immédiatement le suffixe, et qui sont attestés avec les deux variantes :

- (15) Antonioni → antonionano / antonioniano
Flaiano → flaianano / flaianiano
Preziosi → preziosano / preziosiano
Tolkien → tolkienano / tolkieniano
Walter Chiari → walterchiarano / walterchiariano

En général, toutefois, la généralité sémantique du suffixe et la spécialisation à se lier à des noms propres semblent être plus forts de toute restriction sur sa forme phonologique.

3.3.4 *-(i)ano* dans les dictionnaires de l'italien

Dans le dictionnaire Zingarelli (1997) apparaissent 465 dérivés en *-(i)ano*. Une ré-

119 Les deux seules autres bases possibles de *-(i)ano* que j'ai repérées qui peuvent donner lieu au contexte en question sont *Biafra* (un toponyme) et *Viagra* (un nom commercial), et dans les deux cas la forme sans semivoyelle l'emporte largement sur Google (*biafrano* ≈ 6 000 vs. *biafrano* ≈ 20 ; *viagrano* ≈ 230 vs. *viagrano* ≈ 5). En passant, je note que la séquence [jV(C).Crja.no] n'est pas totalement exclue en italien : on la retrouve au moins à la première personne du pluriel du présent de l'indicatif et du subjonctif de certains verbes comme *INCHIOSTRARE* ('encre', *inchiostriamo*).

partition selon le type de base et selon la présence / absence de la semivoyelle [j] donne les résultats suivants :

	[jano]	[ano]	[j]/pal. + [ano]
noms propres	159	9	52
toponymes	54	62	63
noms communs	10	26	7
adjectifs	2	12	2
autres	–	7	–

Tableau 10 : répartition des différents types de dérivés en *-(i)ano* dans Zingarelli (1997)

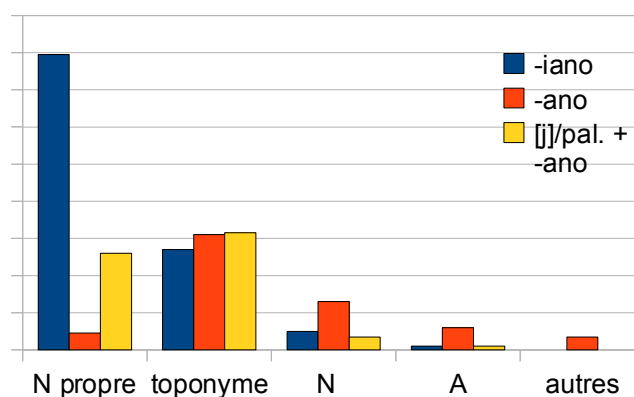


Figure 3

En laissant de côté les cas indécidables (qui apparaissent dans la dernière colonne dans le Tableau 10), les données ci-dessus confirment grosso modo l'intuition que *-iano* est préféré pour la dérivation à partir de noms propres et *-ano* pour la dérivation à partir de noms communs, adjectifs et surtout toponymes. Il existe cependant des exceptions à cela. Parmi les huit exceptions à la généralisation NP+*-iano*, il y en a une qui est probablement un calque direct de l'anglais (*Boole* → *booleano*), et trois autres sont construites sur des bases comportant un hiatus (*Genoa* 'équipe de football de

Gênes' → *genoano*, *Linneo* 'Linné' → *linneano*, *Pompeo* 'Pompée' → *pompeano*, les deux dernières enregistrées aussi avec la forme en [j]), et on peut considérer, comme première hypothèse de travail, qu'une séquence [VjV] est au moins tout aussi marquée qu'un hiatus en italien (cf. 3.3.9 pour un développement). Il reste cinq cas qui sont apparemment des exceptions :

- | | |
|-------------------------------|---------------|
| (16) Copernico 'Copernic' | → copernicano |
| (S.) Domenico 'St. Dominique' | → domenicano |
| (S.) Francesco 'St. François' | → francescano |
| Lutero 'Luther' | → luterano |
| Maometto 'Mahomet' | → maomettano |

Pour ces dérivés, cependant, on peut retenir l'hypothèse de Rainer (1996a), selon qui il s'agit de lexèmes importés en italien directement du latin ecclésiastique ou scientifique. Cela ne signifie pas, naturellement, que tous ces lexèmes ne sont pas transparents, mais simplement qu'il s'agit d'unités qui font partie du lexique italien sans avoir été construites par un procédé actuellement productif.

Passons maintenant au groupe d'exceptions' le plus consistant, celui des 54 lemmes qui violeraient la généralisation toponyme+*-ano*. Une première observation est que de ces 54 mots, seulement 7 sont construits sur la base d'un toponyme italien (par exemple *Canossa* → *canossiano*, *Ponza* → *ponziano*). Puisque, naturellement, l'existence d'un ethnique en italien est systématique pour les toponymes italiens mais pas pour les toponymes étrangers, il est vraisemblable de penser que l'ensemble des premiers est plus ancien que l'ensemble des seconds. Par conséquent, si *-iano* se révélait plus productif avec les toponymes étrangers qu'avec les toponymes italiens (une observation dans ce sens est proposée par Rainer 2004 : 406), on pourrait faire l'hypothèse qu'il est en train de supplanter *-ano* aussi avec ce type de bases. Je reviendrai sur les dérivés de toponymes en 3.3.7.

Il est possible à ce stade de tirer quelques conclusions à partir de l'observation des données du dictionnaire, et d'essayer de donner une réponse à la question si en italien il existe un suffixe *-(i)ano* avec deux allomorphes ou s'il existe bien deux suffixes distincts. Pour résumer, nous avons établi qu'en italien on retrouve :

1. un grand nombre de dérivés de noms propres et un certain nombre de dérivés de toponymes, surtout étrangers, dans lesquels apparaît la forme [jano] ;
2. un nombre important de dérivés dans lesquels apparaît la forme [ano], parmi lesquels :
 - i. un grand nombre de dérivés de toponymes italiens qui sont attestés depuis longtemps ;
 - ii. un certain nombre de dérivés de toponymes étrangers ;
 - iii. quelques dérivés d'autres types de bases (noms communs ou adjectifs) ;
 - iv. des dérivés de bases pour lesquels l'apparition d'une semivoyelle [j] est défavorisée par des contraintes phonologiques (comme dans le cas de *dipietrano*) ;
3. un nombre important de dérivés ambigus pour lesquels la semivoyelle [j] peut aussi bien faire partie de la base que du suffixe (ou des deux).

En particulier pour ce qui concerne les cas listés au point 2., il y a une différence de substance entre les mots qui appartiennent au lexique de l'italien depuis longtemps (comme dans les cas 2.i-iii) et ceux où la variante [ano] apparaît de façon productive (comme dans les cas 2.iv). Seulement dans une optique strictement segmentale (comme celle adoptée par exemple par Seidl 2004), et qui ne fait pas la différence entre morphologie productive et lexique mémorisé, on peut être tenté de considérer tous les éléments qui présentent à la surface la séquence [ano] sans semivoyelle comme issus du même procédé morphologique. Il est probable, au contraire, que la question posée en 3.3.2 (existe-t-il un ou deux suffixes *-(i)ano* ?) ait une double réponse, selon les dérivés considérés : *-ano* et *-iano* sont (ou ont été) deux suffixes différents en italien (comme dans les dérivés 2.i-iii), puisque l'émergence de l'un ou de l'autre n'a pas de justification phonologique, mais il est possible d'identifier des généralisations sémantiques ; *-ano* et *-iano* sont deux variantes du même suffixe (comme dans les dérivés de 2.iv), et le premier fonctionne comme une variante du second lorsque celui-ci est défavorisé pour des raisons phonologiques.

3.3.5 La disponibilité de *-(i)ano*

Avant d'approfondir l'aspect phonologique des dérivés en question, il convient de se pencher encore brièvement sur la disponibilité des deux variantes en italien contemporain, et en particulier de *-ano*, hors conditionnements phonologiques. Pour évaluer la disponibilité des deux suffixes, je prends en considération un corpus de dérivés en *-(i)ano* basé sur le Web que j'ai construit entre 2006 et 2007. Ce corpus a été ainsi constitué : sur la base de la version italienne d'une encyclopédie disponible sur le Web (*Encarta*¹²⁰, dont n'est actuellement disponible que le dictionnaire, l'encyclopédie ayant été supprimée en 2009), j'ai obtenu une liste de noms de personnages et de lieux susceptibles de donner lieu à un dérivé en *-(i)ano*, dont l'existence a été vérifiée par des recherches Google. D'autres recherches ont été effectuées à partir du corpus du quotidien *la Repubblica*¹²¹ et de quelques répertoires de néologismes. Une recherche moins systématique a été effectuée sur des noms propres non de personne (noms d'entreprises, produits, associations, logiciels, etc.). De cette manière un corpus d'environ 1.800 entrées a été recueilli, ainsi réparties : 1.412 dérivés de noms propres de personne, 382 dérivés de toponymes étrangers ; 96 dérivés de noms propres non de personne. Le nombre de tokens recueilli pour chaque dérivé est souvent petit. Ce corpus ne nous permet pas d'effectuer des calculs de fréquences sérieux, mais il est tout de même utile pour observer quelques généralisations.

3.3.6 *-(i)ano* déonomastique

Des environ 1.500 dérivés en *-(i)ano* de noms propres présent dans le corpus, 61 présentent la forme sans semivoyelle, si on exclut, encore une fois les cas d'ambiguïté, dans lesquels la base se termine par une diphtongue ou par une consonne palatale. En (17) ces dérivés sont repartis selon que la séquence [ano] est précédée d'une consonne ou d'une voyelle, et dans ce cas on indique la voyelle qui la précède :

120 <http://encarta.msn.com/>.

121 <http://dev.sslmit.unibo.it/corpora/corpus.php?path=&name=Repubblica>.

(17) [Cano] (<i>Wojtyła</i> → <i>wojtylano</i>)	19
[Vano]	42
dont :	
[eano] (<i>Mallarmé</i> → <i>mallarmeano</i>)	32
[uano]/[wano] (<i>Murnau</i> → <i>murnauano</i>)	6
[oano] (<i>Pessoa</i> → <i>pessoano</i>)	4

Je laisse momentanément de côté les cas en [uano]/[wano]¹²². Dans les cas où la voyelle qui précède [ano] est un [e] ou un [o], celle-ci est déjà présente dans la base comme (i) une voyelle finale accentuée, (ii) le premier élément d'un hiatus, (iii) une voyelle finale non accentuée. Le Tableau 11 résume la situation, en indiquant le nombre d'exemples pour chacun de ces cas :

	(i) finale accentuée	(ii) premier él. de hiatus	(iii) finale non accentuée
[eano]	4 (<i>Mallarmé</i> → <i>mallarmeano</i>)	17 (<i>Cattaneo</i> → <i>cattaneano</i>)	11 (<i>Frege</i> → <i>fregeano</i>)
[oano]	2 (<i>Totò</i> → <i>totoano</i>)	2 (<i>Pessoa</i> → <i>pessoano</i>)	–

Tableau 11 : types de bases avec les dérivés en [eano]/[oano]

L'explication qui a été proposée en 3.3.4 pour les cas comme *genoano* et *linneano* est probablement valable aussi pour ceux illustrés dans le Tableau 11. En particulier, la forme [ano] serait préférée dans certains contextes phonologiques, et notamment quand elle permet d'éviter une séquence [VjV], dont on a dit qu'elle peut être considérée comme moins optimale qu'un hiatus (ce point sera repris plus dans le détail en 3.3.9).

Quant aux 19 mots qui présentent la séquence [Cano], 7 d'entre eux sont du type de *dipietrano* et *antonioniano*, discutés en 3.3.3, et doivent partant être considérés comme conditionnés phonologiquement. Curieusement, les 9 restants ont pour base

122 Qui doivent, de toute façon, être réduits à trois, puisque dans trois autres cas le *u* graphique apparaît avec des noms d'origine française dans la transcription du phonème [ø] (comme dans *Richelieu* → *richelieuano*). Il s'agit de formes éminemment écrites, et j'aurais du mal à décider comment on prononce *richelieuano*, mais les prononciations les plus plausibles me sembleraient être [riʃ:elje'a:ɲo] ou [riʃ:eljø'a:ɲo], aucune desquelles ne comporte un [u].

un nom qui se termine en [a] (*Giacosa* → *giacosano*, *Wojtyła* → *wojtylano*¹²³, cf. 3.3.9).

3.3.7 *-(i)ano* dans les ethniques

Dans la discussion des données dictionnaires j'ai observé que la proportion de dérivés en *-ano* sans semivoyelle est plus importante pour les dérivés qui ont des toponymes, notamment italiens, comme bases. Cette observation est confirmée dans l'analyse proposée par Crocco Galèas (1991), qui a recensé 822 ethniques italiens formés au moyen du suffixe *-(i)ano*, et de ceux-ci 38 comportent une semivoyelle [j] qui n'est pas déjà présente dans la base. Cependant, les données recueillies dans le corpus décrit en 3.3.5 nous permettent de redimensionner ce calcul. Des 382 ethniques formés sur la base d'un toponyme étranger dans le corpus considéré, 251 présentent la variante [jano] sans que la semivoyelle ne soit pas déjà présente dans la base, et seulement 92 présentent la variante [ano] sans ambiguïtés. Les données précises sont résumées dans le Tableau 12¹²⁴ :

[jano]	[ano]		cas ambigu
	[Cano]	[Vano]	
251	69	23	39
<i>(Hannover</i> → <i>hannoveriano)</i>	<i>(Florida</i> → <i>floridano)</i>	<i>(Borneo</i> → <i>borneano)</i>	<i>(Pennsylvania</i> → <i>pennsylvaniano)</i>

Tableau 12 : types de dérivés de toponymes

De plus, si nous laissons de côté les 23 cas comme *borneano* (qui peuvent être assimi-

123 Dans tous ces cas, à côté de la forme dépourvue de semivoyelle en est attestée aussi une qui la comporte (*giacosiano*, *wojtyliano*).

124 Dans le calcul proposé je n'ai pas tenu compte de l'influence éventuelle que la forme de l'ethnique dans la langue originaire peut avoir sur l'ethnique en italien. Un calcul plus précis serait probablement utile, mais quelques observations suggèrent de ne pas surestimer cet aspect : (i) les textes présents sur le Web sont extrêmement hétérogènes, et il est vraisemblable d'imaginer qu'au moins une partie des locuteurs qui ont utilisé les ethniques recensés ne possèdent pas les compétences linguistiques pour construire la forme italienne par un calque sur une forme alloglotte ; (ii) dans toutes les langues où il existe un suffixe dérivé du *-(i)anus* latin le même problème de la variante avec une semivoyelle se pose ; (iii) dans une très grande majorité des cas pour un même toponyme j'ai recensé plusieurs ethniques, et dans certains cas aucun d'entre eux ne correspond à celui employé dans la langue originale (cf. *Ibiṣṣa* → *ibizano* / *ibiziano* vs. esp. *ibicenco* / cat. *eivissenc*).

lés au type *genoano*, *linneano* cité en 3.3.4) et nous repartissons les dérivés dans lesquelles apparaissent les deux variantes avec et sans semivoyelle selon le segment final qui apparaît dans la base, nous obtenons les résultats suivants :

	[jano]	[Cano]
[C]	133	7
[a]	52	55
[e]	17	–
[i]	26	–
[o]	15	5
[u]	2	2

Tableau 13 : dérivés ethniques selon le segment final de la base

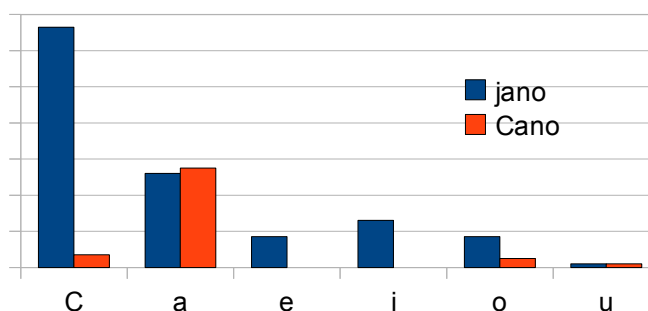


Figure 4

Comme on le voit, la proportion de dérivés dans lesquels le suffixe a la forme [ano] est significative seulement dans les cas où la base se termine déjà par la voyelle [a] – ce qui coïncide avec le comportement des noms propres comme *Wojtyła* cités en 3.3.4. Dans tous les autres cas la prépondérance de la forme [jano] ne fait pas de doute. D'autant plus que pour 35 des 69 bases pour lesquelles un dérivé en [Cano] a été enregistré, celui-ci coexiste avec un dérivé en [jano] (cf. *Minnesota* → *minnesotano* / *minnesotiano*).

La présence de dérivés dans lesquels le suffixe apparaît sans la semivoyelle est tout

de même plus importante avec les ethniques qu'avec les déonomastiques, ce qui démontre que l'interaction entre les différentes contraintes, phonologiques et lexicales, qui pèsent sur la construction de mots dérivés est complexe. La proportion importante de dérivés du type *floridano* ou *minnesotano* est clairement le fruit de l'interaction de la contrainte qui favorise la variante sans semivoyelle lorsque la base se termine par [a] et de celle qui la favorise avec les bases qui sont des toponymes.

3.3.8 La sémantique de *-(i)ano*

Les données recueillies non pas à partir du lexique enregistré dans les dictionnaires, mais sur la capacité du suffixe *-(i)ano* à former de nouveaux lexèmes montrent assez clairement que, si nous acceptons l'idée que *-ano* et *-iano* sont deux variantes du même suffixe, la forme comportant la semivoyelle est celle qui fonctionne comme le défaut, indépendamment de toute restriction d'ordre lexical et / ou phonologique.

Avant de poursuivre dans l'analyse du corpus, quelques observations générales sur ce suffixes sont utiles. Comme on l'a vu, la proportion de dérivés avec *-(i)ano* à partir de base qui ne sont ni des noms propres ni des toponymes est très réduite. La construction de dérivés en *-(i)ano* à partir de noms communs semble n'être productive, actuellement, que dans la langue scientifique (cf. *mammifero* 'mammifère' → *mammiferiano*, *dinosauro* 'dinosaur' → *dinosauriano*) ; au contraire, dans la langue courante les bases presque exclusives de ces dérivés sont, comme on l'a vu, les noms propres et les toponymes. La caractéristique principale de ces deux types d'unités est celle d'être des lexèmes aptes à la désignation d'un référent unique, et celle-ci est justement la seule restriction sémantique qui pèse sur ce suffixe en dehors du langage scientifique. Il existe quelques cas de dérivés en *-(i)ano* qui sont apparemment construits sur un nom commun ; cependant, dans la plupart des cas le nom en question possède un seul référent potentiel. En italien, par exemple, il existe un adjectif *areniano*, qui s'est spécialisé avec le sens de 'relatif à l'Arène de Vérone', et le même est vrai dans l'exemple suivant, où *giubileiano* renvoie spécifiquement au Jubilé de l'année 2000 à Rome :

(18)

Patrizia Sentinelli (Segreteria Nazionale di RC e già connivente del Rutelli sindaco di giubileiana memoria).

‘Patrizia Sentinelli (Secrétariat National de Réfondation Communiste et ex complice du maire Rutelli de l’époque du Jubilé’¹²⁵.

La caractérisation sémantique sommaire qui a été proposée pour le suffixe *-(i)ano* a une répercussion sur son comportement phonologique. La restriction sémantique à laquelle la construction morphologique dont il est l’exposant est soumise, ainsi que son instruction sémantique très générale, ont pour conséquence que tout nom à référent unique puisse fonctionner comme base de cette règle. Puisque la catégorie des noms propres est particulièrement propice à accueillir des termes d’origine alloglotte, le pourcentage de combinaisons phonologiques non communes est fréquente. Le fait que, dans ce cas, les locuteurs sont obligés d’adapter un schéma constructionnel à du matériel inattendu, en fait un bon point d’observation des mécanismes qu’ils mettent en œuvre, au delà des conditionnements lexicaux.

3.3.9 La phonologie de *-(i)ano*

Pour résumer les observations conduites jusqu’à présent, lorsque l’application de la construction en *-(i)ano* n’est pas soumise à d’autres contraintes d’ordre phonologique, la seule option vraiment disponible en italien contemporain est que les dérivés présentent la variante avec la semivoyelle [j]. Comme on l’a rappelé, le cas par défaut dans la dérivation dénominale et désadjectivale de l’italien est que dans un dérivé la base apparaisse sans sa voyelle (atone) finale. Par conséquent, on peut considérer que le schéma par défaut pour la suffixation en *-(i)ano* en italien est celui illustré en (19), avec un nom à finale vocalique atone, et avec un nom à finale consonantique :

- (19) Pirandello → pirandelliano
Iran → iraniano

125 <http://www.siporcuba.it/mcarc-vit.htm>.

La très grande majorité des 1.800 entrées du corpus que j'ai considéré se conforment simplement à la généralisation en (19), et ne présentent donc pas un grand intérêt pour une étude phonologique de la question. Les données le plus intéressantes, de mon point de vue, sont celles dans lesquelles le schéma en question n'est pas respecté parce que le suffixe ne comporte pas la semivoyelle [j], dans la plupart des cas à cause de l'interaction avec la forme phonologique de la base. Il y a un certain nombre de configurations qui sont plus fréquentes que d'autres. En particulier, les bases qui favorisent la non apparition de [j] sont les suivantes :

- bases en voyelle [u] ou [e] atone ;
- bases en voyelle tonique ;
- bases qui comportent un hiatus en finale ;
- bases qui comportent une diphtongue ou un hiatus incluant le segment [w]/[u].

Le Tableau 14 résume le nombre de dérivés présents dans le corpus qui présentent (aussi) une forme différente de [jano], répartis selon le format de la base. Pour chaque type j'indique le nombre de formes attestées dans le corpus et un exemple.

		[jano]	[Vano]	[Vjano]	[VVano]	[VVjano]
V atones	[u]	8 <i>Lussu</i> <i>lussiano</i>	3 <i>Nauru</i> <i>nauruano</i>	1 <i>Cossu</i> <i>cosuiano</i>	2 <i>Ryukyu</i> <i>ryukyiano</i>	
	[e]	20 <i>Goethe</i> <i>goethiano</i>	13 <i>Frege</i> <i>fregeano</i>	5 <i>Kripke</i> <i>kripkeiano</i>		
V toniques	[a]	2 <i>Derrida</i> <i>derridiano</i>		5 <i>Carrù</i> <i>carraiano</i>		
	[e]/[ɛ]	6 <i>Poincaré</i> <i>poincariano</i>	6 <i>Mallarmé</i> <i>mallarmeano</i>	10 <i>De André</i> <i>deandreiano</i>		
	[o]/[ɔ]		2 <i>Totò</i> <i>totoano</i>	12 <i>Ardigò</i> <i>ardigoiano</i>		
	[u]		1 <i>Timbuctu</i> <i>timbuctuano</i>	4 <i>Pelù</i> <i>peluiano</i>		
diphth. / hiatus en [w]	[wa]/[ua]		5 <i>Papua</i> <i>papuano</i>	1 <i>Bevilacqua</i> <i>bevilacquiano</i>		
	[we]/[ue]			2 <i>Zimbabwe</i> <i>zimbabwiano</i>	3 <i>Cimabue</i> <i>cimabueano</i>	1 <i>Ligabue</i> <i>ligabueiano</i>
	[aw]			1 <i>Tokelau</i> <i>tokelaiano</i>	3 <i>Murnau</i> <i>murnauano</i>	2 <i>Spandau</i> <i>spandauiano</i>
hiatus	[ao]			2 <i>Ingrao</i> <i>ingraiano</i>		2 <i>Mao</i> <i>maoiano</i>
	[ea]/[ɛa]		8 <i>Ikea</i> <i>ikeano</i>	3 <i>D'Andrea</i> <i>dandreiano</i>		
	[eo]/[ɛo]	2 <i>Borromeo</i> <i>borromiano</i>	16 <i>Matteo</i> <i>matteano</i>	13 <i>Galileo</i> <i>galileiano</i>		
	[oa]/[ɔa]		4 <i>Samoa</i> <i>samoano</i>	2 <i>Pessoa</i> <i>pessoiano</i>		

Tableau 14 : types de dérivés en *-(i)ano* selon la forme de la base

Lorsque la base se termine par une voyelle atone simple ou par une voyelle tonique, la forme du dérivé choisie plus fréquemment est celle attendue d'un point de vue strictement concaténatif, si on admet que la forme du suffixe productive est effectivement [jano]. Puisqu'en italien, comme on l'a rappelé, une voyelle atone finale est systématiquement effacée devant à un suffixe qui commence par voyelle, on a *Goethe* → *goethiano* comme on a *fiore* 'fleur' → *fiorista* 'fleuriste' ; au contraire, une voyelle finale tonique est généralement maintenue devant un suffixe, et donc on a *De André* → *deandriano* comme on a *virtù* 'vertu' → *virtuoso* 'vertueux'. Cependant, cette 'régularité' apparente est perturbée par le fait que, lorsque la base se termine par un hiatus, bien que le schéma attendu soit attesté (cf. *Cattaneo* → *cattaneiano*), il y en a un autre qui est plus fréquent, à savoir celui dans lequel la forme [ano] est précédée de la voyelle qui constitue le premier élément du hiatus (comme dans *cattaneano*), et donc a la forme [eano], [wano] ou [oano]. Jusqu'à quel point, cependant, est-il utile de se placer dans une optique rigidement concaténative pour l'analyse des dérivés en *-(i)ano* ? Si on considère les données uniquement dans cette perspective, en effet, il y a plusieurs aspects qui risquent de nous échapper, et notamment, la nature des bases auxquelles *-(i)ano* se lie dans les dérivés du Tableau 14 n'est pas claire : parfois elle correspond à une racine (c'est-à-dire, le lexème de base sans la voyelle atone finale), d'autres fois elle est plus grande ou plus petite que celle-ci. La base des dérivés en *-(i)ano* moins canoniques, tels ceux qu'on a illustrés au Tableau 14, est donc difficilement définissable aussi bien en termes morphologiques (parfois elle correspond à l'objet morphologique qui fonctionne de base pour les autres dérivés) qu'en termes phonologiques (le nombre et le type des voyelles effacées sont plutôt imprévisibles). Une position raisonnable est celle de considérer que les locuteurs identifient effectivement un format prototypique pour les dérivés, qui correspond à la base par défaut qu'on a définie plus haut (sans une voyelle atone finale, mais avec une voyelle tonique). Cependant, cela n'empêche pas que, selon l'interaction avec la forme du lexème de base, d'autres configurations soient possibles, par exemple qui prévoient l'effacement d'une voyelle tonique (*derridiano*), l'effacement de plus d'une voyelle (*borromiano*), le maintien d'une voyelle atone finale (*fregeano*), etc. Il est possible aussi que l'effacement de la voyelle fi-

nale des noms propres (les bases typiques des dérivés en *-(i)ano*), dans lesquelles les voyelles finales ne véhiculent pas d'information flexionnelle, soit considérée comme moins automatique que pour les autres lexèmes nominaux.

En sortant d'une optique rigidement segmentale, nous nous apercevons que la caractérisation de l'exposant de la construction morphologique en question comme ayant la forme [jano] est réductrice. Nous avons vu, en effet, que dans un certain nombre de cas le suffixe a la forme [ano], mais que les segment qui précèdent cette séquence sont aussi prévisibles, et que, de plus, ils suivent un ordre de préférence. Il est donc possible de représenter la forme phonologique du suffixe comme une série de contraintes hiérarchisées, que je donne en (20) :

(20) [jano] > [eano] > [wano] > [oano] > [Cano]

Je propose donc de considérer que les affixes ne sont pas nécessairement des objets qui ont une forme sous-jacente qui permet de dériver ses formes de surface. Plutôt, les constructions morphologiques dérivationnelles spécifient un format optimal pour leurs dérivés (dans le cas qui nous intéresse, en spécifiant la séquence de phonèmes qu'ils doivent comporter à leur marge droite) et un ou plusieurs autres formats, qui peuvent être émerger à la suite de l'interaction avec des bases qui ont des formes différentes. On voit que cette caractérisation de la morphologie dérivationnelle est plus compatible avec le modèle de morphologie émergente et fondé sur le poids du lexique existant caractérisé dans la partie 2. Dans le cas de *-(i)ano* la représentation la plus générale que nous pouvons donner est [Vano], où le timbre du segment V est spécifié par la hiérarchie [i] > [e] > [u] > [o] > [a] (pour la présence de [a] dans la hiérarchie cf. ci-dessous). La motivation pour l'existence de la hiérarchie que j'ai indiquée est une conjonction de facteurs phonologiques et de facteurs extralinguistiques liés à l'histoire de l'affixe. L'assignation par défaut de la valeur [j] au segment vocalique est due à la réanalyse des dérivés latins en *-anus* comportant une semivoyelle qu'on a vue e ; 3.3.1. La hiérarchie en (20) rend compte aussi de l'existence de la variante [Cano] et de sa fréquence avec des bases se terminant par [a], puisque l'apparition de deux segments identiques est particulièrement défavorisée dans les langues du monde.

Naturellement, la hiérarchie proposée interagit avec d'autres facteurs, phonologiques ou pas, pour déterminer l'ensemble des formes qu'on a observées avec les dérivés en *-(i)ano*. En premier lieu, il existe une échelle universelle de naturalité des structure syllabiques, due au fait que les langues tendent à éviter les hiatus, les attaques complexes et que les consonnes pleines constituent de meilleures attaques par rapport aux semivoyelles :

(21) CV > CGV > CVV

Les données observées en relation à *-(i)ano*, de plus, suggèrent pour l'italien l'existence d'un échelon ultérieur dans la hiérarchie de (21), c'est-à-dire la préférence pour des séquences CVV par rapport à des séquences CVGV. En effet, lorsque deux outputs concurrents contiennent l'une des deux séquences, comme dans le cas de *Frege* → *fregeano* / *fregeiano*, le premier est en général plus fréquent. Il est possible, de plus, de rajouter quelques observations en faveur de l'hypothèse CVV > CVGV. En premier lieu, statistiquement, dans l'ensemble du lexique italien, les séquences VV sont largement plus fréquentes que les séquences VGV. Dans le Tableau 15 je résume les résultats d'un calcul approximatif que j'ai effectué à partir des lemmes contenus dans le dictionnaire Zingarelli (1997). Comme on voit, les premières sont environ trois fois et demie plus fréquentes que les secondes :

VV	VjV
<i>ae</i> 363	<i>aie</i> 48
<i>ao</i> 132	<i>aio</i> 753
<i>ea</i> 955	<i>eia</i> 19
<i>eo</i> 1.480	<i>eio</i> 32
<i>oa</i> 418	<i>oia</i> 151
<i>oe</i> 345	<i>oie</i> 69
Tot. 3.693	Tot. 1.072

Tableau 15 : séquences VV et VjV dans le lexique italien (Zingarelli 1997)

Deuxièmement, du point de vue diachronique, dans le passage du latin à l'italien an-

cien à l'italien moderne on a parfois observé la tendance à simplifier les séquences VGV parfois en GV et parfois en VV, comme le montrent les toponymes en (22) (données de Pieri 1913 : 110-111) :

(22) latin	it. ancien	italien
Eleutherianus	Luctuiano (1037)	Lutiano (Florence)
Aterianus	Teiano (1104)	Teano (Caserta)
Serianus	Seiano (1134)	Siano (Salerne) / Seano (Prato)

Enfin, on a documenté une tendance, dans l'oral non surveillé, à réduire les voyelles non hautes (en particulier moyennes) à des glides, d'où la prononciation [sterɛoti'pa:to], possible pour *stereotipato* (cf. Marotta 1987 ; Gili Fivela & Bertinetto 1999).

De plus, un principe général impose aux dérivés morphologiques d'être le plus fidèles possible à leurs bases, tant du point de vue phonologique (c'est-à-dire, qu'on efface ou qu'on rajoute le moins de segments possible) que prosodique. Ce principe aussi, cependant, doit être modulé, puisqu'il dépend de façon cruciale d'autres facteurs, comme la longueur de la base, la position de l'accent, le timbre de la voyelle candidate à l'effacement, etc. Dans Montermini (2003a), par exemple, j'ai suggéré qu'il existe une hiérarchie d'effaçabilité pour les voyelles de l'italien, en vertu de laquelle les voyelles basses sont effacées avec beaucoup plus de facilité que les voyelles hautes ou moyennes. Tous ces principes peuvent être représentés sous la forme de contraintes qui pèsent sur la forme de sortie des dérivés en question (cf. 2.6.3). Non seulement les contraintes en question peuvent être violées, mais le fait qu'en général plusieurs outputs sont observés avec les mêmes bases ou les mêmes types de bases montre que leur ordonnancement est relatif. A titre d'exemple, je présente ci-dessous quel peut être un traitement des formes dérivées de *Frege* qui prenne en compte les différentes contraintes évoquées. Les trois formes *fregghiano*, *freggeano* et *freggeiano* sont effectivement attestées :

(23)

Base : *Frege* [frege]

	[jano]	[Vano]	fidélité	CV	CGV	CVV
[fre'ga:no]	*	*	*			
☞ [fre'gja:no]		*	*	*		
☞ [fre'gea:no]	*			*	*	
☞ [fre'geja:no]				*	*	*

Il est clair que dans le traitement proposé ici, une séparation nette entre base et affixe sur la chaîne segmentale qui serait toujours valide et univoque cesse d'être un enjeu. Il n'y a aucune nécessité, en effet, que aussi bien les bases (les lexèmes) que les affixes soient des éléments de la même nature, à savoir de vrais objets lexicaux auxquels correspond une (ou éventuellement plusieurs) séquences de phonèmes fixe. Si cette caractérisation est généralement acceptable pour les lexèmes, j'ai montré que les affixes sont traités de façon plus satisfaisante comme des contraintes sur la forme que doit avoir un mot complexe construit à partir d'un schéma morphologique spécifique. A chaque construction morphologique peuvent correspondre une ou plusieurs de ces contraintes, et dans le cas où elles sont plusieurs, elles sont en général hiérarchisées, la hiérarchie pouvant dépendre de facteurs phonologiques, de facteurs extralinguistiques, ou de l'interaction des deux, et ce dernier est le cas, on la vu, du suffixe *-(i)ano* de l'italien. Cette caractérisation des affixes est cohérente avec le modèle inférentiel et réalisationnel de la morphologie illustré dans la partie 2, et est particulièrement adaptée à rendre compte du fait que les constructions morphologiques émergent sur la base de généralisations effectuées par les locuteurs sur le lexique existant.

Bibliographie

- Albright A. 2002: « Islands of reliability for regular morphology: evidence from Italian ». *Language* 78: 684-709.
- 2003: « A quantitative study of Spanish paradigm gaps » in G. Garding, M. Tsujimura (eds), *WCCFL Proceedings*. Somerville, Ma: Cascadilla Press: 1-14.
- 2009: « Modeling analogy as probabilistic grammar » in Blevins & Blevins (eds) 2009, 185-213.
- à paraître: « Lexical and morphological conditioning of paradigm gaps » in C. Rice (ed.), *When Nothing Wins: Modeling Ungrammaticality in OT*. London: Equinox.
- Amiot D., Dal G. 2007: « Integrating neoclassical combining forms into a lexeme-based morphology » in G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli, S. Scalise (eds), *On-Line Proceedings of the 5th Mediterranean Morphology Meeting, Fréjus, 15-18 September 2005*. Bologna: Università di Bologna, 323-336 [<http://mmm.lingue.unibo.it/mmm-proc/MMM5/232-336-Amiot-Dal.pdf>].
- Amiot D., Montermini F. 2009: « Préfixes et mots grammaticaux » in Fradin *et al.* (eds) 2009, 127-141.
- Anderson S.R. 1992: *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- 2005: *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.
- Arcodia G.F., Grandi N., Montermini F. 2009: « Hierarchical NN compounds in a cross-linguistic perspective ». *Italian Journal of Linguistics / Rivista di linguistica* 21.1: 11-33.
- Aronoff M. 1976: *Word Formation in Generative Grammar*. Cambridge, Ma: MIT Press.
- 1980: « Contextuals ». *Language* 56: 744-758.
- 1983: « A decade of morphology and word formation ». *Annual Review of Anthropology* 12: 355-375.
- 1994: *Morphology by Itself*. Cambridge, Ma: MIT Press.
- 2007: « In the beginning was the word ». *Language* 83.4: 803-830.
- Aronoff M., Fudeman K. 2005: *What Is Morphology?*. Oxford: Blackwell.

- Baayen R.H. 2003: « Probabilistic approaches to morphology » in Bod R., Hay J., Jannedy S. (eds), *Probabilistic Linguistics*. Cambridge, Ma: MIT Press, 229-287.
- Baayen R.H., Dijkstra T., Schreuder R. 1997: « Singulars and plurals in Dutch: evidence for a parallel dual-route model ». *Journal of Memory and Language* 37: 94-117.
- Baayen R.H., McQueen J.M., Dijkstra T., Schreuder R. 2003: « Frequency effects in regular inflectional morphology: Revisiting Dutch plurals » in R.H. Baayen, R. Schreuder (eds), *Morphological Structure in Language Processing*. Berlin: Mouton de Gruyter, 355-390.
- Baerman M. 2008: « Historical observations on defectiveness: the first singular non-past ». *Russian Linguistics* 32: 81-97.
- Bafile L. 1999: « Antepenultimate stress in Italian and some related dialects: Metrical and prosodic aspects ». *Rivista di linguistica*. 11.2: 201-229.
- Baker M.C. 2003: *Lexical Categories. Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.
- Bates E., Goodman J.C. 1999: « On the emergence of grammar from the lexicon » in B. MacWhinney (eds), *The Emergence of Language*. Mahwah, NJ: Erlbaum, 29-79.
- Bauer L. 1997: « Derivational paradigms » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1996*. Dordrecht: Kluwer, 243-256.
- 2000: « Word » in Booij *et al.* (eds) 2000, 247-257.
- 2001: *Morphological Productivity*. Cambridge: Cambridge University Press.
- 2006: « Compound » in K. Brown (ed.), *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 719-726.
- 2009: « Typology of compounds » in R. Lieber, P. Štekauer (eds), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 343-356.
- Beard R. 1995: *Lexeme-Morpheme Base Morphology*. Albany: SUNY Press.
- Becker T. 1993: « Back-formation, cross-formation and 'bracketing paradoxes' in paradigmatic morphology » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1993*. Dordrecht: Kluwer, 1-25.
- Benincà P., Mioni A., Vanelli L. (eds) 1999: *Fonologia e morfologia dell'italiano e dei dialetti*

d'Italia. Atti del XXXI Congresso della Società di Linguistica Italiana (Padova, 25-27 settembre 1997). Roma: Bulzoni.

Bermúdez-Otero R. 2008: « Phonology is cyclic; cyclicity is emergent ». Handout de la communication présentée à la conférence *Models for Linguistic Model-Makers*, Groningen, 22 février 2008

[<http://www.bermudez-otero.com/Groningen.pdf>].

Bertinetto P.M. 1999: « La sillabazione dei nessi /sC/ in italiano: Un'eccezione alla tendenza universale? » in Benincà *et al.* (eds), 1999, 71-96.

— 2009: *Adequate imperfezioni. Sulla scelta di una lingua comune per l'Europa federata ed altri saggi di linguistica*. Palermo: Sellerio.

Bertinetto P.M., Loporcaro M. 2002: « Italian: A phonetic-phonological sketch ». *Quaderni del Laboratorio di linguistica* 3: 9-25.

Bisetto A. 1999: « Note sui composti VN dell'italiano » in Benincà *et al.* (eds) 1999, 503-538.

Bisetto A., Scalise S. 2005: « The classification of compounds ». *Lingue e linguaggio* IV.2: 319-332.

Blevins J.P. 2001: « Paradigmatic derivation ». *Transactions of the Philologic Society* 99.2: 211-222.

— 2006: « Word-based morphology ». *Journal of Linguistics* 42: 531-573.

Blevins J.P., Blevins J. 2009: « Introduction: Analogy in grammar » in Blevins & Blevins (eds) 2009: 1-12.

Blevins J.P., Blevins J. (eds) 2009: *Analogy in Grammar*. Oxford: Oxford University Press.

Bloomfield L. 1933: *Language*. New York: Rinehart and Winston.

Bolinger D. 1974: « Meaning and memory ». *Forum Linguisticum* 1: 1-14.

Bonami O., Boyé G. 2003: « Supplétion et classes flexionnelles dans la conjugaison du français » *Langages* 152: 102-126.

— 2005: « Construire le paradigme d'un adjectif ». *Recherches linguistiques de Vincennes* 34: 77-98.

— 2007: « Remarques sur les bases de la conjugaison » in E. Delais-Roussarie, L.

La morphophonologie de l'italien

- Labrunne (eds), *Des sons et des sens. Données et modèles en phonologie et en morphologie*. Paris: Hermès-Lavoisier, 77-90.
- Bonami O., Boyé G., Kerleroux F. 2009: « L'allomorphie radicale et la relation flexion-construction » in Fradin *et al.* (eds) 2009: 103-125.
- Booij G. 1977: *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Lisse: Peter de Ridder Presse.
- 1994: « Against split morphology » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1993*. Dordrecht: Kluwer, 27-49.
- 1996: « Inherent versus contextual inflection and the split morphology hypothesis » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 1-16.
- 2005: « Compounding and derivation: evidence for Construction Morphology » in W.U. Dressler, F. Rainer, D. Kastovsky, O. Pfeiffer (eds), *Morphology and Its Demarcations*. Amsterdam – Philadelphia: John Benjamins, 190-132.
- 2007: « Construction morphology and the lexicon » in Montermini *et al.* (eds) 2007, 33-34.
- 2008: « Lexical integrity as a formal universal: A constructionist view » in S. Scalise, E. Magni, A. Bisetto (eds), *Universals of Language Today*. Berlin: Springer, 83-100.
- 2009a: « La morphologie constructionnelle – un aperçu ». *Mémoires de la Société de Linguistique de Paris*. Nouvelle série XVII: 13-32.
- 2009b: « Phrasal names: a constructionalist analysis ». *Word Structure* 2: 219-240.
- Booij G., Lehmann C., Mugdan J. (eds), *Morphologie / Morphology. An International Handbook on Inflection and Word-Formation*. Berlin – New York, de Gruyter.
- Boyé G. 2000: *Problèmes de morpho-phonologie verbale en français, en espagnol et en italien*. Thèse de Doctorat. Paris: Université de Paris VII.
- 2006: « Suppletion » in K. Brown (ed.), *Encyclopedia of Language and Linguistics*. Oxford: Elsevier, 297-299.
- à paraître: « Régularité et classes flexionnelles dans la conjugaison du français » in M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat, *Des unités morphologiques au*

lexique.

- Boyé G., Cabredo Hofherr P. 2008: « Defectivity as stem suppletion in French and Spanish verbs ». Communication présentée au workshop *Defective Paradigms: Missing Forms and What They Tell Us*. Londres, 10-11 avril 2008 [<http://www.defectiveness.surrey.ac.uk/Presentations/Boye&Cabredo-Hofherr.pdf>].
- Boyé G., Montermini F. 2007: « Regularity and irregularity in Italian verbal inflection ». Poster présenté au 15th Manchester Phonology Meeting. Manchester, 24-26 mai 2007.
- Burani C., Arduino L.S. 2004: « Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns ». *Brain and Language* 90: 318-325.
- Burzio L. 1994: *Principles of English Stress*. Cambridge: Cambridge University Press.
- 1996: « Surface constraints versus underlying representations » in J. Durand, B. Laks (eds), *Current Trends in Phonology: Models and Methods*. Salford: University of Salford, 97-122.
- 2002: « Surface-to-surface morphology: when your representations turn into constraints » in P. Boucher (ed.), *Many Morphologies*. Somerville, Ma: Cascadilla Press, 142-177.
- ms.: « Lexicon and grammar: unequal but inseparable ». Manuscrit. Johns Hopkins University [<http://web.jhu.edu/cogsci/templates/images/pdf/burzio/LexGramPaper05.pdf>].
- Bybee J.L. 1985: *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- 1988: « Morphology as lexical organization » in Hammond & Noonan (eds) 1988, 119-141.
- 1995: « Regular morphology and the lexicon ». *Language and Cognitive Processes* 10.5: 425-455.
- 2003: « Mechanisms of change in grammaticization » in Joseph & Janda (eds) 2003, 602-623.

- 2007: *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
- Calderone B., Herreros I., Pirrelli V. 2007: « Learning inflection: The importance of starting big ». *Lingue e linguaggio* VI.2: 175-199.
- Canepari L. 2008: *Dizionario di pronuncia italiana*
[http://venus.unive.it/canipa/pdf/DiPI_3_A-Z.pdf].
- Cappellaro C. 2010: « Pronominal overabundance and language use ». Communication présentée au 14th International Morphology Meeting, Budapest, 13-16 mai 2010.
- Carstairs A. 1987: *Allomorphy in Inflection*. London: Croom Helm.
- 1988: « Some implications of phonologically conditioned suppletion » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1988*. Dordrecht: Kluwer, 67-94.
- Carstairs-McCarthy A. 1992: *Current Morphology*. London – New York: Routledge.
- Celata C., Bertinetto P.M. 2005: « Lexical access in Italian: words with and without palatalization ». *Lingue e linguaggio* 4.2: 293-318.
- Chomsky N. 1965: *Aspects of the Theory of Syntax*. Cambridge, Ma: MIT Press.
- 1993: *Language and Thought*. Wakefield, RI / London: Moyer Bell.
- 1995: *The Minimalist Program*. Cambridge, Ma: MIT Press.
- Chomsky N., Halle M. 1968: *The Sound Pattern of English*. New York: Harper and Row.
- Corbett G.G. 2000: *Number*. Cambridge: Cambridge University Press.
- 2007a: « Canonical typology, suppletion, and possible words ». *Language* 83.1: 8-42.
- 2007b: « Deponency, syncretism, and what lies between » in M. Baerman, G.G. Corbett, D. Brown, A. Hippisley (eds), *Deponency and Morphological Mismatches*. Oxford: Oxford University Press, 21-43.
- Corbin D. 1987: *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer.
- 1992: « Hypothèses sur les frontières de la composition nominale ». *Cahiers de grammaire* 17: 26-55.
- 1997: « Entre les mots possibles et les mots existants: les unités lexicales à faible

- probabilité d'actualisation » in D. Corbin, B. Fradin, B. Habert, F. Kerleroux, M. Plénat (eds), *Silexicales 1. Mots possibles et mots existants*. Villeneuve d'Ascq: Université de Lille III, 79-89.
- 1999: « Pour une théorie sémantique de la catégorisation affixale » *Faits de langue* 14: 65-77.
- 2004: « Préfixes et suffixes: du sens aux catégories » *Journal of French Language Studies* 11.1: 41-69.
- Covington M.A. 1981: *Evidence for Lexicalism. A Critical Review*. Bloomington: Indiana Linguistic Club
[<http://www.ai.uga.edu/mc/research/EvidenceForLexicalism.pdf>].
- Crocco Galèas G. 1991: *Gli etnici italiani. Studio di morfologia naturale*. Padova: Unipress.
- 1998: « La base dei processi derivativi in italiano ». *Studi di grammatica italiana* XVII: 245-272.
- Croft W. 1991: *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: Chicago University Press.
- 2000: « Parts of speech as language universals and as language-particular categories » in P.M. Vogel, B. Comrie (eds), *Approaches to the Typology of Word Classes*. Berlin – New York, Mouton de Gruyter, 65-102.
- Cruse D.A. 1986: *Lexical Semantics*. Cambridge: Cambridge University Press.
- Davis S. 1990: « Italian onset structure and the distribution of *il* and *lo* ». *Linguistics* 28: 43-55.
- 1999: « On the representation of initial geminates ». *Phonology* 16.1: 93-104.
- De Lacy P. 2006: *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.
- Dell F. 1973: *Les règles et les sons. Introduction à la phonologie générative*. Paris: Hermann.
- Den Os E., Kager R. 1986: « Extrametricality and stress in Spanish and Italian ». *Lingua* 69: 23-48.
- Derwing B.L., Skousen R. 1988: « Morphology in the mental lexicon: A new look at analogy » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1989*. Dordrecht: Kluwer, 56-71.

- DiFabio E. 1990: *The Morphology of the Verbal Infix /-isk/ in Italian and Romance*. PhD Dissertation. Harvard University.
- D'Imperio M., Rosenthal S. 1999: « Phonetics and phonology of main stress in Italian ». *Phonology* 16: 1-28.
- Di Sciullo M., Williams E. 1987: *On the Definition of Word*. Cambridge, Ma: MIT Press.
- Dixon R.M.W., Aikhenvald A.Y. 2002: « Word: a typological framework » in R.M.W. Dixon, A.Y. Aikhenvald (eds), *Word: A Cross-Linguistic Typology*. Cambridge, Cambridge University Press: 1-41.
- DOP: *Dizionario italiano multimediale e multilingue d'ortografia e di pronunzia* [<http://www.dizionario.rai.it/>].
- Downing P. 1977: « On the creation and use of English compound nouns ». *Language* 53.4: 810-842.
- Dressler W.U. 1985a: *Morphonology: The Dynamics of Derivation*. Ann Arbor: Karona.
- 1985b: « On the predictiveness of Natural Morphology ». *Journal of Linguistics* 21.2: 321-337.
- 1985c: « Suppletion in word formation » in J. Fisiak (ed.), *Historical Semantics. Historical Word-Formation*. Berlin – New York: Mouton, 97-112.
- 1988: « Preferences vs. strict universals in morphology: word-based rules » in Hammond & Noonan (eds) 1988, 143-154.
- 2005a: « Morphological typology and first language acquisition: some mutual challenges » in G. Booij, E. Guevara, A. Ralli, S.C. Sgroi, S. Scalise (eds), *Morphology and Linguistic Typology. On-line Proceedings of the 4th Mediterranean Morphology Meeting (MMM4). Catania, 21-23 September 2003*. Bologna: Università degli Studi di Bologna: 7-20
[<http://mmm.lingue.unibo.it/mmm-proc/MMM4/007-020-Dressler-MMM4.pdf>].
- 2005b: « Word formation in Natural Morphology » in Štekauer & Lieber (eds) 2005, 267-284.
- 2006: « Compound types » in G. Libben, G. Jarema (eds), *The Representation and Processing of Compound Words*. Oxford: Oxford University Press, 23-44.

- 2009: « Morphologie dynamique et statique des diminutifs ». *Mémoires de la Société de Linguistique de Paris*. Nouvelle série XVII: 141-154.
- Dressler W.U., Kilani-Schoch M., Spina R., Thornton A.M. 2003: « Le classi di coniugazione in italiano e in francese » in Giacomo Marcellesi & Rocchetti (eds) 2003, 397-416.
- Dressler W.U., Libben G., Stark J., Pons C., Jarema G. 2001: « The processing of interfixed Germanic compounds » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1999*. Dordrecht: Kluwer, 185-220.
- Dressler W.U., Mayerthaler W., Panagl O., Rennison J. 1987: *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins.
- Dressler W.U., Merlini Barbaresi L. 1991: « Interradical interfixes: contact and contrast » in V. Ivir, D. Kalogjera (eds), *Languages in Contact and Contrast: Essays in Contact Linguistics*. Berlin: Mouton de Gruyter, 133-145.
- 1994: *Morphopragmatics. Diminutives and Intensifiers in Italian, German and Other Languages*. Berlin: Mouton de Gruyter.
- Dressler W.U., Thornton A.M. 1991: « Doppie basi e binarismo nella morfologia italiana ». *Rivista di linguistica* 3.1: 3-22.
- Durand J. 2009: « On the scope of linguistics: data, intuitions, corpora » in Y. Kawaguchi, M. Minegishi, J. Durand (eds), *Corpus and Variation in Linguistic Description and Language Education*. Amsterdam / Philadelphia: John Benjamins, 25-52.
- Eddington D. 2000: « Spanish stress assignment within the Analogical Modeling of Language ». *Language* 76: 92-109.
- Fabrizio C. 2010: « What's the meaning of a noun converted into a verb? A semantic hypothesis about Italian denominal verbs from conversion ». Communication présentée au colloque *Triple. Word Classes. Nature, Typology, Computational Representation*, Rome, 24-26 mars 2010.
- Floricić F. 2008: « The Italian verb-noun anthroponimic compounds at the syntax / morphology interface ». *Morphology* 18: 167-193.
- Ford A., Singh R., Martohardjono R. 1997: *Pace Panini. Towards a Word-Based Theory of Morphology*. New York: Peter Lang.

- Fradin B. 2000: « Combining forms, blends and related phenomena » in U. Doleschal, A.M. Thornton (eds), *Extragrammatical and Marginal Morphology*. München: Lincom Europa, 11-59.
- 2003: *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.
- 2009: « Morphologie constructionnelle et sémantique ». *Mémoires de la Société de Linguistique de Paris*. Nouvelle série XVII: 89-118.
- Fradin B., Kerleroux F. (2009): « L'identité lexémique » in Fradin *et al.* (eds) 2009, 83-102.
- Fradin B., Kerleroux F., Plénat M. (eds) 2009: *Aperçus de morphologie du français*. Saint-Denis: Presses Universitaires de Vincennes.
- Fradin B., Montermini F. 2009: « La morphologie évaluative » in Fradin *et al.* (eds) 2009, 231-266.
- Fradin B., Montermini F., Plénat M. 2009: « Morphologie grammaticale et morphologie extragrammaticale » in Fradin *et al.* (eds) 2009, 21-45.
- Frege G. 1980 (1914): « Letter to Jourdain » cité dans Aronoff (2007).
- Gaeta L. 1999: « Un buon argomento contro il separatismo: il suffisso italiano *-anza/-enza* » in Benincà *et al.* (eds), 1999, 551-585.
- 2006: « Lexical integrity as a constructional strategy ». *Lingue e linguaggio* V.1: 67-82.
- Gaeta L., Ricca D. 2009: « *Composita solvantur*. Compounds as lexical units or morphological objects? ». *Italian Journal of Linguistics / Rivista di linguistica* 21.1: 35-70.
- Garvin P.L. 1954: « Delimitation of syntactic units ». *Language* 30: 345-348.
- Giacomo Marcellesi M., Rocchetti A. (eds), *Il verbo italiano. Studi diacronici, contrastivi, didattici. Atti del XXV Congresso della Società di Linguistica Italiana, Parigi, 20-22 settembre 2001*. Roma: Bulzoni.
- Gili Fivela B., Bertinetto P.M. 1999: « Incontri vocalici tra prefisso e radice (iato o dittongo?) ». *Archivio glottologico italiano* 84.2: 129-172.
- Giraud H., Montermini F., Pirrelli V. à paraître: « Processi cognitivi nell'analisi delle classi verbali dell'italiano: un approccio sperimentale » in Bertinetto P.M., Bambini V., Ricci I. *et al.* (eds), *Linguaggio e cervello / Semantica. Atti del XLII*

- Congresso della Società di Linguistica Italiana (Pisa, Scuola Normale Superiore, 25-27 settembre 2008)*. Roma: Bulzoni (CD-Rom).
- Greenberg J.H. 1963: « Some universals of grammar with particular reference to the order of meaningful elements », in J.H. Greenberg (ed.), *Universals of Language*. Cambridge: Cambridge University Press, 73-113.
- Goldsmith J. 1976: *Autosegmental Phonology*. PhD dissertation. Cambridge: MIT.
- Goldsmith J., O'Brien J. 2007: « Learning inflectional classes ». *Language Learning and Development* 24.4: 219-250.
- Grandi N. 2002: *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: FrancoAngeli.
- Grandi N., Montermini F. 2004: « Prefix-suffix neutrality in evaluative morphology. *Sicilorum Gymnasium* LVII.1: 183-201 [republié dans: Booij G., Guevara E., Ralli A., Scalise S., Sgroi S.C. (eds), *Morphology and Linguistic Typology. On-line Proceedings of the 4th Mediterranean Morphology Meeting. Catania, 21-23 September 2003*. Bologna: Università degli Studi di Bologna, 2005, 143-156].
- 2005: « Valutativi suffissali e valutativi prefissali: un'unica categoria? » in Grossmann M., Thornton A.M. (eds), *La formazione delle parole. Atti del XXXVII Congresso della Società di Linguistica Italiana, L'Aquila, 25-27 settembre 2003*. Roma: Bulzoni, 271-287.
- Greenberg J.H. 1963: « Some universals of grammar with particular reference to the order of meaningful elements » in Greenberg J.H. (ed.), *Universals of Language*. Cambridge, Ma: MIT Press, 73-113.
- Grossmann M., Rainer F. 2004: « Introduzione » in Grossmann & Rainer (eds) 2004, 1-30.
- Grossmann M., Rainer F. (eds) 2004: *La formazione delle parole in italiano*. Tübingen: Niemeyer.
- Guerrero A. 2008: *Analyse morphophonologique de la flexion et de la dérivation en catalan central*. Mémoire de Master 2. Toulouse: Université de Toulouse le Mirail.
- Guerzoni E. 2000: « Stress and morphology in the Italian verb system ». Communication présentée au colloque *Going Romance 2000*. Utrecht, 30 novembre – 1 décembre 2000.

- Guevara E., Scalise S. 2008: « Searching for universals in compounding » in S. Scalise, E. Magni, A. Bisetto (eds), *Universals of Language Today*. Amsterdam: Springer, 101-128.
- Halle M. 1973: « Prolegomena to a theory of word formation ». *Linguistic Inquiry* 4.1: 3-16.
- Halle M., Marantz A. 1993: « Distributed morphology and the pieces of inflection » in K. Halle, S.J. Keyser (eds), *The View from Building 20*. Cambridge, Ma: MIT Press, 111-176.
- Hammond M., Noonan M. (eds) 1988: *Theoretical Morphology. Approaches in Modern Linguistics*. San Diego: Academic Press.
- Harris Z. 1951: *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
- 1954: « Distributional structure ». *Word* 10.2-3: 146-162.
- Haspelmath M. 2002: *Understanding Morphology*. London: Arnold.
- 2007: « Pre-established categories don't exist – Consequences for language description and typology ». *Linguistic Typology* 11.1: 119-132.
- 2009: « The indeterminacy of word segmentation and the nature of morphology and syntax ». Communication présentée à la conférence *Morphology of the World's Languages*, Leipzig, 11-13 juin 2009
[http://www.eva.mpg.de/lingua/staff/haspelmath/pdf/Leipzig09_MOWL.pdf].
- Hathout N. 2009: *Contribution à la description morphologique du lexique et à l'approche extensive en morphologie*. Mémoire d'habilitation à diriger des recherches. Université de Toulouse.
- Hathout N., Plénat M., Tanguy L. 2003: « Enquête sur les dérivés en *-able* ». *Cahiers de grammaire* 28: 49-90.
- Hathout N. Montermini F., Tanguy L. 2008: « Extensive data for morphology: Using the World Wide Web ». *Journal of French Language Studies* 18.1, 67-85.
- Hathout N., Namer F., Plénat M., Tanguy L. 2009: « La collecte et l'utilisation des données en morphologie » in Fradin *et al.* (eds) 2009, 267-287.
- Hockett C.F. 1954: « Two models of grammatical description ». *Word* 10: 210-234.

- 1958: *A Course in Modern Linguistics*. New York: McMillan.
- Hohenhaus P. 2005: « Lexicalization and institutionalization » in Štekauer & Lieber (eds) 2005, 353-373.
- Iacobini C. 1999: « I prefissi dell'italiano » in Benincà *et al.* (eds) 1999, 369-399.
- 2000: « Base and direction of derivation » in Booij *et al.* (eds) 2000, 865-876.
- 2004: « Prefissazione » in Grossmann & Rainer (eds) 2004, 97-164.
- Jackendoff R. 1975: « Morphological and semantic regularities in the lexicon ». *Language* 51.3, 639-671.
- 1999: *The Architecture of the Language Faculty*. Cambridge, Ma: MIT Press.
- 2002: *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- 2008: « Construction after construction and its theoretical challenges ». *Language* 84:8-28.
- Jacobs H. 1994: « How optimal is Italian stress? » in R. Bok-Benema, C. Cremers (eds), *Linguistics in the Netherlands 1994*. Amsterdam / Philadelphia: John Benjamins, 61-70.
- Janda R.D. 1983: « 'Morphemes' aren't something that grows on trees: Morphology as more the phonology than the syntax of words » in J.F. Richardson, M. Marks, A. Chukerman (eds), *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*. Chicago: Chicago Linguistic Society, 79-95.
- Joseph B.D., Janda R.D. (eds) 2003: *The Handbook of Historical Linguistics*. Oxford: Blackwell.
- Kempson R.M., Quirk R. 1971: « Controlled activation of latent contrast ». *Language* 47.3: 548-572.
- Kenneally C. 2007: *The First Word. The Search for the Origin of Language*. London: Penguin.
- Kenstowicz M. Zuraw K. 2002: « Statistical generalizations in the stress distribution of Italian and Spanish ». Research report. Cambridge, MA: MIT.
[http://web.mit.edu/linguistics/people/faculty/kenstowicz/statistical_generalizations.pdf]

- Kiparsky P. 1982: « From cyclic phonology to lexical phonology » in H. van der Hulst, N. Smith (eds), *The Structure of Phonological Representations*. Dordrecht: Foris, 130-175.
- 1996: « Allomorphy or morphophonology? » in R. Singh (ed.), *Trubetzkoy's Orphan. Proceedings of the Montréal Roundtable « Morphology: Contemporary Responses »*. Amsterdam: John Benjamins, 13-31.
- Kleiber G. 1984: « Dénomination et relations dénominatives ». *Langages* 76: 77-94.
- Krämer M. 2009a: « Crypto-variation in Italian velar palatalization » in P.J. Masullo, E. O'Rourke, C.H. Huang (eds), *Romance Linguistics: Structures, Interfaces and Microparametric Variation*. Amsterdam: John Benjamins.
- 2009b: « Main stress in Italian nonce nouns » in D. Torck, W.L. Wetzels (eds), *Romance Languages and Linguistic Theory 2006*. Amsterdam / Philadelphia: John Benjamins.
- 2009c: *The Phonology of Italian*. Oxford: Oxford University Press.
- Krott A. 2009: « The role of analogy for compound words » in Blevins & Blevins (eds) 2009, 118-136.
- Larsen U.B. 1998: « Vowel length, *raddoppiamento sintattico* and the selection of the definite article in modern Italian » in P. Sauzet (ed.), *Langues et grammaire II & III. Phonologie*. Paris: Université de Paris 8, 87-102.
- Libben G., Boniecki M., Marlies M., Mittermann K., Korecky-Kröll K., Dressler W.U. 2009: « Interfixation in German compounds: What factors govern acceptability judgements? ». *Italian Journal of Linguistics / Rivista di linguistica* 21.1: 149-180.
- Lieber R. 1981: *On the Organization of the Lexicon*. PhD dissertation. Cambridge: MIT.
- 1992: *Deconstructing Morphology*. Chicago: University of Chicago Press.
- Lieber R., Scalise S. 2006: « The Lexical Integrity Hypothesis in a new theoretical universe ». *Lingue e linguaggio* 1: 7-32.
- Lignon S. 2000: *La suffixation en -ien. Aspects sémantiques et phonologiques*. Thèse de Doctorat. Université de Toulouse II le Mirail.
- Lignon S., Montermini F. 2008: « Constraints and analogy in the lexicon. An analysis of *-ien/-iano* in French and Italian. Communication présentée au 13th International

- Morphology Meeting, Vienne, 3-6 février 2008.
- Lignon S., Plénat M. 2009: « Echangisme suffixal et contraintes phonologiques » in Fradin *et al.* (eds) 2009: 65-81.
- LIP: De Mauro T., Mancini F., Vedovelli M., Voghera M. 1993: *Lessico di frequenza dell'italiano parlato*. Milano: Etas libri [disponible en ligne à l'adresse: <http://badip.uni-graz.at/index.php>].
- Loporcaro M. 1996: « On the analysis of geminates in standard Italian and Italian dialects » in B. Hurch, R.A. Rhodes (eds), *Natural Phonology: The State of the Art*. Berlin / New York: Mouton de Gruyter, 153-188.
- Lyons J. 1966: *Structural Semantics*. Oxford: Basil Blackwell.
- 1977: *Semantics*. Cambridge: Cambridge University Press.
- Maiden M. 1992: « Irregularity as a determinant of morphological change ». *Journal of Linguistics* 28: 285-312.
- 2003: « Il verbo italo-romanzo: verso una storia autenticamente morfologica » in Giacomo-Marcellesi & Rocchetti (eds) 2003, 3-21.
- Maiden M., O'Neill P. 2008: « On defective morphemes ». Communication présentée au workshop *Defective Paradigms: Missing Forms and What They Tell Us*. Londres, 10-11 avril 2008
- [http://www.defectiveness.surrey.ac.uk/Presentations/Maiden&O'Neill-Morphomic_Defectiveness.pdf].
- Mancini F., Voghera M. 1994: « Lunghezza, tipi di sillabe e accento in italiano » in T. De Mauro (ed.), *Come parlano gli italiani*. Firenze: La Nuova Italia, 217-244.
- Marotta G. 1987: « Dittongo e iato in italiano: una difficile discriminazione ». *Annali della Scuola Normale Superiore di Pisa*. Serie III, XVII: 847-887.
- 1999: « Degenerate feet nella fonologia metrica dell'italiano » in Benincà *et al.* (eds) 1999, 97-116.
- Martinet A. 1967: *Eléments de linguistique générale*. Paris: Armand Colin.
- Masini F. 2006: « Binomial constructions: Inheritance, specification and subregularities ». *Lingue e linguaggio* V.2: 207-232.
- Masini F., Thornton A.M. 2008: « Italian VeV lexical constructions » in A. Ralli, G.

- Booij, S. Scalise, A. Karasimos (eds), *On-Line Proceedings of the 6th Mediterranean Morphology Meeting. Ithaca, 27-30 September 2007*. Patras: University of Patras, 148-189
[http://www.philology.upatras.gr/LMGD/el/research/downloads/MMM6_Proceedings.pdf].
- Matthews P.H. 1972: *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.
- 1981: « Present stem alternations in Italian » in H. Geckeler, B. Schlieben-Lange, J. Trabant, H. Weydt (eds), *Logos Semantikos. Studia linguistica in honorem Eugenio Coseriu*. Vol. 4, *Gramatik*. Madrid: Gredos / New York – Berlin: de Gruyter, 57-65.
- 1991 (1974): *Morphology*, 2nd edition. Cambridge: Cambridge University Press.
- McCarthy J.J. 1982: « Prosodic templates, morphemic templates, and morphemic tiers » in H. van der Hulst, N. Smith (eds), *The Structure of Phonological Representations*, Part I. Dordrecht: Foris, 191-223.
- 1986: « OCP effects: gemination and antigemination ». *Linguistic Inquiry* 17: 207-263.
- 2005: « Optimal paradigms » in L. Downing, T.A. Hall, R. Raffelsiefen (eds), *Paradigms in Phonological Theory*. Oxford: Oxford University Press, 170-210.
- McCarthy J.J., Prince A. 2001 (1993): *Prosodic Morphology. Constraint Interaction and Satisfaction*. Rutgers Optimality Archive.
[<http://roa.rutgers.edu/view.php3?id=590>].
- McCrary K. 2002: « Syllable structure vs. segmental phonotactics: Geminate and clusters in Italian revisited » in A. Agwuele, H. Park (eds), *TLS VII 2002 Proceedings*. Austin: The University of Texas at Austin
[http://uts.cc.utexas.edu/~tls/2002tls/Kristie_McCrary.pdf].
- Mel'čuk I.A. 2006: *Aspects of the Theory of Morphology*. Berlin – New York: Mouton de Gruyter.
- Migliorini B. 1963: « I prefissoidi (il tipo *aeromobile, radiodiffusione*) » in B. Migliorini, *Saggi sulla lingua del Novecento*. Firenze: Sansoni, 9-60.

- 1988 (1958): *Storia della lingua italiana*. Firenze: Sansoni.
- Mioni A.M. 1993: « Fonetica e fonologia » in A.A. Sobrero (ed.), *Introduzione all'italiano contemporaneo. Le strutture*. Roma / Bari: Laterza, 101-139.
- Mithun M. 2010: « Constraints on compounds and incorporation » in Scalise & Vogel (eds) 2010, 37-56.
- Mohanan K.P., Mohanan T. 2003: « Universal and language-particular constraints in OT-LFG » in M. Butt, T. Holloway King (eds), *Proceedings of the LFG '03 Conference*. Stanford: CSLI Publication
[<http://csli-publications.stanford.edu/LFG/8/lfg03momo1.pdf>].
- Montermini F. 1997: « Materiali per uno studio del lessico di origine inglese nel linguaggio giovanile russo ». *Slavia* 3: 160-178.
- 1999: « L'apocope en italien en tant que phénomène de morphologie évaluative » in Corbin D., Dal G., Fradin B., Habert B., Kerleroux F., Plénat M., Roché M. (eds), *Silexicales 2. La morphologie des dérivés évaluatifs (2^e rencontres). Actes du colloque de Toulouse (29-30 avril 1999)*. Villeneuve d'Ascq: Université de Lille III, 149-159.
- 2001: « The Unitary Base Hypothesis and the semantics of word formation rules » in Bouillon P., Kanzaki K. (eds), *First International Workshop on Generative Approaches to the Lexicon. April 26-28, 2001, Geneva, Switzerland*. Genève: Université de Genève.
- 2002: « Formazione delle parole e variazione sociolinguistica: il caso dell'apocope in italiano » in Bauer R., Goebel H. (eds), *Parallela IX. Testo – variazione – informatica / Text – Variation – Informatik*. Wilhelmsfeld: Gottfried Egert Verlag, 305-320.
- 2003a: « Appunti sulla cancellazione di vocale in derivazione » in Bisetto A., Iacobini C., Thornton A.M. (eds), *Scritti di morfologia in onore di Sergio Scalise in occasione del suo 60° compleanno*. Cesena: Caissa Italia, 171-188.
- 2003b: « Pour un traitement des rencontres vocaliques entre base et affixe en italien ». *Cahiers de grammaire* 28: 113-134.
- 2003c: « Prefissi vs. preposizioni nell'italiano di oggi » in Maraschio N., Poggi Salani T. (eds), *Italia linguistica anno Mille Italia linguistica anno Duemila. Atti del XXXIV Congresso della Società di Linguistica Italiana. Firenze, 19-21 ottobre 2000*.

- Roma: Bulzoni, 639-651
- 2003d: « Suffixation et voyelles finales en italien » in Fradin B., Dal G., Hathout N., Kerleroux F., Plénat M., Roché M. (eds), *Silexicales 3. Les unités morphologiques (3^e rencontres). Actes du colloque de Villeneuve d'Ascq (19-21 septembre 2002)*. Villeneuve d'Ascq: Université de Lille III, 133-141.
- 2005: « La questione della capacità categorizzatrice dei prefissi. Uno studio su *anti-* in alcune lingue europee » in N. Grandi (ed.), *Morfologia e dintorni. Studi di linguistica tipologica e acquisizionale*. Milano: FrancoAngeli, 86-104.
- 2006a: « A new look on word internal anaphora on the basis of Italian data ». *Lingue e linguaggio* V.1: 127-148.
- 2006b: « La représentation phonologique des unités lexicales et la question de l'allomorphie. Sur quelques affixes problématiques de l'italien ». *Cahiers de grammaire* 30: 291-305
- 2007: « Hypocoristiques et minimalité en russe » in Delais-Roussarie E., Labrune L. (eds), *Des sons et des sens. Données et modèles en phonologie et en morphologie*. Paris: Hermès / Lavoisier, 2007, 199-213.
- 2008a: « La composition en italien dans un cadre de morphologie lexématique » in Amiot D. (ed.), *La composition dans une perspective typologique*. Artois: Arras Presses Université, 161-187
- 2008b: *Il lato sinistro della morfologia. La prefissazione in italiano e nelle lingue del mondo*. Milano: FrancoAngeli.
- 2010: « Units in compounding » in Scalise & Vogel (eds) 2010, 77-92.
- Montermini F., Boyé G., Hathout N. (eds) 2007: *Selected Proceedings of the 5th Décembrettes. Morphology in Toulouse*. Somerville, Ma: Cascadilla Press.
- Montermini F., Boyé G., Tseng J. (eds) 2009: *Selected Proceedings of the 6th Décembrettes. Morphology in Bordeaux*. Somerville, Ma: Cascadilla Press.
- Morén B. 2001: *Distinctiveness, Coercition and Sonority: A Unified Theory of Weight*. New York: Routledge.
- Morin Y.C. 1972: « The phonology of echo-words in French ». *Language* 48.1: 97-108.
- 1987: « Remarques sur l'organisation de la flexion des verbes français ». *ITL* –

Review of Applied Linguistics 77: 13-91.

- Muljačić Ž. 1969: *Fonologia generale e fonologia della lingua italiana*. Bologna: Il Mulino.
- Nespor M., Vogel I. 1986: *Prosodic Phonology*. Dordrecht: Foris.
- Nevins A. à paraître: « Phonologically-conditioned allomorph selection » in M. van Oostendorp, C.J. Ewen, E. Hume, K. Rice (eds), *The Blackwell Companion to Phonology*. London: Blackwell
- [<http://www.people.fas.harvard.edu/~nevins/allomorphy-chapter.pdf>].
- Nida E. 1949: *Morphology. The Descriptive Analysis of Words*. Ann Arbor: The University of Michigan Press.
- Ohannesian M., Pons C. 2009: « Shattering paradigms: An attempt to formalize pressures within subparadigms » in Montermini *et al.* (eds) 2009, 76-94.
- Passino D. 2007: « Stringhe fonologiche malformate all'incontro di radice e suffisso: il caso del femminile dei deverbali agentivi in *-ore* » in R. Maschi, N. Penello, P. Rizzolatti (eds), *Miscellanea di studi linguistici offerti a Laura Vanelli da amici e allievi padovani*. Udine: Forum, 147-159.
- Paster M.E. 2006: *Phonological Conditions on Affixation*. PhD Dissertation. University of California, Berkeley.
- Peperkamp S. 1995: « Prosodic constraints in the derivational morphology of Italian » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1994*. Dordrecht: Foris, 207-244.
- Perlmutter D. 1988: « The split morphology hypothesis: Evidence from Yiddish » in Hammond & Noonan (eds) 1988: 79-100.
- Pieri S. 1913: « Appunti toponomastici ». *Studj romanzi* X: 105-122.
- Pinker S. 1991: « Rules of language ». *Science* 253: 530-535.
- Pinker S., Jackendoff R. 2005: « The faculty of language: what's special about it? ». *Cognition* 95: 201-236.
- Pinker S., Ullman M.T. 2002: « The past and future of the past tense ». *Trends in Cognitive Sciences* 6.11: 456-463.
- Pirrelli V. 2000: *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale*

dell'italiano. Pisa: Istituti Editoriali e Poligrafici Internazionali.

- 2007: « On the cognitive autonomy of language morphology » in N. Hathout, F. Montermini (eds), *Morphologie à Toulouse. Actes du colloque international de morphologie 4e Décembrettes*. München: Lincom Europa, 245-269.
- Pirrelli V., Battista M. 2000: « The paradigmatic dimension of stem allomorphy in Italian ». *Rivista di linguistica* 12.2: 307-380.
- Plag I. 1998: « The polysemy of *-ize* derivatives: On the role of semantics in word formation » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1997*. Dordrecht: Kluwer, 219-242.
- 2009: « Morphological complexity: Inflection classes and probabilistic allomorph selection ». Communication présentée au ESF Exploratory Workshop *Words in Action*. Pisa, 11-14 octobre 2009.
- Plank F. ms.: « The direction of derivation and the unity of polysemous lexical items: Thoughts on *EAT* as property, substance, state, process, and action, and whichever is basic ». Manuscript. Universität Konstanz
[http://ling.uni-konstanz.de/pages/home/a20_11/plank/fett.pdf].
- Plénat M. 1987: « Morphologie du passé simple et du passé composé des verbes de l'«autre» conjugaison ». *ITL – Review of Applied Linguistics* 77: 93-150.
- 2000: « Quelques thèmes de recherche actuels en morphologie ». *Cahiers de lexicologie* 77: 27-62.
- 2005: « *Rosinette, cousinette, starlinette, chipinette*. Décalage, infixation et épenthèse devant *-ette* » in I. Choi-Jonin, M. Bras, A. Dagnac, M. Rouquier (eds), *Questions de classification en linguistique. Mélanges offerts au professeur Christian Molinier*. Berne: Peter Lang, 275-298.
- 2008: « Le thème L de l'adjectif et du nom » in J. Durand, B. Habert, B. Laks (eds), *Congrès Mondial de Linguistique Française CMLF '08*. Paris: Institut de Linguistique Française, 1613-1626
[<http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08306.pdf>].
- 2009a: « Le conditionnement de l'allomorphie lexicale en français ». *Mémoires de la*

- Société de Linguistique de Paris*. Nouvelle série XVII: 119-140.
- 2009b: « Les contraintes de taille » in Fradin *et al.* (eds) 2009: 47-63.
- Plénat M., Lignon S., Serna N. 2002: « La conjecture de Pichon ». *Corpus et recherches linguistiques* 1: 105-150.
- Plénat M., Roché 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: Universitat Pompeu Fabra, 285-299.
- Postal P.M. 1969: « Anaphoric islands » in R.I. Binnick, A. Davison, G. Green, J. Morgan (eds), *Papers from the Fifth Regional Meeting*. Chicago: Chicago Linguistic Society, 205-239.
- Prince A., Smolensky P. 2002 (1993): *Optimality Theory. Constraints Interaction in Generative Grammar*. Rutgers Optimality Archive.
[<http://roa.rutgers.edu/view.php3?id=845>].
- Pustejovsky J. 1995: *The Generative Lexicon*. Cambridge, Ma: MIT Press.
- Quirk R. 1970: « Aspect and variant inflection in English verbs ». *Language* 46.2: 300-311.
- Rainer F. 1996a: « Copernicano e luterano: sul ruolo del latino nella derivazione deantroponimica italiana ». *Lingua nostra* LVII.1-2: 48-49.
- 1996b: « Inflection inside derivation: Evidence from Spanish and Portuguese » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1995*. Dordrecht: Kluwer, 83-91.
- 2001: « Compositionality and paradigmatically determined allomorphy in Italian word-formation » in C. Schaner-Wolles, J. Rennison, F. Neubarth (eds), *Naturally! Linguistic Studies in Honour of Wolfgang Ulrich Dressler Presented on the Occasion of His 60th Birthday*. Torino: Rosenberg & Sellier, 383-392.
- 2004: « Etnici » in Grossmann & Rainer (eds) 2004: 402-408.
- 2005: « Constraints on productivity » in Štekauer & Lieber (eds) 2005, 335-352.
- 2009: « La influencia latina, francesa y inglesa en el desarrollo del sufijo *-iano* » in F. Sánchez Miret (ed.), *Romanística sin complejos. Homenaje a Carmen Pensado*. Bern:

- Peter Lang, 237-258.
- Ramat P. 1990: « Définition de 'parole' e sua tipologia » in M. Beretta, P. Molinelli, A. Valentini (eds), *Parallela 4. Morfologia / Morphologie. Atti del V Incontro Italo-Austriaco della Società di Linguistica Italiana. Bergamo, 2-4 ottobre 1989*. Tübingen: Narr, 3-15.
- Renzi L. Vanelli L. 1993: « Storia e struttura dell'articolo italiano *il* » in Hilty G. (ed.), *Actes du XXe Congrès de linguistique et philologie romanes*. Tome III. Section IV. Tübingen: Francke, 293-305.
- Repetti L. 1993: « The integration of foreign loans in the phonology of Italian ». *Italica* 70.2: 182-196.
- Ricca D. 2005: « Al limite tra sintassi e morfologia: i composti aggettivali V-N dell'italiano » in Grossmann & Thornton (eds) 2005, 465-486.
- 2010: « Corpus data and theoretical implications: With special reference to Italian V-N compounds » in Scalise & Vogel (eds) 2010, 237-254.
- Rice C. 2007: « Gaps and repairs at the phonology-morphology interface ». *Journal of Linguistics* 43.1: 197-221.
- Roca I.M. 1999: « Stress in the Romance languages » in H.G. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe*. Berlin: Mouton de Gruyter, 659-811.
- Roché M. 2003: « De la 'bonne formation' des dérivés en *-on* ». *Cahiers de grammaire* 28: 91-112.
- 2004: « Mot construit ? Mot non construit ? Quelques réflexions à partir des dérivés en *-ier(e)* ». *Verbum* XXVI.4: 459-480.
- 2007: « Logique lexicale et morphologie: la dérivation en *-isme* » in Montermini *et al.* (eds) 2007, 45-58.
- 2008: « Structuration du lexique et principe d'économie: le cas des ethniques » in J. Durand, B. Habert, B. Laks (eds), *Congrès Mondial de Linguistique Française – CMLF '08*. Paris: Institut de Linguistique Française, 1571-1585
[<http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08064.pdf>].
- 2009: « Pour une morphologie *lexicale* ». *Mémoires de la Société de Linguistique de Paris*.

- Nouvelle série XVII: 65-87.
- à paraître: « Quelle morphologie ? » in M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat, *Des unités morphologiques au lexique*.
- Rohlf G. 1966: *Grammatica storica della lingua italiana e dei suoi dialetti. Fonetica*. Torino: Einaudi.
- 1968: *Grammatica storica della lingua italiana e dei suoi dialetti. Morfologia*. Torino: Einaudi.
- Russell K. 1997: « Optimality Theory and morphology » in D. Archangeli, D.T. Langendoen (eds), *Optimality Theory: An Overview*. London: Blackwell, 102-133.
- Sadler L., Arnold D.J. 1994: « Prenominal adjectives and the phrasal / lexical distinction ». *Journal of Linguistics* 30: 187-226.
- Scalise S. 1983: *Morfologia lessicale*. Padova: Clesp.
- 1984: *Generative Morphology*. Dordrecht: Foris.
- 1994: *Morfologia*. Bologna: Il Mulino.
- Scalise S., Guevara E. 2005: « The lexicalist approach to word-formation and the notion of the lexicon » in Štekauer & Lieber (eds) 2005, 147-187.
- Scalise S., Vogel I. (eds) 2010: *Cross-Disciplinary Issues in Compounding*. Amsterdam: John Benjamins.
- Schwarze C. 1999: « Inflectional classes in Lexical Functional Morphology: Latin -sk- and its evolution » in M. Butt, T. Holloway King (eds), *Proceedings of the LFG 99 Conference*. Stanford: CSLI Publications.
- Schweickard W. 1992: *Deonomastik. Ableitungen auf der Basis von Eigennamen im Französischen (unter vergleichender Berücksichtigung des Italienischen, Rumänischen und Spanischen)*. Tübingen: Niemeyer.
- Seidl C. 2004: « Deantroponimici » in Grossmann & Rainer (eds) 2004: 409-419.
- Selkirk E.O. 1982: *The Syntax of Words*. Cambridge, Ma: MIT Press.
- Sgroi S.C. 2003: « Per una ridefinizione di ‘confisso’: composti confissati, derivati confissati, parasintetici confissati vs etimi ibridi e incongrui ». *Quaderni di semantica* XXIV.1, 81-153.
- Skousen R. 1989: *Analogical Modeling of Language*. Dordrecht: Kluwer.

- Sluyters W. 1990: « Length and stress revisited: A metrical account of diphthongization, vowel lengthening, consonant gemination and word-final vowel epenthesis in Modern Italian ». *Probus* 2.65-102.
- Spagnoletti C., Dominicy M. 1992: « L'accent italien et la cliticisation de la terminaison verbale *no* ». *Revue québécoise de linguistique* 21.2: 9-30.
- Spina R. 2007: *L'evoluzione della coniugazione italo-romanza. Uno studio di morfologia naturale*. Catania: ed.it.
- Štekauer P., Lieber R. (eds), *Handbook of Word-Formation*, Dordrecht: Springer.
- Stemberger J.P., MacWhinney B. 1986: « Frequency and lexical storage of regularly inflected forms ». *Memory and Cognition* 14.1: 17-26.
- 1988: « Are inflected forms stored in the lexicon? » in Hammond & Noonan (eds) 1988, 101-116.
- Stump G.T. 2001: *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- Tekavčić P. 1972: *Grammatica storica dell'italiano*. Vol. I. *Fonematica*. Bologna: Il Mulino.
- Thornton A.M. 1988: *Sui nomina actionis in italiano*. Thèse de doctorat. Pisa: Università di Pisa.
- 2001: « Perché non possiamo dirci *dipietriani? ». *Rivista italiana di onomastica* VII.1: 128-132.
- 2004: « Conversione » in Grossmann & Rainer (eds) 2004, 501-533.
- 2005: *Morfologia*. Roma: Carocci.
- à paraître: « Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology » in M. Goldbach, M.O. Hinzelin, m. Maiden, J.C. Smith (eds), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*. Oxford: Oxford University Press.
- TLFi: *Trésor de la langue française informatisé* [<http://atilf.atilf.fr/tlfi.htm>].
- Tranel B. 1981: *Concreteness in Generative Phonology. Evidence from French*. Berkeley – Los Angeles: University of California Press.
- Traugott E. 2003: « Constructions in grammaticalization » in Joseph & Janda (eds) 2003, 625-647.

- van de Veer B. 2006: *The Italian 'Mobile Diphthongs'. A Test Case for Experimental Phonetics and Phonological Theory*. Utrecht: Lot.
- Vincent N. 1988: « Non-linear phonology in diachronic perspective: Stress and word-structure in Latin and Italian » in P.M. Bertinetto, M. Loporcaro (eds), *Certamen Phonologicum. Papers from the 1987 Cortona Phonology Meeting*. Torino: Rosenberg & Sellier: 421-432.
- Vogel I. 1993: « Verbs in Italian morphology » in G. Booij, J. van Marle (eds), *Yearbook of Morphology 1993*. Dordrecht: Kluwer, 219-254.
- Vogel I., Napoli D.J. 1995: « The verbal component in Italian compounds » in J. Amastae, G. Goodall, M. Montalbetti (eds), *Contemporary Research in Romance Linguistics. Papers from the 22nd Linguistic Symposium on Romance Languages*. Amsterdam: Benjamins, 367-381.
- Ward G., Sproat R., McKoon G. 1991: « A pragmatic analysis of so-called anaphoric islands ». *Language* 67.3: 439-474.
- Westermann G., Plunkett K. 1986: « Connectionist models of inflection processing ». *Lingue e linguaggio* VI.2, 291-311.
- Xu Z. 2007: *Inflectional Morphology in Optimality Theory*. PhD dissertation. Stony Brook University.
- Yip M. 1982: « Reduplication and CV skeleta in Chinese secret languages ». *Linguistic Inquiry* 13.4: 637-661.
- 1998: « Identity avoidance in phonology and morphology » in S.G. Lapointe, D.K. Brentari, P.M. Farrell (eds), *Morphology and Its Relation to Phonology and Syntax*. Stanford, Ca: CSLI Publications, 216-246.
- Yu A.C.L. 2005: « Quantity, stress and reduplication in Washo » *Phonology* 22.3: 437-475.
- Zingarelli N. 1997: *Lo Zingarelli 1998. Vocabolario della lingua italiana*. Bologna: Zanichelli.
- Zwicky A. 1985: « Heads » *Journal of Linguistics* 21.1: 1-29.
- 1988: « Morphological rules, operations, and operation types » in A. Miller, J. Powers (eds), *Proceedings of the Fourth Eastern State Conference on Linguistics*,

La morphophonologie de l'italien

Columbus, Oh: The Ohio State University, 318-334.

— 1995: « Some choices in the theory of morphology » in R.D. Levine (ed.), *Formal Grammar: Theory and Implementation*. Oxford: Oxford University Press, 327-371.