La construction de lexiques de formes fléchies et l'analyse morphologique du coréen

Université Paris 7
Laboratoire d'Automatique Documentaire et Linguistique
Institut Gaspard Monge

Thèse de Doctorat en Informatique

ChangYeol LEE

Directeur de thèse : Maurice GROSS

Membres du Jury : Dominique PERRIN
Eric LAPORTE
Gaston GROSS
Maurice GROSS
Maxime CROCHEMORE

Juin 1997

La construction de lexiques de formes fléchies et l'analyse morphologique du coréen

Université Paris 7
Laboratoire d'Automatique Documentaire et Linguistique
Institut Gaspard Monge

Thèse de Doctorat en Informatique

ChangYeol LEE

Directeur de thèse : Maurice GROSS

Membres du Jury : Dominique PERRIN
Eric LAPORTE
Gaston GROSS
Maurice GROSS
Maxime CROCHEMORE

Juin 1997

La construction de lexiques de formes fléchies et l'analyse morphologique du coréen

Résumé

Ce travail a pour objectif de développer un système d'analyse morphologique du coréen. La particularité de cette étude est qu'il repose sur un dictionnaire du coréen construit systématiquement pour cette application informatique. Les dictionnaires traditionnels, qu'ils soient destinés à un usage humain ou à une consultation informatique, n'ont pas de principes formels de représentation. Le système du dictionnaire utilisé dans notre étude est par contre défini par des principes généraux. Par exemple, il ne contient pas plusieurs fois les mêmes entrées, c'est un des critères formels de recensement linguistique que nous avons choisis.

Le mot "simple" est la plus petite unité dans nos dictionnaires. Dans les phrases coréennes, il est fréquent qu'une séquence délimitée par des séparateurs (comme des blancs, des points, ...) soit composée de plusieurs mots "simples". Ces mots simples peuvent être analysés et construits comme un jeu de LEGO et son démontage est facile.

Il n'y a pas de langues naturelles qui n'ait pas d'ambiguïté. Les méthodes de résolution d'ambiguïté utilisées actuellement sont basées sur un calcul statistique ou sur des règles. C'est parce que la connaissance qu'on a actuellement des langues naturelles ne garantit pas la complétude que l'on a dans un monde clos. Autrement dit, la connaissance linguistique ne peut pas être accumulée automatiquement, et donc elle doit être faite par l'être humain. C'est un des problèmes du monde ouvert.

Dans notre étude donc, la connaissance linguistique est accumulée seulement par les experts humains, non par une méthode statistique ou heuristique. Cela veut dire que la méthode linguistique est plus assurée, même si elle demande beaucoup de temps. L'information de désambiguïsation construite par les experts humains est utilisée dans notre système d'Analyseur morphologique. Pour la construction de l'information, il est inévitable d'analyser les structures internes ambiguës qu'on observe dans les phrases réelles.

Actuellement, on a construit des dictionnaires de Noms simples, de Postpositions nominales, de Préfixes, de Suffixes, de Verbes, d'Adjectifs et de Postpositions prédicatives. Sur la base de ces dictionnaires, on a dérivé deux dictionnaires des formes fléchies: le dictionnaire des Formes Nominalisées et celui des Formes fléchies prédicatives, aussi que des règles pour les Noms Dérivés. Les dictionnaires et l'information de désambiguïsation sont utilisés dans le processeur d'Analyseur morphologique.

Mots Clés:

Dictionnaire de mots simples, Dictionnaire de formes fléchies, Syllabes finales du radical, Postpositions prédicatives, Formes Nominalisées, Formes fléchies prédicatives, Lexique-grammaire, Consultation du dictionnaire, Analyseur morphologique.

한국어 활용어 사전 구축과 형태소 분석

요약

이 논문은 한국어 문장 분석에 필요한 사전 구축과 형태소 분석에 관한 연구 논문이다. 본 논문의 가장 큰 특징은 체계적으로 구성된 사전에 있다. 전통적 사전이 인간을 위한 것이든 기계를 위한 것이든 그 내용이 임의성이 많은데 비하여 여기의 사전은 단순어에 기반을 가지고 체계적이며, 중복성이 없고, 항목 선택의 명확한 기준이 있으며, 또한 구성적으로 구성되었다.

단순어란 의미의 최소 단위로서 통사적 독립성을 유지하고 있다. 한국어에 있어서 컴마나 빈칸등에 의해서 만들어지는 글자들은 여러개의 이들 단순어로 구성되어있다. 이 단순어들은 레고를 가지고 임의의 인조물을 만드는 것과 같이 문장의 합성과 분해에서 어휘의 확장성을 가지고 있으며 어휘의 특징을 추론할 수 있게 한다. 그들은 특별한 변화없이 문장에서 활용어 형태와 일치된다. 그러므로 활용어 사전의 구축은 형태소 분석기의 대부분을 차지하게 된다.

문장을 분해(해석)하는 데 있어서 모호성이 없는 시스템은 물론 없다. 모호성에 대하여 많은 시스템은 통계나 규칙을 가지고 문제에 대하여 접근하고 있다. 그것은 닫힌 세계에서의 지식의 완전성과 건전성 (Completeness and Soundness)을 자연어에서는 보장하지 않기 때문에 생기는 문제이다. 다시말하면 언어적 지식은 자동적으로 축적될 수 없으며 그러므로 인간에 의해서 행해져야 한다.

우리의 연구에 있어서, 언어적 지식은 통계나 휴리스틱한 방법이 아닌 오직 언어적 전문가에 의해서 축적을 한다. 다시말하면 이런 언어적 방법이 비록 많은 시간이 걸리지만 가장 정확성을 보장해준다. 언어적 방법에 의해서 구축된 모호성 해결 정보는 우리의 형태소 분석기에서 사용될 것이고, 그 정보의 구축을 위하여 어떤 모호성 구조가 있는가를 분석하는 것이 물론 전제 조건임은 언급할 필요가 없다.

현재 구축한 단순어 사전으로 단순 명사, 명사 구성 접두사, 명사 구성 접미사, 명사 후치사, 동사, 동사 활용어, 형용사, 형용사 활용어 사전등이 제공되고 있다. 이 사전들을 바탕으로, 문장 분석에 필요한 두개의 활용형 사전인 명사화 유형 사전, 술어-활용어 사전, 그리고 명사를 위한 문법적 규칙을 이 논문에서 구축하였다. 이 사전 정보와 모호성 해결 정보가 본 논문의 형태소 해석기인 품사 표시기(EPL)에서 활용되고 있다.

중요 단어

단순어 사전, 활용어 사전, 어간 끝 음절, 어미, 명사화 유형, 술어와 활용어, 사전 문법, 사전 검색, 형태소 해석기

The construction of lexicons of inflectional forms and morphological analysis in Korean

Abstract

This paper contains a study about the Korean sentence analysis. The characteristics of this study is that it is based on a well-formed dictionary system. Traditional dictionaries, either for human use or in a machine readable form, do not have any clear principle in their construction, but the dictionary system used in this study does not have any duplicate entries between sub-dictionaries; it has clear principles in the selection of entries, and has been systematically constructed with linguistic knowledge.

A simple word is the smallest unit in our dictionaries. A string made up by separators can be composed of several simple words in Korean sentences. These simple words permit to be analyzed and constructed like LEGO system which can freely be constructed and dis-assembled to any artifacts. The constructed words are matched with the inflectional forms in sentence.

There is no system which has no ambiguity. Several systems are based on statistics or heuristics methods to the ambiguity problems. It is caused by the facts that linguistic knowledge doesn't guarantee the completeness and soundness of the closed world assumption. In other words, linguistic knowledge can not be automatically accumulated, and therefore it must be constructed by human. It is an open world problem.

In this paper, the linguistic knowledge is accumulated only by human experts, not by statistics and heuristics methods. It means that the linguistic method is the best way we assure, although it consumes much time. Disambiguisation information constructed by humans is used in our morphological analyser. For the construction of this information, it is prerequisite to analyze the ambiguity structures in real sentences.

Currently, we have several simple dictionaries such as Simple Noun, Nominal-Postposition, Prefix, Suffix, Verb, Adjective, and Inflectional Suffix. Upon these dictionaries, we constructed two inflectional dictionaries, such as Nominalized Form dictionary and Predicate containing Inflectional Suffixes dictionary, and one grammatical rule for Noun phrases. The dictionaries and the disambiguitation information will be used in our lexical language processor.

Keywords

Simple Dictionary, Dictionary of Inflectional Forms, Final Syllables of Root, Inflectional Suffix, Nominalized Form, Predicate with Inflectional Suffix, Lexicon-Grammar, Dictionary Lookup, Morphological Analyser.

TABLE DES MATIERES

Principes de notation

Notations

Coréen	Français	Exemples
명사	Nom(N)	국세청장, 나무
단순명사	Nom Simple(NS)	음악, 나무
명사 접두어	Préfixe(PF)	국 de (국세청장)
명사 접미어	Suffixe(SF)	청, 장 de (국세청장)
파생명사	Noms Dérivés(ND)	신세대
동사	Verbe(V)	가다
형용사	Adjectif(A)	아름답다
서술명사	Nom Prédicatif (NP)	공부 de (공부하다)
명사 어미	Postpositions nominales (PPN)	은, 는, 이, 가, 께서도
명사화 유형	Forme Nominalisée(FN)	잠, 자기, 자는것,
명사화 동사	Verbes Nominalisés(VN)	잠
명사화 형용사	Adjectif Nominalisé(AN)	아름다움
술어와 활용어	Formes fléchies prédicatives(FFP)	자-시거나
어절	Séquences de mots	학교에
음절	Syllabes	학, 교, 에
단어	Mots	학교, 에
기본형	Formes de Base	기쁘다
어간	Radicaux	기쁘
활용어미	Postpositions prédicatives (PPP)	느냐, 구나, 으므로
어간끝 음절	Syllabes Finales du Radical (SFR)	쁘 de (기쁘다)
초성	Premières Consonnes(PC)	ㅎ de (학)
중성	Voyelles(VO)	ㅏ de (학)
종성	Consonnes Finales(CF)	ㄱ de (학)

Symboles

X^+ (= X^n $(n \geq 1)$

X^* (= X^n $(n \geq 0)$

$X^?$ (= X^n $(n = 0)$ ou $(n = 1)$)

$\prod_{i=1}^{n} Xi$ (= $X_1 * X_2 * X_3 * \ldots X_{n-1} * X_n$)

1. Introduction

Nous développons ici un système d'analyse automatique de textes écrits en coréen. A l'heure actuelle, de nombreux documents sont disponibles sous forme électronique que l'on peut consulter directement par ordinateur. Ainsi, pour obtenir des informations contenues dans des textes libres, il sera indispensable d'analyser la langue naturelle dans un système électronique.

Toutes les activités en traitement automatique des langues naturelles, notamment l'analyse automatique du texte, la documentation automatique et la traduction ou la génération automatique de textes nécessitent des dictionnaires électroniques[33]. Dans les applications à base d'analyse de textes, si une unité segmentée n'est pas trouvée dans le dictionnaire, l'analyse plus approfondie du texte qui devrait suivre la première phase sera bloquée. Il est donc indispensable de disposer de bases de données ainsi complètes que possible.

Dans le cas du français, la segmentation d'un texte s'effectue en unités graphiquement séparées par les blancs (et/ou aussi par les tirets, apostrophe ou point). Une chaî ne de caractères comprise entre deux séparateurs sera alors considérée comme mot simple, plus précisément comme forme fléchie de mot simple canonique. Ainsi, le dictionnaire électronique des mots simples du LADL (DELAS) [4] contient le codage méthodique des entrées qui permet d'engendrer automatiquement les formes fléchies du dictionnaire électronique des mots fléchis du LADL (DELAF) [5], c'est-à-dire les formes rencontrées dans les textes.

Pour le coréen, la situation est différente. Les unités séparées par les blancs et/ou par le point (les tirets et apostrophe n'existant pas en coréen) ne sont pas toujours des formes fléchies d'unités lexicalement simples, elles peuvent être constituées de séquences composées, éventuellement des syntagmes, des types :

- Nom - Postposition nominale(PPN)
- (Verbe+Adjectif) - Postposition prédicatif (PPP) (suffixes flexionnels)

Les postpositions, comparables aux prépositions en français, se trouvent obligatoirement soudées à un substantif pour indiquer le rôle grammatical du substantif associé ; les postpositions prédicatives sont soudées aux racines des verbes et des adjectifs (notons que les adjectifs coréens se conjuguent, comme les verbes, à l'aide de postpositions prédicatives, sans prendre de copule).

Le fait que certains morphèmes se trouvent attachés à des éléments lexicaux complique l'analyse morphologique. Surtout quand il existe plusieurs possibilités de segmenter des mots qui forment une seule séquence, les problèmes d'ambiguï tés deviennent alors plus complexes. Supposons qu'on ait une séquence de type :

abcd

Nous pouvons la segmenter des 8 manières suivantes :

(i)	a / b / c / d
(ii)	ab / c / d
(iii)	ab / cd
(iv)	a / b / cd
(v)	a / bc / d
(vi)	abc / d
(vii)	a / bcd
(viii)	abcd

Pour obtenir la(les) segmentation(s) correcte(s), un dictionnaire aussi complet que possible doit être fourni : s'il n'existe pas de mots de type "a" ni de type "d" dans le dictionnaire, les possibilités (i), (ii), (iv), (v), (vi), et (vii) seront rejetées, et donc on n'aura que deux segmentations possibles (iii) et (viii). Si on observe ces deux cas, c'est-à-dire si on trouve les mots "ab", "cd" et "abcd" dans le dictionnaire, le choix entre deux analyses (iii) et (viii) ne peut pas être effectué au niveau lexical ou morphologique, mais il faudra le traiter au niveau syntaxique.

Par exemple, la séquence suivante présente une ambiguïté qu'on ne peut pas lever au niveau morphologique :

정치가 *jengchiga*

car, ce peut être une des deux séquences suivantes :

a / b (**a** = 정치 (*jengchi*) [politique] / **b** = 가 (*ga*) [PPP.*nmtf*],
donc **ab** = 정치-가 [politique-Sujet])
ab (**ab** = 정치가 (*jengchiga*) [politicien])

Les phrases suivantes illustrent respectivement ces deux interprétations :

(1) 한국 정치가 변하고 있다.
hangug ***jengchi-ga*** *byenha-go iss-da*
Corée **politique**-PPN changer-PPP être en train de-PPP
(**La politique** coréene est en train de changer)

(2) 한국 정치가 협회가 결성되었다.
hangug ***jengchiga*** *hyebhoi-ga gyelsengdoi-ess-da*
Corée **politicien** association-PPN être organisé-PPP
(Une association des **politiciens** coréens a été organisée)

Il est évident que ces deux interprétations ne peuvent être obtenues que quand le dictionnaire fournit tous les mots acceptables. Or, les dictionnaires traditionnels sont loin d'être complets de ce point de vue. D'une part, la liste des postpositions qui

s'attachent à un nom (e.g. postpositions casuelles comme *nmtf*, *Acc*, *Dat* ou *Loc*) n'est pas complète, et surtout les combinaisons de postpositions ne sont pas étudiées systématiquement. D'autre part, des noms dérivés comme celui dans (2) (i.e. 정치가 (*jengchi-ga*) [politique-*Sfx*:personne = politicien]) ne sont pas recensés d'une manière exhaustive dans dictionnaires classiques. Dès lors, il serait difficile d'espérer un étiquetage correcte pour les séquences de type (1) ou (2).

Dans ce travail, nous allons construire des dictionnaires aussi complets que possible, dans le but d'analyser toutes les séquences complexes, y compris les formes fléchies de termes prédicatifs (i.e. verbes et adjectifs), formes directement observables dans des textes écrits. Dans le paragraphe suivant, nous allons présenter la structure de notre système d'analyse morphologique comportant des dictionnaires de formes fléchies.

1.1. Modèle conceptuel

Le système d'analyse morphologique que nous allons implémenter dans ce système est construit de la manière suivante <figure 1-1> :

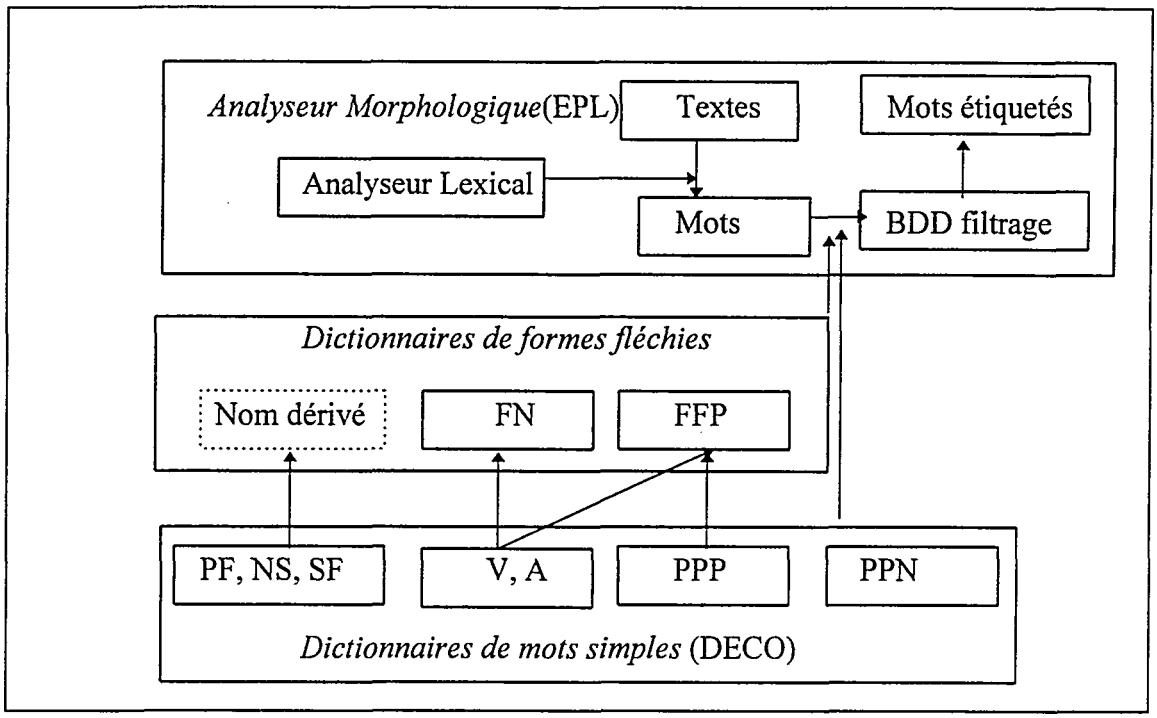


Figure 1-1. Modèle conceptuel du système de l'analyse morphologique.

Notre système est constitué de trois parties : le système lexical DECO, le dictionnaire des formes fléchies et l'analyseur morphologique. Le système DECO contient les sous-lexiques suivants qui ont été décrits et qui figurent dans [24], [25], [27], [30] :

Dictionnaire des	Noms simples	NS
Dictionnaire des	Adjectifs simples	A
Dictionnaire des	Verbes simples	V
Dictionnaire des	Préfixes et Suffixes	PF/SF
Dictionnaire des	Postpositions prédicatives et Postpositions nominales	PPP/PPN

Les deux derniers dictionnaires (PF/SF, PPP/PPN) permettent d'engendrer des séquences complexes en associant avec les trois premiers : il y a d'une part, des formes dérivées basées sur des noms simples (une association de NS avec PF/SF) et d'autre part, des formes fléchies basées sur des prédicatives (une association de A ou V avec PPP).

Basée sur le système lexical DECO, on va construire une grammaire et engendrer les dictionnaires suivants :

Grammaire équivalente au dictionnaire des Noms dérivés	ND
Dictionnaire des Formes Nominalisées des termes prédicatifs	FN
Dictionnaire des Formes fléchies prédicatives	FFP

Les deux derniers dictionnaires comporteront toutes les formes conjuguées des verbes et des adjectifs, y compris toutes les nominalisations basées sur les suffixes dérivationnels -음 (*-eum*) et -기 (*-ki*). La grammaire locale que nous allons construire sera équivalente au dictionnaire des noms dérivés. Autrement dit, cette grammaire permettra de reconnaî tre toutes les formes dérivées, et donc de distinguer ces formes, par exemple, les séquences composées de type Nom-PPN. Considérons les exemples suivants figure 1-2 :

Cas 1	Cas 2	Cas 3	Cas 4
모양가	예술가	친구가	정치가
moyangga *moyang-ga*	*yesulga* *yesul-ga*	*chinguga* *chingu-ga*	*jengchiga* *jengchi-ga*
*	art-PPN	ami-PPN	<ambiguë>
	artiste	ami-***Sujet***	(a) politicien (b) politique-***Sujet***

Figure 1-2. Cas divers de type "*X-ga*" .

Le Cas 1 représente des séquences inacceptables, bien que la partie qui précède le morphème *ga* soit un nom simple (= *apparence*). C'est parce que les substantifs dont la dernière lettre est une consonne (comme dans le *Cas 1*) ne peuvent pas être suivis d'une postposition *ga* que le mot est interdit. Mais, le morphème *ga* n'est pas non plus un suffixe compatible.

Par contre, dans le Cas 2, *ga* est un suffixe désignant une personne, et ainsi la séquence 예술-가 (*yesul-ga*) est un nom dérivé à partir d'un nom simple 예술 (*yesul*) [art], et il figure dans notre dictionnaire. Notre grammaire permettra donc de reconnaître ce mot.

De même, le morphème *ga* dans le Cas 3 ne sera reconnu que comme postposition de nominatif, puisque le nom 친구 (*chingu*) [ami] ne peut pas être suivi du suffixe *ga*, et qu'on ne trouvera pas cette séquence dans la liste des formes dérivées. Notons que nous ne construirons pas de dictionnaires des formes dérivées, mais nous comparerons la liste des suffixes et celle des postpositions, et nous n'établirons les listes de noms dérivés que quand il existe des homographes entre ces deux catégories.

Par exemple, le morphème *ga* est un des homographes typiques, nous établirons la liste des noms dérivés à base de suffixe *ga* : le nombre de ces noms est de 597 dans la version actuelle de notre dictionnaire. Cette liste est intégrée dans le module BDD

<figure 1-1> et sera utilisée pour la désambiguation. Voici un extrait des noms de type *N-ga* <figure 1-3> :

공예가 (*gongyega*)	계몽가 (*gyemongga*)
극작가 (*geugjagga*)	달변가 (*dalbyenga*)
무용가 (*muyongga*)	문예가 (*munyega*)
세력가 (*selyegga*)	수필가 (*supilga*)
연설가 (*yenselga*)	이론가 (*ilonga*)

Figure 1-3. Extrait des noms de type *N-ga*.

Par conséquent, toutes les séquences de type *X-ga* qui ne se trouvent pas dans le module BDD seront considérées automatiquement comme des séquences constituées d'un nom et d'une postposition de nominatif *ga*. La situation est en fait plus complexe, car il y a des noms dérivés et composés qui s'attachent à *ga*. Nous en parlerons plus tard.

Le Cas 4 représente des cas ambigus dûs à des homographes (i.e. le suffixe *-ga* et la postposition *-ga*), étant donné que le nom 정치 (*jengchi*) [politique] est compatible avec le suffixe *-ga*. Cette sorte d'ambiguïté ne peut pas être levée au niveau morphologique, l'information syntaxique doit être prise en compte.

Avec les dictionnaires que nous avons mentionnés au-dessus, nous allons implémenter un analyseur morphologique du coréen (Etiquetage des Parties du discours basé sur le Lexique : EPL). Quand un texte est donné, la première phase sera de le transformer en des chaînes de séquences où les séparateurs sont reconnus. Les structures de toutes les séquences ainsi obtenues seront analysées sur la base des dictionnaires de formes fléchies : tous les découpages possibles seront ainsi proposés. Le Base de Données de Désambiguïsation (BDD) filtrage sera appliquée pour lever des ambiguïtés d'étiquetage des mots.

1.2. Organisation du travail

Ce travail est constitué de trois chapitres. Dans le chapitre 2, nous allons présenter les dictionnaires que nous avons utilisés et engendrés. Comme nous l'avons mentionné ci-dessus, nous nous sommes basé sur des dictionnaires de mots simples et des lexiques d'affixes et de postpositions. Les dictionnaires des formes fléchies sont alors engendrés à partir de ces dictionnaires et de la grammaire que nous avons construite dans ce travail.

Ensuite, nous classons toutes les séquences qu'on observe dans les textes écrits du coréen. A part les *adverbes* que nous ne traiterons pas dans ce travail, on peut diviser les séquences en deux types : séquences *nominales* et séquences prédicatives (*verbales* et *adjectivales*). Plus précisément, on les classera en trois catégories, comme suit :

Type 1 **(T1)**	Toutes les séquences comportant au moins un nom simple. On peut les formaliser comme PF* $NS^{+}SF^{*}$ $PPN^{?}$. Ces séquences sont éventuellement constituées d'un ou plusieurs préfixe/s/ et/ou d'un ou plusieurs suffixes. Les PPN sont des postpositions qui s'attachent à des substantifs. Voici un exemple : 국세청장이 *gugseichengjangi* [PF-NS-SF-SF-PPN].
Type 2 **(T2)**	Toutes les séquences comportant au moins une forme nominalisée. On peut les formaliser comme $(PF^{?}$ (FN | NS)$^{+}$ $SF^{?}$ | Formes Spéciales) $PPN^{?}$. Ces séquences ressemblent à celles du Type 1, sauf qu'elles comportent une forme nominalisée à la place d'un nom simple. Par exemple : 늦잠자기가 *neujjamjagineun* [PF-FN-FN-PPN].
Type 3 **(T3)**	Toutes les séquences comportant un verbe ou un adjectif. On peut les formaliser comme ($PF^{?}$ (FN | NS)$^{+}$ $SF^{?}$)* (V_{root} | A_{root}) PPP. Ces séquences correspondent à des formes fléchies de verbe en français, mais en coréen, ce type de séquences présente une complexité très importante : d'une part, les adjectifs prennent autant de formes fléchies que les verbes ; et d'autre part, le nombre des combinaisons de suffixes de conjugaison est extrêmement élevé. Prenons un exemple : 노래하다가 *jabilhayeneundei* [NS-V-PPP].

Table 1-1. Les types de mot coréen.

Dans le chapitre 3, nous discuterons de la structure de dictionnaire et de la méthode qu'on utilisera pour retrouver des mots dans les dictionnaires. Nous allons comparer plusieurs méthodes déjà proposées ainsi que la méthode que nous utilisons dans ce travail.

Le chapitre 4 est consacré à la construction de notre système d'analyse morphologique du coréen. Ce système permettra de reconnaî tre toutes les analyses possibles pour une séquence donnée : ce système garantira tous les découpages corrects. Nous allons le discuter en détail avec des exemples concrets.

2. Structure du dictionnaire établi pour ce travail

Dans les dictionnaires à usage humain, une entrée lexicale contient plusieurs types d'information [17], [21]. Les dictionnaires utilisables par la machine pour le TNL doivent comporter des informations plus explicites que les dictionnaires faits pour l'usage humain. Une entrée lexicale peut contenir au moins quatre types d'information: la forme phonologique, la catégorie syntaxique, l'indication des irrégularités morphologiques, et le sens [8], [40].

Cependant les dictionnaires coréens existants (sous forme imprimée ou électronique) ont été conçus pour l'usage humain, ils ne sont donc guère satisfaisant pour notre projet. La classification des entrées lexicales en parties du discours n'est pas faite de façon explicite et cohérente. Par exemple, il n'y a pas de critères formels qui permettent de distinguer les verbes des adjectifs. L'information sur la relation dérivationnelle entre les entrées n'a pas été étudiée de manière systématique et exhaustive [24].

Notre système DECO prend en considération de telles informations [24]. L'Analyseur Morphologique, EPL, basé sur DECO, analysera les phrases entrées. L'unité minimale de l'analyse de EPL est le mot et celui-ci est classifié en trois types de structures **T1**, **T2** et **T3**.

Toutes ces structures sont définies dans la table 1-1 et le mot est analysé selon l'une d'elles par l'analyseur morphologique. Chaque classe contient au moins un dictionnaire de formes fléchies la grammaire équivalente au dictionnaire du ND appartient à la classe T1, le dictionnaire du FN à la classe T2, et le dictionnaire du FFP à la classe T3. Nous discuterons plus en détail ces classes dans les sections suivantes.

2.1. Structure de type 1

Le nom se trouve dans T1. Le mot classé T1 est défini ainsi:

$$T1 \quad = \quad PF^* \, NS^+ SF^* \, PPN^?$$

La structure T1 contient au moins un NS. Les autres composantes sont facultatives. Lors de l'analyse d'un mot, nous rencontrons nécessairement des PF et SF qui ont des structures complexes dans T1.

En réalité, l'occurrence répétitive des catégories PF et SF, représentée comme PF^* et SF^*, est très rare, mais NS est très productif. Cela nous donne l'idée de changer T1 en T1' qui n'a pas de structures répétitives dans les cas de PF' et SF'. T1' sera donc défini de la façon suivante :

$$T1' \quad = \quad PF'^? \, NS^+ SF'^? \, PPN^? = ND \; PPN^? \; (ND = PF'^? \, NS^+ SF'^?)$$

ND correspond aux noms dérivés de PF, NS et SF. PF' et SF' dans T1' doivent inclure les cas rares de PF et SF dans T1. Par exemple, "상" (haut) et "품" (qualité) sont des PF de T1, mais "상품" (haute qualité) n'est pas un PF de T1, parce que PF est dans le dictionnaire simple. Dans ce cas, nous ajoutons le mot "상품" (haute qualité) dans PF' de T1' avec la marque de répétition.

Pour construire les dictionnaires PF' et SF' de T1', nous devons examiner linguistiquement toutes les possibilités d'entrées répétitives qui peuvent se produire dans PF et SF de T1. Les structures répétitives peuvent être formulées de la façon suivante :

R1 $(\exists x)\,(\exists y)\,((X \subset Fact(T1)) \wedge x \in X \wedge y \in X \quad \text{et } (x \bullet y \in X))$

($\bullet$ est l'opérateur de la concaténation)

alors $Fact(T1) = PF^* \mid NS^+ \mid SF^* \mid PPN^? \mid PF^* NS^+ \mid NS^+ SF^* \mid SF^* PPN^? \mid PF^* NS^+ SF^* \mid NS^+ SF^* PPN^? \mid T1$

R2 $(\exists x)\,(\exists y)\,((X^* \subset Fact(T1)) \wedge x \in X \wedge y \in X \text{ et } (x \bullet y \notin X))$

Elles peuvent produire des ambiguïtés interne et externe.

R1 est la structure concernant l'ambiguïté interne dans le dictionnaire. Par exemple, chaque "연" (total) (x) et "두" (tête) (y) sont dans PF, et la concaténation "연두" (vert pâle) (x•y) est aussi dans PF. Dans ce cas, il génère les marques ambiguës. Mais toutes les entrées des dictionnaires contenant PF, SF et NS sont des *unités simples*. Cela signifie qu'aucune entrée des dictionnaires ne peut être séparée en deux entrées. "연두" (vert pâle) n'a donc ni le sens de "연" (total), ni celui de "두" (tête). "연두" (vert pâle) est l'unité autonome la plus petite. Notre système DECO ne génère pas le problème **R1,** car il est basé sur des *unités simples*.

R2 est la structure concernant l'ambiguï té externe dans le dictionnaire. PF et SF de T1 ont rarement des cas de répétition. Nous pouvons alors énumérer linguistiquement toutes les entrées répétitives dans PF et SF. Les nouvelles données obtenues seront gardées dans PF et SF de T1'. Ainsi on supprime les problèmes de répétition dans PF et SF de T1'.

D'autres types d'ambiguï tés existent encore dans T1'. Il ne s'agit plus d'ambiguï tés de répétition, mais d'ambiguï tés d'interférence entre les dictionnaires. Pour NS, il existe toujours des ambiguï tés de répétition, mais nous ne sommes pas intéressé au problème des répétitions dans ce travail, à cause de leur caractère productif.

Les ambiguïtés d'interférence sont définies ainsi [19]:

P1 $(x \in X \text{ et } x \in Y)$ et $(X \neq Y)$ $X \subset \text{Fact}(T1')$ et $Y \subset \text{Fact}(T1')$

P1 est le problème de l'homonymie entre les dictionnaires x et y. Par exemple, "가" est dans SF et aussi dans PPN.

P2 $((XY \subset \text{Fact}(T1') \wedge x \bullet z \in X \wedge y \in Y \text{ et } (x \in X \wedge z \bullet y \in Y))$

P2 est le problème du croisement mutuel. Par exemple, le mot "복사실" (Salle de photocopie) peut être marqué au niveau de la morphologie comme les trois candidats suivants :
1) "복사"(la pêche) (PF) (le mot sino-coréen) et "실" (la salle) (NS),
2) "복사" (la photocopie) (NS) et "실" (la salle) (SF),
3) "복"(duplicatif) (PF) et "사실" (le fait) (NS).

Les cas 1) et 3) produisent l'ambiguï té **P2**. Pour ces cas, nous devons fournir l'information linguistiquement correcte. Le cas 2) sera sélectionné et cette information sera gardée dans BDD.

P3 $((XY \subset \text{Fact}(T1') \text{ et } XZ \subset \text{Fact}(T1') \wedge x \bullet z \in X \wedge y \in Y \text{ et } (x \in X \wedge z \bullet y \in Z) \text{ et } (X \neq Y \neq Z))$

P3 est le problème de l'interférence. Ce problème peut se produire seulement entre NS, SF, et PPN.

EPL peut détecter les ambiguï tés dans l'analyse du mot. Ces ambiguï tés sont dues à la répétition, l'interférence, la syntaxe et la sémantique, etc. Dans notre étude, nous nous limiterons à la répétition et l'interférence. La sélection de l'information correcte par rapport aux problèmes d'interférence **P1**, **P2** et **P3** se fera de manière semi-automatique.

Toutes les séquences acceptables dans T1' appartiennent à la sphère 1 de la figure 2-1. La sphère 1 est un ensemble. Les données de cette sphère sont produites automatiquement.

Toutes les données comprises dans la sphère 2 sont ambiguës. Les données proviennent automatiquement de la répétition et de l'interférence que nous venons de mentionner.

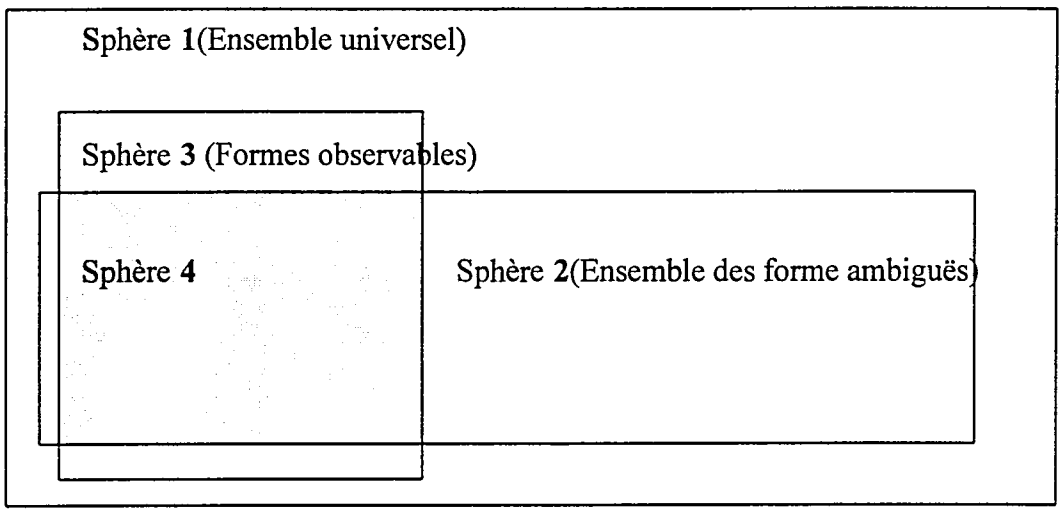


Figure 2-1. Procédure de désambiguïsation dans notre système.

Les données de la sphère 2 ne sont pas toutes observables dans la langue réelle. Par exemple, "우두통" est analysé en deux formes comme "우" (la droite ou la vache) (PF) avec "두통" (mal à la tête) (NS) et "우두" (le chef) (PF) avec "통" (le tonneau) (NS). Bien que la forme présente théoriquement une ambiguï té **P2** entre PF et NS, cette ambiguï té n'est pas effective. Nous devons donc choisir linguistiquement les données réelles dans la sphère 2 et elles seront gardées dans le module BDD. Par conséquent, nous re jetterons l'exemple donné ci-dessus.

Les données sélectionnées linguistiquement occupent la sphère 4 dans la figure 2-1. La procédure de sélection dans les sphères 2 à 4 est exécutée par la sphère 3. La sphère 3 représente l'être humain. Autrement dit, elle implique 'linguistiquement'. L'expression 'linguistiquement' peut signifier 'par des linguistes' ou ' par des êtres humains à l'aide des dictionnaires publiés dans le commerce'[37].

2.1.1. Nom Simple

Le nom est la plus importante des parties du discours. Il peut être forme d'une combinaison de parties du discours telles que PF-SN, SF-SN, ou PPN-SN. Il est difficile de reconnaî tre chaque catégorie sans ambiguï té dans le mot qui comporte des éléments. Les exemples suivants illustrent le type NS <Figure 2-2> [24].

NS	CODE	Information
가	NS	
가간	NS	
가감	NS	/PRED1/PRED3
가객	NS	/HUM
가게	NS	/*PREDHA
가격	NS	
가결	NS	/PRED1/PRED3

Figure 2-2. Exemples de entrées de NS.

Notre dictionnaire a été amélioré à plusieurs reprises, et nous avons obtenu finalement 13 331 entrées de NS.

2.1.1.1. Structure de répétition (NS^+)

Les structures de noms coréens peuvent être décrites de la manière suivante (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; NS désigne un nom simple qui est un élément autonome) (Pour le détail, consulter *Nam 94* [24]). Considérons :

(1) a. NS
b. PF-NS
c. PF-NS-SF
d. PF-PF-NS
e. NS-SF-SF

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

Seule la forme (1a) sera traitée comme nom simple; les formes (1b) ~ (1e) sont dérivées; (2) sont des noms composés. La distinction entre (1a) et (1b) ~ (1e) est faite en fonction de l'existence de préfixes et/ou de suffixes en dehors d'un élément autonome (i.e. NS).

La distinction entre (1) et (2) est effectuée selon le nombre de noms(NS), 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 différents niveaux morphologiques, et d'autre part considèrent des unités de type PF-X ou X-SF (X étant un élément non-autonome) comme dérivées, alors que pour nous ce sont des noms simples que nous ne décrivons pas autrement que NS (ainsi, nous ne intéressons pas à l'analyse en morphèmes des mots qui ne possèdent aucune partie autonome à l'intérieur). Par conséquent, même avec le même préfixe, si l'élément qui suit est une unité autonome, cette combinaison sera reconnue comme dérivée (ou affixée), sinon comme simple, donc nom simple.

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

(3) a. 가을 (automne)
b. 재-결합 (ré-unification = réunification)
c. 포도-주 (alcool de raison = vin)
d. 동-성-애 (même-sexe-amour = amour homosexuel)
e. 가-출-옥 (provisoire-sortir-prison = sortie provisoire de prison)
f. 경리-부-장 (comptabilité-département-responsable
= responsable du département de comptabilité)

(4) a. 구슬-땀 (perle-sueur = goutte de sueur de forme perle)
b. 반-독재-투쟁 (anti-dictature-révolte = révolte anti-dictature)
c. 간호-보조-원 (infirmier-assistance-personne = assistant d'infirmier)
d. 가운뎃-손-가락 (centre-main-doigt = majeur)

On pourrait supposer a priori d'autres types de combinaisons possibles (e.g. PF-PF-NS-SF), mais le nombre des formes autre que (1) et (2) est extrêmement réduit. Par contre, des séquences chaînées de noms simples de type (5) s'observent fréquemment:

(5) 국어 맞춤 법 통일 개정 안
coréen orthographe règle unification réforme proposition
(La proposition de réforme pour l'unification des règles d'orthographe du coréen)

La séquence (5) est composée de six noms simples (NS NS NS NS NS NS). Rappelons que la d'elimitation des mots typographique pas de blancs ne peut par servir de critère opératoire, car on observe également:

(6) a. 국어 맞춤법 통일 개정안
coréen orthographe-règle unification réforme-proposition

b. 국어맞춤법 통일개정안
coréen-orthographe-règle unification-réforme-proposition

Pourtant, bien que cette d'elimitation typographique ne soit pas normalis釬, il existe des contraintes:

(7) a. *국어맞춤 법 통일개정 안
*coréen-orthographe règle unification-réforme proposition

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

Ce type de nom composé est la structure que nous avons définie comme **R2** de NS. Il s'agit d'une structure où plusieurs noms apparaissent successivement. Ce type (symbole + dans NS^{+}) peut engendrer plusieurs problèmes. Bien que peu productive cette structure ne peut pas être ignorée. La répétition de noms simples peut avoir lieu deux ou trois fois selon la règle générale [15]. Nous n'énumérerons pas ici toutes les entrées de ce type de Nom Composé.

NS sera utilisé obligatoirement dans T1', mais PF' et SF' sont différents. C'est pourquoi PF' et SF' sont définis comme $PF'^{?}$ et $SF'^{?}$, alors que NS est défini comme NS^{+} dans la structure T1'. PF' se trouve toujours en tête de NS, alors que SF' est placé à la fin de NS. S'il y a un conflit entre les séquences répétitives de NS et PF' ou SF', on pourra résoudre le problème par le principe que la structure répétitive est autorisée uniquement pour NS.

2.1.2. Préfixe

Le PF, en tant qu'affixe, s'attache en tête du nom. La forme complexe basée sur le nom restera toujours nom. Les entrées du dictionnaire de PF sont des unités simples. Autrement dit, aucune entrée du dictionnaire de PF ne peut être divisée en plus petites unités sans changer le sens de l'entrée. Actuellement, il y a 505 entrées distinctes dans le dictionnaire de PF. La figure 2-3 montre les entrées simples de PF. Par exemple, l'entrée '가등기' (l'enregistrement provisoire) dans la figure 2-3 est composée du PF '가' (provisoire) et du NS '등기' (l'enregistrement).

PF	CODE	Information	Exemples
가	PF.		가극, 가곡
가	PF.		가등
가	PF.		가등기, 가석방
가	PF.		가법, 가례
가	PF.		가홍
가	PF.	PREDH	가해, 가열
가랑	PF.		가랑비, 가랑잎
가시랑	PF.		가시랑비

Figure 2-3. Exemples des entrées de PF.

2.1.2.1. Structure répétitive PF^n (si n >= 2))

Le PF est utilisé de manière récursive dans la structure T1. Cela rend difficile l'analyse des mots contenant des PF et peut produire des ambiguïtés. Nous allons transformer la structure répétitive du PF en celle non-répétitive qui a été définie dans T1'. Cela veut dire que nous allons chercher toutes les entrées répétitives dans le dictionnaire des PF de T1. Les entrées trouvées seront insérées dans le dictionnaire de PF. La procédure consiste en la transformation de PF^* dans T1 en $PF'^?$ dans T1'. Cette procédure permettra de supprimer la structure **R2** dans PF.

Nous avons extrait linguistiquement toutes les entrées combinées de PF dans T1. Les entrées sélectionnées se trouvent dans la sphère 4 de la figure 2-1.

Supposons que le nombre de reprises est deux(si, n=2). Le nombre candidats de **R2** correspond au *Numéro_de* (PF) multiplié par le *Numéro_de* (PF). C'est-à-dire que 505*505 (255,025) est le nombre d'entrées de l'expression PF^2. Parmi les 255,025 entrées, nous avons trouvé de façon manuelle 26 nouvelles entrées qui sont représentées dans la liste 2-1 ci-dessous, et elles sont ajoutées au dictionnaire PF'. Les autre cas PF^n (si, n >= 2)) n'existent pas dans le monde réel des mots.

고품 극단 극대 극소 극한 급회 남고동 냉난 다갈 맨헛 맹혹 상품 생까막 순까막 순헛 역순 온갖 온난 저품 적녹 중품 진고동 청녹 최단 하품 한냉(26)

Liste 2-1. Nouvelles entrées du PF'.

2.1.2.2. Ambiguïté **P1**

Il y a 262 entrées ambiguës de **P1** produites entre PF et NS comme le montre l'Annexe A. Les PF sont toujours employés immédiatement avant un NS, les NS peuvent être employés sans PF. Cela permet de déterminer si une forme est PF ou NS en cas d'ambiguï té. Une autre solution serait d'énumérer toutes les formes possibles entre PF et NS. Cela demandera beaucoup de temps.

2.1.2.3. Ambiguïté **P2**

Afin d'établir la liste des données ambiguës de **P2** entre PF et NS autrement dit, afin d'extraire toutes les données se trouvant dans la sphère 2 sous la structure d'ambiguï té **P2** de T1', nous avons construit le programme de la figure 2-5 qui est la forme algorithmique de la figure 2-4. Les nombres maximum de syllabes du PF et du NS sont réciproquement 3 et 6.

$$\text{mots} = \prod Xm, \text{ donc } \text{PF} = \prod_{k=1}^{j} X_k, \text{NS} = \prod_{k=j+1}^{n} Xk \text{ SF} = \prod_{k=1}^{l} X_k, \text{PPN} = \prod_{k=l+1}^{m} Xk$$

Figure 2-4. Représentation des mots incluant les entrées de PF et de NS.

for(j = 0; j < Maximum_Length_Of_PF; j++)
    for(i = j+1; i <= Maximum_Length_Of_PF;i++)

$$\textit{if}(\quad (\text{Look\_Up}(\text{PF}, \prod_{k=1}^{j} X_k,) \;\&\&\; \text{Look\_Up}(\text{NS}, \prod_{k=j+1}^{n} Xk) \;\&\&$$

$$(\text{Look\_Up}(\text{PF}, \prod_{k=1}^{i} X_k) \;\&\&\; \text{Look\_Up}(\text{NS}, \prod_{k=i+1}^{n} X_k \;))$$

$$\textit{printf}(\text{“\%s \%s »}, \prod_{k=1}^{j} X_k \prod_{k=j+1}^{n} Xk, \prod_{k=1}^{i} X_k \prod_{k=i+1}^{n} X_k); \text{ /* Ambiguïté } \mathbf{P2} \text{ */}$$

Figure 2-5. Programme d'extraction des données **P2**.

Il n'y a pas d'ambiguï té **P2** entre j =3 et j =2, et entre j = 3 et j = 1. Dans d'autres cas, le mot peut être marqué selon l'un des trois cas suivants:

1) PF (longueur 0), NS (la longueur est n) ou
2) PF (longueur 1), NS (longueur n-1), ou
3) PF (longueur 2), NS (longueur n-2).

Les formes qui présentent une ambiguï té **P2** sont 2,739. Elles sont extraites automatiquement à partir du programme dans la figure 2-5. Ces formes se trouvent dans la sphère 2 de la figure 2-1 sous l'ambiguï té **P2**. La liste figure dans l'Annexe B-1. Nous avons extrait manuellement de cette liste les formes susceptibles d'être dans la

sphère 4. l'Annexe B-2 montre la liste de ces formes. Elle sera gardée dans BDD de la figure 1-1 avec les marques correctes. La marque 'S' dans l'Annexe B-2 signifie que le nom se trouve dans les dictionnaires publiés [37].

2.1.3. Suffixe

Le SF s'attache à la fin du nom. La forme combinée avec le nom est classée comme nom. Le problème de la répétition se pose comme pour le PF. Actuellement, 468 entrées distinctes sont dans le dictionnaire SF. La figure 2-6 montre les exemples de SF dans les entrées. Par exemple, 대학가 (le quartier universitaire) est composé de 대학 (l'université) et de 가 (le quartier) et il est marqué comme NS-SF.

SF	CODE	Information	Exemples
가	SF.		병가
가	SF.		영양가, 주가
가	SF.		왕가
가	SF.		은행가, 응원가
가	SF.	HUM	음예가, 무용가
가+	SF.		주택가, 대학가
가+	SF.		하늘가, 화롯가
가+	SF.	HUM	김가, 이가

Figure 2-6. Les exemples des entrées du SF

2.1.3.1. Ambiguïtés de répétition dans SF ^{n} (n >= 2)

P1 = {댁 (chez), 뻘 (camarade), 석 (place) , 실(bureau),
역(membre) , 측(parti), 파(parti)}
P2 = {네 (chez) , 하 (cher)}
P3 = {전 (à), 후 (après)}
S1 = {님 (cher)}
S2 = {들 (-s)}
S3 = {간 (entre), 외 (sauf)}
T1 = {선생 (professeur), 여사 (madame) }
T2 = {양 (mademoiselle, 씨 (monsieur) , 군 (garçon)}
E = Vide

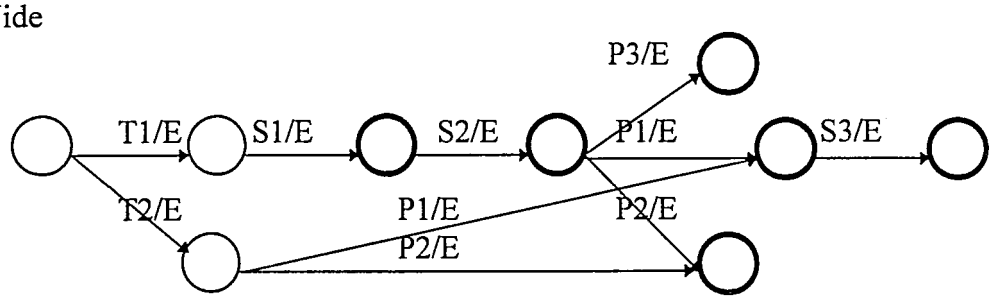


Figure 2-7. Le diagramme de transition pour le cas vocatif dans la structure SF^{+}.

Nous avons établi deux règles de répétition dans la structure SF^{n}. Certaines entrées dans le dictionnaire des SF sont très productives. Elles sont attachées de façon répétitive à la fin du NS. Dans ce cas, il n'est pas simple de représenter la structure SF avec les diagrammes de transition des figures 2-7 et 2-8. La figure 2-7 illustre l'exemple

des formes du *vocatif*. Les éléments terminaux du diagramme sont des éléments vocatifs.

Certains SF représentent *la position sociale*. Pour exprimer les positions sociales variées des gens, plusieurs types des suffixes sont susceptibles d'être attachés à la fin du nom. Ils correspondraient à monsieur, docteur, président, membre, etc. en français.

La figure 2-8 est un diagramme de transition des suffixes de *position* pour le nom. Dans ce diagramme, S2(pluriel) est utilisé librement. Il se trouve en tête ou en fin des entrées P1 ou P2.

A l'exception de ces deux cas, SF n'a pas de structure répétitive dans le mot. Nous pourrions établir le processus pour les deux cas avec des règles, car ils comprennent des formes très régulières. Les règles figurent dans le Module BDD filtrage de la figure 1-1. Dans le Lexique-Grammaire, on appelle cette opération régulière une grammaire locale[13] [18]. Sauf pour les deux cas que nous venons de mentionner, nous pouvons redéfinir facilement la structure SF^* dans T1 comme $SF^?$ dans T1'.

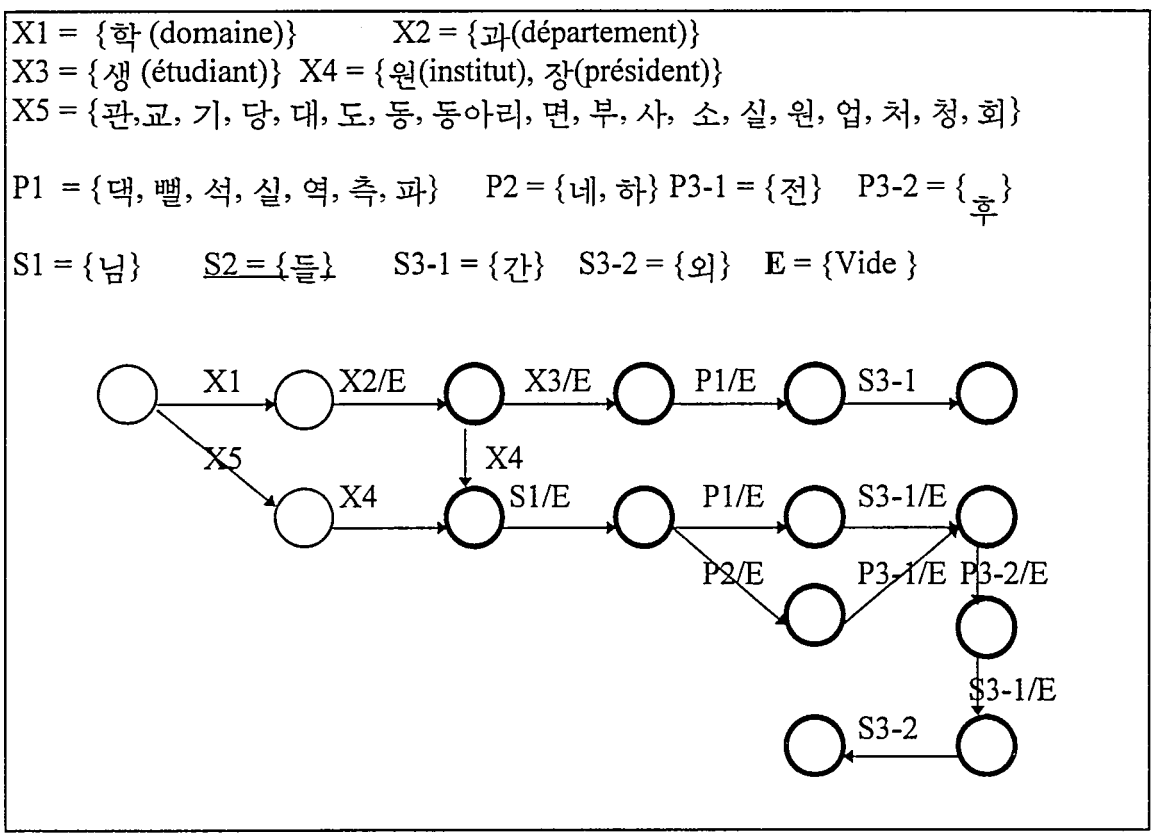


Figure 2-8. Diagramme de transition pour *position sociale*.

2.1.3.2. Ambiguïté **P1**

Il y a 253 entrées P1 entre NS et SF comme le montre la liste 2-2. Le conflit entre NS et SF est résolu de la même manière que pour les ambiguï tés **P1** entre PF et NS. Les SF sont toujours employés immédiatement après un NS, les NS peuvent être employés sans SF. Cela permet de déterminer si une forme est SF ou NS en cas d'ambiguïté. Nous pouvons donc distinguer le SF de NS sans difficulté.

가 가락 가리 각 간 감 갑 강 개 객 거리 거미 건 건너 검 격 결 경 계 곡 골 공 과 곽 관 광 구리 구원 국 군 굽 궐 궤 귀 금 급 기 길 깨 꺼풀 꼬리 꼴 끌 난 남 납 낱 내 내기 널 노 농 눈 님 다리 단 담 답 당 대 대가리 대기 댁 도 독 동 동아리 동이 동자 들 딱지 띠 마 마루 막 만 말 망 망나니 매 맥 머리 면 명 모 목 무 무리 묵 문 물 미 바가지 바구니 바라지 바람 바지 박 반 발 밥 방 방면 배 배기 백 번 벌 범 벼락 벽 변 별 보라 복 본 봉 부 부리 부지 북숭이 분 불 비 빵 사 산 살 상 새 생원 서리 서방 선 선생 설 성 세 소 속 손 솔 수 순 술 시 식 신 실 씨 아비 악 안 암 애 액 약 양 업 역 연 염 영 오리 요 욕 용 우 운 원 위 윤 의 이 이전 인 임 입 자 자락 자위 잔 장 재 쟁 저 적 전 절 점 정 젖 조종 좌 주 주부 주사 주의 죽지 중 지개 지기 지방 진 질 집 짝 차 참 창 채 책 처 천 첩 청 체 초 총 추 축 춤 치기 침 칼 탄 태기 턱 토 톱 통 통수 파 판 패기 펄 포 품 풍 피 학 한 함 합 해 향 호 혼 화 회 효 훈 (253)

Liste 2-2. Les entrées ambiguës **P1** entre NS et SF

2.1.3.3. Ambiguïté **P2** avec NS.

Dans l'expression de la figure 2-4, une ambiguï té **P2** entre NS et SF peut se produire. La procédure d'extraction de **P1** entre NS et SF est la même que pour le cas de **P2** entre PF et NS. Les longueurs maximum du NS et du SF sont réciproquement 6 et 3 syllabes. Les variables *l*, et, *m* sont utilisées dans l'expression à la place de j, et *n* pour le cas de PF.

Nous avons obtenu automatiquement 2,272 données **P2** qui appartiennent à la sphère 2 sous la structure T1'. La liste de ces données est dans l'Annexe C-1. A partir de cette liste, nous avons extrait les mots réels de la sphère 4. Les marques correctes des formes sont enregistrées dans BDD. La liste est dans l'Annexe C-2.

2.1.3.4. Ambiguïté **P1** avec PPN

Il existe 13 SF qui peuvent produire l'ambiguï té **P1** avec PPN comme dans la liste 2-3 Si le mot contient un des 13 SF, EPL l'analysera comme forme ambiguë. Nous avons appliqué la méthode précédente à cette forme. D'abord, nous avons obtenu automatiquement les formes combinées de NS avec les 13 SF. La liste de ces formes combinées se trouvent dans la sphère 2 et elles peuvent être marquées comme NS-SF ou NS-PPN. Le nombre minimum de formes atteint 173,303 (= 13,331 * 13) (Nombre de (NS) * Nombre de (Ambiguïté **P1** entre SF et PPN)).

Nous avons examiné cette liste et extrait manuellement les formes correctes, qui sont dans l'Annexe D. Les données seront sera enregistrées dans BDD. L'extrait est présenté dans la table 2-1.

가과 관께도로론만서선이 의이(13)

Liste 2-3. L'ambiguïté **P1** entre SF et PPN.

NS-SF forme	Nombre	NS-SF forme	Nombre	NS-SF forme	Nombre
NS + 가(SF)	597	NS + 과(SF)	352	NS + 관(SF)	366
NS + 께(SF)	13	NS + 도(SF)	456	NS + 로(SF)	94
NS + 론(SF)	885	NS + 만(SF)	8	NS + 서(SF)	410
NS + 선(SF)	442	NS + 아(SF)	32		
NS + 의(SF)	16	NS + 이(SF)	18		

Table 2-1. L'information BDD pour les formes NS-SF.

Par exemple, les 597 formes pour "가 "" sont marquées comme NS-SF dans la table 2-1. Cela signifie que si on trouve un mot qui peut être marqué comme NS-SF et NS-PPN (i.e. il est ambigu), et s'il n'est pas dans BDD, il sera marqué comme NS-PPN. Le nombre de formes dans la table 2-1 est celui des formes de NS qui peuvent être combinées avec les formes SF.

2.1.4. Postposition du Nom

En français, l'unité minimale pour l'analyse lexicale est, en général, la phrase contenant les blancs. Le dictionnaire doit donc contenir de l'information au delà du niveau morphologique.

Le dictionnaire de PPN est la base de données pour les noms, il fournit les indications sur les variantes morphologiques. Il existe des marques de fonction grammaticale comme les cas : *nominatif, accusatif, datif, locatif,* etc. qui se combinent avec les noms sans blanc. Une séquence composée de deux éléments comme ' à Paris' correspond à un élément 'Paris-Locatif' en coréen. PPN se situe après le nom. La figure 2-9 donne des exemples de la liste du dictionnaire PPN.

PPN	CODE	Information	
가	PPN	\PN1	.nmtf .postp
가라도	PPN	\PN1	.nmtf .postp
같이	PPN	\PN1 \PN2	.postp
까지	PPN	\PN1 \PN2	.nmtf .acc .postp
까지나	PPN	\PN1 \PN2	.nmtf .acc .postp

Figure 2-9. Exemples de la liste PPN.

2.1.4.1. Ambiguïté **P3**.

L'ambiguï té **P3** comporte des formes très complexes. Les deux formules ci-dessous peuvent générer l'ambiguï té **P3** entre NS, SF, et PPN. Nous avons déjà décrit **P3** en détail. On a les deux possibilités ci-dessous pour P3:

1) x•z∈NS et y∈SF et x∈NS et z•y∈PPN.
2) x•z∈NS et y∈PPN et x∈NS et z•y∈SF.

Les 189 formes ont été extraites automatiquement au moyen des deux formules (Liste 2-4). Les 189 candidats de la liste peuvent produire l'ambiguï té **P3**. De cette liste, nous avons obtenu manuellement les trois formes de la liste 2-5.

가동아 가동이 각주의 간주의 감동아 감동이 감주의 강박이 객주의 격동아 격동이 결박이 경쟁이 경주의 골동아 골동이 공동아 공동이 공박이 공주의 군주의 금주의 기동아 기동이 난동아 난동이 냉동아 냉동이 노동아 노동이 논박이 논쟁이 단주의 대동아 대동이 도박이 도주의 독주의 뒤주의 맥주의 면박이 명주의 목동아 목동이 무동아 무동이 물주의 미동아 미동이 밀주의 반동아 반동이 반박이 반주의 발동아 발동이 발주의 백주의 범주의 변동아 변동이 변주의 부동아 부동이 부주의 분쟁이 사동아 사동이 사주의 상동아 상동이 상주의 선동아 선동이 선박이 선주의 소동아 소동이 소박이 소주의 속박이 수동아 수동이 수박이 시동아 시동이 시주의 신동아 신동이 악동아 악동이 안주의 애주의 약주의 양주의 연동아 연동이 연주의 염주의 영주의 요동아 요동이 우동아 ㅇ동이 ㅇ박이 ㅇ주의 운동아 운동이 위주의 음주의 이동아 이동이 이주의 인주의 임박이 입주의자동아 자동이 자주의 재주의 저주의 전동아 전동이 전쟁이 전주의 점주의 정박이 정쟁이 정주의 종주의 주동아 주동이 진동아 진동이 진주의 질주의 찬동아 찬동이 책동아 책동이 청동아 청동이 청주의 초동아 초동이 축주의 탈주의 태동아 태동이 터주의 토주의 투쟁이 파동아 파동이 패주의 편주의 폐갱이 포박이 포주의 폭동아 폭동이 폭주의 피동아 피동이 학동아 학동이 함박이 합동아 합동이 합주의 해동아 해동이 향동아 향동이 호박이 호주의 혼동아 혼동이 활동아 활동이 활주의 황동아 황동이 회동아 회동이 동동주의 조물주의 구세주의 두레박이 (183)

Liste 2-4. Les candidats à l'ambiguïté **P3** entre NS, SF et PPN.

대동아	:	대(PF) 동아(NS), 대동아(Nom Propre)
반동아	:	반동(NS) 아(SF)
부주의	:	부(PF) 주의(NS)

Liste 2-5. La liste des ambiguïtés **P3** avec les marques.

2.2. Structure de type 2

La Forme Nominalisée est construite à partir du verbe ou de l'adjectif avec un suffixe de nominalisation(SFN). SFN est un cas spécial de PPP. FN est défini comme la classe T2 avec PPN. FN comporte le Verbe Nominalisé(VN), l'Adjectif Nominalisé (AN), et les Formes Spéciales.

Les FN fonctionnent comme le nom dans la phrase. Mais elles sont un peu différentes de la notion conventionnelle de nom. C'est-à-dire, FN peut représenter une séquence phrastique. Dans cette séquence, FN joue le rôle de l'infinitif en français ou celui du gérondif en anglais.

Par exemple, '뛰기' (courir) fonctionne comme l'infinitif en position de sujet en français ou comme le gérondif en anglais dans la phrase "뛰기가 어렵다"(Courir est difficile.) (Running is difficult). Dans une autre phrase, comme "뛰는것이 어렵다"(Courir est difficile.) (To run is difficult), "뛰는것 (Courir) (To run) est une autre forme de FN correspondant à l'infinitif en anglais.

Au niveau de la morphologie, il n'est pas important de distinguer la notion de FN de celle de NS. Dans l'interprétation syntaxique, la différence d'information est importante, parce que FN exige une structure de phrase différente de celle de NS.

T2 a une structure supplémentaire par rapport à T1'. FN fonctionne comme un nom dans la phrase. L'analyse de la structure T2 est la même que celle de T1' :

$$T2 = (PF^{?}(FN \mid NS)^{+} SF^{?} \mid \text{Formes Spéciales}) PPN^{?}$$

FN est composé d'un radical et d'un SN. Par exemple, le FN "뛰기" (courir) est analysé comme "뛰" (cour) (*radical*) et "기" (suffixe complémentaire). Dans la structure T2, PPN peut être soudé à la fin de FN. "뛰기가" (courir est) est analysé comme la combinaison de FN "뛰기" (courir) et de PPN "가" (*nominatif*).

La liste 2-6 donne des suffixes typiques pour FN[36]. "이", "개" et "애" qui s'utilisent dans des expressions figées. Les mots contenant ces suffixes sont considérés comme des noms et leur liste est déjà incluse dans NS. Dans le cas de '기', cette syllabe peut être soudée indépendante de la syllabe finale de radical à la fin du radical. Mais le choix des autres éléments de la liste comme '음' (d'un *suffixe complémentaire*) et 'ㅁ'(d'un *suffixe complémentaire*) dépend de la Syllabe Finale du Radical (SFR). Les formes fléchies des prédicats dépendent de SFR. Les formes fléchies des SFN dépendent aussi de SFR.

이 기 개 애 음 ㅁ(6)

Liste 2-6. SFN pour FN.

2.2.1. Verbes Nominalisés

Pour construire le lexique VN, on ajoute tout simplement SFN "기"(*suffixe complémentaire*) à la fin du radical des verbes. La forme de base de l'exemple 1) est "받다"(recevoir). Le radical "받" est la forme dont la dernière syllabe "다" est supprimée. La forme nominalisée du verbe se fait en ajoutant le SFN "기" après le radical "받". Dans ce cas, SFR et le radical sont identiques. Nous obtenons la forme nominalisée 1). La même procédure fournit l'exemple 2).

1) 받기 (la reçue)
2) 하기 (faire)

TYPE	CF dans SFR	Substitution	Nombre	Total	Exemple
NVT1		음	353	353	많다(많음)
NVT2-1		ㅁ	7,857	7,860	하다(함)
NVT2-2		non exister	3		건늬다
NVT3	ㄷ	ㄷ + 음	44	79	받다(받음)
NVT4		ㄹ + 음	35		깨닫다(깨달음)
NVT5-1	ㄹ	ㅁ	259	267	길다(김)
NVT5-2		ㄹ + 음			길다(길음)
NVT6		ㄺ	8		살다(삶)
NVT7	ㅂ	ㅂ + 음	68	84	꼬집다(꼬집음)
NVT8		음	14		눕다(누움)
NVT9		ㅁ	2+1(NVT8)		뵈옵다(뵈음)
NVT10-1	ㅅ	ㅁ	28	47	낫다(남)
NVT10-2		음			낫다(나음)
NVT11		ㅅ + 음	19		벗다(벗음)
NVT12-1	ㅎ	ㅁ	73	73	놓다(놈)
NVT12-2		ㅎ + 음			놓다(놓음)
				8,763	

Table 2-2. Les Classes des VN.

Il existe d'autres formes de SFN, comme "음" et "ㅁ". En principe, la première forme se combine avec SFR qui se termine par une consonne, alors que la dernière se combine avec SFR à voyelle.

Examinons les exemples 3), 4) et 5). Ils ont la même consonne finale (CF) de SFR "ㄷ". Tous les verbes qui ont comme CF du SFR "ㄷ" sont classés NVT3 ou NVT4 dans la table 2-2. L'exemple 3) est classé comme NVT3, mais les deux autres verbes sont classés comme NVT4. Il n'y a pas de principe pour choisir les formes correctes. Parmi 79 verbes, 44 verbes entrent dans la classe NVT3 et le reste des verbes entrent dans NVT4.

3) 받다 (recevoir) (V) : 받음 (la reçue) (FN)
4) 듣다 (écouter) (V) : 들음(l'écoute) (FN)
5) 깨닫다(comprendre) (V) : 깨달음(la compréhension) (FN).

L'irrégularité des formes nominalisées dépend de SFR. Autrement dit, si deux verbes différents ont la même SFR, ils auront la même forme nominalisée.

Nous avons établi linguistiquement la liste des VN dérivés du verbes. Ce sont les formes du présent, et leur classification est décrite dans la table 2-2.

En ce qui concerne la construction des VN, l'élément le plus important est le SFR. Le choix des PPP dépend de la SFR. NVT1 et NVT2(-1, -2) sont des classes typiques des suffixes nominalisateurs. Les verbes qui se combinent avec une forme de NVT1 comportent la CF dans SFR, à la différence des verbes de NVT2-1. NVT2-2 est une classe spéciale. Selon le système de code du coréen, appelé **KSC5601-1992**, il n'existe pas de forme nominalisée pour les trois verbes de NVT2-2 [3].

Les formes des VN de NVT5-1, NVT10-1 et NVT12-1 sont obtenues respectivement de NVT5-2, NVT10-2 et NVT12-2. Bien qu'elles s'utilisent très rarement dans les phrases écrites, ces formes sont acceptables.

NVT2-2: 건늬다 뇌늬다 뙤다(3)
NVT4 : 겉묻다 곁다 곧이듣다 귀넘어듣다 귀담아듣다 귀여겨듣다 그러묻다 긷다 깨닫다 껴묻다 내닫다 눋다 덧묻다 되묻다 되받아묻다 듣다 묻다 새겨듣다 새기어듣다 알아듣다 얻어듣다 여겨듣다 엿듣다 일컫다 주어듣다 주워듣다 줏어듣다 지내듣다 처싣다 치닫다 캐묻다 캐어묻다 헛듣다 설듣다 싣다(35)
NVT6 : 되살다 막살다 못살다 살다 숨어살다 알다 얹혀살다 쥐여살다(8)
NVT8 : 가로눕다 곱다 굽다 깁다 꼽다 눕다 돌아눕다 돕다 뒤눕다 드러눕다 몸져눕다 줍다 짜깁다 여쭙다(14)
NVT9 : 뵈옵다 여쭈웁다 여쭙다(2 + 1)
NVT11: 내솟다 드솟다 들솟다 들이웃다 발가벗다 벌거벗다 벗다 비웃다 빗다 빨가벗다 빼앗다 뺏다 뻘거벗다 솟다 씻다 앗다 웃다 치솟다 헐벗다(19)

Liste 2-7. Liste des verbes de la table 2-2.

La liste 2-7 donne les verbes des classes de la table 2-2. Les verbes à CF comme "ㄷ" sont classés soit dans NVT3, soit dans NVT4. Les verbes qui entrent dans NVT4 sont énumérés dans la liste 2-7. Remarquons que les verbes qui ne se trouvent pas dans NVT4 tout en comportant CF "ㄷ" sont classés dans NVT3.

2.2.2. Adjectifs Nominalisés

Les procédures de construction des AN sont les mêmes que pour les VN. "기" (*suffixe complémentaire*) peut être attaché à SFR de façon automatique, alors que "음" et "ㅁ" se combinent différemment avec les adjectifs en fonction des formes de SFR. Les suffixes engendrent beaucoup de formes fléchies et ils dépendent des formes de SFR des adjectifs.

Les deux exemples ci-dessous sont classés comme NAT10 et NAT11. Les 90 adjectifs qui ont le CF "ㅎ" de SFR entrent normalement dans NAT10 comme dans le cas 1). Mais 3 de ces 90 adjectifs entrent dans NAT11. C'est le cas 2). Il n'existe pas de critère pour classer comme NAT10 ou NAT11.

1) 노랗다 (est jaune) (A) : 노람 (le jaune) (FN)
2) 좋다 (est bon) (A) : 좋음 (la bonté) (FN).

Le dictionnaire de AN est construit linguistiquement. Les résultats sont décrits dans la table 2-3.

TYPE	CF dans SFR	Substitution	Nombre	Total	Exemple
NAT1		음	276	276	맑다(맑음)
NAT2-1		ㅁ	4,303	4,304	가깝하다(가깝함)
NAT2-2		non exister	1		애교다
NAT3-1	ㄹ	ㅁ	40	40	가늘다(가늠)
NAT3-2		ㄹ + 음	40		가늘다(가늘음)
NAT4		름	1		서툴다(서투름)
NAT5	ㅂ	움	391 +143	538	노엽다(노여움)
NAT6		ㅂ + 음	4		비좁다(비좁음)
NAT7		ㅁ	391 + 70		노엽다(노염)
NAT8	ㅅ	음	1	2	낫다(나음)
NAT9		ㅅ + 음	1		헐벗다(헐벗음)
NAT10	ㅎ	ㅁ	87	90	노랗다(노람)
NAT11		ㅎ + 음	3		좋다(좋음)
				5,250	

Table 2-3. Les Classes des AN.

La liste des adjectifs classés dans la table 2-3 se trouve dans la liste 2-8. Les AN de NAT4 dans la table 2-3 sont les formes qui ont subis la liaison à partir des adjectifs de NAT3-2. Les 70 entrées dans NAT7 ont été sélectionnées parmi les 143 entrées de NAT5. Il s'agit d'exemples typiques qui ne peuvent être sélectionnés que par le linguiste. NAT2-2 est un cas particulier. La forme de AN correspondante n'existe pas dans le système de code du coréen standard[32].

Certains VN et AN s'utilisent très rarement mais on en trouve des formes dans les dialectes et dans la langue parlée.

NAT6 : 비좁다 수줍다 좁다 좁디좁다
NAT7*: 가렵다 가소롭다 가파롭다 감미롭다 건조롭다 경사롭다 경이롭다 고롭다 공교롭다 괴까다롭다 괴롭다 굽다 권태롭다 귀엽다 까다롭다 꼬깝다 꼴사납다 꾀까다롭다 날카롭다 노엽다 놀랍다 다사롭다 다채롭다 단조롭다 도탑다 두렵다 두텁다 따사롭다 따습다 똥마렵다 뜨습다 마렵다 매섭다 매스껍다 메스껍다 명예롭다 무섭다 반지랍다 번거롭다 보드랍다 보배롭다 사사롭다 상서롭다 새롭다 순조롭다 신비롭다 아니꼽다 아쉽다 애처롭다 어둡다 영예롭다 외롭다 ᄋᆞ애롭다 위태롭다 의롭다 의외롭다 이롭다 자비롭다 자애롭다 자혜롭다 쟁그랍다 정겹다 정교롭다 정의롭다 조화롭다 지혜롭다 평화롭다 한가롭다 향기롭다 호화롭다(70)
NAT12: 맷집좋다 의좋다 좋다(3)

Liste 2-8. Liste des adjectifs de la table 2-3.

2.2.3. Formes Spéciales

Les verbes que nous avons discutés dans les paragraphes 2.2.1 et 2.2.2 sont au présent. Nous discuterons ici les autres *temps* ou *modalités*.

Les trois formes de NPT1 dans la liste 2-6 représentent trois SFN du *mode*. La première "냐기" exprime le mode *interrogatif* et le *discours indirect*. La forme nominalisée avec suffixe flexionnel implique le sens *'demander si ...'*. La deuxième "다기" est la forme transformée du prédicat de *discours indirect*. La dernière "자기" est le SFN *prépositif*. "가자기" (d'aller ensemble) (VN), par exemple, il est composé de "가" (le radical) et de "자기" (SFN).

NPT2, NPT3, NPT4, NPT5, NPT6 et NPT7 sont aussi complexes que NPT1. Les séries "겠" (autrement dit, les SFN dont la première syllabe est "겠") dans NPT3 dénotent la *supposition* comme modalité et aussi le *futur*. Les séries '셨' dans NPT4, '시' dans NPT5, '우' dans NPT6 et '으' dans NPT7 expriment la *politesse* comme *modalité*. Toutes les formes de la liste 2-9 s'attachent après le SFR et les mots qui les contiennent seront marqués comme FN.

NPT1 : 냐기 다기 자기(3)
NPT2 : 는데(1)
NPT3 : 겠기 겠냐기 겠다기 겠음(4)
NPT4 : 셨겠기 셨겠냐기 셨겠다기 셨겠음 셨기 셨냐기 셨다기 셨었겠냐기 셨었겠다기 셨었기 셨었냐기 셨었다기 셨었음 셨음(14)
NPT5 : 시기 시냐기 시다기 시겠기 시겠냐기 시겠다기 시겠음 시□解未□ 시었겠냐기 시었겠다기 시었겠음 시었기 시었냐기 시었다기 시었었겠기 시었었겠냐기 시었었겠다기 시었었겠음 시었었기 시었었냐기 시었었다기 시었었음 시었음 시자기(24)
NPT6 : 으라기 으셨겠기 으셨겠냐기 으셨겠다기 으셨겠음 으셨기 으셨냐기 우셨다기 으셨었겠냐기 으셨었겠다기 우셨었기 우셨었냐기 으셨었다기 우셨었음 우셨음 으시겠기 우시겠냐기 우시겠다기 으시겠음 우시었겠기 우시었겠냐기 으시었겠다기 우시었겠음 우시었기 으시었냐기 우시었다기 우시었었겠기 우시었었겠냐기 으시었었겠다기 우시었었겠음 우시었었기 우시었었냐기 으시었었다기 우시었었음 우시었음 우시기 우시냐기 우시다기 우시자기 으심 (40)
NPT7 : 으라기 으셨겠기 으셨겠냐기 으셨겠다기 으셨겠음 으셨기 으셨냐기 으셨다기 으셨었겠냐기 으셨었겠다기 으셨었기 으셨었냐기 으셨었다기 으셨었음 으셨음 으시겠기 으시겠냐기 으시겠다기 으시겠음 으시었겠기 으시었겠냐기 으시었겠다기 으시었겠음 으시었기 으시었냐기 으시었다기 으시었었겠기 으시었었겠냐기 으시었었겠다기 으시었었겠음 으시었었기 으시었었냐기 으시었었다기 으시었었음 으시었음 으시기 으시냐기 으시다기 으심 으시자기 (40) (Total 218)

Liste 2-9. Les formes particulières de SFN

Pour dénoter le *passé*, six syllabes sont disponibles. Il s'agit de 렀 (NPT8), 았 (NPT9) , 었 (NPT10) , 였 (NPT11), 왔 (NPT12), et 웠 (NPT13) dans la liste 2-10. Ces suffixes sont attachés après SFR.

Les verbes et les adjectifs peuvent être classés en fonction de la forme de leur SFR. Pour construire les Formes Spéciales de FN, nous devons savoir *quel SFR se combine avec quel SFN*. Nous devons donc avoir l'information sur la connexion entre SFR et le prédicat. Dans ce but, nous en avons fait une liste des SFR dans la figure 2-10. Dans cette figure, SFN est inclue dans PPP.

NPT8 : 렀겠기 렀겠냐기 렀겠다기 렀겠음 렀기 렀냐기 렀다기 렀었겠기 렀었겠냐기 렀었겠다기 렀었겠음 렀었기 렀었냐기 렀었다기 렀었음 렀음(16)
NPT9 : 았겠기 았겠냐기 았겠다기 았겠음 았기 았냐기 았다기 았었겠기 았었겠냐기 았었겠다기 았었겠음 았었기 았었냐기 았었다기 았었음 았음(16)
NPT10: 었겠기 었겠냐기 었겠다기 었겠음 었기 었냐기 었다기 었었겠기 었었겠냐기 었었겠다기 었었겠음 었었기 었었냐기 었었다기 었었음 었음(16)
NPT11: 였겠냐기 였겠다기 였기 였냐기 였다기 였었겠냐기 였었겠다기 였었기 였었냐기 였었다기 였었음 였음(12)
NPT12: 왔겠기 왔겠냐기 왔겠다기 왔겠음 왔기 왔냐기 왔다기 왔었겠기 왔었겠냐기 왔었겠다기 왔었겠음 왔었기 왔었냐기 왔었다기 왔었음 왔음(16)
NPT13: 웠겠기 웠겠냐기 웠겠다기 웠겠음 웠기 웠냐기 웠다기 웠었겠기 웠었겠냐기 웠었겠다기 웠었겠음 웠었기 웠었냐기 웠었다기 웠었음 웠음(16)

Liste 2-10. Les formes du *passé* pour SFN.

Classe 4	**SFR**	**Type PPP**
듣다	듣	SU1-5 6
		SU2-1
		SU3-1 2 4 5 7 8 9 10 12 13 14 15 17 21 23 24
		SU4-1
		PV1-1 3
		PV4-1
	들	SU2-2 3 4 5
		PV4-1

Figure 2-10. Exemple d'un verbe '듣다' du dictionnaire SFR.

La figure 2-10 est un exemple de (e.g. '듣다'(écouter)). Le dictionnaire comporte l'information sur la connexion entre le verbe et le PPP. Le verbe a deux formes de SFR "듣" et "들". Dans le cas de '듣', il a l'information incluant SU3-1, 2, 4, ... dans la zone **Type PPP**. Cela signifie que les données de SU3-1, 2, 4, ... peuvent être soudées après SFR '듣'.

Un autre exemple, celui de '겠기, a l'information SU3-4 dans la zone **Type PPP** de la figure 2-11. Par conséquent, SFR "듣" peut être combinée avec le suffixe "겠기". Le résultat "듣겠기" est marqué comme FN.

Il existe une autre structure qui correspond à l'*infinitif* en français. C'est celle de la forme *modifieur*-"것" (suffixe complémentaire) en coréen. Elle comporte plusieurs formes fléchies selon le temps. Nous appelons ce type de modifieur PPP modificatif. PPP modifieur est un sous-ensemble de PPP.

SFN entré	PDD[1]	**de V/A**	**PPN Type**	**Type PPP**
겠기	SFN	V et A	PPN	SU3-4-NM
겠냐기	SFN	V et A	PPN -PN1	SU3-4-ID
겠다기	SFN	V et A	PPN -PN1	SU3-4-ID
겠음	SFN	V et A	PPN	SU3-4-NM
기	SFN	V et A	PPN	SU3-7-NM
냐기	SFN	V et A	PPN -PN1	SU3-9-ID
는데	SFN	V	PPN	SU3-13-CJ-OV
다기	SFN	V et A	PPN -PN1	SU3-14-ID
렀겠기	SFN	V et A	PPN	PP7-2-NM
렀겠냐기	SFN	V et A	PPN -PN1	PP7-2-ID

Figure 2-11. L'exemple des SFN.

[1] PDD : parties du discours

2.2.4. Procédures de traitement des ambiguïtés

2.2.4.1. Ambiguïté **P1**

Nous avons déjà discuté la différence entre NS et FN. L'ambiguï té **P1** n'est pas importante au niveau morphologique. La différence d'information a lieu aux niveaux syntaxique ou pragmatique. Dans la liste 2-11 les formes d'ambiguï té **P1** sont énumérées, en vue de l'analyse syntaxique ultérieure.

가늠 가뭄 간지럼 갈기 감 감기 개기 검 검침 결함 고기 고름 고함 곰 괴로움 구름 귀염 그리움 금 기름 기울기 기함 김 깨달음 꿈 끄스름 남 내기 노름 노여움 녹음 놀음 담 대감 대기 대함 댐 덤 도움 돌기 동침 두려움 뒤집기 딸기 땀 땜 맘 매기 맴 메기 명함 모기 몸 무서움 미끄럼 미움 밀림 바람 반가움 배기 뱀 범 봄 봉함 부끄러움 부끄럼 부럼 부음 부품 사기 살기 살림 삶 삼 샘 서기 서러움 석기 섬 세기 셈 속기 슬기 시기 식기 식음 신기 신음 실기 앎 어두움 어둠 어름 어림 어지러움 어지럼 얼뜨기 얼띠기 엉큼 연임 염 오금 오기 옴 움 이기 이름 이음 임 자기 잡기 재기 잼 전함 절기 점 조림 좀 주기 줄기 지겨움 지기 지름 지침 직함 집기 짬 차기 참 추기 춤 치기 침 탐 태기 틈 파기 패기 폐함 품 함 희기(143)

Liste 2-11. Les ambiguïtés **P1** entre NS et FN.

감 겸 곰 금 남 담 범 심 엄 염 점 좀 참 품 함(15)

Liste 2-12. Les ambiguïtés **P2** entre PF et FN.

감 검 금 남 내기 담 대기 데기 때기 뜨기 뛈 바람 배기 범 부림 빔 빼기 심 싸대기 씀 염 임 잠 점 지기 짐 짜기 참 춤 치기 침 태기 패기 품 함 (35)

Liste 2-13. Les ambiguïtés **P1** entre SF et FN.

Les listes 2-12 et 2-13 figurent dans la liste des ambiguï tés **P1**. SF et PF sont utilisés facultativement, mais FN est obligatoire dans la structure T2. Nous pouvons utiliser cette information pour lever l'ambiguïté.
Nous avons choisi en priorité le FN dans le cas d'ambiguïté **P1** avec SF et PF.

2.2.4.2. Ambiguïté **P2**

FN peut provoquer l'ambiguï té **P2** entre PF, SF, et PPN. La liste 2-14 est la liste des candidats localisés dans la sphère 2 ayant l'ambiguï té **P2** entre PF et FN. Nous n'avons pas découvert les données susceptibles d'appartenir à la sphère 4 sous T2. Autrement dit, il n'y a pas d'ambiguï té **P2** dans FN. Les variables *x*, *y*, et *z* dans la liste sont expliquées dans la formule **P2**.

강남(xy)	국적이기 국적임 기 기기 김 다르기 다름 루하기 루함 부끄러움 부끄럼 부끄럽기 부럽잖기 부럽잖음 상거리기 상거림 상남상하기 상남상함 성적이기 성적임 실거리기 실거림 실남실하기 실남실함 아넘치기 아넘침 아돌기 아돌아가기 아돌아감 아돌음 아돔 음짓하기 (z) (33)
보금(xy)	실금실하기 실금실함 쳐놓기 하기 함 (z) (5)
어금(xy)	실금실하기 실금실함 쳐놓기 하기 함 (z) (5)
징검(xy)	기 누래지기 누래짐 누럼 누렇기 누르기 누름 디검기 디검음 뜯기 뜯기기 뜯김 뜯음 붉기 붉음 붉히기 붉힘 뿌염 뿌옇기 소하기 소함 숭검숭하기 숭검숭함 숭하기 숭함 실거리기 실거림 실검실하기 실검실함 음 적검적하기 적검적함 치기 침 퍼래지기 퍼래짐 퍼럼 퍼렇기 푸르기 푸르접접하기 푸르접접함 푸르죽죽하기 푸르죽죽함 푸름 (z) (44)

Liste 2-14. Les candidats à l'ambiguïté **P2** dans FN.

2.3. Structure de type 3

Le type 3 concerne le prédicat à PPP dans l'interprétation traditionnelle de la phrase, la prédicat joue le rôle principal dans la structure de la phrase. Les entrées prédicatives relatifs aux éléments lexicaux de PPP peuvent être attachées à [10] [28]. En français, il y a trois types des séquences prédicatives. PPP est attaché seulement au verbe (la deuxième colonne la table 2-4 est le cas du verbe ordinaire, la troisième colonne horizontale est pour la copule, et la quatrième est lexicalement vide mais il est grammaticalement le verbe) [28].

En coréen, nous avons observé trois types de séquences prédicatives. PPP n'est pas seulement attaché aux verbes mais aussi aux adjectifs et aux unités grammaticales liées (GL) aux noms.

PDD	Type français	Coréen	Ex. en français	Ex. Coréen
verbes	N V-**PPP**	N V-**PPP**	Jean **dort**	존은 잔다
adjectif	N Etre-**PPP**Adj	N A-**PPP**	Jean **est** grand	존은 크다
nom prédicatif	N V-**PPP** Npréd	N N-GL-**PPP**	Jean **fait** une plaisanterie (= Jean **plaisante**)	존은 농담한다

Table 2-4. L'exemple des types des mots prédicatifs.

A partir de ces trois types, nous avons défini la structure T3. Les mots qui contiennent un prédicat et un SF sont classés comme T3.

$$T3 = (PF^{?}(NS \mid FN)^{+} SF^{?})^{*} (V_{root} \mid A_{root})$$

(Postposition verbale | Postposition adjectivale)

$$= (ND \mid PF^{?} FN^{+}SF^{?})^{*} (V_{root} \mid A_{root}) PPP$$

$$= (ND \mid PF^{?} FN^{+}SF^{?})^{*} FFP$$

T3 est constitué de trois parties. La première concerne le nom prédicatif. Le nom prédicatif est le nom qui autorise l'attachement du verbe. Il y a d'autres formes à la place des noms de la première partie, comme avec les quatre types suivants :

1) "물컹하다" (être mou) est formé de l'adverbe "물컹" (mou) et du prédicat (dans ce cas, l'adjectif) "하다"(faire). Pour "하다", nous avons observé que les adverbes qui peuvent être combinés avec "하다" afin de former un adjectif n'acceptent pas le suffixe verbal "하다". Par contre, ces adverbes peuvent être combinés avec le suffixe verbal "거리다". Ainsi une verbe comme "물컹거리다" peut être produit. Ce mécanisme est observé très fréquemment. Cependant, même si la procédure est très productive, elle n'est pas applicable à tous les adverbes qui entrent dans l'une de ces relations dérivationnelles.

2) Certains prédicats peuvent être attachés à la tête de l'autre prédicat. Par exemple, "놀고먹다" (manger en s'amusant) est composé des deux verbes "놀고" (s'amuser) et "먹다" (manger).

3) Les deux autres formes complexes du prédicat telles que préfixe-prédicat et prédicate-suffixe peuvent être considérées. Pour la composition de préfixe-prédicat, par exemple "엇물리다" est composé du préfixe "엇" (obliquement) et du verbe "물리다" (lier).

4) La forme prédicat-suffixe est dérivée du verbe simple par l'ajout des affixes. Le suffixe peut être un morphème grammatical qui transforme *le verbe transitif* au *passif*, ou *le verbe intransitif* à la forme *causative*. "박히다" (être enfoncé) est un exemple du premier cas. La forme 'enfoncer-suffixe.pass' transforme en 'être enfoncé'. "굶기다" représente l'exemple du second cas. La forme 'être affamé-suffixe.caus' se transforme en 'affamer'.

Nous avons enregistré les quatre types complexes, ci-dessus, dans notre lexique avec la marque VC (Verbe Complexe).

La deuxième partie concerne la radical des prédicats. Cette partie ne connaî t pas de variations. Pour ajouter la partie finale de T3 à la deuxième partie, il nous faudra définir un nouveau concept appelé SFR. C'est la dernière syllabe du radical. Pour attacher la dernière partie à la fin du radical, il est nécessaire de connaî tre l'information sur le type de SFR. La forme PPP de la dernière partie dépend de SFR. SFR joue le rôle du connecteur de l'interface entre le radical (deuxième partie) et le PPP (dernière partie) dans T3.

SFR permet de classer 8,763 verbes en plusieurs sous-classes. Si deux radicaux possèdent le même caractère final, elles ont généralement le même type de SFR.

La dernière partie est sur PPP. PPP ne peut exister seul. Il a toujours besoin du verbe ou de l'adjectif. Les verbes et les adjectifs utilisent environ 6,000 formes de PPP dans les mots réels [26]. Nous donnerons l'information exacte sur la connexion entre la partie du prédicat et celle de PPP.

2.3.1. Noms Prédicatifs

Le nom prédicatif est un nom qui peut s'attacher du prédicat. Par exemple, "공부하다" (étudier) est composé du nom (N) "공부" (étude) et du verbe (V) "하다" (faire). Ce mot peut être transformé en deux formes disjointes N-PPN et V, comme "공부를 하다" (N-PPN V). Ces deux unités disjointes sont des éléments syntaxiques, car on peut insérer d'autres éléments grammaticaux entre elles. C'est pourquoi nous ne considérons pas cette forme comme un verbe simple. Par contre, nous n'avons pas observé deux unités disjointes du point de vue syntaxique, quant il s'agit d'une forme basée sur l'adjectif [28] [29].

Voici un exemple de phrase à un verbe support '주다' (*juda*) [donner] :

(1) 민우는 인아에게 도움을 주었다
Minu-PPN Ina-PPN aide-PPN donner-PPP
(Minu a donné de l'aide à Ina)

Cette phrase correspond à la construction à verbe distributionnel '돕다' (*dobda*) [aider], lexicalement relié au substantif '도움'(*doum*) [aide] :

(2) = 민우는 인아를 도왔다
Minu-PPN Ina-PPN aider-PPP
(Minu a aidé Ina)

On peut décrire cette relation comme suit [14] [31] :

(3a) *N0* *N1* *Npréd Vsup*
(3b) = *N0* *N1* *V*

La relation (3) s'observe quand il existe un verbe morphologiquement relié au substantif prédicatif. Cependant, avec le verbe '하다' (*hada*) [faire], on a systématiquement ce type de paires de phrases. Ainsi, la phrase suivante [18] :

(4) 민우가 산책을 한다
Minu-PPN promenade-PPN faire-PPP
(Minu fait une promenade)

est en relation transformationnelle avec :

(5) = 민우가 산책한다
Minu-PPN se promener-PPP
(Minu se promène)

Or, la relation entre la phrase à support *hada* et celle à verbe distributionnel correspondant est particulière du point de vue morphologique : l'une contient une séquence Nom-PPN *hada*, et l'autre Nom-*hada*. Autrement dit, le changement de forme

entre ces deux types de prédicats est minimal, il se limite à la présence d'une postposition de l'accusatif. Par ailleurs, cette transformation est permise, a priori, dans toutes les phrases à support *hada,* sans exception

Dans des dictionnaires actuels, les formes Nom-*hada* dont les substantifs (Nom) sont considérés comme prédicatifs sont intégrées sous une entrée de verbe dérivé. Par conséquent, on compte d'une part ces substantifs dans le lexique des substantifs, et d'autre part les verbes basés sur ces substantifs (Nom-*hada*) dans le lexique des verbes. Cette situation est certainement gênante quand les noms de ce type sont nombreux et surtout quand leur liste n'a pas été établie par un procédé systématique. On risquerait de dédoubler le lexique partiellement avec les noms associables à *hada.*

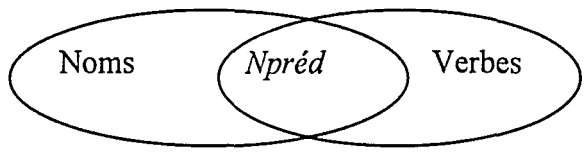


Figure 2-12. Position des Noms Prédicatifs

Il y a des noms qui ne fonctionnent pas comme nom prédicatif. Mais s'ils sont accompagnés d'un suffixe "질" (acte) ou "화" (devenir), les formes N-SF-Prédicat deviennent acceptables, comme ci-dessous :

(1) 가위-질-하다 (ciseaux- acte -faire) (N-SF- Prédicat)
(2) 세계-화-하다 (monde-devenir -faire) (N-SF- Prédicat).

Nous possédons le dictionnaire des noms prédicatifs qui peuvent être accompagnés du verbe '하다' (faire) ou '되다' (devenir), mais, il n'existe ni dictionnaire de tous les noms prédicatifs ni liste complète des prédicats qui peuvent être attachés aux noms prédicatifs.

가다 거리다 끼치다 나다 내다 당하다 되다 떨다 미치다 받다 부리다 시키다 입다 주다 지다 짓다 치다 피우다 하다

Liste 2-15. Liste des verbes utilisés fréquemment avec les noms prédicatifs.

가지다 같다 아니다 없다 이다 있다(6)

Liste 2-16. Liste des adjectifs utilisés fréquemment avec les noms prédicatifs.

Comme le dictionnaire complet des noms prédicatifs n'existe pas, nous utilisons la règle T3. Les listes 2-15 et 2-16 contiennent des prédicats très productifs, qui, entre autres, se combinent avec le nom prédicatif Nom-*hada* sous T3. Mais on n'a pas observé de restrictions sur les prédicats.

Dans le cas de l'adjectif, les noms prédicatifs sont ajoutés avec des restrictions à la tête de "하다" (faire). Même si la liste est longue, nous l'avons construite par un recensement linguistique des données [26]. En effet, l'information est décrite dans le dictionnaire des adjectifs. Celui-ci possède 3,231 entrées dont la forme est composée d'un Nom Prédicatif et de "하다".

2.3.2. Verbes

Notre système comporte une liste de 8,763 verbes simples ou complexes. Comme nous l'avons déjà remarqué, les verbes complexes peuvent avoir quatre types de marque, comme par exemple VC dans le dictionnaire des verbes.

Il existe 80 verbes simples qui sont formés d'éléments non-autonomes et du suffixe "하다", comme "택하다". Tous ces éléments non-autonomes sont sino-coréens à l'origine.

VERBE	PDD	Information		
가까와지다	VS	/INT	/JMA	
가꾸다	VS	/TRA		
가닐거리다	VS	/INT	/ADVKM	
가동거리다	VS	/TRA	/INT	/RKM
가꾸로박히다	VC	/INT	/PS	
가라앉히다	VC	/TRA	/CS	
나자빠지다	VC	/INT	/JM	

Figure 2-13. Extrait de la liste des verbes simples.

Dans la figure 2-13 sont donnés quelque exemples de la liste des verbes. Pour établir l'information de la connexion avec PPP, nous classerons les verbes en plusieurs groupes au paragraphe 2.3.4.

2.3.3. Adjectifs

La catégorie Adjectif représente les formes prédicatives qui se conjuguent au moyen de certains suffixes grammaticaux, de façon analogue au verbe. Dans notre système, il y a 5,250 adjectifs et ils sont classés en fonction de plusieurs formes finales. Par exemple, la classe CM du dictionnaire indique l'adjectif avec le suffixe "이다" (être). Les adjectifs entrant dans cette classe connaissent 800 formes flexionnelles environ. Les formes conjuguées permettent nécessairement de reconnaî tre les entrées lexicales. La figure suivante est un extrait de la liste des adjectifs [23].

Adjectif	PDD	Information
가공적이다	ADJ	/CM
가깝다	ADJ	/RM
가깝디가깝다	ADJ	/RM
가깝하다	ADJ	/HM
가난하다	ADJ	/HM

Figure 2-14. Exemples d'adjectifs.

Pour établir l'information de la connexion avec PPP, nous allons classer les adjectifs en plusieurs groupes dans ce qui suit.

2.3.4. Syllabe Finale du Radical

Quel prédicat peut être combiné avec quelle série flexionnelle ? Supposons que les combinaisons soient sans restriction. Le nombre N de combinaisons possibles (Prédicat + PPP) sera le suivant:

N = Nombre des verbes * Nombre de Postposition verbale
+ Nombre des adjectifs * Nombre de Postposition adjectivale
= 8,763 * 20,853 + 5,250 * 18,893 = 281,923,089.

Dans le cas normal, un prédicat peut être combiné avec environ 6,000 PPP [26]. Le nombre estimé des formes combinées est le suivant:

Le nombre estimé
= Nombre des verbes * 6,000 + Nombre des adjectifs * 6,000
= 8,763 * 6,000 + 5,250 * 6,000 = 84,078,000.

Pour extraire de manière exacte environ 84,078,000 formes de la combinatoire entre les prédicats et PPP, nous devons examiner 281,923,089 données. Le nombre est trop énorme pour une recherche manuelle.

Il est indispensable d'appliquer une méthode semi-automatique pour sélectionner les formes correctes FFP. Après avoir classé les prédicats et les PPP, nous fournirons l'information de connexion entre le groupe des prédicats et celui des PPP. La classification des SF sera discutée dans le paragraphe suivant.

Pour classer les prédicats, le concept de SFR est indispensable. Postposition verbale (PPV) et Postposition adjectivale (PPA) s'attachent après la radical du verbe ou de l'adjectif. Si deux prédicats comportent la même SFR avec le même sens, ils auront le même PPP. Par exemple, avec les deux verbes "나가다" (sortir) et "가져가다" (emporter). Les radicaux de ces verbes sont "나가" et "가져가". Ils ont la même SFR "가". Ils ont également les mêmes formes fléchies, comme dans la figure 2-15.

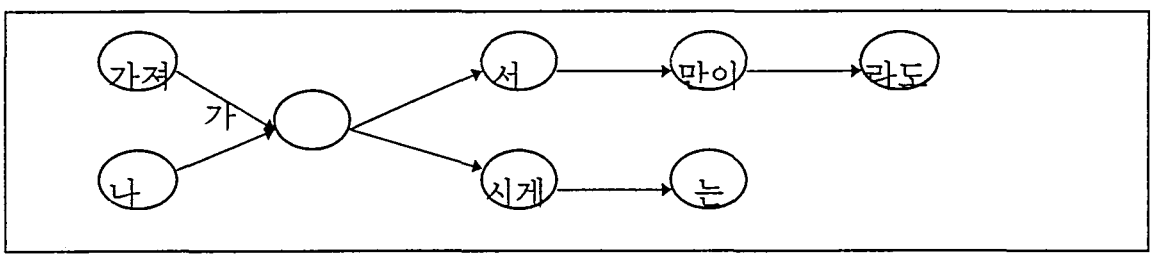


Figure 2-15. L'exemple pour les deux verbes ayant la même SFR.

SFR précède PPP, parce que SFR fait partie du prédicat, et le prédicat précède PPP dans la structure T3. Il se réfère toujours à la forme de PPP. Dans certains cas, une compression entre SFR et la première syllabe du PPP peut avoir lieu, et dans d'autre cas, la dernière consonne du SFR peut être supprimée ou changée.

Les figures 2-16-1 et 2-16-2 donnent des exemples de formes fléchies. La différence entre ce deux figures porte sur un problème de code du coréen. Actuellement on utilise plusieurs systèmes de code [3] [32]. Cela provoque des représentations différentes et des analyses différentes pour les formes de SFR.

Nous avons décidé d'utiliser le système de code à 2 octets (appelé **KSC5601 WANSUNG Code**) comme code standard du coréen pour représenter les syllabes [33]. Le concept est hérité du système **UniCode**. Il est difficile de représenter l'unité phonémique en coréen avec ce dernier car le système est basé sur la syllabe.

La figure 2-16-1 comporte les valeurs de phonème "ㄴ", "ㄹ", et "ㅆ". Le premier noeud "가" et les phonèmes dans la figure 2-16-1 ne peuvent pas se combiner naturellement, parce que le code standard n'a pas de règle pour la combinaison entre une syllabe et un phonème. Selon les règles d'écriture du coréen, la syllabe "가" et le phonème "ㄴ" doivent produire la syllabe "간". C'est impossible dans le code standard du coréen sans programme particulier de composition. Dans d'autres systèmes de code du coréen, ces formes se combinent sans difficulté. C'est-à-dire que la syllabe "간" contient l'information sur la syllabe "가" et sur le phonème "ㄴ" dans le système du code.

La présentation de code **WANSUNG** de la figure 2-16-1 est décrit dans la figure 2-16-2. Même si la forme est différente de celle de la figure 2-16-1, le contenu est identique. On peut reconstituer les formes de la figure 2-16-2 au moyen de celles la figure 2-16-1. Par exemple, le premier noeud "가" et l'élément phonémique "ㄴ" du noeud "ㄴ 지" de la figure 2-16-1 sont combinés linguistiquement, et le résultat "간" est inscrit dans le nouveau noeud de la figure 2-16-2. Nous avons appliqué le même principe à tous les éléments phonémiques dans la figure 2-16-1 et nous avons obtenu la figure 2-16-2. Dans la figure 2-16-1, il n'y a qu'un seul noeud de SFR "가", mais dans l'autre, il y a les noeuds de SFR "간", "갈"et "갔".

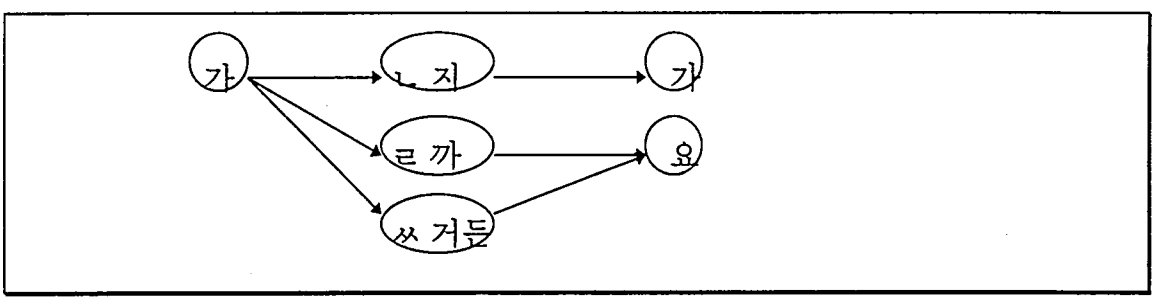


Figure 2-16-1. Une SFR et PPPs avec l'information phonémique.

Par conséquent, dans la figure 2-16-1, il y a un noeud SFR "가", mais dans la figure 2-16-2, il y a trois noeuds SFR. Si on ne tient pas compte de la notation et du système de code, il n'y a pas de différence entre les figures 2-16-1 et 2-16-2.

Il n'y a pas de principe général de variation des SFR. Dans certains cas, la compression entre SFR et la première syllabe de la SF se produit mais dans d'autres elle n'a pas lieu. Nous devons donc vérifier toutes les variations.

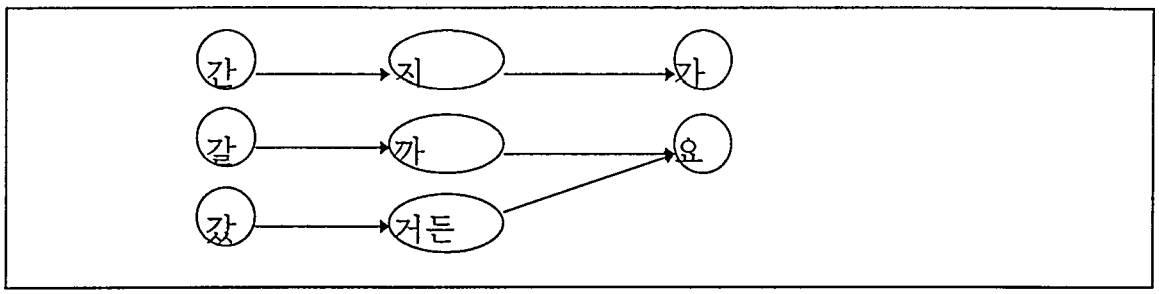


Figure 2-16-2. Trois SFRs et PPPs avec l'information syllabique.

Nous avons énumérés 333 types de SFR pour les verbes et 151 types de SFR pour les adjectifs. Les listes 2-17 et 2-18 donnent les SFR respectivement compatibles avec les verbes et avec les adjectifs. Nous remarquons que le SFR "르" que l'on trouve dans les verbes et dans les adjectifs a des variations particulières. Dans ce cas, le nombre des SFR est 2. Le détail est décrit dans les classes de SFR. Supposons que deux formes différentes de SFR aient le même PPP. Les deux SFR entrent dans la même classe de SFR.

가 갈 갉 감 갖 갚 개 걷 걸 게 겠 겪 견 결 고 골 곪 곯 곱 괴 구 굳 굴 굶 굽 귀 그 글
긁 금 긋 기 긷 길 깁 까 깎 깔 깜 깨 꺾 꼬 꼽 꽂 꾀 꾸 꿇 꿰 뀌 끄 끊 끌 끓 끗 끼 나
낚 날 남 낫 낳 내 너 널 넘 넣 네 녹 놀 놓 뇌 누 눋 눕 뉘 느 늘 늙 늬 니 닐 다 닦 닫
달 닮 닳 담 닿 대 덜 덮 데 돋 돌 돕 되 두 둘 듣 들 듬 디 딛 딪 따 딸 땅 때 떨 떼 뙤
뚫 뛰 뜨 뜯 뜰 띄 띠 라 래 러 레 뢰 루 르 리 막 말 맞 맡 매 맺 먹 멀 멎 메 모 몰 묵
묶 묻 물 미 믿 밀 박 받 밟 밭 배 뱉 벋 벌 벗 베 보 볶 뵈 붇 불 붓 붙 비 빌 빗 빚 빨
빻 빼 뺏 뻗 뽑 뿜 삐 사 삭 살 삶 삼 새 서 석 섞 설 세 속 솎 솟 쇠 수 숙 숨 숫 쉬 슬
슴 시 식 신 싣 실 싸 쌓 쌔 썩 썰 쏘 쏟 쏠 쐬 쑤 쓰 쓸 씌 씹 씻 안 앉 않 알 앍 앓 암
앗 애 얹 얻 얼 얽 업 엎 에 엮 열 오 옭 옮 옵 외 욍 울 웁 웃 위 으 을 읊 의 이 익 일
읽 잃 입 잇 있 잊 자 잡 잣 잦 재 적 절 접 젓 젖 졸 좇 좋 죄 주 죽 줄 줍 쥐 지 집 짓
짖 짚 짜 째 쩔 쪼 쫓 쬐 쭈 쭙 찌 찍 찝 찢 찧 차 참 찾 채 추 치 캐 컫 켜 퀴 크 클 키
타 태 털 투 튀 트 틀 틔 티 파 팔 패 펴 푸 풀 품 피 하 핥 헐 홅 훑 휘 흩 히(333)

Liste 2-17. Liste des SFR apparaîssant dans les verbes.

감 갑 같 갛 걸 검 겁 겋 결 겹 곧 곱 괠 교 굳 굵 굽 궂 글 기 길 깊 깜 깝 껌 껍 꼽 나
낡 납 낫 낮 넓 넘 녹 높 눅 늘 늙 늦 니 달 닮 답 대 덥 되 둡 들 디 딸 땁 떫 뚤 뜨 띠
라 랍 랗 래 럽 렇 렵 롭 르 리 립 많 맑 맙 맞 맣 맵 먹 멀 멓 묵 물 묽 미 밉 밝 밭 벋
벗 볍 보 붉 비 쁘 살 석 설 섧 섭 세 쉽 습 시 싫 싶 싸 쌔 쌨 쓰 약 얇 얕 얗 없 엷 엽
엾 엿 옇 옳 이 익 있 작 잖 잘 잦 재 적 젊 좁 좋 죽 줍 지 질 짙 짜 짧 쩍 차 찮 춥
치 칠 크 탑 태 텁 툴 프 하 희 히(151)

Liste 2-18. Liste des SFR apparaissant dans les adjectifs.

Avec cette méthode, nous avons classé des SFR dans la liste en petits groupes avec l'information de PPP. Pour le cas du verbe, les SFR et l'information sur leur PPP sont décrites dans l'Annexe E, l'adjectif, dans l'Annexe F. Un extrait de la liste avec l'information sur PPP ont été donné dans le paragraphe 2.2.

2.3.5. Postpositions Prédicatives

L'information sur les formes fléchies des noms et les PPP des prédicats doit être indiquée dans le dictionnaire, car le dictionnaire contenant seulement les entrées canoniques ne permettrait pas de reconnaî tre les formes conjuguées telles que '*goes*', '*went*' et '*gone*' pour le verbe '*go*' en anglais.

Dans le cas du français, des informations plus complexes sont exigées. Un nom masculin peut avoir les formes de féminin, masculin pluriel, et féminin pluriel, comme *chanteur* (singer:m.sing), *chanteuse* (singer:f:sing), *chanteurs* (singer:m:pl), et *chanteuses* (singer:f:pl). Un adjectif peut aussi avoir plusieurs formes fléchies. Un verbe apparaî t rarement sous sa forme canonique. Il est plus souvent conjugué, par exemple, *vais*, *vas*, *va*, *allons*, *allez*, *vont*, *allais* et *allait* pour le verbe '*aller*'. Le dictionnaire doit donc fournir l'information sur toutes les formes conjuguées.

Examinons maintenant le cas du coréen. Il n'existe ni de genre pour le nom ni de forme conjuguée pour le verbe. Par contre, il existe des marques de fonction grammaticale, comme les marques de *temps,* de *modalité,* d'*aspect*, de *politesse* et de mode de la phrase. L'ordre et les contraintes de combinaison sont extrêmement complexes. Les adjectifs en coréen doivent être suivis de PPP. Les suffixes indiquent également les fonctions grammaticales des adjectifs, alors qu'en français et en anglais, c'est une copule, comme *'être'* ou '*be*' ou des verbes équivalents, qui prennent les marqueurs indiquant les fonctions grammaticales des mots adjectivaux.

Il était évident qu'un analyseur morphologique du coréen ne peut reconnaî tre les formes fléchies des noms, des verbes et des adjectifs sans l'information associée. Par exemple, "갔다" (FFP) (est allé) est la forme du passé de "가다" (aller). L'entrée est composée du radical et des suffixes grammaticaux. Donc, il faut avoir le dictionnaire qui donne l'information sur les formes fléchies.

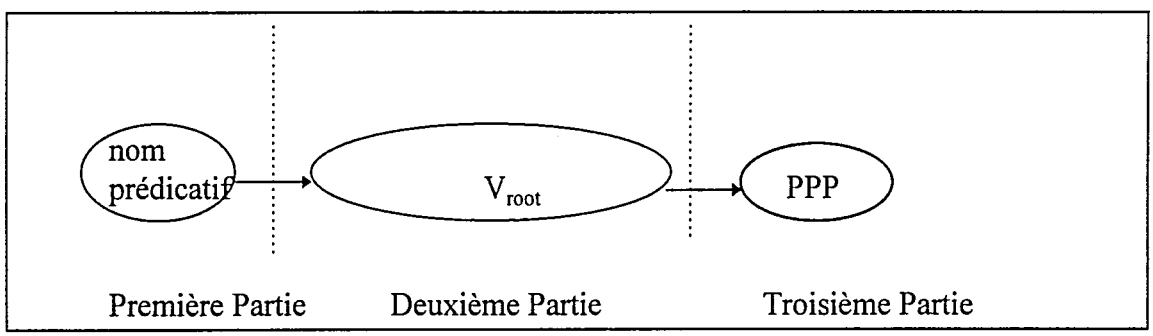


Figure 2-17. Forme de la structure T3.

FFP inclut toutes les formes combinées du radical avec PPP dans la figure 2-17. SFR occupe la fin de la deuxième partie. Pour construire FFP, nous construirons l'information de la connexion entre SFR et PPP, puisque la troisième partie dépend du SFR.

Nous avons classé les entrées PPP à l'aide de l'information de connexion avec le SFR dans la table 2-5.

Par exemple, les données dans la troisième colonne horizontale dans la table 2-5 sont classées comme type SU3-1. Toutes les entrées entre l'entrée "거나" (dans la colonne 'de') et l'entrée "건만은" (dans la colonne 'à') dans le dictionnaire de PPP selon l'ordre lexical sont classées comme types SU3-1. Pour comprendre les types des classes, examinons l'extrait de la liste des PPP de la figure 2-16.

de	à	Type	de	à	Type
ㄴ가	ㄴ지조차도	SU1-1	ㄹ것같으면	ㄹ텐데도	SU1-2
ㅁ세		SU1-3	ㅂ니까	ㅂ시다	SU1-4
거나	건만은	SU3-1	게	게조차도	SU3-2
겐들		SU3-3	겠거나	겠지요	SU3-4
고	고저	SU3-5	구나	구면	SU3-6
기	길레	SU3-7	나		SU3-8 PA1-1 SU1-5
나니	나니라	SU3-8	나마	나마나	PA1-1 SU1-5
나이다		SU3-8	냐		SU1-6
냐거나	냐하	SU3-9	냔	냔듯이	SU3-10
냘	냘텐데도	SU3-11	네	느라고	SU3-12
너라		PV1-2 /* V */	는	는지조차도	PA1-3 SU3-13 /* V */
니	니이까	SU1-6	다	다하	SU3-14
단	단듯이	SU3-15	달	달텐데도	SU3-16
대요		SU3-17	더구나	든지	SU3-17
들		PA1-4	듯이	디	SU3-18
라		PA1-5	라거나	란듯이	SU1-7
랄	랴	SU1-8	러	러조차도	PA7-1 SU1-9 /* V */
렀거나	렀지요	PA7-2	려	리오	SU1-9
마		SU1-10	만	만이라도	PA1-6
매	므로	SU1-11	사오니	삽고	SU3-19 /*V*/
서	서야말로	PA1-7	세		SU3-20
셔	셨지요	SU1-12	소		SU3-21
소서		SU1-13	쇼		SU1-14
습니까		SU4-1	습니까요		SU2-1
습니다		SU4-1	시거나	시지요	SU1-15
신	신지조차도	SU1-16	실	실텐데도	SU1-17
십니까	십시오	SU1-18	아	았지요	PA3-1
야	야만이	PA1-8	어	었지요	PA4-1
여	였지요	PA5-1	오		SU1-19

오니		SU3-22	와	왔지요	PA8-1
요		PA1-9	으나	우이	PA6-1
운	운지조차도	PA6-2	훌	울텐데도	PA6-3
움세		PA6-4	웁니까요	웁시다	PA6-5
워	웠지요	PA6-6	으나	으이	SU2-2
은	은지조차도	SU2-3	을	을텐데도	SU2-4
음세	웁니다	SU2-5	이		SU1-20
자	잔듯이	SU3-23	잘	잘텐데도	SU3-24
잡고		SU3-25 /* V */	지	지요	SU3-26

Table 2-5. Classification des PPP.

Entrée	**PDD**	**derivé de V/A**	**Type PPP**
거나	PPP	V et A	SU3 .Conj
는구료	PPP	V	SU3 .TmDec
서야말로	PPP	V et A	PV1 .Conj
시어요?	PPP	V et A	SU1 .TmInt

Figure 2-18. Extrait de la liste des PPP.

Les significations de 'SU1' et 'TmInt' de la dernière entrée "시어요?" dans la figure 2-18 sont les suivantes [30]:

1) L'entrée peut être combinée avec les prédicats qui finissent par une voyelle '*claire*'.
2) L'entrée peut être combinée avec les prédicats qui se terminent par une voyelle '*obscure*'.
3) L'entrée peut être combinée avec les prédicats qui se terminent par une voyelle '*obscure*' et par une consonne.
4) L'entrée peut être combinée avec les prédicats qui finissent par "ㅂ다".
5) L'entrée est le suffixe terminal au mode *interrogatif*.

FFP est construit facilement avec le dictionnaire PPP et l'information SFR (l'extrait de la liste est donné dans la figure 2-10), parce que PPP et SFR contiennent la même colonne verticale " **Type PPP**" qui connecte les deux.

3. Organisation du dictionnaire

Dans ce chapitre, nous discuterons des techniques de consultation du dictionnaire. Une méthode est celle de l'Automate Fini (AF) [9] [34] [38] qui est la technique appliquée au DELAS et au DELAF du LADL [4] [5] [6] [7] [11], et à Lexique-Compiler de XEROX [16]. L'autre technique est celle qui sera développée dans notre étude. Dans l'état actuel de cette technique, l'utilisation d'automates présente beaucoup d'avantages, alors que l'autre technique que nous appellerons la Consultation Basée sur la Position (CBP) est très compétitive. La méthode CBP a été programmée sur une machine **HP**.

Le technique d'automates ressemble beaucoup à celle des Trie structures [1] [2] [20]. La complexité du temps des deux systèmes est du même ordre. La vitesse de consultation dans les deux structures dépend uniquement de la longueur de la séquence, pas du volume du dictionnaire. L'automate a l'avantage d'occuper moins de place en mémoire que dans la méthode Trie structure [12] [13].

Le Trie représente les entrées du dictionnaire 'ab', 'abc', 'ade', et 'afg' comme dans la figure 3-1-1. Dans ce cas, l'automate fini représente les entrées comme dans la figure 3-1-2. Il y a aussi un autre exemple pour le cas des séquences 'abc' et 'efc'. Le Trie structure est décrit dans la figure 3-2-1, mais l'automate est représenté dans la figure 3-2-2. La différence est nette. Le Trie a besoin d'un plus grand nombre de noeuds que l'automate fini.

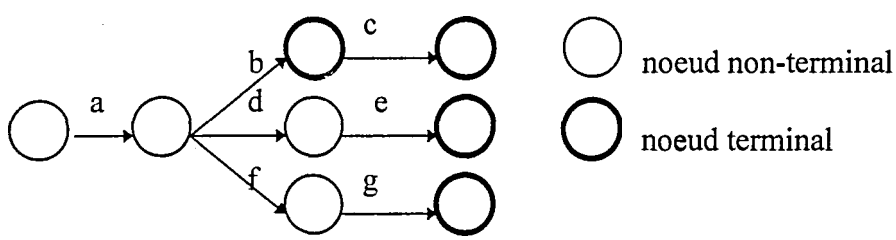


Figure 3-1-1. Diagramme de transition dans Trie structure.

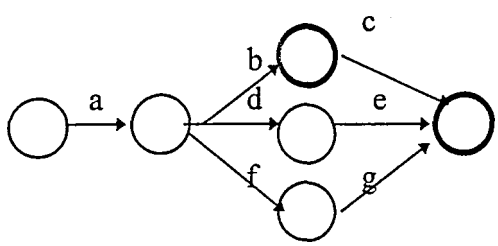


Figure 3-1-2. Diagramme de transition dans Automate Fini.

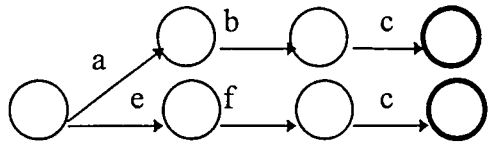


Figure 3-2-1. Représentation par Trie structure pour les séquences 'abc' et 'efc'.

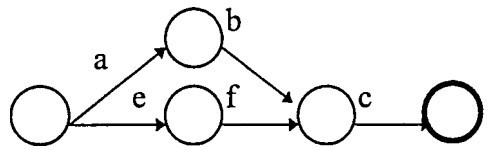


Figure 3-2-2. Représentation par Automate pour les séquences 'abc' et 'efc'.

Dans AF, le caractère 'c' commun en fin de séquence dans la figure 3-2-2 est compressé à la différence de Trie dans la figure 3-2-1. L'Automate permet de compresser les caractères identiques des séquences dans les deux directions. L'Automate est donc plus efficace que Trie structure du point de vue de l'espace mémoire occupé.

3.1. Plan de la spécification de l'Automate

3.1.1. Définition des fonctions

char * **input** /* séquence d'entrée */ [V]

_nd *start_node* [F]
T or **F**

S'il y a les noeuds qui transfèrent un noeud à un autre noeud identique sous les entrées différentes, il y a alors un **nd** (noeud duplicatif) entre les noeuds.

_flm **input** [F]
0 ou *status* et *string*

la tête longue cherche les fils de l'**input** à partir du noeud de départ courant (0) dans le diagramme de transition et retourne au noeud assorti le plus long et au fils assorti. Si il n'y a pas de fils assorti avec l'**input**, il retourne à **0**.

_llm **input** [F]
0 *or status* and *string*

L'ensemble à fin longue cherche les fils de l'**input** à partir du noeud terminal dans le diagramme de transition en ordre inverse et retourne au noeud assorti le plus long et au fils assorti. Si il n'y a pas de **nd** entre les noeuds assortis, il retourne 0.

_sep *start_node end_node* [F]
Automaton

S'il y a un **nd** entre le noeud de départ et le noeud final, la fonction **_sep** sépare les noeuds. Si le noeud final est le noeud terminal, il est redéfini comme le noeud de tête du noeud final et exécute à nouveau **_sep**.

_create *first_node last_node string* [F]
NULL

_create crée le nouveau noeud pour l'input du fils entre le premier noeud et le dernière noeud.

_main *input_strings* [F]
void

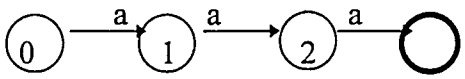
_main construit une structure de l'automate pour l'input des fils. La structure est l'automate fini déterminatif (DAEF).

```
while(_gets(input, file_pointer) != NULL) {
        if(_flm(input)!= 0)    _sep();
        _llm(input);
        if(input != NULL)  _create();
}
```

3.1.2. Exemples de procédures

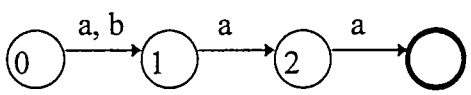
/* **input** : NULL, current_nodes : start →0, terminal →t */

Figure 3-3. L'état initial.



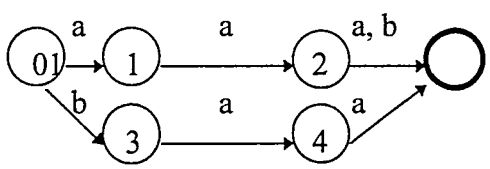
Procédure : /* **input** : "aaa", current_nodes : 0, t */
_flm(input) 0
_llm(input) 0
_create(current_nodes, **input**)

Figure 3-4. Diagramme de transition après l'insertion de la séquence 'aaa'.



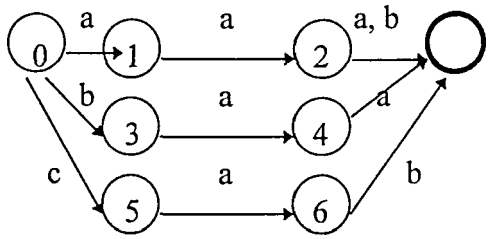
Procédure : /* **input** : "baa", current_nodes : 0, t */
_flm(input) 0
_llm(input) 1, "aa" /* **input** : "b", current_nodes : 0, 1 */
_create(current_nodes, **input**)

Figure 3-5. Diagramme de transition après l'insertion de la séquence 'baa'.



Procédure : /* **input** : "aab", current_nodes : 0, t */
_flm(input) 2, "aa" /* **input** : "b", current_nodes : 2, t */
_sep(0, 2);
_llm(input) 0
_create(current_nodes, **input**)

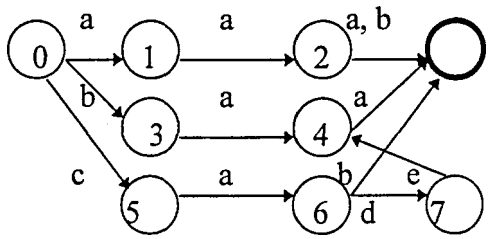
Figure 3-6. Diagramme de transition après l'insertion de la séquence 'aab'.



Procédure : /* **input** : "cab", current_nodes : 0, t */
_flm(input) 0 /* **input** : "cab", current_nodes : 0, t */

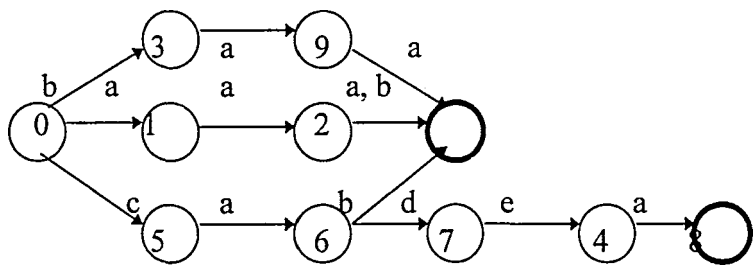
_llm(input) 0
_create(current_nodes, **input**)

Figure 3-7. Diagramme de transition après l'insertion de la séquence "cab".



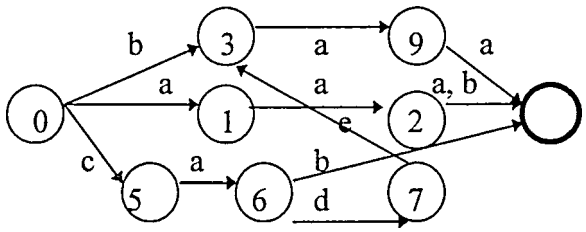
Procédure : /* **input** : "cadea", current_nodes : 0, t */
_flm(input) 6, "ca" /* **input** : "dea", current_nodes : 6, t */
_sep (0,6)
_llm(input) 4, "a" /* **input** : "a" : current_nodes : 6, 4 */
_create(current_nodes, **input**)

Figure 3-8. Diagramme de transition après l'insertion de la séquence "cadea".



Procédure : /* **input** : "cadeaa", current_nodes : 0, t */
_flm(input) 8, "cadea" /* **input** : "a", current_nodes : 8, t */
_sep(0,8)

Figure 3-9-1. Diagramme de transition après l'insertion de la séquence "cadeaa".



_llm(input) 3, "aa" /* **input** : NULL, current_nodes : 4, 3 */
_create(current_nodes, **input**)

Figure 3-9-2. Diagramme de transition après l'insertion de la séquence "cadeaa".

3.1.3. Algorithme pour l'automate fini

Pour construire la structure de l'automate, il est nécessaire de donner certaines définitions et présenter les algorithmes utilisés [13]. Ce sont des algorithmes pour la construction de l'Automate Lexical (qui sera défini ultérieurement). L'optimisation de la structure n'y est pas prise en compte. Ici nous ne discuterons que des types d'opérations nécessaires, c'est-à-dire, nous ne implémentons pas l'Automate lexical, mais nous ne donnons que des spécifications nécessaires pour l'Automate lexical.

Pré-hypothèse

Cet automate sera construit sur la base des hypothèses suivantes:

- Les entrées de l'input sont triées
- Durant l 'exécution de la fonction **_sep**, il y a au plus un **nd** dans l'automate.

L'Automate Lexical est l'automate de la description du lexique. Il comporte les 5 structures suivantes.

L'Automate Lexical = (**I**, **Q**, **P**, $\mathbf{Q_0}$, $\mathbf{Q_t}$)

I = série des caractères anglais {a, b, c, ..., z}
Q = série des états, par exemple { 0, 1, 2, ... }
$\mathbf{P(Q_i, I) \rightarrow Q_{i+1}}$
$\mathbf{Q_0}$ = état de départ
$\mathbf{Q_t}$ = état terminal.

```
char*  input;          /* input string,  x ∈ input, x ∈ I */        [V]
int    position = 0;   /* current position in input */              [V]
int    x, y;           /* current start node and terminal node */
typedef                int    status;
int    table[STATUS][ALPHABET]                                      [V]
status top;            /* current_top positon of the table */
```

```
_p(state, pos)
status state, char pos;
{      return table[state][pos - 'a'];
}
```

```
_rp(state, pos)
status state, char pos;
{      int i;
       for (i = 0; i< STATUS; i++)
              if(table[i][pos- 'a'] == state) return i;
       return 0;
}
```

```
_nd(Qi)
staus Qi
{       status  next; static int i;
        for (next = Qi, until next < position, step next = _p(Qi, input[i++]), do
                { if  ∃a, b, (_p(Qn, a)  ==  next) && ( _p(Qn,b) == next),  if a ≠ b ∈ I,
                        then  return TRUE; }
        return FALSE
}

_flm() {
status  current, temp;
int     i, j;
        if(*input == NULL) return 0;
        current← temp ← Q0;
        j = _strlen(input); i = position;
        while((i < j) && (current ← _p(current, input[i++])) != 0 &&
                        (current != Qt)) temp ← current;
        if(current == Q0) return 0;
        x ←temp; position = --I; strcpy(input, subcat(input, 0, position));
        if( x == Qt)
                x ← table[temp][input[--i] - 'a') ←
                        _create_terminal(temp, input[i]); /* undefined_function */
        /* _sep(Qs, temp); */
}       /* end of the function _flm */

_llm() {
status  current, temp;
int     i, j;
        i = position;
        if(input == NULL) return 0;
        current ← temp ← x;
        j = _strlen(input); j--;
        while((i < j) && (current ← _rp(current, input[j--])) != 0 &&
                        (current != Qs)) temp ← current;
        if(current == Qt) return 0;
        y ←temp; position = ++j;
        if(y == Qs) return 0;
        if(!_dup(current, y)) {
                y ← current;  _set_input(input, position, j);
        }
}       /* end of the function _llm */

_sep(start, end)
status  start, end;
{
status  current;
```

```
current = start;
while(current != end) {
        s'il y a un noeud courant à partir de deux autres noeuds
                then _create(start, current);
}

_create(start, end)
status start, end;
{
int i, j;
if(input == NULL) return 0;
j = strlen((input);
for(i = 1; i < j; i++) {
        table[start][input[i] - 'a'] = top;
        start = top++;
        }
table[start][input[i] - 'a'] = end;
}
```

Figure 3-10. Pseudo algorithmes pour l'automate lexical.

3.2. Consultation Basée sur la Position

Pour une consultation efficace du dictionnaire, la structure optimisée du lexique est cruciale. CBP est basée sur la localisation du caractère dans le lexique. Les procédures de consultation ne sont pas basées sur les caractères mais sur les numéros des positions de ces caractères. Bien que cela nécessite plus d'états que AF, le temps et l'espace nécessaires sont aussi efficaces qu'avec AF.

L'automate a beaucoup de cellules vides [35] et pour réduire ces cellules vides, il exige la technique [35] [39]. Mais CBP n'a pas besoin de procédure de réduction comme dans le cas d'AF.

CBP n'a pas non plus besoin de re-construction comme dans la table de transition avec la méthode AF, bien que ce soit un travail **off-line(batch)** [35] [39]. Par exemple, l'input "aab" dans la figure 3-6 détruit la structure de la figure 3-5 et re-construit un nouveau diagramme comme dans la figure 3-6.

CBP est constituée seulement de deux tables; que l'on appelle Table de l'Index (TI) et Table de colonne Horizontalement (TH), avec des procédures liées.

Pour le dictionnaire de 'on-line', nous pouvons supposer que toutes les entrées sont classées, et que certains caractères identiques apparaissent dans les mêmes positions dans les entrées adjacentes. Cela peut nous indiquer qu'il est possible que l'opération oriente la position. La méthode supprime les mêmes caractères lorsqu'ils apparaissent sur la même position dans des entrées adjacents.

Elle est plus efficace que Trie structure. Dans le cas des inputs 'abc' et 'efc' dans la table 3-1 la structure est identique à AF. Autrement dit, CBP a une capacité de mémoire plus efficace que Trie, bien que CBP comprenne une description en table, mais la forme de la représentation est différente de AF. AF utilise la forme de la table 3-1. Chaque cellule de cette structure correspond à une transition de AF, ce qui implique que cette méthode est plus efficace que Trie structure. Chaque cellule de la table correspond aux noeuds de AF ou de Trie.

	Première position	Deuxième pos.	Troisième pos.
Première entrée: abc	a	b	c
Deuxième entrée: efc	e	f	

Table 3-1. Représentation de CBP pour les inputs "abc" et "efc".

Pour d'autres inputs comme "abcd" et "aded", cette méthode les représente comme la table 3-2.

	Première	Deuxième	Troisième	Quatrième
Première entrée: abcd	a	b	c	d
Deuxième entrée: aded		d	e	

Table 3-2. Représentation de CBP pour les inputs “abc” et “efc”.

3.2.1. Procédures de la construction CBP

La première étape de cette méthode est très simple. Si des caractères identiques apparaissent dans la même position que l'entrée précédente, le caractère spécial '#' s'y substitue. Grâce au caractère spécial, la position est réduite.

N° d'entrée	Entrée originale	1ère transformation
0	aaron	aaron
1	aas	##s
2	aback	#back
3	abaculus	####ulus
4	abaddon	###ddon
5	abandon	###n###
6	abandoned	#######ed
7	abandonee	########e
8	abase	###se
9	abash	####h
10	abask	####k
11	abate	###te
12	eagle	eagl#
13	eaglet	#####t
14	ear	##r
15	eardrop	###drop
16	earl	###l
17	early	####y
18	earmark	###mark
19	earn	###n

Table 3-3. Exemples d'entrées de base et après la première transformation.

Si l'on examine les positions des '#' en lisant ligne par ligne, on peut se rendre compte que les positions marquées sont distribuées relativement au hasard, mais plus souvent à gauche qu'à droite. Mais si on lit colonne par colonne, on peut comprendre facilement l'idée de la compression dans cette méthode. La compression du caractère dans les entrées dépend de la présence du caractère # dans la même position que l'entrée précédente.

Les numéros dans la première colonne de la table 3-3 correspondent aux numéros des entrées originales. Pour la représentation numérique des positions marquées, nous devons passer de cette forme à la forme illustrée par dans la table 3-4.

La dernière ligne contient les numéros des entrées qui finissent par la position correspondance à la colonne. Par exemple, la longueur de l'entrée N° 2, 'aback' correspond à 4 (la longueur est comptée à partir de 0). Cette information est inscrite dans la cellule à l'intersection de la colonne numéro 4 (cinquième colonne) et de la

dernière ligne, comme pour l'information de l'état dans le diagramme de transition (par exemple, le noeud gras dans la figure 3-1). Cette information est gardée en mémoire pour être utilisée dans la fonction **finish_or_not** lors de la consultation du pseudo-code dans la figure 3-11.

	0	1	2	3	4	5	6	7	8
a	0, 11	0, 1 12, 19	2, 11		18, 18				
b		2, 11							
c				2, 3					
d				4, 4 15, 15	4, 7				6, 6
e	12, 19				8, 8 11, 13			6, 7	7, 7
f									
g			12, 13						
h					9,9				
i									
j									
k					2, 2 10, 10		18, 18		
l				12, 13 16, 17		3, 3			
m				18, 18					
n				5, 7 19, 19	0, 0		4, 7		
o				0, 0		4,7 15, 15			
p							15, 15		
q									
r			0, 0 14, 19		15, 15	18, 18			
s			1, 1	8, 10				3,3	
t				11, 11		13, 13			
u					3, 3		3, 3		
v									
w									
x									
y					17, 17				
z									
			1, 14	16, 19	0, 2, 8, 9, 10, 11, 12, 17	13	4, 5, 15,18	3	6, 7

Table 3-4. Deuxième forme de la table dans CBP.

Les paires de numéros dans les cellules représentent la première et la dernière d'une serie d'entrées adjacentes comportant le même caractère dans la même colonne. Par exemple, les deux paires (5,7) et (19, 19) dans la table 3-4 [n] [3] signifient que toutes les entrées entre la 5^{e} et la 7^{e} lignes, qui sont 'abandon', 'abandoned', et 'abandonee' comportent le même caractère 'n' dans la colonne n^{o} 3. Mais les deux

entrées adjacentes (18^{e} et 20^{e} entrées) de la 19^{e} entrée 'earn' n'ont pas de caractère 'n' dans la position n^{o} 3.

La taille des cellules de la table 3-4 varie et il y a des cellules vides. La forme n'est pas efficace pour la gestion de la mémoire. Elle peut être remplacée par les tables 3-5 et 3-6. La table 3-5 est appelée TI. La taille maximum de TI est fixée. C'est 26 (nombre de caractères de l'alphabet) * 9 (longueur maximum des entrées). Dans la table 3-3, 'abandoned' et 'abandonee' ont la longueur maximum, soit 9.

	0	1	2	3	4	5	6	7	8
a	1	4	6		22				
b		5							
c				11					
d				13	23				44
e	2				25			42	45
f									
g			7						
h					26				
i									
j									
k					28		38		
l				15		33			
m				16					
n				18	29		39		
o				19		35			
p							40		
q									
r			9		30	36			
s			10	20				43	
t				21		37			
u					31		41		
v									
w									
x									
y					32				
z									

Table 3-5. TI dans CBP.

Chaque numéro de la table 3-5 est le total des nombres accumulés des paires qui sont données dans la table 3-4 dans l'ordre colone par colonne. Par exemple, le numéro 4 dans TI [a][1] dans la table 3-5 est le nombre cumulé des données (0,11) (la paire dans la table 3-4 [a][0]), (12,19) (la paire dans la table 3-4 [e][0]), (0,1) (table 3-4 [a][1] et

(12,19) (table 3-4 [a][1]). La table 3-6 montre seulement la liste des séquences de la table 3-4 dans l'ordre ligne par ligne. L'information de la table 3-5 est connectée avec celle de la table 3-6. Pour consulter une entrée, les deux tables sont nécessaires. L'information donnée par TI est l'indice de la table 3-6, appelée TH. TH donne uniquement la représentation des séquences de la table 3-4 avec le concept qu'une cellule de TH est une paire de la table 3-4.

Cette méthode a été testée sur une machine **HP**. Pour l'implanter sur un **PC**, TI doit être en mémoire, la taille de TI n'est pas grande. Dans cet exemple, la taille de TI est 234 (= 26 * 9, la longueur maximum des séquences dans cet exemple d'entrées). Mais TH peut exister dans le mécanisme externe du système de fichier. Nous espérons que la méthode CBP figurera dans l'environnement **DOS** dans l'avenir.

0, 11	12, 19	0, 1	12, 19	2, 11	2, 11	12, 13	0, 0	14, 19
1, 1	2, 3	4, 4	15, 15	12, 13	16, 17	18, 18	5, 7	19, 19
0, 0	8, 10	11, 11	18, 18	4, 7	8, 8	11, 13	9, 9	2, 2
10, 10	0, 0	15, 15	3, 3	17, 17	3, 3	4, 7	15, 15	18, 18
13, 13	18, 18	4, 7	15, 15	3, 3	6, 7	3, 3	6, 6	7, 7

Table 3-6. TH.

```
procedure look_up(entry)
begin
int start, end, from, to, i;
start=0;
end= total_ number_ of_ the_ entries;
for (i = 0; i < strlen(entry); i++) {
        if(i == 0) from = 0; else
        from =          TI[entry[i-1] -'a'][i]; /* TI dans la table 3-5 */
        to =    TI[entry[i  ] -'a'][i];
        if(!intersection(TH, start, end, from, to))
/*      re-adjust the values start and to,
        start = first value of ( interval(start, end) ∩ interval(from, to)),
        end   = last value of ( interval(start, end)  ∩ interval(from, to)) */
        return False;
}
return finish_or_not(start, end);
end.
```

Figure 3-11. Pseudo-algorithme pour CBP.

3.2.2. Evaluation de la performance

Nous avons testé l'espace et le temps nécessaires pour la consultation avec l'exemple du dictionnaire des noms acquis à partir de **WordNet** [22]. L'exemple du dictionnaire possède 20,329 entrées qui ne représentent pas le nombre total des noms. La longueur maximum des entrées est 23. Après exécution du programme, nous avons obtenu 72,345 paires. L'espace nécessaire dans la méthode CBP est estimé ainsi:

Espace nécessaire	=	(nombre d'entrées + taille de TI
		+ taille de TH) * taille d'un entier
	=	(20,329 + 26 * 23
		+ 2 (taille d'une cellule dans TH) * 72,345)
		* 4 octets (taille du type **int** en langage C).
	=	662,468 octets
	=	soit 662 KO.

L'espace total nécessaire est d'environ 700 KO. La taille de la table TI n'est pas toujours grande et elle est de 2 402 octets (= 26 * 23 * 4 octets), parce qu'il n'y a qu'un facteur qui affecte l'espace nécessaire dans CBP. C'est la longueur maximum des entrées dans le lexique. Dans cet exemple, elle est de 23.
La taille TH n'a pas de cellule vide. Elle ne nécessite pas de technique de compression de la table, comme pour l'Automate lexical [35].
Pour programmer la technique de CBP sur **DOS**, on a besoin de deux niveaux de mémoire : TI est chargé au niveau mémoire (par exemple, commande '*malloc*'), et TH est au niveau fichier (par exemple, commande '*fopen*'). En effet, **DOS** ne fournit pas beaucoup de mémoire (il n'a pas de mémoire virtuelle).

Calculons le temps nécessaire avec CBP. Comme avec l'Automate lexical, le temps de consultation dans CBP dépend de la longueur des séquences saisies, mais pas du volume du dictionnaire. Pour mesurer la vitesse de consultation dans CBP, la fonction '*ftime*' a été utilisée sur **HP9000/735**. La mesure est au 1/1000 de seconde près. Nous l'avons testée à plusieurs reprises pour obtenir le temps moyen. La méthode CBP permet de rechercher 18 000 entrées par seconde.

CBP se caractérise par la méthode de la position orientée. Même si elle a plus de paires que l'automate de transition, elle ne nécessite ni la table des procédure de reconstruction, ni la table de compression. Quand un nouvel input est entré dans la méthode de l'automate, tous les états de la transition doivent être reconstruits[8]. Cela demande beaucoup de temps.

Dans l'état actuel de nos recherches, nous ne pouvons pas dire quelle est la méthode la plus efficace. Les deux méthodes possèdent la même complexité de temps pour la consultation. La complexité de mémoire dans CBP est très compétitive.

4. Etiquetage des parties du discours basé sur le lexique

EPL est le marqueur de parties du discours(PDD) qui analyse les phrases entrées par le biais du Lexique-Grammaire [12]. Celui-ci utilise systématiquement cette approche a l'analyser.. EPL, basé sur le système DECO [25], ne couvre pas toutes les PDD de la langue coréenne, car DECO n'est pas un système de dictionnaires complets. L'Analyseur morphologique, EPL, analysera les mots en entrée, s'ils comprennent les types de nom (T1', T2) ou les types de prédicats (T3). D'autres cas, comme (T4), n'ont pas été considérés dans l'analyseur morphologique EPL.

EPL se compose de modules : analyseur lexical, T1', T2, T3 et BDD filtrage. La figure 4-1 montre le déroulement du contrôle.

L'analyseur lexical examine les erreurs syntaxiques et les paires de phrases mal combinées. Il envoie les tokens aux modules T1', T2 et T3.
Ces modules examinent les formes de noms et de prédicats, puis envoient les résultats au module BDD filtrage. Celui-ci gère des règles spéciales comme les règles du *cas vocatif* de SF et du *cas de position* de SF, aussi que l'information BDD. Le mot analysé est finalement retourné avec des informations.

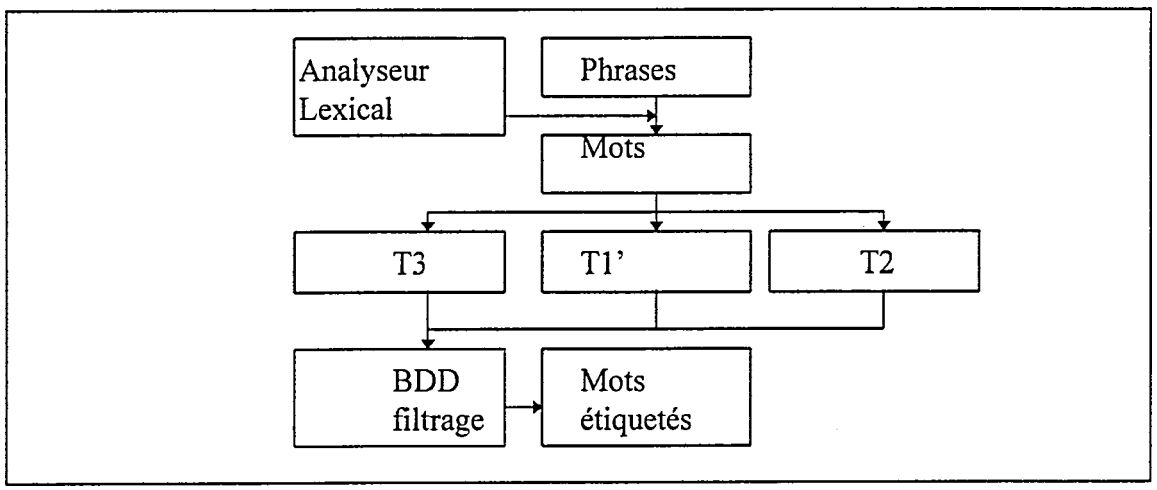


Figure 4-1. Les composantes du système **EPL** avec le déroulement du contrôle.

4.1. Analyseur lexical

L'analyseur lexical analyse les phrases entrées et les renvoie sous forme de tokens. Le système vérifie aussi les fautes et les erreurs syntaxiques. Il connaît les conditions de fin de phrase. Il y a trois symboles pour marquer la fin d'une phrase : *le point final* (.), *le point d'interrogation* (?), et *le point d'exclamation* (!). Cependant, il n'est pas simple de segmenter un texte en phrases, parce que les symboles spéciaux comme *les guillemets*, *les parenthèses*, *le point final*, *le point d'interrogation*, et *le point d'exclamation* peuvent avoir une incidence sur ce découpage.

Par exemple, certains symboles s'utilisent normalement par paires dans des phrases. Nous les appelons Symboles Pairs. L'état de Symboles Pairs est exprimé par OPEN ou CLOSE comme dans la table 4-1. Durant l'analyse des phrases, si l'analyseur lexical rencontre un nombre impair de symboles, l'état est OPEN, sinon CLOSE. Autrement dit, si l'on rencontre un symbole de fin de phrase à l'état OPEN, la phrase est considérée comme non achevée.

Il existe une exception dans les cas de ` (OPE_Q) et ' (CLO_Q) dans la table 4-1. Par exemple, dans une séquence comme *l'expression*, les Symboles Pairs (OPE_Q et CLO_Q) ne sont pas utilisés deux fois. Dans ce cas, si le programme rencontre un symbole de fin de phrase, il est considéré comme terminé, même s'il est OPEN.

	OPEN	NAMING	CLOSE	NAMING
cas 1	(	LOW_O	)	LOW_C
cas 2	[	MID_O	]	MID_C
cas 3	{	HIG_O	}	HIG_C
cas 4	<	COM_O	>	COM_C
cas 5	“	DOU_Q	”	DOU_Q
cas 6	`	OPE_Q	’	CLO_Q
cas 7	‘	CLO_Q	’	CLO_Q

Table 4-1. Les symboles pairs.

4.1.1. Exemples de test

Le programme **snt_chk** est l'analyseur lexical typique. Il vérifie des erreurs syntaxiques et l'agencement des phrases entrées, il reconnaî t alors les phrases bien formées.

Les figures suivantes illustrent des exemples de phrases où on a marqué des informations comme '*%FROM, %FILE NAME, ...*' dans la figure 4-2 et '*#TITL*' dans la figure 4-3. Elles indiquent la formule du corpus. Par exemple, la ligne de '*%FROM*' dans la figure 4-2 décrit un texte extrait du journal '한계레'(*HanGurLe*). La ligne '*#TITL*' dans la figure 4-3 montre que le titre du texte est 'Science et Médecine'. Les lignes marquées sont ignorées dans le programme **snt_chk**. Elles s'utiliseront dans d'autres logiciels, comme la recherche d'information'.

Le programme **snt_chk** a analysé et examiné les phrases dans la figure 4-2 et la figure 4-3. La figure 4-4 montre le résultat.

Figure 4-2. Un type de corpus

Les lignes 2 et 4 de la figure 4-3 sont arrangées comme la ligne 2 de la figure 4-4. Les lignes 4 et 8 de la figure 4-3 sont arrangées comme la ligne 3 de la figure 4-4. Les lignes de 9 à 13 de la figure 4-3 sont ajustées comme la ligne 4 de la figure 4-4. Bien qu'il y ait deux *points finals* pour la phrase présentée dans les lignes 9-13, le premier *point final* qui est dans *les guillemets* n'est pas le signal de la fin d'une phrase.

La ligne 14 de la figure 4-3 comporte CLO_Q dans la ligne. Le guillemet est traité comme une marque spéciale dans cette phrase. La ligne 14 de la figure 4-3 est réajustée comme les lignes 5 et 6 de la figure 4-4.

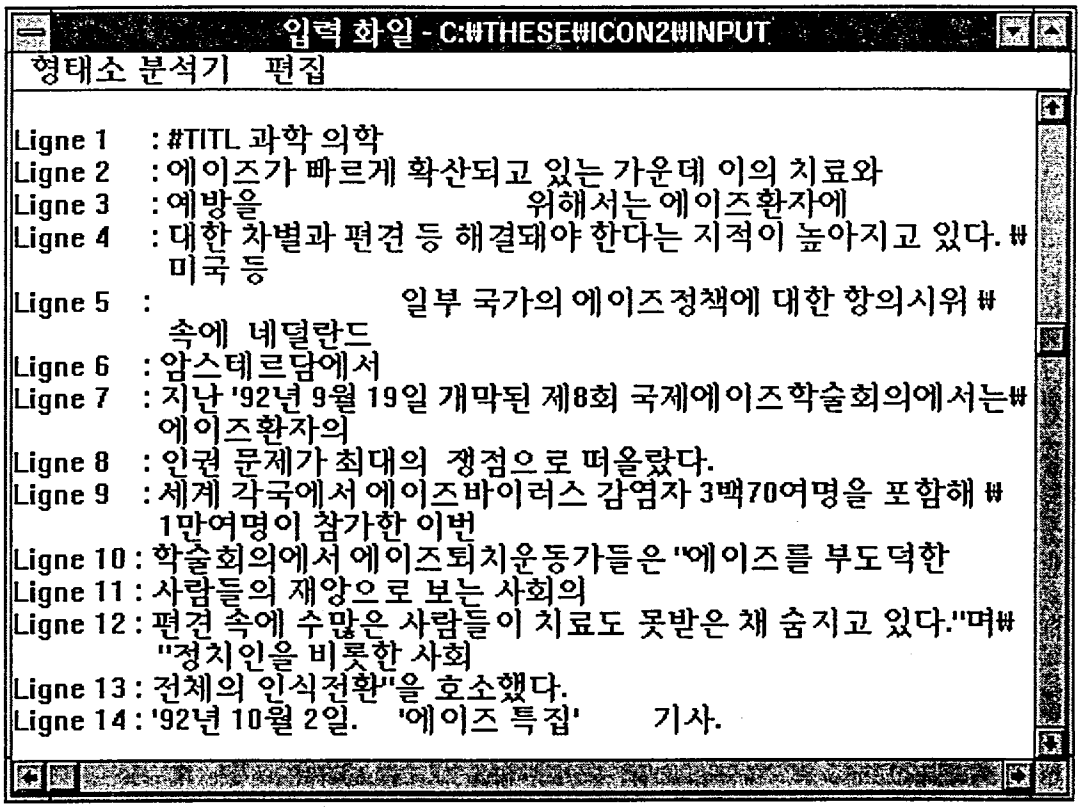


Figure 4-3. Suite de la figure 4-2.

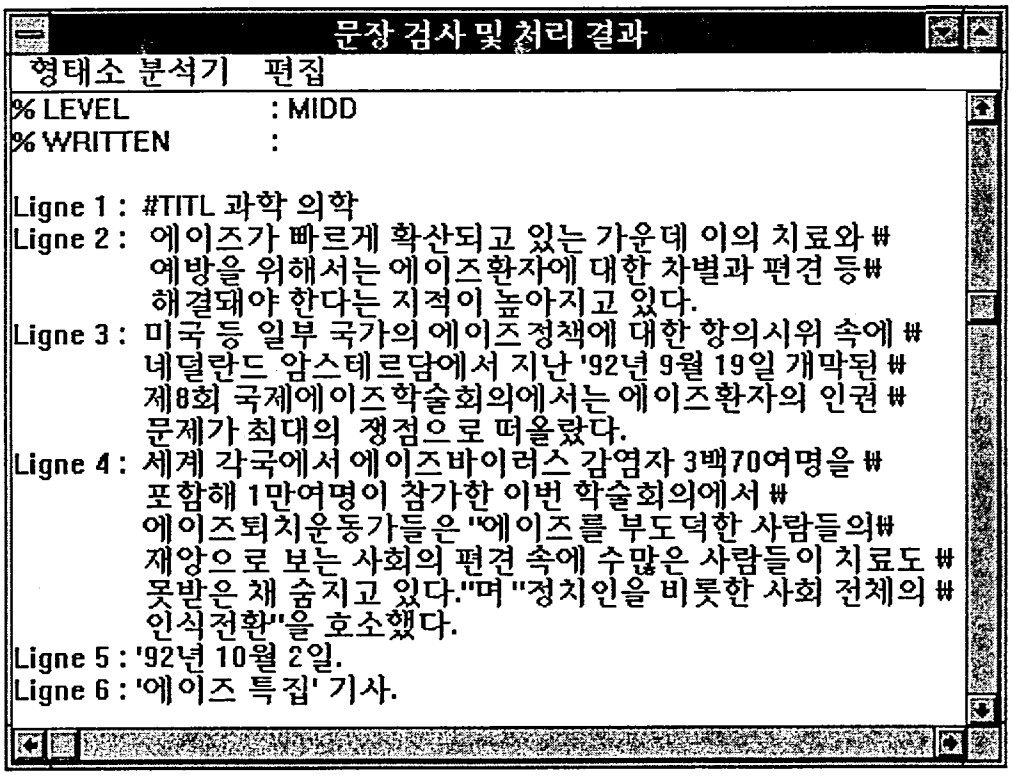


Figure 4-4. Sortie de l'analyseur lexical de la phrase d'entrée de la figure 4-2 et la figure 4-3.

snt_chk met en ordre toutes les phrases en entrée selon le principe " *une phrase sur une ligne*". Ce principe est utilisé dans l'étape d'analyse morphologique appelée **EPL** dans ce travail.

4.2. Analyseur morphologique

Dans la structure T1', NS peut être situé dans n'importe quelle position du mot. Ainsi, pour trouver les candidats de NS dans le mot, nous n'utilisons pas la recherche exacte, mais une méthode de recherche des sous-ensembles du mot d'entrée.

Supposons, par exemple, que l'entrée est "abc". "a", "b", "c", "ab", "bc" et "abc" sont tous des sous-ensembles de l'entrée. Cela suggère de consulter le dictionnaire dans l'ordre tête vers fin. La méthode de la recherche du sous-ensemble est décrit dans la figure 4-5.

Dans la figure 4-5, la fonction *str_same* guide la recherche dans une position au hasard. Dans les arguments de la fonction, *instr* est un mot en entrée, *str* est un candidat NS, et *i* est un argument de positionnement dans *instr*. Si la recherche entre *instr* et *str* est déclenché (dans ce cas, la fonction *str_same* retourne 1), le candidat de NS(*str*) est enregistré dans les structures *sn->start* et *sn->end* avec les marques de positions *i*. Avec la variable *i*, la fonction *str_same* est appliqué à plusieurs reprises. Par conséquent, tous les candidats NS sont enregistrés dans les structures *sn->start* et *sn->end*. Chaque candidat NS est vérifié avec PF, SF et PPN, puisque le mot est composé de NS, PF, SF et PPN dans la structure T1'.

```
initializing();
j = k = 0;
while(fscanf(dict, "%s", str) != EOF) {        /* dict est NS */
        leng= strlen(str);
        i = 0;
        while ( i < leng) {
                if(str_same(instr, str, i) == 1)  {
                        sn->start[k][j] = i;
                        sn->end[k++][j] = i + leng;
                }
                i = i + 2; /* 2 byte */
        }
        break;
}
```

Figure 4-5. Fonction des sous-ensembles.

Dans les structures T1', T2 et T3, tous les candidats entrées de dictionnaires sont gardés temporairement dans la structure *cand* de la figure 4-6. Dans cette structure, les résultats sont analysés.

Dans T1', NS est la PDD principale dans le mot. Cela veut dire que l'on cherche tout d'abord NS dans le mot, puis les autres particules avec l'information de NS. FN et FFP sont les PDD principaux, respectivement dans T2 et T3. Dans la structure *cand*, la PDD principale est située dans les positions *start*[*k*][0], et *end*[*k*][0]. La variable *k* de

chaque structure dénote les possibilités d'interprétation du mot en entrée. Les autres particules, tels que PF, SF et PPN dans le cas de T1', peuvent présenter des ambiguï tés avec chaque candidat du NS. L'information est gardée dans la position *j* de la figure 4-5. Le détail est représenté dans la figure 4-6.

```
struct cand {
int     tota;                   /* total des candidats */
int     leng;                   /* taille de l'entrée */
char    data[WORD];             /* l'entrée */
int     start[NUM][CAND];       /* les premières positions des candidats */
int     end[NUM][CAND];         /* les dernières positions des candidats */
} *sn, *sf, *npp, *is, *fsr, *v, *a, *nis, *np;

struct cant {
int     vert;                   /* total des candidats */
int     hori;                   /* indicateur de la position courante */
char    data[WORD];             /* l'entrée */
int     flag[WORD];             /* le nombre de l'ambiguïté */
int     type[WORD][WORD];       /* types de PD */
int     loca[WORD][WORD];       /* les positions de la type */
} *result;
```

Figure 4-6. Structure des données de EPL.

Nous allons examiner la procédure d'étiquetage sur un exemple. D'abord, nous avons trouvé tous les candidats possibles pour le NS à partir du mot, concept simple. Le concept n'autorise pas la répétition interne **R1** dans le dictionnaire. Si le mot est "정치가는" (le politicien est) (NS), les candidats du NS sont "정치" (la politique) (NS), " 가 " (NS), et "정치가" (NS-NS). Pour les candidats du NS, la structure *cand* comporte l'information suivante :

sn->tota	= 3, (le nombre de *k*)	
sn->leng	= 8	
	(parce que nous utilisons 2 bytes du système de code),	
sn->data	= "정치가는",	
sn->start[0][0]	= 0	/* k = 1 정치 */
sn->end[0][0]	= 4	
sn->start[1][0]	= 4	/* k = 2 가- */
sn->end[1][0]	= 6	
sn->start[2][0]	= 0	/* k = 3 정치가- */
sn->end[2][0]	= 6	

1) *k* = 1 cas:

Le premier candidat du NS "정치" est situé dans (0, 4) de la *data* (0, 8). (0, 4) spécifie la forme de la quatrième position à partir de zéro de *sn->data*. La taille de la donnée en entrée de *sn->data* est (0, 8).

Dans une autre position comme (4, 8), on peut trouver SF ou PPN. PF est situé à la tête du NS, mais le candidat NS occupe la première place du mot. Donc, il n'y a pas de PF.

SF diffère de PPN par le fait que SF est toujours en tête du PPN. Cela signifie que les candidats SF, "가는"(4, 8) et "가"(4, 6) sont possibles, mais pas "는"(6, 8). Si SF est utilisé, il doit commencer en position 4. D'ailleurs, "가는"(4, 8) et "는"(6, 8) peuvent être les candidats PPN, qui doit se terminer à la position 8. Les autres formes contreviennent à ce principe dans la structure T1'.

Dans le dictionnaire de SF, on a les entrées "가"(4, 6) et "는"(6, 8). Dans SF, la valeur de *j* est 2 et cela signifie qu'il y a deux interprétations possibles. L'information est enregistrée provisoirement dans la structure *cand*, ainsi :

sf->start[0][0] = 4
sf->end[0][0] = 6
sf->start[0][1] = 6
sf->end[0][1] = 8.

Dans le dictionnaire PPN, les entrées "가"(4, 6) et "는"(6, 8) existent aussi. Dans *cand*, la forme PPN est la même que dans le cas de SF. Par conséquent, pour la place (4, 8), SF (가) (4, 6) - PPN (는) (6, 8) est seulement autorisé. Les autres combinaisons violent le principe de T1', " SF se situe toujours à la tête de PPN ".

Pour le premier candidat du NS, nous pouvons proposer la forme NS (정치) (0,4) - SF (가) (4, 6) - PPN (는) (6, 8).

Les résultats positifs de l'analyse sont enregistrés dans la structure *cand* de la figure 4-5 le contenu est comme suit:

```
vert = 1          /* La donnée actuelle est le premier candidat de NS */
hori = 3          /* Le nombre des types est 3                        */
data = "정치가는"
flag[0] = 1       /* Pour le premier candidat, il y a seulement une forme,
                     cela signifie qu'il n'y a pas d'ambiguïté
                     Pour le premier candidat de NS                   */
type[0][0] = 2    /*    type 2 dénote NS                              */
loca[0][0] = 4    /*    la dernière position de SN dans data          */
type[0][1] = 3    /*    type 3 dénote SF                              */
loca[0][1] = 6    /*    la dernière position de SF dans data          */
```

type[0][2] = 4 /* type 4 dénote PPN */
loca[0][2] = 8 /* la dernière position de PPN dans *data* */

2) k = 2 cas:

Le deuxième candidat NS est "가"(4, 6). Pour analyser cette deuxième possibilité, nous devons la rechercher dans le dictionnaire de PF car seule la forme PF peut se situer à la tête du NS. Mais, il n'y a pas d'entrée "정치" dans le dictionnaire. C'est un cas d'échec.

3) k = 3 cas:

Le troisième cas est une combinaison des deux cas précédents. Le troisième candidat NS, "정치가"(0, 6) (NS-NS), est la forme répétitive de NS. Cette forme est acceptable dans T1' comme "정치" (la politique) (NS) - "가" (l'addition) (NS) - "는" (PPN) (nominatif). Bien que la forme marquée, NS-NS-PPN, n'ait pas d'erreur syntaxique, le deuxième NS (l'addition) "가" est en conflit avec SF et n'est sémantiquement pas acceptable. BDD filtrage dans la figure 4-1 n'a pas d'information pour lever l'ambiguï té **P1** entre NS et d'autres interprétations possibles.

T2 est le même que le cas de T1'. La PDD principale de T2 est PN. T2 possède en plus le dictionnaire par rapport à T1'.

T3 utilise aussi les structures *cand* et *cant* pour l'analyse du mot. D'abord PPP, puis SFR, le prédicat et les séquences du nom sont appliqués. La structure T3 est analysée avec les séquences en ordre inverse dans le mot, parce que la particule principale de T3 est PPP qui est placé à la fin du mot. Les procédures d'analyse sont les mêmes que dans les cas T1' et T2.

Par exemple, la séquence "비오는"(dans la pluie) est analysée comme "비" (NS) (0, 2), "오 " (SFR) (2, 4), et "는" (PPP) (4, 6) dans l'ordre inverse de la structure T3. Un autre exemple comme "고독하게"(solitairement) est composé de "고독하" (le radical, 하 est SFR) (0, 6), et "게" (PPP) (6, 8). Même si la forme "고독" (0, 4)(solitude) est analysée, rappelons que nous avons enregistré que toutes les formes d'adjectifs comportant '하다'(faire) sont dans le dictionnaire des adjectifs.

4.2.1. Exemples de test

Notre méthode détecte beaucoup d'ambiguïtés, mais elles sont presque inévitables au niveau de la morphologie, dans la mesure où nous n'utilisons pas d'heuristiques approximatives. Pour le NS, la séquence "광산은"a deux possibilités d'étiquetage. L'une est "광산"(la mine) (NS) et "은" (l'argent) (NS). L'autre est "광산"(la mine) (NS) and "은" (nominatif) (PPN). L'étiquetage correct dans ce cas dépend du sens ou de la structure de la phrase.

Mis à part NS, les autre particules incluant PF, SF, PPN, FN, PPP, SFR, V et A ne posent pas de problème. Tous les problèmes morphologiques ont été filtrés dans le module BDD filtrage comportant l'information BDD et les règles.

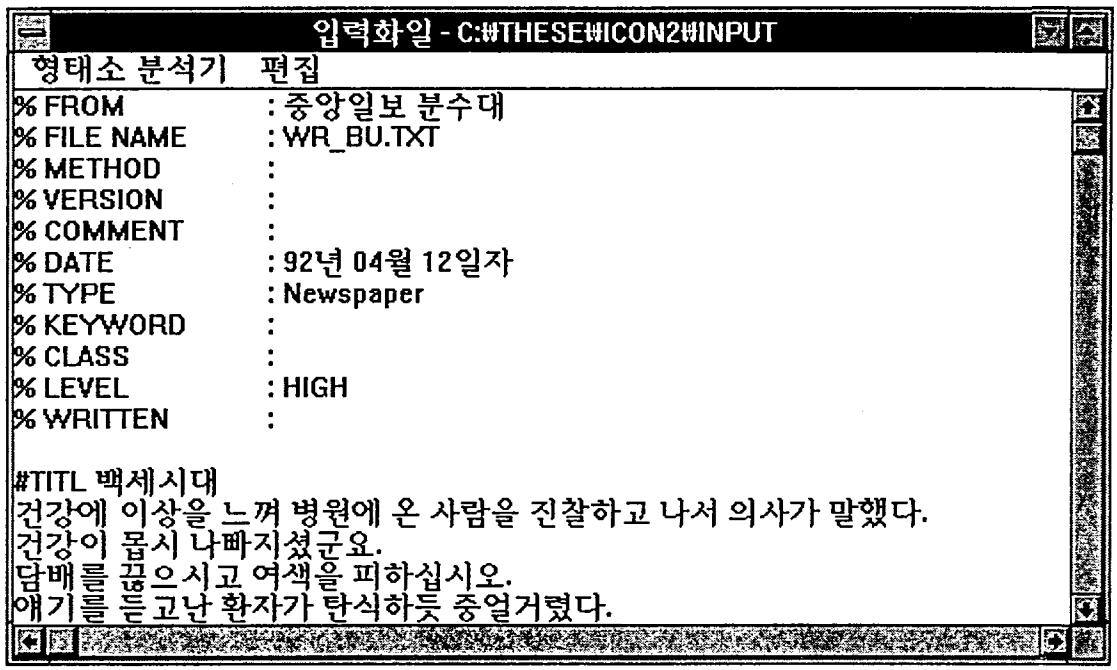


Figure 4-7. Texte entré pour le système EPL.

Plusieurs exemples sont donnés dans la figure 4-7. Les figures 4-8-1, 4-8-2 et 4-8-3 montrent les résultats d'exécution de **EPL** pour les entrées de la figure 4-7. Mais ces résultats sont obtenus avant de passer BDD filtrage (cf. figure 4-1). Le premier exemple, "건강에" peut produire trois candidats, comme NS-SF-NPP, NS-PPN, et PF-NS-NPP. On ne peut pas choisir l'un des trois candidats, parce que nous n'avons pas de dictionnaire ND (Nom Dérivé). Dans ce cas, nous fournissons les trois candidats.

Un autre exemple est un cas de "의사가" (médecin est) dans figure 4-8-3. Il y a sept possibilités d'analyse. Mais, après passage dans le module BDD Filtrage, les deux interprétations '*1th*' et '*4th*' seront supprimées, car cette entrée "의사가" (médecin est) n'est pas trouvés dans le module. C'est-à-dire, on n'interprètera pas certaines entrées comme *noms quelconques* + *SF*. Si on avait un dictionnaire ND, on pourrait enlever la

première et la dernière interprétation comme "의" (NS) + "사" (SF) + "가" (PPN) et "의" (PF) + "사가" (NS).

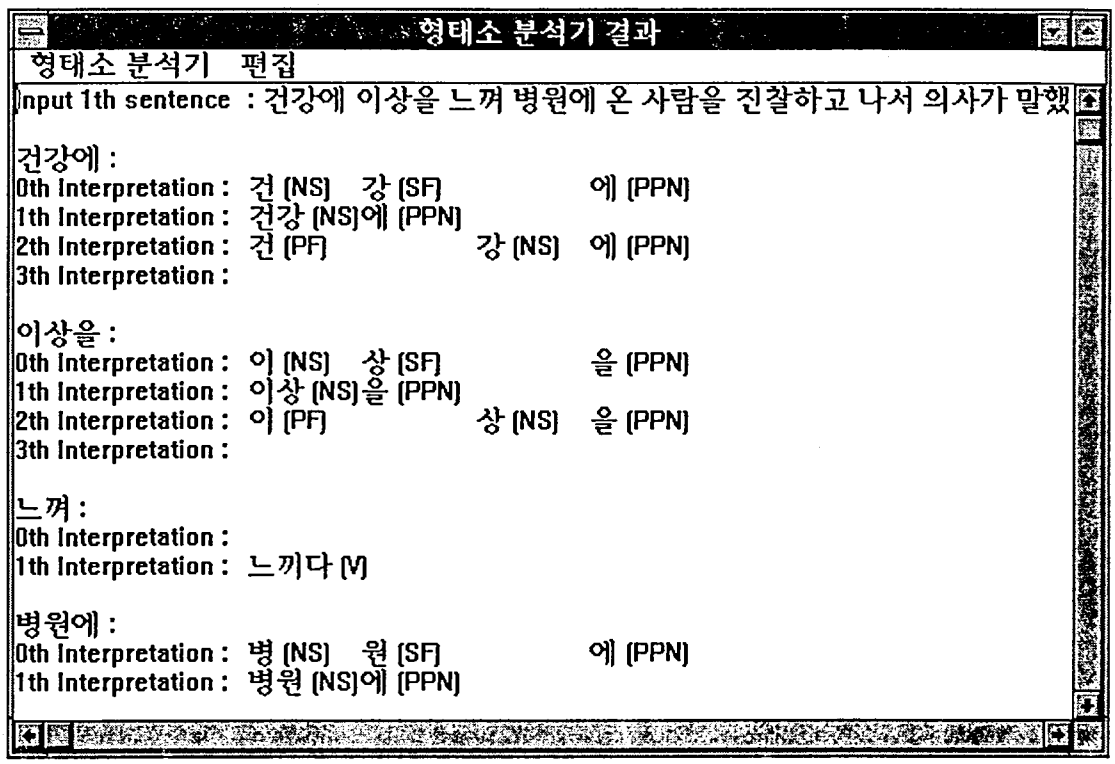


Figure 4-8-1. Résultats d'exécution de EPL des entrées de la figure 4-7.

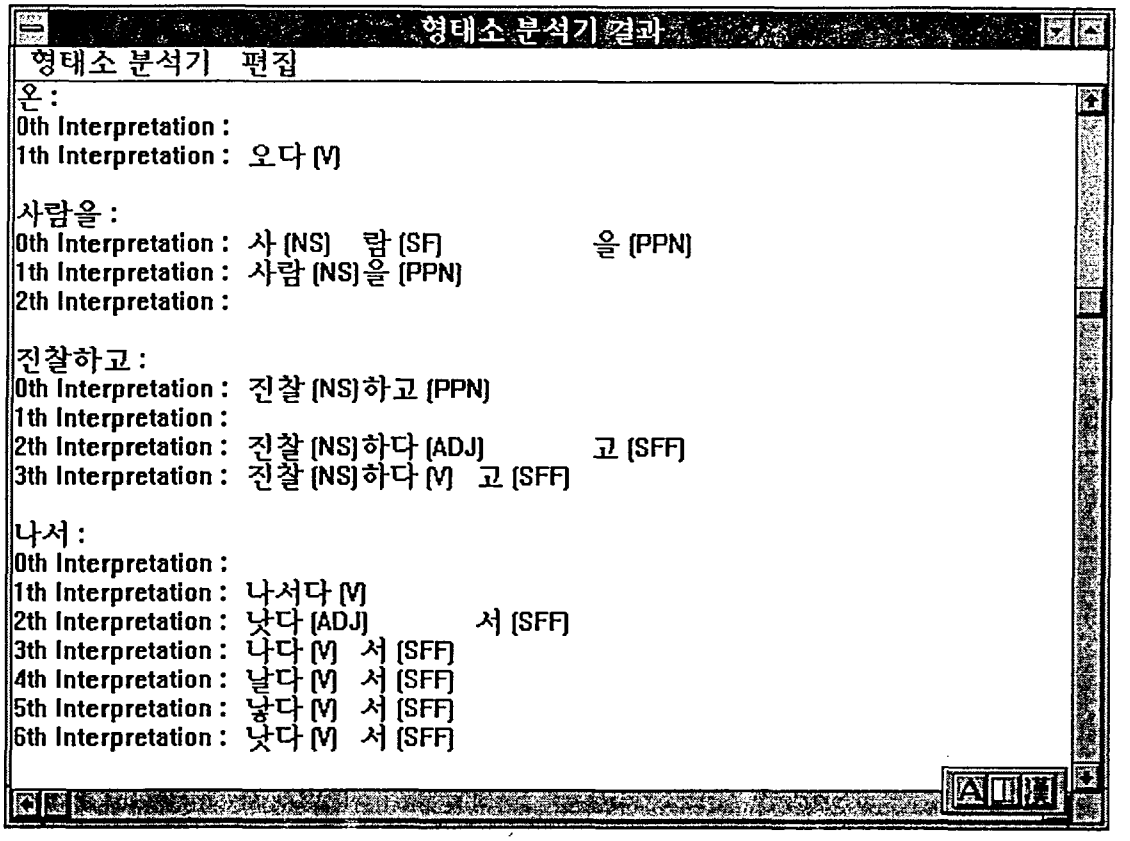


Figure 4-8-2. Suite de la figure 4-8-1.

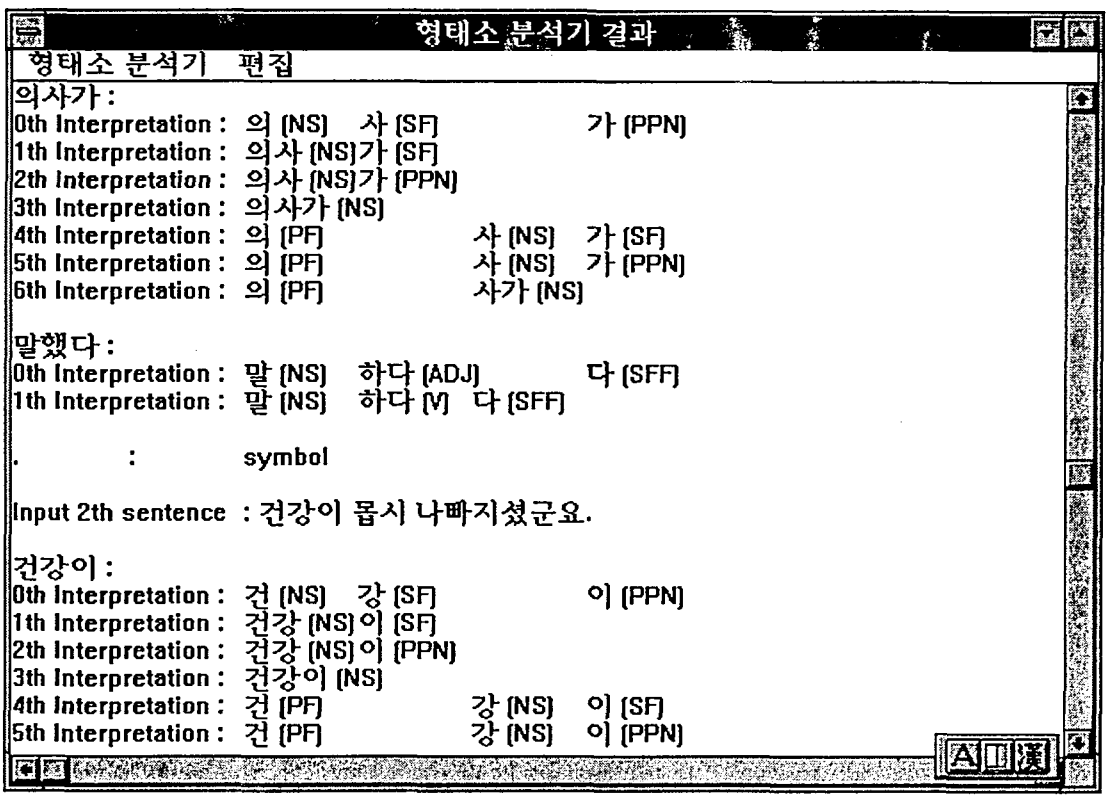


Figure 4-8-3. Suite de la figure 4-8-2.

5. Conclusion

Dans ce travail, nous avons développé un analyseur morphologique du coréen. La particularité de notre système consiste en l'utilisation de dictionnaires construits systématiquement. Etant donné qu'en coréen, les unités délimitées par des séparateurs ne sont que rarement des unités lexicalement simples, mais des séquences plus complexes (e.g. les postpositions casuelles sont soudées aux substantifs, et par ailleurs elles se combinent entre elles), des dictionnaires qui ne comportent que des mots simples ne sont pas suffisants pour reconnaî tre tous les mots et analyser correctement un texte donné.

Pour ce travail, nous avons engendré des dictionnaires de formes fléchies à partir des dictionnaires des *mots simples*, et des dictionnaires d'*affixes* et de *postpostions* : Dictionnaire des *nominalisations* des termes prédicatifs (FN), et Dictionnaire des formes fléchies des *verbes* et des *adjectifs* (FFP). Pour les cas où il existe des homographes (e.g. entre *suffixe* et *postposition*), nous avons établi des listes complètes de noms dérivés comportant ce type de suffixes. Cette liste est intégrée dans la module BDD. Quand un nom se trouve à gauche d'un morphème qui peut entrer éventuellement dans la liste des suffixes, on cherchera cette séquence dans le dictionnaire des mots dérivés de type N-SF : si on ne le trouve pas dans ce dictionnaire, cette séquence sera analysée comme N-PPN.

En nous basant sur ces dictionnaires, nous avons implémenté un système d'analyse morphologique baptisé **EPL.** Ce système propose toutes les interprétations possibles d'une séquence donnée. Voici un exemple :

janeun : 1>	자 (*ja*) [V:dormir]	+ 면	(*neun*) [PPP:Déterminatif]
2>	자 (*ja*) [NS:règle]	+ 면	(*neun*) [PPN:Nominatif]
saineun: 1>	새 (*sai*) [NS:oiseau]	+ 면	(*neun*) [PPN:Nominatif]
2>	새 (*sai*) [V:fuir]	+ 면	(*neun*) [PPP: Déterminatif]
jugessda: 1>	죽 (*jug*) [V:mourir]+ 었다(*essda*) [PPP:Déclaratif_Passé]		

Figure 5-1. Exemples de l'analyse morphologique du coréen.

Nous devons ajouter quelques données lexicales dans la version actuelle de notre système : il s'agit d'une liste d'éléments relativement courte. Les dictionnaires qui comportent des *adverbes*, des *pronoms*, des *noms incomplets* et des *numéraux* doivent être construits selon les mêmes principes.

Les dictionnaires des *noms propres* et des *noms composés* sont à établir à part : ils sont importants en nombre d'entrées et ils ont leur syntaxe particulière. Dans l'analyse morphologique de textes écrits et la documentation automatique, ces dictionnaires seront indispensables.

Nous nous sommes limité, dans ce travail, à la construction d'un analyseur morphologique du coréen, mais l'analyse syntaxique d'un texte donné sera nécessaire dans tous les systèmes de traitement de documents écrits en langue naturelle. Lors de l'analyse syntaxique et sémantique de textes coréens, le *lexique-grammaire* et des *grammaires locales* doivent être utilisés[18], aussi bien que les dictionnaires que nous construisons.

BIBLIOGRAPHIE

[1] Cho Young-Hwan, 1992, *Hangul TRIE and Bidirectional Access Method*, RT, Kaist CS Lab.

[2] Choi Ki-Sun et al., 1993, *Le développement de la documentation informelle*, Rapport du Departement Commercial, Séoul,

[3] Choi, Sung Woo, 1996, "Hangul Code Manager", A meeting of Korean Linguistics in CERIL : *Mémoires du CERIL* N° 14.

[4] Courtois, Blandine, 1987, *Dictionnaire Electronique du Laboratoire d'Automatique Documentaire et Linguistique pour les mots simples du francais* (DELAS), RT-17, LADL, Université Paris 7.

[5] Courtois, Blandine, 1989, *Dictionnaire Electronique du LADL pour les mots flechis du francais* (DELAF), RT-17, LADL, Université Paris 7.

[6] Crochemore, Maxime et Rytter, Wojciech, 1994, *Text Algorithms*, Oxford University Press.

[7] Crochemore, Maxime et Hancart, C., 1997, *Automata for matching patterns*, IGM-97-06, Université de Marne-La-Vallée..

[8] Gazdar, G., Klein, E., Pullum, G., et Sag, I., 1985, *Generalized Phrase Structure Grammar*, Oxford : Blackwell.

[9] Gazdar G. and Mellish, 1989, *Natural Language Processing in LISP*, Addison Wesley.

[10] Gross, Gaston, 1994, "Classes d'objects et description de verbes", *Langages* 115.

[11] Gross, Maurice, 1975, *Méthodes en Syntaxe*, Hermann,.

[12] Gross, Maurice, 1987, Electronic Dictionaries and Automata in Computational Linguistics, *Lecture Notes in Computer Science of the LITP Spring School on Theoretical Computer Science*, Berlin: Springer-Verlag.

[13] Gross, Maurice, 1987, The use of finite automata in the lexical representation of natural language, Electronic Dictionaries and Automata in Computational Linguistics, *Lecture notes in Computer Science* 377, Berlin : Springer-Verlag.

[14] Han, Sun-Hae, 1993, "Sur la construction nominale *N0 N1*-leul *Nprédé*-leul Hada", *Mémoires du CERIL*, N°12, Institute Gaspard Monge, Université de Marne-la-Vallée.

[15] Kang, HyunKyu, LEE, ChangYeol et al, 1994, "An Implementation of an automatic Keyword Extraction System", *The 3rd Pacific Rim International Conference on Artificial Intelligence* (PRICAI), Beijing, China.

[16] Karttunen, Lauri, 1993, *Finite-State Lexicon Compiler*, Xerox, Palo Alto Research Center, ISTL-NLTT-1993-04-02.

[17] Krovetz, Robert et W. Bruce Croft, 1992, Lexical Ambiguity and Information Retrieval, *ACM Trans. on Information Systems* Vol.10 N° 2.

[18] Laporte, Eric, 1994, "Experiences in lexical disambiguration using local grammars", *In COMPLEX 94*, Papers in Computational Lexicography.

[19] Lee, ChangYeol, 1996, "Ambiguity Processing in DECO system", A meeting of Korean Linguistics in CERIL : *Mémoires du CERIL* N° 14.

[20] Liang , Franklin Mark, 1983, *Word Hy-phen-a-tion by Com-put-er*, PhD. thèse, Department of Computer Science, Report N° STAN-CS-83-977 : Stanford University.
[21] Longman Language Activator,1993, The World's First Production Dictionary, Longman Group UK.
[22] Miller, George A. et al., 1990, *Introduction to WordNet*: An On-line Lexical Database, Cognitive Science Lab., Université Princeton, CSL Report N° 43.
[23] Nam, Jee-Sun, 1994, *Classification syntaxique des constructions adjectives en Coréen*, Vol II-annexes, Ph.D. Université Paris 7.
[24] Nam, Jee-Sun, 1994, *Dictionnaire Des Noms Simples Du Coréen*, RT-46, LADL.
[25] Nam, Jee-Sun, 1996, *Dictionary of Korean Simple Verbs*, RT-49, LADL.
[26] Nam, Jee-Sun, 1996, *Building A Korean On-Line Dictioanry Adequate to Parsing Systems*, RT-50, LADL.
[27] Nam, Jee-Sun et Lee, ChangYeol, 1996, *Korean Electronic Dictionary I/II*, IGM96-3/4, Université de Marne-La-Vallée.
[28] Nam, Jee-Sun, 1996, *Lexicon of Korean predicative terms classified by mophological ending forms*. Vol I., IGM 96-8, Université de Marne-La-Vallée.
[29] Nam, Jee-Sun, 1996, *Lexicon of Korean predicative terms classified by mophological ending forms*. Vol II., IGM 96-9, Université de Marne-La-Vallée.
[30] Nam, Jee-Sun, 1996, *Dictionary of N-Postpositions, A-Postpositions, and V-Postpositions in Korean*. RT-51, LADL.
[31] Nam, Jee-Sun, 1997, *Traitement des séquences comportant hada dans un lexique électronique du coréen*, RT-IGM 97, Université de Marne-la-Vallée.
[32] Park Dong-Sun, 1989, "La standardisation de Hangul et Hanza", *OA Lecture*, Séoul, p98~ p106.
[33] Pereira, Fernando C.N. et Barbara J. Grosz, 1994, *Natural Language Processing*, Cambridge, Mass. : The MIT Press.
[34] Perrin, Dominique, 1989, "Automates et algorithmes sur les mots", *Annales des télécommunications*, p20- p33.
[35] Revuz, Dominique, 1991, Dictionnaires et Lexiques Methodes et Algorithmes, Université Paris 7, LITP 91-44.
[36] Shim Jae-Ki, 1980, "La fonction de la nominalisation", *Langue*, Vol. 5, N°1, Séoul, p79 ~ p102.
[37] Shin Ki-Chul, Shin Yong-Chul, 1990, Nouveau Dictionnaire de la Langue Coréenne, Séoul, *Samsung Chulphansa*.
[38] Silberztein, Max, 1993, *Le système INTEX*, Masson.
[39] Tarjan, Robert Endre et Andrew Chi-Chih Yao, 1979, "Storing a Sparse Table", *CACM* Vol. 22 N°11, Stanford University.
[40] Walker, Donald E., et al., 1995, *Automating The Lexicon: Research and Practice in a Multilingual Environment*, Oxford : Clarendon Press

ANNEXE A. Les entrées ambiguës P1 entre PF et NS

(de PF) (252) 가 가장 각 간 감 갑 강 개 객 건 격 결 경 계 고동 곡 골 곰 공 과 관 광 광대 구정 국 군 귀 균 극 극대 극소 극한 금 급 기 길 낙 난 날 남 낱 내 냉 노 녹 논 농 다 다갈 단 단장 담 당 대 대접 도 도마 독 동 되 뒤 득 들 등 때 마 막 만 말 망 매 면 명 모 모시 목 묘 무 무당 묵 문 미 밀 박 반 발 방 배 백 범 법 변 별 보라 복 복사 본 봉 부 분 불 비 빗 사 사시 산 살 삽 상 상고 상품 새 선 설 성 세 소 속 솔 수 수양 순 술 쉬 시 식 신 실 쌍 악 안 알 암 애 액 약 양 양재 얼 얼룩 역 역순 연 연두 열 염 영 예 오지 옥 온난 올 왕 요 욕 용 용수 ᄋᆞ ᄋᆞ두 운 원 위 윤 융 은 음 의 이 인 입 자 잔 장 재 쟁 저 적 전 ᄎᆞᆫ ᄌᆞᆫ 절 점 접 정 제비 조 조약 좀 종 좌 주 죽 중 중품 진 진 진도 질 징 짝 차 참 창 채 처 천 철 청 체 초 촉 촌 최단 추 축 축사 탈 태 토 통 투 파 판 패 편 평 폐 포 표 품 풍 피 하품 학 한 함 함박 합 해 향 허 현 형 호 호로 호리 혹 혼 황 회 효 훈 흉

de PF de T1' (10) 극대 극소 극한 다갈 상품 역순 온난 중품 최단 하품

(262 au total)

ANNEXE B-1. Les candidats de l'ambiguïté P2 entre PF et NS

Les candidats suivants de l'ambiguïté P2 sont produits automatiquement à partir de la liste **de** 131 PF et 2,739 NS. Ces candidats sont les formes acceptables dans T1'.

가장(PF) + 가 골 관 군 기 난 남 내 단 담 대 도 마 면 모 문 물 미 발 복 본 인 부 비 사 삼 성 소 손 수 식 신 악 안 애 원 인 자 전 정 죽 천 티 푸 스 파 품 학 해 화 (NS).

강남(PF) + 매 바 위 발 비 색 자 장 창 포(NS).

강낭(PF) + 군 만 비 색 인 패 (NS).

고동(PF) + 감 갑 강 결 경 계 공 기 면 명 무 문 물 반 백 복 봉 산 상 설 안 업 우 원 위 의 이 인 자 전 정 조 참 창 체 침 태 토 티 포 품 한 향 호 인 화 (NS).

고품(PF) + 격 계 귀 명 목 별 사 성 위 절 (NS).

곱사(PF) + 가 각 감 건 격 계 골 공 과 관 교 복 교 성 기 나 이 내 다 리 단 담 당 대 도 돈 동 마 귀 막 망 면 명 모 무 문 물 발 방 범 변 별 복 본 부 분 비 사 살 상 색 생 자 생 화 선 설 시 식 신 실 양 업 역 연 열 영 욕 용 원 위 은 의 이 인 임 자 재 적 전 절 정 조 죄 주 죽 중 진 질 창 채 체 촌 춘 기 태 통 투 팔 표 학 형 화 활 회 (NS).

곱수(PF) + 간 감 갑 건 검 공 관 구 가 국 금 급 난 납 뇌 다 단 달 당 도 동 라 마 면 모 목 묵 미 박 반 배 복 부 분 비 사 사 학 산 상 색 성 세 소 속 수 술 식 신 액 양 양 녀 업 염 영 예 요 욕 용 용 액 은 음 의 인 입 장 재 저 절 정 제 비 제 품 종 질 집 차 첩 청 초 축 탈 태 평 포 표 표 피 하 물 학 호 화 화 기 화 인 화 자 (NS).

광대(PF) + 가 가 리 각 강 개 결 경 계 공 과 관 국 궐 금 기 농 담 답 대 도 독 마 망 면 모 목 물 변 본 비 사 상 성 세 소 수 양 역 용 응 위 음 의 인 입 장 적 전 접 조 좌 중 진 질 차 책 첩 청 체 추 파 판 패 포 폭 표 풍 피 학 한 함 합 해 화 회 (NS).

구공(PF) + 간 감 개 격 경 공 관 극 금 급 기 납 막 매 명 모 무 문 물 박 방 백 벌 범 복 부 비 사 산 상 설 세 소 수 시 식 안 약 양 업 역 연 영 예 용 인 임 자 장 저 적 전 정 주 중 채 책 천 치 사 토 통 판 평 포 표 학 한 지 해 화 제 회 (NS).

구정(PF) + 각 간 경 계 곡 공 관 글 기 당 도 독 돈 무 물 미 소 밀 도 박 방 벌 범 변 복 본 부 비 사 산 선 설 성 세 수 수 리 식 신 액 약 양 열 염 예 원 의 인 자 장 쟁 적 전 절 정 조 종 진 책 처 청 체 초 탐 토 통 품 학 합 해 형 혼 화 (NS).

극단(PF) + 가 결 계 골 기 도 독 문 발 복 산 상 소 속 식 신 열 원 위 일 자 장 전 정 조 종 좌 죄 주 청 체 축 풍 합 화 (NS).

극대(PF) + 가 가 리 각 강 개 결 경 계 공 과 관 국 궐 금 기 농 담 답 대 도 독 마 망 면 모 목 물 변 본 비 사 상 성 세 소 수 양 역 용 응 위 음 의 인 입 장 적 전 접 조 좌 중 진 질 차 책 첩 청 체 추 파 판 패 포 폭 표 풍 피 학 한 함 합 해 화 회 (NS).

극소(PF) + 각 감 강 개 경 계 관 국 극 금 급 기 대 독 동 득 등 라 매 맥 명 모 묘 문 박 반 방 복 산 설 속 식 신 심 성 심 증 액 양 요 원 위 의 이 인 일 임 장 정 조 주 진 질 집 추 침 통 파 포 품 풍 피 (NS).

극한(PF) + 계 과 극 기 대 도 문 방 복 설 수 역 옥 응 의 자 정 참 창 천 탄 파 풍 학 화 (NS).

급회(PF) + 개 계 관 국 귀 금 담 도 동 복 부 비 사 장 선 수 순 식 신 안 연 오 리 우 원 의 자 장 전 조 중 진 피 한 합 향 화 (NS).

나막(PF) + 내 대 동 이 바 지 사 (NS).

나박(PF) + 격 대 봉 사 살 색 수 애 자 장 진 차 탈 학 해 (NS).

냉난(PF) + 간 관 국 독 동 무 발 봉 사 산 세 소 입 자 정 투 파 포 해 황 (NS).

누비(PF) + 가 결 계 관 단 등 라 만 명 문 밀 방 상 성 소 속 화 약 용 운 위 자 장 재 전 축 탈 토 파 판 평 품 호 화 황 (NS).

다갈(PF) + 고 리 구 리 기 대 등 망 매 기 무 리 비 삼 수 피 (NS).

단장(PF) + 가 골 관 군 기 난 남 내 단 담 대 도 마 면 모 문 물 미 발 복 본 인 부 비 사 삼 성 소 손 수 식 신 악 안 애 원 인 자 전 정 죽 천 티 푸 스 파 품 학 해 화 (NS).

대접(PF) + 경 골 대 변 속 수 시 영 전 종 촉 합 (NS).

도마(PF) + 감 개 고 자 구 간 귀 당 마 모 무 리 부 비 사 회 소 수 술 음 의 적 중 (NS).

도요(PF) + 가 강 건 결 금 기 도 동 망 변 부 사 새 소 약 양 원 인 절 정 철 체 통 해 (NS).

동아(PF) + 가 가 리 가 미 국 귀 기 낙 내 들 명 미 범 부 비 사 성 씨 악 양 연 우 이 쟁 전 집 첨 침 편 (NS).

딱정(PF) + 각 간 경 계 곡 공 관 글 기 당 도 독 돈 무 물 미 소 밀 도 박 방 벌 범 변 복 본 부 비 사 산 선 설 성 세 수 수 리 식 신 액 약 양 열 염 예 원 의 인 자 장 쟁 적 전 절 정 조 종 진 책 처 청 체 초 탐 토 통 품 학 합 해 형 혼 화 (NS).

마제(PF) + 가 감 강 계 공 과 관 국 기 답 도 독 동 막 매 명 목 물 반 방 법 복 본 분 비 사 세 소 시 약 염 위 의 재 전 정 조 주 창 초 패 한 해 화 (NS).
맨헛(PF) + 간 (NS).
맹혹(PF) + 사 성 세 자 정 한 (NS).
모시(PF) + 가 각 골 공 국 굴 내 달 도 동 무 발 범 비 사 사실 사회 상 설 세 술 식 신 역 위 음 의 인 장 접 정 조 종 중 청 체 초 추 침 판 한 해 현 효 (NS).
무당(PF) + 국 나귀 도 면 목 번 부 분 사 선 위 의 일 좌 질 초 (NS).
배내(PF) + 각 계 공 국 기 당 막 복 분 색 성 세 속 수 시 실 알 연 용 의 자 장 재 재인 접 정 조 종 진 처 통 포 피 한 향 흉 (NS).
보금(PF) + 기 납 단 물 수 실 액 욕 융 주 혼 (NS).
복사(PF) + 가 각 감 건 격 계 골 공 과 관 교 복 교 성 기 나이 내 다리 단 담 당 대 도 돈 동 마귀 막 망 면 명 모 무 문 물 발 방 범 변 별 복 본 부 분 비 사 살 상 색 생 자 생화 선 설 시 식 신 실 양 업 역 연 열 영 욕 용 원 위 은 의 이 인 임 자 재 적 전 절 정 조 죄 주 죽 중 진 질 창 채 체 촌 춘 기 태 통 투 팔 표 학 형 화 활 회 (NS).
부시(PF) + 가 각 골 공 국 굴 내 달 도 동 무 발 범 비 사 사실 사회 상 설 세 술 식 신 역 위 음 의 인 장 접 정 조 종 중 청 체 초 추 침 판 한 해 현 효 (NS).
사시(PF) + 가 각 골 공 국 굴 내 달 도 동 무 발 범 비 사 사실 사회 상 설 세 술 식 신 역 위 음 의 인 장 접 정 조 종 중 청 체 초 추 침 판 한 해 현 효 (NS).
삽살(PF) + 기 모 사 무 사 상 의 인 포 해 (NS).
상고(PF) + 가 개 객 고 학 공 관 국 군 금 기 난 뇌 대 도 독 동 명 모 목 무 문 물 발 배 백 별 분 비 사 사리 성 소 속 수 시 안 역 옥 용 원 의 인 자 자질 장 저 적 전 정 조 종 주 질 집 참 채 초 추 투 풍 학 함 향 혹 (NS).
상품(PF) + 격 계 귀 명 목 별 사 성 위 절 (NS).
새앙(PF) + 금 등 살 양 탈 (NS).
생안(PF) + 개 경 광 내 대 도 마 면 목 무 부 산 식 염 위 전 정 주 중 질 (NS).
선들(PF) + 통 (NS).
설농(PF) + 가 간 경 공 기 노 담 도 목 본 부 사 산 상 성 악 액 업 장 축 토 학 한기 (NS).
송골(PF) + 격 동 목 무 반 상 수 절 초 통 (NS).
수양(PF) + 가 각 계 곡 국 궁 난 단 도 도체 돈 동이 말 명 모 미간 병 복 봉 부 분 산 상 서류 성 수 식 악 어장 옥 용 위 은 인 자 장 재 재기 주 처 키 품 풍 학 해 화 (NS).
순헛(PF) + 간 (NS).
신접(PF) + 경 골 대 변 속 수 시 영 전 종 측 합 (NS).
심심(PF) + 계 금 기 뇌 도 독 문 복 부 사 상 성 술 신 심증 연 열 의 장 저 정 천 통 해 화 회 흉 (NS).
안간(PF) + 결 계 극 난 단 담 만 목 발 벌 부 사 선 소 소화 수 식 신 은 음 접 조 주 질 책 첩 촉 통 투 사 판 호 (NS).
양재(PF) + 가 간 개 계 귀 기 난 단 담 독 목 무 물 미 발 벌 범 변 봉 사 산 상 색 선 수 수생 실 액 연 욕 원 의 인 임 자 적 정 주 질 천 촉 축 침 판 패 편 학 합 향 혼 화 회 흥 (NS).
어금(PF) + 기 납 단 물 수 실 액 욕 융 주 혼 (NS).
얼간(PF) + á 계 극 난 단 담 만 목 발 벌 부 사 선 소 소화 수 식 신 은 음 접 조 주 질 책 첩 촉 통 투 사 판 호 (NS).
역순 (PF) + 간 결 국 대 도 라 방 백 산 수 시 열 위 장 종 차 풍 화 회 (NS).
연두 (PF) + 건 둑 목 미 발 부 상 통 (NS).
오막 (PF) + 내 대 동이 바지 사 (NS).
오목 (PF) + 가 각 격 금 기 단 동 마 사 수 욕 이 장 재 적 전 조 질 차 축 침 탄 표 향 화 회 (NS).
오지 (PF) + 가 각 개 게 경 관 구 성 구 전 국 급 기 단 대 도 명 목 묵 물 상 반 방 배 부 분 불 사 성 세 속 시 식 애 양 역 연 우 원 위 은 인 자 장 재 적 정 정 학 조 주 진 질 청 체 침 탄 평 표 학 향 회 (NS).
온난 (PF) + 간 관 국 독 동 무 발 봉 사 산 세 소 입 자 정 투 파 포 해 황 (NS).
용수 (PF) + 간 감 갑 건 검 공 관 구 가 국 금 급 난 납 뇌 다 단 달 당 도 동 라 마 면 모 목 묵 미 박 반 배 복 부 분 비 사 사학 산 상 색 성 세 소 속 수 술 식 신 액 양 양녀 업 염 영 예 요 욕 용 용액 은 음 의 인 입 장 재 저 절 정 제비 제품 종 질 집 차 첩 청 초 축 탈 태 평 포 표 피 하물 학 호 화 화기 화인 화자 (NS).
우두 (PF) + 건 둑 목 미 발 부 상 통 (NS).
원두 (PF) + 건 둑 목 미 발 부 상 통 (NS).
월계 (PF) + 간 기 명 모 발 산 속 수 시 약 장 절 정 진 집 책 통 피 획 (NS).

이복(PF) + 각 강 귀 도 면 부 사 상 사 선 수 역 용 원 잡 화 장 종 창 채 판 학 합 화 술 (NS).
이종(PF) + 가 강 결 국 기 다 리 단 대 말 목 묘 무 반 복 비 사 손 식 신 씨 양 업 용 이 자 적 전 주 창 판 합 (NS).
자봉(PF) + 건 급 기 분 양 오 리 인 창 책 토 투 합 화 황 (NS).
재종(PF) + 가 강 결 국 기 다 리 단 대 말 목 묘 무 반 복 비 사 손 식 신 씨 양 업 용 이 자 적 전 주 창 판 합 (NS).
저품(PF) + 격 계 귀 명 목 별 사 성 위 절 (NS).
적녹(PF) + 각 내 장 말 음 초 화 (NS).
전전(PF) + 각 개 격 경 골 과 관 광 국 기 농 단 달 답 당 대 도 도 금 동 라 말 망 매 모 무 반 방 범 병 복 부 분 불 비 사 설 성 소 속 수 술 시 신 실 업 역 염 용 운 원 위 유 물 의 이 인 임 입 자 장 재 쟁 전 정 조 주 진 집 채 체 초 축 토 통 투 파 패 폐 학 합 해 향 형 화 황 (NS).
제비(PF) + 가 결 계 관 단 등 라 만 명 문 밀 방 상 성 소 속 화 약 용 운 위 자 장 재 전 축 탈 토 파 판 평 품 호 화 황 (NS).
조약(PF) + 간 골 과 관 도 방 본 분 사 소 속 술 식 자 정 조 진 질 체 탈 호 흔 (NS).
주마(PF) + 감 개 고 자 구 간 귀 당 마 모 무 리 부 비 사 회 소 수 술 음 의 적 중 (NS).
주안(PF) + 개 경 광 내 대 도 마 면 목 무 부 산 식 염 위 전 정 주 중 질 (NS).
중품(PF) + 격 계 귀 명 목 별 사 성 위 절 (NS).
진도(PF) + 강 공 굴 금 난 달 당 덕 둑 마 막 망 면 미 박 배 벌 벽 사 산 살 성 시 식 안 약 음 읍 입 장 적 전 주 중 처 체 태 통 포 피 해 화 화 지 회 (NS).
청녹(PF) + 각 내 장 말 음 초 화 (NS).
초생(PF) + 가 각 강 계 기 도 동 감 득 명 모 물 부 불 사 산 선 수 식 신 애 원 자 장 질 채 청 체 태 판 포 화 활 (NS).
초승(PF) + 가 강 구 강 장 객 계 낙 마 무 복 부 상 선 세 소 용 운 인 전 진 패 (NS).
최단(PF) + 가 결 계 골 기 도 독 문 발 복 산 상 소 속 식 신 열 원 위 일 자 장 전 정 조 종 좌 죄 주 청 체 축 풍 합 화 (NS).
축사(PF) + 가 각 감 건 격 계 골 공 과 관 교 복 교 성 기 나 이 내 다 리 단 담 당 대 도 돈 동 마 귀 막 망 면 명 모 무 문 물 발 방 범 변 별 복 본 부 분 비 사 살 상 색 생 자 생 화 선 설 시 식 신 실 양 업 역 연 열 영 욕 용 원 위 은 의 이 인 임 자 재 적 전 절 정 조 죄 주 죽 중 진 질 창 채 체 촌 춘 기 태 통 투 팔 표 학 형 화 활 회 (NS).
투명(PF) + 궁 기 단 당 도 망 맥 명 목 문 물 부 분 사 상 색 성 수 시 암 의 인 장 저 절 주 중 창 태 함 (NS).
표주(PF) + 가 간 객 검 곡 과 관 군 기 단 대 도 동 막 먹 목 무 문 물 미 반 발 범 벽 변 부 불 사 사 위 산 상 선 성 소 술 식 악 업 연 요 위 인 임 입 자 자 학 장 재 저 저 리 전 점 접 정 조 찬 창 책 체 춤 파 판 형 화 황 (NS).
하품(PF) + 격 계 귀 명 목 별 사 성 위 절 (NS).
한냉(PF) + 기 대 동 소 수 이 장 전 정 조 채 (NS).
함박(PF) − 격 대 봉 사 살 색 수 애 자 장 진 차 탈 학 해 (NS).
보라(PF) + 디 오 이 트 일 락 (NS).
애벌(PF) + 거 숭 이 레 목 이 초 (NS).
옹달(PF) + 개 비 걀 관 구 지 래 러 변 음 박 질 음 질 인 팽 이 필 (NS).
자물(PF) + 가 감 건 고 권 량 레 력 리 망 문 물 상 성 심 욕 자 정 주 증 질 체 품 (NS).
종달(PF) + 개 비 걀 관 구 지 래 러 변 음 박 질 음 질 인 팽 이 필 (NS).
징검(PF) + 거 댕 둥 이 류 문 불 사 색 술 시 식 안 역 열 인 정 증 찰 침 토 (NS).

ANNEXE B-2. Désambiguïsation entre PF et NS pour les données de P2

PF - NS	Etiquetage correct	PF - NS	Etiquetage correct
가장물(S)	가장물	고품명	고(PF) -품명
가장비(S)	가장비	고품목	고(PF) -품목
고동기(S)	고동기	고품위(S)	고(PF) -품위
고품격	고(PF) -품격	광대가(S)	광대(PF) -가
고품계	고(PF) -품계	광대역	광대
광대파	광대(PF) -파	구공식	구(PF) -공식
광대폭	광대(PF) -폭	구공약	구(PF) -공약
구공관	구(PF) -공관	구공장	구(PF) -공장
구공문	구(PF) -공문	구공채	구(PF) -공채
구공사	구(PF) -공사	구정관	구(PF) - 정관
구정물	구정(PF) -물	극대파	극(PF) -대파
구정체(S)	구정(PF) -체	극소설	극(PF) -소설
극단가(S)	극단(PF) -가	극소액	극(PF) -소액
극단주	극단(PF) -주	극한대	극 - 한대
극대조	극(PF) -대조	급회전(S)	급(PF) -회전
누비관(S)	누비관	대접시	대접시
단장사(S)	단장사	도요지(S)	도요지
단장정(S)	단장정	마제초(S)	마제초
단장품(S)	단장품	맹혹사	맹(PF) -혹사
단장화(S)	단장화	모시가(S)	모(PF) - 시가
무당질(S)	무당(NS)-질	복사도	복사(PF) -도
배내기(S)	배내(PF) -기	복사면	복사(PF) -면
복사각	복사(PF) -각	복사본	복사(PF) -본
복사건	복사(PF) -건	복사비	복사(PF) -비
복사대	복사(PF) -대	복사색	복사(PF) -색
복사선(S)	복사(PF) -선	복사자	복사(PF) -자
복사시	복사(PF) -시	복사계(S)	복사(PF) -계
복사실	복사(PF) -실	복사파(S)	복사(PF) -파
복사업	복사(PF) -업	복사기(S)	복사(PF) -기
복사열(S)	복사(PF) -열	복사담(S)	복사담
복사무(S)	복사무	사시도(S)	사시도
복사상(S)	복사상	상고공(S)	상고공
복사체(S)	복사체	상고대(S)	상고대
부시인(S)	부시인	상고배(S)	상고배
부시장(S)	부시장	상고사(S)	상고사

PF - NS	Etiquetage correct	PF - NS	Etiquetage correct
상고인(S)	상고인	상품명	상품(NS) -
상고모	상고모	상품목	상품(NS) -
상고참	상(PF) -고참	상품별	상품(NS) -
상품격	상(PF) - 품격	상품성	상품(NS) -
상품계	상품(NS) -	생안경	생안경
수양모(S)	수양모	수양재	수양재
수양부(S)	수양부	양재기(S)	양재기
수양산(S)	수양산	양재사(S)	양재사
수양수(S)	수양수	역순산	역(PF) -순산
수양자(	수양자	역순위	역(PF) -순위
오목장 (S)	오목장	용수난	용수(NS) -난
오지대 (S)	오지대	용수비	용수(NS) -비
온난관	온난(NS) -관	용수액	용수(NS) -액
용수공	용수(NS) -공	용수정	용수(NS) -정
용수관	용수(NS) -관	용수초 (S)	용수(NS) -초
우두목	우(PF) -두목	저품계	저(PF) -품계
월계수 (S)	월계(PF) -수	저품명	저(PF) -품명
이종손	이종(PF) -손	저품목	저(PF) -품목
재종손(S)	재종손	전전관(S)	전전관
저품격	저(PF) -품격	전전기(S)	전전기
전전부(S)	전전부	중품격	중(PF) -품격
전전파(S)	전전파	중품계	중(PF) -품계
제비가(S)	제비가	중품명	중(PF) -품명
조약도(S)	조약도	중품목	중(PF) -품목
주마감(S)	주마감	중품별	중(PF) -품별
최단가	최단(NS) -가	투명문	투명문
최단기	최단(NS) -기	투명색(S)	투명색
축사도(S)	축사도	투명성	투명성
축사주(S)	축사주	하품격	하(PF) -품격
투명도(S)	투명도	하품계	하(PF) -품계
하품명	하(PF) -품명	함박살(S)	함박살
하품목	하(PF) -품목		
하품별	하(PF) -품별		
한냉기	한냉기		
한냉대	한냉대		

ANNEXE C-1. Les candidats de l'ambiguïté P2 entre NS et SF

Les candidats suivants de l'ambiguï té P2 sont produits automatiquement à partir de la liste **de** 2271 NS et 253 SF. Ces candidats sont les formes acceptables dans T1'.

감개관국귀낙천농단대매목무물미배부분비사선세시애양연염영요 가락(SF)
원인자장재조종주진찬참창첨초추축패폐향허호호전화흉(NS)+
감개관국귀낙천농단대매목무물미배부분비사선세시애양연염영요 가량(SF)
원인자장재조종주진찬참창첨초추축패폐향허호호전화흉(NS)+
감개관국귀낙천농단대매목무물미배부분비사선세시애양연염영요 가리(SF)
원인자장재조종주진찬참창첨초추축패폐향허호호전화흉(NS)
폐(NS)+ 갱이(SF)
검과군기논대동별선소수은의전점정주집천철추(NS)+ 거리(SF)
검과군기논대동별선소수은의전점정주집천철추(NS)+ 거미(SF)
개경다대등묵백병부사상성수정약인전주청초통투폐한핵효(NS) 과리(SF)
가강개공극내노농당도마문미방배백불사살선속식안역연영요욕인 구녁(SF)
입자장장신전절정창천첨청촉추축침탐투폐포표학함해허호화(NS)+
가강개공극내노농당도마문미방배백불사살선속식안역연영요욕인 구리(SF)
입자장장신전절정창천첨청촉추축침탐투폐포표학함해허호화(NS)+
가강개공극내노농당도마문미방배백불사살선속식안역연영요욕인 구원(SF)
입자장장신전절정창천첨청촉추축침탐투폐포표학함해허호화(NS)+
가각감관국궐기막사시안암옥원이인장체흉(NS)+ 내기(SF)
가각감관국궐기막사시안암옥원이인장체흉(NS)+ 내미(SF)
경광극기냉대독등막무박부분사상선성세소순식악안연ㅇ의인임입 대기(SF)
장적전절접종주중차천초추축침편포피학한함허위현화(NS)+
가감강경과국군궤기노농다단당대도마면명무무인묵반복부사선세 도리(SF)
속수순시식신안애약양역영요용위ㅇ명의인장적전절정주중진집천
축파판편포학한해향호황회(NS)+
가감격골기난냉노목무미반발변부사상선소수시신악요ㅇ운이자전 동아(SF)
주진찬책청초태파폭피학합해향혼활황회(NS)+
가감격골기난냉노목무미반발변부사상선소수시신악요ㅇ운이자전 동이(SF)
주진찬책청초태파폭피학합해향혼활황회(NS)+
가감격골기난냉노목무미반발변부사상선소수시신악요ㅇ운이자전 동자(SF)
주진찬책청초태파폭피학합해향혼활황회(NS)+
가리감강경기낙대도마목무병부수안애얼연ㅇ이장처천테해(NS) 마님(SF)
가리감강경기낙대도마목무병부수안애얼연ㅇ이장처천테해(NS)+ 마리(SF)
가감국당대도동만모방복불사상수안영장체ㅍ(NS)+ 면피(SF)
가강골공국난동문병분사상시실쌍악안애업연열용운의임잔재전정 무리(SF)
종주집채책총판편풀학해허훈(NS)+
강결공논도두레면반선소속수ㅇ임정포함호(NS)+ 박이(SF)
강결공논도두레면반선소속수ㅇ임정포함호(NS)+ 박질(SF)
가개건공국낙만모문변비사선ㅈ순약연예ㅇ이일전정처추탐토표한 방면(SF)
해향혼(NS)+
건담도분선성세수시정연위진집참첨초축패(NS)+ 배기(SF)
가간갑결공과광국기납농당마망매명발배백범복본부분사상선세수 부리(SF)
신악안양열요욕음의인임임산자장전정조조산주중질창천첨청초포
풍피학현화회흉(NS)+
가간갑결공과광국기납농당마망매명발배백범복본부분사상선세수 부림(SF)
신악안양열요욕음의인임임산자장전정조조산주중질창천첨청초포
풍피학현화회흉(NS)+

가간갑결공과광국기납농당마망매명발배백범복본부분사상선세수 부지(SF)
신악안양열요욕음의인임임산자장전정조조산주중질창천첨청초포
풍피학현화회흉(NS)+

가 가중 간 감 감리 갑 강 개 객 검 결 경 공 관 국 급 기 난 농 답 당 대 도 독 등 막 매 명 모 목 묘 무 밀 박 반 발 방 백 변 병 복 본 부 분 불 사 상 선 설 속 수 시 신 악 애 약 역 연 열 영 예 예상 요 용 은 의 이 이용 일 입 장 재 전 정 조 종 주 중 중대 진 질 집 찬 참 처 천 축 탄 탐 토 투 판 패 평 폐 품 학 해 향 허 형 호 혹 혼 황 회 훈 (NS) + 사귀(SF)

갱 공 기 난 동 무 발 방 사 선 소 수 식 양 영 위 자 재 재수 중 창 천 축 탄 태 파 학 (NS) + 생원(SF)

각 간 경 금 낙 단 대 도 독 동 문 밀 배 봉 분 비 사 선 성 수 순 신 악 양 용 원 음 의 이 장 전 정 조 진 진정 질 첨 총 추 투 판 폭 피 혹 (NS) + 서리(SF)

각 간 경 금 낙 단 대 도 독 동 문 밀 배 봉 분 비 사 선 성 수 순 신 악 양 용 원 음 의 이 장 전 정 조 진 진정 질 첨 총 추 투 판 폭 피 혹 (NS) + 서방(SF)

간 개 객 결 관 국 기 낙 당 독 등 배 병 복 본 사 상 솔 신 알 우 인 자 재 쟁 적 정 조 주 참 총 침 파 폐 풍 피 함 화 회 (NS) + 선생(SF)

간 갱 걸 경 과 군 귀 단 독 만 무 문 미 반 발 배 변 분 불 사 소 수 시 실 역 영 운 위 은 인 임 입 자 장 전 정 종 중 처 체 총 추 통 투 피 화 회 (NS) + 신료(SF)

관 기 기타 남 농 무 미 발 방 부 사생 산 상 소 우 수 원 자 지진 (NS) + 아비(SF)

관 기 기타 남 농 무 미 발 방 부 사생 산 상 소 우 수 원 자 지진 (NS) + 아치(SF)

국 단 대 독 문 밀 반 방 병 북 불 비 상 상투 속 술 신 악 역 연 염 영 용 원 인 조 청 추 타이 투 표 풍 활 (NS) + 어치(SF)

결 관 급 대 부 상 소 수 잔 참 투 (NS) + 여사(SF)

가 각 설 구 렁 굽 길 냉 노 랑 놀 동 먹 모 목 미장 바 둑 벌 변 사 상 소 스키 식 애 송 역 전 종 천 키 티 파 편 풀 (NS) + 이개(SF)

가 각 설 구 렁 굽 길 냉 노 랑 놀 동 먹 모 목 미장 바 둑 벌 변 사 상 소 스키 식 애 송 역 전 종 천 키 티 파 편 풀 (NS) + 이전(SF)

감 강 격 공 국 군 기 난 남 내 노 단 독 동 망 모 문 물 박 방 배 우 백 부 분 비 사 사생 상 선 지 성 세 손 수화 식 신 실 암 액 약 양 용 원 융 의 인 임 입 장 재 저 적 전 정 종 주 진 창 철 첩 청 초 심 침 태 투 패 편 포 풍 학 한 함 현 혹 혼 화 활 회 (NS) + 자락(SF)

감 강 격 공 국 군 기 난 남 내 노 단 독 동 망 모 문 물 박 방 배 우 백 부 분 비 사 사생 상 선 지 성 세 손 수화 식 신 실 암 액 약 양 용 원 융 의 인 임 입 장 재 저 적 전 정 종 주 진 창 철 첩 청 초 심 침 태 투 패 편 포 풍 학 한 함 현 혹 혼 화 활 회 (NS) + 자위(SF)

가 가부 강 개 계 공 공판 과 관 광 극 기 김 남 내 냉 노 농 단 답 대 도 등 만 매 명 목 무 문 미 박 변 별 복 비 빗 선 선인 성 소 수 순 시 신 악 암 양 연 연하 염 영 용 원 위 의 이 인 입 저 전 정 조 주 차 천 초 총 추 춘부 통 파 폐 포 학 함 합 해 현 화 회 훈 (NS) + 장이(SF)

경 논 분 전 정 투 (NS) + 쟁이(SF)

강 경 국 대 수 중 포 (NS) + 졸중(SF)

각 간 감 객 공 구 세 군 금 단 도 뒤 맥 명 밀 반 발 백 범 부 사 상 선 소 안 애 양 연 염 영 우 위 음 이 입 자 재 저 전 종 진 질 청 축 탈 터 토 패 편 포 폭 합 호 활 (NS) + 주부(SF)

각 간 감 객 공 구 세 군 금 단 도 뒤 맥 명 밀 반 발 백 범 부 사 상 선 소 안 애 양 연 염 영 우 위 음 이 입 자 재 저 전 종 진 질 청 축 탈 터 토 패 편 포 폭 합 호 활 (NS) + 주사(SF)

각 간 감 객 공 구 세 군 금 단 도 뒤 맥 명 밀 반 발 백 범 부 사 상 선 소 안 애 양 연 염 영 우 위 음 이 입 자 재 저 전 종 진 질 청 축 탈 터 토 패 편 포 폭 합 호 활 (NS) + 주의(SF)

가 가락 각 간 감 객 경 공 굴 금 기 낙 녹 논 농 단 대 도화 동 먹 명승 무 바라 반 방 백 벽 변 본 봉 부 분 비 사 선 소 수 실 악 양 연 영 예 요 용 우 세 음 의 이 인 입 자 장 저 적 전 정 종 주 죽 중 차 처 천 첩 초 촌 축 탈 탐 토 통 투 파 패 팬 페이 편 평 폐 함 해 현 화 황 휴간 (NS) + 지개(SF)

가 가락 각 간 감 객 경 공 굴 금 기 낙 녹 논 농 단 대 도화 동 먹 명승 무 바라 반 방 백 벽 변 본 봉 부 분 비 사 선 소 수 실 악 양 연 영 예 요 용 우 세 음 의 이 인 입 자 장 저 적 전 정 종 주 죽 중 차 처 천 첩 초 촌 축 탈 탐 토 통 투 파 패 팬 페이 편 평 폐 함 해 현 화 황 휴간 (NS) + 지방(SF)

가 강 공 대 알 진 (NS) + 짜기(SF)

가 강 공 대 알 진 (NS) + 짜리(SF)

감 골 기 녹 당 말 문 발 벌 불 수 시 식 애 연 원 자 전 정 좌 태 해 화 (NS) + 초리(SF)

대 불 살 성 식 악 우 절 촌 편 회 (NS) + 충이(SF)

가 물 경 골 국 극 기 김 난 날 납 대 도 동 등 망 명 방 배 백 불 비 사 삼 상 새 설 소 안 양 염 영 예 운 위 의 이 인 자 자갈 잔 장 재 전 정 조 중 차 참 천 통 폐 풍 합 (NS) + 치기(SF)

강 격 기 낙 난 배 안 연 이 자 장 질 (NS) + 타리(SF)

가 간 개 계 골 공 관 골 내 도 들 무 변 불 사 상 소 염 요 융 전 정 진 체 총 호 화 (NS) + 통수(SF)
가 간 경 구 미 국 기 노 문 방 백 변 부 비 산 수 신 암 애 약 우 원 절 점 참 토 포 표 (NS) + 호지(SF)
부 사 자 폐 (NS) + 활량(SF)
깨 다 래 도 미 이 장 조 토 (NS) + 끼리(SF)
사 이 수 (NS) + 다귀(SF)
사 이 수 (NS) + 다리(SF)
무 (NS) + 덤불(SF)
가 운 (NS) + 데기(SF)
가 귀 녹 대 만 박 백 부 선 소 수 쌍 양 연 염 우 운 이 자 절 종 초 태 호 화 황 (NS) + 두리(SF)
기 무 천 (NS) + 둥이(SF)
기 무 천 (NS) + 둥치(SF)
감 강 균 급 미 분 비 상 소 열 차 평 폭 (NS) + 등신(SF)
때 본 (NS) + 때기(SF)
가 리 감 강 경 기 낙 대 도 마 목 무 병 부 수 안 애 얼 연 우 이 장 처 천 테 해 (NS) + 마루(SF)
유 우 타 이 (NS) + 머리(SF)
과 낙 상 정 (NS) + 오리 (SF)
반 사 장 폭 (NS) + 죽지(SF)
가 가 락 각 간 감 객 경 공 굴 금 기 낙 녹 논 농 단 대 도 화 동 먹 명 승 무 바 라 반 방 백 벽 변 본 봉 부 분 비 사 선 소 수 실 악 양 연 영 예 요 용 우 세 음 의 이 인 입 자 장 저 적 전 정 종 주 죽 중 차 처 천 첩 초 촌 축 탈 탐 토 통 투 파 패 팬 페 이 편 평 폐 함 해 현 화 황 휴 간 (NS) + 지기(SF)
가 별 추 (NS) + 첨지(SF)
조 (NS) + 카락(SF)
건 과 도 동 망 명 모 배 변 사 상 세 수 용 자 추 해 (NS) + 태기(SF)
대 방 부 성 실 연 위 재 쟁 전 참 호 흥 (NS) + 패기(SF)
대 빵 애 (NS) + 꾸머리(SF)
경 광 극 기 냉 대 독 등 막 무 박 부 분 사 상 선 성 세 소 순 식 악 안 연 우 의 인 임 입 장 적 전 절 접 종 주 중 차 천 초 추 축 침 편 포 피 학 한 함 허 위 현 화 (NS) + 대가리(SF)
노 대 도 명 물 불 사 선 쇠 신 실 열 요 욕 원 인 전 절 조 책 측 축 패 피 흥 (NS) + 망태기(SF)
유 우 타 이 (NS) + 머거리(SF)
가 가 락 각 간 감 객 경 공 굴 금 기 낙 녹 논 농 단 대 도 화 동 먹 명 승 무 바 라 반 방 백 벽 변 본 봉 부 분 비 사 선 소 수 실 악 양 연 영 예 요 용 우 세 음 의 이 인 입 자 장 저 적 전 정 종 주 죽 중 차 처 천 첩 초 촌 축 탈 탐 토 통 투 파 패 팬 페 이 편 평 폐 함 해 현 화 황 휴 간 (NS) + 지거리(SF)
고주 (NS) + 망태기(SF)
방 (NS) + 부성이(SF)

ANNEXE C-2. Désambiguïsation entre NS et SF pour les données de P2.

NS - SF	Etiquetage correct	NS - SF	Etiquetage correct
첨가량(S)		사가리(S)	
초가량(S)		대거리(S)	
추가량(S)		등거리(S)	
대가리(S)		소거리(S)	
비가리(S)		전거리(S)	
개구녁	- 구녁(SF)	공구리	- 구리(SF)
문구녁	- 구녁(SF)	내구리(S)	
안구녁	- 구녁(SF)	방구리(S)	
창구녁	- 구녁(SF)	자구리(S)	
개구리(S)		투구리(S)	
청구원	- 원(SF)	화구원(S)	
연구원(S)		감내기(S)	
화구원(S)	- 원(SF)	사내기	
감내기(S)		장내기(S)	
화구원(S)		극대기	-기(SF)
기대기	-기(SF)	막대기(S)	-기(SF)
냉대기	-기(SF)	부대기(S)	-기(SF)
대대기	-기(SF)	분대기	-기(SF)
독대기	-기(SF)	소대기(S)	-기(SF)
등대기	-기(SF)	연대기(S)	-기(SF)
입대기	-기(SF)	현대기	
중대기	-기(SF)	가도리(S)	
편대기(S)	-기(SF)	당도리(S)	
포대기(S)	-기(SF)	신도리(S)	
함대기(S)	-기(SF)	장도리(S)	
중도리(S)		선동이(S)	
중도리(S)		수동이(S)	
변통수(S)		악동이	-이(SF)
방호지		난동자	-자(SF)
무동이	-이(SF)	노동자(S)	-자(SF)
반동자(S)	-자(SF)	진동자(S)	-자(SF)
변동자	-자(SF)	찬동자(S)	-자(SF)
선동자(S)	-자(SF)	책동자(S)	-자(SF)
이동자	-자(SF)	폭동자(S)	-자(SF)
주동자(S)	-자(SF)	피동자	-자(SF)

NS - SF	Etiquetage correct	NS - SF	Etiquetage correct
부동자(S)	-자(SF)	결박질	-질(SF)
수동자(S)	-자(SF)	논박질	-질(SF)
전동자(S)	-자(SF)	도박질	-질(SF)
소박이(S)		두레박질	-질(SF)
정박이	-이(SF)	선박질(S)	
가방면(S)		매부리	
이방면	-방면(SF)	가부리(S)	
시정배기	-배기(SF)	발부리(S)	
집배기	-기(SF)	광부림	
도배기(S)		기부지	-지(SF)
납부囚	-지(SF)	각서리	
명부지(S)	-지(SF)	각서방	-서방(SF)
발부지	-지(SF)	독서방	-방(SF)
자부지(S)	-지(SF)	비서방	-방(SF)
갱생원		대서방(S)	-방(SF)
전서방(S)		수호지(S)	
정서방(S)		반신료(S)	-료(SF)
독선생(S)		발신료	-료(SF)
갱신료	-료(SF)	전신료(S)	-료(SF)
산호지(S)		추신료	-료(SF)
통신료	-료(SF)	소이전(S)	
투신료	-료(SF)	물자위(S)	
회신료	-료(SF)	가장이(S)	-이(SF)
놀이개	-개(SF)	만장이(S)	-이(SF)
동이개	-개(SF)	목장이(S)	-이(SF)
미장이(S)	-이(SF)	종주부(S)	
염장이 (S)	-이(SF)	도주사(S)	
통장이(S)	-이(SF)	진주사(S)	
경쟁이(S)	-이(SF)	부주의	
정쟁이(S)	-이(SF)	각지개	
감지개	-개(SF)	저지방	
기지개(S)		가지방(S)	
무지개		함지방(S)	
자지개(S)		강짜기	
각지방(S)		말초리	-초리(SF)

NS - SF	Etiquetage correct	NS - SF	Etiquetage correct
문초리	-초리(SF)	소치기	
식충이(S)	-이(SF)	양치기(S)	-치기(SF)
날치기(S)	-치기(SF)	자치기(S)	-치기(SF)
등치기(S)	-치기(SF)	재치기(S)	
배치기(S)	-치기(SF)	차치기	-치기(SF)
합치기		계통수(S)	
백치기(S)		폐활량(S)	
새치기(S)			
장치기(S)			
골통수			

ANNEXE D. La list de des-ambiguïtés P1 **entre** SF et PPN

Les données suivantes sont extraits manuellement à partir de la liste de l'ambiguï té P1 entre SF et PPN. Elles sont marquées comme NS-SF.

가공가 각본가 각색가 감독가 감상가 감시가 감식가 감정가 강가 개량가 개발가 개사가 개선가
개울가 개척가 개혁가 건축가 검사가 검색가 검술가 검시가 검안가 검역가 검열가 검증가
검침가 격양가 경력가 경매가 경연가 경영가 경작가 경제가 경험가 경호가 계략가 계몽가
계약가 계획가 고문가 고시가 고용가 고유가 고전가 고지가 곡예가 공격가 공급가 공론가
공매가 공명가
공모가 공사가 공상가 공시가 공약가 공업가 공연가 공예가 공장가 공정가 공증가 공지가
공출가 관상가 관찰가 관청가 관측가 광고가 광신가 광업가 괴력가 교역가 교육가 교정가
교차가 교화가 교훈가 구매가 구원가 구입가 구제가 구조가 구직가 구축가 구출가 국사가
국악가 국역가 국졸가 국학가 군가 권농가 권력가 권세가 권투가 궤변가 균열가 극작가 금연가
금욕가 금융가 기고가
기공가 기교가 기술가 기업가 기예가 기입가 기준가 기획가 길가 끽연가 낙농가 낙찰가 낙향가
난동가 납세가 납품가 내정가 냉혈가 노동가 노력가 노예가 노획가 농경가 농민가 농성가 눈가
능력가 다식가 다작가 달변가 답가 대농가 대변가 대소가 대역가 대체가 대출가 대필가 대학가
대행가 도덕가 도매가 도박가 도배가 도벌가 도산가 도식가 도심가 도안가 도입가 도학가
독립가
독살가 독선가 독식가 독재가 독점가 독주가 독학가 동결가 동맹가 동심가 등반가 마술가
만담가 만화가 망명가 망상가 매각가 매매가 매수가 매입가 면담가 멸문가 명가 명망가 명문가
명상가 명언가 명창가 명필가 모반가 모사가 모색가 모시가 모험가 목동가 목축가 못가 몽상가
무당가 무도가 무명가 무산가 무속가 무술가 무심가 무역가 무용가 무장가 문벌가 문예가
문장가 문필가
문학가 미신가 미장가 미필가 밀매가 박제가 반역가 반출가 발권가 발명가 발주가 발행가
방범가 방어가 방청가 방탕가 배급가 배당가 배상가 백가 번역가 벌목가 법가 변론가 변절가
병가 보급가 보상가 보유가 복사가 봉인가 봉합가 부대가 부력가 부속가 부식가 부양가 분석가
분쇄가 분양가 분업가 분절가 불가 불교가 불참가 불평가 불허가 비원가 비평가 비행가 빈농가
사대가
사랑가 사발가 사상가 사색가 사업가 사육가 사제가 사진가 사창가 사친가 사학가 사행가 삼가
삽가 삽화가 상담가 상속가 상식가 상업가 상점가 상품가 생명가 생산가 서도가 서예가 서화가
석별가 선동가 선지가 선험가 설계가 섭외가 성가 성악가 성조가 세도가 세력가 소국가 소농가
소설가 소액가 소작가 송별가 송환가 쇼핑가 수가 수거가 수색가 수송가 수심가 수완가 수용가
수입가 수집가 수채화가 수출가 수탈가 수평가 수필가 수혈가 숙련가 순시가 순찰가 시술가
시중가 시추가 시판가 식당가 실력가 실무가 실습가 실업가 실종가 실지가 실질가 실천가 안가
안무가 암벽가 암흑가 애견가 애국가 애도가 애연가 애조가 애주가 애창가 애향가 애호가
액션가 야심가 약정가 양도가 양돈가 양봉가 양산가 양상가 양생가 양식가 어부가 언론가
여행가 역사가 역설가
연구가 연락가 연설가 연쇄가 연주가 연출가 영농가 영양가 영업가 예술가 예시가 예언가
예지가 예탁가 예편가 완력가 왕가 외환가 요람가 ㅇ가 ㅇ물가 운동가 운반가 운송가 웅변가
원예가 원자가 원작가 원정가 ㅇ람가 ㅇ랑가 ㅇ림가 ㅇ통가 ㅇ행가 육상가 육식가 윤락가
은행가 음양가 응결가 응원가 의술가 의창가 이가 이농가 이론가 이별가 이주가 인권가 일가
임대가 입양가
입주가 입찰가 입하가 자력가 자본가 자산가 자선가 자작가 자찬가 자탄가 작명가 작사가
장식가 재력가 재봉가 재산가 재상가 재정가 저력가 저술가 저작가 적가 적정가 전도가 전략가
전문가 전범가 전술가 전용가 절약가 접견가 정력가 정상가 정열가 정찰가 정책가 정치가
정탐가 제련가 제작가 제조가 조각가 조난가 조력가 조언가 종교가 종횡가 주문가 주술가
주연가 주제가 주택가
중심가 증권가 증산가 지도가 지식가 지정가 지평가 직류가 직행가 진료가 집필가 찬미가
찬불가 찬송가 참전가 창업가 창작가 채권가 채무가 채식가 채집가 채취가 책가 책동가 책략가
처가 처분가 처사가 천황가 철거가 철학가 첩보가 청춘가 청취가 초빙가 초심가 촉망가 총가
총원가 최고가 최신가 최저가 추도가 추산가 추정가 축조가 출고가 출발가 출입가 출자가
출하가 취득가
침몰가 컬러가 쾌속가 탁송가 탈출가 탐방가 탐사가 탐색가 탐정가 탐험가 탑승가 택지가
통과가 통상가 통솔가 통신가 통역가 통제가 투시가 투자가 특급가 특매가 특지가 판매가

편도가 편력가 평행가 폐품가 포획가 폭등가 폭락가 표류가 표준가 풍류가 풍수가 피난가 피랍가 하늘가 하역가 하천가 학가 학도가 학원가 한복가 한정가 할인가 합작가 합창가 항해가 해변가 해설가 해안가
해학가 행상가 행진가 향락가 혁명가 현지가 현학가 협상가 협잡가 협정가 호객가 호사가 호송가 호식가 화전가 확정가 환락가 환송가 환전가 활동가 활용가 황가 황조가 회가 회도가 흡연가

가공과 가물치과 가상과 가스과 가오리과 가자미과 가재과 가전학과 가정과 가지과 각과 간호과 갈매기과 갈치과 감리과 감사과 감시과 감식과 감찰과 개구리과 개미과 개발과 거머리과 거미과 거위과 건과 건설과 건축과 검사과 검색과 검시과 검역과 검열과 검증과 검찰과 검침과 겨자과 격과 경제과 계획과 고고학과 고등어과 고사리과 고슴도치과 고양이과 고용과 곰과 공매과 공모과 공미리과 공보과
공시과 공업과 공연과 공예과 공증과 관과 구매과 국문과 국사과 국세과 국악과 국어과 국학과 국화과 굴과 귀뚜라미과 귤과 기계과 기린과 기상과 기술과 기획과 까마귀과 꼴뚜기과 꽁치과 꾀꼬리과 꿩과 끈끈이과 나리과 나방과 나비과 낙과 낙지과 낙타과 날치과 남생이과 내과 내외과 너구리과 넙치과 논과 농공과 농과 느타리과 다람쥐과 다래과 다랭이과 다슬기과 다시마과 단과 단풍과 단화과
달팽이과 담당과 대구과 대승과 대출과 도과 도라지과 도롱뇽과 도루묵과 도미과 독수리과 돌고래과 두꺼비과 두더지과 두루미과 등에과 따개비과 따오기과 딱다구리과 뜸부기과 마취과 망과 망둥이과 매과 매미과 맹꽁이과 멍게과 메기과 메뚜기과 멧돼지과 멸구과 멸치과 명과 모기과 모델과 목련과 무과 무궁화과 무역과 무용과 문과 문란과 문리과 문어과 문학과 물리과 물상과 미나리과 미역과 민어과 밀과 밀봉과
발권과 백로과 백부과 백사과 백합과 뱀과 뱅어과 벌레과 벌목과 법과 베짱이과 벼과 벼룩과 별과 병어과 보육과 보장과 복과 본과 봉선화과 부사과 부유과 부인과 부평초과 분과 불가사리과 불과 불문과 비뇨기과 비둘기과 비어과 비자과 비취과 빈대과 빈사과 사마귀과 사생과 사슴과 사학과 사회과 산과 산실과 산업과 산호과 살모사과 살무사과 삼치과 생강과 생원과 서무과 석류과 석이과
선과 선인장과 성악과 세제과 소과 소라과 소아과 송백과 송사리과 송이과 수공과 수국과 수리과 수선과 수선화과 수의과 수학과 숭어과 신경과 실과 실업과 쑥과 아귀과 아욱과 악어과 안과 야자과 양귀비과 양미리과 양조과 여치과 연수과 연어과 연작과 영문과 영사과 영어과 예과 예능과 예술과 오랑우탄과 오리과 오미자과 오징어과 옥돔과 올빼미과 외래과 원숭이과 위과 응서과 응인원과 응화과
윤리과 음양과 의과 이화 인문과 인사과 잉어과 자라과 자본과 자연과 자유과 잠자리과 장과 장미과 전갈과 전갱이과 전문과 전자과 정보과 정신과 정치과 조과 조사과 조선과 족제비과 종다리과 종사과 종친과 주목과 지네과 지빠귀과 진과 진달래과 진드기과 진디과 진료과 진사과 진찰과 질경이과 찌르레기과 참벌과 철쭉과 철학과 첨단과 청각과 청어과 청정과 체육과 촉수과 촌충과 총무과 칠면조과
칠엽수과 카멜레온과 캥거루우과 코브라과 콩과 타조과 토끼과 토란과 토목과 통계과 통신과 통역과 투어과 파래과 파리과 파인애플과 판매과 펭귄과 편충과 포도과 표피과 풍뎅이과 피부과 하마과 학과 학무과 학생과 한의과 해과 해동과 해사과 해양과 해파리과 행정과 향과 홍어과 홍합과 화공과 화학과 환경과 회양목과

가공관 가스관 가정관 가족관 가치관 감독관 감리관 감사관 감시관 감식관 감지관 감찰관 강관 개봉관 개선관 개원관 개찰관 개표관 검사관 검색관 검시관 검안관 검역관 검열관 검인관 검증관 검찰관 검침관 검토관 결재관 결혼관 경리관 경연관 경영관 경찰관 계몽관 계수관 계엄관 계획관 고무관 고문관 고시관 고압관 고용관 공매관 공보관 공사관 공시관 공업관 공예관 과학관
관계관 관리관 관찰관 관측관 관할관 교도관 교련관 교양관 교육관 교정관 교직관 교화관 구리관 국가관 국사관 군관 굴관 굴착관 귀관 귀빈관 근로관 금관 급수관 급유관 기구관 기념관 기록관 기상관 기술관 기업관 기포관 기획관 나무관 나팔관 내관 내세관 냉각관 냉관 냉수관 네온관 노관 노무관 노예관 뇌관 달팽이관 담당관 대독관 대사관 대청관 도덕관 도서관 도장관
독립관 돌관 동관 동맥관 드럼관 매관 매몰관 매설관 메시아관 명령관 명사관 목관 목사관 목회관 무관 무역관 문무관 문서관 문정관 문학관 문화관 미관 미래관 민족관 밀관 박물관 반응관 발광관 발사관 발행관 방역관 방전관 배관 배기관 배설관 배수관 배심관 배전관 배출관 번역관 법관 변조관 변환관 병관 병기관 보급관 보도관 보안관 보좌관 보호관 복합관 봉관
부검관 부모관 부부관 분류관 분수관 불교관 비닐관 비서관 사교관 사랑관 사령관 사료관 사무관 사물관 사생관 사서관 사업관 사진관 사회관 산란관 삼림관 상무관 상설관 상연관

상영관 생명관 생사관 생활관 서기관 서무관 석공관 선관 선교관 선전관 선출관 세계관 세관 소기관 소화관 송수관 송신관 송풍관 수도관 수령관 수배관 수사관 수상관 수생관 수석관 수세관 수송관 수술관
수신관 수압관 수정관 수족관 수직관 수축관 수평관 수학관 시공관 시국관 시찰관 시추관 시험관 신경관 신기관 신문관 신부관 신앙관 심사관 심의관 심판관 아동관 언론관 언어관 에너지관 역사관 연기관 연무관 연방관 연성관 연소관 연수관 연예관 열관 영빈관 영사관 영화관 예관 예술관 오줌관 온수관 올리브관 왕관 외교관 요리관 우주관 운동관 운용관 원소관 월동관 위관 응교관
유도관 응리관 응물관 응수관 응압관 육상관 윤리관 의정관 이관 이발관 이재관 인간관 입문관 자녀관 짜아관 응자연관 응자위관 응자치관 장려관 재무관 재판관 전도관 전령관 전시관 전자관 전쟁관 전조관 전통관 정류관 정무관 정보관 정부관 정세관 정압관 정치관 조각관 조사관 종교관 종말관 종사관 주무관 주임관 증기관 증발관 지구관 지방관 지하관 지휘관 직업관 진공관 진관 진리관
진찰관 질서관 집사관 집정관 집행관 징수관 참모관 참사관 처세관 척수관 척추관 천황관 철관 철인관 철학관 체육관 총독관 총무관 총재관 추기관 추수관 축적관 춘추관 출납관 충전관 취조관 측후관 타인관 타작관 통기관 통역관 통제관 특약관 판무관 편관 편수관 평형관 하사관 하수관 학생관 행정관 향도관 허파관 현실관 홍보관 화분관 화학관 황동관 회화관 후생관 훈련관 흑인관
흡기관 흡수관 흡입관

그믐께 만조께 며칠께 명절께 무릎께 상순께 상오께 새벽께 여명께 오후께 저녁께 중순께 초순께

가계도 가공도 가상도 가속도 가입도 가청도 가치도 각도 각색도 감광도 감리도 감성도 감소도 감속도 감손도 감시도 감압도 감영도 감응도 감정도 감지도 강화도 개구도 개도 개략도 개산도 개요도 개폐도 개표도 갱도 건습도 건조도 검도 검사도 검시도 견적도 견제도 결빙도 결선도 결속도 결시도 결정도 결집도 결핍도 결함도 결합도 결항도 경기도 경사도 경상도 경위도
경작도 경직도 경험도 계도 계보도 계절도 계제도 계통도 계획도 고가도 고난도 고속도 고온도 고용도 고지도 고착도 골격도 공감도 공극도 공사도 공수도 공시도 공유도 공작도 공정도 공평도 공학도 과세도 과학도 관내도 관리도 관심도 관여도 관제도 관찰도 관측도 광산도 광속도 광신도 교역도 교차도 교환도 구교도 구비도 구상도 구제도 국사도 국선도 국학도 굴곡도 굴절도 궁도
근친도 기독교도 기마도 기명도 기본도 기사도 기상도 기여도 기초도 기호도 기후도 긴장도 난행도 남도 내도 내륙도 내수도 내장도 노력도 노송도 노자도 노화도 누진도 대륙도 대마도 대모도 대부도 대사도 대진도 대첩도 대청도 대칭도 대화도 독도 동요도 동위도 동인도 동학도 득도 등속도 마모도 마찰도 만족도 매설도 매화도 명시도 명제도 명중도 무녀도 무명도 무사도 무신도
무용도 무학도 문갑도 문학도 밀집도 반발도 반사도 반역도 반응도 발달도 발사도 발생도 발육도 발전도 발행도 발효도 방사도 방전도 배분도 배선도 배차도 배치도 법계도 법도 변성도 변이도 변천도 변화도 병풍도 보살도 보유도 보존도 보증도 복사도 복원도 부품도 부호도 분리도 분말도 분산도 분지도 분출도 분포도 분할도 분해도 분화도 불교도 불능도 비례도 비만도 사교도 사양도
사출도 산도 산호도 삽도 상상도 상태도 상황도 색도 생성도 생육도 생존도 서간도 서한도 선광도 선적도 설계도 설명도 설치도 성공도 성사도 성악도 성장도 성취도 세공도 수렵도 수사도 수심도 수용도 수직도 수행도 숙련도 숙성도 순서도 순위도 순응도 습작도 신교도 신뢰도 신망도 신빙도 신사도 신선도 신수도 신용도 신임도 신장도 실감도 실용도 실천도 실측도 실태도 심취도
악성도 안내도 안마도 안면도 안전도 안정도 암혈도 양산도 양화도 연결도 연마도 연화도 열전도 염기도 염도 염전도 영상도 영속도 영정도 영향도 오동도 오식도 오염도 완결도 완공도 완벽도 완성도 완숙도 왕도 용맹도 용해도 응산도 원산도 위험도 응료도 응인도 응적도 응전도 은유도 응고도 응집도 의존도 의학도 이교도 이용도 이해도 이행도 인기도 인물도 인식도 인지도 인체도
자립도 자유도 작약도 장애도 재배도 재생도 재현도 저속도 저온도 저위도 적응도 전개도 전도도 전라도 전염도 전위도 전이도 절연도 점도 점착도 접속도 정교도 정립도 정온도 정제도 정찰도 정합도 정확도 제련도 제어도 제작도 제주도 조각도 조감도 조도 조립도 조명도 종속도 죽도 준공도 중심도 중위도 중학도 증가도 증감도 증산도 지명도 지방도 지분도 지세도 지적도 지지도

지질도 지출도 지하도 직선도 진입도 진지도 진척도 진취도 집약도 집적도 집중도 찬성도 참가도 참여도 채광도 천마도 천이도 천주교도 철도 첨가도 첨도 첨삭도 청도 청렴도 청산도 청정도 청취도 청포도 체감도 초과도 초속도 촉도 최고도 축소도 축적도 축적도 축척도 춘화도 출현도 충성도 충전도 충족도 침습도 쾌속도 쾌적도 탄도 태권도 태도 통계도 투과도 투사도 투시도
투영도 파손도 판매도 팽창도 편광도 편리도 편재도 편집도 평안도 포화도 폭도 품계도 풍류도 풍속도 풍수도 풍화도 피로도 하수도 함유도 항공도 항성도 항해도 항행도 해류도 해부도 해안도 핵도 행선지도 향상도 허용도 호응도 혼잡도 혼합도 화랑도 활성도 활용도 회교도 회로도 회전도 흐름도 흡수도

가스로 간선로 감염로 감응로 강변로 개폐로 갱로 건조로 경사로 경수로 공격로 공항로 관통로 교역로 교차로 교통로 귀빈로 귀항로 귀향로 귀환로 급수로 급유로 기항로 내왕로 냉각로 농로 대상로 대학로 도입로 도주로 등교로 등반로 말로 매립로 매설로 무역로 반사로 반출로 발사로 방수로 방어로 배수로 보급로 보행로 분쇄로 산책로 소각로 수색로 수송로 순환로 용수로 용해로 우회로 운반로 운송로 원자로 유출로 융해로 이동로 저수로 전기로 전송로 전용로 전철로 전환로 접속로 주행로 중앙로 증식로 직선로 진입로 진출로 차단로 철로 추격로 탐험로 통상로 통신로 통풍로 통행로 통화로 퇴각로 투입로 판매로 하강로 하수로 항공로 항해로 해안로 행진로 호송로 활주로 회전로 횡단로

가상론 가설론 가속론 가입론 가정론 가족론 가치론 가톨릭론 각론 각색론 갈등론 감각론 감동론 감사론 감상론 감성론 감시론 감염론 감응론 감정론 감축론 갑론 강화론 개국론 개념론 개량론 개발론 개선론 개연론 개전론 개조론 개척론 개최론 개축론 개편론 개폐론 개항론 개혁론 개화론 객관론 거사론 거세론 거품론 건강론 건축론 게임론 격리론 격하론 견제론 견책론 결과론
결별론 결사론 결의론 결전론 결정론 결투론 결핍론 결혼론 겸용론 경각론 경계론 경영론 경쟁론 경제론 경질론 경합론 경험론 계몽론 계발론 계산론 계승론 계약론 계엄론 계책론 계통론 계획론 고갈론 고립론 고문론 고생론 고시론 고용론 고전론 고증론 고착론 고행론 공간론 공개론 공격론 공리론 공백론 공범론 공산론 공상론 공생론 공세론 공습론 공시론 공식론 공약론 공업론 공연론
공유론 공익론 공적론 공적론 공지론 공직론 공채론 공천론 공출론 공평론 공포론 공해론 공황론 과반론 과세론 과장론 관계론 관권론 관념론 관련론 관리론 관망론 관상론 관성론 관습론 관여론 관조론 관철론 관측론 광고론 광명론 광신론 교감론 교배론 교본론 교섭론 교양론 교역론 교육론 교재론 교전론 교차론 교체론 교화론 교환론 교훈론 구국론 구매론 구문론 구사론 구실론 구원론
구인론 구전론 구제론 구조론 구축론 구혼론 국가론 국권론 국론 국방론 국부론 국수론 국시론 국악론 국영론 국유론 국익론 국지전론 국책론 국학론 군론 군세론 군신론 군자론 군주론 군집론 권고론 권력론 권세론 권위론 궤도론 궤변론 귀국론 귀속론 귀양론 규약론 규정론 규제론 규탄론 균등론 균열론 균형론 극대론 극락론 극성론 극우론 극찬론 근본론 근절론 금기론 금융론 금주론
금지론 급등론 긍정론 기계론 기권론 기능론 기만론 기반론 기본론 기술론 기약론 기업론 기여론 기원론 기제론 기준론 기질론 기초론 기호론 기획론 길흉론 낙담론 낙인론 낙제론 낙지론 낙찰론 낙태론 납세론 납치론 납품론 낭만론 내각론 내세론 내조론 내통론 냉전론 노론 노소론 녹화론 농경론 단계론 단원론 단자론 단축론 달관론 달변론 담력론 담판론 담화론 당위론 대론 대망론
대부론 대비론 대상론 대세론 대승론 대응론 대전론 대책론 대처론 대체론 대출론 대치론 대표론 대필론 대학론 대화론 덕론 도교론 도덕론 도심론 도용론 도입론 도출론 도피론 독단론 독립론 독선론 독자론 독재론 독해론 동거론 동결론 동기론 동맹론 동심론 동의론 동정론 동조론 동참론 동향론 등극론 등정론 등제론 마녀론 마찰론 만전론 만족론 만찬론 말세론 망론 맹신론 맹아론
면역론 면책론 멸공론 명령론 명망론 명명론 명목론 명분론 명칭론 모반론 모방론 모사론 모색론 모순론 목사론 목적론 목차론 목회론 묘론 무론 무산론 무신론 무심론 무아론 무역론 무용론 무인론 무장론 무한론 무효론 묵비론 문단론 문명론 문신론 문예론 문자론 문장론 문제론 문체론 문학론 문헌론 문화론 물량론 물성론 물질론 미래론 미분론 미신론 미정론 미추론 민권론 민중론
박애론 박해론 반공론 반대론 반사론 반성론 반역론 반영론 반전론 반향론 발달론 발론 발명론 발생론 발원론 발전론 발췌론 방비론 방위론 방임론 방전론 방조론 방황론 배급론 배반론 배상론 배우론 배제론 배치론 번역론 범위론 범주론 법론 변동론 변별론 변신론 변증론 변혁론

병용론 병행론 보관론 보유론 보장론 보전론 보증론 보험론 보호론 본성론 본질론 본체론 부국론 부권론
부인론 부재론 부족론 분권론 분단론 분담론 분류론 분립론 분배론 분석론 분업론 분자론 분출론 분할론 분화론 불능론 불멸론 불사론 불성론 불식론 불신론 불용론 불찰론 불행론 붕괴론 비관론 비교론 비용론 비유론 비판론 비평론 사랑론 사분론 사유론 사행론 사회론 살해론 삼매론 상대론 상론 상식론 상조론 상징론 상환론 상황론 색론 생기론 생명론 생산론 생성론 생체론 선거론
선교론 선동론 선별론 선언론 선제론 선택론 선험론 설계론 설립론 섭식론 성격론 성결론 성과론 성불론 성쇠론 성장론 성화론 세계론 세례론 소각론 소론 소멸론 소설론 소아론 소작론 속결론 속도론 쇄국론 수거론 수기론 수도론 수반론 수성론 수습론 수양론 수업론 수용론 수지론 수출론 숙명론 숙청론 순리론 순수론 순화론 순환론 숭배론 승계론 시가론 시인론 시작론 시장론 시조론
식별론 신문론 신성론 신앙론 신의론 신장론 신정론 신중론 실각론 실재론 실적론 실존론 실증론 실질론 실천론 실체론 실학론 실험론 심사론 심성론 심신론 심판론 심화론 안전론 암행론 애환론 약식론 양생론 양심론 양자론 어원론 어음론 어의론 어휘론 언어론 언쟁론 에너지론 여권론 여심론 역모론 역설론 역적론 역학론 연결론 연구론 연기론 연대론 연맹론 연쇄론 연임론 영상론 영향론
예견론 예방론 예술론 예찬론 예치론 예편론 오성론 옹호론 와전론 왜곡론 외교론 외설론 외환론 요인론 욕구론 ㅇ경론 ㅇ연론 ㅇ위론 ㅇ주론 운동론 운송론 원리론 원자론 원정론 원조론 원칙론 위계론 ㅇ기론 ㅇ조론 ㅇ증론 ㅇ헌론 위험론 ㅇ기론 ㅇ명론 ㅇ무론 ㅇ문론 유물론 ㅇ배론 ㅇ상론 ㅇ신론 유실론 유아론 유용론 유임론 유지론 육감론 윤리론 율격론 은사론 은총론 은행론 은혜론
음기론 음양론 음주론 의거론 의무론 의미론 의식론 의지론 이간론 이기론 이데아론 이상론 이성론 이행론 인간론 인과론 인구론 인권론 인문론 인성론 인식론 인지론 일반론 임관론 임박론 임전론 임차론 입론 입자론 입지론 자각론 자격론 자만론 자멸론 자본론 자생론 자아론 자원론 자위론 자유론 자조론 자질론 자치론 작가론 작용론 작품론 장론 재건론 재량론 재무론 재벌론 재정론
재질론 재현론 적응론 전략론 전용론 전이론 전자론 전쟁론 전파론 절망론 절멸론 점유론 접합론 정리론 정벌론 정보론 정설론 정수론 정신론 정원론 정의론 정전론 정책론 정체론 정치론 정토론 정통론 제도론 제물론 제한론 조공론 조상론 조어론 조직론 조화론 족벌론 존재론 존폐론 졸업론 종교론 종말론 종속론 종식론 주권론 주기론 주리론 주술론 주유론 주의론 주자론 중용론 증감론
증거론 지대론 지원론 지진론 지행론 직관론 직립론 진리론 진압론 진화론 진흥론 질서론 집합론 징발론 착오론 찬론 참관론 참배론 참전론 참회론 창업론 창작론 창조론 책략론 책임론 처리론 처방론 척결론 천기론 천도론 천황론 철칙론 철학론 청춘론 청탁론 초월론 총장론 총체론 추계론 추대론 추방론 추이론 추정론 추측론 추태론 축적론 출가론 취입론 침략론 침묵론 침체론 쾌락론
타살론 타파론 탄생론 탄성론 탕감론 태몽론 토착론 통신론 통제론 통합론 통화론 퇴적론 투시론 투여론 투자론 투쟁론 특혜론 파경론 파국론 파급론 파면론 파병론 파악론 파업론 판공론 패망론 편재론 평등론 평론 평행론 평화론 폐간론 폐기론 폐모론 폐지론 포장론 폭격론 폭등론 폭락론 폭론 폭발론 표절론 표현론 품사론 풍자론 필연론 학대론 학문론 학벌론 학습론 학원론 한계론
한방론 할당론 함몰론 함수론 합류론 합명론 합의론 항변론 항해론 해방론 해빙론 해석론 해운론 해체론 행동론 행복론 행위론 행정론 행태론 허가론 허무론 헌신론 헤게모니론 혁명론 혁신론 현상론 현실론 현장론 현정론 협력론 형성론 형식론 혼돈론 화공론 화두론 화성론 화용론 화폐론 화해론 확률론 환경론 환매론 활용론 회계론 회상론 회의론 회화론 후보론 휴전론 흥망론 흥미론 흥행론
희곡론 희망론

강화만 경기만 고성만 동경만 임기만 진주만 진해만 하루만

가문서 가입서 가톨릭서 간행서 감리서 감정서 감축서 개론서 개봉서 건의서 검사서 검안서 검역서 검인서 검증서 검침서 검토서 견본서 견적서 결말서 결산서 결서 결약서 결의서 결재서 결정서 경력서 경찰서 계량서 계몽서 계발서 계산서 계시서 계약서 계율서 계통서 계획서 고문서 고발서 고백서 고용서 고증서 고지서 고찰서 공매서 공문서 공서 공소서 공시서 공식서 공약서
공인서 공증서 공탁서 과세서 관찰서 관측서 관할서 교과서 교리서 교본서 교부서 교양서 교역서 교화서 교훈서 구독서 구매서 구문서 구약서 구원서 구인서 구입서 구혼서 국서 권고서

권두서 궤변서 규약서 규제서 규탄서 급여서 기결서 기도서 기록서 기명서 기밀서 기별서
기부서 기상서 기술서 기원서 기탄서 기획서 낙찰서 날조서 납세서 납품서 내규서 내력서
내부서 내통서 논문서
논술서 논증서 농업서 다짐서 담론서 담화서 답변서 답서 대관서 대변서 대본서 대부서 대여서
대작서 대출서 대필서 대화서 도교서 도덕서 도식서 도안서 도용서 도입서 도출서 독립서
독학서 독해서 동맹서 동봉서 동의서 동정서 동참서 동행서 동향서 등극서 등록서 매각서
매립서 매매서 매입서 매출서 면담서 명령서 명명서 명언서 명작서 명저서 모범서 무용서
문명서 문학서 물품서
미결서 미필서 민권서 밀매서 반박서 반출서 발간서 발권서 발달서 발매서 발명서 발부서
발언서 발의서 발행서 방언서 배급서 백서 번역서 법문서 변명서 변천서 변환서 병력서 보고서
보증서 보충서 보험서 복원서 봉인서 봉합서 분석서 비결서 비방서 비준서 비평서 사망서
사무서 사문서 사양서 사업서 사용서 사유서 사의서 사직서 사퇴서 상속서 서식서 서약서
선고서 선언서 선적서
설계서 설명서 설문서 설정서 성명서 성문서 송달서 수령서 수신서 수취서 습작서 승낙서
승인서 식별서 신고서 신앙서 실태서 심리서 심문서 심의서 악필서 안경서 안내서 약정서
언약서 역사서 역순서 역학서 연감서 연구서 연작서 연재서 열람서 영문서 예산서 예언서
외교서 요구서 원류서 원작서 위문서 위탁서 유언서 유통서 의견서 의결서 의뢰서 의정서
의학서 의향서 이력서 이론서
이별서 이유서 이직서 이혼서 익명서 인가서 인계서 인권서 인수서 인용서 인정서 인준서
인증서 인출서 임명서 임용서 입문서 입증서 입찰서 자백서 자습서 자필서 작명서 잠언서 저서
저술서 적요서 전단서 전도서 전말서 전출서 접견서 정산서 정찰서 정책서 제공서 제의서
제작서 제출서 조사서 조제서 조퇴서 조회서 주문서 주서 주술서 증명서 증빙서 증여서 증정서
지급서 지령서
지분서 지시서 지원서 지참서 지출서 지침서 진단서 진료서 진술서 진찰서 진필서 질문서
질의서 집필서 차압서 착공서 참가서 참고서 참전서 창업서 채용서 처방서 철학서 청구서
청문서 청산서 청원서 청취서 체결서 초심서 초판서 촉탁서 총람서 총론서 추대서 추천서
출간서 출판서 취임서 취하서 친필서 타결서 타협서 탄원서 탐구서 탐사서 토론서 통계서
통고서 통보서 통신서 통지서
퇴직서 특집서 파면서 판결서 판독서 판매서 판정서 편찬서 평서 포고서 포기서 필독서 하달서
학술서 학습서 한문서 합격서 합의서 항의서 항해서 해명서 해몽서 해부서 해설서 해임서
허가서 허락서 헌혈서 혁명서 현황서 협상서 협약서 협의서 협정서 형식서 호명서 호출서
혼약서 혼인서 화혼서 확약서 확인서 확정서 회고서 회의서 휴가서 휴직서

가공선 가상선 가슴선 가시선 각선 간조선 간첩선 갈대선 감독선 감마선 감선 감시선 강선
개량선 갱선 거래선 거북선 건축선 검사선 검역선 격리선 견사선 결속선 결승선 결합선 경계선
경기선 경도선 경보선 경비선 경선 경위선 경의선 경전선 경호선 계류선 계선 계약선 고가선
고대선 고속선 고압선 고조선 곡선 공격선 공공선 공작선 공중선 관광선 관측선 광선
교역선 교외선 교정선 교차선 교환선 구강선 구두선 구리선 구분선 구원선 구제선 구조선
구출선 구호선 국경선 국내선 군선 굴곡선 귀국선 귀선 귀항선 귀향선 귀환선 균열선 극선
급수선 급유선 급전선 급행선 기준선 기중기선 기함선 기획선 까치선 나사선 낙도선 낙하선
난파선 남포선 내수선 내항선 냉동선 냉장선 노선 노예선 노획선 단선 단속선 단절선 대각선
대동선 대륜선
대선 대장선 대칭선 대피선 도선 도안선 도입선 도착선 도킹선 도화선 돌파선 동력선 동문선
동선 둘레선 등록선 등분선 등속선 마감선 만선 만원선 만조선 망명선 매각선 매매선 명시선
모선 목선 목제선 무선 무역선 무임선 무장선 밀매선 밀수선 밀어선 밀항선 바지선 반품선
발동선 발명선 발육선 발전선 발주선 방범선 방비선 방사선 방어선 방위선 방적선 방패선
방화선
배급선 배전선 배치선 버스선 벌목선 범선 변경선 보고선 보급선 보물선 보유선 보조선 보초선
복사선 봉쇄선 봉인선 봉합선 봉황선 부력선 부설선 부속선 부식선 분계선 분기선 분단선
분비선 분쇄선 분할선 비상선 비행선 사령선 상승선 상품선 상하선 생명선 생식선 생장선
설계선 성장선 세포선 소화선 속선 송전선 송환선 수거선 수리선 수비선 수상선 수색선 수송선
수인선 수입선
수직선 수차선 수출선 수평선 수호선 순라선 순시선 순찰선 순환선 승객선 승하선 시발선
시추선 시험선 식량선 신간선 신흥선 실선 실습선 실종선 실험선 심해선 안전선 안테나선
알루미늄선 암선 에나멜선 여객선 연결선 연락선 연속선 연습선 연장선 연해선 열선 영업선
영해선 예방선 예비선 완성선 왕복선 왜구선 외곽선 외항선 우송선 우수선 우승선 우주선
우편선 우회선 운반선 운송선

운항선 운행선 원자선 원정선 위도선 ㅇ람선 ㅇ물선 ㅇ방선 ㅇ치선 윤곽선 융기선 이민선
이온선 이주선 이중선 이하선 인출선 일ㅈ선 임신ㅈ선 입사선 입양선 입자선 입하선 자력선 자매선
잡역선 장선 재단선 재봉선 재수선 저속선 저압선 저지선 저항선 전기선 전라선 전력선 전류선
전선 전송선 전신선 전압선 전용선 전위선 전이선 전자선 전철선 전투선 전화선 절개선 절선
절연선
점선 접경선 접선 접속선 접착선 접촉선 정박선 정지선 정찰선 정탐선 제련선 제어선 제작선
조난선 조사선 조준선 종선 주사선 주유선 주파선 주행선 중간선 중심선 중앙선 증기선 지도선
지평선 직간선 직류선 직행선 진료선 진주선 진해선 차단선 차양선 착륙선 참치선 채집선
채취선 천이선 철거선 첩보선 초선 촉각선 최고선 최상선 최신선 최전선 추적선 출발선 충적선
측후선
침몰선 컬러선 쾌속선 탁송선 탄소선 탈선 탈출선 탈피선 탐방선 탐사선 탐색선 탐지선 탐험선
탑승선 토사선 통과선 통신선 통운선 통제선 통행선 통화선 투사선 투시선 투영선 특급선
판매선 판목선 편도선 평행선 평화선 폐색선 폐품선 포경선 포선 포위선 포획선 폭등선 표류선
표선 표적선 피난선 피랍선 피복선 피부선 피아노선 하역선 하행선 한계선 한정선 함몰선 합선
합작선
합판선 항공선 항로선 항해선 해골선 해안선 해적선 해체선 해태선 허리선 호르몬선 호송선
호위선 혼선 혼용선 화물선 화학선 활개선 회귀선 회유선 회조선 후송선 훈련선 휴전선 흉선
흐름선 흡수선

건아 고립아 기린아 냉혈아 만삭아 망아 문란아 문제아 반항아 방탕아 불구아 비만아 사산아
성숙아 야생아 양아 ㅇ랑아 장애아 재능아 정박아 정상아 지체아 천재아 촉망아 출산아 출생아
태아 태아 패륜아 풍운아 학대아 혼혈아

가결의 개업의 개원의 검역의 무성의 소아의 수련의 실습의 양의 임상의 전문의 접골의 지구의
천구의 측정의 한방의

간난이 거북이 검댕이 겨드랑이 깜장이 껌정이 누렁이 능청이 맹추이 빨강이 식충이 얼룩이
염장이 우렁이 우롱이 이월이 지은이 황진이

ANNEXE E. Classes de SFR dans le verbe

CLASSE	SFR	Information
1 감다	감	SU3 : 1 2 3 4 5 7 8 12 13 14 17 18 23 24 25 26
		SU1 : 5
		SU2 : 1 2 3 4 5
		SU4 : 1
		PV1 : 1 3
		PV3 : 1
		PV4 : 1
2 녹다	녹	SU3 : 1 2 3 4 5 7 8 12 13 14 17 18 23 24 26
		SU1 : 5
		SU2 : 1 2 3 4 5
		SU4 : 1
		PV1 : 1 3
		PV3 : 1
3 익다	익	SU3 : 1 2 4 5 7 8 9 10 12 13 14 15 17 21 23 24
		SU2 : 1 2 3 4 5
		SU4 : 1
		PV4 : 1
		SU1 : 5 6
		PV1 : 1 3
4 듣다	듣	SU3 : 1 2 4 5 7 8 9 10 12 13 14 15 17 21 23 24
		SU2 : 1
		SU4 : 1
		SU1 : 5 6
		PV1 : 1 3
	들	SU2 : 2 3 4 5
		PV4 : 1
5 곱다	고	PV6 : 1 2 3 5 6
		PV8 : 1
	곱	SU1 : 4 5
		SU3 : 1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18 26
		SU2 : 1
		SU4 : 1
		PV1 : 1
6 옵다	오	SU3 : 8 12
		SU1 : 5 6 9 15 16 17 18
		PV1 : 1
	옵	SU3 : 1 2 4 5 7 14 17 18 21 23 26
		SU1 : 6 9 13
		SU4 : 1
		SU2 : 3 4 5
7 갈다	가	SU3 : 8 9 12 13 20
		SU1 : 5 6 12 13 14 15 16 17 18
		PV1 : 1 3
	간	SU1 : 1
	갈	SU1 : 2 9 10 11
		SU2 : 2 3 4 5
		SU3 : 1 2 3 4 5 6 7 14 15 16 17 18 23 24 26
		PV3 : 1
		PV7 : 1 2

8 낳다 나 SU3 : 1 4 7 8
SU1 : 5 6 11 12 13 15 16 17 18 19
PV1 : 1 7 8 9
PV3 : 1
난 SU1 : 1
날 SU1 : 2
낳 SU3 : 1 2 3 4 5 7 8 9 12 13 14 15 17 18 21 23 26
SU1 : 5 6
SU2 : 1 2 3 4 5
SU4 : 1
PV1 : 1 3 8 9
PV3 : 1

9 낫다 나 SU1 : 6 7 12 13 15 16 17 18
PV3 : 1
SU3 : 4 5 7 17 18
PV1 : 7
난 SU1 : 1
날 SU1 : 2
낫 SU3 : 1 2 3 4 5 6 7 8 9 10 12 13 14 15 17 18 26
SU1 : 5
SU2 : 1 2 4 5
SU4 : 1
PV1 : 1 7

10 뢰다 뢴 SU1 : 1
뢸 SU1 : 2
뢰 SU3 : 1 2 4 5 7 8 12 13 17
SU1 : 5 6 7 8 9 15 16 17 18
PV4 : 1
PV1 : 1 3
PV7 : 1

11 걸다 건 SU1 : 1
거 SU3 : 8 12 13 20
SU1 : 5 6 12 15 16 17 18
PV1 : 1 2 3
걸 SU1 : 2 7 8 9 10 11
SU2 : 2 3 4 5
SU3 : 1 2 3 4 5 7 14 15 17 18 23 26
PV1 : 5
PV4 : 1
PV7 : 1

12 숫다 순 SU1 : 1
술 SU1 : 2
숫 SU3 : 1 2 3 4 5 7 8 12 13 14 15 17 18 19
SU2 : 1
SU4 : 1
SU1 : 5 6
PV1 : 1 3
수 SU1 : 9 19 11 12 13 15 16 17 18 23
SU3 : 20 21
PV4 : 1

13 넣다	넌	SU1 : 1
	널	SU1 : 2
	너	PV1 : 5
		SU1 : 7 8 9
		PV7 : 1
	넣	SU3 : 1 2 3 4 5 6 8 12 13 14 17 18 23 26
		SU1 : 5 6 9 10 11 12 13 15 16 17 18
		SU2 : 1 2 3 4 5
		SU4 : 1
		PV4 : 1
		PV1 : 1 2 3 7 8
		PV7 : 1
14 긋다	근	SU1 : 1
	글	SU1 : 2
	긋	SU3 : 1 2 4 5 7 8 12 13 14 17 18 23 26
		SU1 : 5 6
		PV1 : 1 3
	그	SU1 : 9 10 11 12 13 15 16 17 18
		SU3 : 20 21
		SU2 : 2 3 4 5
		PV4 : 1
15 가르다	가른	SU1 : 1
	가를	SU1 : 2
	가르	SU3 : 1 2 3 4 5 6 7 8 9 12 14 15 17 18 23 24 26
		SU1 : 5 6 9 11 12 13 14 15 16 17 18
		PV1 : 1
		PV7 : 1
	갈랐	SU3 : 1 4 5 6 7 8 9 12 21 26
		SU1 : 5 6 12 13 14
		SU2 : 3 4 5
		PV1 : 1
		PV4 : 1
16 꼬다	꼰	SU1 : 1
	꼴	SU1 : 2
	꼬	SU3 : 1 2 4 5 7 8 12 13 14 17 18 19 20 21 26
		SU1 : 5 6 7 8 9 10 11 12 13 15 16 17 18
		SU2 : 2 3 4 5
		PV3 : 1
		PV5 : 1
		PV1 : 1 3 5
		PV7 : 1
	꽜	SU3 : 1 4 5 7 12 14 15 17 18 21 26
		SU2 : 1 2 4 5
		SU4 : 1
		PV4 : 1
17 개다	갠	SU1 : 1
	갤	SU1 : 2
	개	SU3 : 1 2 4 5 7 8 12 13 14 17 18 19 20 21 26
		SU1 : 5 6 7 8 9 10 11 12 13 15 16 17 18
		SU2 : 2 3 4 5
		PV4 : 1
		PV5 : 1
		PV1 : 1 3 5 8
		PV7 : 1
	갰	SU3 : 1 4 5 7 12 14 15 17 18 21 26
		SU2 : 1 2 4 5
		SU4 : 1
		PV4 : 1

18 굽다	구	PV1 : 5
		SU1 : 7 8 9 10 11 12 13 15 16 17 18
		SU2 : 2 3 4 5
		PV4 : 1
		PV6 : 1 2 3 4 5 6
		PV7 : 1
	군	SU1 : 1
	굴	SU1 : 2
	굽	SU3 : 1 2 3 4 6 7 8 9 10 11 12 13 14 15 17 18 23 26
		SU2 : 1
		SU4 : 1
		SU1 : 5 6
		PV1 : 1 3
19 가다	간	SU1 : 1
	갈	SU1 : 2
	가	SU3 : 1 2 3 4 5 7 8 9 10 12 13 14 15 17 18 22 23 24 26
		SU1 : 5 6 7 9 10 12 13 15 16 17 18 19
		PV1 : 1 3 7 8 9
		PV7 : 1
	갔	SU3 : 1 3 4 5 7 8 12 13 14 15 17 18 21 26
		SU1 : 5 6 13
		SU2 : 1 2 4 5
		SU4 : 1
		PV4 : 1
		PV1 : 1 3
20 웁다	우	SU3 : 1 4 5 8 12 13 17 18 29 21 26
		SU1 : 5 6 9 10 11 12 13 15 16 17 18 19 20
		PV7 : 1
		PV1 : 1 3
	울	SU1 : 2
	운	SU1 : 1
	웁	SU3 : 1 2 4 5 7 8 9 12 13 14 17 18 26
		PV1 : 1 3
		SU1 : 5 13 15
	웠	SU1 : 5 6
		SU2 : 1 2 4 5
		SU4 : 1
		SU3 : 1 2 3 4 6 7 8 9 12 14 15 16 17 18 26
		PV1 : 1 2
21 하다	하	SU3 : 1 2 3 4 5 7 8 10 11 12 13 14 15 17 18 21 23 24 26
		SU1 : 5 6 8 9 10 11 12 13 15 16 17 18 19 20
		PV5 : 1
		PV1 : 1 3
		PV7 : 1
	한	SU1 : 1
	할	SU1 : 2
	해	PV1 : 7
	했	SU3 : 1 2 3 4 5 7 8 9 10 12 13 14 15 16 17 18 19 26
		SU1 : 5 6
		PV1 : 1 3 7
		PV4 : 1
		SU2 : 2 4 5
99 기타		

ANNEXE F. Classes de SFR dans l'adjectif

CLASSE	SFR	Information
1 감다	감	SU3 : 1 2 4 5 7 8 9 12 14 17 18 23 24 26
		SU1 : 5 6
		SU2 : 1 2 3 4 5
		SU4 : 1
		PA1 : 1 3 4
		PA3 : 1
		PA4 : 1
2 녹다	녹	SU3 : 1 2 4 5 7 8 9 12 14 17 18 23 24 26
		SU1 : 5 6
		SU2 : 1 2 3 4 5
		SU4 : 1
		PA1 : 1 3 4
		PA3 : 1
3 검다	검	SU3 : 1 2 3 4 5 6 7 8 9 12 14 15 17 18 21 26
		SU1 : 4 5 6
		SU2 : 2 3 4 5
		PA1 : 1
		PA4 : 1
4 없다	없	SU3 : 1 2 3 4 5 6 7 8 9 12 14 15 17 18 21 24
		SU1 : 3 4 5 6 19 20
		SU2 : 1 2 4 5
		SU4 : 1
		PA1 : 1 3
		PA4 : 1
5 갑다	가	PA6 : 1 2 3 5 6
		PA8 : 1
	갑	SU1 : 4 5
		SU3 : 1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18 26
		SU2 : 1
		SU4 : 1
		PA1 : 1
6 둡다	두	PA6 : 1 2 3 5 6
	둡	SU1 : 4 5
		SU3 : 1 2 3 4 5 6 7 8 9 10 11 12 14 15 16 17 18 26
		SU2 : 1
		SU4 : 1
		PA1 : 1
7 달다	다	SU1 : 5 6
		SU3 : 8 12
	단	SU1 : 1
	달	SU1 : 2
		SU2 : 2 3 4 5
		SU3 : 1 2 4 5 7 15 17 18 26
		PA3 : 1
8 걸다	거	SU1 : 5 6 9 11 12 13 14 15 16 17 18 19
		SU3 : 12 20 21 22
		PA1 : 1 3
	건	SU1 : 1
	걸	SU1 : 2 9 11
		SU2 : 2 3 4 5
		SU3 : 1 2 3 4 5 6 7 14 15 17 18 23 24 26
		PA4 : 1

9 되다	되	SU3 : 1 2 3 4 5 7 8 9 12 14 15 17 18 26
		SU1 : 5 6 7 9 10 11 12 13 14 15 16 17 18
		PA1 : 1 3 5
	된	SU1 : 1
		SU3 : 14 15
		PA1 : 4
	될	SU1 : 2
10 기다	기	SU3 : 1 2 3 4 5 6 7 8 9 10 12 14 17 18 20 21 23 24 26
		SU1 : 5 6 9 11 12 13 15 16 17 18
		PA1 : 1 3
		PA4 : 1
		PA6 : 1 2 3 4 5 6
	긴	SU1 : 1
	길	SU1 : 2
	겼	SU3 : 1 4 5 6 7 8 9 10 12 14 15 17 18 23 24 26
		SU1 : 5
		SU2 : 1 2 4 5
		SU4 : 1
		PA1 : 1
		PA4 : 1
11 나다	나	SU3 : 1 2 3 4 5 7 8 9 10 12 14 15 17 18 21 26
		SU1 : 5 6 13 14 15 16 17 18
		SU2 : 2 3 4 5
		PA1 : 1 3 8 9
		PA3 : 1
	난	SU1 : 1
	날	SU1 : 2
	났	SU3 : 1 4 5 6 7 8 9 10 12 14 15 17 18 21 23 26
		SU1 : 5 13 14 19
		SU2 : 1 2 4 5
		SU4 : 1
		PA1 : 1
		PA4 : 1
12 차다	차	SU3 : 1 2 3 4 5 6 7 8 9 10 12 14 17 18 21 26
		SU1 : 5 6 8 9 10 11 12 13 15 16 17 18
		SU2 : 3 4 5
		PA1 : 1 3 5 7 8
		PA3 : 1
		PA8 : 1
	찬	SU1 : 1
	찰	SU1 : 2
	찼	SU3: 1 4 5 6 7 8 9 10 12 14 15 17 18 21 26
		SU1 : 5 13 14 19
		SU2 : 1 2
		SU4 : 1
		PA1 : 1
		PA4 : 1
13 쌨다	쌔	SU1 : 6
	쌘	SU1 : 1
	쌜	SU1 : 2
	쌨	SU3 : 1 4 5 6 7 8 12 14 17 18 26
		SU2 : 1 2 4 5
		SU1 : 5 19
		PA1 : 1 3
		SU4 : 1
		PA4 : 1

14 낫다	나	SU1 : 6 9 12 13 15 16 17 18
		PA3 : 1
		SU3 : 4 5 7 17 18
		PA1 : 7
	난	SU1 : 1
	날	SU1 : 2
	낫	SU3 : 1 2 3 4 5 6 7 8 9 10 12 14 15 17 18 26
		SU1 : 5
		SU2 : 1 2 4 5
		SU4 : 1
		PA1 : 1 7
15 하다 (동사와 같음)		
16 그르다	그르	SU3 : 1 2 3 4 5 6 7 8 9 12 14 15 17 18 23 24 26
		SU1 : 5 6 9 11 12 13 14 15 16 17 18
		PA1 : 1
	그른	SU1 : 1
	그를	SU1 : 2
	글렀	SU3 : 1 4 5 6 7 8 9 12 21 26
		SU1 : 5 6 12 13 14
		SU2 : 3 4 5
		PA1 : 1
		PA4 : 1
17 갛다	가	SU3 : 8 17
		SU1 : 5 6
		PA1 : 1
	간	SU1 : 1
	갈	SU1 : 2
	갛	SU3 : 1 2 3 4 5 6 7 8 9 10 12 14 15 21 22 26
		SU1 : 3 4 5 6 19
		SU2 : 1
		SU4 : 1
		PA1 : 1
	갰	SU3 : 1 4 6 7 8 9 10 12 14 15 17 18 21 26
		SU1 : 5 19
		SU2 : 1 2 4 5
		SU4 : 1
		PA1 : 1
		PA4 : 1
18 ETC.		