

INSTITUT GASPARD MONGE, UNIVERSITÉ DE MARNE-LA-VALLÉE

Système électronique de lexiques coréens *DECO*

Mémoire présenté en vue de
l'obtention du diplôme d'habilitation à diriger des recherches
en *Linguistique Informatique*

1997

NAM Jee Sun

Vol I. <i>Mémoire</i>

INSTITUT GASPARD MONGE, UNIVERSITÉ DE MARNE-LA-VALLÉE

Système électronique de lexiques coréens *DECO*

Mémoire présenté en vue de
l'obtention du diplôme d'habilitation à diriger des recherches
en *Linguistique Informatique*

1997

NAM Jee Sun

Vol I. <i>Mémoire</i>

SYSTÈME ÉLECTRONIQUE DE LEXIQUES CORÉENS

Mémoire présenté en vue d'obtenir le diplôme d'habilitation à diriger des recherches

Dans ce mémoire, les travaux que nous avons publiés postérieurement à notre thèse sont résumés et rassemblés. Nos travaux sont particulièrement consacrés à l'élaboration d'un dictionnaire électronique du coréen, nécessaire à tous les programmes informatiques du traitement automatique de textes écrits, ils comportent souvent de longues listes de mots, codés et classés par sous-ensembles. En raison de l'importance du nombre des lexiques établis de façon exhaustive et modulée, ils se présentent, la plupart du temps, sous forme de rapports techniques du laboratoire (LADL et IGM).

TABLE DES MATIÈRES

Vol. I. *Mémoires*

PRINCIPES DE TRANSCRIPTION

PRINCIPES DE NOTATION

0. Introduction

- 0.1. Dictionnaire électronique vs. dictionnaire éditorial
- 0.2. Unités de base dans le traitement automatique du coréen
- 0.3. Organisation du système lexical *DECO*

1. Dictionnaire électronique des mots simples *DECOS*

- 1.1. Lexique des noms *DECOS-NS*
 - 1.1.1. Constitution du lexique des noms simples
 - 1.1.2. Informations associées aux noms et codage
- 1.2. Lexique des adjectifs *DECOS-AS*
 - 1.2.1. Constitution du lexique des adjectifs
 - 1.2.2. Extension à d'autres données
 - 1.2.2. Informations associées aux adjectifs et codage
- 1.3. Lexique des verbes *DECOS-VS*
 - 1.3.1. Problèmes liés aux séquences *N-V*
 - 1.3.2. Constitution du lexique des verbes
 - 1.3.3. Informations associées aux verbes et codage
- 1.4. Lexique des mots invariables *DECOS-IS*

2. Dictionnaire électronique des formes fléchies *DECOF*

- 2.1. Système de suffixes de conjugaison et de postpositions
- 2.2. Lexique des postpositions *DECO-POST*
 - 2.2.1. Estimation des effectifs des formes fléchies
 - 2.2.2. Lexique des N-Postpositions *DECO-PostN*
 - 2.2.3. Lexiques des A-Postpositions et des V-Postpositions *DECO-PostA/PostV*
- 2.3. Dictionnaire des formes fléchies *DECOF*

3. Dictionnaire électronique des formes complexes *DECOC*

- 3.1. Noms affixés et noms composés
- 3.2. Typologie des structures de noms
- 3.3. Définition des catégories grammaticales
 - 3.3.1. Déterminant (*Dét*) vs. *Adj-Sd / V-Sd*

- 3.3.2. Déterminant (*Dét*) vs. Préfixe (*PF*)
- 3.3.3. Préfixe (*PF*) vs. Nom (*N*)
- 3.3.4. Suffixe (*SF*) vs. Nom (*N*)
- 3.3.5. Nom incomplet (*NI*) vs. Suffixe (*SF*)
- 3.3.6. Numéral (*Num*) vs. Déterminant (*Dét*)
- 3.4. Lexique des affixes *DECO-AFX*
- 3.5. Typologie des noms composées
- 3.6. Noms composées de type [*Adj-Sd*] *N* et [*V-Sd*] *N*
- 3.7. Noms composées de type *NN*
 - 3.7.1. Analyse structurale des *NN*
 - 3.7.2. Formes comportant un *PN*

4. Analyseur lexical et problèmes d'ambiguïtés

- 4.1. Construction des grammaires locales : les *Numéraux*
- 4.2. Grammaires locales d'expressions horaires
- 4.3. Séquences nominales comportant des *Dnum*
 - 4.3.1. Structure *Dnum Nmes-Gén N*
 - 4.3.2. Structure *N Dnum Nmes*
 - 4.3.3. Structure *Dnum N*

5. Analyseur syntaxique et Lexique-grammaire

- 5.1. Description syntaxique des éléments lexicaux
- 5.2. Représentation par matrices binaires
- 5.3. Construction d'un dictionnaire syntaxique du coréen

REFERENCES

Vol. II. Annexes

I. Lexique des mots simples *DECOS*

(*DECOS-NS* / *DECOS-AS* / *DECOS-VS* / *DECOS-NFA-M* / *DECOS-NFV-M*)

II. Lexique des affixes *DECO-AFX*

(*DECO-PF* / *DECO-SF* / *DECO-PN*)

III. Lexique des postpositions *DECO-POST*

(*DECO-PostN* / *DECO-PostA* / *DECO-PostV*)

Vol. III. Sélection de travaux

LISTE DE PUBLICATION

SELECTION DE TRAVAUX

PRINCIPES DE TRANSCRIPTION

Les 14 consonnes et 10 voyelles simples du coréen ont été transcrites phonétiquement de la façon suivante :

Consonnes

ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅇ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅎ
g	n	d	l	m	b	s	ng	j	ch	k	t	p	h

Voyelles

ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ
a	ya	e	ye	o	yo	u	yu	eu	i

et les consonnes et voyelles complexes :

Consonnes complexes

ㄲ	ㄸ	ㅃ	ㅆ	ㅉ
gg	dd	bb	ss	jj

Voyelles complexes

ㅐ	ㅒ	ㅖ	ㅘ	ㅙ	ㅚ	ㅜ	ㅠ	ㅝ	ㅞ
ai	yai	ei	yei	oi	wai	wei	wi	eui	wa

Écriture syllabique

L'écriture coréenne est alphabétiquement combinatoire. Un mot est une suite de syllabes, une syllabe étant une unité constituée d'une voyelle et plus d'une consonne. Cependant, on notera une particularité du système d'écriture du coréen. Une ligne est une architecture de boîtes (invisibles) qui contiennent, chacune, une syllabe. Soit :

S	S	S	S	S
---	---	---	---	---

Chaque boîte syllabique a une structure intérieure déterminée par le type de la voyelle et la présence d'une deuxième consonne.

L'absence phonétique de la première consonne ne conditionne pas la forme architecturale de la syllabe, puisqu'une consonne muette "o" doit être présente automatiquement, quand il n'y a pas de consonne prononcée en tête de la syllabe. Les voyelles *o*, *yo*, *u*, *yu*, *eu* se situent en bas de la boîte, alors que les autres voyelles simples se trouvent à droite de la consonne. Soit :

vs.

우	/u/	부	/bu/
아	/a/	다	/da/

Quand on a une deuxième consonne, cette consonne se trouve toujours en bas de la boîte. Parallèlement aux exemples précédents, on a :

음	/um/	bum	/bum/
알	/al/	달	/dal/

Par exemple, une phrase :

hai journée - ga nmtf gil long - da St
(La journée est longue)

s'écrit en quatre boîtes de forme :

해	가	길	다
--------------	--------------	--------------	--------------

PRINCIPES DE NOTATION

<i>Acc</i>	Postposition du complément <i>accusatif</i>
<i>Adj</i>	Adjectif
<i>Adv</i>	Adverbe
<i>E</i>	Séquence vide
<i>Gén</i>	Postposition du <i>génitif</i>
<i>Mnom</i>	Suffixe (ou morphème) de <i>nominalisation</i>
<i>Mpas</i>	Suffixe (ou morphème) de <i>passé</i>
<i>Mpl</i>	Suffixe (ou morphème) de <i>pluriel</i>
<i>N</i>	Substantif
<i>nmtf</i>	Postposition du <i>nominatif</i>
<i>Postp</i>	Postposition d'un complément <i>non-accusatif</i>
<i>Pro</i>	Pronom
<i>Sd</i>	Suffixe <i>déterminatif</i>
<i>St</i>	Suffixe <i>terminal</i> du mode <i>déclaratif</i>
<i>Sti</i>	Suffixe <i>terminal</i> du mode <i>interrogatif</i>
<i>V</i>	Verbe
<i>Vsup</i>	Verbe <i>support</i>
<i>W</i>	Complément circonstanciel ou adverbe quelconque
<i>*</i>	Phrase inacceptable
<i>?, ??</i>	Phrase dont l'acceptabilité est douteuse

0.

Introduction

0.1. Dictionnaire électronique vs. dictionnaire éditorial

L'ensemble des travaux que nous présentons ici consiste à établir une couverture lexicale de façon systématique et exhaustive, pour traiter automatiquement des textes écrits en coréen. Il est certain que « les dictionnaires électroniques entrent dans la plupart des programmes qui mettent en jeu une procédure de reconnaissance d'unités linguistiques significatives. Dans de tels programmes, un texte est, dans une phase initiale, soumis à une procédure de segmentation en unités graphiques (les mots). Un dictionnaire est alors consulté afin de déterminer la nature de chacune de ces unités. » [Gro89]

La notion de dictionnaire électronique ne doit pas être confondue avec celle de dictionnaire usuel disponible sur support informatique [Gro89], [Sil93]. il est facile de vérifier que même si les listes de mots existent sous forme électronique, les informations que comportent des dictionnaires usuels n'ont aucun rapport avec celles des dictionnaires électroniques : étant donné que les dictionnaires électroniques sont conçus pour l'utilisation dans divers traitements automatiques des langues naturelles, ils ne contiennent que des codes informatiques, alors que les dictionnaires usuels se présentent sous forme de textes en langue naturelle. Par ailleurs, les entrées établies dans ces deux types de dictionnaires ne se superposent que partiellement. Une des raisons qui expliquent cette situation est que les dictionnaires usuels ont tendance à ne pas donner d'informations trop évidentes pour l'utilisateur: en français, par exemple, des adjectifs en *-able* dérivés de verbes, des adjectifs en *-ant* ou des adverbes en *-ment* ne sont pas tous répertoriés dans les dictionnaires usuels. La situation est comparable concernant le coréen : des noms en -기 *-gi* ou -음 *-eum* dérivés de verbes ou d'adjectifs, ou des verbes en -어하다 *-ehada* dérivés d'adjectifs¹ ne sont pas systématiquement représentés.

Cependant, la construction d'un dictionnaire électronique ne se limite pas au seul établissement des listes de mots. Un tel dictionnaire doit également comporter des

¹ Notons que, d'ailleurs, nous ne considérons pas comme formes dérivées des verbes en -어하다 *-ehada* : pour nous, ce sont deux unités (*Adj-e* et *hada*) éventuellement soudées. Voir [Nam94b].

informations syntaxiques et sémantiques. Mais la représentation de ces informations n'y prendra pas la même forme que celles dans un dictionnaire éditorial.

En fait, « les différences entre les dictionnaires électroniques et les dictionnaires usuels sont dues au fait que les dictionnaires électroniques sont avant tout construits pour être utilisés par des programmes informatiques qui n'ont aucune connaissance linguistique a priori, tandis que les dictionnaires du commerce ont été construits pour être utilisés par des humains disposant d'un acquis énorme de connaissances.[Sil93] » Les caractéristiques des dictionnaires électroniques peuvent donc être résumées de la manière suivante [Sil93] :

- ♣ les dictionnaires électroniques doivent être aussi **complets** que possible ;
- ♣ les informations données par les dictionnaires électroniques doivent être **explicites** ;
- ♣ les dictionnaires électroniques sont étroitement liés à des programmes de traitement automatiques. Toutes les informations contenues dans les dictionnaires électroniques doivent donc être **cohérentes**, aussi bien du point de vue de la forme des entrées, que du point de vue des informations associées.

0.2. Unités de base dans le traitement automatique du coréen

Dans le traitement automatique des langues naturelles, l'analyse de textes écrits prend une place importante, notamment pour la documentation automatique, la traduction automatique et la correction orthographique. L'analyse automatique est souvent considérée comme composée de trois modules : analyse morphologique (ou lexicale), analyse syntaxique et analyse sémantique. Dans les langues comme le français ou l'anglais, l'unité de base pour l'analyse morphologique correspond globalement à ce qu'on appelle un **mot**, c'est-à-dire à une chaîne de caractères comprise entre deux blancs ou entre un blanc et un signe de ponctuation.

Dans le cas du coréen, l'analyse morphologique ne pourra pas être effectuée directement sur les unités délimitées par ces séparateurs typographiques. Une des raisons principales est que des postpositions grammaticales, correspondant à des prépositions en français, peuvent se souder à des substantifs. Ainsi, aux deux mots français : *à Paul*, correspondra une séquence en coréen : 폴에게 *poleigēi*. Sans s'interroger sur les notions linguistiques en jeu (i.e. morphème, mot ou locution?), on pourrait théoriquement décider de prendre comme unités minimales de l'analyse morphologique les formes délimitées par des séparateurs. Cette solution est adoptée pour le français dans le traitement des séquences verbales. Un verbe canonique peut correspondre à environ 50 formes fléchies au maximum (cf. DELAF)², et ces dernières, enregistrées dans un dictionnaire électronique, sont considérées comme des **mots** au

² Dictionnaire Electronique du Laboratoire d'Automatique Documentaire et Linguistique pour les mots fléchis du français. Voir [Cou89].

même niveau que à ou *Paul*. La phrase *Marie parlait à Paul* est donc composée de 4 de ces unités de base. Les verbes en coréen se comportent de façon similaire. Plusieurs suffixes de conjugaison (temps, aspect ou modalité) s'attachent aux racines verbales pour engendrer des formes fléchies. Il est théoriquement possible de traiter ces formes fléchies comme des unités de base, et si l'on étend cette solution aux substantifs, une séquence nominale comme 폴에게 *poleigēi* serait également une unité minimale. Ainsi, la phrase 마리가 폴에게 말하였다 *maliga poleigēi malhayessda* (Marie-nmtf Paul-à parlait) serait considérée comme ayant 3 composants, et non pas 5 (si on sépare les postpositions 가 *ga* [nmtf: nominatif] et 에게 *eigēi* [Postp: à] des noms) :

Cependant, ce choix est lié au nombre et à la productivité de ces unités de base. Par exemple, le dictionnaire des mots simples du français [DELAS] qui contient 90 000 entrées lexicales donne un dictionnaire de 700 000 formes fléchies [DELAf]. Or, dans le cas du coréen, le nombre des formes fléchies de toutes les parties du discours dépasse 100 millions, ce qui constitue un ensemble d'une taille trop élevée. Notre dictionnaire électronique des items simples du coréen (DECOS) contient 35 000 entrées canoniques, réparties en parties du discours, i.e., *noms*, *verbes*, *adjectifs* et *mots invariables* (nous y reviendrons au paragraphe suivant) : 1 500 postpositions (*Postp*) et 6 000 suffixes de conjugaison (*SC*) peuvent s'y attacher (Cf. [Nam96b]). Plus précisément, les noms simples peuvent prendre dans les textes 22 millions (15 000 x 1 500) de formes différentes, et les verbes canoniques apparaissent sous 45 millions (7 500 x 6 000) de formes fléchies. Par ailleurs, les adjectifs, qui prennent directement des suffixes de conjugaison comme les verbes sans être accompagnés d'une copule de type *être*, donnent 32 millions (5 300 x 6 000) de formes fléchies.

Le problème de la taille ne s'arrête pas ici : étant donné que la formation de noms dérivés et de noms composés est importante en nombre et en complexité, l'effectif des séquences nominales *Nom-Postp* devient rapidement explosif. Quant aux éléments prédicatifs, i.e. verbes et adjectifs, les estimations ci-dessus ne sont pas définitives, car des compléments (c'est-à-dire des noms) peuvent se trouver soudés aux racines verbales ou adjectivales. Les exemples suivants illustrent respectivement *PF-N-Postp*, *N-V-SC* et *N-Adj-SC*³ :

시어머니에게 <i>siemenieigēise</i>	<PF:si-N:emeni-Postp:eigēise =belle-mère-de>
욕먹으면서 <i>yogmegeumyense</i>	<N:yog-V:meg-SC:eumyense = insulte-recev(oir)-ant>
힘세면서도 <i>himseimyensedo</i>	<N:him-Adj:sei-SC:myensedo =force-fort-étant>

Il est donc évident qu'en coréen les unités minimales pour l'analyse morphologique ne peuvent pas être les formes définies au moyen des séparateurs typographiques. De plus, les séquences *N-(V+Adj)-SC* ne peuvent être déterminées que quand des informations syntaxiques et sémantiques sont disponibles de façon systématique.

³ *PF* signifie *préfixe*; *N*, *V* et *Adj* représentent respectivement *nom*, *verbe* et *adjectif*.

Le fait qu'on ne puisse pas prendre comme unités minimales les séquences délimitées par les séparateurs typographiques complique les algorithmes d'analyse lexicale. Chaque fois qu'on rencontre une unité typographique dans un texte, l'analyseur doit consulter un dictionnaire, et donc la segmenter en tenant compte de toutes les entrées disponibles dans le dictionnaire. D'où des problèmes d'ambiguïté plus complexes que dans des langues comme le français ou l'anglais. Supposons qu'on rencontre dans un texte français une séquence comme :

enfermer

Si l'unité typographique n'était pas la forme minimale de l'analyse, l'analyseur devrait la segmenter en tenant compte de toutes les entrées qui se trouvent dans un dictionnaire. Pour cet exemple, on trouverait 6 entrées lexicales (ou 8 solutions, en séparant les deux homographes *en* et *fermer*) dans le dictionnaire (cf. DELAF) :

en .Pron(Ppv) .PREP
fer .N1
mer .N21
enfer .N1
fermer .V3(t) .V3E(se)(i)
enfermer .V3(t)

En conséquence, les 4 analyses seraient présentées comme suit (Figure 1) :

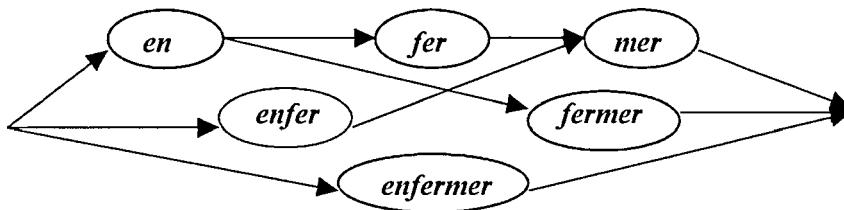


Figure 1

C'est en fait de cette manière que l'analyseur lexical du coréen fonctionne. Si on rencontre, dans un texte coréen, la séquence :

값비싸다 *gabsbissada* [(coûter) cher]

elle doit être segmentée en tenant compte de toutes les entrées disponibles du dictionnaire. Ainsi, on trouvera 5 entrées formellement différentes (*gab*, *bi*, *ssa*, *da*, *bissa*), et au moins 24 unités distinguables sémantiquement :

값 *gabs* (prix), NS,
 비 *bi* (pluie), NS,
 싸 *bi* (proportion), NS,
 다 *bi* (ramasse-poussière), NS,

- ㅂㅣ *bi* (reine), NS, Hum,
- ㅂㅣ *bi* (tort), NS,
- ㅂㅣ *bi* (B), NS,
- ㅂㅣ *bi* (vide), ADJS-R, <*bida*> ,
- ㅂㅣ *bi* (vider), VS-R, TRA, <*bida*>
- ㅂㅣ *bi* (prier), VS-R, TRA, <*bilda*>
- ㅂㅣ *bi* (emprunter), VS-R, TRA, <*bilda*>
- ㅅㅅㅏ *ssa* (envelopper), VS-R, TRA, <*ssada*>
- ㅅㅅㅏ *ssa* (uriner), VS-R, TRA, <*ssada*>
- ㅅㅅㅏ *ssa* (léger), ADJS-R, <*ssada*>
- ㅅㅅㅏ *ssa* (bon marché), ADJS-R, <*ssada*>
- ㄷㅏ *da* (tout), ADV,
- ㄷㅏ *da* (à), N-Postp, Loc,
- ㄷㅏ *da* (/déclaratif/), A-Postp, Dec,
- ㄷㅏ *da* (s'exciter), VS-R, INT, <*dalda*>
- ㄷㅏ *da* (installer), VS-R, TRA, <*dalda*>
- ㄷㅏ *da* (mesurer), VS-R, TRA, <*dalda*>
- ㄷㅏ *da* (sucré), ADJS-R, <*dalda*>
- ㄷㅏ *da* (digne), ADJS-R, <*dabda*>
- ㅂㅣㅅㅅㅏ *bissa* (cher), ADJS-R, <*bissada*>

Le graphe suivant (Figure 2) représente 328 analyses possibles (i.e. $(1 \times 10 \times 4 \times 8) + (1 \times 1 \times 8) = 328$) pour la séquence *ㅂㅣㅅㅅㅏㅂㅣㅅㅅㅏ* *gabsbissada* [(être) cher] :

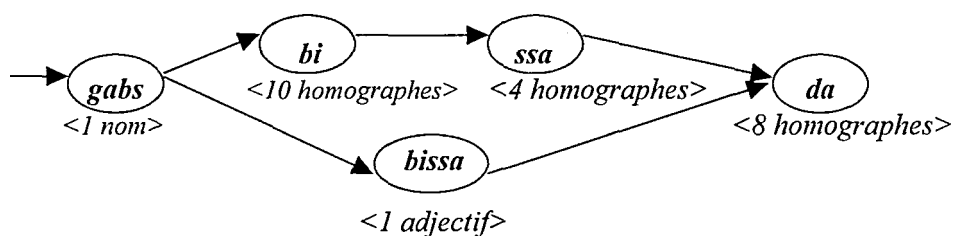


Figure 2

Il est à noter qu'une telle complexité d'analyse n'est pas seulement due au fait que les unités lexicales peuvent être soudées, mais aussi qu'un grand nombre de mots du coréen sont **monosyllabiques** et extrêmement ambigus. Cette caractéristique est liée au fait que, dans le lexique du coréen moderne, il existe beaucoup d'emprunts d'idéogrammes chinois, et que chaque idéogramme (un sens est associé à un caractère) est transcrit selon une syllabe en écriture coréenne. D'où la nécessité, dans le cas du coréen, de construire de façon cohérente des sous-lexiques modulaires à partir desquels on pourra établir un lexique complet qui couvrira toutes les séquences obtenues par des séparateurs typographiques. Une fois ce lexique construit, l'analyseur lexical du coréen ne sera plus équivalent à celui du français ou de l'anglais au même niveau, mais il sera

intermédiaire entre un analyseur lexical et un analyseur syntaxique, puisqu'un nombre considérable d'informations syntaxiques auront déjà été intégrées du système.

0.3. Organisation du système lexical DECO

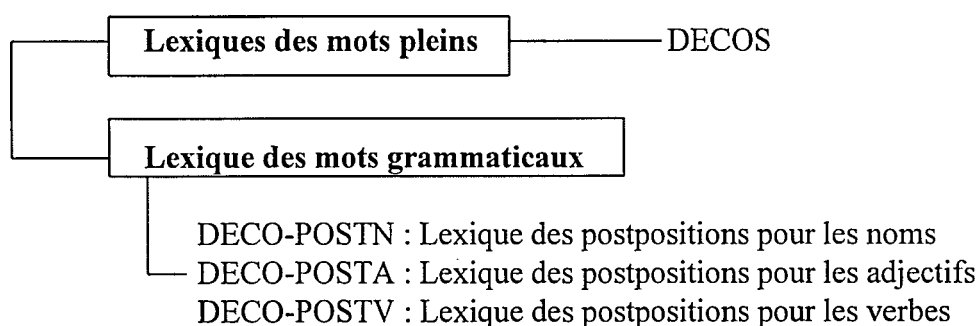
Le système lexical du coréen *DECO* est principalement constitué des trois lexiques suivants, établis de façon modulaire :

DECOS : Lexique des mots *simples*
 DECOF : Lexique des formes *fléchies*
 DECOC : Lexique des formes *complexes*

Ces lexiques comportent plusieurs sous-lexiques, répartis en fonction des parties du discours. Le DECOS est composé de quatre sous-lexiques suivants :

DECOS-NS : Lexique des *noms* simples
 DECOS-VS : Lexique des *verbes* simples
 DECOS-AS : Lexique des *adjectifs* simples
 DECOS-IS : Lexique des mots *invariables*⁴

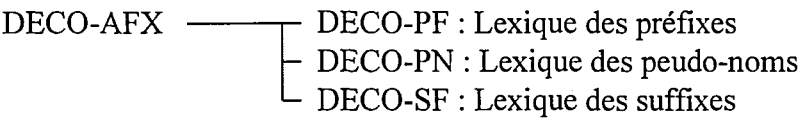
Le DECOF comporte toutes les formes fléchies des mots simples. Ce lexique est obtenu par combinaison à partir du DECOS et des lexiques de mots grammaticaux (DECO-POST) :



Le DECOC est un dictionnaire comportant des entrées *non simples*, ces items sont en deux natures : **mots affixés** et **mots composés**. Une entrée non simple constituée d'un élément autonome (i.e. un mot simple) et d'un ou plusieurs éléments non autonomes (i.e. affixe(s)), est un mot affixé ; une entrée comportant au moins deux éléments autonomes est défini comme un mot composé. Pour établir un lexique des mots affixés, nous avons construit en premier lieu un lexique des affixes (DECO-AFX). L'utilisation

⁴ La version du *DECOS* présentée dans ce travail ne contient pas encore le *DECOS-IS*. Ce lexique sera intégré dans la version ultérieure.

de ce dernier et au DECOS permettra d'établir un lexique des mots affixés. Voici le DECO-AFX :



Nous présentons dans le tableau suivant le système lexical *DECO* (Figure 3). Les parties en boîte grise représentent les lexiques que nous allons discuter en détail dans ce travail.

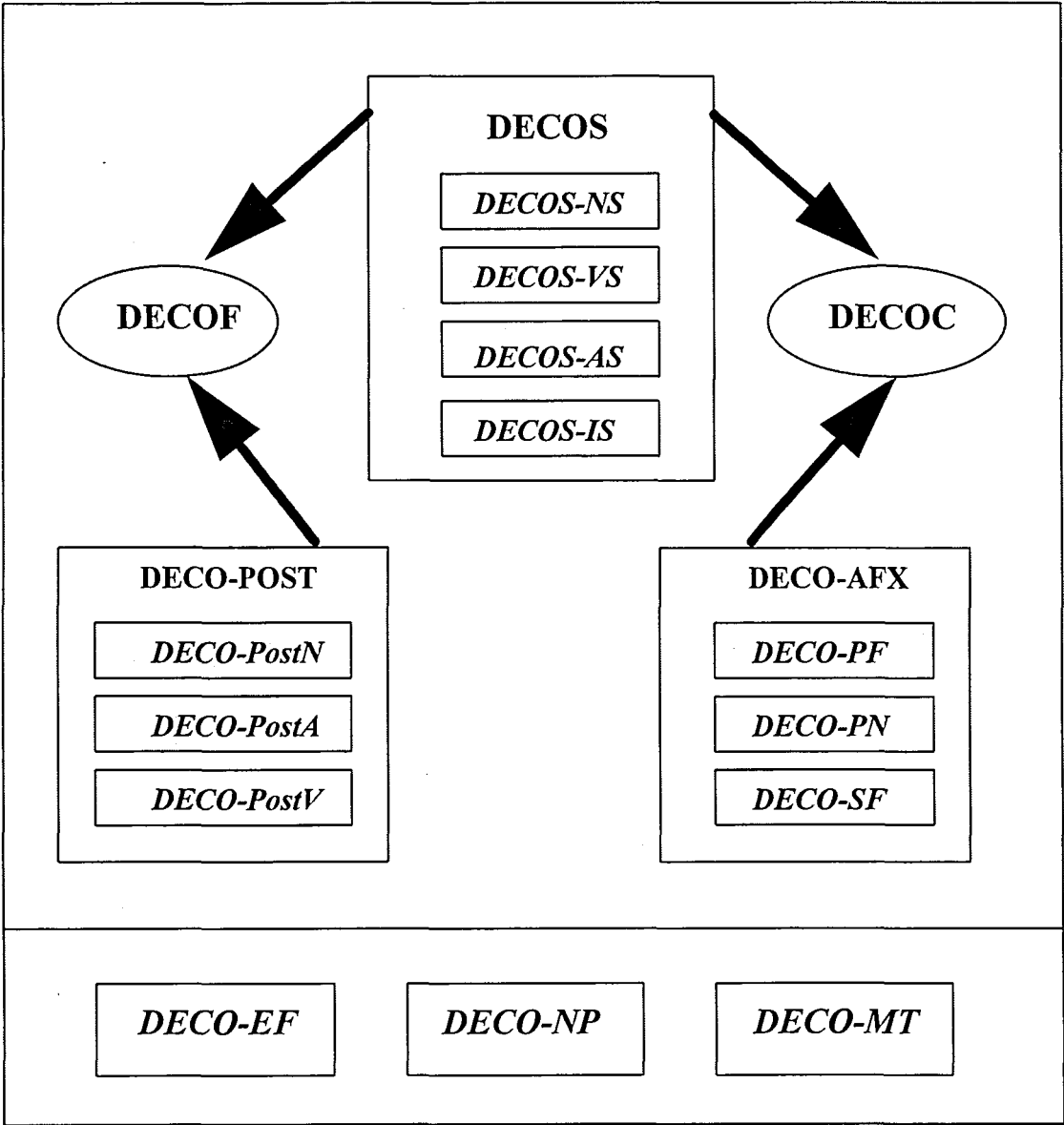


Figure 3. Système lexical du coréen *DECO*

1.

Dictionnaire électronique des mots simples DECOS

1.1. Lexique des noms *DECOS-NS*

1.1.1. Constitution du lexique des noms simples

Le lexique des noms simples DECOS-NS contient 15 000 entrées que nous avons recensées d'abord à partir de grands dictionnaires éditoriaux du coréen (*cf.* [Sin92], [Lee92]). Cependant, les dictionnaires classiques ne sont pas directement utilisables dans un système électronique, faute de principes cohérents et explicites. Voici quelques-uns des principaux problèmes que nous avons rencontrés dans la consultation des dictionnaires éditoriaux.

1.1.1.1. Noms propres

De nombreux noms propres sont intégrés dans des dictionnaires actuels du coréen, alors qu'ils doivent être traités autrement que les noms communs, en raison des particularités de leur syntaxe et de leur productivité. Considérons les exemples suivants:

noms de personnes comme 니콜라이이세 *nikollai isei*, 예이츠 *yeiicheu*,
 noms géographiques comme 노트르담 드 랭스 *noteuleudam deu laingseu*,
 noms d'oeuvres d'art comme 리어왕 *liewang*, 야상곡 *yasanggog*,
 noms d'événements historiques comme 가덕도 해전 *gadeugdo haijen*,
 noms propres divers comme 에밀레종 *eimilei jong*.

Nous les traitons dans les lexiques des noms propres (DECO-NP) et des mots technologiques (DECO-MT).

1.1.1.2. Termes archaïques, techniques et scientifiques

Les termes archaïques, techniques ou scientifiques dont l'emploi est extrêmement réduit couvrent pourtant, dans les dictionnaires usuels, un grand nombre de noms :

Termes archaïques comme 흰노 *hwennyu*, 휘철 *hweichel*,
 Mots dialectaux comme 가개비 *gagaibi*, 아적 *ajeg*,

Noms de plantes comme 닭의 덩굴 *dalgeui denggul*, 월하미인 *welhamiin*,
 Noms d'animaux comme 게르치 *geileuchi*, 더러브렛 *delebeuleis*,
 Termes techniques comme 갠타민 *gaillentamin*, 모넬메탈 *moneilmeital*,
 Emprunts étrangers comme 개프 *gaipeu*, 나딩 *nading*,
 Sigles, symboles et chiffres comme 가네포 *ganeipo*, 언커어크 *enkeekeu*.

Dans notre système, ces termes sont traités dans le lexique des mots technologiques (DECO-MT).

1.1.1.3. Formes non-autonomes

Dans des dictionnaires éditoriaux, on rencontre des entrées de noms qui ne sont pas des unités autonomes, donc qui, en réalité, ne sont pas des noms : cet étiquetage est probablement dû au fait que ces mots sont issus de caractères chinois, et gardent certains traits sémantiques (en général, identifiables par l'analogie avec d'autres mots comportant ces unités), bien qu'ils aient perdu leur autonomie syntaxique. Ce sont, par exemple :

- [i] Des racines adjectivales comme 영리 *yengli*, 예민 *yeimin* ou 모호 *moho*. On les observe, en fait, dans les formes adjectivales 영리하다 *yenglihada* [intelligent], 예민하다 *yeiminhada* [susceptible] ou 모호하다 *mohohada* [ambigu] respectivement.
- [ii] Des unités en 적 *jeg* comme 과학적 *gwahagjeg*, 이성적 *isengjeg* ou 이기적 *igijeg*.
- [iii] Des unités sino-coréennes non-autonomes comme 이복 *ibog*, 공산 *gongsan* ou 청량 *chenglyang*.

Etant donné que ce ne sont pas des mots (i.e. unités autonomes), les séquences comportant les formes [i] et [ii] sont directement traitées dans le lexique des adjectifs (DECO-AS) et les formes [iii] figurent dans le lexique des pseudo-noms (DECO-PN).

1.1.1.4. Formes affixées et composées

Dans les dictionnaires actuels, on observe que les noms affixés (ou dérivés) et composés ne sont pas listés systématiquement. Par exemple, il s'agit de noms comme 시어머니 *siemeni*, 재등록 *jaidéunglog* ou 탈옥 *talog* d'une part, 밤낮 *bamnaj*, 집사람 *jibsalam*, ou 예술사진 *yeisulsajin* d'autre part. Vu que la formation des noms affixés et composés est extrêmement productive en nombre d'entrées, pour compléter leurs listes de façon systématique, il est nécessaire de les séparer des noms simples à l'aide de critères opératoires (cf. Voir 3.1.). Nous les considérons dans le lexique des formes complexes (DECOC).

1.1.2. Informations associées aux noms et codage

Certains traits sémantiques et syntaxiques des noms sont considérés dans le lexique DECOS-NS, ces informations sont indiquées sous forme de code comme suit :

NS	Noms simples
HUM	Substantif <i>humain</i>
ANM	Substantif désignant un <i>animal</i>
PLT	Substantif désignant une <i>plante</i>
PRED1	Substantif <i>prédicatif</i> associable avec <i>hada</i> dont l'ensemble est équivalent à un verbe <i>transitif</i> (i.e. N-하다 <i>hada</i>)
PRED2	Substantif <i>prédicatif</i> associable avec <i>hada</i> dont l'ensemble est équivalent à un verbe <i>intransitif</i> (i.e. N-하다 <i>hada</i>)
PRED3	Substantif <i>prédicatif</i> associable avec <i>doida</i> dont l'ensemble est équivalent à un verbe <i>intransitif</i> (i.e. N-되다 <i>doida</i>)
PREDH	Substantif <i>prédicatif</i> associable avec <i>hada</i> (i.e. N-하다 <i>hada</i>), non indiqués dans les dictionnaires classiques
PREDHA	Substantif <i>prédicatif</i> associable avec <i>hada</i> exigeant une postposition de l'accusatif (i.e. N-를 <i>leul</i> 하다 <i>hada</i>), non indiqués dans les dictionnaires classiques
ADJH	Substantif dont dérive un adjectif en -하다 <i>hada</i>
ETR	Substantif transcrit d'un alphabet <i>étranger</i>
NFA-M	Nominalisation d'adjectif à l'aide du suffixe -음 <i>eum</i>
NFV-M	Nominalisation de verbe à l'aide du suffixe -음 <i>eum</i>
NVK	Nominalisation d'un élément <i>prédicatif</i> à l'aide du suffixe -기 <i>gi</i> (entrée enregistrée dans les dictionnaires éditoriaux, leur liste est donc incomplète)
NVM	Nominalisation d'un élément <i>prédicatif</i> à l'aide du suffixe -음 <i>eum</i> (entrée enregistrée dans les dictionnaires éditoriaux, leur liste est donc incomplète)
NVS	Nominalisation de verbe (ou d'adjectif) à l'aide d'un suffixe nominal comme -보 <i>bo</i> ou -개 <i>gai</i>
*	Symbole désignant une nouvelle adjonction

1.2. Lexique des adjectifs DECOS-AS

1.2.1. Constitution du lexique des adjectifs

Etant donné qu'aucun critère explicite n'existe, dans des dictionnaires actuels, pour définir la catégorie *Adjectif*, les entrées indiquées adjectifs comportent des éléments non adjectivaux, idiomatiques, archaïques ou mal connus. Voici quelques types de données que nous ne traitons pas dans le lexique DECOS-AS :

1.2.1.1. Données défectives

Il s'agit de formes qui ne sont compatibles avec aucun suffixe terminatif :

Epithètes comme 가공할 *gagonghal*, 애꿎은 *aiggujeun*,

Formes figées comme 오죽하면 *ojughamyen*, 여차하면 *yechahamyen*,
 Formes négatives comme 대수롭지않다 *daisulopjiahda*, 달갑지않다 *dalgabjiahda*.

Nous les traitons dans le lexique des expressions figées (DECO-EF).

1.2.1.2. Données non-adjectivales

Il s'agit d'éléments lexicaux classés comme adjectifs dans les dictionnaires mais qui, en fait, doivent être traités soit comme des verbes soit comme des locutions syntaxiquement composées. Par exemple :

Verbes comme 봄비다 *bumbida*, 해당하다 *haidanghada*,
 Noms appuyés d'un terme support comme 남자답다 *namjadabda*,
 귀신같다 *gwisingatda*.

1.2.1.3. Données idiomatiques

Les locutions figées comme 온데간데없다 *on dei gan dei ebsda*, 남녀유별하다 *namnyeyubyelhada* et les mots d'origine étrangère transcrits en coréen comme 애브노멀하다 *aibeunomelhada*, 이지하다 *ijihada* doivent être traités ailleurs. Nous les rangeons dans le lexique des expressions figées (DECO-EF).

1.2.1.4. Données non-synchroniques

On relève dans les dictionnaires un grand nombre d'adjectifs archaïques ou très peu employés dont la nature syntaxique et le sens ne sont pas attestables. La plupart du temps, ce sont des adjectifs sino-coréens vieillissés. Ainsi on ne comprend plus guère le sens de 무편하다 *mupyenhada*, 알매하다 *almaihada* ou 슬므 |다 *seulmeuida*. Et même si la connaissance des idéogrammes chinois nous aide à deviner le sens d'origine, il n'est pas question d'établir les constructions qui leur sont propres, ni de déterminer le type de leurs sujets et compléments. Ils sont donc considérés ailleurs.

1.2.2. Extension à d'autres données

1.2.2.1. Adjectifs dérivés en -이다 *ida*

Certains substantifs associés à *이다* *ida*, du point de vue logique, ne fonctionnent pas comme attributs du sujet, contrairement à la plupart des substantifs qui entrent dans la construction *N1-ida* [être N1]. Par exemple, il s'agit de formes comme 열성이다 *yelsengida*, 안달이다 *andalida*. Du point de vue syntaxique, ils ne répondent pas à la question par 누구 *nugu?* (qui?) ou 무엇 *mues?* (quoi?), mais à celle par 어떠하 *eddeha?* (comment?) ; par ailleurs, on ne peut insérer aucun modifieur avant le substantif accompagné de *ida*, alors qu'on peut insérer des adverbes de degré comme 매우 *maiu* (très) ou 아주 *aju* (très) qui portent sur l'ensemble *N1-ida*. Autrement dit, les séquences basées sur ce type de substantifs sont des adjectifs lexicalement simples. Environ 80 adjectifs en *이다* *-ida* de ce type ont été recensés, et intégrés dans le lexique des adjectifs simples DECOS-AS.

1.2.2.2. Adjectifs en 적이다 *jeg-ida*

Il s'agit de séquences comportant 적이다 *jeg-ida*, comme 이기적이다 *igijegida* ou 합리적이다 *hablijegida*. Dans les dictionnaires actuels, les formes *X-jeg* (où *X* est un substantif ou un élément non-autonome) sont classées comme substantifs ou déterminants, alors que ces formes ne sont pas autonomes : l'ensemble *X-jeg-ida* se comporte comme un simple adjectif du point de vue sémantique et syntaxique. Le nombre de ce type d'adjectifs est environ 700. Ils sont intégrés dans le lexique DECOS-AS.

1.2.3. Informations associées aux adjectifs et codage

Le lexique des adjectifs simples DECOS-AS contient 5300 unités. Certaines informations sont attachées aux entrées lexicales conformément au codage suivant :

AADVHM	Adjectif en suffixe 하다 <i>hada</i> basé sur un adverbe
ANHM	Adjectif en suffixe 하다 <i>hada</i> basé sur un nom
ARHM	Adjectif en suffixe 하다 <i>hada</i> basé sur une racine lexicale
CM	Adjectif en suffixe 적이다 <i>jegida</i>
SM	Adjectif en suffixe 슥다 <i>seulebda</i>
IM	Adjectif en suffixe 이다 <i>ida</i>
JM	Adjectif en suffixe 지다 <i>jida</i>
LM	Adjectif en suffixe 롭다 <i>lobda</i>
MM	Adjectif en suffixe 맞다 <i>majda</i>
OYA	Adjectif en suffixe 있다/없다 <i>issda/ebsda</i>
RAM/A-20	Adjectif sans suffixe spécifique dont la classe est numérotée (e.g. la classe A-20).

1.3. Lexique des verbes *DECOS-VS*

1.3.1. Problèmes liés aux séquences *N-V*

Lors de la construction d'un lexique des verbes simples, des dictionnaires éditoriaux sont consultés en premier lieu. Comme dans les cas du lexique des noms et du lexique des adjectifs, les dictionnaires classiques ne sont pas utilisables pour notre but, faute de principes cohérents. En particulier, concernant les entrées classées verbes, un des problèmes les plus difficiles est de traiter des séquences constituées d'un substantif et d'un verbe comme 하다 *hada*.

On observe que, dans les dictionnaires éditoriaux, un grand nombre de substantifs qui peuvent être accompagnés d'un verbe 하다 *hada* (e.g. 일을 하다 *il-eul hada* (travail-Acc faire)) donnent des verbes dérivés à suffixe 하다 *hada* (e.g. 일하다 *il-hada* (travail-faire = travailler)). Or, la relation *N-Acc hada* = *N-hada* étant très régulière, garder ces noms (*N*) dans un lexique des noms d'une part, et ces verbes (*N-hada*) dans un lexique des verbes d'autre part nous apparaît peu naturel. Nous ne considérons pas que les unités comportant 하다 *hada* (i.e. *N-hada*) justifient systématiquement la variation d'entrées différentes de celles des séquences *N-Acc hada*, pour les raisons suivantes :

1.3.1.1. Régularité de la relation entre *N-hada* et *N-Acc hada*

Sur 14 000 noms simples, environ 5 000 apparaissent sous une forme de type *N-hada*. Or, ils permettent tous la construction *N-Acc hada* sans aucune exception. Il est donc plus naturel d'attacher cette information à ces noms que de dédoubler ces unités (noms et verbes) pour les traiter dans plusieurs lexiques.

1.3.1.2. Productivité de la séquence *N-hada*

La productivité de la séquence *N-hada* constitue un problème, tant que d'une part un lexique complet des noms dérivés n'est pas disponible, et que d'autre part, un grand nombre de noms concrets sont également susceptibles d'apparaître dans la construction *N-hada*, bien qu'ils ne soient indiqués que partiellement dans les dictionnaires actuels. Par exemple, à côté de la séquence 선생하다 *sensaing-hada* (enseignant-faire), on observera également 선생질하다 *sensaing-jil-hada* (enseignant-acte-faire), 선생노릇하다 *sensaing-noleus-hada* (enseignant-rôle-faire), 선생행세하다 *sensaing-haingsei-hada* (enseignant-apparence-faire), qui sont formés sur des noms dérivés de type *N(sensaing)-Sfx*. Par ailleurs, des noms concrets comme 밥 *bab* (riz), 장조림 *jangjolim* (plat de boeuf) peuvent être accompagnés de *hada*, bien que le rôle et le sens de ce *hada* soient probablement différents du *hada* apparu avec des noms prédicatifs. La liste des séquences basées sur ces noms est à établir.

1.3.1.3. Autres séquences permettant la même relation

La relation entre *N-V* et *N-Acc V* s'observe également avec d'autres verbes que *hada*. Par exemple, on a :

벌주다 / 벌을 주다 <i>bel-juda / bel-eul(Acc) juda</i>	(= punition-donner)
사랑받다 / 사랑을 받다 <i>salang-badda / salang-eul(Acc) badda</i>	(= amour-recevoir)
잠자다 / 잠을 자다 <i>jam-jada / jam-eul(Acc) jada</i>	(= sommeil-dormir)

Il serait absurde de considérer des verbes de ce type à côté de ces noms : on marquera l'information de ces constructions autour des entrées **noms**.

1.3.1.4. Adjectifs entrant dans la même relation : *N-Adj* et *N-Postp Adj*

On observe la même relation transformationnelle avec des séquences de type *N-Adj*, comme 맛있다 *mas-issda* (goût-présent), 용기없다 *yonggi-ehsda* (courage-absent),

꿈같다 *ggum-gatda* (rêve-comme) : ils correspondent respectivement à 맛이 있다 *mas-i issda* (goût-Postp présent), 용기가 없다 *yonggi-ga ebsda* (courage-Postp absent), 꿈과 같다 *ggum-gwa gatda* (rêve-Postp comme). Si on traitait les *N-hada* comme des verbes simples, il serait nécessaire de considérer ces séquences adjectivales (i.e. *N-Adj*) comme des adjectifs lexicalisés à côté des adjectifs (i.e. *Adj*) qui peuvent apparaître avec un *N-Postp*, ce qui serait une solution fort redondante.

1.3.2. Constitution du lexique des verbes

1.3.2.1. Verbes défectifs et non autonomes

Dans les dictionnaires éditoriaux, on observe certaines entrées marquées verbes, qui sont en fait des formes morphologiquement défectives ou syntaxiquement non autonomes. D'abord, des formes considérées comme verbes comme 다그다 *dageuda* (s'approcher), 덩달다 *dengdalda* (accompagner), ou 데리다 *deilida* (apporter) n'apparaissent pas seules, mais sous forme complexe comme 다가가다 *daga-gada* (s'approcher-aller), 덩달아올다 *dengdala-ulda* (accompagner-pleurer), ou 데리고오다 *deiligo-oda* (apporter-venir). Quant aux verbes comme 굴다 *gulda* (se comporter) ou 보다 *boda* (essayer), ils ne sont pas syntaxiquement autonomes. Ainsi, sans une séquence (*Adj+V*)-*Sconj*, des phrases comportant ces verbes ne sont pas acceptables comme le montre la phrase suivante :

그들은 누구에게나 (비굴하게 + *E) 굴었다
*geudeul-eun nugu-eigeina (bigulha-gei + *E) gul-ess-da*
 ils-nmtf tous-Postp (lâche-Sconj + *E) se comporter-Pas-St
 (Ils se comportaient (lâchement + E) avec tout le monde)

Les formes *morphologiquement* défectives sont exclues du lexique des verbes DECOS-VS.

1.3.2.2. Séquences de type Nom-Verbe

Comme des formes *Nom-hada*, des séquences comme 거름주다 *geleum-juda*, 방울지다 *bangul-jida*, ou 산통깨다 *santhong-ggaida* ne seront pas représentées dans le lexique DECOS-VS : il s'agit de séquences syntaxiquement complexes obtenues par tranformation à partir des séquences *Nom-Postp Verbe* (e.g. 거름을 주다 *geleum-eul juda*, 방울이 지다 *bangul-i jida*).

1.3.2.3. Séquences de type (Verbe + Adjectif)-Sconj Verbe

Il s'agit d'un autre type de verbes complexes, comme 기뻐하다 *gibb-e-hada*, 가까와지다 *gagga-wa-jida* ou 울게하다 *ul-gei-hada*. Ce type de séquences est parfois très productif, il n'y aurait pas de sens à les intégrer dans le lexique des verbes simples sans en établir la liste complète.

1.3.2.4. Séquences de type Adverbe-Verbe

Les séquences comme 따로서다 *ddalo-seda* (séparément-se lever), 막깎다 *mag-ggagda* (rapidement-couper) sont des formes complexes, constituées d'un adverbe et d'un verbe. Bien qu'ils soient relativement plus figés que des séquences comme 똑바로앉다 *ddogbalo-anjda* (droit-s'asseoir), 분명히하다 *bunmyenghi-hada* (clair-décider), il est difficile de les séparer rigoureusement les uns des autres. Nous écartons ces formes complexes tant que nous n'en avons pas une liste complète.

1.3.2.5. Séquences de type Verbe-go(et)-Verbe

Il s'agit de verbes complexes constitués de deux verbes simples. Deux verbes sont reliés à l'aide d'un suffixe conjonctif 고 *go(et)* comme 놀고먹다 *nol-go-megda* (jouer-et-manger), 싸고돌다 *ssa-go-dolda* (envelopper-et-tourner). La productivité de ce type de séquences étant fort élevée comme le montrent les exemples suivants, on les traitera en dehors du lexique DECOS-VS :

먹고자다	<i>meg-go-jada</i>	(manger-et-dormir)
먹고마시다	<i>meg-go-masida</i>	(manger-et-boire)
죽고살다	<i>jug-go-salda</i>	(mourir-et-vivre)

1.3.2.6. Séquences affixées

Le nombre des verbes affixés comme 엇비슷하다 *es-biseushada*, 맞대다 *maj-daida* ou 엇물리다 *es-mullida* est faible, mais en l'absence d'une liste exhaustive nous les traitons à part. De la même manière, les verbes contenant les suffixes passifs ou causatifs comme 박히다 *bag-hi-da*, 굼기다 *gulm-gi-da*, ou 기울이다 *giul-i-da* sont accompagnés d'un code indiquant que leur liste est à compléter. Les verbes complexes de type 갈아입다 *gal-a-ibda* (changer-Sconj-habiller) ou 건너뛰다 *gen-ne-ddwuida* (traverser-Sconj-sauter) étant peu nombreux, on les a intégrés dans le lexique DECOS-VS, avec le code VC (verbes complexes) : leur liste est relativement complète.

1.3.3. Informations associées aux verbes et codage

VS	Verbe simple
VC	Verbe complexe
INT	Verbe intransitif
TRA	Verbe transitif
ADVHM	Verbe en suffixe 하다 <i>hada</i> basé sur un adverbe
RHM	Verbe en suffixe 하다 <i>hada</i> basé sur une racine lexicale
ADVKM	Verbe en suffixe 거리다 <i>gelida</i> basé sur un adverbe
RKM	Verbe en suffixe 거리다 <i>gelida</i> basé sur une racine lexicale
JM	Verbe en suffixe 지다 <i>jida</i>
TM	Verbe en suffixe 뜨리다 <i>ddeulida</i>
PS	Verbe comportant un suffixe passif
CS	Verbe comportant un suffixe causatif
DE	Verbe incomplet

1.4. Lexique des mots invariables *DECOS-IS*

Le lexique des mots invariables (DECOS-IS) contient des adverbes, des mots fonctionnels et des pronoms. Contrairement aux autres lexiques, le lexique DECOS-IS ne donne pas de formes fléchies et ne contribue pas à la constitution du DECOF. Les détails de l'organisation et des principes de ce lexique ne seront pas traités dans ce travail. Certains problèmes dans la distinction entre les mots invariables et les affixes ou les particules (i.e. unités non autonomes) seront discutés dans les paragraphes suivants.

2.

Dictionnaire électronique des formes fléchies DECOF

2.1. Système de suffixes de conjugaison et de postpositions

La complexité des formes fléchies des éléments prédicatifs se présente comme le premier obstacle à surmonter dans l'analyse automatique des langues naturelles. Étant donné que les verbes et les adjectifs qu'on rencontre dans un texte donné ne sont pas des formes canoniques, mais des formes fléchies, un dictionnaire électronique qui ne fournirait que des formes de base n'aurait pas de pertinence pour l'analyse automatique. Cependant, dans les dictionnaires classiques du coréen, on ne trouve aucune information sur les formes fléchies des éléments canoniques. C'est parce que, d'une part, ils ont été conçus pour être utilisés par des humains qui ont la compétence de cette langue, et d'autre part, le mécanisme des combinaisons de suffixes de conjugaison varie en fonction des éléments prédicatifs. Les grammaires scolaires donnent quelques règles locales qui s'appliquent à certains éléments prédicatifs, mais elles sont très incomplètes et en partie implicites. Par exemple, on parle de suffixes non-terminaux comme le suffixe de politesse *si* ou les suffixes de temps *ess* ou *geiss*, mais l'ordre et les contraintes dans leurs combinaisons (on a par exemple *시었 si-ess*, *시겠 si-geiss* ou *있겠 ess-geiss*, et non **있시 ess-si* ou **겠시있 geiss-si-ess*) ne sont indiqués que partiellement. Par ailleurs, les variations morphologiques aux frontières des suffixes et de la racine sont indiquées de façon peu systématique. À titre d'exemple, considérons les variations morphologiques des adjectifs dont la racine finit par la consonne *ㅂ /b/*. Les grammairiens disent que lorsque ces adjectifs s'associent avec les voyelles comme *ㅏ /e/* ou *ㅗ /eu/*, la finale *ㅂ /b/* devient *ㅑ /u/* et incorpore les voyelles qui suivent. Ainsi, on observe :

- (1) 가볍(다) - 음 *gabye/b/(da) - /eu/m* ---> 가벼움 *gabye-/u/-m*
 [(être) léger] - [Suffixe de nominalisation] ---> [légèreté]

Or, on ne peut pas appliquer cette règle à tous les adjectifs finissant par *ㅂ /b/*, car on observe par exemple :

- (2) 비좁(다) - 음 *bijo/b/(da) - /eu/m* ---> *비조움 *bijo-/u/-m*
 비좁음 *bijo/b/-/eu/m*
 [(être) étroit] - [Suffixe de nominalisation] ---> [étroitesse]

Cependant, aucun principe général ne peut être déduit de ces deux exemples. Il n'y a d'autre recours que d'examiner tous ces adjectifs un par un. Selon notre recensement [Nam97a], seuls 5 adjectifs sur 524 finissant par ㅁ /b/ se comportent comme celui de (2). Cette observation ne permet pas de généralisation aux verbes finissant par ㅁ /b/ : contrairement aux adjectifs, la variation (1) n'est observée qu'avec 6 verbes sur 22, et cette liste n'est point prédictible.

En coréen, le système des suffixes de conjugaison est extrêmement riche et complexe. Les combinaisons sont nombreuses et s'écrivent sans espace typographique. Aucune règle qu'on puisse appliquer de façon générale sur les contraintes des combinaisons de suffixes n'est connue à l'heure actuelle. Nous les avons donc examinés un par un, et nous avons construit des grammaires locales. Par exemple, les combinaisons des suffixes :

있, 겠, 다, 나, 던, 가
ess, geiss, da, nya, den, ga,

sont représentées sous la forme d'un automate fini ⁵ (Figure 1) :

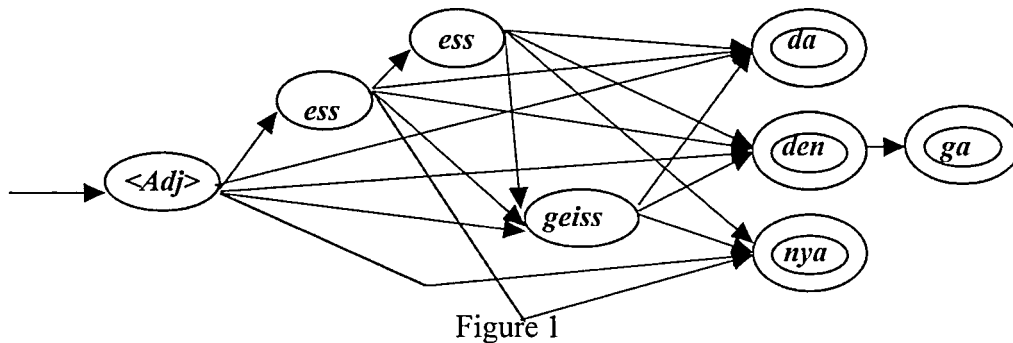


Figure 1

L'automate de la Figure 2 contient les 24 combinaisons suivantes :

<Adj>-ass-ess-da	<Adj>-ass-geiss-da
<Adj>-ass-ess-den	<Adj>-ass-geiss-den
<Adj>-ass-ess-den-ga	<Adj>-ass-geiss-den-ga
<Adj>-ass-ess-nya	<Adj>-ass-geiss-nya
<Adj>-ass-ess-geiss-da	<Adj>-da
<Adj>-ass-ess-geiss-den	<Adj>-den
<Adj>-ass-ess-geiss-den-ga	<Adj>-den-ga
<Adj>-ass-ess-geiss-nya	<Adj>-nya
<Adj>-ass-da	<Adj>-geiss-da
<Adj>-ass-den	<Adj>-geiss-den
<Adj>-ass-den-ga	<Adj>-geiss-den-ga
<Adj>-ass-nya	<Adj>-geiss-nya

Figure 2

⁵ Les cercles doubles représentent des états finaux, donc des séquences reconnues (ou acceptées) par l'automate. Notons que, dans cet automate, la première occurrence du suffixe '있 *ess*' varie en fonction des syllabes finales de la racine adjectivale (e.g. 적었 *je/g/+e/ss-->jeg/e/ss*, 잤 *ja/g/+e/ss-->jag/a/ss* ou 추웠 *chu/b/+e/ss-->ch/we/ss*). Les transducteurs finis permettent de représenter ce type de variation morphologique (cf. [Cle93], [Sil93] et [Lap97]).

En répétant cette procédure, nous avons obtenu environ 6 000 combinaisons de suffixes de conjugaison pour chaque élément prédicatif canonique, et 1 500 combinaisons de postpositions pour chaque substantif. Dans ces combinaisons, on a rendu compte de toutes les variations morphologiques entre les suffixes, et ainsi nous n'avons pas besoin d'ajouter de règles morphologiques ad hoc.

Les adjectifs et les verbes sont répartis en une dizaine de classes en fonction des variations morphologiques des racines. Voici, par exemple, la typologie des classes d'adjectifs (Figure 3) :

Classe [code]	Type de la lettre finale	Variation	Effectifs	Total
NAT1 [010]	<i>Voyelle</i>	<i>sans variation</i>	4462	4462
NAT2 [020]	<i>Consonne sauf les suivantes</i>	<i>sans variation</i>	200	200
NAT3-1 [031]	ㅁ /b/	<i>sans variation</i>	5	524
NAT3-2 [032]		ㅍ /ul/	519	
NAT4-1 [041]	ㅅ /s/	<i>sans variation</i>	1	2
NAT4-2 [042]		<i>E</i>	1	
NAT5-1 [051]	ㅎ /h/	<i>sans variation</i>	1	88
NAT5-2 [052]		<i>E</i>	87	
NAT6-1 [061]	ㄹ /l/	<i>sans variation</i>	29	34
NAT6-2 [062]		<i>E</i>	5	
Total				5300

Figure 3

2.2. Lexique des postpositions *DECO-POST*

2.2.1. Estimation des effectifs des formes fléchies

Les lexiques de postpositions *DECO-POST* sont établis pour qu'on puisse engendrer les lexiques des formes fléchies *DECOF* à partir des *DECOS*. Sauf les entrées du lexique des mots invariables *DECOS-IS*, les formes canoniques des *DECOS* doivent être associées à toutes les formes fléchies qu'on rencontre dans des textes : Le *DECOS-NS* et le *DECO-PostN* permettent d'engendrer le *DECOF-NS* ; le *DECOS-AS* et le *DECO-PostA* donnent le *DECOF-AS* ; l'association du *DECOS-VS* et du *DECO-PostV* produit le *DECOF-VS*. Les effectifs de ces lexiques sont estimés comme suit (Figure 4) :

	DECOS	DECO-POST	DECOF
Nom	15 000	1 500	2.2×10^7
Verbe	7 500	6 000	4.5×10^7
Adjectif	5 300	6 000	3.2×10^7
Mot invariable	7 200		7 200
Total	35 000	13 500	10^8

Figure 4

On notera que ces formes fléchies sont basées sur les items simples. Comme nous l'avons mentionné au-dessus, si on considère des mots affixés (ou dérivés) et composés, le nombre des formes fléchies correspondant sera fortement augmenté. Par ailleurs, les relations morphologiques entre les parties du discours sont complexes, les formes fléchies correspondant doivent être examinées parallèlement. Le graphe suivant montre ces relations (Figure 5) :

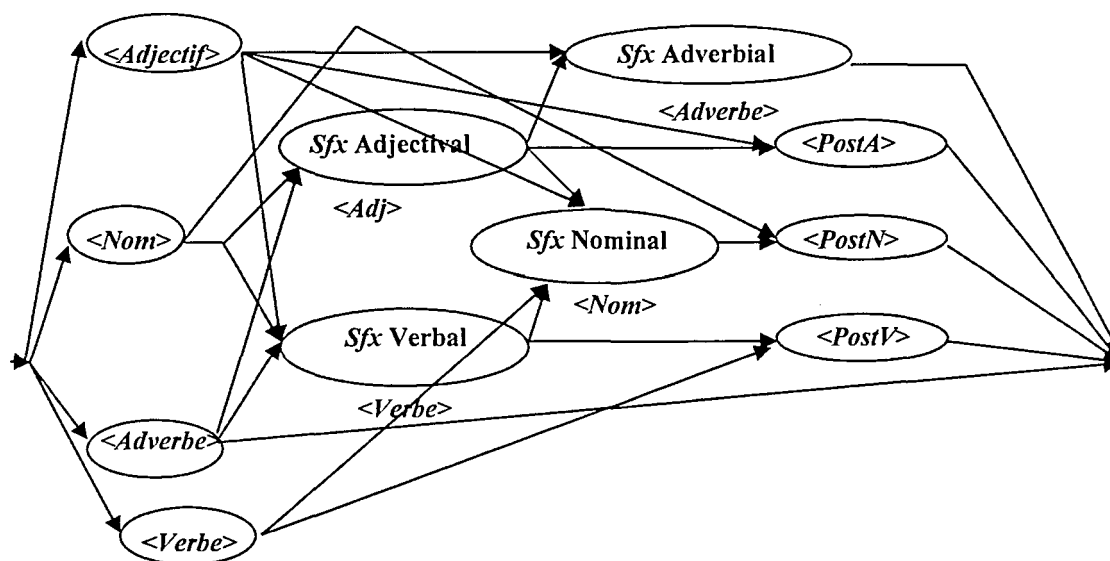


Figure 5

Soit par exemple (Figure 6) :

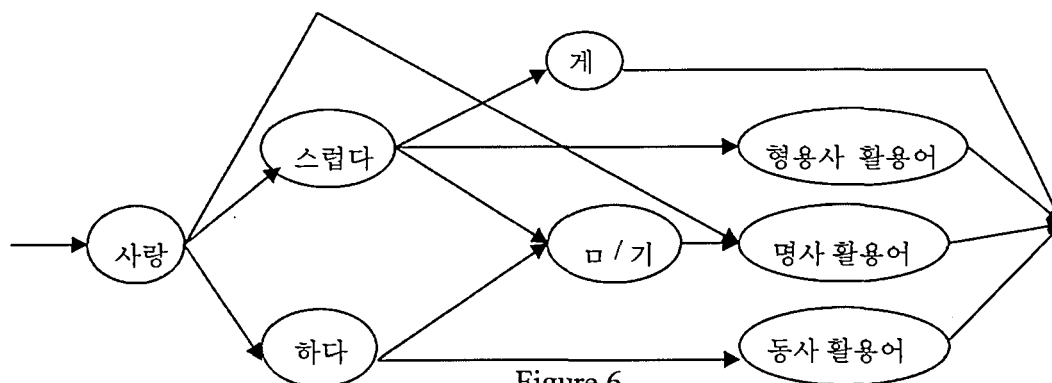


Figure 6

Dans la version actuelle (V01-97) du lexique DECOS que nous présentons dans l'annexe de ce travail, nous avons inséré un lexique des nominalisations d'adjectifs et de verbes par suffixation en 음 -eum (codées NFA-M et NFV-M respectivement). Etant donné que tous les éléments prédicatifs engendrent une nominalisation en 음 -eum, le nombre de ces formes correspond à celui des adjectifs et des verbes (i.e. 7 500 et 5 300). Les variations morphologiques dans la formation des nominalisations en 음 -eum ne sont pas prévisibles en fonction de règles, on doit les examiner une par une. Notons que la version du DECOS que nous présentons ici comporte donc environ 40 600 entrées (i.e. NS, AS, VS, NFA-M et NFV-M).

2.2.2. Lexique des N-Postpositions DECO-PostN

Le lexique DECO-PostN contient 1 500 formes représentées sous forme d'automate fini. Les associations entre les noms et les *PostN* se font sans variations morphologiques. Pour le moment, les *PostN* sont répartis en deux classes : PN1 indique les postpositions qui s'attachent à des substantifs dont la lettre finale est une voyelle, comme 국가 *gugga* (pays) ; PN2 est associé aux substantifs qui se terminent par une consonne, comme 학문 *hagmun* (étude). Quand le marqueur de pluriel 들 *deul* (-s) est inséré, un PN2 apparaît systématiquement, puisque ce marqueur contient une consonne finale. Le graphe suivant représente ces combinaisons (Figure 7).

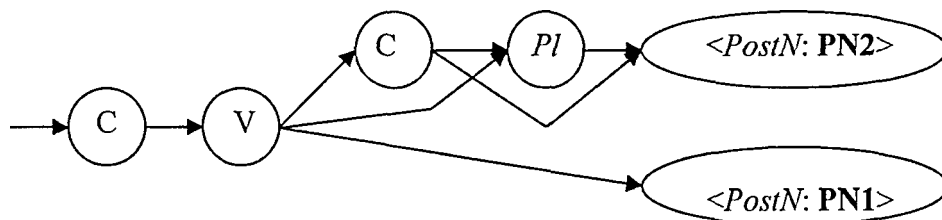


Figure 7

2.2.3. Lexiques des A-Postpositions et des V-Postpositions DECO-PostA/PostV

Les lexiques des A-Postpositions et des V-Postpositions comportent 6 000 entrées. Ces entrées sont obtenues par combinaison de plusieurs postpositions simples. Or, les combinaisons entre les racines verbales ou adjectivales et les postpositions se font avec des changements morphologiques. Il existe trois cas suivants :

- [i] Variation de la syllabe finale des racines ;
- [ii] Variation de la première syllabe des postpositions ;
- [iii] Variation de la syllabe finale des racines et de la première syllabe des postpositions.

Voici des exemples illustrant ces trois cas respectivement :

[i] La syllabe finale du verbe 듣다 *deud-da* (entendre) devient 들 *deul-* devant les postpositions qui commencent par une voyelle (donc typographiquement, une consonne vide), comme 어라 *-ela*, 으면 *-eumyen*, alors qu'elle ne change pas de forme devant les postpositions dont la première syllabe commence par une consonne, comme 고 *-go*, -다가 *daga* (i.e. 듣다 *deudda* --> 들어라 *deul-ela* / 듣고 *deud-go*).

[ii] L'adjectif 착하다 *chagha-da* (gentil) reste invariable contrairement au cas précédent, mais la première syllabe des postpositions qui commencent par une voyelle (donc typographiquement, une consonne vide) subit un changement morphologique : ainsi, les postpositions 어서 *-ese*, 오히려 *-esseumeulo* deviennent 여서 *-yese*, 오히려 *-yesseumeulo* après les racines dont la syllabe finale est 하 *ha-* (i.e. 착하다 *chaghada* --> 착하여서 *chagha-yese*).

[iii] La syllabe finale du verbe 굽다 *gub-da* (cuire) change en même temps que la première syllabe des postpositions qui commencent par une voyelle (donc typographiquement, une consonne vide), comme 으면 *-eumyen*, ou 어 *-e*. La racine devient 구 *gu-* et ces postpositions deviennent 우면 *-umyen* et 워 *-we*. (i.e. 굽다 *gubda* --> 구우면 *gu-umyen*).

2.3. Dictionnaire des formes fléchies *DECOF*

Le dictionnaire des formes fléchies (*DECOF*) est engendré automatiquement par l'association du *DECOS* et du *DECO-POST*. Ce lexique et le *DECOS* sont reliés par des transducteurs finis dans notre système informatique. Des informations morphologiques, nécessaires pour engendrer des formes fléchies à partir des formes canoniques, seront attachées aux entrées du *DECOS* et du *DECO-POST*. Nous n'aborderons pas ici les principes de codage et les méthodes techniques qui sont, pourtant, essentielles pour l'implémentation d'un système informatique.

Pour l'analyse morphologique et syntaxique de textes, les informations que nous venons de discuter doivent être représentées de façon explicite et cohérente. Les formes fléchies des adjectifs et des verbes illustreront la nécessité de présenter d'une manière adéquate ces informations pour éviter des données trop lourdes. Voici un extrait du lexique des adjectifs *DECOS-AS* :

가공적이다 *gagongjegida*, AS06-CM,
 가깝다 *gaggabda*, AS11-RAM/24
 가깝디가깝다 *gaggabdigaggabda*, AS11-RAM/24
 가깝하다 *gaggabhada*, AS03-ARHM
 가난하다 *gananhada*, AS02-ANHM
 가날프다 *ganyalpeuda*, AS11-RAM/143
 가느다랗다 *ganeudalahda*, AS11-RAM/59

3.

Dictionnaire électronique des formes complexes DECOC

3.1. Noms affixés et noms composés

Une des difficultés qu'on rencontre dans l'analyse automatique de textes écrits est due au fait que des formes affixées et composées sont enregistrées dans les dictionnaires actuels de façon incomplète et peu systématique sans être distinguées des items simples. Dans les cas des verbes, adjectifs et adverbes, étant donné que les nombres des formes affixées et composées sont relativement limités, la distinction entre les formes simples et les formes complexes n'a probablement pas beaucoup d'utilité : nous avons actuellement une liste satisfaisante, obtenue à partir des dictionnaires classiques.

Dans le cas des noms, le besoin de séparer les noms complexes des noms simples pourra se justifier de plusieurs manières. Du point de vue morphologique, il est évident qu'il existe certaines règles de dérivation ou de composition, bien qu'une description de ces règles ne soit pas simple. Examinons :

- | | | | |
|--------|------|---------------------|---------------------------|
| (1) a. | 시어머니 | <i>si-emeni</i> | (belle-mère) |
| b. | 재등록 | <i>jai-deunglog</i> | (ré-inscription) |
| c. | 탈옥 | <i>tal-og</i> | (sortir-prison = évasion) |

Le préfixe 시 *si-* (beau-) de (1a) ne s'ajoute qu'à un certain nombre de noms désignant des relations familiales comme 아버지 *abeji* (père), 어머니 *emeni* (mère), 동생 *dongsaing* (frère) ou 누이 *nui* (soeur) : la liste de ces noms dérivés semble fermée. Le préfixe 재 *jai-* (re-) de (1b) crée plus facilement de nouveaux termes dérivés, mais il ne s'associe probablement qu'avec des substantifs prédicatifs : ainsi, l'existence d'un verbe morphologiquement lié (la plupart du temps, ce sont des formes en 하다 *-hada* (i.e. *N-hada*)) est très régulière. En revanche, le préfixe 탈 *tal-* (sortir-) de (1c) rend prédicatif le substantif attaché, qui peut être soit un nom concret soit un nom abstrait : ainsi, le substantif concret 옥 *og* (prison) qui n'a pas de relation morphologique avec un verbe

A partir de ce dictionnaire, on engendre le dictionnaire des formes fléchies DECOF-AS.
Soit par exemple :

가까운 *gaggaun*, <*gaggabda+eun*>, AS11-RAM/24-2, V-Det
 가깝니 *gaggabni*, <*gaggabda+ni*>, AS11-RAM/24-1, C-Int
 가깝워서 *gaggawese*, <*gaggabda+ese*>, AS11-RAM/24-3, V-Conj
 가깝한 *gaggabhan*, <*gaggabhada+eun*>, AS03-ARHM/NAT1, V-Det
 가깝하므로 *gaggabhameulo*, <*gaggabhada+meulo*>, AS03-ARHM/NAT1, C-Conj
 가날퍼 *ganyalpe*, <*ganyalpeuda+e*>, AS11-RAM/143/NAT1, V-Conj
 가날프니 *ganyalpeumi*, <*ganyalpeuda+ni*>, AS11-RAM/143/NAT1, C-Conj/Int
 가날픈 *ganyalpeun*, <*ganyalpeuda+eun*>, AS11-RAM/143/NAT1, V-Det

Pour le moment, 5 300 adjectifs canoniques engendrent 32 millions (5 300 x 6 000) de formes fléchies. Mais, quand des informations sur les contraintes morpho-syntaxiques plus détaillées seront disponibles, l'effectif des formes fléchies sera fortement diminué. Le passage entre ces deux dictionnaires exige un module informatique extrêmement efficace, sans lequel on ne peut espérer traiter de gros textes.

aura un verbe dérivé, une fois qu'il est associé au préfixe 탈 *tal-* (sortir) : 탈옥하다 *tal-og-hada* (sortir-prison-faire = s'évader).⁶

Du point de vue syntaxique, la structure interne des noms composés est à étudier d'une manière plus systématique et approfondie. Prenons des exemples :

- (2) a. 밤낮 *bam-naj* (nuit-jour = toujours)
 b. 집사람 *jib-salam* (maison-personne = épouse)
 c. 돼지고기 *dwaiji-gogi* (cochon-viande = porc)
 d. 예술사진 *yeisul-sajin* (art-photo = photo artistique)

Au niveau typographique, pour les formes à moins de 4 syllabes ((2a) et (2b)), il est difficile d'insérer un blanc entre les constituants, alors que les formes à 4 syllabes (c'est le cas le plus fréquent) ou à plus de 4 syllabes se réalisent souvent de deux manières : soit **séparées** en plusieurs unités par le(s) blanc(s) soit **soudées** en unité simple. Ainsi, on observe pour (2c) et (2d) :

- (3) 돼지고기 / 돼지 고기 *dwaijigogi / dwaiji gogi*
 예술사진 / 예술 사진 *yeisulsajin / yeisul sajin*

Rappelons que deux formes typographiques (avec et sans blanc) posent des problèmes complexes non seulement dans la reconnaissance des mots, mais aussi dans l'analyse syntaxique des textes. Quant à la structure interne, le degré de figement est très élevé dans (2a) : deux noms simples en relation d'antonymie constituent un nouveau terme dont le sens n'est pas prévisible à partir de la composition des sens de ces deux noms. Il en est de même avec (2b). Si 사람 *salam* (personne) peut désigner un homme aussi bien qu'une femme, 집사람 *jib-salam* ne désigne que la femme (i.e. épouse) ; on ne pourra non plus remplacer 집 *jib* (maison) par des synonymes de type :

= *(아파트 + 가옥 + 건물)-사람
 *(*apateu*(appartement) + *gaog*(pavillon) + *genmul*(bâtiment))-*salam* (personne)

mais on observe une variante :

안사람 *an-salam* (intérieur-personne = épouse)

Les composés (2c) et (2d) sont moins lexicalisés que ces deux premiers ((2a) et (2b)). Autrement dit, le choix de combinaisons y est plus libre et le changement de sens dans les nouveaux termes est moindre : pour une structure de type (2c), tous les noms d'animaux comestibles sont en principe acceptés en position du premier nom. Ici, on

⁶ On observe le même phénomène avec le préfixe 입 *ib-* (entrer-). Quand les substantifs concrets de type 성 *seng* (château) ou 관 *gwan* (bière) sont associés à ce préfixe, la dérivation des verbes s'effectue alors :

*성하다 **seng-hada* *ib-seng-hada* (entrer dans le château)
 *관하다 **gwan-hada* *ib-gwan-hada* (mettre le mort dans la bière)

pourrait supposer que (2c) et (2d) autorisent deux formes typographiques (avec et sans blanc), non seulement parce qu'ils sont de quatre syllabes, mais aussi parce qu'ils ne sont pas figés, contrairement à (2a) et (2b), qui sont plutôt lexicalisés. Autrement dit, (2c) et (2d) pourraient être considérés comme des groupes nominaux libres comme les exemples suivants :

- (4) a. 돼지머리 *dwaiji-meli* (cochon-tête = tête de cochon)
 b. 동생사진 *dongsaing-sajin* (frère-photo = photo de frère)

Cependant, on observe que si (4a) et (4b) entrent dans la relation " $NaNb = Na-eui Nb$ (Nb de Na)", (2c) et (2d) l'autorisent difficilement. Considérons :

- (5) a. 돼지의 머리 *dwaiji-eui meli* (cochon-Gén tête = tête de cochon)
 b. 동생의 사진 *dongsaing-eui sajin* (frère-Gén photo = photo de frère)
- (6) a. ??돼지의 고기 ??*dwaiji-eui gogi* (cochon-Gén viande = porc)
 b. ??예술의 사진 ??*yeisul-eui sajin* (art-Gén photo = photo artistique)

Dans cette situation, l'identification des séquences (2c) et (2d) devient plus complexe. En dehors des deux cas extrêmes (i.e. totalement figés et *totalement* (supposons-le) libres), il est difficile d'attribuer aux séquences complexes tantôt l'étiquette de **nom composé** tantôt l'étiquette de **groupe nominal** : cette dichotomie n'a peut-être pas de sens.

Cette observation nous suggère que la relation $NaNb = Na-eui Nb$ ne détermine qu'un type parmi les séquences NN , et que cette relation ne peut pas être considérée comme un critère fondamental qui garantit une distinction entre un groupe nominal et un nom composé. Nous proposerons donc, dans le paragraphe suivant, quelques types de phrases élémentaires qui permettront de décrire et classer des NN , sans nous préoccuper de les nommer tantôt nom composé tantôt groupe nominal.

Il est à noter aussi que si on observe 예술사진 *yeisul-sajin* (photo artistique) dans des dictionnaires, on n'a pas 증명사진 *jeungmyeng-sajin* (photo d'identité), malgré leur ressemblance dans l'examen de l'équivalence avec $Na-eui Nb$. L'arbitraire préside fréquemment à l'enregistrement des formes NN dans les dictionnaires. On se demande pour quelle raison les formes (7) sont enregistrées, alors que celles de (8) sont absentes :

- (7) a. 가죽가방 *gajug-gabang* (cuir-sac = sac **en cuir**)
 b. 가죽잠바 *gajug-jamba* (cuir-blouson = blouson **en cuir**)
 c. 가죽조끼 *gajug-joggi* (cuir-gilet = gilet **en cuir**)
- (8) a. 플라스틱가방 *peullaseutig-gabang* (plastique-sac = sac **en plastique**)
 b. 면잠바 *myen-jamba* (coton-blouson = blouson **en coton**)
 c. 털조끼 *tel-joggi* (laine-gilet = gilet **en laine**)

Il en est de même avec la série (9), absente dans les dictionnaires, bien qu'il n'y ait différence avec (7) :

- (9) a. 가죽치마 *gajug-chima* (cuir-jupe = **jupe** en cuir)
 b. 가죽벨트 *gajug-beilteu* (cuir-ceinture = **ceinture** en cuir)
 c. 가죽코트 *gajug-khoteu* (cuir-manteau = **manteau** en cuir)

Tant que la liste des noms composés n'est établie, dans les dictionnaires actuels, ni avec des principes cohérents ni d'une manière exhaustive, il nous apparaît qu'il ne faut pas intégrer cette liste telle qu'elle est dans notre dictionnaire.

De plus, on rencontre des situations encore plus embarrassantes au cours du recensement des noms simples à partir des dictionnaires actuels. Examinons :

- (10)a. 옛추억 *yeis-chueg* (ancien-souvenir = souvenir du passé)
 b. 옛친구 *yeis-chingu* (ancien-ami = vieil ami)
 (11)a. 전남편 *jen-nampyen* (passé-mari = ex-mari)
 b. 전시대 *jen-sidai* (passé-époque = époque passée)
 (12)a. 후세대 *hu-seidai* (futur-génération = génération future)
 b. 후처 *hu-che* (futur-épouse = deuxième épouse)
 (13)a. 새색시 *sai-saigsi* (nouveau-mariée = nouvelle mariée)
 b. 새책 *sai-chaig* (nouveau-livre = nouveau livre)

La partie 옛 *yeis-* (ancien) de (10), dans les dictionnaires actuels, a un double statut : 접두사 *jebdusa* (préfixe) et 관형사 *gwanhyengsa* (déterminant), le critère pour distinguer ces deux emplois n'étant pas précis ; l'élément 전 *jen-* (passé) de (11) est traité comme un préfixe, alors que 후 *hu-* (futur) de (12), relié au précédent par une relation d'antonymie, est considéré comme un nom ; enfin, l'unité 새 *sai-* (nouveau) est nommée déterminant. Ces étiquetages ne semblent pas être basés sur des principes cohérents. Pourtant, ils ont conséquences importantes dans le recensement et la description des noms : les mots (10) deviennent soit des noms dérivés (comportant le préfixe 옛 *yeis-* (ancien)) soit des séquences syntaxiquement complexes (séquence soudée du déterminant 옛 *yeis-* et un nom simple) ; (11) seraient des noms dérivés comportant le préfixe 전 *jen-* (passé) ; (12) seraient des noms composés, puisque 후 *hu-* (futur) est un nom ; (13) seraient des séquences syntaxiquement complexes soudées (i.e. déterminant-nom) comme (10) avec la deuxième analyse.

La situation concernant les exemples suivants est également problématique. Il s'agit de l'utilisation des notions de **noms incomplets** et de **suffixes**. Voici des exemples :

- (14)a. 연구회 *yengu-hoi* (recherche-réunion = équipe de recherche)
 b. 송별회 *songbyel-hoi* (séparation-réunion = pot de départ)
- (15)a. 이십년 *isib-nyen* (vingt-an = vingt ans)
 b. 금년 *keum-nyen* (ce-an = cette année)

Les éléments 회 *-hoi* n'ont pas la même signification dans (14a) et (14b) : le premier désigne une réunion durable et donc souvent une association ou une équipe, et le deuxième une réunion ponctuelle et ici un pot de départ. Cependant, cela ne fournit pas de raison pour que le premier soit considéré, dans les dictionnaires, comme un nom, et le deuxième comme un suffixe. Dans les cas (15a) et (15b), les deux 년 *-nyen* n'ont pas le même statut : le premier est classé nom incomplet, alors que le deuxième est considéré comme suffixe.

En fait, dans aucune grammaire scolaire, les distinctions entre préfixe et déterminant d'une part, et entre suffixe et nom incomplet d'autre part n'ont été effectuées de façon opératoire. Par conséquent, dans les dictionnaires actuels, l'enregistrement des noms comportant ces éléments est souvent aléatoire et donc aucun principe de classification n'y est discernable. Ce présent travail ne sera pas indépendant de ce problème, vu que la grande majorité de la couverture lexicale du coréen est constituée de noms, surtout des noms affixés et composés dont l'effectif maximal n'est pas encore évaluable.

Soulignons que nous envisageons de les séparer des noms simples avec des principes applicables et cohérents et de les répertorier sous une autre forme de dictionnaire, d'une part pour pouvoir compléter ces listes d'une manière plus systématique, et d'autre part pour pouvoir leur associer des informations syntaxiques dont les caractéristiques peuvent être décrites et regroupées en fonction des catégories grammaticales.

3.2. Typologie des structures de noms

Les noms coréens peuvent correspondre aux structures suivantes (PF désigne un préfixe, défini ici comme élément non-autonome s'ajoutant en tête d'un mot ; SF est un suffixe, élément non-autonome s'ajoutant à la fin d'un mot ; N désigne un nom (simple), donc un élément autonome) :

- (1) a. N
 b. PF-N
 c. N-SF

- d. PF-N-SF
- e. PF-PF-N
- f. N-SF-SF

g. PF-SF

- (2) a. N-N
 b. PF-N-N
 c. N-N-SF
 d. N-N-N

Seule la forme (1a) doit être traitée comme nom simple ; les formes (1b)~(1f) sont des noms affixés ; (2) sont des composés. La distinction entre (1a) et (1b)~(1f) est faite en fonction de l'existence de préfixes et/ou de suffixes en dehors d'un élément autonome (i.e. N). La distinction entre (1) et (2) est effectuée selon le nombre de noms, c'est-à-dire dans (2) le nombre de noms est égal ou supérieur à 2. Soulignons donc que la typologie décrite ici ne correspond pas aux modèles présentés par certaines grammaires scolaires qui, d'une part, distinguent plusieurs niveaux morphologiques (e.g. [PF-N]-SF, PF-[PF-N] ou N-[N-N]), et d'autre part considèrent des unités de type PF-X, X-SF ou PF-SF comme dérivées (X étant un élément non-autonome), alors que pour nous ce sont des noms simples que nous ne décrivons pas autrement que N (ainsi, nous ne nous intéressons pas à l'analyse en morphèmes des mots qui ne possèdent aucune partie autonome à l'intérieur). Par conséquent, si une unité autonome suit un préfixe, cette combinaison sera reconnue comme affixée, mais sinon, comme simple.

Voici des exemples qui illustrent respectivement les types de combinaisons (1) et (2) :

- (3) a. 가을 *gaeul* (automne)
 b. 재결합 *jai-gyelhab* (ré-unification = réunification)
 c. 포도주 *podo-ju* (alcool de raisin = vin)
 d. 동성애 *dong-seng-ai* (même-sexe-amour = amour homosexuel)
 e. 가출옥 *ga-chul-og* (provisoire-sortir-prison=sortie provisoire de prison)
 f. 경리부장 *gyengli-bu-jang* (comptabilité-département-responsable
 = responsable du département de comptabilité)
- (4) a. 구슬땀 *guseul-ddam* (perle-sueur = goutte de sueur de forme perle)
 b. 반독재투쟁 *ban-dogjai-tujaing* (anti-dictature-révolte = révolte anti-dictature)
 c. 간호보조원 *ganho-bojo-wen* (infirmier-assistance-personne = assistant d'infirmier)
 d. 가운데손가락 *gaundeis-son-galag* (centre-main-doigt = majeur)

On pourrait supposer a priori d'autres types de combinaisons possibles (e.g. PF-PF-N-SF / N-SF-SF-SF / PF-N-N-SF, etc.), mais le nombre des formes autres que (1) et (2) est

extrêmement réduit.⁷ Par contre, des séquences enchaînées de noms simples de type (5) sont s'observent fréquemment :

- (5) 국어 맞춤법 통일 개정안
guge majchum beb tongil gaijeng an
 coréen orthographe règle unification réforme proposition
 (Proposition de réforme pour l'unification des règles d'orthographe du coréen)

La séquence (5) est composée de six noms simples (N N N N N N). Rappelons que la règle de soudage ne peut pas servir de critère opératoire, car on observe également :

- (6) a. 국어 맞춤법 통일 개정안
guge majchumbeb tongil gaijengan [N NN N NN]
 coréen orthographe-règle unification réforme-proposition
- b. 국어맞춤법 통일개정안
gugemajchumbeb tongilgaijengan [NNN NNN]
 coréen-orthographe-règle unification-réforme-proposition

Pourtant, bien que cette règle ne soit pas fiable, il existe des contraintes :

- (7) a. *국어맞춤법 통일개정안
 **gugemajchum beb tongilgaijeng an* [NN N NN N]
 *coréen-orthographe règle unification-réforme proposition
- b. ?*국어 맞춤법통일개정안
 ?**guge majchumbebtongilgaijeng an* [N NNNN N]
 ?*coréen orthographe-règle-unification-réforme proposition

Cela est probablement dû au fait que les formations de mots composés à l'intérieur de cette séquence ne se font pas au même niveau morphologique.

⁷ Voici des exemples :

- (i) 국세청장실
gugsei-cheng-jang-sil [N-SF-SF-SF]
 (impôt national-département-résponsable-bureau)
 (Bureau du responsable du département national des impôts)
- (ii) 신경제학이론
sin-gyengjei-hag-ilon [PF-N-SF-N]
 nouveau-économie-science-théorie
 (Nouvelle théorie de la science économique)

3.3. Définition des catégories grammaticales

Au cours du recensement de noms, on rencontre fréquemment des cas où il n'est pas possible de décider s'il s'agit d'un dérivé ou un composé, sans faire recours à l'identification grammaticale des éléments concernés. Nous avons mentionné brièvement l'absence d'un système explicite et cohérent dans les dictionnaires actuels sur ce sujet. Prenons un exemple :

- (1) 간밤 *gan-bam* (passé-nuit = cette nuit)

L'élément précédant le nom 밤 *bam* (nuit) est considéré, dans les dictionnaires, comme un 접두사 *jebdusa* [préfixe]. Or, comment peut-on le distinguer d'une part de l'élément 왼 *oin-* (gauche), dans l'exemple suivant, classé 관형사 *gwanhyengsa* [déterminant] ?

- (2) 왼편 *oin-pyen* (gauche-côté = côté gauche)

Et d'autre part, comment le distinguer de 긴 *gin-* (long) de (3), qui n'est indiqué ni comme préfixe, ni comme déterminant, mais implicitement comme adjectif conjugué à l'aide d'un suffixe déterminatif (i.e. 길다 *gilda* (forme canonique de l'adjectif "long") - ㄴ *n* (Sd) = 긴 *gi-n*) ?

- (3) 긴밤 *gi-n-bam* (long-Sd-nuit = nuit longue)

Pour pouvoir répondre à ces questions, il est indispensable de cerner formellement ces catégories. Dans le paragraphe suivant, nous les examinerons plus en détail.

3.3.1. Déterminant (Dét) vs. Adj-Sd / V-Sd

Un déterminant (Dét) est distingué d'un adjectif ou verbe suivi d'un suffixe déterminatif (Sd), par le fait que ce dernier peut s'observer sous une autre forme que celle de déterminant, notamment sous une forme prédicative (avec un suffixe terminal (St) de type 다-*da* ou 는다-*nda*). Ainsi, (1) comportent des Dét, aussi bien que (2) :

- | | | | |
|--------|--------|---------------------------|------------------------------------|
| (1) a. | 여러 학생들 | <i>yele hagsaing-deul</i> | (plusieurs étudiant-Mpl) |
| b. | 매 시간 | <i>mai sigan</i> | (chaque heure) |
| c. | 또한 생각 | <i>ddan sainggag</i> | (différent idée = idée différente) |
| (2) a. | 이 책상 | <i>i chaigsang</i> | (ce table = cette table) |
| b. | 저 소녀 | <i>je sonye</i> | (ce(-là) fille = cette fille-là) |
| c. | 그 책 | <i>geu chaig</i> | (ce livre) |

Par contre, les séquences suivantes comportent un adjectif ou un verbe suivi d'un suffixe déterminatif (Sd), car on observe parallèlement des phrases élémentaires dont cet adjectif ou ce verbe est un élément prédicatif :

- (3) a. 큰 가방 *keu-n gapang* (grand sac)
 b. 단 맛 *da-n mas* (sucré goût = goût sucré)
 c. 다른 생각 *daleu-n sainggag* (différent idée = idée différente)
 d. 쉰 밥 *swi-n bab* (pourri riz = riz pourri)
- (4) a.= 가방이 크다 *gapang-i(nmtf) keu-da(St)* (Ce sac est grand)
 b.= 맛이 달다 *mas-i(nmtf) dal-da(St)* (Ce goût est sucré)
 c.= 생각이 다르다 *sainggag-i(nmtf) daleu-da(St)* (Cette idée est différente)
 d.= 밥이 쉬었다 *bab-i(nmtf) swi-ess(Mpas)-da(St)* (Ce riz a pourri)

Quand ces formes déterminatives sont figées avec un substantif, la transformation de type (4) n'est pas admise, mais on reconnaît l'existence (lexicale) de leurs formes adjectivales ou verbales, par analogie avec (4). Il s'agit des cas comme :

- (5) a. 큰아버지 *keu-n-abeji* (grand-père = grand frère du père)
 = *아버지가 크다 **abeji-ga(nmtf) keu-da(St)* (Le père est grand)
- b. 단잠 *da-n-jam* (sucré-sommeil = sommeil profond)
 = ?*잠이 달다 *?*jam-i(nmtf) dal-da(St)* (Le sommeil est sucré[profond])

Reprenons les exemples cités plus haut :

- (6) a. 간밤 *ga-n-bam* (passé-nuit = cette nuit)
 b. 왼편 *oin-pyen* (gauche-côté = côté gauche)
 c. 긴밤 *gi-n-bam* (long-Sd-nuit = nuit longue)

Pour nous, (6a) et (6c) seront donc analysés comme mots composés, puisque les éléments 간 *gan-*(passé) et 긴 *gin-* (long) sont liés respectivement au verbe 가다 *gada* (passer) et à l'adjectif 길다 *gilda* (long), contrairement à 왼 *oin-* (gauche) de (6b) qui n'a aucune relation de ce type :

- (7) a.= 밤이 갔다 *bam-i(nmtf) ga-ess(Mpas)-da(St)* (La nuit est passée)
 b.= *편이 외다 **pyen-i(nmtf) oi-da(St)* (Le côté est gauche)
 c.= 밤이 길다 *bam-i(nmtf) gil-da(St)* (La nuit est longue)

(6a) et (6c) sont donc de structures :

- (8) *V-Sd N*
Adj-Sd N

Soulignons qu'ici, on appelle les séquences (8) noms composés, par opposition aux noms affixés, et non par opposition aux groupes nominaux.

3.3.2. Déterminant (*Dét*) vs. Préfixe (*PF*)

Un déterminant est un élément indépendant, alors qu'un préfixe n'est pas syntaxiquement autonome. Ainsi, si on peut écrire un élément à gauche d'un substantif en le séparant avec un blanc, on ne peut pas l'appeler préfixe. Notons que si une séquence soudée ne représente pas toujours une unité simple en coréen (étant donné qu'elle s'écrit aussi avec un blanc), une séquence séparée par un blanc ne peut jamais être une seule unité. Examinons :

- (1) a. 새[✓]자가용 *sai-jagayong* (nouveau-voiture = nouvelle voiture)
 b. 시어머니 *si-emeni* (beau-mère = belle-mère)
- (2) a. 새 자가용 *sai jagayong*
 b. ?*시 어머니 ?**si emeni*

(1a) contient un déterminant, (1b) un préfixe. Pour les deux éléments suivants, sémantiquement voisins :

- (3) a. 온세계 *on-segyei* (entier-monde = monde entier)
 b. 전세계 *jen-segyei* (entier-monde = monde entier)

on observe :

- (4) a. 온 세계 *on segyei*
 b. 전 세계 *jen segyei*

Il sont donc tous les deux des mots comportant un déterminant, bien que, dans les dictionnaires actuels, 온 *on* (entier) soit traité comme déterminant, 전 *jen* (entier) comme préfixe.

La relation entre le déterminant et le substantif est, en général, plus libre que celle entre le préfixe et le substantif, la productivité des mots préfixés est donc plus réduite et contrainte que celle des séquences *Dét N*. Ainsi, le nombre des mots contenant 새 *sai* (nouveau) (soudé au substantif ou non) doit être nettement plus élevé que celui des mots commençant par 시 *si-* (beau). Cependant, cette règle n'est pas absolue, quand on compare :

- (5) a. 온세계 *on-segyei* (entier-monde = monde entier)
 b. 여선생 *ye-sensaing* (féminin-professeur = femme professeur)

il n'est pas certain que le préfixe 여 *ye-* (féminin) donne une série de noms plus limitée que celle qui comporte le déterminant 온 *on-* (entier).

Le degré de cohésion entre les constituants de séquences peut être mesuré par la possibilité ou non de l'insertion d'un mot à l'intérieur de séquences. Ainsi, nous proposons un critère : si on peut insérer une unité autonome entre un nom et l'élément qui précède, cet élément sera reconnu comme un déterminant, sinon il s'agira d'un préfixe. Examinons les cas suivants. Dans les dictionnaires actuels, quelques mots ont deux étiquettes comme :

- (6) 별 *byel* (inattendu) : [Dét], [PF]
 a. 별소리 *byel[Dét]-soli* (inattendu-propos = propos absurde)
 b. 별말 *byel[PF]-mal* (inattendu-parole = parole absurde)
- (7) 첫 *ches* (premier = le plus ancien) : [Dét], [PF]
 a. 첫대통령 *ches[Dét]-daitonglyeng* (premier-chef d'Etat)
 b. 첫눈 *ches[PF]-nun* (premier-neige (de la saison))

Cependant, il nous apparaît que tous les exemples (6) et (7) doivent être traités comme des séquences comportant un déterminant, et non des noms préfixés, car on peut insérer un élément autonome à l'intérieur de tous ces exemples. Soit :

- (8) 별 이상한 소리를 다 듣겠네 !
byel isangha-n soli-leul da deut-gessne !
 inattendu **bizarre-Sd** propos-Acc enfin entendre-St
 (Enfin, quels propos **bizarres** et inattendus on entend !)
- (9) 첫 한국의 대통령
ches hangug-eui daitonglyeng
 (premier **Corée-Gén** chef d'Etat = premier chef d'Etat **de la Corée**)

Reprenons les exemples cités ci-dessus :

- (10)a. 새자가용 *sai-jagayong* (nouveau-voiture = nouvelle voiture)
 b. 시어머니 *si-emeni* (beau-mère = belle-mère)
- (11)a. 온세계 *on-segyei* (entier-monde = monde entier)
 b. 전세계 *jen-segyei* (entier-monde = monde entier)
 c. 여선생 *ye-sensaing* (féminin-professeur = femme professeur)

L'examen de l'insertion d'un élément à l'intérieur de ces séquences nous donne le résultat suivant :

- (12)a. 새 멋진 자가용 *sai mesji-n jagayong*
(nouveau magnifique-Sd voiture = une nouvelle voiture magnifique)
- b. *시 멋진 어머니 **si mesji-n emeni*
(beau magnifique-Sd mère = une belle-mère magnifique)
- (13)a. 온 서구 세계 *on segu segyei*
(entier **occident** monde = l'**Occident** en entier)
- b. 전 서구 세계 *jen segu segyei*
(entier **occident** monde = l'**Occident** en entier)
- c. *여 훌륭한 선생 **ye hullyungha-n sensaing*
(féminin **excellent-Sd** professeur = la femme professeur **excellente**)

Ce résultat nous confirme que (12a), (13a) et (13b) comportent un déterminant, alors que (12b) et (13c) sont des mots préfixés qui acceptent difficilement l'insertion d'un élément à l'intérieur. On les considérera ici comme des unités lexicalisées (ou figées) dont la première partie n'est plus un déterminant libre, mais un préfixe morphologiquement et typographiquement soudé à ces substantifs. En fait, le double étiquetage pour un élément donné (déterminant et préfixe) n'a pas de sens à moins d'énumérer les substantifs concernés.

3.3.3. Préfixe (PF) vs. Nom (N)

Examinons :

- (1) a. 왕뱀 *wang-baim* (roi-serpent = serpent géant)
b. 악조건 *ag-jogen* (mal-condition = mauvaise condition)

Le premier élément de chaque exemple est considéré comme un préfixe dans les dictionnaires. Pourtant, on sait qu'il existe des noms homonymes, dont le sens n'est, d'ailleurs, pas complètement indépendant de celui de ces éléments de (1). Soit :

- (2) a. 왕이 되다 *wang-i doi-da*
(roi-nmtf devenir-St = devenir un roi)
- b. 악에 물들다 *ag-e muldeul-da*
(mal-Postp.e influencé-St = être sous l'influence du mal)

Alors, 왕 *wang* et 악 *ag* de (1) ne seront-ils pas des noms, et donc les mots (1) des noms composés ? Sinon, selon quel critère les distinguer ? Maintenant, comparons (1) avec :

- (3) a. 왕세자 *wang-seja* (roi-fils = fils du roi)
 b. 악령 *ag-lyeng* (mal-fantôme = diable)

L'intuition immédiate répugne à identifier la série (3) avec (1). Alors pourquoi ? On pourrait tenter d'expliquer qu'ils sont différents parce que, par exemple, 왕세자 *wang-seija* (fils du roi) garde toujours la distribution de 왕 *wang* (roi), alors que 왕뱀 *wang-baim* (serpent géant) n'a plus rien de commun avec 왕 *wang* (roi). Ainsi, on observe :

- (4) 그 (왕 + 왕세자)-는 영국사람이다
geu (wang + wang-seja)-neun yenggug-salam-i-da
 ce (roi + roi-fils)-nmtf Angleterre-personne-être-St
 (Ce (roi + fils de roi) est une personne de nationalité anglaise)
- (5) 그 (왕 + *왕뱀)-은 영국사람이다
*geu (wang + *wang-baim)-eun yenggug-salam-i-ta*
 ce (roi + *roi-serpent)-nmtf Angleterre-personne-être-St
 (Ce(t) (roi + serpent géant) est une personne de nationalité anglaise)

Par ailleurs, et si (3a) correspond à un groupe nominal de type (6), (1a) n'a pas cette correspondance :

- (6) a. *NaNb*
 b. = *Na-eui Nb* [Nb de Na]

Soit :

- (7) a. 왕세자 *wang-seja* (roi-fils = fils du roi)
 b. = 왕의 세자 *wang-eui seja* (roi-**Gén** fils = fils du roi)
- (8) a. 왕뱀 *wang-baim* (roi-serpent = serpent géant)
 b. = *왕의 뱀 **wang-eui baim* (roi-**Gén** serpent = serpent géant)

Or, il se trouve que cette relation ne peut pas être généralisée. Par exemple, si on a :

- (9) a. 굴껍질 *gul-ggebjil* (huître-coquille=coquille de l'huître)
 b. 굴김치 *gul-gimchi* (huître-salade=salade comportant des huîtres)

On sait que 굴 *gul* (huître) étant concrètement présent dans ces deux cas, il s'agit de noms composés de type *NN* : (9a) signifiant la coquille de l'huître, il partage la distribution avec 굴 *gul* (huître) comme dans (4) ; 굴 *gul* dans (9b) n'est pas le pivot du mot, (9b) désigne une sorte de 김치 *gimchi* (salade) qui comporte des huîtres. La relation (6) n'est admise que pour (9a) :

- (10)a.= 굴의 껍질 *gul-eui ggebjil* (huître- Gén coquille)
 b.= *굴의 김치 **gul-eui gimchi* (huître- Gén saladé)

Malgré cette observation, il est évident qu'on ne peut pas traiter 굴김치 *gul-gimchi* (huître-salade) autrement qu'un mot composé. Par ailleurs, on verra que les structures internes des noms composés de type *NN* sont tellement variées que la relation examinée au-dessus ne suffit pas pour les caractériser vis-à-vis des noms affixés.

Alors s'il en est ainsi, devons-nous traiter 왕뱀 *wang-baim* de (1a) comme un composé qui serait d'une autre structure que (3a) ? Notre réponse est négative et nous parlerons de nom affixé dans le cas de (1a). Voici le principe que nous avons adopté : si on peut reconstituer, pour une forme à deux éléments, une sorte de paraphrase qui comporte ces deux éléments en tant que substantifs, cette forme sera reconnue comme un nom composé, mais s'il est difficile de construire une telle phrase, cette forme sera considérée comme un nom affixé. Ainsi, si on observe pour (3b) :

- (11) 악령 *ag-lyeng* (mal-fantôme = diable)
 = 악이 (들어있는 + 지배하는) 영혼 *ag-i (deuleiss + jibaiha)-neun yenghon*
 (mal-nmtf (exister+dominer)-Sd âme = l'âme dominée par le mal)

on n'observe aucune paraphrase pour (1) :

- (12)a. 왕뱀 *wang-baim* (roi-serpent = serpent géant)
 = *왕의 뱀 **wang-eui baim* (roi-Gén serpent = serpent du roi)
 = *왕과 같은 뱀 **wang-gwa gat-eun baim*
 (roi-Postp.wa pareil-Sd serpent = serpent ressemblant à un roi)
- b. 악조건 *ag-jogen* (mal-condition = mauvaise condition)
 = *악의 조건 **ag-eui jogen* (mal-Gén condition = condition du mal)
 = *악이 지배하는 조건 **ag-i jibaiha-neun jogen*
 (mal-nmtf dominer-Sd condition = condition dominée par le mal)

Les noms (1) ne seront donc pas considérés comme des noms composés. Ici, on pourrait se demander pourquoi on doit attribuer absolument une étiquette préfixe à 왕 *wang* de (12), alors qu'on pourrait le considérer aussi comme un substantif, mais apparu dans un mot composé figé. La distinction entre "le préfixe d'un nom dérivé" et "le nom dans un mot composé figé" a pourtant certaines importances :

- cette décision remet en cause les définitions de noms affixés et de composés, voire la cohérence du recensement de noms affixés et de noms composés : certaines formes risquent d'être traitées tantôt comme des affixés tantôt comme des composés sans garder un principe cohérent et donc applicable à des tâches ultérieures ;

- ce souci n'est pas gratuit dans plusieurs situations concernant des homonymes (ou homographes) monosyllabiques, qui sont issus d'idéogrammes chinois. Examinons :

- (13)a. 학교 *hag-gyo* (école)
b. 요술 *yo-sul* (magie)

Ce sont des mots à deux syllabes en transcription coréenne, qui sont à l'origine deux idéogrammes chinois dont chacun possède un sens autonome. Beaucoup d'idéogrammes ont perdu leur autonomie syntaxique dans le lexique du coréen, la nature des mots ne doit pas être considérée en fonction de l'origine étymologique (si on classait des noms affixés ou des noms composés en fonction des idéogrammes chinois à l'origine, la grande majorité des noms coréens comportant plus de deux syllabes seront des composés, étant donné que la majorité des noms sont sino-coréens). Les mots (13) sont donc, pour nous, des noms simples dont aucune des deux syllabes ne sont autonomes.

Dans ce travail, nous n'entrerons pas dans les détails de ces questions. Il est certain que ce problème demandera une analyse plus raffinée et approfondie, ce qui fera l'objet de travaux ultérieurs.

3.3.4. Suffixe (*SF*) vs. Nom (*N*)

Dans la distinction entre le suffixe et le nom, on rencontre les mêmes problèmes que dans les cas précédents, à cause de l'emprunt de nombreux idéogrammes chinois dans le lexique des noms coréens. Ils créent un nombre considérable d'homographies entre monosyllabes. Nous utilisons, ici aussi, le même principe que celui qu'on a adopté pour la distinction entre le préfixe et le nom (dont l'idéogramme originel est identique à celui du préfixe), bien que ce principe doive être raffiné et complété par d'autres critères sémantico-syntaxiques : l'absence d'une paraphrase comportant l'élément en question en tant que substantif caractérisera des noms affixés vis-à-vis les noms composés. Prenons un exemple. Le mot 머리 *meli* a trois sens, entre autres, dans ces phrases suivantes :

- (1) a. 머리가 크다 *meli-ga(nmtf) keu-da(St)* (La tête est grande) : [*meli* = tête]
b. 머리가 부드럽다 *meli-ga(nmtf) budeuleb-da(St)*
(Les cheveux sont doux) : [*meli* = cheveu]
c. 머리가 둔하다 *meli-ga(nmtf) dunha-da(St)*
(Le degré d'intelligence est bas) : [*meli* = degré d'intelligence]

A partir de ces trois phrases élémentaires, on obtiendra les groupes nominaux de type *Adj-Sd N* (*N Adj*) comme :

- (2) a. 큰 머리 *keu-n(Sd) meli* (la grande tête)
b. 부드러운 머리 *budeuleu-n(Sd) meli* (les cheveux doux)

- c. 둔한 머리 *dunha-n(Sd) meli* (la tête dure = le degré d'intelligence bas)

Comparons maintenant (2) avec les formes suivantes qui, comportant la même unité 머리 *meli*, ne sont pas des séquences composées, mais des formes complexes qui comportent un élément non-autonome :

- (3) a. 대머리 *dai-meli* (PF-tête = chauve)
 b. 더벅머리 *debek-meli* (PF-cheveu = cheveux mal coupés)

Il s'agit de noms affixés dont la distribution est identique à celle du nom 머리 *meli* :

- (4) (민우 + 민우의 머리)-는 (대머리 + 더벅머리)-이다
 (*Minu + Minu-eui meli*)-neun (*dai-meli + debek-meli*)-i-da
 (Minu + Minu-Gén tête)-nmtf (chauve + cheveux mal coupés)-être-St
 ((Minu + Les cheveux de Minu) est(sont) (chauve(s) + des cheveux mal coupés))

comme les noms composés suivants :

- (5) a. 곱슬머리 *gobseul-meli* (frisé-cheveu = cheveux frisés)
 b. 파마머리 *pama-meli* (permanente-cheveu = cheveux ayant subi une permanente)
- (6) (인아 + 인아의 머리)-는 (곱슬머리 + 파마머리)-이다
 (*Ina + Ina-eui meli*)-neun (*gobseul-meli + pama-meli*)-i-da
 (Ina + Ina-Gén cheveu)-nmtf (frisé-cheveu + permanente-cheveu)-être-St
 ((Ina + Les cheveux de Ina) est(sont) des cheveux (frisés + permanentés))

Les phrases (4) et (6) montrent que (3) et (5), désignant une partie du corps humain 머리 *meli* (tête/cheveu/etc.), sélectionne comme sujet un nom humain métonymiquement employé ou une séquence classificatrice (*Nhum-eui Nclas* [Nclas de Nhum]). Comparons cette observation avec :

- (7) a. 진절머리 *jinjel-meli* (ras-le-bol)
 b. 소갈머리 *sogal-meli* (dignité)
- (8) a. 체신머리 *cheisin-meli* (élégance)
 b. 인정머리 *injeng-meli* (pitié)

Les mots (7) et (8) décrivent un état ou une qualité humaine, ils n'ont rien de commun avec le nom même 머리 *meli* (tête ou cheveux) :

- (9)a. *(인아 + 인아의 머리)-는 (진절머리 + 소갈머리)-이다
 *(*Ina + Ina-eui meli*)-neun (*jinjel-meli + sogal-meli*)-i-da
 *(Ina + Ina-Gén cheveu)-nmtf (ras-le-bol + dignité)-être-St
 ((Ina + Les cheveux de Ina) est(sont) un(e) (ras-le-bol + dignité))

- b. *(인아 + 인아의 머리)-는 (체신머리 + 인정머리)-이다
 *(Ina + Ina-eui meli)-neun (chesin-meli + injeng-meli)-i-ta
 *(Ina + Ina-Gén cheveu)-nmtf (élégance + pitié)-être-St
 ((Ina + Les cheveux de Ina) est(sont) une (élégance + pitié))

On les observe dans des constructions comme suit :

- (10)a. 민우는 그 일에 진절머리를 냈다
 Minu-neun geu il-e jinjel-meli-leul nai-ess-da
 Minu-nmtf ce affaire-Postp.e ra-le-bol-Acc montrer-Mpas-St
 (Minu a montré son ras-le-bol vis-à-vis de cette affaire)
- b. 민우는 소갈머리가 없다
 Minu-neun sogal-meli-ga eps-da
 Minu-nmtf dignité-Postp.i absent-St
 (Minu n'a pas de dignité)
- (11)a. 그 남자는 체신머리가 없다
 geu namja-neun chesin-meli-ga eps-da
 ce homme-nmtf élégance-Postp.i absent-St
 (Cet homme n'a pas d'élégance)
- b. 그 여자는 인정머리가 없다
 geu yeca-neun injeng-meli-ga eps-da
 ce femme-nmtf pitié-Postp.i absent-St
 (Cette femme n'a pas de pitié)

Ici, on retrouve les questions qui se sont posées auparavant : 머리 *meli* dans (7) et (8) sont-ils les substantifs (comme dans (3) et (6)), figés avec un autre élément ? Ou bien sont-ils des éléments non-autonomes qui n'ont aucun rapport avec le substantif 머리 *meli* ? Notre analyse reste relativement simple pour le moment : vu qu'il n'existe aucune paraphrase comportant le mot autonome 머리 *meli* pour les mots (7) et (8), nous considérons que ces mots contiennent un suffixe 머리 *-meli*, et donc que les mots (7) sont des noms simples, et (8) des noms suffixés (i.e. *N-meli*).⁸

⁸ Notons qu'il existe aussi d'autres types de séquences comportant 머리 *meli*, que nous ne traiterons pas ici. Soit :

- | | | | | |
|------|----|------|-----------------------|------------------------------------|
| (i) | a. | 책상머리 | <i>chaigsang-meli</i> | (table-tête = l'angle de la table) |
| | b. | 침대머리 | <i>chimtai-meli</i> | (lit-tête = le devant du lit) |
| (ii) | | 양지머리 | <i>yangji-meli</i> | (PF-tête = une partie du boeuf) |

3.3.5. Nom incomplet (NI) vs. Suffixe (SF)

Soulignons que le terme de nom incomplet (NI) désigne, pour nous, des formes qui exigent un modifieur de type :

- **Numéraux** (e.g. 두 *du* (deux), 다섯 *dases* (cinq) ou 삼십 *samsib* (trente))
- **Modifieurs phrastiques** (e.g. 민우가 떠날 [*Minu-ga ddena-l*] (que Minu part))

Voici trois types de noms incomplets :

- (1) a. 마리, 장, 명, 년
mali (Nclas : animal), *jang* (Nclas : feuille), *myeng* (Nclas : personne), *nyen* (année)
- b. 줄, 데, 수, 때문
jul (manière), *de* (endroit), *su* (moyen), *ddaimun* (raison)
- c. 면, 점, 쪽, 측
myen (côté), *jem* (point), *jjog* (direction), *cheug* (côté)

La **série (1a)** n'apparaît qu'avec des numéraux (Num). Ainsi, ces noms incomplets font partie des noms de mesure (Nmes). Soit :

- (2) a. 다섯 마리의 늑대 . *dases mali-eui neugdai*
(cinq Nmes[animal]-Gén loup = cinq loups)
- b. 늑대 다섯 마리 *neugdai dases mali*
(loup cinq Nmes[animal] = cinq loups)
- (3) a. 백 장의 종이 *paig jang-eui jongi* (cent Nmes[feuille]-Gén papier=cent feuilles)
- b. 종이 백 장 *jongi baik jang* (papier cent Nmes[feuille] = cent feuilles)

Les numéraux ne peuvent pas apparaître facilement sans ces noms incomplets :

- (4) ?다섯 늑대 ?*dases neugdai* (cinq loups)
- (5) ?*백 종이 ?**paig jongi* (cent feuilles)

La **série (1b)** demande un modifieur phrastique obligatoire :

- (6) a. 민우는 (친구들의 부탁을 거절할 +*E) 줄을 모른다
*Minu-neun (chingu-deul-eui butak-eul gejelha-l + *E) jul-eul moleu-nda*
Minu-nmtf [(ami-Mpl-Gén demande-Acc refuser-Sd + *E) **manière**]-Acc ignorer-St
(Minu ne sait pas comment (refuser les demandes des amis + E))
- b. 인아는 도저히 (민우를 떠날 +*E) 수가 없었다
*Ina-neun tojehi (Minu-leul ddena-l + *E) su-ga eps-ess-da*
Ina-nmtf vraiment [(Minu-Acc quitter-Sd + *E) **moyen**]-Postp.i absent-Mpas-St

(Ina ne trouvait pas le **moyen** (de quitter Minu + E))

Les **noms (1c)** sont accompagnés soit d'un numéral soit d'un modifieur phrastique :

- (7) a. 민우는 여자를 무시하는 면이 있다
Minu-neun yeca-leul musiha-neun myen-i iss-da
 Minu-nmtf [femme-Acc mépriser-Sd **côté**]-Postp.i avoir-St
 (Minu a un **côté** méprisant envers les femmes)
- b. 그 상자곽의 세 면이 빨간 색이다
geu sangjagwag-eui se myen-i bbalga-n saig-i-da
 ce boîte-Gén trois **côté**-nmtf rouge-Sd couleur-être-St
 (Trois **côtés** de cette boîte sont de couleur rouge)

Nous avons recensé les noms incomplets et les avons répertoriés dans une liste à part. Leur statut particulier (autonome syntaxiquement, mais non-autonome lexicosémantiquement) pose certains problèmes dans le recensement des noms affixés et composés. Soit par exemple :

- (8) a. 양쪽 *yang-jjog* (deux-côtés)
 b. 쓸데 *sseul-de* (utiliser-endroit = utilité)
 c. 탈것 *tal-ges* (monter-chose = moyen de transport)

Nous décrivons les formes (8) comme noms composés comportant des noms incomplets, et non noms affixés comportant des suffixes. Cependant, cette description demandera des examens plus rigoureux.

3.3.6. Numéral (*Num*) vs. Déterminant (*Dét*)

Dans les grammaires scolaires et les dictionnaires usuels, on distingue deux classes suivantes :

- (1) a. 수사 : 하나, 둘, 셋, 넷, 다섯
susa (Num) : hana (un), dul (deux), ses (trois), nes (quatre), dases (cinq), etc.
- b. 관형사 : 한, 두, 세, 네, 다섯
gwanhyengsa (Dét) : han(un), du(deux), se(trois), ne(quatre), dases(cinq), etc.

Les exemples suivants illustrent les phrases comportant un 수사 *susa* (Num) et un 관형사 *gwanhyengsa* (Dét) respectivement :

- (2) a. 그 사람의 나이는 서른이다
geu salam-eui nai-neun seleun-i-da
 ce homme-Gén âge-nmtf **trente**-être-St
 (L'âge de cet homme est **trente**)
- b. 그 사람은 서른 살이다
geu salam-eun seleun sal-i-da
 ce homme-nmtf **trente** an-être-St
 (Cet homme a **trente** ans)

Mais il nous apparaît qu'il s'agit plutôt de deux fonctionnements syntaxiques d'une seule catégorie grammaticale, et non deux catégories grammaticales différentes, comme nous n'attribuons pas deux catégories aux formes déterminatives et prédicatives d'adjectifs et de verbes, ou bien aux noms modifiant un autre et modifiés par un autre.

De plus, la variation de forme entre ceux qui sont notés *Num* et ceux qui sont considérés comme *Dét* se limite aux quatre premiers chiffres :

- (3) a. 하나 / 한 *hana* (un) / *han* (un)
 b. 둘 / 두 *dul* (deux) / *du* (deux)
 c. 셋 / 세 *ses* (trois) / *se* (trois)
 d. 넷 / 네 *nes* (quatre) / *ne* (quatre)

Plus précisément, tous les nombres terminés par un de ces quatre chiffres ont une autre forme quand ils modifient un substantif. Ainsi, comparons (4a) avec (4b) :

- (4) a. 스물에 (열둘 + 열다섯)-을 더하면, 합이 (서른둘 + 서른다섯)-이다
seumul-e (yel-dul+yel-dases)-eul deha-myen, hab-i (seleun-dul+seuleun-dases)-i-da
 20-Postp (12 + 15)-Acc ajouter-Sconj, total-nmtf (32 + 35)-être-St
 (20 plus (12 + 15) égalent (32 + 35))
- b. (열두 + 열다섯) 소년이 모두 왔다
(yel-du + yel-dases) sonyen-i modu o-ess-da
 [(12 + 15) garçon]-nmtf tous venir-Mpas-St
 (Les (12 + 15) garçons sont tous venus)

Nous considérons donc que les quatre numéraux (*i.e.* 1, 2, 3 et 4) et les numéraux terminés par ces quatre chiffres (*e.g.* 11, 12, 13, 14, 21, 22, 23, 24, etc.) prennent une forme déterminative quand ils apparaissent à gauche d'un substantif, alors que les autres numéraux restent invariables. La catégorie *Dét* ne contiendra donc pas de numéraux.

Les dictionnaires usuels sont incohérents dans la description de cette catégorie. D'abord, pour les numéraux qui sont, à l'origine, des idéogrammes chinois (nous les

appelons numéraux sino-coréens, par opposition aux numéraux coréens qu'on vient d'examiner) :

- (5) 일, 이, 삼, 사, 오
il (un), *i* (deux), *sam* (trois), *sa* (quatre), *o* (cinq), etc.

il n'y a aucune indication de leur emploi déterminatif. Si on observe, pour les Num coréens, une double étiquette [Num] et [Dét] :

- (6) 여섯 *yese* (six) [Num], [Dét]
 일곱 *ilgob* (sept) [Num], [Dét]

pour les Num sino-coréens, on n'en a qu'une [Num]. Comparons (7) avec (6) :

- (7) 육 *yug* (six) [Num]
 칠 *chil* (sept) [Num]

Par ailleurs, à partir de certains chiffres, l'étiquetage devient plus aléatoire comme le montre les exemples suivants⁹ :

- (8) a. 삼십 *sam-sib* (trois-dix = trente) [Num] / [Num], [N]
 b. 사십 *sa-sib* (quatre-dix = quarante) [N] / [Num], [N]
 c. 오십 *o-sib* (cinq-dix = cinquante) [Num] / [Num]

Vu que le nombre des numéraux est a priori illimité, il y a une certaine sélection des numéraux dans les dictionnaires éditoriaux. Mais on ne peut pas appliquer le même principe dans un dictionnaire électronique : au lieu de les enregistrer partiellement à titre exemplaire, il est nécessaire d'intégrer un automate fini, qui permet de reconnaître tous les numéraux dont le nombre est "illimité". La syntaxe des numéraux a une telle complexité qu'il faudra l'étudier sous forme d'une grammaire locale. Nous y reviendrons plus tard.

3.4. Lexique des affixes *DECO-AFX*

Le lexique des affixes *DECO-AFX* comporte des éléments non-autonomes dont le degré de productivité est très varié. Notre objectif ultérieur étant de répertorier aussi complètement que possible les lexiques des noms affixés et des noms composés, et non les simples listes des préfixes, suffixes et pseudo-noms, l'établissement de ces lexiques ne sera qu'une étape provisoire. Les pseudo-noms sont des éléments non autonomes d'origine sino-coréenne. Notons ici que le mécanisme de création des noms comportant

⁹ Cf. Voir [Sin90] et [Lee88].

des pseudo-noms est particulier et que ce mécanisme doit être étudié séparément des règles sous-jacentes aux dérivations par préfixes ou suffixes.

Etant donné que notre recensement des préfixes et suffixes est basé sur des principes purement formels (i.e. tous les composants de noms qui ne sont pas "syntaxiquement" autonomes), le résultat ne correspondra pas tout à fait à celui qu'on obtiendrait en listant les préfixes et suffixes, sur la base de leur étymologie, ou sur la base de leurs traits lexico-sémantiques. Ainsi, notre lexique comportera certainement des éléments non-autonomes dont l'existence a été négligée, dans des grammaires scolaires, à cause de leur faible productivité. Par contre, certains éléments considérés, dans la tradition littéraire, comme suffixes ou préfixes (pour des raisons étymologiques ou d'autres), mais qui ne s'observent qu'avec des éléments non-autonomes ne seront, pour nous, ni des suffixes ni des préfixes : les formes comportant ces éléments seront traitées comme noms simples.

Rappelons que le lexique DECO-AFX est constitué de trois sous-lexiques : DECO-PF, DECO-SF et DECO-PN. Les PN (Pseudo-noms) indiquent des unités sino-coréennes à deux syllabes qui n'ont pas d'autonomie syntaxique : ils ne s'observent qu'en association avec un autre substantif (cf. voir 3.6.3).

Dans le lexique des préfixes *DECO-PF*, on trouvera les codes suivants :

PF	Préfixe
PREDH	Préfixe rendant un substantif <i>prédicatif</i> , l'ensemble peut être associé à <i>hada</i> (i.e. <i>PF-N-hada</i>)
CNI	Préfixe qui se combine avec un nom incomplet (i.e. <i>PF-NI</i>)

Le lexique des pseudo-noms à deux syllabes *DECO-PN* ne comporte, pour le moment, que le code PN :

PN	Pseudo-noms à deux syllabes
-----------	-----------------------------

Le lexique des suffixes *DECO-SF* contient les codes suivants :

SF	Suffixe
ANM	Suffixe de construction d'un nom d' <i>animal</i>
HUM	Suffixe de construction d'un nom <i>humain</i>
PLT	Suffixe de construction d'un nom de <i>plante</i>
PREDH	Suffixe rendant un substantif <i>prédicatif</i> , l'ensemble peut être associé à <i>hada</i> (i.e. <i>N-SF-hada</i>)

3.5. Typologie des noms composés

Les noms composés de deux unités autonomes pourront être formellement classés de la manière suivante (ici, # signifie un blanc ; *E* indique l'absence du blanc, alors les deux unités seront soudées) :

- (1) a. $N-(\# + E)-N$
 b. $[V-Mnom]-(\# + E)-N$
 c. $N-(\# + E)-[V-Mnom]$
 d. $[Adj-Mnom]-(\# + E)-[V-Mnom]$
 e. $[V-Mnom]-(\# + E)-[V-Mnom]$
 f. $[Adj-Sd]-(\# + E)-N$
 g. $[V-Sd]-(\# + E)-N$
 h. $Adv-N$
 i. $Adv-[V-Mnom]$

voici des exemples :

- (2) a. 나이테 *nai-te* (âge-anneau = anneau d'accroissement)
 물거품 *mul-gepum* (eau-écume = écume d'eau)
 b. 맺음말 *maij-eum-mal* (terminer-parole = conclusion)
 쉼표 *swi-m-pyo* (pause-symbole = virgule)
 c. 목걸이 *mog-gel-i* (cou-porter = collier)
 발걸음 *bal-gel-eum* (pas-marcher = pas)
 d. 넓이뛰기 *nelb-i-ddwi-gi* (large-sauter = saut en longueur)
 높이뛰기 *nop-i-ddwi-gi* (haut-sauter = saut en hauteur)
 e. 걸음걸이 *gel-eum-gel-i* (marcher-marcher = marche)
 뛸뛰기 *ddwi-m-ddwi-gi* (courir-courir = course)
 f. 신물 *si-n-mul* (acide-eau = ras-le-bol)
 큰형 *keu-n-hyeng* (grand-frère = grand frère)
 g. 죽을힘 *juk-eul-him* (mourir-force = efforts à la mort)
 뜬소문 *ddeu-n-somun* (flotter-bruit = faux bruit)
 h. 껌소리 *ggig-soli* (aïe-cri = cri de douleur)
 나란히꼐 *nalanhi-ggol* (parallèlement-forme = parallélogramme)
 i. 멀리뛰기 *melli-ddwi-gi* (loin-sauter = saut en longueur)
 깔깔웃음 *ggalggal-us-eum* (éclat-rire = éclat de rire)

Notons que toutes les formes (1) peuvent être regroupées en quatre types de composés : $N-N$ (i.e. (1a), (1b), (1c), (1d) et (1e)), $[Adj-Sd]-N$ (i.e. (1f)), $[V-Sd]-N$ (i.e. (1g)) et $Adv-N$ (i.e. (1h) et (1i)). Quant aux formes $Adv-N$ ou $Adv-[V-Mnom]$, étant donné que l'ordre et l'association des constituants ne sont pas habituels, leur identification est relativement simple. Par contre, les trois premiers types ($N-N$, $[Adj-Sd]-N$ et $[V-Sd]-N$) ont une

structure formellement identique à celle des groupes nominaux *libres*. D'où la difficulté du recensement de ces types de noms.

Nous nous limiterons, dans ce travail, à montrer quelques difficultés rencontrées dans la délimitation des noms composés de ces trois types.

3.6. Noms composés de type *[Adj-Sd] N* et *[V-Sd] N*

Reprenons les exemples cités plus haut :

- (1) a. 신물 *si-n-mul* (acide-eau = ras-le-bol)
 큰형 *keu-n-hyeng* (grand-frère = grand frère)
- b. 뜬소문 *ddeu-n-somun* (flotter-bruit = faux bruit)
 죽을힘 *juk-eul-him* (mourir-force = efforts à la mort)

Ils semblent plus lexicalisés que les séquences suivantes dont la structure est identique :

- (2) a. 신사과 *si-n-sagwa* (acide-pomme = pomme acide)
 큰가방 *keu-n-gabang* (grand sac)
- b. 군만두 *gu-n-mandu* (fire-ravioli = ravioli frit)
 지난가을 *jina-n-gaeul* (passer-automne = automne passé)

Les mots (1) sont de nouveaux termes dont le sens n'est pas prévisible à partir de ceux de ces constituants : 신물 *si-n-mul* signifie l'ennui (ras-le-bol) et 큰형 *keu-n-hyeng* indique un frère aîné plutôt qu'un frère dont la taille est grande, alors que les mots (2a) ne présentent aucun sens supplémentaire comme (1a). Il en est de même avec (1b) : 뜬소문 *ddeu-n-somun* ne garde plus le sens ordinaire du verbe 뜨다 *ddeuda* (flotter), mais signifie un faux bruit ; et 죽을힘 *juk-eul-him* n'est pas la force qui va mourir, mais les efforts qu'on fait à la mort. Par contre, les mots (1b) semblent être des groupes nominaux décomposables, bien qu'ils soient enregistrés comme entrées dans les dictionnaires usuels :

- (3) 군만두 = 만두를 굽다
 gu-n-mandu = *mandu-leul gupda*
 (frيره-Sd ravioli (ravioli frit) = frيره des raviolis)
- (4) 지난가을 = 가을이 지나다
 jina-n-gaeul = *gaeul-i jinada*
 (passer-automne (automne passé) = l'automne passe)

Par ailleurs, on peut substituer à chacun des constituants de (2b) une série de paradigmes de type :

- (5) a. (군 + 찜 + 빗 + 고기를 넣)-은 만두
 (gu + jji + bij + gogi-leul neh)-(eu)n mandu
 (frir + cuire + faire + viande-Acc mettre)-Sd ravioli
 (ravioli (frit + cuit (à la vapeur) + fait (à la main) + fourré à la viande))
- b. 군 (만두 + 사과 + 감자)
 gu-n (mandu + sagwa + gamja)
 frir-Sd (ravioli + pomme + pomme de terre)
 ((ravioli + pomme + pomme de terre) frit(e))
- (6) a. (지난 + 짧은 + 3 년전의) 가을
 (jina-n + jjalb-eun + 3 nyen-jen-eui) gaeul
 (passer-Sd + court-Sd + 3 an-avant-Gén) automne
 (automne (passé + court + d'il y a 3 ans))
- b. 지난 (가을 + 봄 + 해)
 jina-n (gaeul + bom + hai)
 passer-Sd (automne + printemps + année)
 ((automne + printemps + année) passé(e))

On peut insérer un élément à l'intérieur des séquences (2), ce qui est interdit pour (1) :

- (7) a. 군 고기 만두 gu-n gogi mandu
 frir-Sd viande ravioli (= ravioli de viande frit)
- b. 지난 빠리에서의 가을 jina-n Paris-ese-eui gaeul
 passer-Sd Paris-Postp.ese-Gén automne (= automne passé à Paris)
- (8) a. *신 과일 물 *si-n gwail mul (acide-Sd fruit eau)
 b. *큰 멋진 형 *keu-n mesji-n hyeng (grand-Sd chic-Sd frère)
- (9) a. *뜬 그 소문 *ddeu-n geu somun (flotter-Sd ce bruit)
 b. *죽을 민우의 힘 *jug-eul Minu-eui him (mourir-Sd Minu-Gén force)

Comme nous l'avons mentionné plus haut, la décision simple de savoir s'il s'agit de noms composés, par opposition à des groupes nominaux libres n'a pas de sens en tant que telle (du moins dans la construction d'un dictionnaire électronique), vu que la grande majorité des séquences composées se trouvent en fait entre ces deux extrêmes. Tout ce continuum doit être décrit en détail à l'aide de critères syntaxiques et formels.

3.7. Noms composés de type *NN*

3.7.1. Analyse structurale des *NN*

Les séquences composées de type *NN* constituent la grande majorité des noms composés. Ici aussi, en fonction de leur degré de figement, ces séquences forment un continuum entre les noms composés figés et les groupes nominaux libres. Les relations entre deux constituants sont plus complexes que dans le cas des composés [*Adj-Sd*]-*N* et [*V-Sd*]-*N* où le premier élément, sauf les cas très lexicalisés, joue le rôle d'élément prédicatif et où le deuxième joue le rôle d'argument (soit sujet soit complément de l'élément prédicatif).

Dans ce travail, nous n'entrerons pas dans le détail de la description et de la classification des noms composés. Mais il est évident que dans la construction d'un dictionnaire électronique des noms composés, on ne pourra pas éviter des problèmes fondamentaux dans la délimitation des noms composés vis-à-vis des groupes nominaux libres, qu'on estime ne pas pouvoir répertorier dans un premier temps. Il nous apparaît, répétons-le, que la distinction dichotomique entre les noms composés et les groupes nominaux n'a pas d'intérêt, à moins que toutes les séquences *NN* soient formellement décrites et classées, qu'elles soient appelées noms composés ou groupes nominaux.

Voici les principales relations qui existent entre deux *N* constituant la structure *NN*. Ces relations permettront de classer l'ensemble des *NN* de façon formelle :

$$\begin{aligned} [1] \text{ a. } Na-Nb &= (Na)_0 (Nb)\text{-hata} [(Na)_0 \text{ faire } (Nb)] \\ &= Na\text{-eui } Nb \text{ [Nb de Na]} \end{aligned}$$

대통령연설 = 대통령이 연설한다
daitonglyeng-yensel = *daitonglyeng-i yensel-ha-nda*
 président-discours = **président-nmtf discours-faire-St**
 (Discours du président = Le président fait un discours)

대통령연설 = 대통령의 연설
daitonglyeng-yensel = *daitonglyeng-eui yensel*
 président-discours = **président-Gén discours**
 (Discours du président = Discours **du** président)

- b. Na-Nb = (Na)₀ N₁-Acc (Nb)-hata [(Na)₀ faire (Nb) Prép N₁]¹⁰
 = Na-eui Nb [Nb de Na]

경찰수사 = 경찰이 사건을 수사한다
 gyengchal-susa = **gyengchal-i** sagen-eul **susa-ha-nda**
 police-enquête = **police-nmtf** affaire-Acc **enquête-faire-St**
 (Enquête de la police = La police fait une enquête sur cette affaire)

경찰수사 = 경찰의 수사
 gyengchal-susa = **gyengchal-eui** susa
 police-enquête = **police-Gén** enquête
 (Enquête de la police = Enquête de la police)

- [2] a. Na-Nb = N₀ (Na)-Acc (Nb)-hata [N₀ (Na)₁ faire (Nb)]
 = Na-eui Nb [Nb de Na]¹¹

국회의원선거 = (국민이) 국회의원을 선출한다
 gughoi euiwen-senchul = (gugmin-i) **gughoi euiwen-eul** **senchul-ha-nda**
 député-élection = (peuple-nmtf) **député-Acc** **élection-faire-St**
 (Election des députés = (Le peuple) élit les députés)

국회의원선거 = 국회의원의 선출
 gughoi euiwen-senchul = **gughoi euiwen-eui** **senchul**

¹⁰ Il faut souligner que l'inverse n'est pas toujours vrai : le sujet d'une phrase transitive admet la relation NaNb = Na N₁-Acc Nb-hata plus difficilement que Na en position du complément. Soit :

- (i) ?독일파괴 ≠ 독일이 그 나라를 파괴했다
 ?dogil-pakoi ≠ **dogil-i** geu nala-leul **pakoi-ha-ess-da**
 ?Allemagne-destruction ≠ **Allemagne-nmtf** ce pays-Acc **destruction-faire-Mpas-St**
 (Destruction de l'Allemagne ≠ L'Allemagne a fait la destruction de ce pays)

La séquence même *dogil-pakoi* peut être acceptée, mais avec l'interprétation "C'est l'Allemagne qui a été détruite".

¹¹ Notons que quand Nb est morphologiquement relié à un adjectif, la construction NaNb (ou Na-(Adj-Mnom)b) est interdite. Seule la forme Na-eui Nb (ou Na-eui (Adj-Mnom)b) est autorisée :

- (i) *Na-Nb = (Na)₀ ((Nb)-hata)_{Adj} [(Na)₀ être Adj]
 = Na-eui Nb [Nb de Na]
 (ii) *Na-(Adj-Mnom)b = (Na)₀ Adj [(Na)₀ être Adj]
 = Na-eui (Adj-Mnom)b [(Adj-Mnom)b de Na]

Soit par exemple :

- (iii) *여자불행 = 여자는 불행하다 = 여자의 불행
 *yeja-bulhaing = **yeja-neun** **bulhaing-ha-da** = **yeja-eui** **bulhaing**
 *femme-malheur = **femme-nmtf** **malheur-hata-St** = **femme-Gén** **malheur**
 (Malheur de femme = La femme est malheureuse = Malheur de femme)
 (iv) *주인상냥함 = 주인이 상냥하다 = 주인의 상냥함
 *juin-(sangnyangha-m) = **juin-i** **sangnyangha-da** = **juin-eui** **sangnyangha-m**
 *patron-(gentil-Mnom) = **patron-nmtf** **gentil-St** = **patron-Gén** **gentil-Mnom**
 (Gentillesse du patron = Le patron est gentil = Gentillesse du patron)

député-élection = député **-Gén** élection
(Election des députés = Election **des** députés)

$$\begin{aligned} \text{b. Na-(V-Mnom)b} &= N_0 \text{ (Na)-Acc } V \text{ [N}_0 \text{ (Na)}_1 \text{ V]}^{12} \\ &= *Na\text{-eui (V-Mnom)b [(V-n)b de Na]} \end{aligned}$$

구두닦기 = (민우가) 구두를 닦는다
gudu-(dagg-gi) = (*Minu-ka*) *gutu-leul dagg-neunda*
chaussure-(polir-Mnom) = (Minu-nmtf) **chaussure-Acc** **polir-St**
(Polissage de chaussures = (Minu) polit des chaussures)

구두닦기 = *구두의 닦기
gudu-(dagg-gi) = **gudu-eui dagg-di*
chaussure-(polir-Mnom) = *chaussure-**Gén** polir-Mnom
(Polissage de chaussures = Polissage **de** chaussures)

$$\begin{aligned} [3] \text{ a. Na-Nb} &= (Nb)_0 \text{ (Na)-ita [(Nb)}_0 \text{ être (Na)}_1\text{]} \\ &= *Na\text{-eui Nb [Nb de Na]} \end{aligned}$$

가짜돈 = 그 돈은 가짜다
gajja-don = *keu don-eun gajja-i-da*
faux-billet = ce **billet-nmtf** **faux-être-St**
(Faux billet = Ce billet est faux)

가짜돈 = *가짜의 돈
gajja-don = **gajja-eui don*
faux-billet = *faux-**Gén** billet
(Faux billet = Billet **de** faux [faux billet])

$$\begin{aligned} \text{b. Na-Nb} &= (Na)_0 \text{ (Nb)-ita [(Na)}_0 \text{ être (Nb)}_1\text{]} \\ &= *Na\text{-eui Nb [Nb de Na]} \end{aligned}$$

살인죄 = 살인은 죄다
salin-joi = *salin-eun joi-i-da*
meurtre-crime = **meurtre-nmtf** **crime-être-St**
(Crime de meurtre [Assassinat] = Le meurtre est un crime)

살인죄 = *살인의 죄
salin-joi = **salin-eui joi*

¹² Soulignons que cette relation (i.e. *Nb* est une nominalisation du verbe de la phrase $N_0 \text{ } N_1\text{-Acc } V$) ne s'observe pas quand *Na* est le sujet de la phrase de base (i.e. $Na \text{ } N_1\text{-Acc } V$).

meurtre-crime = *meurtre-**Gén** crime
 (Crime de meurtre [Assassinat] = Crime de meurtre)

- [4] a. Na-Nb = (Na)₀ (Nb)-Acc *kacikoissta* [(Na)₀ avoir (Nb)₁]
 = Na-eui Nb [Nb de Na]

여우털 = 여우는 털을 가지고 있다
yeu-tel = *yeu-neun tel-eul gajigoiss-da*
 renard-poil = renard-nmtf poil-Acc avoir-St
 (Poils de renard = Le renard a des poils)

여우털 = 여우의 털
yeu-tel = *yeu-eui tel*
 renard-poil = renard-**Gén** poil
 (Poils de renard = Poils **de** renard)

- b. Na-Nb = (Nb)₀ (Na)-Acc *kacikoissta* [(Nb)₀ avoir (Na)₁]
 = *Na-eui Nb [Nb de Na]¹³

독버섯 = 버섯이 독을 가지고 있다
tog-beses = *beses-i dog-eul gajigoiss-da*
 poison-champignon = champignon-nmtf poison-Acc avoir-St
 (Champignon vénéneux = Ce champignon a du poison)

독버섯 = *독의 버섯
tog-peses = **tog-eui beses*
 poison-champignon = *poison-**Gén** champignon
 (Champignon vénéneux = Champignon **de** poison)

- [5] Na-Nb = (Nb)₀ (Na)-Postp *Vsup* [(Nb)₀ *Vsup* Prép (Na)₁]
 = (Na)-Postp *Vsup-Sd* (Nb) [(Nb) *Vsup-pp* Prép (Na)]
 = *Na-eui Nb [Nb de Na]

¹³ Notons que, pour cette structure Na-Nb, on observe qu'il existe une séquence de type Nb-eui Na [Na de Nb]. Mais cette séquence n'est pas équivalente à la structure de départ :

독버섯 ≠ 버섯의 독
tog-beses ≠ *beses-eui dog*
 poison-champignon ≠ champignon-**Gén** poison
 (Champignon vénéneux ≠ Poison **du** champignon)

a. (Nb)₀ (Na)-lo toi eissta [(Nb)₀ être fait en (Na)₁]

고무공 = 그 공은 고무로 되어 있다

gomu-gong = *geu gong-eun gomu-lo doi eiss-da*

caoutchouc-ballon = ce **ballon**-nmtf **caoutchouc**-Postp.lo être fait-St

(Ballon en caoutchouc = Ce ballon est en caoutchouc)

고무공 = 고무로 되어 있는 공

gomu-gong = *gomu-lo doi eiss-neun gong*

caoutchouc-ballon = **caoutchouc**-Postp.lo être fait-Sd **ballon**

(Ballon en caoutchouc = Ballon fait en caoutchouc)

고무공 = *고무의 공

gomu-gong = **gomu-eui gong*

caoutchouc-ballon = *caoutchouc-**Gén** **ballon**

(Ballon en caoutchouc = Ballon de caoutchouc)

b. (Nb)₀ (Na)-e (issta+salta+etc.) [(Nb)₀ (exister+vivre+etc.) Loc (Na)₁]

들새 = 그 새는 들에 산다

deul-sai = *geu sai-neun deul-e sa-nda*

champs-oiseau = ce **oiseau**-nmtf **champs**-Postp.e vivre-St

(Oiseau des champs = Cet oiseau vit dans les champs)

들새 = 들에 사는 새

deul-sai = *deul-e sa-neun sai*

champs-oiseau = **champs**-Postp.e vivre-Sd **oiseau**

(Oiseau des champs = Oiseau qui vit dans les champs)

들새 = ?*들의 새

deul-sai = *? *deul-eui sai*

champs-oiseau = ?*champs-**Gén** **oiseau**

(Oiseau des champs = Oiseau des champs)

c. (Nb)₀ (Na)-e (issta+ota+etc.) [(Nb)₀ (exister+arriver+etc.) à (Na)₁]

가을비 = 그런 비는 가을에 온다

gaeul-bi = *geulen bi-neun gaeul-e o-nda*

automne-pluie = ce **pluie**-nmtf **automne**-Postp.e arriver-St

(Pluie d'automne = (Ce genre de) pluie tombe en automne)

가을비 = 가을에 오는 비

gaeul-bi = *gaeul-e o-neun bi*

automne-pluie = **automne**-Postp.e arriver-Sd **pluie**
 (Pluie d'automne = Pluie qui tombe en automne)

가을비 = ?* 가을의 비
gaeul-bi = ?**gaeul-eui bi*
 automne-pluie = ?*automne-**Gén** pluie
 (Pluie d'automne = Pluie d'automne)

d. (Nb)₀ (Na)-lo *Vsup* [(Nb)₀ *Vsup* Prép (Na)₁]

겉모양 = 그 모양이 겉으로 보인다
get-moyang = *geu moyang-i ket-eulo boi-nda*
 extérieur-forme = ce **forme**-nmtf **extérieur**-Postp.lo manifester-St
 (Forme extérieure = Cette forme se voit de l'extérieur)

겉모양 = 겉으로 보이는 모양
ket-moyang = *ket-eulo boi-neun moyang*
 extérieur-forme = **extérieur**-Postp.lo manifester-Sd **forme**
 (Forme extérieure = Cette forme qui se voit de l'extérieur)

겉모양 = ?*겉의 모양
ket-moyang = ?**ket-eui moyang*

extérieur-forme = ?*extérieur-**Gén** forme
 (Forme extérieure = Forme de l'extérieur)

[6] Na-Nb = (Nb)₀ (Na)-(katta+moyang-ita)[(Nb)₀ être (comme+de forme) (Na)₁]
 = (Na)-(kat-eun+moyang-i-n) (Nb) [(Nb) (comme+de forme) (Na)]
 = *Na-eui Nb [Nb de Na]

구슬땀 = 땀이 구슬-(같+모양이)-다
guseul-ddam = *ddam-i guseul-(kat + moyang-i)-da*
 perle-sueur = **sueur**-nmtf **perle**-(comme + forme-être)-St
 ((Goutte de) sueur de forme perle=Les gouttes de sueur sont (comme+de forme) perle)

구슬땀 = 구슬-(같은 + 모양인) 땀
guseul-ddam = *guseul-(kat-eun + moyang-i-n) ddam*
 perle-sueur = **perle**-(comme-Sd + forme-être-Sd) **sueur**
 ((Goutte de) sueur de forme perle = Gouttes de sueur (comme+de forme) perle)

구슬땀 = *구슬의 땀
guseul-ddam = **guseul-eui ddam*
 perle-sueur = *perle-**Gén** sueur

((Goutte de) sueur de forme perle = Les gouttes de sueur de perle)

- [7] Na-Nb = (Nb)₀ (Na)-(sik+pung)-ita [(Nb)₀ être de style de (Na)₁]
 = (Na)-(sik+pung) (Nb) [(Nb) de style de (Na)]
 = *Na-eui Nb [Nb de Na]

프랑스식당 = 그 식당은 프랑스-(식+풍)-이다
peulangseu-sigdang = *geu sigdang-eun peulangseu-(sig+pung)-i-da*
 France-restaurant = ce restaurant-nmtf France-(style)-être-St
 (Restaurant français = Ce restaurant est de style français)

프랑스식당 = 프랑스-(식+풍) 식당
peulangseu-sigdang = *peulangseu-(sig+pung) sigdang*
 France-restaurant = France-(style) restaurant
 (Restaurant français = Restaurant de style français)

프랑스식당 = *프랑스의 식당
peulangseu-sigdang = **peulangseu-eui sigdang*
 France-restaurant = *France-Gén restaurant
 (Restaurant français = Restaurant de France)

- [8] Na-Nb = (N₀) (Na)-leul kaciko (Nb)-Acc Vsup [(N₀) Vsup (Nb)₁ avec (Na)₂]¹⁴
 = (Na)-leul kaciko Vsup-Sd (Nb) [(Nb) Vsup-pp avec (Na)]
 = *Na-eui Nb [Nb de Na]

칼춤 = (여자가) 칼을 가지고 춤을 춘다
kal-chum = (*yeja-ga*) *kal-eul gajigo chum-eul chu-nda*
 canif-danse = (femme-nmtf) canif-avec danse-Acc danser-St
 (Danse avec les canifs = Cette femme fait une danse avec les canifs)

칼춤 = 칼을 가지고 추는 춤
kal-chum = *kal-eul gajigo chu-neun chum*
 canif-danse = canif-avec danser-Sd danse

¹⁴ Quand Na se trouve en position d'un complément de types N-ei ou N-wa, la structure NaNb n'est pas admise. Soit par exemple :

- (i) a. 친구편지 /= 민우는 친구에게 편지를 했다
chingu-phyenji /= *Minu-neun chingu-ege phyenji-leul ha-ess-da*
 ami-lettre /= Minu-nmtf ami-Postp.e lettre-Acc faire-Mpas-St
 (Lettre de l'ami /= Minu a écrit une lettre à un ami)
 b. ??친구말다툼 /= 민우는 친구와 말다툼을 했다
 ??*chingu-maldatum* /= *Minu-neun chingu-wa maldatum-eul ha-ess-da*
 ??ami-dispute /= Minu-nmtf ami-Postp.wa dispute-Acc faire-Mpas-St
 (Dispute de l'ami /= Minu a fait dispute avec son ami)

(Danse avec les canifs = Danse faite avec les canifs)

칼춤 = *칼의 춤

kal-chum = **kal-eui chum*

canif-danse = *canif-Gén danse

(Danse avec les canifs = Danse des canifs)

- [9] Na-Nb = (N₀) (Na)-ese (Nb)-Acc V^{sup} [(N₀) V^{sup} (Nb)₁ Loc (Na)₂]
 = (Na)-ese V^{sup}-Sd (Nb) [(Nb) V^{sup}-pp Loc (Na)]
 = *Na-eui Nb [Nb de Na]

뒷거래 = (그 사람들이) 뒤에서 거래를 한다

dwis-gelai = (*geu salamdeul-i*) *dwi-ese gelai-leul ha-nda*

derrière-marché = (ce gens-nmtf) derrière-Postp.ese marché-Acc faire-St

(Marché de derrière[noir] = Ces gens font du marché par derrière)

뒷거래 = 뒤에서 하는 거래

dwis-gelai = *dwi-ese ha-neun gelai*

derrière-marché = derrière-Postp.ese faire-Sd marché

(Marché de derrière[noir] = Marché fait par derrière)

뒷거래 = *뒤의 거래

dwis-gelai = **dwi-eui gelai*

derrière-marché = *derrière-Gén marché

(Marché par derrière[noir] = Marché de derrière)

- [10] Na-Nb = (Nb)₀ (Na)-Postp V-ki-wihan kes-ita
 [(Nb)₀ être pour V-inf Prép (Na)₁]
 = (Na)-Postp V-ki-wihan (Nb) [(Nb) pour V-inf Prép (Na)]
 = *Na-eui Nb [Nb de Na]

술통 = 그 통은 술을 넣기위한 것이다

sul-tong = *keu tong-eun sul-eul neh-gi-wihan ges-i-da*

vin-tonneau = ce tonneau-nmtf vin-Acc mettre-pour chose-être-St

(Tonneau de vin = Ce tonneau est pour mettre du vin)

술통 = 술을 넣기위한 통

sul-tong = *sul-eul neh-gi-wihan tong*

vin-tonneau = vin-Acc mettre-pour tonneau

(Tonneau de vin = Tonneau pour mettre du vin)

술통 = *술의 통

sul-tong = **sul-eui tong*

vin-tonneau = *vin-**Gén** tonneau
(Tonneau de vin = Tonneau de vin)

- [11] Na-Nb = (Nb)₀ (Na)-*etaihan kes-ita* [(Nb)₀ être sur (Na)₁]
= (Na)-*etaihan* (Nb) [(Nb) sur (Na)]
= *Na-*eui* Nb [Nb de Na]

영어사전 = 그 사전은 영어에 대한 것이다
yenge-sajen = *geu sajen-eun yenge-eidaihan ges-i-da*
anglais-dictionnaire = ce **dictionnaire**-nmtf **anglais**-sur chose-être-St
(Dictionnaire de l'anglais = Ce dictionnaire est sur l'anglais)

영어사전 = 영어에 대한 사전
yenge-sajen = *yenge-eidaihan sajen*
anglais-dictionnaire = **anglais**-sur **dictionnaire**
(Dictionnaire de l'anglais = Dictionnaire sur l'anglais)

영어사전 = *영어의 사전
yenge-sajen = **yenge-eui sajen*
anglais-dictionnaire = *anglais-**Gén** dictionnaire
(Dictionnaire de l'anglais = Dictionnaire de l'anglais)

- [12] Na-Nb = *Constructions résiduelles*

고추잠자리	<i>gochu-jamjali</i>	(piment-libellule = libellule de couleur piment)
길동무	<i>gil-dongmu</i>	(chemin-compagnon = compagnon de voyage)
달나라	<i>dal-nala</i>	(lune-monde = lune)
눈꼴	<i>nun-ggol</i>	(oeil-forme = apparence)
발걸음	<i>bal-geleum</i>	(pied-marche = pas)
밤낮	<i>bam-nac</i>	(nuit-jour = toujours)

3.7.2. Formes comportant un PN

Examinons :

- (1) a. 민주 사회 *minju sahoi* (société démocratique)
b. 잠정 조치 *jamjeng jochi* (décision provisoire)
c. 정기 집회 *jenggi jibhoi* (réunion régulière)
d. 장기 치료 *janggi chilyo* (traitement de longue durée)
e. 공산 정권 *gongsan jenggwen* (régime communiste)

Ces mots sont composés de deux unités sino-coréennes à deux syllabes dont la première est un pseudo-nom qui n'a pas d'autonomie syntaxique (*i.e.* structure *PN N*). Ainsi, par exemple :

- (2) a. *민우는 민주를 지지한다
 *Minu-neun **minju**-leul jijiha-nda
 *Minu-nmtf démocratie-Acc soutenir-St
 (Minu soutient la démocratie)
- b. *회원들은 집회의 정기를 원한다
 *hoiwen-deul-eun jibhoi-eui **jenggi**-leul wenha-nda
 *membre-Mpl-nmtf réunion-Gén régularisation-Acc désirer-St
 (Les membres désirent une régularisation de la réunion)
- c. *공산이 ^우리의 목표이다
 ***gongsan**-i uli-eui mogpyo-i-da
 *communisme -nmtf nous-Gén but-être-St
 (Le communisme est notre but)

Etant donné que ces formes sont syntaxiquement non-autonomes, doit-on les considérer comme des préfixes ? La réponse est un choix terminologique, mais le consensus sur la notion de préfixe nous fait hésiter à les intégrer dans la classe des préfixes. Nous avons recensé tous les éléments sino-coréens à deux syllabes qui n'ont aucune autonomie syntaxique : nous les appelons pseudo-noms à deux syllabes (codés *PN*), ils sont répertoriés à part dans le lexique des pseudo-noms à deux syllabes DECO-PN. La comparaison entre les séries suivantes révèle la productivité des séquences comportant des *PN*, aussi importante que celle des noms composés de type *NN* :

- (3) a. 민주 정치 **minju**-jengchi (*PN-N* = régime démocratique)
 b. 민주 제도 **minju**-jedo (*PN-N* = système démocratique)
 c. 민주 국가 **minju**-gugga (*PN-N* = pays démocratique)
 d. 민주 사상 **minju**-sasang (*PN-N* = pensée démocratique)
- (4) a. 자유 정치 **jayu**-jengchi (*N-N* = régime libéral)
 b. 자유 제도 **jayu**-jedo (*N-N* = système libéral)
 c. 자유 국가 **jayu**-gugga (*N-N* = pays libéral)
 d. 자유 사상 **jayu**-sasang (*N-N* = pensée libérale)

Les *PN* sont susceptibles de constituer des chaînes de noms (sino-coréens), la répétition de cette opération étant a priori sans limites :

- (5) a. 공산 독재 체제 **gongsan dogjai chejai**
 (système de dictature communiste)
 b. 자유 민주 주의 실현 **jayu minju jueui silhyen**

(réalisation de la société démocratique)

- c. 원시 공동체 사회 구성 요건 *wensi gongdongche sahoi guseng yegen*
(condition pour la formation d'une société commune primitive)

Il est à noter que le recensement des pseudo-noms à deux syllabes pose le problème des formes dites locutions sino-coréennes, qui comportent, en général, quatre syllabes. Par exemple :

- (6) a. 오비 이락 *obi ilag* (Le corbeau s'envole, la poire tombe)
 b. 세월 여류 *sewel yelyu* (L'écoulement du temps est comme celui de l'eau)
 c. 겔초 보은 *gyelcho boeun* (remerciement pour une générosité)
 d. 남녀 유별 *namnye yubyel* (L'homme et la femme ont des différences)

Les quatre syllabes de chaque séquence (6) s'écrivent avec ou sans blanc au milieu (i.e. entre les deux parties à deux syllabes), les deux parties n'ayant en général aucune autonomie syntaxique: le figement entre ces deux parties est, d'ailleurs, extrêmement élevé. Ces parties à deux syllabes doivent-elles faire l'objet de notre premier recensement de PN ? Il nous apparaît que ces éléments doivent être étudiés à part sous forme d'une grammaire locale. D'une part, il est difficile de construire une phrase élémentaire comportant des séquences composées de type (6). En effet, ce sont des locutions figées souvent *phrastiques*, qui énoncent des histoires, des légendes ou des proverbes d'origine chinoise ; et d'autre part, il est rare d'observer des paradigmes substituables à des constituants de ces expressions. Il est également difficile de former des séquences composées plus étendues à partir de ces locutions, comme (5). Cette caractéristique nous fournit un critère provisoire, mais opératoire : nous n'avons ainsi répertorié comme PN que ceux qui sont susceptibles d'entrer dans une construction étendue de type (5). Les séquences (6) qui admettent difficilement une extension de type (5) seront donc considérées comme des locutions figées à quatre syllabes qui ont chacune une syntaxe fixe et une sémantique propre. Elles feront l'objet d'une étude détaillée et spécifique.

4.

Analyseur lexical et problèmes d'ambiguïtés

La consultation d'un dictionnaire électronique lors de l'analyse automatique de textes ne permet pas de lever des ambiguïtés lexicales des mots. D'où la nécessité de construire des grammaires locales [Lap96], [Moh94]. Nous présentons ici les systèmes de numéraux en coréen en vue de les présenter systématiquement dans le DECOS, et de construire des grammaires locales de séquences nominales comportant des numéraux.

4.1. Construction des grammaires locales : les Numéraux

Rappelons qu'il existe, en coréen, deux systèmes de numéraux : l'un est constitué d'unités issues d'idéogrammes chinois (*numéraux sino-coréens*) ; l'autre contient des formes proprement coréennes (*numéraux coréens*) (Cf. Voir 3.3.6.).

Parmi les numéraux sino-coréens, il n'y a aucune variation morphologique entre le déterminant et le substantif, alors que certains numéraux coréens subissent des changements morphologiques. La figure 1 représente ces deux systèmes de numéraux.

Dans les dictionnaires éditoriaux, étant donné que la place est limitée, seulement certains numéraux sont présents. Il s'agit souvent de numéraux représentant des nombres qui ont un sens particulier ou une connotation particulière ou une particularité historique. Dans l'usage réel, les numéraux sont en nombre élevé mais fini; nous nous arrêtons en fait à :¹⁵

- (1) 구천 구백 구십 구조 구천 구백 구십 구역 구천 구백 구십 구만 구천 구백 구십 구
gu-chen-gu-baig-gu-sib-gu-jo gu-chen-gu-baig-gu-sib-gu-eg
gu-chen-gu-baig-gu-sib-gu-man gu-chen-gu-baig-gu-sib-gu
 (9999 9999 9999 9999 = en convention française : 9 999 999 999 999 999)

¹⁵ Notons qu'un *kyeng* sera probablement le successeur de (1), mais ce mot n'est guère employé dans l'usage réel. Quant au cas du français, voir [Sil93].

	한글 [Système coréen]		한자 [Système sino-coréen]	
	관사 [Dnum]	명사 [Nnum]	관사 [Dnum]	명사 [Nnum]
0	E	E	영 yeng	
1	한 han	하나 hana	일 il	
2	두 du	둘 dul	이 i	
3	세 se	셋 ses	삼 sam	
4	네 ne	넷 nes	사 sa	
5	다섯 dases		오 o	
6	여섯 yeses		육 yug	
7	일곱 ilgob		칠 chil	
8	여덟 yedelb		팔 pal	
9	아홉 ahob		구 gu	
10	열 yel		십 sib	
11	열한 yel-han	열하나 yel-hana	십일 sib-il	
12	열두 yel-du	열둘 yel-dul	십이 sib-i	
13	열세 yel-se	열셋 yel-ses	십삼 sib-sam	
14	열네 yel-ne	열넷 yel-nes	십사 sib-sa	
15	열다섯 yel-dases		십오 sib-o	
16	열여섯 yel-yeses		십육 sib-yug	
20	스물 seumul		이십 i-sib	
30	서른 seleun		삼십 sam-sib	
90	아흔 aheun		구십 gu-sib	
99	아흔아홉 aheun-ahob		구십구 gu-sib-gu	
100	E		백 baig	
1 000	E		천 chen	
10 000	E		만 man	
11 000	E		만천 man-chen	

Figure 1. Système coréen vs. système sino-coréen

Il est évident qu'il n'y a pas de sens à enregistrer quelques numéraux dans un dictionnaire électronique à titre d'exemples. Il faudrait les présenter d'une manière systématique et exhaustive. En fait, le mécanisme de formation des unités numérales est régulier et répétitif, et il est possible de les présenter dans un dictionnaire électronique sans énumérer la liste complète jusqu'au numéral de (1), en utilisant des automates récursifs (RTN, Recursive Transition Networks). [Sil93]

On notera que ce double système (i.e. système coréen et système sino-coréen) n'existe que pour les chiffres entre 1 et 99 : à partir de 100, les nombres en position de dizaine se réalisent difficilement par des numéraux coréens (e.g. dans 110, 10 est difficilement écrit par le numéral coréen 열 *yel* <10>, mais plutôt par le numéral sino-coréen 십 *sip* <10>) ; et ceux en position de centaine et plus n'acceptent pas le numéral coréen (e.g. dans 2500, le nom de nombre pour 500 est celui du système sino-coréen (오백 *o-baig* : 5x100), et non celui du système coréen (*다섯백 *dases-baig* : 5x100)).

Voici encore quelques exemples :

	Systeme Coréen	Systeme Sino-coréen
23	스물 셋 [<i>seumul</i> (20)- <i>ses</i> (3)]	이십 삼 [<i>i</i> (2)- <i>sib</i> (10)- <i>sam</i> (3)]
82	여든 둘 [<i>yedeun</i> (80)- <i>dul</i> (2)]	팔십 이 [<i>pal</i> (8)- <i>sib</i> (10)- <i>i</i> (2)]
101	<i>E</i>	백-(하나 + 일) [<i>baig</i> (100)-(hana(1) + il(1))]
244	<i>E</i>	이백-(?마흔 넷 + 사십 사) [<i>i</i> (2)- <i>baig</i> (100)-(?maheun(40)- <i>nes</i> (4) + sa(4)- <i>sib</i> (10)-sa(4))]
3562	<i>E</i>	삼천 오백 (?예순 둘 + 육십 이) [<i>sam</i> (3)- <i>chen</i> (1000)-o(5)- <i>baig</i> (100)-(?yesun(60)- <i>dul</i> (2)+ <i>yug</i> (6)- <i>sib</i> (10)-i(2))]

Le graphe *NumSin99* de la figure 2 contient les numéraux du système sino-coréen représentant les nombres compris entre 0 et 99 ; le graphe *NumCor99* dans Figure 3 définit les numéraux du système coréen qui représentent les nombres compris entre 1 et 99 :

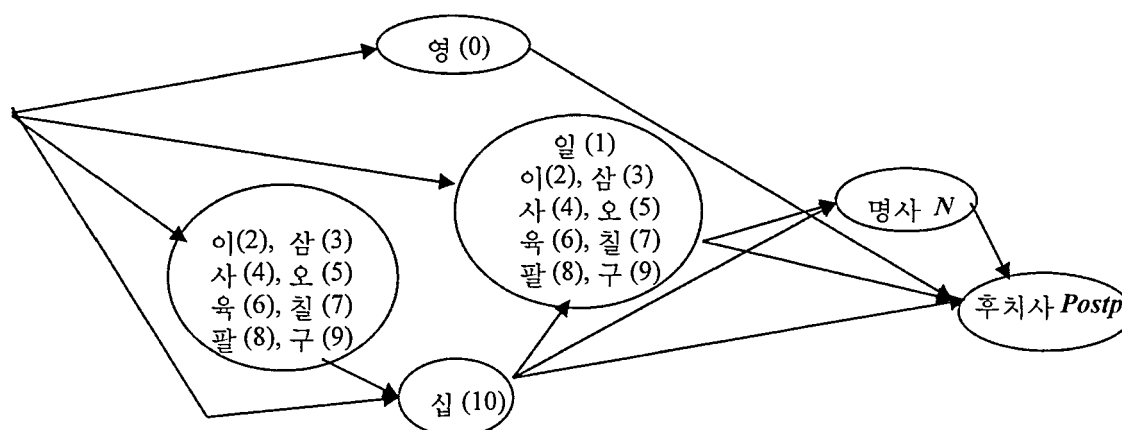


Figure 2 - Graphe *NumSin99*

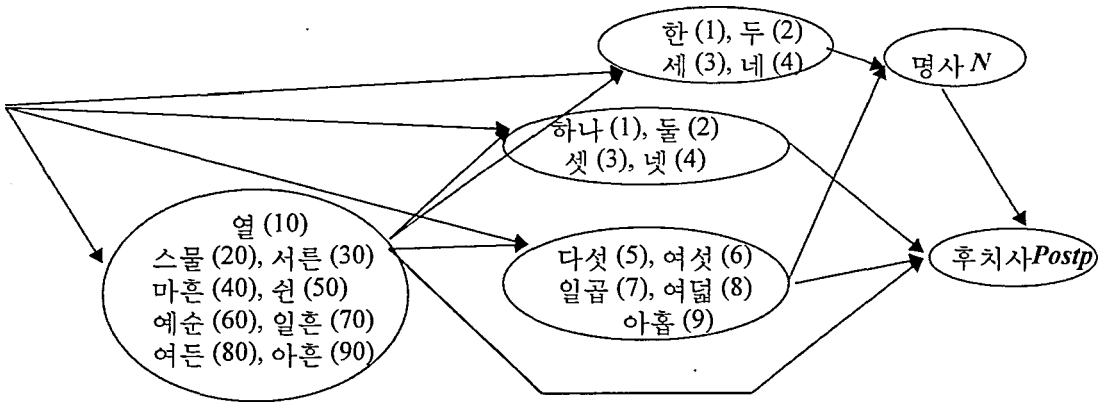


Figure 3. Graphe *NumCor99*

Le graphe *NumSin9999* de la figure 4 montre l'automate des numéraux sino-coréens représentant des nombres compris entre zéro et 9999 :

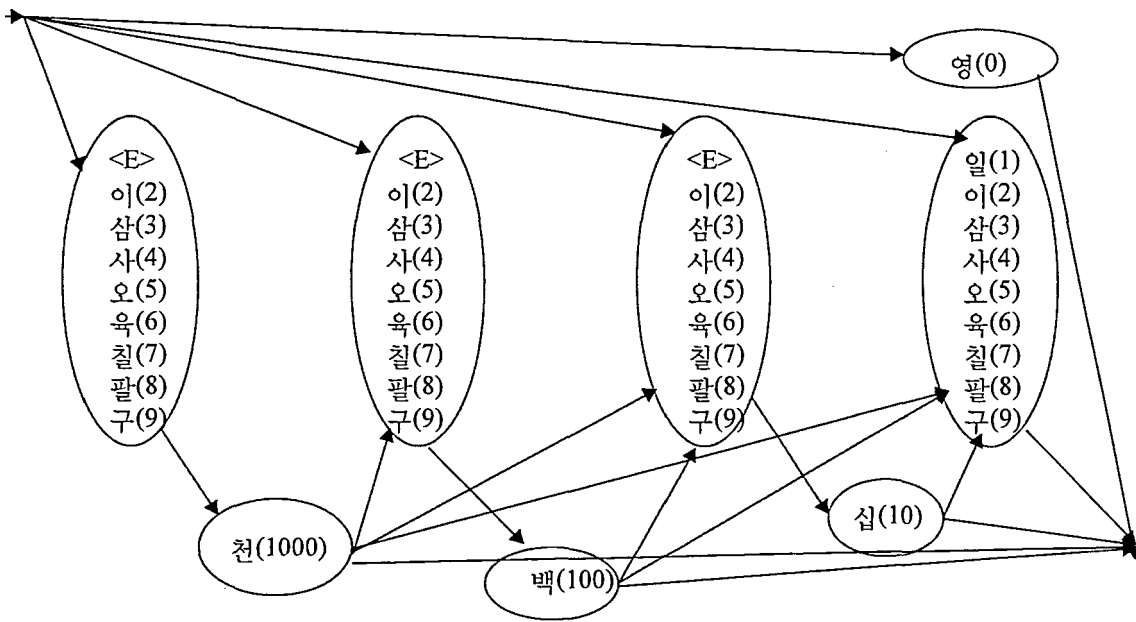


Figure 4. Graphe *NumSin9999*

Les noms de nombre du graphe *NumSin9999* se retrouvent dans d'autres nombres plus élevés. On peut donc obtenir un seul graphe qui représente tous les numéraux jusqu'à celui de (1) en répétant le graphe *NumSin9999* comme suit. La figure 5 contient le graphe de tous les numéraux sino-coréens. Il s'agit d'un graphe récursif RTN : les noeuds gris sont des noms de graphes auxiliaires.

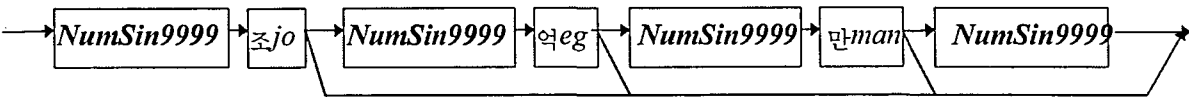


Figure 5 - Graphe *NumAllSin*

4.2. Grammaires locales d'expressions horaires

Quand les numéraux s'emploient comme déterminants, l'utilisation des systèmes sino-coréen et coréen obéit à des contraintes extrêmement complexes, liées au choix des substantifs modifiés. Par exemple, dans les expressions horaires, on utilise le système coréen pour compter l'heure (l'autre y étant interdit), et le système sino-coréen pour compter la minute. Soit (les unités coréennes sont soulignées) :

- (1) "3-*si*(heure) 3-*pun*(minute)" :
 (세 시 삼 분 / *삼 시 삼 분 / *세 시 세 분 / *삼 시 세 분)
 (se-si sam-bun / *sam-si sam-bun / *se-si se-bun / *sam-si se-bun)
- (2) "10-*si*(heure) 12-*pun*(minute)" :
 (열 시 십이 분 / *십 시 십이 분 / *열 시 열두 분 / *십 시 열두 분)
 (yel-si sib-i-bun / *sib-si sib-i-bun / *yel-si yel-du-bun / *sib-si yel-du-bun)

En terminologie scientifique ou technique, on peut également utiliser le système sino-coréen pour compter l'heure, mais plutôt à partir de 13 heures (jusqu'à 10 heures, c'est presque inacceptable ; pour 11 heures et 12 heures, c'est meilleur, mais peu naturel) :

- (3) 7h 5 : (일곱 ilgob(7) + *칠 chil(7)) - 시 *si* 오 o(5) - 분 *bun*
 9h 10 : (아홉 ahob(9) + *구 gu(9)) - 시 *si* 십 sib(10) - 분 *pun*
 12h 10 : (열두 yel(10)-du(2) + ?? 십이 sib(10)-i(2)) - 시 *si* 십 sib(10) - 분 *bun*
 22h 8 : (스물두 seumul(20)-du(2) + 이십이 i(2)-sib(10)-i(2)) - 시 *si* 팔 pal(8) - 분 *bun*

Quant à la seconde, au jour, au mois et à l'année, seul le système sino-coréen est accepté. Ainsi on observe :

- (4) [7h 7' 7", le 7 juillet 1777]
 = 1777-*nyen*(an) 7-*wel*(mois) 7-*il*(jour) 7-*si*(heure) 7-*bun*(minute) 7-*cho*(seconde)
 = (천 칠백 칠십 칠년 칠월 칠일 (*칠 + 일곱)-시 칠 분 칠초)
 = (chen-chil-baig-chil-sib-chil-nyen chil-wel chil-il(*chil+ilgob)-si chil-bun chil-cho)
- (10b) [5h 45' 52", le 20 novembre]
 = 11-*wel*(mois) 20-*il*(jour) 5-*si*(heure) 45-*bun*(minute) 52-*cho*(seconde):
 = (십 일 월 이 십 일 (*오 + 다섯)-시 사십 오 분 오십 이 초)
 = (sib-il-wel i-sib-il (*o + dases)-si sa-sib-o-bun o-sib-i-cho)

Il est à noter qu'il existe d'autres substantifs qui désignent également le mois et l'année: 달 *dal* (mois) et 해 *hai* (an). Or, ils exigent des numéraux coréens. Soit :

- (6) (일곱 ilgob + *칠 chil) - 달 *dal* (7 mois)
 (7) (일곱 ilgob + *칠 chil) - 해 *hai* (7 ans)

L'apparition d'un déterminant numéral coréen dans ces séquences n'est pas indépendante du fait que ces substantifs 달 *dal* et 해 *hai* sont aussi des termes proprement coréens (contrairement aux substantifs 월 *wel*(mois) et 년 *nyen*(an) dans (4), qui sont sino-coréens). Cependant, (6) ne signifie pas le 7ème mois (= juillet), mais une durée de 7 mois ; (7) n'indique pas la 7ème année, mais une durée, contrairement à 칠년 *chil-nyen* (7 ans) qui est ambigu (la 7ème année ou une durée de 7 ans).

Le graphe dans Figure 1 représente l'automate des expressions de date et d'heure (sauf les unités indiquant l'année).

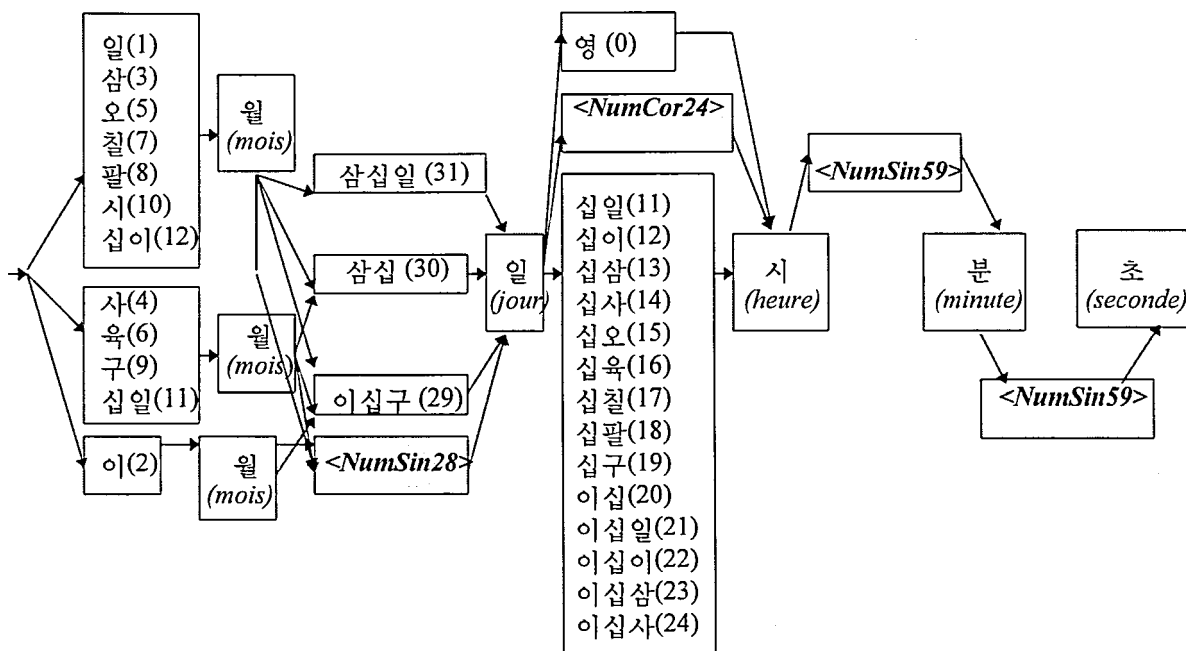


Figure 1. Graphe des expressions horaires *ExpHoraire*

4.3. Séquences nominales comportant des *Dnum*

4.3.1. Structure *Dnum Nmes-Gén N*

Dans des séquences nominales, les déterminants numériques exigent souvent un **nom de mesure** (*Nmes*), approprié aux substantifs modifiés. Autrement dit, on observe difficilement une séquence nominale d'un déterminant numéral directement suivi d'un nom (i.e. *Dnum N*). Par exemple, les séquences suivantes semblent peu naturelles :

(?* 세 *se* + * 삼 *sam*) 돼지 *dwaiji* (trois cochons)
 (*? 다섯 *dases* + * 오 *o*) 연필 *yenpil* (cinq crayons)

mais celles de (1) sont parfaites :

- (1) a. 세 *se* 마리 *mali*[Nmes]-의 *eui*[Gén] 돼지 *dwaiji* (trois Nmes de cochons)
 b. 다섯 *dases* 자루 *jalu*[Nmes]-의 *eui*[Gén] 연필 *yenpil* (cinq Nmes de crayons)

La structure des séquences (1) peut être décrite comme :

- (2) **Dnum Nmes-Gén N** (Dnum Nmes de N)

Ici, la nature de *Nmes* (Nom de mesure) est difficile à définir. D'une manière générale, il s'agit des noms d'unité que l'on utilise pour mesurer ou compter des substantifs. Ainsi, on compte 맥주 *maigju* (la bière) de plusieurs manières :

- (3) a. 세 병의 맥주
se byeng[Nmes:bouteille]-*eui*[Gén] *maigju* (3 bouteilles de bière)
 b. 세 잔의 맥주
se jan[Nmes:verre]-*eui*[Gén] *maigju* (3 verres de bière)
 c. 삼 리터의 맥주
sam lite[Nmes:litre]-*eui*[Gén] *maigju* (3 litres de bière)

On observe ici trois Nmes 병 *byeng* (bouteille), 잔 *jan* (verre) et 리터 *lite* (litre) respectivement.

♣ Dépendance entre *Nmes* et *N*

Les types de *Nmes* sont déterminés en fonction des noms associés. Ainsi, si la série de *Nmes* utilisée pour le substantif 맥주 *maigju* (bière) est comme celle de (3), on observe avec le nom 고기 *gogi* (viande) une autre série :

- (4) a. 네 덩어리의 고기
ne dengeli[Nmes:morceau]-*eui*[Gén] *gogi* (4 morceaux de viande)
 b. 사 킬로의 고기
sa killo[Nmes:kilo]-*eui*[Gén] *gogi* (4 kilos de viande)
 c. 네 근의 고기
ne geun[Nmes:poids]-*eui*[Gén] *gogi* (4 Nmes[poids] de viande)

Pourtant, les Nmes semblent parfois peu étroitement appropriés. Comparons les exemples suivants avec (4) :

- (5) a. 네 상자의 고기
ne sangja[Nmes:boîte]-*eui*[Gén] *gogi* (4 boîtes de viande)
 b. 네 봉지의 고기
ne bongji[Nmes:sac]-*eui*[Gén] *gogi* (4 sacs de viande)

- c. 네 트럭의 고기
ne teuleg[Nmes:camion]-eui[Gén] *gogi* (4 camions de viande)

Néanmoins, les **Nmes** comme ceux de (6) ne peuvent pas être utilisés pour mesurer une substance solide. On doit examiner plus en détail à quel degré les **Nmes** sont appropriés aux substantifs :

- (6) a. ?* 네 병의 고기
 ?**ne byeng[Nmes:bouteille]-eui[Gén]* *gogi* (4 bouteilles de viande)
 b. ?* 사 리터의 고기
 ?**sa lite[Nmes:litre]-eui[Gén]* *gogi* (4 litres de viande)

♣ Dépendance entre *Dnum* et *Nmes*

Le choix entre le déterminant numéral sino-coréen et coréen dépend du nom de mesure. Comme on a vu dans les exemples ci-dessus, les **Nmes** comme 덩어리 *dengeli* (morceau), 상자 *sangja* (boîte) exigent un **Dnum** coréen, alors que les **Nmes** comme 킬로 *killo* (kilo), 리터 *lite* (litre) demandent un **Dnum** sino-coréen. Ainsi, les combinaisons suivantes ne sont pas acceptées :

- (7) a. * 사 덩어리의 고기
 **sa dengeli[Nmes:morceau]-eui[Gén]* *gogi* (4 morceaux de viande)
 b. * 사 상자의 고기
 **sa sangja[Nmes:boîte]-eui[Gén]* *gogi* (4 boîtes de viande)
 c. * 네 킬로의 고기
 **ne killo[Nmes:kilo]-eui[Gén]* *gogi* (4 kilos de viande)

Notons que cette contrainte ne semble pas être sémantiquement régulière, on ne peut guère la prévoir en fonction du sens des noms de mesure. L'origine étymologique des **Nmes**, emprunts d'idéogrammes chinois ou termes proprement coréens, joue probablement un rôle important: si un **Nmes** est, à l'origine, un caractère chinois, il s'associe avec un **Dnum** sino-coréen ; s'il s'agit d'un **Nmes** proprement coréen, il se trouvera avec un **Dnum** coréen. Mais cette observation n'est pas toujours valable, car, par exemple, pour un **Nmes** sino-coréen comme 권 *gwen* (**Nmes** pour compter des livres), on a obligatoirement un **Dnum** coréen. Soit :

- (네 + *사) 권의 책
 (*ne + *sa*) *gwen[Nmes]-eui[Gén]* *chaig* (4 **Nmes** de livres)

Les deux systèmes sont nettement complémentaires quand il s'agit d'un nombre, en général, de moins d'une dizaine, mais ils commencent à entrer en concurrence pour les nombres plus élevés. Plus précisément, les **Dnum** sino-coréens deviennent aussi acceptables dans les séquences exigeant un **Dnum** coréen à partir de certains nombres. Comparons (7a) et (7b) avec les exemples suivants :

- (8) a. (마흔 다섯 + 사십 오) 덩어리의 고기
 (*maheun-dases+sa-sib-o*) *dengeli[morceau]-eui gogi* (45 morceaux de viande)
 b. (췌 일곱 + 오십 칠) 상자의 고기
 (*swin-ilgob + o-sib-chil*) *sangja[boîte]-eui gogi* (57 boîtes de viande)

Cependant, il est difficile de prévoir à partir de quel nombre cette concurrence apparaît : cela dépend du contenu lexical de la séquence.

4.3.2. Structure *N Dnum Nmes*

Notons que la structure (2) est en relation transformationnelle avec une autre structure où la partie *Modifieur* se trouve à droite du *Nom*. Soit :

- (9) *N Dnum Nmes* (N Dnum Nmes)

Par exemple, pour (1), on aura :

- (10)a. 돼지 세 마리 *dwaiji se mali[Nmes]* (cochons trois **Nmes**)
 b. 연필 다섯 자루 *yenpil dases jalu[Nmes]* (crayons cinq **Nmes**)

Cette relation, fort systématique en coréen, l'est moins en français. Comparons (11) avec (12):

- (11)a. 세병의 맥주 *se byeng[Nmes:bouteille]-eui[Gén] maigju* (3 bouteilles de bière)
 b. = 맥주 세병 *maigju se byeng[Nmes:bouteille]* (bières 3 bouteilles)

- (12)a. 3 bouteilles de bière
 b. = ?*la bière de 3 bouteilles

Maintenant, examinons la paire suivante :

- (13)a. 삼 일간의 휴가 *sam ilgan[Nmes:jour]-eui[Gén] hyuga* (3 jours de vacances)
 b. = 휴가 삼일간 *hyuga sam ilgan[Nmes:jour]* (vacances (de) 3 jours)

On a une paire voisine en français (cf. [Bor85], [Gro77]) :

- (14a) 3 jours de vacances
 (14b)= des vacances de 3 jours

dont les structures sont :

- (15a) *Dnum Nmes de N*
 (15b)= *N de Dnum Nmes*

Voici un autre exemple :

- (16)a. 오 만 프랑의 월수입
o-man peulang[Nmes:fr]-eui[Gén] welsuib (50 000 francs de salaire)
 b. = 월수입 오 만 프랑
welsuib o-man peulang[Nmes:fr] (salaire 50 000 francs)

qui correspond également à une paire similaire en français :

- (17)a. 50 000 francs de salaire
 b. = un salaire de 50 000 francs

Les natures des constituants de la séquence (12a) et de leur relation seraient probablement différentes de celles qui existent dans (17a). En coréen aussi, bien qu'on obtienne le même type de transformation pour la paire (11) et (16), ces deux paires doivent être analysées de façon différente.

L'existence de la relation (2)-(9) (en français, (15a)-(15b)) distingue la structure comportant un Nmes approprié de la structure transformée à partir d'une phrase *N0 avoir N1*. Comparons la paire (16) avec :

- (18)a. 오 만 명의 월수입
o-man myeng[Nmes:personne] - eui[Gén] welsuib (le salaire de 50 000 personnes)
 b. = *월수입 오 만 명
 **welsuib o-man myeng[Nmes:personne]* (50 000 personnes de salaire)

Les séquences (16a) et (18a) sont formellement identiques, mais, le Nmes 명 *myeng* (personne) dans (18a) n'est pas relié à N 월수입 *welsuib* (salaire), comme le Nmes 프랑 *peulang* (franc) dans (16a). Il s'agit d'une séquence comportant un modifieur génitif (possesseur), qui est équivalente à :

- (19)a. = 오 만 명이 가지고있는 월수입
o-man myeng-i gajigoissneun welsuib (le salaire que 50 000 personnes ont)
 (19)b. = 오 만 명이 일정 월수입을 가지고있다
o-man myeng-i iljeng welsuib-eul gajigoissda (50 000 personnes ont certains salaires)

Le sujet 오만명 *o-man myeng* (50 000 personnes) est une forme obtenue par l'effacement d'un nom. Ainsi, on peut restituer, dans (18a), un substantif après le Nmes 명 *myeng*, alors qu'il est difficile de l'observer dans (16a). Soit pour (18a) :

- (20)a. = 오 만 명의 성인 남자의 월수입
o-man myeng-eui sengin namja-eui welsuib
 (le salaire de 50 000 Nmes adultes masculins)

- (20)b.= 성인 남자 오만 명의 월수입
sengin namja . o-man myeng-eui welsuib
 (le salaire d'adultes masculins 50 000 Nmes(personnes))

En français, on observe la même différence que (16)-(18). Comparons (17) avec (21) :

- (21)a. *le salaire de 50 000 personnes*
 b.= **50 000 personnes de salaire*

Nous n'allons pas entrer ici dans une discussion plus détaillée. Mais, notons qu'il existe plusieurs différences syntaxiques entre la structure *Dnum Nmes-Gén N* et *N Dnum Nmes*. Prenons l'exemple du marqueur de pluriel (qui n'est jamais obligatoire) : on observe que, dans la première structure, le marqueur *들 deul* (morphème pluriel comme <s> en français) s'ajoute seulement aux N, alors que dans la deuxième, on peut répéter *들 deul* après N et Nmes :

- (22)a. 세 분 - (E + *들) - 의 어른들
*sei bun-(E + *deul)-eui eleun-deul* (3 Nmes-(E + *pl<s>) de adulte-s)
 b.= 어른들 세 분 - (E + 들)
eleun-deul sei bun-(E + deul) (adulte-s 3 Nmes-(E + pl<s>))

4.3.3. Structure *Dnum N*

La structure *Dnum N* est difficilement observable pour les paires mentionnées ci-dessus. Par exemple, si on a :

- (23) 다섯 명의 아이 *dases myeng[Nmes:personne]-eui ai* (5 Nmes d'enfants)
 = 아이 다섯 명 *ai dases myeng[Nmes:personne]* (enfants 5 Nmes)
 = 다섯 아이 *dases ai* (5 enfants)

Il est difficile d'observer :

- (24) 다섯 마리의 돼지 *dases mali[Nmes:animal]-eui dwaiji* (5 Nmes de cochons)
 = 돼지 다섯 마리 *dwaiji dases mali[Nmes:animal]* (cochons 5 Nmes)
 = ??다섯 돼지 ?? *dases dwaiji* (5 cochons)

En français, on observe une situation similaire. Si la relation (25) est permise pour certain cas comme (26), elle est interdite pour des cas comme (27) :

- (25)a. *Dnum Nmes de N*
 b.= *Dnum N*

Soit :

- (26)a. 3 bouteilles de Calsberg
b.= 3 Calsbergs

- (27)a. 3 paquets de macaroni
b.= *3 macaronis

Aucune des séquences en coréen correspondant à (26a) et (27a) n'acceptent la structure **Dnum N** : ¹⁶

- (28)a. 세 병의 맥주 *sei byeng[Nmes:bouteille]-eui maigju* (3 bouteilles de bière)
b.= *세 맥주 **sei maigju* (3 bières)
- (29)a. 세 봉지의 국수 *sei bongji[Nmes:paquet]-eui gugsu* (3 paquets de macaroni)
b.= *세 국수 **sei gugsu* (3 macaroni)

Récapitulons les trois structures discutées jusqu'ici. Les exemples de la figure 1 représentent trois cas différents en coréen, et ceux de la figure 2 illustrent trois cas en français, étudiés ici :

	Dnum Nmes-Gén N	N Dnum Nmes	Dnum N
3 명의 아이 <i>3 myeng-eui ai</i> (3 Nmes d'enfants)	+	+	+
3 킬로의 고기 <i>3 killo-eui gogi</i> (3 Nmes[kilo] de viande)	+	+	-
3 명의 월수입 <i>3 myeng-eui welsuib</i> (le salaire de 3 personnes)	+	-	-

Figure 1. Séquences nominales comportant **Dnum** en coréen

	Dnum Nmes de N	N de Dnum Nmes	Dnum N
3 jours de vacances	+	+	-
3 paquets de macaroni	+	-	-
3 bouteilles de bière	+	-	+

Figure 2. Séquences nominales comportant **Dnum** en français

¹⁶ Même lorsqu'un des deux systèmes de numéraux est nettement interdit dans la séquence **Dnum N**, l'interdiction devient moins claire à partir de certains nombres assez élevés. Ainsi, on observe :

- (ia) (세 + *삼) 아이 *(sei + *sam) ai* (3 enfants)
(ib) (마흔 + ?*사십) 아이 *(maheun + ?*sa-sib) ai* (40 enfants)
(ic) (백 스물 + ?백 이십) 아이 *(baig-seumul + ?baig-i-sib) ai* (120 enfants)
- (iia) (다섯 + *오) 국가 *(dases + *o) gugga* (5 pays)
(iib) (마흔 + ?사십) 국가 *(maheun + ?sa-sib) gugga* (40 pays)
(iic) (?백 스물 + 백 이십) 국가 *(?baig-seumul + baig-i-sib) gugga* (120 pays)

Les grammaires locales doivent être construites et attachées à chaque substantif et les noms de mesure appropriés à ce substantif. Elles peuvent être représentées sous forme de graphes et converties en automates finis dans un système électronique d'analyse grammaticale. Par exemple, le premier cas en coréen dans Figure 1 sera être représenté par un graphe comme dans Figure 3 :

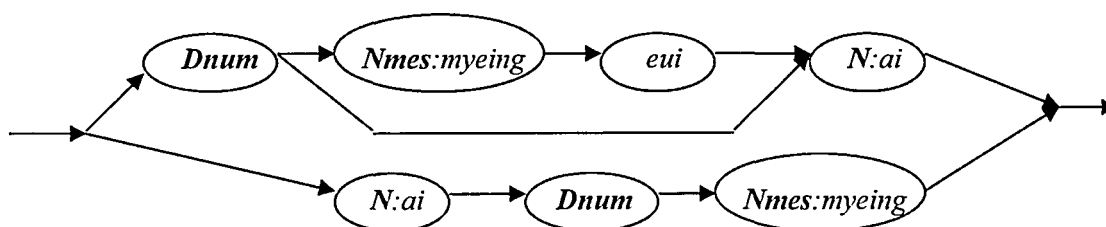


Figure 3. Graphe de '3 myeng-eui ai'

Les groupes nominaux comportant des numéraux sont à étudier à part et en détail. L'interdiction des deux systèmes de numéraux dans les séquences nominales, et l'association entre noms de mesure et substantifs feront l'objet d'une étude ultérieure.

5.

Analyseurs syntaxiques et Lexique-grammaire

La construction d'un dictionnaire électronique ne se limite pas au seul établissement des listes de mots. Un tel dictionnaire doit aussi donner des informations sur les contextes dans lesquels un mot donné peut être utilisé. Autrement dit, des informations syntaxiques concernant chaque élément lexical et enfin des indications sur le sens doivent figurer dans un dictionnaire électronique.

Vu que la description du sens est un problème encore ouvert, il nous faudra, dans un premier temps, décrire systématiquement des informations syntaxiques. La description des caractéristiques syntaxiques des éléments lexicaux a un enjeu particulier dans des applications telles que la désambiguïsation des homographes rencontrés fréquemment dans un texte. En fait, "la résolution des ambiguïtés des mots est la première fonction des programmes de traitement des langues naturelles. La méthode la plus courante consiste à effectuer une analyse syntaxique des phrases, autrement dit, les contextes où apparaissent les mots sont pris en compte. Pour qu'une telle prise en compte puisse se faire systématiquement, la totalité des contextes ou formes de contextes de chaque mot doit être disponible dans les programmes. La construction de grammaires formelles et de lexiques électroniques des langues répond à ce besoin. [Gro89]"

5.1. Description syntaxique des éléments lexicaux

Nous allons présenter, à titre d'exemple, une description syntaxique d'une classe d'adjectifs, définie formellement. Une phrase simple construite autour d'un adjectif prédicatif a la structure :

$$N_0\text{-}nmtf \quad (N_1\text{-}Postp + E) \quad Adj\text{-}St$$

Le modèle de description est le **Lexique-grammaire** dont l'idée principale est que les unités de sens ne sont pas les mots isolés mais les phrases élémentaires. Ainsi, les adjectifs seront décrits au niveau phrastique, c'est-à-dire, dans leurs relations avec leur sujet et leurs éventuels compléments. Une structure syntaxique, si elle n'est pas

attachée à des éléments lexicaux, peut rester très ambiguë. Par exemple, la structure syntaxique (1) définit toutes les constructions à adjectifs **symétriques**¹⁷ :

- (1) $N_0\text{-nmtf } N_1\text{-wa } Adj\text{-St} =$:

A 선은 B 선과 평행하다
A sen-eun B sen-gwa pyenghaingha-da
 A ligne-nmtf B ligne-Postp.wa parallèle-St
 (La ligne A est parallèle à la ligne B)

La structure (2) est ambiguë :

- (2) $N_0\text{-nmtf } N_1\text{-i } Adj\text{-St}$

Comparons :

- (3a) 민우는 눈이 간질간질하다
Minu-neun nun-i ganjilganjilha-da
 (Minu, les yeux sont démangeants = Minu a les yeux qui démangent)
- (3b) 민우는 눈이 파랗다
Minu-neun nun-i palah-da
 (Minu, les yeux sont bleus = Minu a les yeux bleus)

Ces deux phrases sont syntaxiquement identiques, et elles entrent dans la structure (2) avec les mêmes actants : aucune marque formelle ne permet de prévoir la présence de différences entre elles. Ces adjectifs se combinent tous les deux avec un actant **humain** et un actant **partie du corps**, quels que soient leurs rôles grammaticaux. Les deux

¹⁷ Notons que cette construction ne peut pas caractériser les constructions à verbes symétriques, car il y a plusieurs types de complément $N_1\text{-wa}$ dans les constructions verbales, contrairement aux cas des constructions adjectivales. Ainsi, si le complément $N_1\text{-wa}$ dans (i) est un actant du verbe symétrique *ihonhada* (divorcer), celui dans (ii) ne détermine pas la nature du verbe *ilhada* (travailler) :

- (i) 민우는 인아와 이혼했다
Minu-neun Ina-wa ihonha-ess-da
 Minu-nmtf Ina-Post.wa divorcer-Pas-St (Minu a divorcé avec Ina)
- (ii) 민우는 인아와 일했다
Minu-neun Ina-wa ilha-ess-da
 Minu-nmtf Ina-Post.wa travailler-Pas-St (Minu a travaillé avec Ina)

En fait, le complément dans (ii) est un complément circonstanciel qui peut être supprimé ou remplacé par *honja* (seul), alors que l'autre est obligatoire (cf. [Bor71], [Hon87], et [Nam92b]):

- (iii) *민우는 (E + 혼자) 이혼했다
 **Minu-neun (E + honja) ihonha-ess-da*
 *Minu-nmtf (E + seul) divorcer-Pas-St (Minu a divorcé (E + seul))
- (iv) 민우는 (E + 혼자) 일했다
Minu-neun (E + honja) ilha-ess-da
 Minu-nmtf (E + seul) travailler-Pas-St (Minu a travaillé (E + seul))

actants sont en relation inaliénable. Ainsi, on ne peut y insérer un complément de nom qui n'est pas coréférentiel au sujet comme :

* N_0 -nmtf [N_2 -Gén N_1]-i Adj-St (* N_0 , [N_1 de N_2] être Adj)

Soit :

*민우는 인아의 눈이 간질간질하다

*Minu-neun Ina-eui nun-i ganjilganjilha-da

(*Minu-nmtf Ina-Gén yeux-i démangeant-St = Minu a les yeux de Ina qui démangent)

*민우는 인아의 눈이 파랗다

*Minu-neun Ina-eui nun-i palah-da

(*Minu-nmtf Ina-Gén yeux-i bleu-St = Minu a les yeux de Ina qui sont bleus)

Comparons ces deux phrases avec une autre construction adjectivale qui permet ce type de complément de nom. La phrase suivante est constituée des mêmes actants que les précédentes, et elle est parfaitement acceptable :

민우는 인아의 눈이 좋다

Minu-neun Ina-eui nun-i joh-da

(Minu-nmtf Ina-Gén yeux-i plaisant-St = Minu aime les yeux de Ina)

Malgré la ressemblance entre les adjectifs de (3a) et (3b), on observe plusieurs différences. Sémantiquement, l'adjectif 간질간질하다 *ganjilganjilhada* (démangeant) dans (3a) exprime un symptôme physique éprouvé par le sujet humain *Minu*, et donc le siège du sentiment est intérieur au sujet¹⁸, alors que l'adjectif 파랗다 *palahda* (bleu) dans (3b) ne décrit aucune expérience du sujet grammatical, l'observation faite par le locuteur reste extérieure. Syntaxiquement, plusieurs différences sont relevées entre ces deux constructions. Voici quelques propriétés syntaxiques qui caractérisent la construction (3a) :

♣ Restructuration en groupe nominal

La construction (3a) ne permet pas la relation transformationnelle suivante, alors qu'elle est acceptée pour (3b) :

(4) N_0 -nmtf N_1 -i Adj-St
= [N_0 -Gén N_1]-nmtf Adj-St ([N_1 de N_0] être Adj)

Soit :

(5a) *민우의 눈이 간질간질하다

*[Minu-eui nun]-i ganjilganjilha-da

¹⁸ C'est pourquoi le sujet à la première personne améliore la phrase :

나는 눈이 간질간질하다

na-neun nun-i ganjilganjilha-da

(Moi, les yeux sont démangeants = J'ai les yeux qui démangent)

(*[Minu-Gén yeux]-nmtf démangeant-St = Les yeux de Minu démangent)

- (5b) 민우의 눈이 파랗다
 [Minu-eui nun]-i palah-da
 ([Minu-Gén yeux]-nmtf bleu-St = Les yeux de Minu sont bleus)

En fait, (5b) est la phrase-source de la construction (3b) dont les deux actants sont obtenus à partir d'un groupe nominal par la transformation (4) : 파랗다 *palahda* (bleu) dans (3b) est un adjectif qui demande un seul argument ¹⁹ (i.e. N_0 -nmtf Adj-St).

♣ Construction en 가지고있다 *gajigoissda* (avoir)

La construction (3a) n'accepte pas la relation transformationnelle suivante, alors qu'elle est acceptée pour (3b) :

$$\begin{array}{l} N_0\text{-nmtf} \quad N_1\text{-i} \quad \text{Adj-St} \\ = \quad N_0\text{-nmtf} \quad [\text{Adj-Sd } N_1\text{-Acc}] \quad \text{가지고있다 } gajigoiss(da)\text{-St} \quad (N_0 \text{ avoir } [\text{Adj } N_1]) \end{array}$$

*민우는 간질간질한 눈을 가지고있다
 *Minu-neun [ganjilganjilha-n nun]-eul gajigoiss-da
 (*Minu-nmtf [démangeant-Sd yeux]-Acc avoir-St = Minu a les yeux qui démangent)

민우는 파란 눈을 가지고있다
 Minu-neun [pala-n nun]-eul gajigoiss-da
 (Minu-nmtf [bleu-Sd yeux]-Acc avoir-St = Minu a les yeux bleus)

♣ Adverbe aspectuel inchoatif

La construction (3a) accepte les adverbes de type 지금 막 *jigeummag* (tout d'un coup) ou 점점 더 *jemjemde* (de plus en plus), alors qu'ils sont interdits dans (3b). Ainsi on aura :

- (6a) 민우는 지금 막 눈이 간질간질하다
 Minu-neun jigeummag nun-i ganjilganjilha-da
 (Minu, les yeux sont tout d'un coup démangeants)

- (6b) *민우는 지금 막 눈이 파랗다
 *Minu-neun jigeummag nun-i palah-da
 (*Minu, les yeux sont tout d'un coup bleus)

¹⁹ Etant donné que la postposition *o* *i* (*i* représentant deux variantes *o* *i* et *ga* *ga*) est formellement identique à celle du nominatif (*nmtf* =: *i* + *ga*), la construction de type N_0 -nmtf N_1 -i (Adj+V)-St est souvent appelée **phrase à double sujet**, ou à **double nominatif**. Cette construction est fréquemment observée : il s'agit d'une restructuration en séquence en *o* *-i* soit d'un complément du génitif dans un groupe nominal, soit d'un complément quelconque (essentiel ou circonstanciel) en vue d'un effet emphatique. Les phrases-sources de la construction à double sujet sont tellement variées qu'une étude plus détaillée est nécessaire. Dans le cas des constructions adjectivales, il apparaît que seuls les adjectifs de type (3a) exigent pour structure canonique N_0 -nmtf N_1 -i Adj-St.

La construction adjectivale décrivant en général une qualité ou un état durable du sujet, il est souvent difficile d'insérer un adverbe qui introduit un aspect ponctuel. L'aspect ponctuel, traduit explicitement par des adverbes comme celui de (6a), est une caractéristique importante de la construction (3a). Cette particularité, qui la rapproche d'une phrase verbale, est probablement liée à la sémantique de cet adjectif : il manifeste un *symptôme physique* que le sujet éprouve.²⁰

Les propriétés syntaxiques de l'adjectif de (3a) que nous venons d'examiner sont la traduction formelle de notre intuition de départ. Cette intuition sémantique, confirmée par des tests syntaxiques, nous amène à établir une définition formelle qui permet de caractériser les adjectifs de type (3a) et à les distinguer d'autres qui pourtant entrent dans la même structure phrastique (2). La relation transformationnelle suivante sert de **critère définitionnel** pour ces adjectifs :

- (7) N_0 -nmtf N_1 -i Adj-St
 = N_0 -nmtf [N_1 -nmtf Adj-고 go] 느끼다 neuggi(da)-St (N_0 sentir [que N_1 être Adj])

Ainsi on observe :

- (8a) 민우는 눈이 간질간질하다고 느낀다
 Minu-neun nun-i ganjilganjilha-da-go neuggi-nda
 (Minu sent que les yeux sont démangeants)
- (8b) *민우는 눈이 파랗다고 느낀다
 *Minu-neun nun-i palah-da-do neuggi-nda
 (*Minu sent que les yeux sont bleus)

²⁰ Or, dans certaines phrases adjectivales comme la suivante, on observe aussi des adverbes aspectuels non-duratifs comme '지금은 jigeumeun (maintenant)' :

- (i) 민우는 지금은 똥똥하다
 Minu-neun jigeumeun ddungddungha-da
 (Minu-nmtf maintenant gros-St = Minu est gros maintenant)

L'adverbe '지금은 jigeumeun (maintenant)' introduit un aspect ponctuel, à l'opposé des adverbes comme 늘 neul (toujours), 항상 hangsang (toujours) ou 변함없이 byenhameksi (constamment). Cependant, l'adverbe de (6) n'est pas accepté dans (i). Ainsi on ne peut avoir (ii) :

- (ii) ?*민우는 지금 막 똥똥하다
 ?*Minu-neun jigeummag ddungddungha-da
 (?*Minu-nmtf tout d'un coup gros-St = Minu est gros tout d'un coup)

En fait, la différence entre (i) et (ii) est due au fait que l'adverbe de (ii) introduit un **aspect inchoatif**, aspect qui marque le début d'un procès sans indiquer une durée quelconque, ce qui est incompatible avec les adjectifs comme celui de (ii). L'aspect ponctuel de l'adverbe de (i) doit donc être interprété comme un **effet contrastif**. (i) implique une phrase telle que :

민우는 예전에는 똥똥하지 않았는데, 지금은 똥똥하다
 Minu-neun yejeneineun ddungddungha-jianh-ass-neunde, jigeumeun ddungddungha-da
 Minu-nmtf avant gros-Mnég-Mpas-Conj, maintenant gros-St
 (Avant, Minu n'était pas gros, mais maintenant il est gros)

La construction (8a) est une phrase plutôt théorique, elle n'est pas très naturelle. Mais, elle est clairement interdite pour l'adjectif de (8b). Nous avons recensé environ 200 adjectifs qui entrent dans la relation définitionnelle (7). Ils présentent une homogénéité sémantique, et ils ont des propriétés syntaxiques communes : nous les appelons **adjectifs symptomatiques**. Voici quelques exemples :

간질간질하다	<i>ganjilganjilhada</i>	[(être) démangeant]
쓰라리다	<i>sseulalida</i>	[(être) cuisant]
파갑다	<i>ddagabda</i>	[(être) lancinant]
어질어질하다	<i>ejilejihada</i>	[(être) tournant]

La classe ainsi obtenue fera partie de notre classification complète des adjectifs: ces classes sont toutes rigoureusement définies par des critères formels et syntaxiques, mais montrent en général une certaine homogénéité sémantique. Les appellations comme '*adjectifs symétriques* (codés **AWS**)', '*adjectifs psychologiques* (codés **AEP**)' ou '*adjectifs symptomatiques* (codés **AIC**)' ne sont donc attribuées qu'après classement par des critères formels. Ces classes ne correspondront pas toujours parfaitement à notre intuition sémantique.

Nous avons réparti 5 300 adjectifs en 15 classes définies syntaxiquement (cf. [Nam96c]). Le tableau suivant²¹ récapitule les critères formels de cette classification (Figure 1) :

Code	Classe des adjectifs	Critères définitionnels
AWS	Adjectifs symétriques	$N_0 \ N_1\text{-}wa \ Adj$ $=N_1 \ N_0\text{-}wa \ Adj$
AIC	Adjectifs symptomatiques	$N_0 \ N_1\text{-}i \ Adj$ $=N_0 \ [N_1 \ Adj\text{-}go] \ neuggida$
AED	Adjectifs à complément datif	$N_0 \ N_1\text{-}ei \ Adj$ $=N_0 \ N_1\text{-}ei \ Adj\text{-}gei \ gulda$
AEV	Adjectifs à complément vecteur	$N_0 \ N_1\text{-}ei \ Adj$ $=*N_0 \ N_1\text{-}ei \ Adj\text{-}gei \ gulda \ / \neq N_0 \ N_1\text{-}eidaihae \ Adj$
AEG	Adjectifs scéniques-génitifs	$N_0 \ N_1\text{-}ei \ Adj$ $=N_1 \ N_0\text{-}lo \ Adj \ / \ = [N_1\text{-}eui \ N_0] \ Adj$
AES	Adjectifs scéniques	$N_0 \ N_1\text{-}ei \ Adj$ $=N_1 \ N_0\text{-}lo \ Adj \ / \ =*[N_1\text{-}eui \ N_0] \ Adj$
AEP	Adjectifs psychologiques	$(Qu \ P)_0 \ N_1\text{-}ei \ Adj$ $=N_1 \ (Qu \ P)_0\text{-}lo \ Adj$
AEE	Adjectifs évaluatifs	$(Qu \ P)_0 \ N_1\text{-}ei \ Adj$ $=*N_1 \ (Qu \ P)_0\text{-}lo \ Adj \ / \neq N_1 \ (Qu \ P)_0\text{-}Acc \ Adj\text{-}e \ hada$
AER	Adjectifs à complément repère	$N_0 \ N_1\text{-}ei \ Adj$ (cf. Les relations définitionnelles des classes AED, AEV, AEG, AES, AEP et AEE y sont interdites)

²¹ Pour simplifier, les symboles *nmtf* et *St* sont omis dans ce tableau.

ARR	Adjectifs à compléments résiduels	(a) N_0 N_1 -eise Adj (b) N_0 N_1 -boda Adj (c) N_0 P-go Adj
AP	Adjectifs à sujet-complétive réduite en -gi	$[N_1\text{-Acc VR-gi}]_0$ Adj $= (N_1)_0$ $[VR-gi]_1$ - i Adj
AS	Adjectifs à sujet-complétive et à sujet humain	$[(Nhum)_0 VR\text{-}Sd Ncomp]_0$ Adj $= (Nhum)_0 VR\text{-}da\text{-}ni, (Nhum)_0$ Adj
AQ	Adjectifs à sujet-complétive	$(Qu P)_0$ Adj
AH	Adjectifs à sujet humain	$(Nhum)_0$ Adj
AN	Adjectifs à sujet non-humain	$(N\text{-}hum)_0$ Adj

Figure 1

Et l'organisation de ces classes selon la distribution des actants, i.e. sujet et complément, est présentée dans la Figure 2 :

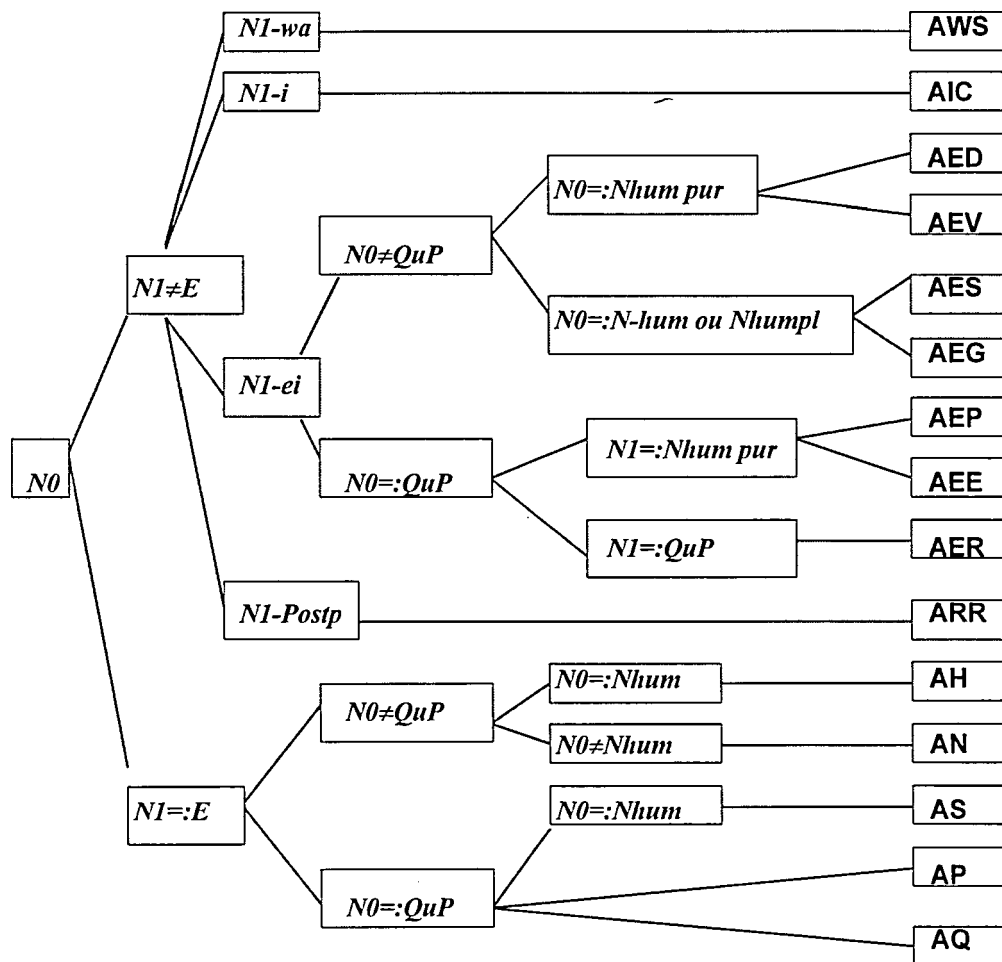


Figure 2

5.2. Représentation par matrices binaires

Toutes les classes d'adjectifs se présentent sous forme d'une matrice binaire, où chaque colonne (i.e. propriété syntaxique) représente une structure de phrase. La matrice des entrées lexicales d'adjectifs ne comporte que des marques binaires (signe '+' ou signe '-'), représentant une évaluation de l'acceptabilité. On notera que "les marques d'acceptabilité (signe '+') signifient qu'au moins une séquence acceptable répondant à la définition et constituant une phrase a pu être trouvée" [BGL76]). Voici un extrait de la tables de la classe des adjectifs AIC (Figure 3).

N0				N1								Cf.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
N0 =: Nhum	N0 =: N-hum	N0 =: Qu P		N1 =: Nhum	N1 =: N-hum	N1 =: Npc (parle du corps)	N1 =: Nlocatif (type "sangchô")	N1 =: Nfigé résiduel (type "côngsin")	N1 =: Qu P																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
+	-	-	가운하다	-	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

Figure 3

Dans le dictionnaire électronique, les informations syntaxiques sont présentées sous forme de codes qui correspondent à notre classement syntaxique. Ainsi, le lexique-grammaire des adjectifs permettra de construire un dictionnaire syntaxique DECO-SYN. A titre d'exemple, DECO-SYN basé sur le dictionnaire des formes canoniques sera de la forme :

가공적이다 *gagongjegida*, **AQ**, AS06-CM
 가깝다 *gaggabda*, **AWS, ARR**, AS11-RAM/24
 가깝디가깝다 *gaggabdigaggabda*, **AWS, ARR**, AS11-RAM/24
 가깝하다 *gaggabhada*, **AEP**, AS03-ARHM
 가난하다 *gananhada*, **AH**, AS02-ANHM
 가냘프다 *ganyalpeuda*, **AH**, AS11-RAM/143
 가느다랗다 *ganeudalahda*, **AH**, AS11-RAM/59

Les codes indiquant les classes syntaxiques doivent permettre l'accès aux matrices binaires qui comportent les informations sur les propriétés syntaxiques des entrées lexicales concernées. La structure de données adoptée détermine l'efficacité de l'algorithme d'analyse syntaxique (*cf.* [Roc93] et [Lap96]). Les choix informatiques faits lors de l'implémentation d'un dictionnaire prennent une importance considérable, surtout quand la taille des données lexicales est très élevée. Nous en reparlerons dans une étude ultérieure.

5.3. Construction d'un dictionnaire syntaxique du coréen

Cette étude s'inscrit dans l'objectif global de décrire systématiquement les variations morphologiques et les propriétés syntaxiques des éléments lexicaux coréens, en vue d'implémenter un système qui traite automatiquement les textes écrits en coréen. Nous nous sommes limitée ici à illustrer une classe d'adjectifs définie formellement. Il est certain qu'une description complète des structures phrastiques basées sur des adjectifs sera un modèle important pour les autres éléments lexicaux. Une description complète du lexique du coréen sera effectuée de façon reproductible et cohérente dans notre système électronique DECO.

RÉFÉRENCES

Dictionnaires

- I, Hi-Seung, 1988, *Guge Dai Sajen* (Grand Dictionnaire de la Langue Coréenne), Séoul: Minjungselim.
- I, Hi-Seung, 1989, *Minjung Eisseinseu Guge Sajen* (Dictionnaire du coréen), Séoul : Minjungselim.
- I, Eung-Baig, 1992, *Guge Dai Sajen* (Grand Dictionnaire du Coréen), Séoul : Gyoyugdose.
- Sin, Gi-Chel ; Sin, Yong-Chel, 1990, *Sai Ulimal Keun Sajen* (Nouveau Dictionnaire de la Langue Coréenne), Séoul : Samseng Chulpansa.
- Grand Dictionnaire Encyclopédique Larousse*, 1982, Paris : Larousse.
- Grand Robert de la Langue Française*, 1986, Paris : Le Robert.
- Petit Robert : Dictionnaire universel des noms propres*, 1983, Paris : Le Robert.

- Boons, Jean-Paul ; Alain Guillet ; Christian Leclère, 1976, *La structure des phrases simples en français : constructions intransitives*, Genève : Droz.
- Borillo, Andrée, 1971, Remarques sur les verbes symétriques français, *Langue Française* 11, Paris : Larousse.
- Choi, Hyun-Bai, 1929 (ré-édité 1989), *Uli Malbon* (Grammaire du coréen), Séoul : Jengeum Munhwasa.
- Choi, Key-Sun et al., 1994, A Two-Level Morphological Analysis of Korean, *Proceedings of The 15th International Conference on Computational Linguistics* (COLING '94), Kyoto, Japan.
- Choi, Key-Sun et al., 1996, *Korean Language Information Base*, Balanced Corpus of Korean Texts : KAIST.
- Clemenceau, David, 1993, *Structuration du lexique et reconnaissance de mots dérivés*, Thèse de Doctorat, Paris : Université Paris 7.
- Courtois, Blandine, 1987, *Dictionnaire électronique du Laboratoire d'Automatique Documentaire et Linguistique pour les mots simples du français* (DELAS), Rapport Technique du LADL, n°17, Paris : Université Paris 7.
- Courtois, Blandine, 1989, DELAF: *Dictionnaire électronique du LADL pour les mots fléchis du français*, Rapport Technique du LADL, N°20, Paris: Université Paris 7.
- Crochemore, Maxime et Rytter, Wojciech, 1994, *Text Algorithms*, Oxford University Press.
- Crochemore, Maxime et Hancart, C., 1997, *Automata for matching patterns*, Rapport Technique IGM-97-06, Université de Marne-la-Vallée.
- Danlos, Laurence, 1984, *Génération automatique de textes en langues naturelles*, Thèse de doctorat d'Etat, Université Paris 7.
- Dugas, André, 1990, La dictionnaire électronique des mots simples, *Langue Française* 87, Paris : Larousse.

- Fabre, André, 1986, Les adjectifs en *seulebda*, *Cahiers d'études coréennes*, N-4, Paris : Centre d'études coréennes, Collège de France.
- Garrigues, Mylène, 1993, Prépositions et noms de pays et d'îles : une grammaire locale pour l'analyse automatique des textes, *Linguisticae Investigationes*, XVII : 2, Amsterdam : J. Benjamins B.V..
- Giry-Schneider, Jacqueline, 1987, *Les prédicats nominaux en français : Les phrases simples à verbe support*, Genève : Droz.
- Gross, Gaston, 1989, *Les constructions converses du français*, Genève : Droz.
- Gross, Gaston, 1990, Définition des noms composés dans un lexique-grammaire, *Langue Française* 87, Paris : Larousse.
- Gross, Maurice, 1975, *Méthodes en syntaxe*, Paris ; Hermann.
- Gross, Maurice, 1981, Les bases empiriques de la notion de prédicat sémantique, *Langages* 63, Paris: Larousse.
- Gross, Maurice, 1987, The use of finite automata in the lexical representation of natural language, *Lecture Notes in Computer Science* 377, Springer-Verlag.
- Gross, Maurice, 1989, La construction de dictionnaires électroniques, *Annales des Télécommunications*, tome 44 N° 1:2, Issy-les-Moulineaux / Lannion : CNET.
- Guillet, Alain ; Christian Leclère, 1992, *La structure des phrases simples en français - Constructions transitives locatives*, Genève-Paris : Droz.
- Harris, Zellig S, 1976, *Notes du cours de syntaxe*, traduit par M. Gross, Paris : Le Seuil.
- Hong, Chai-Song, 1985, *Syntaxe des verbes de mouvement en coréen contemporain*, Amsterdam/Philadelphia : John Benjamins.
- Hong, Chai-Song, 1987, *Hyundai hanguge dongsagumuneui yengu* (Etudes de constructions verbales en coréen contemporain), Séoul : Tap Chulpansa.
- I, Ik-Seb ; Im, Hong-Bin, 1983, *Guge munbeblon* (Grammaire du coréen), Séoul : Hagyensa.
- Kim, Seg-Deug, 1992, *Ulimal Hyengtailon* (La morphologie du coréen), Séoul : Tap Chulpansa.
- Ko, Yeng-Geun, 1989, (ré-édit.1991), *Guge Hyengtailon Yengu* (Etude morphologique du coréen), Séoul : Editions de Séoul Université.
- Laporte, Éric, 1988, *Méthodes algorithmiques et lexicales de phonétisation de textes, Applications au français*, Thèse de doctorat, Université Paris 7.
- Laporte, Éric, 1997, Context-free Parsing with Finite-state Transducers, *Proceedings of The Third South American Workshop on String Processing: WSP 1996*, Ottawa, Carleton University Press.
- Meunier, Annie, 1981, *Nominalisations d'adjectifs par verbes supports*, Thèse de troisième cycle, LADL, Université Paris 7.
- Mohri, Mehryar, 1993, *Analyse et représentation par automates de structures syntaxiques composées. - Application aux complétives*, Thèse de doctorat, Université Paris 7.
- Molinier, Christian, 1984, Etude syntaxique et sémantique des adverbes en *-ment*, Thèse de doctorat d'Etat, Université de Toulouse-Le-Mirail.
- Monceaux, Anne, 1993, *La formation des noms composés de structure NOM ADJECTIF : Elaboration d'un lexique électronique*, Thèse de doctorat, Université Paris 7.
- Nam, Jee-Sun, 1990, Sur une construction $N_0 N_1$ -ita en coréen, *Linguisticae Investigationes*, Vol.XIV ; N° 2, Amsterdam : J. Benjamins B.V.

- Nam, Jee-Sun, 1991, *Etablissement du corpus des adjectifs coréens* : Rapport technique N° 30, Paris : Institut Blaise Pascal, Université Paris 7.
- Nam, Jee-Sun, 1992a, Corpus des adjectifs coréens : Constitution et classification, *XVème Congrès International des Linguistes*, Québec : Université Laval.
- Nam, Jee-Sun, 1992b, Formalisation des données lexico-syntaxiques dans le dictionnaire : adjectifs symétriques en coréen, *Papers in Computational Lexicography Complex '92*, ed. by Ferenc Kiefer, Gabor Kiss et Julia Pajzs, Budapest : Linguistics Institute, Hungarian Academy of Sciences.
- Nam, Jee-Sun, 1996a, *Dictionary of Noun-Postpositions and Predicate-Postpositions in Korean: DECOS-PostN / DECOS-PostA / DECOS-PostV*, Rapport technique N-51, LADL, Université Paris 7.
- Nam, Jee-Sun, 1996b, Construction of Korean electronic lexical system DECO, *Papers in Computational Lexicography Complex '96*, ed. by Ferenc Kiefer, Gabor Kiss et Julia Pajzs, Budapest : Linguistics Institute, Hungarian Academy of Sciences.
- Nam, Jee-Sun, 1996c, *Classification syntaxique des constructions adjectivales en coréen*, Amsterdam / Philadelphia : John Benjamins Publishing Company.
- Nam, Jee-Sun; Lee Chang-Yeol, 1997, *Lexicon of Nominalized Forms of Korean Predicates*, Rapport Technique, LADL, Université Paris 7.
- Nam, Ki-Shim, 1989, *Guge Wanhyeng Bomunbeb yengu* (Etudes des phrases complétives en coréen), Séoul : Gugehaghoi.
- Park, Se-Young et al., 1994, An Implementation of an Automatic Keyword Extraction System, *Proceedings of Pacific Rim International Conference on Artificial Intelligence '94*, Beijing, Chine.
- Park, Se-Young et al., 1996, *Corpus based on Korean News Papers* : ETRI.
- Perrin, Dominique, 1989, Automates et algorithmes sur les mots, *Annales des Télécommunications*, tome 44 N 1:2, Issy-les-Moulineaux, Lannion : CNET.
- Picabia, Lélia, 1978, *Les constructions adjectivales du français*, Genève : Droz.
- Revuz, Dominique, 1991, *Dictionnaires et Lexiques : Méthodes et algorithmes*, Thèse de doctorat, Université Paris 7.
- Roche, Emmanuel, 1993, *Analyse syntaxique transformationnelle du français par transducteurs et lexique-grammaire*, Thèse de doctorat, Université Paris 7.
- Silberztein, Max, 1990, Le dictionnaire électronique des mots composés, *Langue Française* 87, Paris : Larousse.
- Silberztein, Max, 1993, *Dictionnaires électroniques et analyse automatique de textes - le système INTEX*, Paris : Masson.

