

Université de Montréal

**La morphologie du pluriel nominal du persan
d'après la théorie Whole Word Morphology**

par
Pegah Faghiri

Département de linguistique et de traduction
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maîtrise ès arts (M. A.)
en linguistique

août, 2010

© Pegah Faghiri, 2010

Université de Montréal
Faculté des arts et des sciences

Ce mémoire intitulé :

**La morphologie du pluriel nominal du persan
d'après la théorie Whole Word Morphology**

présenté par
Pegah Faghiri

a été évalué par un jury composé des personnes suivantes :

.....Pr Mireille Tremblay.....
président-rapporteur

.....Pr Rajendra Singh
directeur de recherche

.....Pr Jean-Yves Morin.....
membre du jury

.....
examineur externe

.....
représentant du doyen

Résumé

Ce mémoire présente une étude de la morphologie de ce qui est généralement appelé le pluriel nominal du persan (parler de Téhéran) dans le cadre d'une théorie de la morphologie basée sur le mot : *Whole Word Morphology*, développée par Ford et Singh (1991). Ce modèle lexicaliste adopte une position plus forte que les modèles proposés par Aronoff (1976) et Anderson (1992) en n'admettant aucune opération morphologique sur des unités plus petites que le mot.

Selon cette théorie, une description morphologique consiste en l'énumération des Stratégies de Formation de Mots (SFM), licenciées chacune par au moins deux paires de mots ayant la même covariation formelle et sémantique. Tous les SFM suivent le même schéma.

Nous avons répertorié 49 SFM regroupant les pluriels et les collectifs. Nous constatons qu'il est difficile de saisir le pluriel nominal du persan en tant que catégorie syntaxique et que les différentes « marques du pluriel » présentées dans la littérature ne constituent pas un ensemble homogène : elles partagent toutes un sens de pluralité qui cependant varie d'une interprétation référentielle à une interprétation collective non-référentielle.

Cette étude vise la description de la compétence morphologique, ce qui ne dépend d'aucune considération extralinguistique. Nous argumentons notamment contre la dichotomie arabe/persan généralement admise dans la littérature. Nous avons également fourni des explications quant à la production des pluriels doubles et avons discuté de la variation supposée du fait d'un choix multiple de « marques du pluriel ».

Mots clés : morphologie basée sur le mot, stratégie de formation de mots, pluriel nominal, collectif, langue persane.

Abstract

This thesis presents a word-based study of what is generally called the nominal plural morphology of Persian (Tehrani dialect) within the framework of the Whole Word Morphology developed by Ford & Singh (1991). This lexicalist model takes up a stronger position than that proposed by Aronoff (1976) and Anderson (1992), by not allowing any morphological operation on units smaller than the word.

According to this theory a morphological description consists of the listing of the Word Formation Strategies (WFS), each licensed by at least two pairs of words having the same formal and semantic covariation. All WFS's follow the same schema.

We have listed 49 WFS's of plurals and collectives. We note that it is difficult to understand the import of the plural nominal as a syntactic category in Persian and that different "marks of plural" presented in the literature do not make a homogeneous unity: they all share a plurality meaning but it varies from referential interpretation to collective and non-referential interpretation.

This study's aim is to describe the morphological competence, which does not depend on any extra-linguistic criteria. In particular, we argue against the generally admitted Arabic/Persian dichotomy. We also provide explanation with regards to the utterance of double plurals and to the variation assumed since more than one choice of "plural marks" are available.

Key-words : word-based morphology, word formation strategy, nominal plural, collective, Persian.

Table des matières

Résumé	i
Abstract	iv
Table des matières	v
Introduction	1
1. Considérations préliminaires	3
1.1. La « langue interne » comme objet d'étude	4
1.2. Sur la morphologie générative	7
1.2.1. De Bloomfield à <i>Sound Patterns of English</i> (Chomsky et Halle 1968)	7
1.2.2. Après <i>Remarks on normalization</i> (Chomsky 1970)	7
1.2.2.1. Halle (1973)	14
1.2.2.2. Jackendoff (1975)	16
1.2.2.3. Kiparsky (1982)	18
1.2.2.4. Halle et Marantz (1993)	19
1.2.3. Morphologie basée sur le mot	20
1.2.3.1. Aronoff (1976)	22
1.2.3.2. Anderson (1992)	24
1.3. Word Whole Morphology	27
1.3.1. Le modèle WWM	28
1.3.2. Le mot, unité de base de la morphologie	29
1.3.3. Morphophonologie	30
1.3.4. Dérivation <i>vs</i> flexion	30
1.4. Clarifications sur le pluriel en persan	32
1.5. A propos du lexique	41
1.6. Aperçu sur les travaux existants sur le persan	43
1.6.1. L'apparition de la grammaire du persan	43
1.6.2. Appréciation générale	45
1.6.3. Sur la morphologie	46
1.6.4. Dichotomie arabe/persan	46

2. Description et discussion	52
2.1. Phonologie du persan	53
2.1.1. Inventaire des phonèmes	53
2.1.1.1. <i>Phonèmes syllabiques</i>	54
2.1.1.2. <i>Traits des voyelles</i>	54
2.1.1.3. <i>Phonèmes non-syllabiques</i>	55
2.1.1.4. <i>Quelques clarifications sur les segments</i>	56
2.1.2. Les alternances automatiques	56
2.1.2.1. <i>L'attaque branchante</i>	63
2.1.2.2. <i>L'attaque vide</i>	64
2.1.2.3. <i>La coda branchante</i>	65
2.1.2.4. <i>Les géminées</i>	65
2.2. Liste des stratégies de formation du pluriel nominal	66
2.3. Considérations sémantiques à prendre en compte	76
2.3.1. Les formes en <i>-an</i> , <i>-in</i> et <i>-un</i>	77
2.3.2. Les formes en <i>-iat</i> et <i>-dʒat</i>	81
2.4. Les pluriels doubles	84
2.5. La variation	88
Conclusion	91
Références	94

Introduction

Far earlier than the nineteenth century, words were already the very foundation of language study. (Matthews 1974 : 3)

Ce travail présente une description de la formation du pluriel nominal du persan dans le cadre d'une théorie de la morphologie basée sur le mot.

Bien que le mot ait été l'unité centrale de l'étude grammaticale traditionnelle, les linguistes, au cours du développement de la linguistique moderne, ont négligé ce statut au profit d'unités plus abstraites, *i.e.* le morphème, le lexème, etc. Ce travail est réalisé dans un cadre théorique qui plaide pour le nettoyage de la linguistique de son héritage paninien.

Dans le développement des études morphologiques, depuis l'institutionnalisation de la linguistique moderne par Saussure, deux notions ont été toujours centrales : les unités de base et les alternances. Nous pouvons faire remonter l'origine de la conception de ces notions dans la pratique actuelle de la morphologie jusqu'à Panini dont l'ontologie atomistique a fait de ce domaine « a matter of *divide and rule* » (Ford *et al.* 1997 : xiii). Le cadre théorique adopté dans ce travail, le *Whole Word Morphology* (ci-après WWM) est une théorie générative « radicalement non-Paninienne » (Ford *et al.* 1997 : xiii), qui résulte d'un examen critique de ces notions.

Le but de ce travail est de présenter une description la plus exhaustive possible de la morphologie du pluriel nominal du persan en tant que « langue interne » de l'auteur. Selon la théorie WWM, la morphologie dépend entièrement du lexique. Cependant, « le lexique » n'est qu'une fiction et, en réalité, il existe autant de lexiques que de locuteurs d'une langue qui dépendent de leur biographie particulière et qui sont constamment sujets aux changements. La description

présentée ici est extraite de mon lexique actuel du persan, elle peut également être généralisée à la population urbaine et éduquée de la ville de Téhéran.

Dans le cadre de cette théorie, une description morphologique équivaut à l'énumération des stratégies de formation de mots, un dispositif formel pour capter la relation en forme et sens entre deux mots d'une langue donnée. Chaque stratégie doit être licenciée par, au moins, deux paires de mots ayant la même covariance formelle et sémantique. A titre d'exemple :

$/X/_N \leftrightarrow /Xidan/_V$

tars 'peur' \leftrightarrow *tarsidan* 'avoir peur'

xab 'sommeil' \leftrightarrow *xabidan* 'dormir'

Ainsi, leur différence est représentée par des variables, *i.e.* *X*, et leur similitude par des constantes, *i.e.* *-idan*.

Ce travail est organisé en deux parties. La première partie comprend un ensemble de considérations préliminaires destinées à déterminer ce que cette tâche représente précisément. En premier lieu, je présente l'approche *I-language* vis-à-vis de l'approche *E-language* et, ensuite, un aperçu de la morphologie générative et du développement du regard critique sur le morphème comme unité de base de la morphologie, pour arriver enfin à la théorie WWM que je présente en détail. Après avoir présenté le cadre théorique, je fournis certaines clarifications afin de déterminer en quoi consiste la description présentée : tout d'abord, à propos de ce que signifie le pluriel nominal en persan, *i.e.* la raison d'être d'un tel regroupement de stratégies et, en second lieu, au niveau du lexique. Finalement, je présente un aperçu des travaux existants sur le persan.

Dans la seconde partie, je présente l'énumération des stratégies de formation du pluriel, en la faisant précéder d'un aperçu de la phonologie du persan. Par la suite, je présente quelques remarques d'ordre sémantique sur cette description et discute de phénomènes tels que le pluriel double et la variation.

1

Considérations préliminaires

1.1

La « langue interne » comme objet d'étude

Afin d'étudier une langue, il faut tenter de dissocier un ensemble de facteurs divers qui interfèrent avec la compétence sous-jacente pour déterminer la performance effective. (Chomsky 1969 : 126)

Une étude linguistique scientifique, dans le sens exact du terme, ne peut se faire sur une langue en tant qu'entité sociopolitique. La grammaire que l'on peut décrire concerne la langue interne, *I-language* (Cf. Chomsky 1986), et doit viser la compétence d'un locuteur plutôt que sa performance¹.

¹ Chomsky (1965) présente le cadre théorique de la théorie générative en faisant la distinction entre compétence et performance :

We must make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations). (Chomsky 1965 : 4)

Cette distinction renvoie à la distinction langue/parole élaborée par Saussure, en rejetant toutefois sa conception de la langue comme « merely a systematic inventory of items » (Chomsky 1965 : 4). En revanche, il faut reprendre la vision de Humboldt, c'est-à-dire comme « underlying competence as a system of generative processes » (Chomsky 1965 : 4). Selon Chomsky (1969), on trouve chez Humboldt l'expression la plus convaincante de l'importance accordée par les cartésiens à l'aspect créateur du langage (« le propre de cette faculté est d'ouvrir des possibilités sans limite et de ne dépendre d'aucun stimulus » (Chomsky 1969 : 20) : la caractérisation du langage comme « une production active » plutôt que comme « un produit sans vie ». Pour Humboldt, la seule vraie définition du langage est une

Il est bien évident que [...] une personne qui possède une langue a, d'une certaine manière, intériorisé le système de règles qui détermine et la forme phonétique de la phrase et son contenu sémantique intrinsèque ; cette personne a développé ce que nous appellerons une compétence linguistique spécifique. Mais il est clair que l'utilisation du langage telle qu'on l'observe – la performance effective – ne reflète pas seulement les relations intrinsèques entre le son et le sens établi par le système des règles linguistiques. La performance implique aussi bien de nombreux autres facteurs [...]. Des considérations extralinguistiques, portant sur le locuteur et sur la situation, jouent un rôle fondamental lorsqu'il s'agit de déterminer comment le discours est produit, identifié et compris. (Chomsky 1969 : 125-126)

Les travaux de Chomsky et le développement de la théorie de la grammaire universelle ont déplacé l'attention des études linguistiques de la langue externe vers la langue interne : “from the study of language regarded as an externalized object to the study of knowledge of language attained and internally represented in mind/brain” (Chomsky 1986 : 24).

Les recherches du type *E-language* considèrent la langue comme “a collection of actions, or utterances, or linguistic forms (words, sentences) paired with meanings” (Chomsky 1986 : 19)² et conçoivent une grammaire pour décrire les régularités que l'on trouve dans une telle collection. Ainsi, la description linguistique consiste à faire l'inventaire des unités de la langue et à en établir ensuite des règles qui régissent leurs relations. Selon cette approche, étudier une langue équivaut donc à réunir un corpus dans le but de faire apparaître des régularités formelles. C'est cette vision

définition « génétique » (cité par Chomsky 1969 : 41). La propriété fondamentale du langage est donc de « faire, à partir d'un moyen fini, un usage infini » (Chomsky 1969 : 42).

² Pour Bloomfield, une langue est “the totality of utterances that can be made in a speech-community”, où *utterance* se définit comme “act of speech” (Bloomfield 1926 : 155) et la grammaire d'une langue se constitue de “meaningful arrangement of forms” (Bloomfield 1933 : 163).

structuraliste de la langue et de la grammaire qui prédominait en pratique au moment où Chomsky publie son ouvrage fondateur de la grammaire générative qui représente une rupture radicale avec la vision structuraliste de la linguistique.

Dans une approche *I-language*, on s'occupe de "what exactly one knows when one knows a language" (Chomsky 1986 : 24) en prenant le langage comme une propriété de la pensée humaine "abstracting from sociopolitical and normative-teleological factors" (Chomsky 1986 : 28), plutôt que comme quelque chose d'externe, une construction artificielle, un « artefact », un « épiphénomène ».

Dans le sillage de l'approche de Chomsky, l'objet d'étude de ce travail porte sur la langue interne de l'auteure en tant que locutrice native du persan³ de Téhéran, que l'on peut généraliser à la population urbaine et éduquée de la ville de Téhéran.

³ Le persan est une langue indo-iranienne parlée en Iran, en Afghanistan, au Tadjikistan, en Ouzbékistan et au Bahreïn. Le persan pris en compte ici est celui parlé à Téhéran. En Iran, malgré la situation de multilinguisme, le persan est la seule langue officielle et la seule langue de l'enseignement.

1.2

Sur la morphologie générative

Dans ce mémoire, avant d'arriver à la présentation de la théorie WWM que j'adopte pour cette étude, je donne un aperçu, tout d'abord, du développement des théories morphologiques au sein de l'école générative et, par la suite, de l'émergence du regard critique sur le morphème. Pour ce faire, étant donné l'héritage structuraliste de cette école, je commence par présenter brièvement la pratique de la morphologie du structuralisme américain.

1.2.1 De Bloomfield à Chomsky et Halle

Chez Bloomfield, toute forme complexe est entièrement constituée de morphèmes et l'inventaire de tous les morphèmes d'une langue couplé avec l'inventaire des patrons de leurs arrangements – car une partie du sens est donnée par l'arrangement des formes – permet de décrire le sens de toutes les *utterances* d'une langue (Cf. Bloomfield 1933 : 163).

Bloomfield définit le morphème comme « a linguistic form which bears no partial phonetic-semantic resemblance to any other form » (1933 : 161) dont toute forme complexe est entièrement composée. C'est pourquoi, si l'on considère que *-berry* de *cranberry* est un morphème (étant donné *blackberry*, *strawberry*, etc.), *cran-* en est forcément un aussi. Bloomfield reconnaît qu'il n'est pas toujours évident de décider si, pour les formes phonétiquement identiques (par exemple, *straw-* dans *strawberry* et *strawflower*), une ressemblance sémantique existe, il considère cependant que cette difficulté est « part of universal difficulty of semantics » (1933 : 160). De plus, il souligne que le sens ne peut être analysé que par un « omniscient observer » (1933 :

162). C'est en appliquant cette logique jusqu'au bout que Bloomfield (1933 : 245) est amené à poser *gl-*, de *glitter*, *gleam*, *glow*, etc., comme un morphème (« root-forming morphemes of vague signification »), aussi contre-intuitif que cela puisse paraître.

La conception du morphème, en tant que « minimal same of form and meaning », exige donc une correspondance biunivoque entre la forme et le sens. Or, comment expliquer le fait que le même sens, par exemple la pluralité en anglais, soit réalisé par une série de formes différentes, *i.e.* *dog/dog[z]*, *cat/cats*, *house/hous[ɪz]*, *ox/oxen*, *sheep/sheep*, *focus/foci*, etc. Pour rendre compte de cette ambiguïté, les structuralistes ont été amenés à raffiner la définition du morphème par l'introduction d'un niveau d'analyse intermédiaire entre l'analyse phonémique et morphémique : le niveau morphophonémique⁴, composé de *morphes*⁵, des unités réelles de forme constante (Hocket 1947 : 323). Le morphème est ainsi conçu comme une classe d'allomorphes en distribution complémentaire.

La morphologie post-bloomfieldienne consiste en deux étapes : tout d'abord, l'identification des morphèmes ('morphemics', Cf. Harris 1942, repris par Hocket 1947 : 322) et, ensuite, la description des arrangements dans lesquels les morphèmes se produisent ('tacitcs'⁶, Cf. Hocket 1947 : 321). Pour ce faire, la première étape est de trouver les formes récurrentes ayant le même sens, les *morphes*, et la seconde étape

⁴ La morpho(pho)nologie est un terme introduit par Troubetzkoy pour « l'étude de l'emploi en morphologie des moyens phonologiques d'une langue » (Troubetzkoy 1949 : 337). J'y reviendrai plus loin.

⁵ Terme introduit par Hocket (1947 : 322) par analogie à *phones*.

⁶ Dans la tradition post-bloomfieldienne, la distinction entre la syntaxe et la morphologie n'est plus respectée et *tactics* s'applique indépendamment du domaine du mot ; les termes tels que *tactics*, *syntax*, *morphology*, *grammar* utilisés dans la littérature de cette tradition réfèrent presque tous au même phénomène (Cf. Matthews 1972 : 44).

consiste à regrouper les *morphes* qui ont le même sens et qui sont en distribution complémentaire dans un seul morphème ; par exemple : 'PLURIEL' {-z ; -en ; Ø, ...}.

Trois modèles ont été proposés pour la description morphologique : *Item-and-Process*, IP (Hocket 1954), qui est une approche procédurale, *Item-and-Arrangement*, IA⁷ (Hocket 1954), qui est une approche combinatoire, et *Word-and-Paradigme*, WP (Robins 1959), pour lequel, contrairement aux deux précédents, le mot est l'unité de base.

Dans les années 1960, avec l'arrivée de la théorie générative, la morphologie a perdu du terrain au profit de la syntaxe. Le module morphologique était inexistant dans la grammaire conçue par Chomsky qui se composait d'une composante syntaxique, d'une composante sémantique et d'une composante phonologique⁸. Le fait que « the ratio of what is learned to what is innate appeared higher in morphology than syntax » est une raison importante qui a causé la mise à l'écart de ce domaine dans une théorie dont l'intérêt portait sur les aspects innés plutôt que sur les aspects appris de la grammaire (Carstairs-McCarthy 1992 : 5).

En effet, bien que les années 1940 et le début des années 1950 ont été une époque de progrès pour la morphologie, les années 1960 ont été une « decade of syntax » où le

⁷ « IA has been formulated at least in part because of a feeling of dissatisfaction with the 'moving-part' or 'historical' analogy implicit in IP » (Hocket 1954 : 211).

⁸ Cf. entre autres : « A grammar contains a syntactic component which is a finite system of rules generating an infinite number of syntactic descriptions of sentences. Each such a syntactic description contains a deep structure and a surface structure that is partially determined by the deep structure underlying it. The semantic component of the grammar is a system of rules that assigns a semantic interpretation to each syntactic description, making essential reference to the deep structure and possibly taking into account certain aspects of surface structure as well. The phonological component of the grammar assigns a phonetic interpretation to the syntactic description, making reference only to properties of the surface structure, so far as we know » (Chomsky & Halle 1968 : 6-7).

mot n'était plus l'unité de base puisque le niveau où les relations syntaxique étaient traitées, c'est-à-dire la structure profonde, était considéré comme trop abstrait pour les « words to serve any usefull purpose » (Matthews 1974 : 4).

Dans les débuts du générativisme, la morphologie, telle que pratiquée par les structuralistes, a donc été distribuée entre la syntaxe et la phonologie – il ne restait alors pour le morphologue qu'à faire la liste des supplétions arbitraires.

Syntaxe :

La syntaxe s'occupait de l'organisation des éléments plus petits dans les constructions plus grandes sans faire de distinction entre le domaine du mot et le domaine au-dessus du mot⁹ ; il était supposé que la division du mot en unités plus petites se faisait selon les mêmes principes que la division du syntagme.

Seulement les mots « simples » étaient répertoriés et les mots « complexes » étaient construits dans la composante transformationnelle où les règles de réécriture et les règles transformationnelles manipulaient aussi bien les mots que les morphèmes ; la dérivation et la flexion ont été traitées pareillement. Par exemple, *destruction* ne constitue pas une entrée lexicale et résulte de la transformation de la nominalisation : *nom + destroy_v* (en combinaison avec une règle phonologique).

Néanmoins, quelques années plus tard, Chomsky met en cause l'approche transformationnelle dans son article paru en 1970 qui représente un tournant décisif

⁹ Comme nous l'avons déjà noté, la disparition de la distinction entre les deux domaines avait déjà débuté chez les distributionnalistes. Bien que Bloomfield ait conservé une division entre la morphologie et la syntaxe, « the constructions in which free forms appear in phrases differ very decidedly from the constructions in which the free or bound forms appear in words » (Bloomfield 1933 : 183), dans les modèles antérieurs, notamment chez Harris (1951), la morphologie « covers the analysis by distributional criteria of both » (Matthews 1993 : 86). Cf. le terme "tactics" mentionné plus haut.

dans l'histoire de la morphologie (Cf. Scalise et Guevara 2005 : 7), sur lequel je reviendrai dans la section suivante.

Phonologie :

La phonologie s'occupait donc de tout ce qui était relatif à la variation dans la forme (allomorphie) avec les règles phonologiques dont la forme devrait être (Halle 1962 : 54) :

$$A \rightarrow B \text{ in the environment } X \underset{Z}{\rule{1cm}{0.4pt}} Y$$

where A, B, X, Y, Z are symbols of a particular language or zero,
and « \rightarrow » can be read « is to be rewritten as ».

Les règles phonologiques de *Sound Patterns of English* (SPE) traitent toutes des alternances de la même manière, tandis qu'il s'agit d'ensembles fondamentalement différents ; certains relèvent des « processus naturels »¹⁰ qui s'appliquent de manière automatique et systématique, et d'autres représentent des règles arbitraires issues de la généralisation faite sur un ensemble limité de mots liés. Il faudrait donc séparer, d'une manière ou d'une autre, les alternances dues aux conditions purement phonologiques, *i.e.* les alternances automatiques, des alternances où une information grammaticale quelconque joue un rôle, *i.e.* les alternances non-automatiques qui ont été traditionnellement appelées morphophonologiques.

Au sein de l'école générative, différentes théories se sont développées en réaction à la manière dont la phonologie générative s'occupait des alternances et quant à la

¹⁰ Stampe (1972) a posé comme principe que les données substantives devraient toujours avoir plus d'importance que les préoccupations d'ordre formel en phonologie. Il a fondé la phonologie naturelle selon laquelle le domaine de la phonologie est limité à tout ce qui relève d'un processus naturel (par exemple, l'assimilation du lieu d'articulation ou le voisement) ; les processus innés et universels basés sur les contraintes d'ordre physiologique que l'enfant acquiert et dont la phonologie de sa langue maternelle lui apprend à en éliminer une partie.

démarcation du domaine de la phonologie *vs* la morphologie : la *Phonologie Naturelle Générative* (Cf. Hooper 1976, entre autres) a revendiqué un statut autonome pour la morphophonologie, la *Lexical Phonology* (Cf. Kiparsky 1982, entre autres) a proposé le traitement des règles phonologiques « in tandem » avec les règles morphologiques. Il faut noter que l'introduction de la morphophonologie en tant que module à côté de la phonologie et de la morphologie remonte à Troubetzkoy :

A côté de la phonologie, qui étudie le système des phonèmes considérés comme étant les idées acoustico-motrices, significatives dans une langue donnée, les plus simples, et de la morphologie, qui étudie le système des morphèmes, la grammaire doit comprendre encore un chapitre particulier, qui étudie l'utilisation morphologique des différences phonologiques, et qui peut être appelée la morphonologie. (Troubetzkoy 1929 : 85)

La morphophonologie, telle qu'a été postulée par Troubetzkoy, a ensuite été intégrée par le structuralisme américain et est devenue une partie intégrale de la grammaire dans la tradition bloomfieldienne, appelée également « morphophonémique ».

1.2.2 Après Chomsky, *Remarks on nominalization*

Le point de départ pour la « renaissance de la morphologie »¹¹ au sein de la grammaire générative est l'article de Chomsky, dans lequel il a soulevé plusieurs problèmes que le traitement transformationnel de la nominalisation déverbale engendre. Chomsky (1970) soutient qu'au moins une partie de la formation du mot, à cause des irrégularités sémantiques, doit être prise en charge par le lexique¹² plutôt que par la syntaxe, afin de conserver la portée générale de cette dernière. De cette

¹¹ « The revival of morphology (...) » (Carstairs-McCarthy 1992 : 3).

¹² Car « This range of variation and its rather accidental character are typical of lexical structure » (Chomsky 1970 : 189). Cf. la définition bloomfieldienne du lexique : « The lexicon is really an appendix of grammar, list of basic irregularities » (Bloomfield 1933 : 274).

observation découle « l'hypothèse lexicaliste » (vs « hypothèse transformationnaliste ») selon laquelle la syntaxe ne manipule pas la structure interne des mots.

L'hypothèse lexicaliste connaît deux versions : l'hypothèse forte de Jackendoff (1972, 1975) et de Halle (1973) exclut de la syntaxe tous les phénomènes morphologiques ou, comme cela a été formulé ultérieurement, considère le mot comme un « atome syntaxique » (Cf. Di Sciullo et Williams 1987), tandis que l'hypothèse faible de Aronoff (1976) et de Anderson (1982, 1992) maintient la distinction entre la flexion et la dérivation et garde la flexion pour le domaine de la syntaxe : « for the purposes of syntax, the word (sans inflection) is the minimal sign » (Aronoff 1976 : 9). En outre, certains lexicalistes comme Kiparsky (1982) ont traité la dérivation et la flexion dans le lexique et comme instance de même opération, mais, en les ordonnant de manière hiérarchique, ont maintenu la distinction entre les deux.

Plus récemment, Halle et Marantz (1993) ont, à l'extrême opposé, opté pour l'hypothèse transformationnelle et ont proposé la théorie de la *Morphologie Distribuée* selon laquelle la syntaxe manipule la structure interne des mots.

Au sein de l'approche lexicaliste, le « ré-établissement » (Anderson 1988 : 147) du domaine de la morphologie a été entrepris par Halle (1973), dans son article intitulé « Prolegomena to a theory of word formation », le premier ouvrage à fournir l'aperçu d'une approche générative de la morphologie, qui a été suivi du modèle du lexique de Jackendoff (1975). L'établissement de la morphologie générative est concrètement dû à Aronoff (1976) avec la première monographie dans le domaine intitulé *Word-formation in generative grammar* qui a de plus été le premier modèle génératif basé sur le mot.

Par la suite, je présente brièvement dans l'ordre chronologique les modèles de Halle (1973), Jackendoff (1975), Kiparsky (1982) et celui de Halle et Marantz (1993).

J'aborderai le modèle d'Aronoff (1976) dans la section suivante qui sera consacrée à la morphologie basée sur le mot.

1.2.2.1 Halle (1973)

Le point de départ de Halle est l'observation selon laquelle, à l'intérieur d'un mot, les morphèmes ne peuvent être mis ensemble que dans un ordre précis : *trans-form-at-ion-al* est un mot possible d'anglais, mais non *ion-trans-al-at-form* ni *al-form-at-ion-trans* (Halle 1973 : 3), et que la compétence linguistique d'un locuteur comprend des connaissances de ce type.

Dans la proposition de Halle, la morphologie se base sur trois composantes :

- la liste des morphèmes y compris les affixes (dont les affixes flexionnels¹³) accompagnés des informations grammaticales ainsi que combinatoires ;
- les règles de formation des mots, « *word formation rules* » (WFR)¹⁴, qui, couplées avec la liste des morphèmes, permettent de générer les « *potential words* » de la langue (Halle 1973 : 6), le trait [+/-Lexical Insertion] est employé pour distinguer les mots effectifs, « *actual words* », des mots

¹³ Halle traite de la flexion de la même manière que la dérivation et insiste sur le fait qu'il ne trouve aucune justification pour continuer la tradition (*i.e.* de séparer la morphologie flexionnelle). Il suppose qu'au niveau profond, au lieu d'un seul item lexical, tout ou partie de son paradigme flexionnel est inséré et que les mots fléchis inappropriés sont éliminés au niveau de la surface par un mécanisme appelé « *paradigme pressure* » (Halle 1973 : 9).

¹⁴ Halle propose deux sortes de WFR : les règles qui précisent la séquence linéaire des morphèmes (comme [STEM+al]_N basé sur les mots tels que *tot-al*) et celles qui rendent compte de la dérivation d'un mot à partir d'un autre (comme [ADJ(+i)+ty]_N basé sur les mots tels que *active-ity*).

possibles mais inexistants¹⁵ et, par ce biais, de rendre compte de la productivité d'un processus¹⁶, et ;

- un filtre qui fournit les informations imprédictibles et idiosyncrasiques¹⁷ ainsi que la valeur du trait [+/-Lexical Insertion]. La sortie de cette composante morphologique est le dictionnaire des mots effectifs de la langue – la morphologie étant conçue comme « producing a long list of words » (Halle 1973 : 9).

La transformation d'insertion lexicale est ainsi conçue comme la sélection des éléments dans ce dictionnaire et leur insertion aux places appropriées au niveau de la représentation sous-jacente de la phrase. Halle souligne que ces composantes ne sont pas « always fully involved in every speech act » et qu'une bonne partie du dictionnaire est mémorisée et qu'un locuteur n'a pas besoin de faire appel à la « word formation » composante que lorsqu'il entend un mot inconnu (*unfamiliar*) ou qu'il emploie un nouveau mot, *i.e.* qu'il vient d'inventer (Halle 1973 : 16).

Halle est obligé de supposer un filtre pour rendre compte des informations imprévisibles qui ne sont pas fournies par des règles de formation, étant donné qu'il prend la morphologie comme une composante qui **produit** la liste de « only and all » mots de la langue. Cependant, pour accomplir cette tâche, la composante morphologique doit avoir accès au dictionnaire. L'on remarque donc une sorte de circularité et de redondance entre le filtre et le dictionnaire. Par exemple, le mot

¹⁵ Par exemple **derival*, **arrivation* qui sont des résultats réguliers d'application des règles de formation de mot mais qui ne sont pas des mots d'anglais.

¹⁶ Les processus non-productifs génèrent plus d'éléments marqués [-Lexical insertion] que [+Lexical insertion].

¹⁷ Par exemple, le fait que *recital*, malgré sa formation (Xal = 'the fact of being V-ed'), veut dire 'a concert by a solist', ou que le mot *obesity* contrairement à *serenity*, *obscenity*, *sincerity*, etc. ne doit pas subir la règle phonologique de *trisyllabic shortening*.

recital est listé dans le dictionnaire comme '*a concert by a solist*' et, pour cela, il faut que le filtre fournisse cette irrégularité sémantique imprévisible. Mais d'où sort cette information si ce n'est depuis le dictionnaire même ? L'ambiguïté du modèle est que le dictionnaire est à la fois le résultat, *i.e.* la sortie, de la composante morphologique et supposé par celui-ci.

1.2.2.2 Jackendoff (1975)

Une autre réponse à l'hypothèse lexicaliste est Jackendoff (1975), « Morphological and Semantic regularities in the lexicon », qui, pour rendre compte des relations¹⁸ entre des mots comme *decide/decision* en dehors de l'hypothèse transformationnaliste, *i.e.* sans aucune relation transformationnelle entre les deux, propose une théorie du lexique basée sur les règles de redondance lexicale.

Selon la position transformationnelle, étant donné qu'une bonne partie de l'information dans le nom *decision* est déjà comprise dans le verbe *decide*, l'existence de deux entrées lexicales séparées est contre le principe d'économie. Toutefois, l'intervention de Chomsky (1970) a montré les problèmes du traitement transformationnel de la nominalisation. Or, une prise en compte lexicaliste crée forcément des redondances au niveau du nombre de symboles utilisés dans la grammaire (Cf. par exemple le modèle de Halle (1973), dans lequel certaines informations sont listées trois fois).

Pour résoudre le paradoxe d'économie, Jackendoff (1975) propose une autre méthode pour évaluer la complexité que la méthode courante du comptage de

¹⁸ Il s'agit de l'intuition du locuteur quant à la relation lexicale entre les deux mots, c'est-à-dire le fait qu'il est plus facile de mémoriser/apprendre l'un étant donné l'autre, *i.e.* *decision* étant donné *decide* que *decision* étant donné par exemple *jelly*.

nombre de symboles utilisés (Cf. Chomsky & Halle 1968)¹⁹, une méthode basée sur « independent information content » où seulement les informations imprévisibles rentrent dans le comptage.

Le modèle du lexique de Jackendoff, « full-entry theory » (Jackendoff 1975 : 643), comprend des entrées lexicales complètes pour tous les mots effectifs de la langue ainsi qu'un ensemble de règles de redondance lexicale : un couplage de règles de redondance morphologiques et sémantiques²⁰ qui permet simplement de formaliser le fait que les paires de mots liés comportent moins d'informations indépendantes que les mots sans rapport.

Par conséquent, la nouveauté dans le modèle du lexique de Jackendoff est le fait qu'il s'est débarrassé de la liste des morphèmes²¹ et des règles dérivationnelles, c'est-à-dire que pour lui le lexique n'est pas **produit** par une composante. Par ailleurs, Jackendoff précise que ces règles de redondance permettent également de rendre compte de la créativité lexicale, car « after a redundancy rule is learned, it can be

¹⁹ Cette conception d'économie existait également chez les structuralistes. Pour Hockett, « if several different analyses equally satisfy the other requirement, that which establishes the smallest number of phonemes is the one to be preferred. This is a corollary of the general scientific principle that the simplest description which accounts adequately and accurately for all the facts is to be preferred » (Hockett 1942 : Joos 1958 : 101), et, comme le mentionne Ford *et al.* (1997 : 23), cela remonte à Panini.

²⁰ Il peut y avoir des cas où seules les règles morphologiques jouent un rôle parce qu'il n'y a pas de régularité sémantique. Par contre, quand les règles sémantiques existent, le modèle ne tranche pas quant au fait qu'elles doivent être utilisées en couplage avec les règles morphologiques ou non.

²¹ Bien que le découpage en morphèmes soit présent avec les mêmes symboles de frontière que SPE, *i.e.* +, =

used generatively, producing a class of partially specified possible lexical entries » (Jackendoff 1975 : 668). Cependant, il ajoute que cela n'est pas leur rôle habituel²².

Bien que l'intervention de Jackendoff ne constitue pas un modèle de la morphologie, sa conception du lexique et sa façon de rendre compte des relations entre les mots a eu sûrement un impact important dans le développement des théories morphologiques.

1.2.2.3 Kiparsky (1982)

Kiparsky a proposé un modèle pour traiter les règles phonologiques « in tandem » avec les règles morphologiques. Les deux groupes de règles se trouvent dans le lexique, hiérarchiquement organisés dans les blocs appelés « *strata* ». A chaque strate est associé un ensemble de règles morphologiques liées à un ensemble de règles phonologiques qui déterminent comment l'*output* de chaque strate doit être prononcé. Les règles morphologiques sont ordonnées de manière à ce que les règles flexionnelles s'appliquent après les règles dérivationnelles et les règles de la composition, *i.e.* dans les *strata* plus élevées.

Les règles phonologiques sont divisées en deux ensembles : lexical et post-lexical. Les règles lexicales sont assignées aux *strata* et sont de nature cyclique, mais les règles post-lexicales s'appliquent seulement au moment où les mots se combinent dans les phrases et sont donc de nature non-cyclique. En morphologie lexicale, comme dans SPE, les alternances morphophonologiques font partie de la phonologie, sauf qu'il propose un mécanisme pour rendre compte de leur dépendance morphologique. Il semble que Kiparsky reprenne les règles de SPE et en propose une organisation plus sophistiquée, notamment pour mieux rendre compte

²² « (...) the normal mode for syntactic rules is creative, and normal mode for lexical rules is passive » (Jackendoff 1975 : 668). Cf. également Halle (1973 : 16)

du placement de l'accent en anglais. Le modèle de Kiparsky étant très restreint et spécifié, il souffre d'exceptions et de contre-exemples multiples.

1.2.2.4 Halle et Marantz (1993)

Halle et Marantz rejettent l'hypothèse lexicaliste au profit de l'hypothèse transformationnelle et proposent la théorie de la *Morphologie Distribuée* selon laquelle les mots et leurs sous-parties sont simplement des constituants syntaxiques plus petits que les constituants tels que la phrase ou le syntagme, et ne diffèrent pas de ces derniers de manière fondamentale. La structure interne de tout constituant – bien que chacun ait ses propres propriétés – est donnée par la même composante grammaticale qui est la syntaxe. Notamment, la structure interne des mots découle de l'opération syntaxique de mouvement de tête, X° .

Le lexique, tel qu'il est conçu par les lexicalistes, n'existe pas dans ce modèle et toutes les opérations morphologiques sont distribuées entre différentes composantes et n'appartiennent pas à un seul module. La grammaire conçue dans le cadre de cette théorie contient deux parties : la première comprend les différents ensembles d'informations listées (une liste de morphèmes, une liste de vocabulaire et une encyclopédie) ; la seconde partie est le moteur génératif qui comprend les mécanismes syntaxiques et post-syntaxiques. Ce moteur construit les expressions linguistiques en prenant les morphèmes dans la première partie et fournit leurs interprétations phonologiques et sémantiques avec les informations disponibles dans le vocabulaire et dans l'encyclopédie.

Dans ce modèle, les expressions de toute taille font partie de l'encyclopédie si leur sens n'est pas complètement prédictible à partir de leur structure. Le mot n'a aucun statut particulier et l'idiosyncrasie d'un mot comme *transmission* est prise en compte de la même manière que celui d'une phrase comme « *kick the bucket* ».

1.2.3 Morphologie basée sur le mot

Words, once formed, persist and change; they take on idiosyncrasies, with the result that they are soon no longer generable by a single algorithm of any generality. The word gravitates towards sing. (Aronoff 1976 : 18)

Au cours du développement de la linguistique moderne, les linguistes ont négligé le statut privilégié du mot au profit des unités plus petites tel que le morphème.

The 'morpheme', as identified traditionally (« smallest unit having meaning »), (...) is a by-product of structuralism and its two-faced sign, and stems from certain principles such as Martinet's double articulation or Hjelmslev's isomorphism, as well as from a taxonomic methodology based on the principles of segmentation and classification. (Ford et al. 1997 : 8)

L'on peut tracer déjà l'origine du développement du regard critique sur le morphème dans les travaux de structuralistes comme Hocket (1947). Les structuralistes post-bloomfieldiens étaient sensibles aux problèmes soulevés par le morphème et celui de la vision de la morphologie comme arrangement de ces entités dans le mot. Sauf que, plutôt que de chercher des solutions et de développer un regard critique profond sur la notion de morphème, leur effort était concentré sur l'identification et l'étiquetage des morphèmes problématiques (morphème zéro, morphème porte-manteau, morphème vide, etc. (Cf. Hocket 1947), ainsi que sur le choix du modèle le plus pertinent, *i.e.* entre IA et IP ou encore WP, à adopter pour mieux décrire un corpus donné. Cela s'explique par le fait que la téléologie structurelle était simplement de décrire une langue, *i.e.* un corpus d'*utterances*, de manière la plus adéquate et la plus claire.

Dans sa monographie consacrée à la conjugaison verbale du latin, Matthews (1972) met en cause la validité du modèle IA, bien qu'il soit peut-être parfait dans le cas des langues agglutinantes. Il met en question la relation biunivoque et le caractère linéaire et séquentiel des constructions présupposées par ce modèle, qui ne peut être entièrement respecté et qui réduit la morphologie à l'agglutination. Il critique la façon dont les cas problématiques ont été contournés pour garder à tout prix une

description IA, par exemple en proposant des morphes « replacive » (Cf. Matthews 1972 : 57-59), et montre ainsi l'incohérence de la logique d'un tel modèle. Selon lui, il y a lieu de se demander si la méfiance envers le processus morphologique justifie un traitement plus complexe :

- une formation où un segment est ajouté au milieu plutôt qu'une formation dans laquelle un segment est ajouté à la fin ? ... ou ;
- une distinction par une opposition vocalique plutôt qu'une distinction par absence ou présence d'un suffixe ?

Matthews démontre la possibilité de poser le mot²³ comme l'unité morphologique et argumente en faveur du modèle *Word-and-Paradigm*.

Premièrement, certaines particularités du mot (en latin) lui procurent un statut théorique spécial (Cf. Matthews 1972 : 96-97) :

- il est l'unité la plus petite qui peut à lui tout seul constituer une phrase, ce qui n'est pas le cas pour le morphème ;
- il ne connaît pas de construction récursive ;
- la cohésion des morphèmes à l'intérieur du mot est beaucoup plus importante que celle des mots à l'intérieur de la phrase ;
- l'arrangement des éléments à l'intérieur d'un mot est fixe et non contrastif, c'est-à-dire qu'une opposition de sens est formée par le changement d'arrangement des éléments à l'intérieur du mot.

Deuxièmement, c'est le mot qui détermine le domaine du *morphemic conditioning* (Cf. Matthews 1972 : 99) :

²³ Matthews distingue trois conceptions du mot : *word-form* au niveau phonologique, *word* au niveau grammatical et *lexème* au niveau lexical. Selon lui, c'est le lexème qui constitue l'unité de la morphologie (Matthews 1972 : 160-161).

- les éléments s'influencent les uns les autres à l'intérieur du mot : par exemple le choix d'allomorphe d'un morphème est déterminé par les morphes voisins ;
- il n'y a pas de cas où un morphème détermine la forme phonologique d'un autre morphème à travers la frontière du mot.

Il en conclut donc que c'est le mot (*i.e.* le lexème) plutôt que le morphème qui est digne du statut d'unité morphologique.

Au sein de l'école générative, différents auteurs ont argumenté contre le morphème comme unité de base de la morphologie et ont proposé des modèles de morphologie basée sur le mot. Dans le cadre de ce travail, je me limite aux travaux d'Aronoff (1976) et à ceux d'Anderson (1992), pour arriver au modèle WWM de Ford *et al.* (1997).

1.2.3.1 Aronoff (1976)

Dans la continuité du rétablissement du domaine morphologique dans le cadre de la théorie générative, après Halle (1973) et Jackendoff (1975), Aronoff (1976) propose son modèle de la formation du mot à partir d'une mise en question de la notion de morphème comme « signe minimal ». Aronoff se base sur le critère biunivoque du signe minimal (dans le sens saussurien du terme : l'unité minimale de sens, une association arbitraire du son et du sens) et présente trois ensembles de données pour démontrer que le morphème, tel qu'il est défini, viole ce critère.

Comme nous l'avons vu auparavant, avec la définition structuraliste du morphème, l'on est obligé de donner le statut du morphème à *cran-* de *cranberry*. Cependant ce morceau n'apparaît nulle part ailleurs et il n'est donc possible de lui associer un sens que de manière circulaire.

Bien que le *black-* de *blackberry* se produise ailleurs, l'on ne peut pas lui assigner un sens indépendant du mot dans lequel il se trouve. Autrement dit, *-black* n'a pas de sens constant. (Bloomfield fait mention du problème d'identifier le « même sens »

mais ne s'y s'attarde pas sous prétexte que définir le sens dépasse le domaine de la linguistique. Cf. 1.2.1). La sous-spécification du sens ne peut non plus nous fournir une solution, car il n'y a presque aucun élément de sens commun à toutes les instances de *-black* que l'on puisse isoler.

Les suffixes comme *-mit*, pour lequel les mots dans lesquels il se trouve ne partagent pas d'élément de sens commun, présentent le même comportement morphophonologique devant certains suffixes (i.e. $t \rightarrow s / _ +ion$), de telle manière que l'on est tenté de supposer qu'il s'agit d'un même morphème.

Aronoff conclut que la notion de morphème comme « unité minimale de sens » n'est pas tenable, puisqu'elle présuppose que le sens des mots est compositionnel. Or, le mot est une unité qui connaît l'idiosyncrasie : dans le temps, le sens des mots change indépendamment de leur formation de départ. Il souligne que le seul élément auquel l'on peut assigner un sens constant est le mot et c'est le mot qui satisfait le critère du signe minimal. La propriété essentielle des morceaux que l'on a traditionnellement identifiés comme morphèmes n'est pas leur sens mais le fait qu'ils sont reconnaissables.

A partir de ces observations, Aronoff propose un modèle de morphologie basé sur le mot – plus précisément sur le lexème²⁴ – en adoptant l'hypothèse lexicaliste faible. Ainsi, pour Aronoff, le domaine de la morphologie n'inclut pas la flexion, et, comme chez Matthews, l'unité de base de la morphologie est le lexème.

Aronoff propose comme Jackendoff (1975) et contrairement à Halle (1973) un lexique unique de mots. Le morphème est conçu comme une séquence sonore reconnaissable faisant partie d'une règle de formation du mot (*Word Formation Rule*, ci-après WFR) et n'a pas de statut indépendant de la règle dont il fait partie. Les WFR, comme les

²⁴ « throughout (...) this work word should be taken to mean 'word sans inflection' or *lexeme* in the sense of Matthews (1974) » Aronoff (1976 : 9).

règles de redondance de Jackendoff, s'appliquent aux mots plutôt qu'aux morphèmes, mais sont des règles génératives, à usage unique, utilisées pour produire des mots nouveaux à partir des mots existants ou pour analyser des mots momentanément oubliés.

Le modèle d'Aronoff comprend deux autres types de règles regroupées sous l'étiquette de « règles de réajustement » à la SPE (mais qui chez Aronoff, contrairement à SPE, ne sont pas prises en charge par la phonologie) : tout d'abord, les règles de troncation pour rendre compte des formations de mots tel que *nomin-ate/nomin-ee* ou *formul-ate/formul-able* où, puisque la « base » n'est pas un mot, une WFR ne peut être proposée, et, enfin, les règles d'allomorphie.

1.2.3.2 Anderson (1992)

Anderson (1988, 1992) a rassemblé un ensemble de problèmes qu'engendre la notion structuraliste du morphème et a repris notamment les remarques faites auparavant par Matthews (1974) et Aronoff (1976). Comme je l'ai mentionné précédemment, les structuralistes avaient déjà relevé les difficultés d'isoler le morphème dans les cas qui dévient du cas canonique compositionnel, *i.e.* additif, concaténatif et monotonique. Les cas problématiques (à l'exclusion de ceux déjà mentionnés par Aronoff) sont les suivants :

- l'infixe : comme dans le latin *rumpo* 'je casse' où une nasale est insérée entre la base et la flexion (base : *rup-*, Cf. 'je cassai' *rupi*) ;
- les morphèmes vides : une forme qui n'a aucun contenu, comme les voyelles thématiques dans les langues romanes ;
- les morphèmes superflus : où la forme a un contenu, mais ce contenu n'a aucun rapport au mot : l'adverbe en *-ment* en français qui a comme base l'adjectif au féminin ; la marque du féminin est considérée comme non pertinente pour l'adverbe ;

- les morphèmes porte-manteau : le cas du cumul du contenu par la même forme comme dans 'au' = 'à le' en français ;
- les morphèmes zéro : l'absence de forme tandis que le contenu est attendu comme dans les cas du pluriel anglais de type *sheep* ;
- la soustraction : dans le cas canonique exige qu'un ajout de contenu (catégorie dérivée) soit associé à l'adjonction à la forme de base, les cas où, par exemple, le pluriel, le féminin ou le passé comporte moins d'éléments de forme que le singulier, le masculin ou le présent, sont problématiques ;
- l'apophonie : où le sens est associé à la modification d'un élément phonique plutôt qu'à l'ajout d'un élément nouveau, comme dans les paires : *sing/sang, man/men, food/feed, prove/proof* en anglais ;
- de plus, le cas où l'opération de modification ne se fait pas au niveau segmental mais suprasegmental : le déplacement d'accent, la méthatèse, ainsi que les cas tels que la réduplication, où le morceau qui se double dépend de la structure phonologique de chaque mot, et donc il n'y a pas de segment phonologique en commun.

En se basant sur ces observations, Anderson affirme que la relation biunivoque entre le sens et la forme, *i.e.* le critère définitoire du morphème, n'est pas tenable.

Anderson (1992) propose un modèle de morphologie basé sur le mot, concrètement basée sur le *stem* : « word minus (productive) inflectional affixation » (Anderson 1988 : 163), comme chez Aronoff (1976) et Matthews (1972). Il reformule l'hypothèse lexicaliste comme : « The syntax neither manipulate nor has access to the internal form of words » (Anderson 1988 : 165).

Anderson argumente que certains aspects de la structure interne des mots sont pertinents pour la syntaxe, notamment les informations concernant la catégorie syntaxique. Etant donné que ces aspects sont « flexionnels » plutôt que « dérivationnels », Anderson opte pour une dichotomie dérivationnelle/flexionnelle où la morphologie flexionnelle se définit comme la composante qui s'occupe des

aspects de la structure interne du mot qui sont manipulés par la syntaxe (Cf. Anderson 1988 : 167-168).

Dans un objectif de débarrasser complètement la morphologie du morphème (d'où le nom de sa théorie : *A-morphous Morphology*) ainsi que de la vision de la morphologie comme l'affixation, les règles de formation de mots chez Anderson sont conçues comme des processus réalisationnels. Ainsi, contrairement au modèle d'Aronoff, les opérations morphologiques ne sont pas basées sur la segmentation mais sont conçues en termes de processus. De ce fait, la partie de formation du mot, qui est traitée chez Aronoff avec les règles de troncation, est traitée chez Anderson par un seul et unique type de règle : $/X(At)/_V \rightarrow /X(\əbl)/_{Adj}$ (Cf. Anderson 1992 : 187). Néanmoins, les règles d'allomorphie sont maintenues pour modifier un radical, *stem*, lorsque nécessaire : *applicable* résulte de $/X/_V \rightarrow /Xable/_{Adj}$ couplé avec la règle d'allomorphie $[ply] \rightarrow [plic] /_{-} \{ABLE ; ATE, \dots\}$ (Cf. Anderson 1992 : 188).

1.3

Whole Word Morphology

Morphology is the study of formal relationships amongst words. (Ford et al. 1997: 1)

Une fois les problèmes conceptuels causés par l'attribution d'un statut privilégié non-fondé au morphème rendu intelligible, il faut proposer un modèle débarrassé de ce dernier et qui sera basé sur le mot. Bien que les modèles proposés par Aronoff (1976) et Anderson (1992) rejette le morphème, ils ne sont pas pour autant entièrement basés sur le mot, étant donné qu'ils ne rejettent pas les unités plus petites que le mot de manière catégorique. Notamment, ils ne prennent déjà pas en compte le mot mais le *stem* ou le *lexème*, qui sont des entités abstraites, comme unité de base des opérations morphologiques.

Un modèle morphologique pleinement basé sur le mot et complètement débarrassé de toute unité plus petite que le mot est celui de la *Whole Word Morphologie* (ci-après WWM) proposé par Ford et Singh (1991)²⁵. Je présente par la suite ce modèle en détail.

²⁵ Le modèle WWM a été esquissé premièrement dans Ford et Singh (1991) et ensuite dans une présentation complète sous forme d'une monographie parue dans Ford *et al.* (1997). Cf. Singh (2006).

1.3.1 Le modèle WWM

La morphologie est l'étude de la relation formelle entre les mots ou plus précisément l'étude de la covariance systématique de forme et de sens chez les mots (Cf. Singh 1998 : 127, et également Haspelmath 2002 : 2). Dans le domaine du mot, toute alternance qui doit être décrite en faisant référence à de l'information grammaticale fait partie de la morphologie.

Les relations formelles entre deux mots d'une langue peuvent être décrites par une *Stratégie de Formation de Mots* (ci-après SFM), formulée de la façon suivante :

$$/X/_{\alpha} \leftrightarrow /X'/_{\beta}$$

où :

X et X' sont des mots ;

α et β sont des catégories lexicales majeures ;

\leftrightarrow indique une relation d'équivalence (une implication bidirectionnelle) ;

X' est une fonction sémantique d'X ;

['] indique une différence formelle entre les deux pôles de la relation morphologique ;

['] peut être nul si $\alpha \neq \beta$.

Une stratégie formalise une généralisation qui peut être faite à partir d'un fait particulier et, pour établir une SFM, un minimum de deux paires de mots est nécessaire. La théorie ne se prononce pas sur la question de savoir si deux paires de mots sont suffisantes. En fait, la question de savoir si un locuteur donné, dont le lexique comprend **deux** paires de mots donnés, est en mesure de faire la généralisation nécessaire pour en faire une stratégie ou, s'il en lui faut trois, quatre, cinq ou davantage de paires, exige une réponse empiriquement. Ce qui est indiscutable, c'est qu'il en faut au moins deux. Donc, pour faire une description

morphologique à l'aide du modèle WWM, chaque stratégie doit être appuyée par deux paires de mots.

Les SFM formalisent la relation morphologique entre les mots. Elles permettent à un locuteur de faire une analyse morphologique d'un mot qu'il n'avait pas analysé auparavant, ou de créer un mot nouveau qu'il n'a peut-être pas encore rencontré ou qu'il a oublié. L'énumération des SFM constitue donc une partie de la description d'une langue. Elles relèvent de la compétence linguistique et sont une composante de la grammaire.

1.3.2 Le mot, unité de base de la morphologie

Les opérations morphologiques ne s'appliquent qu'au mot et à aucune autre unité. Le modèle ne reconnaît aucune catégorie plus petite que le mot. Les stratégies ont le mot comme base et celui-ci comporte deux parties : une partie constante²⁶ (qui peut être nulle) et une partie variable ; et aucun statut ontologique n'est donné à ces deux parties. Le mot n'est pas donné et doit être défini par la théorie. Les propriétés que le modèle exige comme des critères pour reconnaître une entité linguistique comme mot sont :

- une structure phonologique ;
- une catégorie syntaxique ;
- un emploi sémantique.

²⁶ L'on peut identifier la définition du morphème comme « active part in word formation processes » (le point de vue défendu par Kurylowicz 1949) à la partie constante d'une stratégie morphologique qui n'a donc aucun statut en dehors de la stratégie en question.

Toutefois, il faut noter que le mot reste une entité difficile à saisir. Il n'y a pas toujours consensus sur les unités que l'on peut identifier comme mots et ce choix influence complètement l'analyse que l'on peut en faire.

1.3.3 Morphophonologie

Certaines relations formelles entre les mots sont l'objet de la phonologie qui s'occupe exclusivement et sans exception d'alternances automatiques et globales gouvernées par la phonotactique. L'aspect de la relation formelle entre les mots qui relèvent de la phonologie doit être éliminé de la morphologie. Les alternances, que l'on a qualifiées de morphophonologiques, sont toujours cadrées avec les affixes qui les déclenchent et ne peuvent pas évoluer de manière indépendante. De plus, ces alternances ne causent pas d'interférences dans l'acquisition d'une langue seconde, comme c'est le cas des alternances phonologiques, mais font partie des opérations morphologiques (Cf. Ford *et al.* 1997). Ces auteurs rejettent donc l'existence de la morphophonologie : les alternances automatiques et globales relèvent du domaine de la phonologie et les alternances non automatiques du domaine de la morphologie.

1.3.4 Dérivation *vs* flexion

La théorie WWM n'admet aucune morphologie multiple. Donc, il n'y a pas de morphologie dérivationnelle *vs* flexionnelle, qui est une distinction sous-entendue dans les modèles relevant du lexicalisme faible.

Par la suite, je présente les critères, extraits de Stump (1998 : 13-18), avancés pour distinguer entre la flexion et la dérivation, suivis des exceptions pour chaque critère :

- La dérivation, contrairement à la flexion, change la catégorie syntaxique de la base :

D'un côté, il y a des exemples considérés comme dérivationnels qui ne changent pas pour autant la catégorie syntaxique de la base : faire_v/refaire_v ;

jaune_{Adj}/jaunasse_{Adj} ; maison_N/maisonnette_N. De l'autre côté, il existe un cas comme le participe passé qui est considéré comme partie intégrante du paradigme flexionnel, mais qui a un caractère adjectival dans beaucoup de langues.

- La flexion est déterminée par la syntaxe :

Stump ajoute que « but of course not all inflectional morphology is directly relevant to syntax » et conclut que l'utilité d'un tel critère dépend de la précision du principe utilisé pour distinguer entre les propriétés lexicosémantiques et les propriétés morphosyntaxiques.

- La flexion est plus productive et régulière :

L'adverbe en *-ment* en français ou *-ly* en anglais sont « productifs », mais étant donné qu'ils changent la catégorie de la base ils sont considérés comme dérivationnels. De plus, l'existence de paradigmes flexionnels défectueux, les verbes défectueux par exemple pour lesquels certains cas du paradigme flexionnel sont vides (*i.e.* en français : *frir_v*), démontre que la flexion n'est pas toujours productive.

- La flexion empêche la dérivation, mais pas l'inverse :

En français, l'adverbe, qui est considéré une catégorie dérivationnelle, est formé sur la base de la catégorie flexionnelle du féminin : *lour-de-ment*.

Somme toute, il existe des contre-exemples pour presque tous les critères que les tenants de la dichotomie flexion/dérivation proposent, c'est pourquoi la théorie WWM refuse d'accepter une telle distinction.

1.4

Clarifications sur le pluriel en persan

Avant de procéder à l'énumération des SFM, il nous reste à préciser ce que le terme « pluriel nominal » représente ici. Autrement dit, il faudrait préciser la raison d'être de ce regroupement de stratégies fait sous l'étiquette « pluriel ».

Comme tous les autres éléments linguistiques, le « pluriel » n'est pas une étiquette pourvue d'une valeur, dans le sens saussurien du terme, universelle. Ce que la catégorie « pluriel » signifie pour le français n'est pas la même chose que pour l'anglais, ni pour le persan, etc. Il faudrait donc déterminer quelle est la valeur de pluriel à l'intérieur de la langue persane et en opposition avec d'autres éléments dans cette langue.

De plus, Corbett (2000) fait remarquer que, bien que la catégorie de nombre semble être universelle, les valeurs numériques varient à l'intérieur de chaque système. Il existe différents systèmes numériques et les catégories, comme singulier, pluriel, etc., ne sont pas universelles et prennent leur valeur à l'intérieur de chaque système numéral.

En ce qui concerne le persan, la catégorie connue comme « singulier » porte un sens générique. « *ketab* » veut dire livre en général, un sens proche de *books* de l'anglais dans une phrase comme « reading books is a good thing », *ketab xandan kar-e xubi-e*, ou « le livre » du français dans une phrase comme « le salon du livre aura lieu cette

semaine », *namajefgah-e ketab in hafte bargozar mifavad*; i.e. il est non-référentiel dans ces contextes²⁷. Pour obtenir une interprétation référentielle, dans le cas de l'indéfini, *je(k) ketab ou ketabi* et, dans le cas du défini, *ketabe* sont employés. L'opposition *un livre/des livres* du français ou *a book/books* de l'anglais se fait en persan simplement par une opposition au niveau du quantifieur *je(k)/tfand-ta ketab*.

« *ketab(h)a* », la forme qui appartient à la catégorie connue comme « pluriel » est employée pour parler de plusieurs livres précis, cette forme engendre forcément une interprétation référentielle et définie. En effet, le persan connaît un système numéral « general/singular *versus* plural » (Cf. Corbett 2000 : 13-16) que les exemples ci-dessous permettent d'illustrer :

(1) xeyli vaqt bud mixastam ketab bexaram emruz je ketab/tfand-ta ketab
xardiam vali ketabe/ketaba dʒa mund/(an) tu maqaze

ça faisait longtemps que je voulais acheter des livres (livre)²⁸ aujourd'hui j'ai acheté un livre/quelques livres mais le livre/les livres a/ont été (est/sont) oublié/s dans le magasin

(2) ketabi/ketabai ke xaride budam tu maqaze dʒa mund/(an)

le livre/les livres que j'avais acheté/s a/ont été (est/sont) oublié/s dans le magasin

N.B. : L'emploi non-référentiel du pluriel n'est pas grammatical²⁹ :

(3) namajefgah-e ketab/*ketaba hafte-e digast

²⁷ Il existe également des cas où le contexte nécessite une interprétation référentielle, et donc « *ketab* » renvoie à un livre précis : *ej vaj ketab xis fod*: « Ah ! le livre a été mouillé ».

²⁸ J'ai ajouté la traduction exacte des termes ou une précision entre parenthèses.

²⁹ Cependant, il existe des contextes où le pluriel non-référentiel est utilisé comme intensifiant : *u ketabha xunde* 'lui, il a lu beaucoup de livres'. Toutefois, il s'agit d'un emploi marginal et plutôt vieilli.

le salon du livre aura lieu (est) la semaine prochaine

- (4) **foruṣ-e ketab/ketaba daramad-e xubi daṣt**

*la vente du livre/des livres (et l'on sait de quels livres il s'agit) donnait/a donné (avait)
un bon revenu*

- (5) **xarid-o-foruṣ-e ketab/*ketaba daramad-e xubi dare**

le commerce de livres rapporte bien

Par ailleurs, il existe, ou plus correctement il y a des locuteurs dont le lexique contient, un autre mot (formellement) lié au mot *ketab* qui porte également un sens de pluralité : il s'agit du mot *kotob*, traditionnellement appelé « pluriel brisé³⁰ (d'origine) arabe », par exemple dans un énoncé comme :

- (6) **dar kotob-e xat:i in kalme be surat-e digari neveṣte ṣode ast**

dans les (livres) manuscrits ce mot est orthographié différemment

Cependant, *kotob* ne peut pas remplacer *ketab(h)a* partout :

- (7) **diruz tḡand-ta ketab xaridam *kotob-i/ketab(h)a-i ke xaridam tu-je taksi
ḡa mund(an)**

j'ai acheté quelques livres hier, les livres que j'ai achetés ont été (sont) oubliés dans le taxi

La phrase de l'exemple (7) n'est pas une phrase bien formée du persan. Cependant, le fait que son inacceptabilité vient d'un problème grammatical ou est liée au mélange du registre (formel et informel) reste à vérifier.

³⁰ Il faut noter que l'étiquette « brisé », qui est une autre manière de dire « non-concaténatif », n'a aucun statut ontologique dans la théorie WWM. Je continuerai cependant à utiliser ce terme pour référer à ce type de pluriel par simple commodité.

Le « pluriel brisé » que l'on peut qualifier de [+formel] est une forme qui n'est pas accessible à tous les locuteurs et qui est apprise tardivement dans le cours de l'acquisition, normalement par le biais de l'enseignement scolaire à partir du niveau secondaire. Il faut noter qu'une forme « brisée » peut très bien faire partie du lexique d'un locuteur mais être sous-analysée par celui-ci (Cf. Tiersma 1982, également discuté dans la section 2.4), c'est-à-dire ne pas présenter pour lui une relation formelle avec un nom singulier. En outre, beaucoup de mots analysables comme « pluriel brisé » ont obtenu des sens différents, parfois en plus du sens du mot singulier correspondant.

Parallèlement, il existe d'autres oppositions formelles qui présentent une interprétation de pluralité. Je présente ces oppositions à l'aide des exemples ci-dessous :

- (8) man nazar-am-o tu baxf-e nazarat neveftam : nazar_N 'commentaire'
j'ai écrit mon commentaire dans la section des commentaires
- (9) in mozaxrafat tji-e migi ? : mozaxraf_{N,Adj} 'absurde', 'absurdité'
c'est quoi ces absurdités que tu dis ?
- (10) sefare[at-etun ro ba mo[axasat-e kamel bar-am imejl konin : sefaref_N
 'commande', mojaxas_{Adj} 'précis' (*mojaxas_N)
envoyez-moi par mail vos commandes avec toutes les informations
- (11) inɟa masraf-e ma[rubat-e alkoli ɟorm-e : ma[rub_N 'boisson (alcoolisée)'
ici consommer les boissons alcoolisées est interdit
- (12) alkolɟat tji tu basat-et hast ? : alkol_N 'alcool'
qu'est-ce que t'as comme boissons alcooliques ?
- (13) ahl-e ma[rubɟat hasti ? : ma[rub_N 'boisson (alcoolisée)'
est-ce que tu bois/t'aimes boire ?

anva?-e sandevitʃɒʒat o pitzadʒat be xord-emun dad : sandevitʃ_N ‘sandewich’
pitzɑ_N ‘pizza’

il³¹ nous a fait manger toutes sortes de sandwiches et de pizza

(14) xaste ʃodam az in tʃarndiat : tʃarnd_{Adj,N} ‘absurde’, ‘absurdité’

je suis fatigué de ces absurdités

(15) mazhabiun emruz ruze gerefteand : mazhabi_{Adj,N} ‘religieux’, ‘personne religieuse’

les religieux jeûnent aujourd’hui

(16) mosaferin-e mohtaram xof amadin : mosafer_N ‘voyageur’

chers voyageurs bienvenue

(17) rahbaran-e asli-e moxalefan dastgir ʃodand : rahbar_N ‘guide, tête’,
moxalef_{N,Adj} ‘opposant’

les principaux guides (leaders) des opposants ont été arrêtés

Les formes en *-in* et *-an* (Cf. les exemples (16) et (17)) diffèrent de leur contrepartie en *-(h)a* en ce qu’elles sont simplement caractérisées comme formelles et on les entend rarement dans les discours spontanés et informels. Je reviendrai sur ces formes dans la partie 2 (Cf. 2.3.1).

Les formes en *-ɒʒat* et *-iat* diffèrent de leur contrepartie en *-(h)a* en ce qu’elles ont une interprétation générique ou « collective » (Cf. Corbett 2000 : 117-119), signifiant

³¹ L’opposition du genre en persan ne se fait ni au niveau du nom, ni du pronom, ni de l’adjectif. J’utilise le masculin par commodité.

l'appartenance à un groupe, et ne permettent pas une interprétation référentielle. Je reviendrai sur ces formes également dans la partie 2 (Cf. 2.3.2).

Les formes en *-un* se différencient également en ce qu'elles désignent l'appartenance à un groupe. *-un* partage certaines particularités sémantiques avec *-in* et *-an* sur lesquelles je reviendrai dans la partie 2 (Cf. 2.3.1).

Quant aux formes en *-at* et les « pluriels brisés », comprendre leur comportement demande un examen plus détaillé. Pour ce faire, je voudrais en dire plus sur le trait [+défini] mentionné ci-dessus.

Dans les ouvrages qui traitent du pluriel et de ses particularités sémantiques ou syntaxiques, *i.e.* ceux qui considèrent la définitude comme un trait du pluriel (Cf., entre autres, Ghomeshi 2003), bien que le corps d'argumentations, les exemples et les contre-exemples portent exclusivement sur la forme *-(h)a*, il est mentionné d'une manière ou d'une autre que les conclusions et les arguments avancés sont également valables pour les autres « marques du pluriel ». Néanmoins, je trouve que cette vision est simpliste.

Pour clarifier les problèmes posés par cette vision du pluriel, je voudrais présenter quelques exemples. Le plus frappant est un morceau de discours que j'ai entendu à la télévision. Le sujet de conversation est *mavane?-e demokrasi dar iran* 'les obstacles (contre) la démocratie en Iran' et, après que l'un des interlocuteurs eut présenté sa liste de ces obstacles, *mavane?*, un autre pose une question dans laquelle l'on entend les énoncés suivants :

(18) mavane?-e ziadi hast... albate in mavane?ha-i ke joma efare kardin... vali

mane?-e asli... jeki az mane?ha-i ke voçjud dare...

*il existe beaucoup d'obstacles... cependant les obstacles que vous avez mentionnés...
mais l'obstacle principal... l'un des obstacles qui existe...*

Nous remarquons quatre formes *māneʔ*, *māneʔha*, *mavāneʔ* et *mavāneʔha*, et notons que trois formes s'opposent au singulier *māneʔ*. Bien que *mavāneʔ* porte le sens de pluralité, là où une référence précise est attendue, la forme *mavāneʔha* est utilisée, et il semble que *mavāneʔ* ne satisfait pas le critère [+défini] pour ce locuteur.

Le deuxième exemple est un énoncé que j'ai trouvé dans la partie commentaires d'un blog sur l'Internet. Le scripteur voulait télécharger une chanson, mais, étant donné les commentaires, il s'est rendu compte que la chanson était trop lourde :

(19) ba in nazarata-i ke xundam paʃimun ʃodam

étant donné les commentaires que j'ai lus je laisse tomber

nazarat est le pluriel de *nazar* 'commentaire'. Ce mot est utilisé dans le jargon 'Internet' pour désigner la fenêtre destinée à écrire des commentaires. On constate qu'ici, comme dans le premier exemple (20), le locuteur emploie le (double) pluriel – *(h)a* quand le trait [+défini] est exigé. Ainsi, *nazarat* est un pluriel (en opposition à *nazar*) mais avec un usage générique plutôt qu'avec un usage référentiel.

Le troisième exemple est un énoncé que j'ai trouvé dans la partie commentaires d'un forum sur le *body-building*. Quelqu'un y a posé une question :

(20) ... mixastam bebinam tʃe harekati hast... ?

... je voulais savoir quels mouvements exist(ent)... ?

En réaction à une des réponses, la même personne écrit :

(21) dorost-e in harekatha ham taʔsir darand...

c'est vrai ces mouvements ont également un effet...

harekat est le pluriel de *harekat* 'mouvement'. Ici aussi, comme dans les deux exemples précédents, il semble que le locuteur emploie le pluriel (double) –*(h)a* quand le trait [+défini] est exigé.

Si l'on accepte l'hypothèse selon laquelle la définitude est un trait du pluriel nominal, qui s'applique indépendamment de la marque, comment expliquer ces cas de pluriel double en $-(h)a$? Et les exemples de ce type ne sont aucunement exceptionnels ; le fait que les grammaires prescriptives sanctionnent les pluriels doubles laisse supposer qu'ils se produisent assez fréquemment pour attirer l'attention.

Néanmoins, cette observation ne suffit pas à conclure que $-ha$ est nécessaire et suffisant pour le trait [+défini]. Les pluriels, autres que $-(h)a$, **peuvent** être employés dans des cas où un trait [+défini] est nécessaire. La seule affirmation que l'on peut faire sans risque est que $-ha$ implique forcément la définitude. Il est intéressant de noter que Samvelian (2007) considère le « plural suffix $-ha$ » et le « definite suffix $-(h)e$ » (mentionné plus haut) comme « word-level inflexional affixes » (Samvelian 2007 : 618) en **distribution complémentaire**.

Par ailleurs, l'opposition créée par $-h(a)$ est acquise beaucoup plus tôt dans le cours d'apprentissage de la langue ; les enfants en bas âge reconnaissent déjà cette opposition, tandis que les autres viennent beaucoup plus tard, et certaines oppositions (par exemple parmi les « pluriels brisés ») ne sont jamais apprises.

Il faut noter que la réalisation des pluriels doubles est particulièrement significative dans le cas du « pluriel brisé », mais existe également à un degré nettement moindre chez les formes en $-at$ (et les grammaires prescriptives sanctionnant ces emplois mentionnent des pluriels doubles de ces deux types). Toutefois, je ne suis pas en mesure d'affirmer si cette divergence concerne exclusivement ces deux cas. Etablir les contextes et les critères qui expliquent l'occurrence des pluriels doubles nécessite une recherche empirique poussée qui, à ma connaissance, n'a pas été faite. Dans l'état actuel de cette recherche, toute autre conclusion plus restreinte serait de l'ordre de la spéculation. Le seul élément que je peux avancer est que l'iconicité (Cf.

Dressler *et al.* 1987) y joue un rôle. Je reprendrai la question des pluriels doubles dans la partie 2 (Cf. 2.4).

Rentrer davantage dans le débat concernant le système numéral du persan et le comportement syntaxique des pluriels dépasse les limites de ce travail dont l'intérêt principal est la morphologie. Je m'en tiendrai dans cette étude à regrouper les SFM selon leurs règles d'interprétation (RI) au sein de la catégorie syntaxique du Nom. Etant donné les observations au sujet du statut particulier de *-(h)a*, j'opterai pour réserver l'étiquette *N.pl* à cette forme³² et aux formes en *-an* et *-in* qui de plus appartient à la langue formelle : *karbaran*_{N.pl. formel}. Par ailleurs, j'emploierai l'étiquette *N.collectif*.

³² N. B. : *-(h)a* en persan n'est pas comparable à *-z* en anglais. *-(h)a* s'utilise globalement et sans aucun conditionnement morphologique, *i.e.* en concurrence morphologique avec aucune autre stratégie, contrairement à *-z* de l'anglais qui est une stratégie parmi/en concurrence avec d'autres, mais qui, étant la moins spécifiée, peut être qualifiée de pluriel *par excellence*.

1.5

A propos du lexique

La vision de la théorie WWM en ce qui concerne le lexique diffère radicalement des théories qui supposent une sorte de dictionnaire où l'ensemble des mots d'une langue donnée, accompagnés des informations grammaticales différentes, sont stockés. Selon la théorie WWM, on ne peut pas parler **du** lexique d'une langue mais **des** lexiques (mentaux) des locuteurs de cette langue. Pour chaque individu, le lexique mental comporte « in their full glory³³ » (Singh et Agnihotri 1997 : 21) tous les mots que l'on peut légitimement supposer que cet individu connaît, qui est donc une fonction des éléments biographiques de cet individu et sujet aux changements.

Pour un locuteur donné dont le dialecte se décrit, la validité d'une stratégie dépend de l'existence des mots à partir desquels la stratégie en question est établie. Lorsqu'un locuteur ne connaît ni les mots à partir desquels la stratégie est établie dans la description, ni d'autres mots correspondant au même schéma, l'on ne peut pas dire que la stratégie en question fasse partie de sa morphologie.

Comme mentionné auparavant, ce travail se donne comme tâche l'énumération des stratégies de formation du pluriel nominal, à partir de mon lexique du persan. Je

³³ Selon la théorie WWM, les mots « complexes » sont mémorisés au même titre que les mots simples et ne contiennent aucune autre structure que la structure phonologique. Autrement dit, l'on n'a pas accès aux mots « complexes » à l'aide d'une règle quelconque et les mots entre eux sont en relation de manière directe et sans aucun intermédiaire. Cette relation est formalisée par les SFM. Toutefois, les SFM du modèle WWM ne sont pas simplement un outil d'aide à la mémorisation, mais aussi un moyen de génération de nouveaux mots.

n'aborderai pas dans ce mémoire la distinction entre les deux variétés formelle et informelle³⁴ puisque je m'intéresse avant tout au lexique interne d'un locuteur donné que je prends en compte toutes sortes de paires de mots, ici les noms, qui représentent une opposition formelle couplée avec une interprétation de pluralité.

La description proposée ici peut être donc généralisée à toute personne dont le lexique contient ces mots ou d'autres mots correspondant au même schéma.

³⁴ Le persan connaît deux variétés : formelle et informelle. Ces termes ne sont que des étiquettes approximatives pour désigner les extrémités d'un *continuum* et pour lesquelles aucune définition délimitative n'est disponible. Le persan formel correspond à la langue écrite, c'est-à-dire à la langue de la littérature, et se différencie de la langue familière à tous les niveaux : phonologique, syntaxique et lexical, et toute personne familière avec cette langue admet cette divergence. Les règles du bon usage du persan formel s'enseignent à l'école, tandis que le persan familier s'acquiert naturellement. En effet, le persan connaît la situation que Ferguson (1959) a qualifiée de diglossie. Mais, il faut noter qu'une telle situation existe dans une certaine mesure dans toutes les langues, au moins toutes les langues dotées de l'écrit, et que la diglossie est une question de degré et non une caractéristique dichotomique. Autrement dit, qualifier une langue de diglossique n'implique pas grand-chose. Bien que la variété informelle corresponde à l'*I-language*, la part de la compétence internalisée de la variété formelle varie d'un individu à l'autre du fait de chaque biographie particulière.

1.6

Aperçu des travaux existants sur le persan

Il n'y a pas de tradition grammaticale ancienne pour le persan. Les Iraniens ont participé au développement de la tradition grammaticale de l'arabe³⁵, mais n'ont vraisemblablement pas trouvé le persan digne du même traitement, bien que la production littéraire et poétique en persan était considérable. Il a existé cependant, de manière sporadique, des références aux éléments de la grammaire du persan, écrites pour la plupart en arabe, dont la plus ancienne date du XIII^e siècle³⁶. Depuis le XVII^e siècle, certaines allusions à la morphologie du persan étaient faites dans les introductions aux dictionnaires du persan publiés en Inde.

1.6.1 L'apparition de la grammaire du persan

Ce n'est qu'à partir de la fin du XIX^e siècle que le développement de l'étude grammaticale sur le persan a démarré et cela à travers l'élaboration de manuels d'apprentissage et de dictionnaires bilingues, notamment en Turquie et en Inde, par

³⁵ Pendant l'époque où l'Islam dominait en Iran, les linguistes iraniens étudiaient la langue arabe, à l'époque la langue de la religion, du savoir, de l'éducation, de la politique et du prestige social. Le plus connu est Sibvayh (VIII^e siècle), dont la grammaire de l'arabe est devenue une référence. (Cf., entre autres, Meshkatod Dini 1992 : 261)

³⁶ Voir l'œuvre de Sham Qeis Râzi (XIII^e siècle) qui porte sur la poésie et la métrique du persan. Ensuite viennent l'œuvre perdue d'Abu Hayân Nahi (XIV^e siècle) et celle d'Ibn-e Mohanna (XIV^e siècle) sur la grammaire de l'arabe, du turc, du persan et du mongol.

des savants qui étaient familiers avec les traditions grammaticales occidentales et qui avaient une certaine maîtrise des langues telle que l'arabe, le turc et le français³⁷.

Les premières grammaires du persan ayant introduit les parties de discours ont paru dans la première moitié du XX^e siècle³⁷. Ensuite, un certain nombre d'ouvrages grammaticaux descriptifs détaillés sont publiés dans les années 1950-70 : Mohammad Mo'in a publié depuis 1952 (jusqu'en 1960) une série de monographies sur le nom et a fourni une multitude d'exemples s'étendant de la littérature aussi ancienne que celle du XIII^e siècle jusqu'à la langue courante ; P. Natel Khanlari (1972) a publié une grammaire du persan destinée à l'enseignement qui est la première à faire une analyse de la phrase au moins partiellement en fonction du cadre du constituant immédiat (Cf. Meshkatod Dini 1992 : 266).

La discipline de la linguistique moderne s'est développée dans les années 1970 surtout par l'intermédiaire d'une élite qui a étudié en Europe (notamment en France, en Angleterre ou en Allemagne) ou aux Etats-Unis et qui a été nourrie des théories

³⁷ En 1890, Mirza Habib Isfahani, qui enseigné le persan à Istanbul et qui maîtrisait la langue française, a publié la première grammaire (iranienne) du persan en introduisant dix parties de discours. Tandis que les travaux antérieurs (*i.e.* Abdul-Karim Irvani 1846, Mohammad-Karim Kermani 1859, Hosein-Mas'ud Ansari 1872, etc.) consistaient essentiellement en des copies des règles grammaticales à partir des grammaires d'autres langues à la portée des auteurs. Par ailleurs, c'est lui qui a introduit le mot *dastur* 'l'ordre' comme équivalent du mot grammaire. Abdul-Azim Qarib (1899) a continué l'œuvre de Mirza Habib et a publié une grammaire plus détaillée, destinée à l'enseignement, qui est à la base du manuel de la grammaire de persan *pandz ostad* 'les cinq maîtres' (année de publication inconnue), auquel ont contribué quatre autres chercheurs : J. Homai, B. Foruzânfâr, M. Bahâr, Q. Rashid-Yasemi.

linguistiques occidentales comme le fonctionnalisme européen³⁸, le structuralisme américain et finalement la grammaire générative.

1.6.2 Appréciation générale

L'étude plus profonde de ces travaux laisse l'impression que, dans beaucoup de cas, l'impact des théories contemporaines de la linguistique se limite à l'adaptation et à l'application d'une formalisation, sans réelle intégration de la théorie elle-même. De plus, on remarque qu'il existe une sorte de mélange et ce qui est fait en guise de grammaire descriptive peut être qualifié de description de la norme. En outre, beaucoup d'auteurs bataillent contre la différence entre la compétence linguistique des locuteurs et la langue standardisée, *i.e.* la variété formelle. Même les travaux qui semblent prendre la langue courante comme objet d'étude ne l'étudient pas en tant que langue interne.

Par ailleurs, bien que toute personne familière avec la langue reconnaisse la divergence entre les deux variétés, formelle et informelle, une distinction claire entre les deux n'est jamais faite dans les travaux linguistiques. Les travaux consacrés essentiellement à la variété formelle mentionnent de manière sporadique des faits de la variété informelle, et les travaux dont l'objet est la langue courante font l'inverse. L'objet d'étude n'est de ce fait jamais défini de manière exclusive et les descriptions/analyses sont à cheval entre les deux variétés.

³⁸ Le premier chercheur à publier une étude grammaticale sur le persan dans le cadre d'une théorie linguistique moderne est M. R. Bateni (*towsif-e saxteman-e dasturi-e zaban-e farsi* 'la description de la structure de grammaire du persan', 1969) qui a étudié en Angleterre et dans le cadre de la grammaire fonctionnelle d'Halliday. Ensuite, A. A. Sadeqi et Q. Arjang (1976) ont adopté le cadre de la théorie fonctionnelle de Martinet (Meshkatod Dini 1992 : 266).

1.6.3 Sur la morphologie

En ce qui concerne la morphologie, les différentes analyses qui se trouvent dans la littérature relèvent toutes d'une approche atomistique, c'est-à-dire qu'elles font appel, d'une manière ou d'autre, aux notions telles que : morphèmes, allomorphes, affixes, dérivation *vs* flexion, racines, radicaux, etc.

La description morphologique y est essentiellement consacrée à l'énumération des affixes et de leurs variantes : Majidi (1986), paru en allemand, l'étude la plus élaborée que j'ai rencontrée, qui présente en deux volumes une description structuraliste de la phonologie et de la morphologie du persan ; Kalbasi (2008 [1993]) qui n'adapte pas un cadre théorique explicite mais qui relève plutôt du structuralisme américain ; Meshkatod Dini (2008[2005]) dans un cadre génératif transformationnel.

Je considère que ces études ne reflètent pas la réalité de la langue. Leur défaut commun est de reposer sur une approche atomistique et de ne porter que sur l'*E-language*.

1.6.4 Dichotomie arabe/persan

Les grammaires normatives sanctionnent l'emploi des marques du pluriel qualifiées d'arabe avec les mots (d'origine) persans (non arabes) et cette sanction ne se limite pas à la marque du pluriel. A titre d'exemple, je cite deux « erreurs », selon l'article *Les erreurs orthographiques courantes et la grammaire du persan* (Nafisi 1965) publié sur le site du Conseil de la langue et de la littérature persane³⁹ (10/10/2007) : « *ostad* : ce mot est persan et il faut dire son pluriel *ostadan* et non pas *asatid* », et, plus loin,

³⁹ www.persian-language.org, [*oraj gostare*] *zaban va adabiat fârsi*

azmajefat est qualifié de « faux » puisque « *azmajef* est un mot persan » et qu'« il faut le pluraliser en /*ha*/persan ».

Najafi (1992), dans le livre intitulé *N'écrivons pas faux (qalat nanevisim)*, regroupe toutes les fautes liées à l'emploi des formes dites arabes avec les mots persans (non arabes). Ces sanctions sévères ont stigmatisé les emplois de ce type et qualifié les locuteurs qui les emploient d'« incultes » (en persan : *bisavad*, littéralement : illettré/analphabète, et par extension : inculte).

Les mots persans pluralisés en *-at* « contrairement aux règles » sont attestés dans la littérature dès les débuts du persan moderne⁴⁰, entre autres : *baqat*, *dehat*, *bolukat*, *jejlaqat*, *qeslaqat*, *kuhestanat*⁴¹, *gomrokat*, *filat*, *nemudarat*, *farmajefat*, etc. (Cf. Mo'in 1952, 1961) Par ailleurs, à un moment donné dans l'histoire de la langue persane, les pluriels « brisés » proliféraient (Farshidvard 2003[1969]), les mots suivants ont tous une origine persane *faramin* : *farman* ; *daravif* : *darvif* ; *banader* : *bandar* ; *ronud* : *rend* (n'est plus utilisé) ; *majadin* : *mejdan*⁴².

⁴⁰ La langue persane connaît trois périodes quant à son évolution : le vieux persan, le moyen persan et le persan moderne. La période qualifiée de persan moderne débute par le persan comme la langue officielle de la cour Samanide (875-999 après J.-C.), la dynastie iranienne qui prit le pouvoir après la conquête arabe (632 après J.-C.). La rupture du persan moderne avec le persan moyen correspond donc à la période de l'influence massive de la langue arabe.

⁴¹ Khanlari, dans sa revue de Mo'in (1952), avance qu'en persan le pluriel en *-at* a souvent le sens du collectif, précisément dans les noms de lieu, et que cela est dû au fait que dans les langues iraniennes du Nord-Est - d'où le persan moderne (le dari) est originaire, *-t* ou *-at* a été la marque du pluriel et *-t* apportait surtout un sens du collectif (Khanlari 1952 : 72-73).

⁴² Il faut noter que la présence de certaines de ces formes en persan est due au fait que l'arabe a emprunté également certains mots du persan, et qu'une fois arabisés ils ont été réempruntés par le persan sous leur nouvelle forme du pluriel. C'est le cas, entre autres, de

Les grammaires normatives et philologiques mises à part, il est également coutume dans les travaux linguistiques/descriptifs de faire une séparation entre les marques propres au persan et les marques venues de l'arabe. Par la suite, à titre d'exemples, je présente les remarques de ce type dans les ouvrages relativement récents traitant de manière directe ou indirecte de la morphologie du pluriel.

Dans son livre sur la morphologie du persan, Kalbasi (2008[1993]) fait explicitement la distinction. Dans un paragraphe intitulé « Les suffixes flexionnels du persan »⁴³, on trouve « *-ha*, *-an* et ses variétés (les marques du pluriel qui s'ajoutent au nom) comme *deraxtan*, *mardan*, *salijan*, *raftegan* » (Kalbasi 2008 : 33). Les marques « arabes » sont quant à elles listées dans le paragraphe « Les suffixes flexionnels d'arabe en persan »⁴⁴. L'auteure souligne qu'ils « s'emploient non seulement avec les mots arabes mais aussi occasionnellement avec les mots persans ou non persans » (Kalbasi 2008 : 34). Cinq suffixes sont énumérés, dont trois sont des marques du pluriel, *i.e.* : *-at*, *-in*, *-un*, ainsi que *-ejn* comme marque du duel.

Bien que Meshkatod Dini (2008[2005]) ne fasse pas de séparation persan/arabe aussi explicite que Kalbasi (2008[1993]), il regroupe dans la rubrique « Marque de pluriel » *-hâ* et *-an* et énumère *-dʒat*, *-in*, *-un* et *-at* dans le sous-paragraphe « Les autres marques du pluriel », en les qualifiant de marques du pluriel s'utilisant « avec les mots d'emprunts arabes et certains mots persans ». Le « pluriel brisé » vient à la fin et est qualifié comme « un autre type de marque du pluriel » lié aux emprunts arabes (Meshkatod Dini 2008 : 143).

baqat, *ruznamedʒat*, *anbedʒat*, en arabe le pluriel de *ruznama**dʒ*, *anbadʒ* arabisé de *ruznama*, *anbe*, etc. (Cf., entre autres, Mo'in 1961, Sadeghi 1969).

⁴³ *vandha-je tasrifi-e farsi*.

⁴⁴ *vandha-je tasrifi-e arabi dar farsi*.

Sadeghi (1969) est un des rares travaux à rejeter la distinction persan/arabe. Pour lui, bien que « certaines marques du pluriel ou (certains) pluriels soient à l'origine empruntées de l'arabe, [...] elles ont toutes été soumises aux règles de la grammaire persane » (Sadeghi 1969 : 778).

Comme Sadeghi (1969), nous pensons que faire une telle distinction sur le plan synchronique n'est pas pertinente. Un locuteur naïf n'a pas accès à l'étymologie des mots qu'il utilise. Bien qu'une partie importante du lexique d'un locuteur du persan ait une origine arabe, ces mots sont adaptés au persan. Tout d'abord, ils sont régis par la phonologie du persan et, sur ce plan, ne sont pas reconnaissables⁴⁵. De plus, beaucoup de ces mots ont connu une évolution sémantique et n'ont plus le même usage qu'à l'origine : les paires de mots qu'on peut qualifier de « faux amis » sont innombrables. Il faut également noter que, dans de nombreux cas, l'emprunt adapté en persan correspond à une variante contextuelle et non à la forme canonique. Il existe aussi des cas où deux mots empruntés correspondent à deux variantes contextuelles d'un même mot⁴⁶ (Cf. Perry 1991 et 2005). Autrement dit, à part une

⁴⁵ Bien évidemment, les personnes ayant reçu une bonne éducation littéraire reconnaissent des indices pour repérer les mots d'origine arabe ; cependant une telle reconnaissance n'est pas toujours absolument décisive.

⁴⁶ "The grammatically feminine marker in Arabic is realized phonetically as either /-at/ (in pre-juncture position) or /-a/ (pausal form), according to contextual syntax of Arabic, but written with a single hybrid graph (the *ta' marbuta*). This syntactically determined variation in spoken Arabic was irrelevant to Persian, which is uninflected for grammatical gender. (...) of some 1450 such loans in current Persian, (do) 640 end in *-at* and 810 in *-ah* [la transcription de Perry correspond à l'écrit, à l'oral il s'agit de *-e*, ultérieurement *-a*], including 50 pairs of doublets lexicalized with both endings" (Perry 2005 : 103). A titre d'exemple, *hareka(t)* a donné *harekat* 'mouvement', mais *ḍomla(t)* a donné *ḍomle* 'phrase', et *resala(t)* a donné *resalat* 'mission' ainsi que *resale* 'dissertation, thèse'.

origine lointaine, beaucoup de ces mots n'ont rien d'arabe et même les locuteurs de cette langue auraient du mal à les reconnaître.

En plus de ne pas être pertinent dans une étude synchronique, ces analyses relèvent d'une hypothèse selon laquelle les unités plus petites que le mot sont empruntables d'une langue à l'autre. Nous rejetons cette hypothèse (pour plus de discussion, Cf. Singh 2001a, 2001b, 1998, 1996), et les données du persan nous permettent de le démontrer.

L'observation des exemples des pluriels dits arabes et la comparaison avec l'emploi originel (en arabe) de ces formes démontrent que ces mots ont été fabriqués et généralisés selon le principe de l'analogie. Bien qu'une partie de ces types de mots soit empruntée telle quelle directement de l'arabe, beaucoup d'autres, même d'origine arabe, sont la fabrication des persanophones. Ces mots n'obéissent pas aux règles qui régissent ces formes en grammaire de l'arabe, mais correspondent simplement à l'application des lois de l'analogie. Farshidvard (2003[1969]), dans son ouvrage intitulé « De l'arabe en persan »⁴⁷, dans lequel il a rassemblé, entre autres, un ensemble de mots de fabrication iranienne qui ne suivent pas les règles de grammaire de l'arabe, constate que :

Les lois de l'arabe⁴⁸ sont utilisées en persan de deux manières : soit avec les mots et expressions arabes, soit avec les mots et expressions persanes et non arabes – et parfois certaines modifications et changements sont réalisés sur ces lois par les Iraniens (...). L'analogie, le suivi des habitudes langagières et l'ignorance des règles de la syntaxe et morphologie (sarf o nahv) de l'arabe sont les raisons principales de ces modifications. (Farshidvard 2003 : 147)

⁴⁷ *arabi dar fārsi.*

⁴⁸ *qava'ed arabi.*

Enfin, le persan n'a emprunté aucun suffixe, ni aucun procédé morphologique de l'arabe, mais a adopté un certain nombre de mots. Les locuteurs en ont déduit de nouvelles stratégies de formation de mots qu'ils appliquent indépendamment du trait [+/-arabe].

Autrement dit, il n'y a pas de formes « arabes » en morphologie du persan : « En soi la morphologie ne semble pas être empruntable », « Seuls les mots sont empruntés et la morphologie est à nouveau construite par les locuteurs de L2 » (Singh 2001a : 31).

Il faut mentionner que les formes « arabes » sont considérées plus prestigieuses en comparaison avec leurs contreparties en *-(h)ā* et appartiennent à la langue formelle, où les locuteurs ont plus tendance à respecter les règles normatives restrictives qui restreignent l'emploi de ces formes en fonction de l'origine des mots. Cela ne veut pas dire que les locuteurs ne sont pas en mesure de comprendre et/ou de produire des mots qui ne respectent pas cette norme. A titre d'exemple, voilà une phrase écrite dans un style formel ironique que j'ai rencontrée sur un *weblog* :

(22) *bande vazife-je vaks zadan-e kofu ra bar ohde daftam*

j'avais la responsabilité de cirer les chaussures

kofu n'est pas un mot du persan standard, mais est complètement compréhensible comme étant le pluriel de *kaf*, un mot d'origine persane d'usage ordinaire.

La théorie WWM refuse d'intégrer dans une description morphologique les considérations de type étymologique qu'elle considère comme non pertinentes pour la description de la compétence morphologique.

2

Description et discussion

2.1 Phonologie du persan

Selon la théorie WWM, avant de procéder à une analyse morphologique, il est nécessaire d'établir les propriétés phonologiques d'une langue afin d'éviter d'incorporer en morphologie les alternances automatiques et globales, pour lesquelles la phonologie est responsable, et de pouvoir identifier la part de chacune des parties de la grammaire dans les alternances.

Pour ce faire, je présente tout d'abord l'inventaire des phonèmes. J'apporte ensuite quelques clarifications sur les segments sujets à controverse, à savoir /ʔ/ et /ow/. Je m'attarde également sur le comportement de la consonne /h/ qui joue un rôle dans la formation du pluriel et l'analyse que l'on en fait dépend de son comportement. Je présente enfin les alternances automatiques et cela non de manière exhaustive, mais seulement dans la limite des besoins de ce travail.

2.1.1 Inventaire des phonèmes

L'inventaire des phonèmes que je propose contient 22 consonnes, 6 voyelles et une semi-consonne.

2.1.1.1 Phonèmes syllabiques

Voyelles et diphtongues				
	antérieure		postérieure	
fermée	i			u
	e			o /ow
ouverte	a		ɑ	

2.1.1.2 Traits des voyelles

	i	e	a	u	o	ɑ	ow
haute	+	-	-	+	-	-	- +
basse	-	-	+	-	-	+	-
ouverte	-	-	+	-	-	+	-
fermée	+	-	-	+	-	-	- +
avant	+	+	+	-	-	-	-
arrondie	-	-	-	+	+	-	+
étirée	+	-	+	-	-	-	-
tendue	+	-	-	+	-	+	+
longue	+	-	-	+	-	+	+

2.1.1.3 Phonèmes non-syllabiques

Consonnes et semi-consonnes									
	bilabiale	labio-dentale	alvéolaire	post-alvéolaire	palatale	vélaire	post-vélaire	uvulaire	glottale
occlusive	p b		t d			k g		q	ʔ
affriquée				tʃ dʒ					
nasale	m		n						
fricative		f v					x		h
spirante			s z	ʃ ʒ					
vibrante				r					
approximante					j				
latérale				l					

2.1.1.4 Quelques clarifications sur les segments

2.1.1.4.1 La diphtongue /ow/

Tout d'abord, il faut noter que /w/[-syl] n'est pas permis en persan de Téhéran⁴⁹ et que par conséquent la séquence [ow] n'existe que dans une position tautosyllabique. Cette séquence permet de différencier des paires de mots telles que :

/qol/ 'boule' vs /qowl/ 'promesse' vs /qul/ 'monstre'

/qom/ 'nom d'une ville au sud de Téhéran' vs /qowm/ 'ethnie'

/dowr/ 'tour' vs /dur/ 'loin'

/sowt/ 'voix' vs /sut/ 'sifflement'

De plus, dans le style *allegro*, cette séquence se prononce comme [o] ou [o.].

Etant donné les paires minimales notées ci-dessus, l'analyse de [ow] comme alternance allophonique de /o/ n'est pas acceptable.

Majidi (1986 : 56, 76), Samare (1999[1985] : 100-101) ont proposé d'analyser [w] comme alternance allophonique de /v/ quand il se trouve en coda précédé de /o/, en se basant sur les alternances du type :

/movɒʒ/ → [mowɒʒ] 'vague' vs /amvɒʒ/ 'les vagues'

/nov/ → [now] 'nouveau' vs /novin/ : [novin] 'relatif au nouveau'

→ /ov{C#, #}/ → [ow(C).]

⁴⁹ Pour adapter les mots d'emprunt, le /w/ consonantique est assimilé à /v/, comme /viski/ 'whisky'.

Mais, cette hypothèse n'est pas tenable non plus, parce que :

- bien qu'il n'y ait pas de mot qui contienne une séquence comme [-ovC.], la séquence [-ov#] n'est pas une séquence impossible pour les persanophones, étant donné qu'il existe des mots tels que /qolov/ 'exagération', /olov/ 'le plus haut'.
- dans la séquence /ov/, si, à la suite d'une re-syllabification, le /v/ se trouve en attaque, on pourrait s'attendre à ce que /ov/ se réalise en [o.v], comme c'est le cas dans la paire [now]/[novin]. Cela ne se produit cependant pas de manière systématique. Par exemple, la forme du pluriel des mots singuliers qui se terminent en [ow] n'est pas en [ova] comme on l'attendrait si on postule la forme sous-jacente comme /-ov/. Le pluriel de /mow/ ([mow], [mo:], [mo]) 'vigne' se prononce : [mowa], [mo:a], [moʔa], [moa], [moha], [mo:ha]. Si la forme sous-jacente était /mov/, le pluriel [movɑ] serait attendu, mais [movɑ] ne se dit jamais. La même chose se produit dans le contexte du marqueur du défini, alors que nous attendons [move], si la forme sous-jacente était /mov/, la forme définie de ce mot se prononcerait [mowe], [mo:e], [moʔe], [moe], [mohe], [mo:he] (comme pour les mots qui se terminent en voyelle), mais jamais [move].

Les faits ci-dessus nous amènent à postuler /ow/ comme un phonème syllabique du persan qui se réalise [ow], [o:] et [o] dans le style *allegro*.

2.1.1.4.2 *L'occlusive glottale*

L'analyse du statut phonologique de la consonne glottale⁵⁰ en persan n'est pas close et est sujette à controverse. En ce qui concerne la description (phonétique) des variantes de cette consonne en persan, à ma connaissance, les travaux de Samare (1999[1985]) : 62-65) et de Haghshenas (1991, 1995) sont les plus complets.

En résumé, il existe quatre réalisations différentes de la consonne glottale (Haghshenas 1995 : 24) :

- consonne glottale occlusive [ʔ] :
[ʔabr] 'nouage' ; [bedʔat] 'innovation' ; [ʃarʔ] 'religion'
- consonne glottale fricative [ʕ] :
[sa'id] 'un prénom' ; [bera'at] 'innocence' ; [qa'em] 'droit'
- consonne glottale fricative-occlusive [ʔʕ] :
[ra'ʔd] 'tonnerre' ; [e'ʔtelaf] 'fusion' ; [sa'ʔdi] 'poète iranien'
- prolongation vocalique :
[ɟa:mʔ] 'regroupement' ; [na:re] 'cri' ; [ba:d] 'après'

Ces réalisations sont en « distribution complémentaire », sachant que, dans « le style *lento* ou de citation, l'occlusive peut librement remplacer toutes les autres variantes ».

La description des variantes de la consonne glottale mise à part, il existe différentes analyses phonologiques :

⁵⁰ Dans la littérature persanophone, il est courant de parler de cette consonne sous le nom de *hamze*.

- statut phonémique dans tous les contextes (Samare 1999, Dinmohammadi 1996, Windfuhr 1997) ;
- dans tous les contextes, pour les mots [+arabe]⁵¹, statut phonémique et pour les mots [-arabe] statut d'élément prosodique (Cf. la phonologie polysystémique de Firth) (Haghshenas 1995) ;
- statut non-phonémique dans le contexte initial et intervocalique et statut phonémique ailleurs (Cf. Meshkatod Dini 1995, Sepanta 1973, Lazard 2006[1957]). (Chez Lazard, le statut phonémique est limité au style formel) ;
- la non-existence de ce phonème dans la variété informelle et son remplacement par l'allongement de la voyelle (ou la consonne) précédente (Cf. Lazard 2006).

Tous les travaux sur ce sujet font, d'une manière ou d'une autre, mention du fait que, dans la langue courante, la consonne glottale est instable et qu'elle est souvent réduite ou omise avec un allongement compensatoire le cas échéant (Cf. Lazard 2006[1957], Windfuhr 1997, Meshkatod Dini 1995, Samare 1999[1985], Mâhutiân 1997, Sadeghi 2001, Haghshenas 1995, Sepanta 1973, etc.). L'étude la plus récente est Jahani (2005) qui, dans le but de déterminer le statut phonologique de cette consonne, a étudié la prononciation de la consonne glottale dans différents contextes à l'aide d'entrevues ciblées.

En résumé :

⁵¹ Pour comprendre le point de vue de Haghshenas, il importe de noter qu'il est généralement admis que l'existence de cette consonne en persan moderne, à part en tant que consonne épenthétique, est due à l'adaptation des mots arabes. Les deux consonnes fricatives, la pharyngale voisée *ejn* et l'occlusive glottale non voisée *hamze*, aucune étant alors présente en persan, ont été perçues et réalisées comme une occlusive glottale.

Bien que tout le monde s'accorde pour dire qu'en persan toutes les voyelles en position initiale (de la syllabe) sont réalisées avec l'occlusive glottale comme l'attaque (Cf. 2.1.2.2), Jahani (2005) constate qu'en position d'attaque, **seulement** après une pause, l'occlusive glottale est **nécessairement** insérée. Lazard (2006) fait aussi le constat que, dans le langage informel, deux voyelles, qui en langage formel sont séparées avec l'occlusive glottale, sont en hiatus.

En position coda, la chute de la consonne glottale entraîne un allongement compensatoire de la voyelle précédente (Cf. Haghshenas 1995, Sepanta 1973). Jahani (2005) constate que, dans la paire minimale /bad/ 'mauvais' vs /baʔd/ 'puis', la distinction est normalement retenue avec la quantité vocalique et va jusqu'à souligner, dans une note de bas de page, la nécessité de poser la voyelle longue /a:/ comme un phonème : /a/ vs /a:/ vs /ɑ/.

En position intervocalique, la chute de la consonne glottale peut également engendrer l'allongement de la voyelle suivante (Cf. Sepanta 1973).

Il est important de mentionner qu'il peut exister de la variation dans la réalisation de /ʔ/, même dans des contextes phonologiques et des registres du langage identique. Selon Jahani (2005), ces variations, quand elles sont produites par le même locuteur, ont une explication de l'ordre de la pragmatique : la topicalisation, l'hésitation, la pause ou l'existence d'une paire minimale peuvent renforcer la réalisation de /ʔ/, c'est-à-dire empêcher son affaiblissement.

Dans le style formel, la consonne glottale est toujours prononcée. Dans ce style, elle peut même être réalisée comme consonne épenthétique dans un contexte où [j] ou [w] sont attendus (Cf. Sepanta 1973).

Je pense qu'il faut étudier davantage la question du statut phonologique de la consonne glottale. Aucune de ces études n'est exhaustive étant donné la réalité de la langue parlée, où l'on constate que le comportement de cette consonne diverge dans

le style de citation et rend l'établissement de la forme sous-jacente difficile. Voici un exemple qui en témoigne :

Le mot 'fonction, suiveur' dans le style de citation est prononcé [tabeʔ] c'est-à-dire que le mot se termine en /ʔ/. Toutefois, la même personne peut dire :

(23) man tabe-je ɖʒa:m-am

je suis la majorité

Lorsqu'un mot se termine en consonne /ʔ/, l'on attend que la forme d'*Ezafe*⁵² se dise /tabeʔe/ et non /tabeje/, comme si le mot se terminait par la voyelle /e/, c'est-à-dire que /tabeʔ/ ne peut pas être la forme sous-jacente. Pourtant, /ʔ/ est souvent prononcé dans la variété formelle ou dans le cas de la citation⁵³. Étant donné que cette forme de citation est souvent apprise tardivement dans le cours d'apprentissage de la langue, l'on peut imaginer que l'insertion de /ʔ/ est faite par une règle de lexique quand une forme [+formel] est exigée.

Toutefois, approfondir l'analyse de cette consonne dépasse le cadre de ce travail. Pour les besoins de cette étude, je me limite au style de citation. Je fais les suppositions suivantes :

/ʔ/ est un phonème.

⁵² Littéralement 'addition, adjonction'. L'*Ezafe* est un terme technique utilisé pour référer à un type de construction dans le syntagme nominal. En ce qui concerne le persan, tous les éléments suivis par un autre élément sont réalisés comme X-e ou XV-je : *barun-e bahare-je tehran* : 'la pluie printanière de Téhéran'.

⁵³ Il existe également des cas d'hypercorrection, où un mot qui finit par [e] est prononcé ou orthographié par certains locuteurs avec l'occlusive glottale à la fin : *ma(h)itabe* 'poète' : *mahitabeʔ*

En position initiale, avant une voyelle, l'insertion de /ʔ/ est prévisible :

$$\emptyset \rightarrow \text{ʔ} / \# \text{ ___ } V$$

N. B. : Il existe des mots pour lesquels /ʔ/ initial est présent dans la forme sous-jacente, *i.e.* /ʔadad/ 'chiffre'. Ce constat vient du fait que le pluriel de ce mot se dit /aʔdad/, ce qui correspond à la **SFM 28** :

$$/C_1V[+ \text{avant, -haut}]C_2aC_3/_{N.sg} \leftrightarrow /aC_1C_2aC_3/_{N.pl}$$

En position intervocalique quand aucune des trois voyelles /i/, /o/ et /u/ n'est présente, l'insertion de /ʔ/ est prévisible.

$$\emptyset \rightarrow \text{ʔ} / V \text{ ___ } V$$

N. B. : Il existe des mots pour lesquels /ʔ/ intervocalique est un phonème, *i.e.* /maʔaref/ 'connaissances'. Le constat que /ʔ/ est présent dans la forme sous-jacente vient du fait que le singulier se dit /maʔrefat/, ce qui correspond à la **SFM 39** :

$$/C1aC2C3eC4at/_{N.sg} \leftrightarrow /C1aC2aC3eC4/_{N.pl}$$

2.1.1.4.3 Le phonème /h/

Comme une règle générale de phonologie du persan, dans le style *allegro*, en position de coda, /h/ s'affaiblit, voire s'efface, quand il se trouve dans un groupe consonantique tautosyllabique ; dans les groupes consonantiques hétéro-syllabiques, la chute de /h/ est moins fréquente et s'accompagne toujours d'un allongement compensatoire (Cf. Samare 1999 : 62) :

$$/siah/ \rightarrow [sia] \text{ 'noir'}$$

$$/sobh/ \rightarrow [sob] \text{ 'matin'}$$

$$/mehrda(h)/ \rightarrow [me:rdad]/[merdad] \text{ 'nom propre'}$$

/mahsa/ → [ma:sa] ‘nom propre’

/kohne/ → [ko:wne] ‘vieux’

La variation se fait donc sur un *continuum* entre bien prononcé en style *lento* et très faible, voire effacé en style *allegro*. Il faut noter que, même dans le style *allegro*, l’effacement ne se fait pas de manière systématique⁵⁴.

2.1.2 Les alternances automatiques

2.1.2.1 L’attaque branchante

L’attaque branchante n’est pas permise en persan. En règle générale, un groupe consonantique non-initial est syllabifié de façon à attribuer une consonne à l’attaque et le reste en coda :

/kafʃduz/ → [kafʃ.duz] ‘coccinelle’

Quand plus d’une consonne se trouve en position initiale – par exemple dans le cas d’un emprunt ou d’interférence L2/L1 – si la première consonne du groupe est /s/ ou /ʃ/, la prothèse [e], et dans les autres cas l’épenthèse [e], est insérée ; cette consonne subit une assimilation [+arrondi] et [+haut] en fonction de la voyelle suivante. Dans certains cas, les deux prononciations coexistent, mais il me semble qu’il ne s’agit que d’une matière lexicale :

sport → [esport] ; snack → [esnak] ; style → [estil] ; Stuttgart → [eʃtutgart] ;

⁵⁴ L’effacement total se fait seulement pour certains mots très fréquents pour qui, étant donné que /h/ est présent dans leur forme de citation, l’on pose /h/ dans la forme sous-jacente. Or, il y a des faits qui laissent croire que, pour ces mots dans la variété informelle, /h/ est absent dans la forme sous-jacente. Cf. également la même sorte de problème dans le cas de l’occlusive glottale dans le paragraphe ci-dessus.

structure → [esteraktʃer] ;

cravate → [keravat], [kerevat] ; présentation → [perezentejʃen] ;

projet → [poroʒe] ; blond → [bolond], [belond] ; flute → [folut]/[fulut] ;

clips → [kilips], prise → [piriz].

2.1.2.2 L'attaque vide

En position initiale, si un mot commence par une voyelle dans sa forme sous-jacente, l'occlusive glottale remplit l'attaque (Cf. 2.1.1.4.2) :

$\emptyset \rightarrow ? / \# _ _ V$

En position intervocalique, une consonne épenthétique [ʔ], [w] ou [j] remplit la position d'attaque en fonction de la voyelle suivante. En présence de /i/, le glide [j], en présence de /o/ ou /u/ (et dans l'absence de /i/), le glide [w], et dans les autres cas [ʔ], sont insérés.

Les consonnes épenthétiques⁵⁵ sont réalisées moins prononcées que les phonèmes correspondants (N. B. : [w] n'a pas de statut phonémique en persan).

⁵⁵ Il importe de prendre note que, à part les consonnes mentionnées ci-dessous, autant que je puisse en parler, l'occurrence d'aucune autre consonne dans une position similaire ne peut être considérée comme automatique et donc doit être présentée dans la forme sous-jacente. Bien que, dans un point de vue diachronique, la conservation de /g/, /ɟ/ et /j/ dans les formes du pluriel comme /bandegan/, /miveɟat/ et /xodajan/ peut être vue comme dépendante d'une condition phonologique, cependant, dans une analyse synchronique, ces alternances ne sont pas prédictibles et, en conséquence, appartiennent au domaine de la morphologie et la consonne devrait être présente dans la forme sous-jacente.

Dans le style *allegro*, [ʔ] et [w] peuvent être omis et le groupe vocalique peut être prononcé en hiatus, et, dans le style *lento*, [ʔ] peut être inséré dans un contexte où [j] ou [w] sont attendus.

2.1.2.3 La coda branchante

La coda peut être minimalement branchante. Cependant, il semble y avoir une restriction quant au choix et à la façon dont les consonnes peuvent se regrouper en coda. Il est souvent avancé (Cf., entre autres, Meshkatod Dini 1995, Samare 1999, Haghshenas 2005[1977]) que la coda avec plus de deux consonnes n'est pas possible en persan. Il existe pourtant des contre-exemples tels que :

[orkestr], [marks], etc.

Dans le cadre de ce travail, étant donné que cette question ne se mêle pas à mon analyse, je ne m'attarderai pas davantage sur celle-ci.

2.1.2.4 Les géminées

En persan, les géminées sont réalisées seulement quand elles se trouvent dans la position hétérosyllabique. Mais les géminées tautosyllabiques existent dans la forme sous-jacente :

[kolije] 'rein' vs [kol.li.je.] 'le tout'

[haq] 'droit' vs la forme *ezafe* : [haq.qe] → /haqq/

Cependant, il y a des cas d'assimilation totale qui se produisent dans le style *allegro* et créent des géminées tautosyllabiques :

/dastkef/ : [das:kef] 'gants'⁵⁶

⁵⁶ Ce cas d'assimilation totale crée des cas inverses suite à l'hypercorrection : /massane/ → [mastane] 'vessie'

2.2

Liste des stratégies de formation du pluriel nominal

Dans cette section, je présente la liste des SFM (Cf. la présentation du modèle WWM en 1.3.1) qui sont réparties en deux groupes : **a)** les pluriels et **b)** les collectifs. J'ai établi jusqu'à 49 SFM, dont 8 pour le pluriel.

Chaque SFM est justifiée par deux paires de mots. Les mots sont transcrits dans leurs formes sous-jacentes accompagnées d'une traduction approximative en français. Les alternances phonologiques ne sont pas mentionnées. Par exemple, dans la forme /Xa/_{N,pl} quand le singulier se termine en /i/, étant donné que la séquence /ia/ n'est pas permise (Cf. 2.1.2.2), /j/ est automatiquement inséré et n'est pas présent dans la forme sous-jacente : /irani/ : /irania/ → [iranija]. Les SFM sont formulées de façon à ce qu'elles soient les plus générales et les moins redondantes possible afin de rendre compte des données de manière optimale. Dans certains cas, pour justifier le choix de formulation, plus de deux paires de mots sont présentées.

Aucun sous-regroupement quant aux types d'opération morphologique n'est fait et la liste ne suit pas un ordre donné d'avance, cependant l'ordre dans lequel je fais l'énumération reflète un désir de mettre ensemble les stratégies dont les parties spécifiées, autrement dit les constantes (*vs* les variables), se ressemblent et de les présenter dans l'ordre ascendant quant à la complexité de ces parties spécifiées.

a) Pluriels

1. $/X/_{N.sg} \leftrightarrow /Xa/_{N.pl}$ ⁵⁷

/ketab/ 'livre' /ketaba/

/miz/ 'table' /mizha/

2. $/Xe/_{N.sg} \leftrightarrow /Xeha/_{N.pl}$

/mive/ 'fruit' /miveha/

/gorbe/ 'chat' /gorbeha/

3. $/Xa/_{N.sg} \leftrightarrow /Xaha/_{N.pl}$

/faza/ 'espace' /fazaha/

/baba/ 'papa' /babaha/⁵⁸

⁵⁷ Dans la langue écrite ainsi que dans la variété formelle, *-ha* s'attache à n'importe quel mot et en forme le pluriel. Il est courant dans la littérature aussi bien normative que descriptive de postuler la forme *-ha* comme la forme par excellence du pluriel. Cependant, en ce qui concerne la langue courante, *-ha* s'emploie seulement avec les mots qui se terminent en /a/ ou /e/. Toutefois, synchroniquement parlant, aucune des deux formes ne peut être supposée de dériver de l'autre par une règle phonologique : d'un côté, nulle part ailleurs dans le même contexte phonologique l'effacement de *-h* ne se fait de manière automatique, et, d'un autre côté, l'analyse de *-h* comme consonne épenthétique n'est pas justifiée.

Etant donné que la forme que l'on acquiert le plus tôt dans le cours d'apprentissage de la langue est la forme en *-a* et que la forme avec /h/ qualifié de [+formel] s'apprend plus tardivement, souvent par l'enseignement scolaire, l'on peut supposer que, quand le niveau du langage exige un mot [+formel], l'ajout de /h/ se fait par une règle lexicale purement stylistique: $/Xa/_{N.pl} \leftrightarrow /Xha/_{N.pl.[+formel]}$.

⁵⁸ Etant donné l'observation de Samvelian (2007) (Cf. 1.4), nous pouvons également établir une seule stratégie en prenant comme base le nom défini : $/Xe/_{N.déf.sg} \leftrightarrow /Xa/_{N.déf.pl}$ gorbehe : gorbeha ; mize : miza.

4. /XC/_{N.sg} ↔ /XCan/_{N.pl.formel}
 /zan/ ‘femme’ /zanan/
 /deraxt/ ‘arbre’ /deraxtan/
5. /Xe/_{N.sg} ↔ /Xegan/_{N.pl.formel}
 /honarpiʃe/ ‘comédien’ /honarpiʃegan/
 /fereʃte/ ‘ange’ /fereʃtegan/
6. /XV_[+longue]/_{N.sg} ↔ /X V_[+longue]jan/_{N.pl.formel}
 /geda/ ‘mendiant’ /gedajan/
 /tʃarpa/ ‘quadrupède’ /tʃarpajan/
 /daneʃɖu/ ‘étudiant’ /daneʃɖujan/
 /soxangu/ ‘locuteur, porte-parole’ /soxangujan/
7. /XV_[+arrondie,-ouverte]/_{N.sg} ↔ /Xovan/_{N.pl.formel}
 /pejrow/ ‘partisan’ /pejrovan/
 /xosrow/ ‘roi’ /xosrovan/
 /ahu/ ‘cerf’ /ahovan/
 /banu/ ‘dame’ /banovan/
8. /XC/_{N.sg} ↔ /XCin/_{N.pl.formel}
 /moaven/ ‘assistant’ /moavenin/
 /karbar/ ‘utilisateur’ /karbarin/

b) Collectifs

9. /**Xi**/_{N.sg} ↔ /**Xiun**/_{N.collectif}
 /mazhabi/ ‘religieux’ /mazhabiun/
 /rohani/ ‘clergé’ /rohaniu/
10. /**X**/_N ↔ /**Xɖat**/_{N.collectif}
 /mive/ ‘fruit’ /miveɖat/
 /boronz/ ‘bronze’ /boronzɖat/
11. /**X**/_{N.sg} ↔ /**Xiat**/_{N.collectif}
 /qazal/ ‘une forme de poème’ /qazaliat/
 /haɖv/ ‘satire, vulgarité’ /haɖviat/
12. /**X**/_{N.sg} ↔ /**Xat**/_{N.collectif}
 /tasadof/ ‘accident’ /tasadofat/
 /piɧnahad/ ‘proposition’ /piɧnahadat/
13. /**Xe**/_{N.sg} ↔ /**Xat**/_{N.collectif}
 /moadele/ ‘équation’ /moadelat/
 /kalame/ ‘mot’ /kalamat/
14. /**Xat**/_{N.sg} ↔ /**Xat**/_{N.collectif}
 /adat/ ‘habitude’ /adat/
 /ɖahat/ ‘direction’ /ɖahat/
15. /**XC₁C₂e**/_{N.sg} ↔ /**XC₁aC₂at**/_{N.collectif}
 /ɖomle/ ‘phrase’ /ɖomalat/
 /zarbe/ ‘coup’ /zarabat/

16. /XC₁C₂at/_{N.sg} ↔ /XC₁aC₂at/_{N.collectif}

/hazrat/ 'seigneur' /hazarat/

/neʔmat/ 'richesse' /neʔamat/

17. /C₁eC₂C₃at/_{N.sg} ↔ /C₁eC₂aC₃/_{N.collectif}

/mellat/ 'nation' /melal/

/hekmat / 'sagesse' /hekam/

18. /C₁V_[+arrondie,-ouverte]C₂at/_{N.sg} ↔ /C₁ovaC₂/_{N.collectif}

/dowlat/ 'gouvernement' /doval/

/surat/ 'visage' /sovar/

19. /C₁V_[-longue,-ouverte]C₂C₃e/_{N.sg} ↔ /C₁V_[-longue,-ouverte]C₂aC₃/_{N.collectif}

/qesse/ 'histoire' /qesas/

/herfe/ 'profession' /heraf/

/ʃoʔbe/ 'succursale' /ʃoʔab/

/nosxe/ 'version' /nosax/

20. /(C₁)aC₂C₃/_{N.sg} ↔ /(C₁)oC₂uC₃/_{N.collectif}

/dars/ 'leçon' /dorus/

/haqq/ 'droit' /hoquq/

/asl/ 'principe' /osul/

/ʃart/ 'condition' /ʃorut/

21. /(C₁)aC₂C₃/_{N.sg} ↔ /(C₁)aC₂aʃeC₃/_{N.collectif}

/arz/ 'propos' /arqez/

/ʃart/ 'condition' /ʃarqet/

22. /C₁V_[-longue]C₂C₃/_{N.sg} ↔ /aC₁C₂aC₃/_{N.collectif}

/fard/ 'personne' /afrad/

/lafz/ 'expression' /alfaz/

/lotf/ 'gentillesse' /altaf/

/xolq/ 'attitude' /axlaq/

/fekr/ 'pensée' /afkar/

/qesm/ 'type' /aqsam/

23. /vaC₁C₂/_{N.sg} ↔ /owC₁aC₂/_{N.collectif}

/vaqt/ 'temps' /owqat/

/vaz?/ 'situation' /owza?/

24. /V_[+avant,-haut]C₁C₂/_{N.sg} ↔ /aC₁aC₂i/_{N.collectif}

/ahl/ 'originaire de' /ahali/

/esm/ 'nom' /asami/

25. /C₁aC₂i/_{N.sg} ↔ /aC₁C₂ia/_{N.collectif}

/nabi/ 'prophète' /anbia/

/qani/ 'riche' /aqnia/

26. /vaC₂i/_{N.sg} ↔ /owC₁ia/_{N.collectif}

/vasi/ 'héritier' /owsia/

/vali/ 'tuteur' /owlia/

27. /C₁V_[+longue]C₂/_{N.sg} ↔ /aC₁vaC₂/_{N.collectif}

/mowɟ/ 'vague' /amvaɟ/

/now?/ 'variété' /anva?/

/hal/ 'état' /ahval/

/mal/ 'bien, possession' /amval/

/nur/ 'lumière' /anvar/

28. /C₁V_[+avant,-haut]C₂aC₃/_{N.sg} ↔ /aC₁C₂aC₃/_{N.collectif}

/heram/ 'pyramide' /ahram/

/maraz/ 'maladie' /amraz/

29. /vaC₂aC₃/_{N.sg} ↔ /owC₁aC₂/_{N.collectif}

/valad/ ‘enfant’ /owlad/

/varaq/ ‘papier’ /owraq/

30. /aC₁aC₂/_{N.sg} ↔ /aC₁aC₂/_{N.collectif}

/asar/ ‘œuvre’ /asar/

/adab/ ‘politesse, coutume’ /adab/

31. /(C₁)aC₂iC₃(e)/_{N.sg} ↔ /(C₁)aC₂ajeC₃/_{N.collectif}⁵⁹

/vazife/ ‘devoir’ /vazajef/

/natiǰe/ ‘résultat’ /nataǰeǰ/

/arize/ ‘requête, pétition’ /arajez/

/dalil/ ‘raison’ /dalajel/

/zimir/ ‘pronom’ /zamazjer/

32. /C₁aC₂ie/_{N.sg} ↔ /C₁aC₂aja/_{N.collectif}

/qazie/ ‘postulat’ /qazaja/

/hadie/ ‘cadeau’ /hadaja/

33. /C₁aC₂iat/_{N.sg} ↔ /C₁aC₂aja/_{N.collectif}

/maziat/ ‘avantage’ /mazaja/

/vasiat/ ‘conseil’ /vasaja/

⁵⁹ Le mot du côté droit de cette stratégie peut également avoir la forme /C₁aC₂aeC₃/ : /vazaef/ : [vazaʔef] ; /nataeǰ/ : [nataʔeǰ] ; /dalael/ : [dalaʔel] ; /zamaer/ : [zamaʔer] ; /dalael/ : [dalaʔel] ; /araez/ : [araʔez]. La variation dans la forme du pluriel n’est pas systématique et peut exister même chez la même personne.

34. /C₁aC₂iC₃at/_{N.sg} ↔ /C₁aC₂ajeC₃/_{N.collectif}⁶⁰
 /qanimat/ ‘pillage’ /qanqjem/
 /haqiqat/ ‘vérité’ /haqqjeq/
35. /C₁aC₂iC₃e/_{N.sg} ↔ /C₁oC₂oC₃/_{N.collectif}
 /madine/ ‘ville’ /modon/
 /tariqe/ ‘façon’ /toroq/
36. /C₁eC₂aC₃/_{N.sg} ↔ /C₁oC₂oC₃/_{N.pl}
 /ketab/ ‘livre’ /kotob/
 /mesal/ ‘exemple’ /mosol/
37. /C₁aC₂C₃V_[+avant,-haut]C₄/_{N.sg} ↔ /C₁aC₂aC₃eC₄/_{N.collectif}
 /maxzan/ ‘réservoir’ /maxazen/
 /bandar/ ‘quai’ /banader/
 /masdʒed/ ‘mosquée’ /masaʒed/
 /mahfel/ ‘cercle, compagnie’ /mahafel/
38. /C₁aC₂C₃V_[-longue]C₄e/_{N.sg} ↔ /C₁aC₂aC₃eC₄/_{N.collectif}
 /mantaqe/ ‘région’ /manateq/
 /taʒrobe/ ‘expérience’ /taʒareb/
 /madrese/ ‘école’ /madares/
39. /C₁aC₂C₃eC₄at/_{N.sg} ↔ /C₁aC₂aC₃eC₄/_{N.collectif}
 /maʔrefat/ ‘connaissance’ /maʔaref/
 /manfeʔat/ ‘bénéfice’ /manafeʔ/

⁶⁰ Comme dans la SFM 31, le mot du côté droit de cette stratégie peut également avoir la forme /C₁aC₂aeC₃/ : /qanaem/ : [qanaʔem] ; /haqaeq/ : [haqaʔeq].

40. /C₁aC₂C₃a/ N.sg ↔ /C₁aC₂aC₃i/ N.collectif
 /maʔna/ ‘signification’ /maʔani/
 /daʔva/ ‘dispute’ /daʔavi/
41. /C₁aC₂eC₃/ N.sg ↔ /C₁aC₂aC₃e/ N.collectif
 /saken/ ‘habitant’ /sakane/
 /kaseb/ ‘commerçant’ /kasabe/
42. /(C₁)aC₂eC₃/ N.sg ↔ /(C₁)oC₂C₂aC₃/ N.collectif
 /hazer/ ‘présent’ /hozzar/
 /taɖʒer/ ‘commerçant’ /toɖʒar/
 /aʃeq/ ‘amoureux’ /oʃʃaq/
43. /(C₁)aC₂eC₃/ N.sg ↔ /(C₁)oC₂aC₃a/ N.collectif
 /ʃaʔer/⁶¹ ‘poète’ /ʃoʔara/
 /aqel/ ‘sage’ /oqala/
 /aref/ ‘soufi’ /oraf/
44. /(C₁)aC₂iC₃/ N.sg ↔ /(C₁)oC₂aC₃a/ N.collectif
 /faqir/ ‘pauvre’ /foqara/
 /safir/ ‘ambassadeur’ /sofara/
 /asir/ ‘otage’ /osara/
45. /C₁V_[+arrondie, +longue]C₂eC₃(e)/ N.sg ↔ /C₁avaC₂eC₃/ N.collectif
 /sahel/ ‘côte’ /savahel/
 /zaher/ ‘apparence’ /zavaher/

⁶¹ Ici, [ʔ] fait partie de la forme sous-jacente et n’est pas inséré automatiquement étant donné :

/ʃaʔer/ : /ʃeʔr/ vs /faaʔel/ : /feʔl/, /saher/ : /sehr/, etc.

/mowze?/ ‘position’ /mavaze?/

/mowqe?/ ‘temps’ /mavaqe?/

/tajefe/ ‘tribu’ /tavajef/

/hadese/ ‘accident’ /havades/

46. /C₁aC₂eC₃e/ N.sg ↔ /C₁a C₂a_{je} C₃/ N.collectif

/faɖʒe?e/ ‘catastrophe’ /faɖʒa_{je}?/

/vaqe?e/ ‘accident’ /vaqa_{je}?/

47. /C₁aC₂ie/ N.sg ↔ /C₁avaC₂i/ N.collectif

/haʃie/ ‘marge’ /havaʃi/

/sanie/ ‘seconde’ /savani/

48. /C₁aC₂V_[+longue]C₃/ N.sg ↔ /C₁avaC₂iC₃/ N.collectif

/qanun/ ‘loi’ /qavanin/

/tarix/ ‘histoire, date’ /tavarix/

49. /C₁V_[-longue]C₂C₃V_[+longue]C₄/ N.sg ↔ /C₁aC₂aC₃iC₄/ N.collectif

/farman/ ‘ordre’ /faramin/

/soltan/ ‘roi’ /salatin/

/ʃejtan/ ‘diable’ /ʃajatin/

/mejdan/ ‘place’ /majadin/

/taklif/ ‘devoir’ /takalif/

/darviʃ/ ‘derviche’ /daraviʃ/

/mafhum/ ‘notion’ /mafaim/

/ɖʒomhur/ ‘république’ /ɖʒamahir/

Dans les deux sections suivantes, je présenterai quelques réflexions concernant les considérations sémantiques, les pluriels doubles ainsi que la variation.

2.3

Considérations sémantiques à prendre en compte

La théorie WWM prédit que si une forme satisfait les conditions structurales d'une stratégie, *i.e.* gagne face aux stratégies concurrentes, il n'y a aucune contrainte supplémentaire qui peut exclure la formation de ce mot au niveau de compétence d'un locuteur. L'exclusion d'un mot potentiel se fait par des considérations extérieures à la compétence morphologique.

Par exemple, le fait que **childs* n'est pas acceptable en anglais standard n'a rien à voir avec la morphologie et un locuteur putatif, par exemple un enfant en cours d'acquisition de la langue avant d'être influencé par la norme, peut très bien former et employer *childs*, le cas échéant.

Ainsi, la théorie WWM exclut toutes les conditions sémantiques imposées sur les mots qui participent dans une stratégie de formation du mot et, pour les architectes de cette théorie, « les propriétés sémantiques d'un mot ne sont pour rien dans les opérations de formation du mot » (Ford et Singh 1991 : 564). Néanmoins, prendre cette affirmation au pied de la lettre peut aboutir à une description qui manque à refléter une partie de la réalité.

Dans la description présentée plus haut, je me suis limitée à des étiquettes comme *N.sg*, *N.pl*, *N.pl.formel*, *N.pl.générique* et *N.collectif*, et je n'ai formulé aucune règle d'interprétation (ci-après RI) pour accompagner ces stratégies. Dans cette section, je compte développer les particularités sémantiques sous-jacentes dans ces formes et proposer éventuellement des RI.

2.3.1 Les formes en *-an*, *-in* et *-un*

Dans un premier temps, quand je faisais mon analyse des données du persan avec le souci d'éliminer toutes les contraintes sémantiques de ma description, je cherchais des contre-exemples pour justifier cette démarche : des contre-exemples qui montreraient que ce n'est pas « vrai » que la formation en *-an* est limitée sémantiquement par le trait [+animé], comme le suggérait et préconisait la littérature (grammaires prescriptives et descriptives incluses), malgré les quelques pluriels [-animé], un ensemble fermé maintenu depuis une époque dans la langue où *-an* n'était pas restreint⁶² : *gonahan*, *setaregan*, *sajaregan*, *axtaran*, etc., ainsi que pour les parties de corps en paires : *tfejman*, *dastan*, *bazovan*, *angoftan*, etc.

Les quelques contre-exemples rencontrés notamment dans les textes rédigés sur l'Internet, non au niveau de la production plus spontanée, *i.e.* dans les forums (je n'ai jamais noté de tels contre-exemples dans une production orale), semblaient être soit une tentative « non-réussie » pour parler plus soigneusement, *i.e.* un cas d'hypercorrection, sinon un choix pour respecter une rime, un rythme ou encore une « erreur »⁶³.

Ces emplois frappent sûrement l'oreille, mais cela peut s'expliquer par le fait que les formes en *-an* appartiennent à la variété formelle et, dans cette variété, l'on est normalement tenu de respecter la norme. Autrement dit, l'on peut maintenir qu'au niveau de la compétence morphologique, les contraintes sémantiques ne jouent

⁶² Dans la période du « moyen persan », *-an* était la seule marque du pluriel (Cf. Khanlari 1969 : 83)

⁶³ Par exemple, dans un article où dans un syntagme la forme *-an* est généralisée aux autres mots : *saran-e keshvaran-e...* : *saran-e keshvarha-je...* 'les têtes des pays...'

aucun rôle. Cependant, cette analyse pourrait cacher une partie de la réalité que j'essaye de décrire par la suite.

Si un adjectif, ou un nom générique (en persan, il existe la relation d'identité entre l'adjectif et le nom : $/X/_N \leftrightarrow /X/_{Adj}$), peut caractériser une entité indépendamment de son trait [+/-animé], la forme (potentielle) *-an* correspondante, qui est donc un nom pluriel référentiel, ne peut se référer qu'à un ensemble d'être humain :

sar 'tête' peut, en tant qu'adjectif, caractériser une entité animée ou non animée (in *xune/daneŝamuz bein-e hame sar-e* 'cette maison/cet étudiant est en tête de tous/est la meilleure'), et, en tant que nom, référer à la partie haute d'une entité non-animée quelconque (*sar-e xiabun* 'le bout de la rue'). Mais *saran* ne peut désigner que des hommes avec cette caractéristique, *i.e.* 'en tête de quelque chose', tandis que *sarha* n'a pas de restriction de cette sorte.

Quand un locuteur entend un mot en *-an* qu'il n'a pas entendu auparavant et dont il comprend l'adjectif correspondant (*i.e.* la forme sans *-an*), sa démarche pour comprendre ce mot nouveau consisterait à chercher à imaginer dans quelles situations l'on peut caractériser un homme par un tel adjectif ou quelle sorte d'hommes peut correspondre à une telle caractérisation. Autrement dit, le fait que ce mot désigne un être humain va de soi⁶⁴. Ainsi, à part pour l'ensemble clos et lexicalisé des noms non-animés pour lesquels un pluriel en *-an* est disponible, les **SFM 4**, **SFM 5** et **SFM 6** sont utilisées activement dans la formation des mots nouveaux, *i.e.* des noms pluriels [+formel], qui partagent tous le trait [+humain].

Pour refléter cette réalité, je propose une RI que l'on peut formuler comme :

⁶⁴ Il faut noter que cela n'est pas le cas des formes correspondantes en *-ha* qui peuvent aussi bien référer aux entités [+animé] que [-animé].

$/X(g,j)an/_{N,pl.formel} \leftrightarrow$ 'l'ensemble d'êtres humains/d'hommes avec la propriété
 $/X(e,\{a,u\})/_{Adj}$ '

La forme en *-in* est soumise à des restrictions sémantiques/pragmatiques semblables. En fait, cette forme peut se substituer à la forme *-an* quand le mot se termine par une consonne. Cependant, l'emploi des formes en *-in* est plus limité que celui de *-an*, mais cela dépend de raisons extralinguistiques. Premièrement, au niveau de la norme⁶⁵, la forme *-in* n'est acceptable que pour des mots d'origine arabe ; de plus, aujourd'hui, l'emploi d'un pluriel en *-in* est fortement déconseillé au profit de sa contrepartie en *-an*, dans une tendance générale d'éviter les éléments « d'origine arabe » quand une « solution plus persane » est disponible.

La forme en *-un* est également soumise à des restrictions sémantiques/pragmatiques semblables à la forme en *-an*, mais elle présente une interprétation plus précise que celle pour *-an* et, de plus, cette stratégie, dans son emploi actif dans la formation de nouveaux mots, prend comme base une forme $/Xi/_{N,Adj}$ qui se trouve dans une stratégie comme :

$/X/_{N} \leftrightarrow /Xi/_{Adj}$ 'lié à $/X/$ '

$/mazhab/_{N}$ 'religion' : $/mazhabi/_{Adj,N}$ 'religieux, une personne/un acte religieux'

$/iran/_{N}$: $/irani/_{Adj,N}$ 'iranien', 'Iranien'

⁶⁵ Cette forme est également considérée [+formel] et s'emploie donc essentiellement dans des situations formelles où les locuteurs ont le plus tendance à respecter la norme.

Quand un locuteur entend une forme /Xiun/_{N,pl} qu'il n'avait pas entendue auparavant, il comprend 'l'ensemble de gens qui sont liés, croient ou soutiennent /X/' :

/qejremazhabiun/ 'ceux qui soutiennent la non-croyance, ceux qui ne croient pas à la religion'

/xatamiun/ 'ceux qui soutiennent ou ont soutenu Khatami (l'ancien président de l'Iran)'

/ahmadinezadiun/ 'ceux qui soutiennent Ahmadinejad (l'actuel président de l'Iran)'

Pour refléter cette réalité, je propose une RI comme :

/Xiun/_{N,pl} ↔ 'les partisans de /X/_N'

Somme toute, pour rendre compte de la réalité de la formation de mots nouveaux en ce qui concerne ces trois formes, nous sommes obligés de proposer des RI qui restreignent l'interprétation de ces formes pour ne référer qu'aux hommes. Or, cette RI équivaut à doter /Xan/_{N,pl} d'un trait [+humain]. Cependant, l'ajout d'un tel trait n'est pas toléré par le modèle WWM⁶⁶. Nous ne pouvons cependant pas nous passer de cette RI si nous voulons maintenir une description réaliste de la compétence morphologique et de son pouvoir de prédiction.

⁶⁶ « In WWM, there would be no such [generic] semantic feature in the WFS as the interpretations of the two poles of the strategy need only be 'a function of' each other... saying that X is a function of X' amounts to saying that X' equals X plus some generic semantic feature » (Dubé 2008 : 489).

2.3.2 Les formes en *-iat* et *-dʒat*

Les deux formes *-iat*⁶⁷ et *-dʒat*⁶⁸ se différencient également des autres formes par leurs RI : elles désignent ‘un ensemble de différentes sortes de quelque chose’, ou ‘un ensemble de choses liées à quelque chose’ ou un emploi comme nom générique.

(24) *masraf-e ziad-e firiničʒat/firini/*firiniha barâje salamti mozer ast*

consommer trop de sucreries nuit à la santé

firiničʒat ‘différentes sortes de gâteaux’⁶⁹ ou ‘la famille des gâteaux en général’

taraqečʒat ‘différentes sortes de pétards’, *taraqe* ‘pétard’

talačʒat ‘l’ensemble de ce qui est fait d’or’

⁶⁷ Certaines analyses ignorent *-iat* (Cf., entre autres, Meshkatod Dini 2008[2005] et Majidi 1986) ou l’analysent comme une variante de *-at* (Cf., entre autres, Karimi-Doostan 2004 et Kalbasi 2008[1993]), faisant abstraction d’exemples comme *qazal vs qazaliat* où *-i* n’est pas présent dans le singulier, mais est présent dans le pluriel. La raison en est que la présence de cette forme est due (Cf. Mo’in 1961 : 119) au fait qu’à l’origine (en arabe) *qazaliat* est le pluriel de l’adjectif *qazalie* : le syntagme *qatačʒat-e qazalie* ‘les morceaux (de poésie) de qazal’ repris par *qazaliat*. Mais *qazalie* n’est pas un mot du persan et, pour un persanophone, *qazal* est lié directement à *qazaliat* sans passer par *qazalie*.

⁶⁸ Certains travaux (Cf., entre autres, Karimi-Doostan 2004, Majidi 1986, Kalbasi 2008[1993]) considèrent *-dʒat* comme une variante phonologique de *-at* et [čʒ] comme une consonne épenthétique. Cette analyse non justifiée est due au fait qu’à une certaine période, *-dʒat* se limitait aux pluriels dont le singulier se terminait en *-a*, *-i* ou *-u* : *ruzname-dʒat*, *daru-dʒat* et *firini-dʒat* ; aujourd’hui cette limite ne tient plus. De plus, il existe des cas où les deux formes co-existent : *edare* ‘bureau, administration, service’ : *edarat vs edaredʒat*, *havale* ‘virement’ : *havalat vs havaledʒat*, et *qabale* ‘contrat, bail’ : *qabalat vs qabaledʒat*.

⁶⁹ *firiniha-je moxtalef*.

Bien que d'autres formes puissent, si le contexte l'exige, porter les sortes de sens mentionnées ci-dessus, l'inverse n'est pas vrai et les formes en *-ɖʒat* et *-iat* ne portent jamais un sens référentiel et défini.

(25) emruz raftam mive xaridam vali miveha-i/*miveɖʒat-i ke xaridam ɖʒa mundan tu maqaze

aujourd'hui je suis allé acheter des fruits mais les fruits que j'ai achetés ont été (sont) oubliés dans le magasin

(26) diruz bara-je in belag nazar neveʒte budam am :a nazaraq-i/*nazarat-i ke neveʒte budam tʃap naʒodan

hier, j'ai écrit des commentaires pour ce blog mais les commentaires que j'ai écrits ne sont pas publiés

Les formes en *-iat* réfèrent normalement aux entités abstraites comme la production langagière, les idées, etc. :

haɖʒviat, haɖʒv 'prose humoristique', robaiat, robai 'quatrain', tʃarndiat, tʃarand 'absurdité', etc.

Tandis que les formes en *-ɖʒat* réfèrent plutôt aux entités concrètes.

L'on constate plus clairement cette distinction dans les cas où les deux formes sont employées :

Bien que *mozaxrafɖʒat*_{N.Collectif} peut se référer aussi bien à un ensemble de nourritures ou de films ou d'émissions ou toute autre chose qu'à un ensemble de productions langagières que l'on a qualifié de mauvais, *mozaxrafiat*_{N.Collectif} ne peut se référer qu'à un ensemble d'idées et de discours (*mozaxraf*_{N.Adj} 'absurde, mauvais').

La formation des mots nouveaux démontre également une telle tendance :

sandevitǰat, alkoldǰat, pitzadǰat, kampiuterǰat, esemesǰat, etc.

Jeʔriat (Jeʔr 'poème'), efqiat (efq 'amour'), havasiat (havas 'l'envie'), qalatiat (qalat 'erreur'), etc.

Cependant, contrairement aux cas de *-an*, *-un*, *-in*, avec les données actuelles pour ces formes, l'on ne peut que parler d'une tendance. Pour donner un contre-exemple, j'ai déjà entendu *maǰngǰat* pour parler des gens *maǰng* [+familier] 'fou, qui est dans la lune' : *je moǰt xeng o xol o maǰngǰat* 'une bande d'idiots, de fous, de folles-dingues', également *ahvalǰat* au lieu de *ahval* 'les états (de santé/de l'esprit)' dans la phrase *ahvalǰat-et tǰetor-e?* 'comment va (tes états) ? [très familier]'.

Un cas intéressant est le mot utilisé pour référer aux différentes sortes de *araq* (*araqha-je giahi*) 'l'eau des fleurs/plantes'. Les deux mots *araqiat* et *arqidǰat* sont utilisés parallèlement ; *araqiat* est qualifié de correct et *arqidǰat* d'une formation « erronée » tardive. Toutefois, j'ai observé qu'un locuteur, qui n'a ni l'un ni l'autre dans son lexique interne, peut former le mot *araqǰat* spontanément. Il faut noter que, dans le langage familier, *araq* réfère à la boisson alcoolisée avec une forte connotation péjorative (étant donné l'interdiction de consommer de l'alcool dans l'Islam). Cela peut en partie expliquer la préférence pour la forme *arqidǰat*, qui semble être un hybride de *araqǰat* et de *araqiat*.

2.4

Les pluriels doubles

J'ai déjà mentionné dans la section 1.4. l'existence des pluriels doubles et je me suis basée là-dessus pour justifier le statut particulier réservé aux formes en $-(h)a$. Dans cette section, je voudrais développer davantage ce fait.

En 1.4., j'ai avancé que les pluriels doubles sont particulièrement significatifs dans le cas des « pluriels brisés » et, en deuxième lieu, les pluriels en $-at$, bien que je ne puisse pas affirmer l'inexistence dans les autres cas, je n'étais pas en mesure d'en présenter des exemples. De plus, dans la section précédente, j'ai discuté des particularités sémantiques de toutes les autres formes, qui nous permettent de regrouper les pluriels en $-an$, $-in$ et $-un$, ainsi que les collectifs en $-iat$ et $-dʒat$. Donc, il reste les « pluriels brisés » et les pluriels en $-at$, qui eux partagent ce phénomène de pluriel double⁷⁰.

Ces deux formes peuvent, comme $-an$ et $-in$, être qualifiées de [+formel]. Dans la variété formelle et soutenue, ces deux formes sont utilisées de manière semblable aux pluriels en $-(h)a$ ⁷¹. Il arrive qu'un locuteur reprend son discours pour remplacer un pluriel en $-(h)a$ par sa contrepartie « brisé » dans une situation formelle :

⁷⁰ N. B. : Certains mots de ces formations ont perdu complètement leurs lien sémantique avec le mot singulier de base, c'est le cas de *tazahorat*, *entexabat*, *dʒavaher*, etc.

⁷¹ Il faut noter que ces types de pluriels sont qualifiés d'être « d'origine arabe » et la norme exige donc que le mot singulier soit un mot d'origine arabe. Certains emplois non normatifs,

(27) dar dʒameʔeha dʒavameʔ-e eslami... (dʒameʔe ‘société’)

dans les sociétés islamiques...

Comme nous l’avons vu dans la section précédente, les pluriels doubles se produisent même dans un discours formel. J’y ai expliqué ces emplois – dans les contextes impliquant la définitude – en faisant l’hypothèse que le pluriel en *-(h)a* implique forcément le trait [+défini], tandis que les autres sont seulement compatible avec un tel trait, *i.e.* cela n’est pas automatique.

Néanmoins, ces pluriels doubles se produisent également dans les contextes qui n’impliquent pas forcément la définitude et cela même dans un discours formel. Bien qu’il soit possible pour ces productions d’être de simples cas d’ ‘erreur de parole’, dans tous les cas, la grammaire doit pouvoir en fournir une explication.

Je propose deux autres hypothèses pour expliquer ces emplois :

Premièrement, nous pouvons faire l’hypothèse que *-(h)a* est un marqueur de classe (Ford et Singh 1991 : 564-565) et qu’un locuteur n’interprète pas un mot qui ne se termine pas en *-(h)a* comme un nom pluriel à proprement parler, *i.e.* syntaxiquement pertinent. Cela revient également à dire que [+défini] est un trait du pluriel nominal, en tant que catégorie syntaxiquement pertinente, en persan.

A ce point, je dois mentionner le fait qu’en persan l’accord verbal en nombre n’est pas un élément décisif pour distinguer entre un nom pluriel et un nom non-pluriel. De plus, l’accord en nombre n’existe nulle part ailleurs, *i.e.* il n’y a pas d’accord au niveau de l’adjectif, etc. En persan, l’accord est obligatoire seulement pour les sujets au pluriel [+humain]. La règle prescriptive indique qu’un sujet [-humain] soit

justifié par leur emploi dans la grande littérature ou par leur ancienneté, sont pourtant acceptés. Bien évidemment, les locuteurs emploient cette forme sans souci de l’origine des mots auxquels ils n’ont pas forcément accès.

toujours accompagné d'un verbe au singulier, mais, dans la réalité, les sujets pluriels [-humain] prennent aussi bien des verbes au singulier qu'au pluriel, et les grammaires descriptives soulignent aussi qu'un sujet [-animé] peut s'accompagner d'un verbe au singulier ou d'un verbe au pluriel, sans précision sur les conditions et les contextes précis pour chacun, ce qui laisse l'impression que cet accord verbal jouit d'une variation assez libre. Cependant, il semble que l'interprétation distributive *vs* collective joue un rôle dans cet accord.

Développer davantage ce point dépassait le cadre de ce travail. Etant donné que l'accord verbal en nombre, dans l'état actuel de l'analyse que l'on peut en faire, ne peut être décisif quant à la nature plurielle ou non de son sujet, je ne l'ai pas considéré comme un élément pertinent pour y baser mon analyse.

Ainsi, si l'on considère la définitude comme un trait du pluriel, l'on peut qualifier les autres formes qui présentent une opposition formelle liée avec une interprétation de pluralité comme des « collectifs ».

Deuxièmement, nous pouvons faire l'hypothèse du marquage local du pluriel. D'après Tiersma (1982), l'occurrence des noms doublement pluralisés est due à la sous-analyse (*under-analysis*) du locuteur du fait du « marquage local » de ce mot. C'est-à-dire que, à cause de la fréquence plus élevée par exemple de la forme en *-at*, un locuteur naïf est susceptible de la sous-analyser comme un singulier. Contrairement au cas cité par Tiersma, ce qui se passe ici est que le nom singulier de base ne disparaît pas de la circulation. De plus, la justification de l'existence du « marquage local », basé sur le fait que pour certains mots l'emploi au pluriel est plus fréquent qu'au singulier, par exemple *les bottes vs la botte*, n'est pas toujours aussi convainquante pour les exemples du persan.

Il me semble que dans les exemples du persan la forme du pluriel joue un rôle : moins la relation formelle entre les deux mots est « transparente » (*Cf. Dressler et al. 1987*), plus il y a de chance qu'une telle sous-analyse se produise : dans la plupart des cas de pluriel double, les relations ne sont pas de simples « affixations » (*Cf. de la*

SFM 13 à la **SMF 49**) et le nombre de paires de mots qui représente l'opposition est assez restreint dans un lexique mental donné.

Ces hypothèses ne sont pas mutuellement exclusives, et la vérité et la pertinence de chacune dépendent des mots et des contextes de leur emploi.

2.5 La variation

[I]n order to talk about variability, it is important to realize that we must make a very fundamental assumption, and that is the assumption of sameness. (Singh 1996 : 20)

En général, les différentes formes dont nous avons discutées sont présentées comme différentes « marques du pluriel »⁷², et, étant donné la disponibilité d'un ensemble de formes, la question de la variation est posée. Pour pouvoir discuter de la question de la variation, nous devons nous assurer tout d'abord que les 'candidats' en question sont au moins identiques à un niveau d'analyse donné. Or, nous avons vu dans la section 1.4., ainsi que dans celles de 2.3. et de 2.4., que les différentes formes regroupées sous l'étiquette du pluriel nominal sont de nature différentes.

Il me semble qu'il existe un *continuum* dont un extrême comprend les formes *-h(a)*, *-in* et *-an*, qui, leurs différences stylistiques mises à part, partagent l'interprétation référentielle et engendrent forcément le trait [+défini], et dont l'autre extrême comprend les formes *-dʒat*, *-iat* et *-iun*, qui ne sont pas des pluriels mais des collectifs, avec une interprétation générique plutôt que référentielle de type « l'ensemble dont le caractère est ... » (N. B. : cette interprétation n'est pas identique pour les trois).

⁷² *ne|aneha-je dʒam?*

Somme toute, les formes en *-an* et *-in* versus leurs contreparties en *-(h)a* et les formes en *-a* versus leurs contreparties en *-ha* peuvent être qualifiées comme des variantes sociolinguistiquement parlant.

Les formes en *-at* et les formes « brisées » se trouvent entre ces deux extrêmes, car elles peuvent – notamment dans la variété formelle – engendrer la définitude et être employées comme simples variantes stylistiques de leur contrepartie en *-(h)a*. Mais, comme nous l’avons vu, la définitude chez ces formes n’est pas une caractéristique nécessaire et suffisante. Ainsi, la problématique présentée quant aux pluriels doubles est également valide quant à la question de la variation.

Comme observation générale, ces formes sont considérées comme plus prestigieuses en comparaison de leurs contreparties en *-(h)a*, néanmoins leur emploi – surtout quant aux formes « brisées » – est risqué lorsqu’un locuteur ne connaît pas le mot correspondant (lorsqu’il ne l’a pas dans son lexique ou l’a momentanément oublié). Car, il n’est pas toujours possible de former de tels mots avec la seule aide de la compétence morphologique, étant donné qu’il existe des stratégies concurrentes qui fournissent le locuteur avec plus d’un mot possible dont le choix est arbitraire : Cf., entre autres, **SFM 41, 42 et 43** ; **SFM 20, 21 et 22** ; **SFM 43 et 44** ; etc.

A titre d’exemple, si l’on a oublié que *hokkam* est le pluriel de *hakem*, l’on ne peut pas décider si *hokama* est le pluriel de *hakem* (Cf. **SFM 43**) ou de *hakim* (Cf. **SFM 44**).

SFM 43 : $/(C1)aC2eC3/_{N.sg} \leftrightarrow /(C1)oC2aC3a/_{N.collectif}$

SFM 44 : $/(C1)aC2iC3/_{N.sg} \leftrightarrow /(C1)oC2aC3a/_{N.collectif}$

De plus, il est plausible que la stratégie nécessaire ne soit pas accessible au locuteur – car les paires de mots pertinents ne font pas partie de son lexique. (N. B. : l’exigence de deux paires de mot pour établir une stratégie est une condition nécessaire mais pas suffisante).

Il faut noter que, du fait de leurs biographies particulières, la compétence morphologique des individus diffère, surtout en ce qui concerne la langue formelle ;

les mots [+formels] sont appris tardivement, au cours de l'apprentissage de la langue et, souvent, à travers une éducation poussée.

Il est important de garder à l'esprit que, lorsqu'un locuteur connaît un mot, c'est-à-dire lorsque ce mot se trouve dans son lexique, il n'utilise/ne mobilise pas la morphologie dans le but de produire le mot. La morphologie n'est là que pour aider à retrouver un mot temporairement oublié ou comprendre/produire un mot nouveau.

Conclusion

Ce travail s'est donné comme tâche la description morphologique du pluriel nominal en persan et, pour ce faire, j'ai opté pour le modèle WWM. En guise de conclusion, je présente tout d'abord une appréciation générale quant à ce choix théorique et ensuite ma motivation pour prendre un tel morceau de la morphologie du persan comme cible de la description morphologique.

Le modèle WWM est un modèle génératif et universel qui décrit la compétence morphologique d'un locuteur d'une langue donnée. Par universel, je veux dire que le modèle peut rendre compte de tous types d'opérations morphologiques par le même principe – en ce qui concerne ce travail, la morphologie concaténative et non-concaténative ou gabaritique et non-gabaritique ; par ailleurs, la liste des SFM montre qu'un tel sous-regroupement ne se tient pas étant donné que parmi les SFM qui correspondent à l'« affixe *-at* » (Cf. SFM 12 à 16), il existe des cas mieux qualifiés comme « non-concaténatifs » qu'« affixaux » (Cf. SFM 15 et 16).

La compétence morphologique est conçue comme la capacité de généralisation à partir des relations présentes dans les ensembles de mots, qui sont disponibles aux locuteurs en entier. C'est pour cela que le modèle réussit également à expliquer (et même à prévoir) les phénomènes « externes » liés aux erreurs de parole, de la variation, ainsi qu'au changement : interne ou externe, *i.e.* dû au contact entre les langues. A titre d'exemple :

En anglais, l'existence des stratégies de formations du passé concurrentes prédit que, pour un locuteur en apprentissage, un enfant, le passé de *bring* soit **brang* au lieu de *brought*, étant donné *sing/sang*, *ring/rang*, *sink/sank*, etc.

En français, un apprenant, qui prononce l'énoncé *j'ai *ouvré la porte*, forme le participe passé conformément à la stratégie à la base de *(je) ferme : fermé*, *(je) parle :*

parlé, avant qu'on le corrige et qu'il apprenne donc que *ouvrir* suit la stratégie à la base de *souffrir* : *souffert*, *couvrir* : *couvert*.

En persan, dans la période de l'influence de l'arabe, l'émergence de mots tels que *basatin* (pl. de *bostan* 'jardin') s'explique par une généralisation locale une fois que les mots tels que *soltan/salatin*, *ejtan/ajatin*, etc. sont adaptés par les locuteurs.

De plus, ce modèle permet une adéquation descriptive et explicative, avec une machinerie que l'on peut juger comme étant la plus simple disponible. Il repose simplement sur le répertoire des mots entiers⁷³ ainsi qu'un seul type de processus formel ; en effet, tout ce qui peut être dit en morphologie est résumé par :

$$/X/\alpha \leftrightarrow /X'/\beta$$

L'idée de s'attaquer à ce morceau de la morphologie du persan relève d'un double intérêt personnel pour la morphologie aussi bien que pour la « sociolinguistique » (dans le sillage du regard critique qu'a porté Singh (1996) sur cette discipline). Le pluriel nominal, tel qu'il est présenté dans la littérature, engendre des réflexions sociolinguistiques au niveau de la variation ainsi qu'au sujet de la question des contacts entre les langues – en l'occurrence le persan avec l'arabe depuis la conquête de l'Iran par les Arabes au XII^e siècle.

D'un côté, un ensemble – non homogène – de formes est présenté comme des variantes de la marque du pluriel dont le choix dépend de considérations sociolinguistiques comme : le registre, le niveau de l'éducation du locuteur, etc. (Cf. Ghomeshi 2003 : 57). Mais comment peut-on parler de la variation avant de s'assurer qu'il s'agit de « la même chose » (Cf. Singh 1996 : 20) ? Et, en effet, les

⁷³ En se basant entre autres sur l'intervention de Jackendoff (1975), l'on peut écarter une objection qui juge cela contre le principe d'économie.

observations faites dans ce travail démontrent que traiter ces « marques du pluriel » comme des variantes n'est pas justifié.

D'un autre côté, les « marques de pluriel » sont presque toujours présentées comme étant accompagnées de remarques sur « leur origine » persane ou arabe. J'ai discuté ce point dans la section 1.6 et ai argumenté contre de telles considérations. La position défendue ici est que la morphologie n'est pas empruntable, seuls les mots le sont (*Cf.*, entre autres, Singh 1996), et que les traits qui font allusion à l'origine des mots doivent être complètement exclus d'une description synchronique.

Références

- Anderson S. R., (1982)**, Where's morphology?, *Linguistic Inquiry*, **13**, pp. 571–612.
- Anderson S. R. (1988)**, Morphological Theory, in Newmeyer F. (ed.), *Linguistics: The Cambridge Survey*, vol. 1, Cambridge: Cambridge University Press, pp. 146-191.
- Anderson S. R. (1992)**, *A-Morphous Morphology*, Cambridge: Cambridge University Press.
- Aronoff M. (1976)**, *Word Formation in Generative Grammar*, Cambridge (Mass.): MIT Press.
- Bateni M. R. (1969)**, Towsif-e sâxtemân-e dasturi-e zabân-e fârsi [Une description de la structure grammaticale du persan], Tehran: Amir Kabir Ed.
- Bloomfield, L. (1926)**, A Set of Postulates for the Science of Language, *Language*, **2**, pp. 153–64
- Bloomfield L. (1933[1984])**, *Language*, Chicago: University of Chicago Press.
- Carstairs-McCarthy A. (1992)**, *Current Morphology*, London: Routledge.
- Chomsky N., Halle, M. (1968)**, *The Sound pattern of English*, New York: Harper and Row.
- Chomsky N. (1965)**, *Aspects of the Theory of Syntax*, Cambridge (Mass.): MIT Press.
- Chomsky N. (1969)**, *La linguistique cartésienne*, Paris : Seuil.
- Chomsky N. (1970)**, Remarks on Nominalization, in Jacobs R. A. and Rosenbaum P. S. (eds.), *Readings in Transformational Grammar*, Waltham (Mass.): Ginn and co, pp. 184-221
- Chomsky N. (1986)**, *Knowledge of language: Its nature, origin, and use*, New York: Praeger.
- Corbett G. (2000)**, *Number*, Cambridge: Cambridge University Press.
- Dinmohammadi Gh. R. (1996)**, Jâygâh-e hamze dar sâxt-e âvâi-e zabân-e fârsi bar pâye-ye nazarie-ye farâ-zanjirei [La place du hamze dans la structure phonologique du persan selon la théorie suprasegmentale], *Zabân va adabiat; pajûheš zabânhâ-ye xâreji*, **3**, pp. 127-140.
- Di Sciullo A.-M., Williams E. (1987)**, *On the definition of word*, Cambridge (Mass.): MIT Press.

- Dressler W. U., Mayerthaler W., Panagl O., Wurzel W. U. (1987)**, Leitmotifs in natural morphology, Amsterdam: Benjamins.
- Dubé J. (2008)**, Review of “Explorations in Seamless Morphology”, Singh S. and Starosta S. (eds), *Indian Linguistics*, **69**, pp. 483-496.
- Farshidvard K. (2003[1969])**, Arabi dar fârsi [L’arabe en persan’], Tehran: Tehran University Publications.
- Ferguson Ch. (1959)**, Diglossia, *Word*, **15**, pp. 325-40.
- Ford A., Singh R. (1991)**, Propédeutique morphologique, *Folia Linguistica*, **25**, pp. 549-575.
- Ford A., Singh R., Martohardjono G. (1997)**, Pace Panini, New York: Peter Lang.
- Ghomeshi J. (2003)**, Plural Marking, Indefiniteness, and the Noun Phrase, *Studia Linguistica*, **57(2)**, pp. 47-74.
- Haghshenas A. M. (1991)**, Naqshâ-ye dogâne-ye hamze dar sâxt-e âvâi-ye zabân-e fârsi [Les rôles duals du hamze dans la structure phonologique de la langue persane], in Maqâlât-e adabi-zabânšenâxti, Tehran: Nilufar, pp. 257-284.
- Haghshenas A. M. (1995)**, Hamze dar zabân va xat-e fâris [Le hamze dans la langue et l’écriture persane], *Zabân va adabiât*, **9**, pp. 23-36.
- Haghshenas A. M. (1997[1977])**, Phonetics, Tehran: Âgâh Ed.
- Halle M., Marantz A. (1993)**, Distributed Morphology and the Pieces of Inflection, in Ken Hale K. and Keyser S. (eds.), *The View from Building 20*, Cambridge (Mass.): The MIT Press, pp. 111-176.
- Halle M. (1962)**, Phonology in Generative Grammar, *Word*, **18**, pp. 54-72.
- Halle M. (1973)**, Prolegomena to a Theory of Word-formation, *Linguistic Inquiry*, **4**, pp. 3-16.
- Harris Z. S. (1942)**, Morpheme Alternants in Linguistic Analysis, *Language*, **18**, pp. 169-180.
- Haspelmath M. (2002)**, Understanding Morphology, London: Arnold.
- Hockett Ch. (1947)**, Problems of Morphemic Analysis, *Language*, **23**, pp. 321-343.
- Hockett Ch. (1954)**, Two Models of Grammatical Description, *Word*, **10**, pp. 210-231.
- Hooper J. B. (1976)**, An Introduction to Natural Generative Phonology, New York: Academic Press.
- Jackendoff R. (1972)**, Semantic Interpretation in Generative Grammar, Cambridge (Mass.): MIT Press.
- Jackendoff R. (1975)**, Morphological and Semantic Regularities in the Lexicon, *Language*, **51**, pp. 639-671.

- Jahani C. (2005)**, The Glottal Plosive: A Phoneme in Spoken Modern Persian or Not?, in Csato E. V., Bo I., Jahani C. (ed.), *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, Cuzon: Routledge, pp. 79–96.
- Joos M. (ed.) (1958)**, *Readings in Linguistics*, Washington: American Council of Learned Societies.
- Kalbasi I. (2008[1993])**, *The Derivational Structure of Word in Modern Persian*, Tehran: Institute for Humanities and cultural studies.
- Karimi-Doostan Gh. (2004)**, Gunehâ-ye nešânehâ-ye jam' -ât va -ân dar fârsi [Les variantes des marques du pluriel -ât et -ân en persan], *Journal of social and human sciences*, **21(1)**, Shiraz: University of Shiraz, pp. 29-40.
- Khanlari P. Natel (1972)**, *Dastur-e zabân-e fârsi [La grammaire du persan]*, Tehran: Bonyâd-e farhang-e Iran.
- Khanlari P. Natel (1986[1969])**, *Tarix-e zaban-e fârsi*, vol. 3, Téhéran: Nashr-e now.
- Khanlari P. Natel (1952)**, Review of « qâ'edehâ-ye jam' dar zaban-e farsi », [Les livres récents : Les règles du pluriel en persan], *Soxan*, **4(1)**, Tehran, pp. 71-73.
- Kiparsky P. (1982)**, Lexical Morphology and Phonology, in In-Seok Yang (ed.), *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, Seoul: Hanshin Pub. Co, pp. 3-91.
- Kuryłowicz J. (1966)**, La nature des procès dits “analogiques”, in Hamp E. P., Householder F. W., Austerlitz R. P. and Joos M. (eds), *Readings in Linguistics II*, Chicago: University of Chicago Press, pp. 158-174. *Reprint/Original edition: Kuryłowicz J. (1949)*, La nature des procès dits “analogiques”, *Acta Linguistica*, **5**, pp. 121-138.
- Lazard G. (2006[1957])**, *Grammaire du persan contemporain*, Tehran: Institut Français de Recherches en Iran et Nashr-e dâresh.
- Mahootian Sh. (1997)**, *Persian*, London: Routledge.
- Majidi M. R. (1986)**, *Strukturelle Grammatik des Neupersischen (Fârsi): Band I: Phonologie*, Hamburg: Buske Verlag.
- Majidi M. R. (1986)**, *Strukturelle Grammatik des Neupersischen (Fârsi): Band 2: Phonologie*, Hamburg: Buske Verlag.
- Matthews P. H. (1972)**, *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*, Cambridge: Cambridge University Press.
- Matthews P. H. (1974)**, *Morphology*, Cambridge: Cambridge University Press.
- Matthews P. H. (1993)**, *Grammatical Theory in the United States: from Bloomfield to Chomsky*, Cambridge: Cambridge University Press.
- Martinet A. (1960)**, *Éléments de linguistique générale*, Paris : Armand Colin.

- Meshkatod Dini M. (1990)**, Introduction to Persian Transformational Syntax, Mashhad: Ferdowsi University Ed.
- Meshkatod Dini M. (1995)**, Sound Patterns of Language: An Introduction to Generative Phonology, Mashhad: Ferdowsi University Ed..
- Meshkatod Dini M. (2008[2005])**, Persian Grammar: The Lexical Categories and Merge, Tehran: SAMT.
- Mo'in M. (1952)**, Qâedehâ-ye jam' dar zabân-e fârsi [Les règles du pluriel en persan], in Tarh-e dasturi-e zabân-e fârsi [An outline of Persian Grammar], Téhéran : Ed. inconnu.
- Mo'in M. (1961)**, Mofrad va jam' [Le singulier et le pluriel], in Tarh-e dasturi-e zabân-e fârsi [An outline of Persian grammar], Téhéran : Ed. Inconnu.
- Nafisi S. (1965)**, Qalathâ-ye mašhur-e emlâi va zabân-e fârsi [Les erreurs d'orthographe connues et la langue persane], extrait de *Dar mahzar-eostâd*, publié sur le site de *šorâ-ye gostareš-e zabân va adabiât-e fârsi* [Conseil du développement de la langue et la littérature persane], www.persian-language.org
- Najafi A. (1981)**, Qalat nanevisim [N'écrivons pas faux], Tehran: Centre de Presse Universitaire.
- Perry J. R. (1991)**, Form and Meaning in Persian Vocabulary: The Arabic Feminine Ending, *Bibliotheca Persica Persian Studies Series*, 12. Costa Mesa: Mazda Publishers.
- Perry J. R. (2005)**, Lexical Areas and Semantic Fields of Arabic Loanwords in Persian and beyond, in Csato E. V., Bo I., Jahani C. (eds.), *Linguistic convergence and areal diffusion: Case studeis from Iranian, Semitic and Turkic*, Cuzon: Routledge, pp. 97-111
- Robins R. H. (1959)**, In Defence of WP, *Transactions of the Philological Society*, pp. 116-144.
- Sadeghi A. A. (1969)**, Alâmathâ-ye jam' dar farsi-ye moâser [Les marques du pluriels en persan contemporain], *Soxan*, 19, Tehran, pp. 777-786.
- Sadeghi A. A., Arjang Gh. R. (1976)**, Dastur [La grammaire], Tehran: Sâzmân nashr-e ketâbhâ-ye darsi Ed.
- Sadeghi A. A. (2001)**, Masâel târîxi-ye zabân-e fârsi [Les problèmes historiques de la langue persane], Tehran: Soxan Ed.
- Samare Y. (1999 [1985])**, Âvâšenâsi-ye zabân-e fârsi [Phonologie du persan], Tehran: Centre de presse universitaire.
- Samvelian P. A. (2007)**, (Phrasal) Affix Analysis of the Persian, *Ezafé: Journal of Linguistics*, 43, pp. 605-645.
- Saussure F. (de) (1968[1916])**, Cours de linguistique générale, Paris: Payot.

- Scalise S., Guevara E. (2005)**, The Lexicalist Approach to Word-formation and the Notion of Lexicon, in Stekauer P., Lieber R. (eds), *Hand book of word-Formation*, Dordrecht: Springer, pp. 147-186
- Sepanta S. (1973)**, Hamze dar zabân-e fârsi [Le hamze dans la langue persane], *Zaban va adabiat: dâneškade-ye adabiat va olum ensâni (dânešgâh esfehân)*, **9(1)**, pp. 58-70.
- Singh R., Agnihotri R. K. (1997)**, *Modern Hindi Morphology*, Delhi: Motilal Banarasidass.
- Singh R. (1995)**, *Linguistic Theory: Language Contact, and Modern Hindustani*, New York: Peter Lang.
- Singh R. (1996)**, *Lectures against Sociolinguistics*, New York: Peter Lang.
- Singh R. (2001a)**, Chasing Butterflies in a Multilingual Garden: Grounding Linguistics in Multilingualism, *Conférence donnée à l'Université de Delhi*, 12 janvier.
- Singh R. (2001b)**, Morphological Diversity and Morphological Borrowing in South Asia, in Peri Bhaskararao and Karumuri Venkata Subbarao (eds.), *The Yearbook of South Asian Languages and Linguistics*, Thousand Oaks: Sage, pp. 349-368.
- Singh, R. (2006)**, Whole Word Morphology, in Keith B. (ed.), *Encyclopedia of Language & Linguistics*, Oxford: Elsevier, pp. 578-579.
- Stampe D. (1972)**, *A Dissertation on Natural Phonology*, New York: Garland.
- Stump G. T. (1998)**, Inflection, in A. Spencer and A. M. Zwicky, (eds.), *The Handbook of Morphology*, Oxford & Malden, MA: Blackwell, pp.13-43.
- Tiersma P. M. (1982)**, Local and General Markedness, *Language*, Vol. 58, **4**, pp. 832-849.
- Troubetzkoy N. (1929)**, Sur la 'morphologie', *TCLP*, **1**, pp. 85-88.
- Troubetzkoy N. (1949)**, *Principes de phonologie*, Paris : Klincksieck.
- Windfuhr G. L. (1997)**, Persian Phonology, in Kaye A. S. (ed.), *Phonologies of Asia and Africa 2*, Winona Lake: Eisenbrauns, pp. 675-689.