

# Vers le traitement des groupes nominaux du coréen avec INTEX \*

Sun-Mee BAE

**Abstract:** The processing of Korean noun phrases in INTEX involves the solution of two specific problems induced by the agglutinative characteristics of Korean: the recognition of combinations of nouns and the recognition of combinations of noun suffixes. In order to deal with combinations of noun suffixes, we propose a process of segmentation by a finite-state transducer. The recognition of compound nouns requires describing them in an electronic dictionary according to their structure and to typographic conventions as regards spacing.

**Keywords:** noun suffix, compound noun, finite-state transducer, electronic dictionary      **Mots clés :** suffixe nominal, nom composé, transducteur, dictionnaire électronique.

## 1. Introduction

À cause du caractère agglutinant du coréen, la méthode de traitement automatique fondée sur le dictionnaire électronique des composés nominaux et des postpositions nominales se heurte à des difficultés spécifiques, notamment en ce qui concerne les combinaisons entre postpositions nominales et les combinaisons libres entre noms.

Les exemples de combinaisons de postpositions nominales s'observent fréquemment dans un texte <sup>1</sup> :

---

\* Je tiens à remercier Éric Laporte qui m'a dirigée dans ce travail.

---

✉ Sun-Mee BAE, Institut Gaspard Monge, Université de Marne-la-Vallée  
5, bd Descartes, Champs-sur-Marne  
F-77454 Marne-la-Vallée CEDEX 2 France  
Tel. : 33 - 1 60 95 77 15, Fax : 33 - 1 60 95 75 57      e-mail : sbae@monge.univ-mlv.fr

- (i) *Max-ga chingu-deul-eigei-man-eun senmul-eul ju-ess-da*  
 Max-Nmtf ami-SfxPl-Postp.eigei-Aux-Aux cadeaux-Acc donner-  
 Mpas-St  
 (Max a donné des cadeaux au moins seulement aux amis)

Dans l'exemple (i), le substantif *chingu* prend une combinaison de quatre suffixes nominaux *deul-eigei-man-eun*. Or, lorsque deux éléments se concatènent, il peut se produire des modifications formelles.

- a. *chingu-deul-eigei-neun* (après voyelle)  
 b. *chingu-deul-eigei-man-eun* (après consonne)

Dans les exemples ci-dessus, *-ei* du (a) est une voyelle et *-n* du (b) est une consonne. La postposition finale modifie sa forme morphologique en fonction d'une condition phonologique (*-neun* après voyelle, *-eun* après consonne). Il s'agit d'une variation phonologiquement conditionnée.

Maintenant, considérons un exemple de combinaison de noms :

- (ii) *gughoi euijang tupyos silsi* (N N N N)  
 « l'Assemblée nationale » « président » « vote » « exécution » =  
 « exécution du vote du député qui représente l'Assemblée nationale »

Cette séquence est un groupe nominal libre comprenant une séquence figée *gughoi euijang* (*député qui représente l'Assemblée nationale*). En fonction de leur degré de figement, les séquences de noms forment un continuum entre les séquences figées et les séquences nominales libres. Les conventions typographiques d'espacement compliquent ce problème.

Nous allons examiner ces deux problèmes des groupes nominaux du coréen et présenter des éléments de solutions dans les sections 2.5. Les sections 2 et 3 seront consacrées à la classification et à la combinatoire des postpositions nominales. Les sections 4 et 5 présenteront le problème des séquences nominales et la description des composés

---

<sup>1</sup> Nous utiliserons le tiret pour les frontières des mots et de chaque élément fonctionnel. Toutes les unités séparées par un tiret dans les exemples coréens s'écrivent en fait soudées.

nominaux dans un dictionnaire électronique en fonction de leur structure et des conventions typographiques d'espacement.

## 2. Critères de classification des postpositions nominales

En coréen, la postposition nominale constitue une des parties du discours. Les postpositions sont classées en trois grandes catégories, les postpositions casuelles, auxiliaires et conjonctives<sup>1</sup>, en fonction de critères syntaxiques et distributionnels. Ces deux types de critères ont des relations étroites.

### 2.1. Critère syntaxique

#### 2.1.1. *Postpositions casuelles*

Dans la phrase coréenne, le déplacement des constituants est relativement libre par rapport au français ou à l'anglais. La postposition casuelle marque les rôles syntaxiques des substantifs dans une phrase. Ainsi, la postposition nominative *-ga/-i* attribue le cas nominatif aux substantifs. Les actants, y compris le sujet et l'objet direct, sont donc distinguables grâce aux postpositions<sup>2</sup>. Les postpositions casuelles attachées aux noms sont sous-catégorisées en six cas d'après la forme morphologique propre : le cas nominatif, le cas accusatif, le cas copulatif, le cas vocatif, le cas génitif, et une catégorie résiduelle qualifiée d'adverbiale ((1a) – (1f)).

- (1a) *Max-ga haggyo-ei ga-nda*  
 Max-Nmtf école-Postp.ei aller-St  
 (Max va à l'école)
- (1b) *Max-ga Léa-leul salangha-nda*  
 Max-Nmtf Léa-Acc aimer-St  
 (Max aime Léa)

<sup>1</sup> Cf. KIM 1992, NAM et GO 1998, LEE et LEE 1998.

<sup>2</sup> a. *Max-ga Léa-leul salangha-nda*  
 Max-Nmtf Léa-Acc aimer-St  
 (Max aime Léa)

b. *Léa-leul Max-ga salangha-nda*  
 Léa-Acc Max-Nmtf aimer-St  
 (Max aime Léa)

- (1c) *geuges-eun jongi-ga ani-da*  
Ce-**Aux** papier-**Copu** (ne pas être)-**St**  
(Ce n'est pas du papier)
- (1d) *Max-ya, bbali o-nela*  
Max-**Voc** vite venir-**Stim**  
(Max, viens vite)
- (1e) *Max-eui chag-eul da-o*  
Max-**Gén** livre-**Acc** donner-**Stim**  
(Donne-moi le livre de Max)
- (1f) *Max-ga Léa-eigei ggoch-eul ju-ess-da*  
Max-**Nmtf** Léa-**Postp.eigei** fleurs-**Acc** donner-**Mpas-St**  
(Max a donné des fleurs à Léa)

Le nom propre *Max* de (1a) est marqué comme sujet dans cette phrase par le cas nominatif *-ga*. *Léa* dans (1b) devient l'objet direct dans cette phrase à l'aide du cas accusatif *-leul*. La postposition *-gal-i* accolée au verbe *anida* (*ne pas être*) ou *doida* (*devenir*) marque le cas copulatif ((1c)), non le nominatif. La postposition vocative de (1d) *-ya* donne la valeur d'appel à *Max*. La postposition du génitif dans (1e) fait de *Max* un modifieur du substantif *chag* (*livre*). Dans l'exemple (1f), *Léa* devient le datif par la postposition *-eigei*. *Léa* est un complément en *-eigei*. L'instrumental (*-lossel-eulosse*), le locatif (*-ei, -eise*), le comparatif (*-boda, -mankeum*) etc. sont d'autres exemples. Nous utiliserons *Postp.losselosse*, *Postp.ei*, *Postp.eigei*, etc. pour ce type de postpositions.

### 2.1.2. Postpositions auxiliaires

La postposition auxiliaire ajoute une nuance sémantique à la phrase. Elle n'a pas de rôle syntaxique direct. La fonction syntaxique d'un substantif auquel elle s'attache n'est pas déductible des postpositions auxiliaires. Par exemple, la postposition auxiliaire *-neun* a la même forme dans les différentes positions syntaxiques. Voici quelques exemples :

- (2a) *Léa-neun haggyo-ei ga-ss-da*  
Léa-**Aux** école-**Postp.ei** aller-**Mpas-St**  
(C'est Léa qui est allée à l'école)
- (2b) *Max-ga Léa-neun salangha-nda*  
Max-**Nmtf** Léa-**Aux** aimer-**St**

(C'est Léa que Max aime)

- (2c) *Max-ga Léa-neun ggoch-eul ju-ess-da*  
 Max-Nmtf Léa-Aux fleurs-Acc donner-Mpas-St  
 (C'est Léa à qui Max a donné des fleurs)

Parmi les trois phrases précédentes, le nom propre *Léa* joue des rôles différents : dans (2a), le sujet ; dans (2b), le complément d'objet direct et dans (2c), le complément d'objet indirect. Les postpositions auxiliaires comme *-do* (*aussi*), *-neun* (*c'est...que*), *-jocha* (*même*), *-man* (*ne...que*), *-mada* (*chaque*), etc. apportent à la phrase une nuance sémantique spécifique<sup>1</sup>.

- (3a) *Max-ga Léa-man salangha-nda*  
 Max-Nmtf Léa-Aux aimer-St  
 (Max n'aime que Léa)
- (3b) *Max-ga Léa-do salangha-nda*  
 Max-Nmtf Léa-Aux aimer-St  
 (Max aime aussi Léa)

### 2.1.3. Postpositions conjonctives

Les postpositions conjonctives lient des substantifs ayant la même fonction syntaxique dans la phrase. Des substantifs liés par une postposition conjonctive sont considérés comme un constituant.

- (4) *Max-wa Léa-ga haggyo-ei ga-nda*  
 Max-Conj Léa-Nmtf école-Postp.ei aller-St  
 (Max et Léa vont à l'école)

Dans l'exemple ci-dessus la postposition de coordination *-wa* lie *Max* et *Léa* qui devient une unité suivie de la postposition nominative *-ga*. Les postpositions conjonctives ressemblent à des postpositions casuelles par plusieurs aspects. Mais elles ne peuvent pas attribuer de fonction syntaxique aux substantifs et seulement les coordonner.

---

<sup>1</sup> En revanche, la postposition de cas nominatif, *-ga* de (1a) et la postposition de cas accusatif, *-leul* de (1b) ne portent pas de sens. Le substantif accolé avec le cas nominatif *-ga* est le sujet et *-leul*, l'objet direct. La postposition conjonctive *-wa* (*et*) de (4) n'a que le sens de coordination.

## 2.2. Critères distributionnels

### 2.2.1. Possibilité de combinaison avec l'adverbe

Certaines postpositions auxiliaires peuvent s'attacher à des adverbes pour ajouter une nuance sémantique. Les postpositions casuelles et conjonctives n'ont pas cette propriété.

- (5a) *Léa-ga jal-do ged-neunda*  
Léa-Nmtf bien-Aux marcher-St  
(Léa marche bien aussi)
- (5b) *Léa-ga jal-man ged-neunda*  
Léa-Nmtf bien-Aux marcher-St  
(Léa ne marche que bien)
- (5c) \**Léa-neun jal (-ga, -eul, -eui, -lo, -ya, -wa) ged-neunda*  
\*Léa-Aux bien (Nmtf, Acc, Gén, Postp.lo, Voc, Conj) aller-St

### 2.2.2. Omission

Les postpositions casuelles et conjonctives peuvent être omises dans la langue orale si et seulement si les fonctions syntaxiques des substantifs sans postpositions sont suffisamment claires. En particulier, les cas nominatif et accusatif sont aisément omis. En revanche, les postpositions auxiliaires ne peuvent pas être omises sans changement de sens. Par exemple :

- (6a) *geu namja (ga + E) ddena-ss-ni?*  
Cet homme (Nmtf + E) partir-Mpas-Sti  
(Cet homme, est-il parti ?)
- (6b) *Na-do i chag (eul + E) ilg-ess-e*  
Je-Aux ce livre (Acc + E) lire-Mpas-St  
(Moi aussi, j'ai lu ce livre)
- (6c) *Haggyo (ei + E) gani?*  
école (Postp.ei + E) aller-Sti  
(Est-ce que tu vas à l'école ?)
- (6d) *donggul (eui + E) ibgu*  
grotte (Gén + E) entrée  
(L'entrée de la grotte)
- (6e) *chag (gwa + E), yepil-eul da-o*  
livre (Conj + E), stylo-Acc donner-Stim  
(Donnez-moi un livre et un stylo)

Nous pouvons résumer les critères de classification des postpositions nominales dans la figure 1.

		Postp Casuelle	Postp Aux	Postp Conj
Critères syntaxiques	Marque la fonction syntaxique des actants	Oui	Non	Non
Critères distributionnels	Possibilité de combinaison avec l'adverbe	Non	Oui	Non
	Omission	Oui	Non	Oui

Figure 1. Les critères de classification des postpositions nominales

### 3. Combinatoire des postpositions nominales

#### 3.1. Choix de la forme canonique et variantes phonologiquement conditionnées

Comme nous l'avons mentionné à propos de l'exemple (i) dans l'introduction, certains éléments morphologiques peuvent avoir plusieurs formes en fonction de conditions phonologiques ou morphologiques. Par exemple, le morphème *-neum* qui correspond à l'expression *c'est...que* en français possède deux formes dont l'emploi obéit à des conditions phonologiques sur le contexte gauche :

*-neum* / V \_\_\_\_\_  
*-eum* / C \_\_\_\_\_

On peut sélectionner une des deux formes comme forme canonique. Le choix de la forme canonique est arbitraire, mais il a généralement une motivation liée à l'organisation générale de la description. Comme des variations similaires sont observées avec d'autres suffixes, on a adopté une règle simple : la variante observée après voyelle est choisie comme forme canonique. Ainsi, *-neum*, *-ya*, *-leul*, *-wa*, *-lo*, etc. sont des formes canoniques. Chacune de ces formes est en distri-

bution complémentaire avec une ou plusieurs formes morphologiquement apparentées.

### 3.2. Variantes libres

D'autre part, il existe aussi des variantes interchangeables sans modification de sens. Nous avons considéré les variantes comme libres si les trois conditions suivantes sont réunies :

- (i) Les deux morphèmes appartiennent à la même classe de postpositions
- (ii) La combinatoire avec les autres postpositions est la même
- (iii) Le sens ne change pas en cas de substitution

Par exemple :

- (7a) *Na-neum Léa-leul salangha-nda*  
Na-Aux Léa-Acc aimer-St  
(C'est moi qui aime Léa)
- (7b) *Na-n Léa-leul salangha-nda*  
Na-Aux Léa-Acc aimer-St  
(C'est moi qui aime Léa)

La relation entre *-neum* et *-eum* dans l'exemple (i) est phonologiquement conditionnée. En revanche, dans (7), la relation entre *-neum* et *-n* est libre : *-neum* et *-n* sont interchangeables sans modification de sens. On utilise souvent *-n* dans la langue orale.

### 3.3. Combinatoire des postpositions nominales

Une fois choisie une forme canonique de chaque postposition nominale, la description systématique des combinaisons de postpositions et de leurs variations peut être faite par référence aux formes canoniques. Voici la combinatoire des formes canoniques des postpositions nominales, représentée sous la forme d'un automate fini.



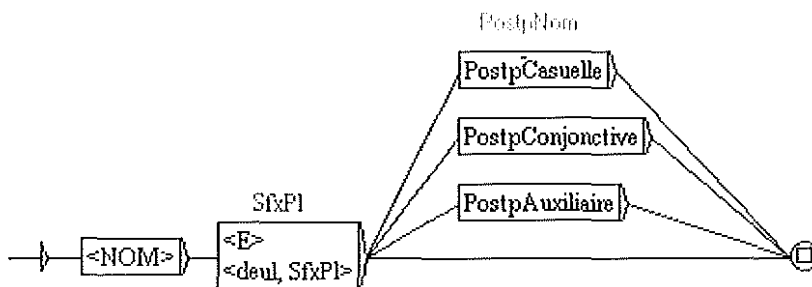


Figure 2. Combinatoire des postpositions nominales

Pour formaliser les conditions phonologiques nous avons classé les formes à l'aide de six codes<sup>1</sup>. Extrayons plusieurs entrées de la liste des postpositions nominales codées :

Forme canonique	Variante	Classe	Code
Ga	Ga	Copu	VV
	I	Copu	CV
Ga	Ga	Nmtf	VV
	I	Nmtf	CV
Gati	Gati	Postp	XV
	Chelem	Postp	XC
:	:	:	:

<sup>1</sup> Classification des formes de postpositions nominales en vue de la formalisation des conditions phonologiques des variations.

	Après voyelle	Après consonne	Après voyelle ou consonne
Fin voyelle	VV	CV	XV
Fin consonne	VC	CC	XC

Enfin, on peut engendrer les 2 400 combinaisons<sup>1</sup> de postpositions nominales à partir des 73 formes canoniques et de leurs 124 variantes. On engendre la combinatoire complète des variantes sans contraintes phonologiques. On construit un transducteur dont les entrées représentent toutes ces formes et dont les sorties donnent les formes canoniques. Les contraintes  $(A^*\backslash((CCIVCIXC)(VCIVV)|(VVICVIXV)(CVICC)))$  sont alors appliquées : l'intersection entre l'entrée du transducteur, qui représente la combinatoire complète, et l'automate qui représente les contraintes donne la combinatoire observable.

#### 4. Séquences de substantifs

Maintenant, considérons le problème des combinaisons de noms coréens. L'agglutination de noms en séquences libres et les irrégularités dans l'emploi de l'espace typographique posent de grandes difficultés dans la délimitation et le recensement des séquences nominales figées. La délimitation des séquences nominales figées par opposition aux séquences nominales libres, qu'on estime ne pas pouvoir répertorier dans un premier temps, est un problème fondamental.

En coréen, les séquences composées de type  $NN^+$ , c'est-à-dire  $NN$ ,  $NNN$ ,  $NNNN$ ..., constituent la grande majorité des composés nomi-

<sup>1</sup> Voici une partie de la liste des combinaisons de postpositions nominales.

Combinatoire des variantes	Combinatoire des formes canoniques
<i>gatimaje</i>	<gati Postp><maje Aux>
<i>gatimajen</i>	<gati Postp><maje Aux><neun Aux>
<i>gatimajendeul</i>	<gati Postp><maje Aux><ndeul Aux>
<i>gatimajenkenyeng</i>	<gati Postp><maje Aux><neunkenyeng Aux>
<i>gatimajeneun</i>	<gati Postp><maje Aux><neun Aux>
<i>gatimajeneunkenyeng</i>	<gati Postp><maje Aux><neunkenyeng Aux>
<i>gatimajedo</i>	<gati Postp><maje Aux><do Aux>
<i>gatimajelado</i>	<gati Postp><maje Aux><lado Aux>
<i>gatimajeindeul</i>	<gati Postp><maje Aux><ndeul Aux>
<i>gatiman</i>	<gati Postp><man Aux>
<i>gatimando</i>	<gati Postp><man Aux><do Aux>
<i>gatimaneun</i>	<gati Postp><man Aux><neun Aux>
<i>gatimaneul</i>	<gati Postp><man Aux><leul Acc>
<i>gatimani</i>	<gati Postp><man Aux><ga Copu>
<i>gatimanilado</i>	<gati Postp><man Aux><lado Aux>

naux. Les combinaisons de noms sont plus libres en coréen qu'en français et sont en nombre théoriquement infini. En fonction de leur degré de figement, ces séquences forment un continuum entre les séquences nominales figées et les séquences nominales libres. Les séquences libres sont celles qui obéissent simplement aux règles combinatoires générales de la langue, et les séquences figées, par définition, celles qui ne sont pas tout à fait libres dans la combinaison syntaxique et sémantique de leurs éléments. Le sens des séquences libres peut être déduit de celui de leurs éléments. On appellera celles-ci les *groupes nominaux libres* et celles-là les *noms figés*. Il n'est pas facile de distinguer les noms figés et les groupes nominaux libres selon des critères précis<sup>1</sup>.

Considérons des exemples de combinaisons de noms :

- (8a) *gughoi* (N)  
« l'Assemblée nationale »
- (8b) *gughoi euijang* (N N)  
« l'Assemblée nationale » « président » =  
« député qui représente l'Assemblée nationale »
- (8c) *gughoi euijang tupyoo* (N N N)  
« l'Assemblée nationale » « président » « vote » =  
« vote du député qui représente l'Assemblée nationale »
- (8d) *gughoi euijang tupyoo silsi* (N N N N)  
« l'Assemblée nationale » « président » « vote » « exécution » =  
« exécution du vote du député qui représente l'Assemblée nationale »

L'exemple (8b) est un nom figé puisque ses deux éléments constituent une unité lexicale figée et que le sens n'est pas tout à fait transparent. En revanche, les séquences (8c) et (8d) sont des groupes nominaux libres contenant le nom figé *gughoi euijang*. Elles sont constituées à partir de la libre combinaison de *gughoi euijang* et *tupyoo* ou de *gughoi euijang* et *tupyoo silsi*.

---

<sup>1</sup> Les informaticiens qui travaillent sur le coréen traitent généralement les noms figés et libres de la même façon et mènent des recherches sur les algorithmes de décomposition parce que les groupes nominaux sans blanc sont considérés comme non enregistrables (cf. KANG 1998, YOON *et al.* 1995, CHOI 1996).

## 5. Classification des composés nominaux dans un dictionnaire électronique

### 5.1. Différences formelles entre les composés français et coréens

Nous allons commencer par mentionner les différences entre les composés français et coréens en ce qui concerne les conventions typographiques.

En coréen, les mots composés ne sont pas reliés par un trait d'union ni par une apostrophe. Tandis qu'un trait d'union ou une apostrophe en français peuvent servir de marqueur de composition des éléments, c'est la soudure qui joue ce rôle en coréen. Par exemple, les mots *si-emeni* (PF-N)<sup>1</sup> (*côté du mari – mère = belle-mère*) ou *bel-sai* (N-N) (*abeille – oiseau = oiseau-mouche*) sont soudés<sup>2</sup>, alors que leurs équivalents français se réalisent graphiquement avec un trait d'union (*belle-mère, oiseau-mouche*). Ainsi, il n'est pas étonnant que les formes graphiquement soudées de la grande majorité des noms dérivés et des noms composés au sens traditionnel de la linguistique coréenne (cf. plus loin) correspondent à la typologie des composés graphiques du français<sup>3</sup>. Deuxièmement, l'espacement entre les mots est très irrégulier. Pour la grande majorité des séquences de noms, en principe, on met un espace, mais il est acceptable de ne pas espacer. Par exemple, une même séquence *huchenseng myenyeg gyelpibjeung* (N N N) (*qualité acquise – immunité – symptôme de déficience = syndrome immuno-déficitaire acquis, sida*) peut se réaliser graphiquement de quatre façons possibles comme suit :

- (9) huchenseng myenyeg gyelpibjeung (N N N)  
 huchenseng myenyeggyelpibjeung (N NN)  
 huchensengmyenyeg gyelpibjeung (NN N)  
 huchensengmyenyeggyelpibjeung (NNN)

<sup>1</sup> PF signifie un préfixe.

<sup>2</sup> La forme *bel sai* (N N), non soudée, signifie « une abeille et un oiseau ».

<sup>3</sup> Cf. MATHIEU-COLAS 1989.

## 5.2. Classification des composés nominaux

### 5.2.1. Noms compacts, obligatoirement soudés

Introduisons la terminologie traditionnelle de la grammaire coréenne sur la formation des noms. Ils sont classés en noms simples et noms complexes. Le nom simple se définit comme toute unité non décomposable (par exemple, *emeni* « mère », *nai* « âge » ou *tei* « anneau »). Tout nom qui n'est pas nom simple comporte au moins deux éléments tels qu'un affixe, un nom simple ou une racine. On l'appelle un nom complexe. Parmi les noms complexes, on distingue les noms dérivés et les noms composés. Un nom dérivé comme *si-emeni* (*belle-mère*) est un nom qui constitue une unité lexicale nouvelle à partir d'éléments qui ne sont pas tous susceptibles d'emploi indépendant. Un nom composé comme *nai-tei* (*anneau d'accroissement*) est un nom qui constitue une unité sémantique mais qui est constitué d'éléments susceptibles d'avoir par eux-mêmes une autonomie dans la langue. Enfin, certains noms qui s'écrivent dans l'usage effectif avec un espacement typographique facultatif, comme (12a) et (14a), sont traditionnellement qualifiés de noms composés.

Pour éviter des ambiguïtés avec la terminologie traditionnelle, nous appelons *noms compacts* les noms obligatoirement soudés, qu'ils soient simples ou complexes (par exemple, *nai*, *si-emeni*, *nai-tei*). Nous appelons *noms figés* les séquences nominales figées à espacement facultatif ou obligatoire dans l'usage effectif<sup>1</sup>. Le dictionnaire des noms compacts du coréen sera donc constitué de noms simples, de noms dérivés et de noms composés au sens traditionnel (ou de la formation des mots).

### 5.2.2. Noms figés, à espacement facultatif ou obligatoire

Par définition, les noms figés ne sont pas entièrement libres dans la combinaison syntaxique et sémantique entre leurs éléments et présentent un espacement facultatif ou obligatoire. Les noms figés à espacement facultatif sont de la forme  $NN^+$ , et les noms figés à espacement obligatoire se répartissent en plusieurs structures comme *N-Postp N*,

---

<sup>1</sup> Une forme comme *gughoi euiwen* (*député*) sera considérée comme un nom figé en raison de l'espacement facultatif.

*A-Sfd N, N V-Sfd N*, etc. Sans entrer dans le détail de leur classification, nous nous contenterons de présenter plusieurs exemples<sup>1</sup> à répertorier dans le dictionnaire électronique des noms figés. Nous tiendrons compte de l'origine sino-coréenne ou proprement coréenne des éléments des noms figés à espacement facultatif.

### 5.2.2.1. *Les noms figés sc sc*<sup>2</sup>

Le vocabulaire sino-coréen est historiquement emprunté au chinois<sup>3</sup>. Chaque idéogramme chinois emprunté est associé à une forme monosyllabique et à une signification, mais en général il n'a pas d'autonomie syntaxique dans le système du coréen<sup>4</sup>. La plupart des noms sino-coréens sont formés par une combinaison de plusieurs idéogrammes, en général deux. Au niveau sémantique, la plupart des noms disyllabiques sino-coréens sont figés.

Les séquences figées sino-coréennes sont peu étudiées, mais elles constituent une grande partie des noms figés du type *NN*<sup>+</sup>. Considérons les exemples suivants :

- (10) *talji ^ bunyu*<sup>5</sup>  
 « écrémage » « lait en poudre » = « lait écrémé en poudre »  
*goso ^ gongpojeug*  
 « lieu élevé » « phobie » = « acrophobie »

Ces exemples sont généralement soudés dans les dictionnaires commerciaux, mais il n'existe pas de consensus, même entre dictionnaires commerciaux, sur les conventions d'espacement à leur appli-

<sup>1</sup> Ces exemples ne représentent pas tous les types de noms figés.

<sup>2</sup> *sc* indique un mot sino-coréen.

<sup>3</sup> Les mots sino-coréens constituent une grande partie du lexique coréen.

<sup>4</sup> Par exemple, *chun* (*printemps*) et *chu* (*automne*) sont des noms autonomes dans le système du chinois. Mais ces formes monosyllabiques deviennent non autonomes en coréen :

\**chun-i* « printemps – *Nmtf* »

\**chu-ga* « automne – *Nmtf* »

*chunchu-ga* « âge – *Nmtf* »

<sup>5</sup> ^ indique l'espacement facultatif.

quer, et on les trouve souvent aussi non soudés dans les textes. Il faut donc les considérer comme des noms figés à espacement facultatif.

Les exemples suivants illustrent un autre cas de désaccord entre l'usage et les dictionnaires :

- (11) *gughoi ^ euijang*  
 « l'Assemblée nationale » « président » = « député qui représente l'Assemblée nationale »  
*chuli ^ sospel*  
 « raisonnement » « roman » = « roman policier »  
*gosog ^ cheldo*  
 « grande vitesse » « chemin de fer » = « chemin de fer rapide, TGV »

Pour ces exemples, on n'insère généralement pas de blanc entre les deux unités dans l'usage effectif, mais ces séquences apparaissent souvent avec un blanc dans les dictionnaires éditoriaux<sup>1</sup>. Pour ces groupes nominaux on ne rencontre pas particulièrement de difficulté à décider de leur figement. Leur sens n'est pas entièrement transparent. Donc nous proposons d'inclure ce genre de séquences dans le dictionnaire des noms figés et non dans le dictionnaire des noms compacts.

#### 5.2.2.2. *Les noms figés pc pc*<sup>2</sup>

Les noms figés proprement coréens figurent en général sans espacement dans les dictionnaires éditoriaux et dans les textes, et le figement sémantique est plus net que pour les séquences sino-coréennes<sup>3</sup>. C'est pourquoi les suites figées proprement coréennes sont plus aisément considérées comme des noms composés au sens traditionnel<sup>4</sup>. Cependant, certaines des formes classées comme telles par les grammairiens peuvent apparaître avec un espace dans les textes :

<sup>1</sup> Du point de vue diachronique, les noms figés du coréen comme *gughoi euiwen* (l'Assemblée nationale – membre = député) tendent peu à peu à se souder typographiquement et à devenir des noms compacts.

<sup>2</sup> *pc* signifie un mot proprement coréen.

<sup>3</sup> Par exemple, *jip-salam (N-N)(pc-pc)* « maison » « personne » = « épouse »

<sup>4</sup> En fait, les études sur les noms composés sont souvent sur les séquences proprement coréennes et les critères de définition du nom composé au sens traditionnel s'appliquent bien aux séquences proprement coréennes.

- (12a) *dwaiji ^ gogi*  
 « cochon » « viande » = « viande de porc »  
*nelbi ^ ddwigi*  
 « largeur » « saut » = « saut en longueur »  
*nopi ^ ddwigi*  
 « hauteur » « saut » = « saut en hauteur »

La séquence nominale *dwaiji gogi* est répertoriée sans blanc dans tous les dictionnaires commerciaux, mais on ne peut pas négliger la possibilité de la rencontrer avec un blanc. Le nombre de syllabes<sup>1</sup> joue également un rôle : lorsqu'un mot est figé avec un mot monosyllabique, ils sont presque toujours soudés. Ainsi, les noms à 3 syllabes sont compacts :

- (12b) *yang-gogi*  
 « mouton » « viande » = « viande de mouton »  
*dag-gogi*  
 « poulet » « viande » = « viande de poulet »  
*ddwim-ddwigi*  
 « course » « saut » = « saut sur place »

En cas de désaccord entre l'usage et les grammairiens, on ne peut analyser les formes que comme erreurs de l'utilisateur ou variantes acceptables. Pour reconnaître les formes qui contiennent un espace et les formes qui sont soudées à d'autres formes, on a besoin d'un dictionnaire qui précise les possibilités d'espacement. On doit donc répertorier *dwaiji gogi* dans le dictionnaire comme nom figé et non comme nom compact. En revanche, pour les cas tels que (12b) où la soudure est clairement obligatoire, nous les traiterons comme noms compacts.

Nous avons aussi rencontré quelques noms répertoriés avec un blanc dans les dictionnaires éditoriaux, mais qu'on trouve aussi soudés dans les textes :

- (13) *pari ^ mogsum*  
 « mouche » « vie » = « vie de peu de valeur »

---

<sup>1</sup> Face à des mots de 4 syllabes ou plus, l'usage coréen est hésitant du point de vue de l'espacement graphique. Moins les éléments sont figés, plus l'hésitation augmente.



*hojumeni* ^ *sajeng*

« poche » « condition » = « condition économique »

Ces exemples sont sémantiquement très figés.

### 5.2.2.3. Les noms figés pc sc et sc pc

Nous ne pouvons exclure les noms figés constitués d'un nom proprement coréen et d'un nom sino-coréen :

(14a) *mailhwa-namu* (sc-pc)

« fleur du prunellier » « arbre » = « prunellier »

(14b) *minsog* ^ *noli* (sc-pc)

« coutume d'un peuple » « jeu » = « jeu transmis par le folklore »

*ius* ^ *sachon* (pc-sc)

« voisinage » « cousin » = « voisin familial comme un cousin »

L'exemple (14a) est traditionnellement qualifié de nom composé. Mais on peut facilement aussi le trouver non soudé dans les textes. Au contraire, les exemples (14b) ne sont pas soudés dans les dictionnaires, alors qu'on peut facilement les trouver soudés.

### 5.2.2.4. Noms figés à espacement obligatoire

Il existe plusieurs structures de noms figés à espacement obligatoire. Une étude approfondie est donc nécessaire. Nous nous contentons ici de citer des exemples représentatifs des structures *N-Postp N*, *A-sfd N* et *N V-sfd N* :

(15a) *bingsan-eui ilgag* (*N-Postp N*)

« iceberg – Gén » « coin » = « la partie émergée de l'iceberg »

(15b) \**bingsan ilgag* (*NN*)

« iceberg » « coin » = « ? »

(16) *bbyojogha-n su* (*A-sfd N*)

« être pointu – sfd » « moyen » = « moyen prodigieux »

(17a) *ko mud-eum don* (*N V-sfd N*)

« morve » « s'attache à – sfd » « argent » = « argent de poche des enfants »

Dans (15a), la postposition du génitif est obligatoire. Dans (17a), le substantif *ko* apparaît obligatoirement sans la postposition du nominatif pour exprimer le sens figé « argent de poche des enfants ».

L'insertion de la postposition fait disparaître ce sens au profit du sens littéral :

(17b) *ko-ga mud-eun don (N-Postp V-sfd N)*

« morve – *Nmtf* » « s'attache à – *sfd* » « argent » = « argent auquel la morve s'attache »

Étant donné les exemples que nous avons vus dans cette section 5.2, nous pouvons comparer l'espacement des noms dans le dictionnaire et l'usage effectif :

		Dictionnaire	Usage	Effectif
Nom compact	Nom simple	Soudé	Soudé	Très nombreux
	Nom dérivé			
	Nom composé au sens traditionnel			
Nom figé (sc sc)		Soudé	Facultatif	Nombreux
Nom figé (sc sc)		Non soudé	Facultatif	Très nombreux
Nom figé (pc pc)		Soudé	Facultatif	Limité
Nom figé (pc pc)		Non soudé	Facultatif	Très limité
Nom figé (pc sc ou sc pc)		Soudé	Facultatif	Nombreux
Nom figé (pc sc ou sc pc)		Non soudé	Facultatif	Limité
Nom figé (N postp. N, Adj N, ...)		Non soudé	Non soudé	Limité

Figure 4. Espacement des noms dans le dictionnaire et dans l'usage

## 6. Conclusion

Dans cet article, nous avons abordé deux problèmes fondamentaux du traitement informatique des groupes nominaux en coréen : les séquences de postpositions nominales et les séquences de noms.

Pour le traitement des combinaisons entre suffixes nominaux, nous les avons segmentées et décrites systématiquement à l'aide d'une grammaire locale, et construit un transducteur fini dont les entrées sont les formes observables, avec prise en compte des variations, et qui les traduit en étiquettes lexicales comportant les formes canoniques des suffixes.

Pour le problème des composés nominaux, nous avons constaté que les combinaisons dans lesquelles entrent les substantifs coréens (surtout les combinaisons de mots sino-coréens) et l'espacement irrégulier posent de grandes difficultés dans la délimitation non seulement des séquences nominales figées vis-à-vis des séquences nominales libres mais aussi entre les différentes catégories de séquences nominales figées. Nous avons esquissé la structure d'un dictionnaire électronique des composés nominaux. Des critères plus élaborés, opératoires et systématiques sont indispensables pour délimiter précisément la frontière entre les noms compacts, les noms figés et les groupes nominaux libres.

## Bibliographie

- CHOI (Jai-Hyug) : 1996, « A Division Method of Korean Compound Noun », in *The 8th Hangeul mich Hanguge jengbocheli hagsul-balpyo nonmunjib*, p. 262–267.
- GROSS (Gaston) : 1996, *Les expressions figées en français* (Paris : Ophrys).
- LEE (Hee-Ja) et LEE (Jong-Hee) : 1998, Tegseuteu Bunsegjeg Guge Josaeui Yeungu ; *Études sur les postpositions nominales pour l'analyse textuelle* (Séoul : Hangug Munhwasa).
- KANG (Seung-Shik) : 1998, « A Decomposition Algorithm of Korean Compound Nouns », in *The Korean Information Science Society: Journal of KISS*, p. 172–182.
- KIM (Seug-Deug) : 1992, Ulimal hyengtailon ; *La Morphologie coréenne* (Séoul : Tapchulpansa).
- MATHIEU-COLAS (Michel) : 1989, *Typologie des noms composés*, Rapport n° 7 du programme de Recherche Coordonnée « Informatique Linguistique » (Paris : LADL).
- MATHIEU-COLAS (Michel) : 1988, « Variations graphiques des mots composés dans le *Petit Larousse* et le *Petit Robert* », dans *Lingvisticae Investigationes*, 17:2, p. 235–280
- MATHIEU-COLAS (Michel) : 1990, « Orthographe et informatique : établissement d'un dictionnaire électronique des variantes graphiques », dans *Langue française*, 87, p. 104–111.

- NAM (Gi-sim) et GO (Yong-geun) : 1998, Pyojun Gugemunbeblon ; *Grammaire Standard du coréen* (Séoul : Topchulpansa).
- SILBERZTEIN (Max) : 1990, « Le dictionnaire électronique des mots composés », dans *Langue française*, 87, p. 71-83.
- SILBERZTEIN (Max) : 1993, *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX* (Paris : Masson).
- YOON (Bo-Hyun) *et al.* : 1995, « A Method of Korean Compound Noun Analysis using Statistical Information », in *The Korean Information Science Society: Journal of KISS*, p. 925-928.

<b>NotationsAux</b>	postposition auxiliaire
<b>Conj</b>	postposition conjonctive
<b>Nmtf</b>	postposition du nominatif
<b>Acc</b>	postposition de l'accusatif
<b>Gén</b>	postposition du génitif
<b>Copu</b>	postposition du copulatif
<b>Voc</b>	postposition du vocatif
<b>SfxPl</b>	suffixe de pluriel
<b>Mpas</b>	suffixe de passé
<b>St</b>	suffixe terminal du mode déclaratif
<b>Sti</b>	suffixe terminal du mode interrogatif
<b>Stim</b>	suffixe terminal du mode impératif
<b>Sfd</b>	suffixe déterminatif
<b>Postp</b>	postposition d'un complément non accusatif
<b>Postp.eigei</b>	postposition du complément en <i>eigei</i>
<b>Postp.lo</b>	postposition du complément en <i>lo</i>