

Analyse syntaxique et traitement automatique du syntagme nominal grec moderne

Athina MICHOU

LATL – Université de Genève

Athina.Michou@lettres.unige.ch

Résumé. Cet article décrit le traitement automatique du syntagme nominal en grec moderne par le modèle d'analyse syntaxique multilingue Fips. L'analyse syntaxique linguistique est focalisée sur les points principaux du DP grec : l'accord entre les constituants fléchis, l'ordre flexible des constituants, la cliticisation sur les noms et le phénomène de la polydéfinitude. Il est montré comment ces phénomènes sont traités et implémentés dans le cadre de l'analyseur syntaxique FipsGreek, qui met en œuvre un formalisme inspiré de la grammaire générative chomskyenne.

Abstract. This article describes an automatic treatment to the Modern Greek noun phrase in terms of the Fips multilingual syntactic parser. The syntactic analysis focuses on the main issues related to the Greek DP: the agreement among the inflected constituents, the relatively free constituent order, noun cliticisation, and the polydefiniteness phenomenon. The paper discusses how these processes are treated and implemented within the FipsGreek parser, which puts forth a formalism relying on Chomsky's generative grammar.

Mots-clés : analyseur grec, analyse morphosyntaxique, syntagme nominal, grec moderne.

Keywords: Greek parser, morphosyntactic analysis, determiner phrase, modern Greek.

1 Introduction

C'est aux années '70 que remonte l'étude du grec moderne sous l'angle des théories linguistiques modernes; c'est la période que marquent la reconnaissance et la propagation de la grammaire générative aux Etats-Unis et en Europe, en tant que modèle dominant de la recherche dans le domaine syntaxique des langues naturelles. Depuis, les études de la morphosyntaxe du grec moderne ont connu une évolution croissante jusqu'à nos jours où la recherche linguistique constitue un champ bien établi.

Parmi d'autres champs de la syntaxe grecque, le syntagme nominal et ses aspects ont souvent offert des sujets de débats et d'analyses linguistiques étendues. En particulier, l'attention des chercheurs a été attirée par la place et l'ordre relativement libre des modificateurs nominaux, le comportement des noms déverbaux, la nominalisation, les pronoms déictiques, la distribution et l'interprétation du génitif possessif. En outre, un défi pour les chercheurs a constitué le phénomène de la répétition de l'article défini accompagné d'un adjectif, appelé construction polydéfinie, détermination multiple, double définitude ou encore *determiner*

spreading. Néanmoins, si la situation au niveau de la recherche linguistique est florissante, le domaine du traitement automatique de la syntaxe grecque n'est pas équivalent. Cela est du au fait de la complexité de la langue, vu l'ordre relativement libre de ses constituants, et au manque des ressources linguistiques pour l'évaluation des travaux. Ainsi, la plupart de programmes de traitement du langage développés jusqu'à aujourd'hui visent à traiter la morphologie mais très peu la syntaxe du grec moderne. De plus, les analyseurs morphosyntaxiques intégrés dans les processeurs lexicaux des systèmes du traitement automatique du grec sont d'habitude destinés à la correction orthographique ou syntaxique ou encore à une application de lexicographie ou de la recherche de l'information.

Dans le cadre de la présente recherche, nous avons envisagé d'étudier les phénomènes intéressants du syntagme nominal en grec moderne dans le but de bâtir un analyseur syntaxique qui puisse traiter efficacement sa structure. La structure du syntagme nominal grec présente des particularités que les langues romanes et germaniques ne partagent pas et dont les analyseurs syntaxiques existants ne tiennent pas compte. Dans notre étude, nous proposons une analyse linguistique de la structure du syntagme nominal grec en vue de son traitement automatique par le modèle FipsGreek (Leoni de León & Michou, 2006). Du point de vue théorique, notre approche repose sur une analyse profonde basée sur l'hypothèse dite analyse DP et du point de vue computationnel la modélisation prend place dans le cadre d'analyse automatique multilingue.

2 État de l'art

Pendant les deux dernières décennies, plusieurs tentatives de traiter le grec moderne, notamment au niveau du traitement morphologique et du traitement syntaxique ont vu le jour dans des instituts et départements universitaires du pays. Parmi ces travaux, il y en a qui traitent des données linguistiques spécifiques sans le support ferme d'un cadre théorique, d'autres qui se réfèrent à des applications au grec moderne des outils puissants déjà développés, comme par exemple le modèle KIMMO (Markopoulos, 2001) et encore d'autres qui recouvrent diverses approches linguistiques théoriques dans le TALN tels que Affix Grammars (Triantopoulou, 1995), HPSG (Kolliakou, 2004), grammaire de validation (Ralli & Galiotou, 1987). Néanmoins, la plupart d'entre eux, même s'ils ont donné des résultats satisfaisants, réalisent un traitement morphologique ou syntaxique partiel. Parmi les programmes développés, nous allons citer ceux qui ont pour but de traiter la syntaxe du grec moderne.

Triantopoulou (1995) a présenté une première tentative du traitement automatique du grec moderne. Son travail consiste en une description du syntagme nominal suivant le modèle formel des grammaires d'affixes AGFL (Affix Grammar over a Finite Lattice) et en utilisant un système orienté vers les corpus.

Draggiotis, Grigoriadou et Philokyprou (1997) ont présenté le système Dinous, dans le but de simuler la manière dont les locuteurs natifs analysent les phrases naturelles. Cet analyseur combine deux sous-systèmes computationnels différemment structurés. Le premier d'entre eux utilise de l'information concernant les préférences des locuteurs natifs et le deuxième traite la connaissance linguistique. Dinous analyse des phrases en fournissant comme résultats des représentations arborescentes annotées avec de l'information morphosyntaxique. Kermanidis, Sgarbas, Fakotakis et Kokkinakis (2000) ont implémenté un analyseur syntaxique pour le grec moderne basé sur le formalisme PC-PATR, ayant pour objectif de traiter plusieurs phénomènes syntaxiques. Dans ce cadre, ils traitent les constructions à verbes copulatifs, l'ensemble des propositions subordonnées, tous les types d'accord (accord entre déterminant - adjectif et nom, sujet - verbe) et la majorité des possibilités de l'ordre des mots

dans les structures syntagmatiques.

Baldzis, Eumeridou et Kolalas (2002) ont développé un système pour le traitement automatique du grec moderne, inscrit dans le domaine de l'enseignement des langues assisté par ordinateur (ELAO). Le modèle proposé essaie, d'une part de formaliser la morphologie, la syntaxe et la sémantique du grec moderne, et d'autre part d'encoder dans un ensemble fini des règles les interactions entre ces niveaux, au cours de l'analyse syntaxique des propositions réelles. Le système a été créé pour couvrir principalement des domaines thématiques importants pour l'approche communicative de l'enseignement et peut être utilisé dans la classe ou par correspondance via Internet.

3 Description du syntagme nominal grec moderne

Dans cette partie, nous allons nous rapporter aux constituants divers du syntagme nominal, et aborder des phénomènes intéressants qui s'y produisent : l'accord entre les constituants fléchis et leur ordre (section 3.1), la cliticisation (section 3.2), la polydéfinitude (section 3.3).

3.1 Les propriétés des constituants du syntagme nominal

Le syntagme nominal du grec moderne est composé d'une tête que constitue le nom, de modificateurs possibles que sont le déterminant, l'adjectif, l'adverbe, et de compléments tels qu'un syntagme nominal au génitif, un pronom clitique ou une proposition relative. Les noms, ainsi que tous ses modificateurs fléchis (déterminants, numéraux, adjectifs) sont dotés d'un genre (masculin - féminin - neutre), d'un nombre (singulier - pluriel), d'un cas (nominatif - accusatif - génitif - vocatif) et appartiennent à plusieurs classes flexionnelles qui déterminent leurs suffixes flexionnels aux cas du singulier et du pluriel. Les caractéristiques majeures des constituants fléchis du syntagme nominal sont : l'accord entre eux en genre, nombre et cas, comme le dénote clairement l'exemple (1a-f), et leur ordre relativement libre :

- | | |
|---|---|
| (1) a. <i>Ο μικρός πρίγκιπας</i> (nom-masc-sg) | d. <i>Οι μικροί πρίγκιπες</i> (nom-masc-pl) |
| b. <i>Του μικρού πρίγκιπα</i> (gen-masc-sg) | e. <i>Των μικρών πριγκίπων</i> (gen-masc-pl) |
| c. <i>Το μικρό πρίγκιπα</i> (acc-masc-sg) | f. <i>Τους μικρούς πρίγκιπες</i> (acc-masc-pl) |
| Det _{def} petit _{adj} prince _{Nom} | Det _{def} petits _{adj} princes _{Nom} |
| 'Le petit prince' | 'Les petits princes' |

Les deux articles, l'article défini (*ο, η, το* 'le, la') et l'article indéfini (*ένας, μια, ένα* 'un, une'), marquent la définitude ou l'indéfinitude d'un syntagme nominal. Il se peut, cependant, que l'article soit absent. En ce qui concerne l'ordre des éléments divers du syntagme nominal, les modificateurs précèdent majoritairement le nom en se plaçant normalement entre l'article et le nom. Un autre modificateur ou un adjectif uniquement peut les séparer par la tête nominale, à l'exception du démonstratif *τέτοιος* 'tel' et du contrastif *άλλος* 'autre' qui peuvent soit précéder soit suivre les noms. Les démonstratifs *αυτός* 'ce', *τότος* 'ce-ci' et *εκείνος* 'ce-là', ainsi que les quantifieurs *όλος* 'tout' et *ολόκληρος* 'entier' diffèrent des autres modificateurs en restant à la périphérie du syntagme nominal. Dans le cas où plusieurs modificateurs sont employés ensemble dans le syntagme nominal, leur ordre canonique est le suivant : *Quantifieur + démonstratif + déterminant + numéral + adjectif + nom + pronom clitique* (Holton et al., 1997).

3.2 La cliticisation dans le syntagme nominal

Les deux types de complément du nom dont nous tenons compte dans cette étude sont :
I. Le syntagme nominal plein au génitif, en l'occurrence le syntagme *της Ειρήνης*, exprimant le possesseur qui peut soit suivre (ex. 2a), soit précéder le nom (ex. 2b) :

- (2) a. *Το αυτοκίνητο της Ειρήνης*
 Det_{def} voiture_{Nom} Det_{def} Irène_{Nom}
 'La voiture d'Irène'
 b. *Της Ειρήνης το αυτοκίνητο*
 Det_{def} Irène_{Nom} Det_{def} voiture_{Nom}
 'La voiture d'Irène'

II. Le génitif des pronoms clitiques (*μου, σου, του/της, μας, σας, τους*) qui, en tant qu'élément possessif, s'attache toujours à un élément hôte qui peut être soit la tête nominale (ex. 3a, 4a), soit un numéral (ex. 3b), soit un adjectif (ex. 4b), mais jamais l'article défini (ex. 3c-d, 4c-d). D'autre part, le clitique *της* ne peut pas apparaître seul sans être soudé à un élément hôte (ex. 3e). A savoir, il ne se comporte ni comme un déterminant, ni comme un adjectif ; il ne partage pas les propriétés distributionnelles de ces catégories et il dispose des propriétés particulières (étant intransitif, ne faisant pas l'accord avec son hôte, portant des traits de personne) et il entretient un lien thématique avec le nom dont ils est sélectionné.

- (3) a. *Το αυτοκίνητό της*
 Det_{def} voiture_{Nom} PRcl_{gén}
 'Sa voiture'
 b. *Το πρώτο της αυτοκίνητο*
 Det_{def} premier_{Num} PRcl_{gén} voiture_{Nom}
 'Sa première voiture'
 c. **Το της αυτοκίνητο*
 d. **Το της πρώτο αυτοκίνητο*
 e. **Της αυτοκίνητο*
 (4) a. *Το ωραίο αυτοκίνητό της*
 Det_{def} belle_{adj} voiture_{Nom} PRcl_{gén}
 'Sa belle voiture'
 b. *Το ωραίο της αυτοκίνητο*
 Det_{def} belle_{adj} PRcl_{gén} voiture_{Nom}
 'Sa belle voiture'
 c. **Το της ωραίο αυτοκίνητο*
 d. **Το της το ωραίο το αυτοκίνητο*

3.3 Les adjectifs et le phénomène de la polydéfinitude

En général, les modificateurs adjectivaux peuvent apparaître en position prénominale ou postnominale. Plus précisément, les adjectifs dans leur usage attributif sont prénominaux étant précédés de l'article défini (ex. 5a) ou indéfini (ex. 6a). Ils peuvent également suivre le nom dans un syntagme nominal indéfini (ex. 6b), mais jamais dans un syntagme nominal défini simple (ex. 5b). La répétition de l'article défini est obligatoire devant l'adjectif qui apparaît en position postnominale seulement si la tête nominale est définie (ex. 5c); au contraire, la propagation de l'article indéfini est exclue (ex. 6c):

- (5) a. *Η πολυτελής κατοικία*
 Det_{def} luxueux_{adj} domicile_{Nom}
 b. **Η κατοικία πολυτελής*
 Det_{def} domicile_{Nom} luxueux_{adj}
 'Le luxueux domicile'
 c. *Η κατοικία η πολυτελής*
 (6) a. *Μια πολυτελής κατοικία*
 Det_{indef} luxueux_{adj} domicile_{Nom}
 b. *Μια κατοικία πολυτελής*
 Det_{indef} domicile_{Nom} luxueux_{adj}
 'Un domicile luxueux'
 c. **Μια κατοικία μια πολυτελής*

Cette possibilité d'avoir à l'intérieur du syntagme nominal grec des occurrences multiples du déterminant défini suivi par des adjectifs, constitue une particularité appelée 'determiner spreading' (Androutsopoulou, 1995) ou 'construction polydéfinie' (Kolliakou, 2004). Ce cas de figure est mieux illustré dans les exemples (7a-c), où l'article défini *το* 'le' précède obligatoirement le nom *αυτοκίνητο* 'voiture' ainsi que les adjectifs *μικρό* 'petit' et *ιταλικό*

‘italien’ qui le modifient. Les adjectifs qui entrent dans la polydéfinitude peuvent être également postnominiaux (ex. 7a), prénominiaux (ex. 7b) ou les deux (ex. 7c-d). Par contre, le syntagme nominal en (ex. 7e) est agrammatical puisqu’il n’existe pas d’article défini pré-adjectival. En outre, le placement des adjectifs dans les exemples ci-dessous est interchangeable en rendant tous les différents ordres possibles :

- (7) a. *To αυτοκίνητο το μικρό το ιταλικό*
 Det_{def} voiture_{Nom} Det_{def} petite_{adj} Det_{def} italienne_{adj}
 ‘La petite voiture italienne’
 b. *Το μικρό το ιταλικό το αυτοκίνητο*
 c. *Το μικρό το αυτοκίνητο το ιταλικό*
 Det_{def} petite_{adj} Det_{def} voiture_{Nom} Det_{def} italienne_{adj}
 d. *Το ιταλικό το αυτοκίνητο το μικρό*
 e. **Το αυτοκίνητο μικρό ιταλικό*

Comme il sera présenté dans la section 4.1, dans notre analyse nous adoptons le cadre théorique de l’analyse DP (Abney, 1987) qui postule que le syntagme nominal est la projection d’une catégorie fonctionnelle D qui est réalisée par un élément appelé déterminant et sélectionne un complément de catégorie NP. Pour ce qui est de l’attachement des adjectifs, lorsqu’ils sont prénominiaux, ils constituent des spécificateurs du nom et ils occupent la place de [Spec, NP] (Stavrou, 1996, 1999). En plus, la position canonique des adjectifs étant à gauche de la tête nominale, l’accord morphosyntaxique réalisé entre l’adjectif et le nom est le type d’accord qui a lieu directement entre Spec - tête nominale. Quant aux adjectifs postnominiaux, ils ont une lecture prédicative et ils sont analysés dans une projection fonctionnelle (FP) de type small clause, adjointe à droite du NP dans laquelle le nom n’est pas réalisé. La procédure de l’analyse Fips de ce type de structure sera décrite plus en détail en fin de la section 4.2.

Dans notre analyse, nous traitons aussi les séquences d’adjectifs, en usage attributif, qui sont soit en modification parallèle du nom, c’est-à-dire une suite linéaire d’adjectifs, sans marque de ponctuation ou de conjonction (ex. 8a), soit en modification avec marque de conjonctions (ex. 8b). Dans la structure en (8b) les adjectifs étant coordonnés par la conjonction de coordination *και* ‘et’, ils constituent des conjoints dans une projection ConjP. En ce qui concerne l’ordre d’apparition des adjectifs dans la projection nominale incluse dans les structures suivantes, il est flexible permettant toutes les combinaisons possibles¹, à l’exception du quantifieur adjectival *πολλά* ‘beaucoup de’, qui doit précéder les autres adjectifs.

- (8) a. *Τα πολλά μικρά εικονογραφημένα παιδικά βιβλία* → [DP [NP [AP][AP][AP][AP] N]]
 Det_{def} nombreux_{adj} petit_{adj} illustré_{adj} enfant_{adj} livre_{Nom}
 ‘Les nombreux petits illustrés livres pour enfants’
 b. *Ενα ωραίο και γρήγορο αυτοκίνητο* → [DP[NP[AP [ConjP[AP] [AP]]] N]]
 Det_{indef} belle_{adj} et rapide_{adj} voiture_{Nom}
 ‘Une belle et rapide voiture’

¹ Sachant qu’au point de vue linguistique il y a des restrictions dans l’ordre des adjectifs qui semblent être des préférences, certaines combinaisons apparaissent plus ou moins acceptables au niveau du parsing.

4 Analyse du groupe nominal proposée dans Fips

4.1 Le formalisme du système Fips et l'algorithme d'analyse

Dans notre système FipsGreek, tout comme dans les autres Fips faisant partie du modèle d'analyse syntaxique multilingue (Wehrli, 2004), le formalisme grammatical utilisé relève de la théorie générative chomskyenne et notamment du module X-barre qui dicte de façon générale et uniforme la géométrie des structures syntaxiques. Néanmoins, le schéma X-barre est simplifié pour faciliter le processus d'implémentation. Comme le nœud 'barre' n'est pas représenté au niveau informatique le schéma est transformé à celui présenté dans la figure 1. Ainsi, la tête syntagmatique X, est une catégorie lexicale (N/nom, V/verbe, P/préposition, Adv/adverbe, Adj/adjectif, Conj/conjonction) ou fonctionnelle (C/complémenteur, T/verbe conjugué, D/déterminant, F/fonctionnelle), ainsi que Left (à gauche de la tête) et Right (à droite de la tête) correspondent à des listes (éventuellement vides) de projections maximales.

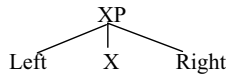


Figure 1 : le schéma X-barre simplifié

Par convention, dans le schéma, la projection DP (Determiner Phrase) est la représentation d'un syntagme dont la tête D (Abney, 1987), réalisée par un déterminant, sélectionne une catégorie syntaxique NP en position de sous-constituant droit. En ce qui concerne l'algorithme d'analyse, il comprend une analyse lexicale, responsable de la segmentation d'une phrase d'entrée en unités lexicales, servant de base à une analyse syntaxique, qui assigne à une phrase une ou plusieurs structures syntaxiques. Puisque la grammaire n'est pas déterministe, c'est-à-dire qu'elle admet plusieurs alternatives possibles, l'analyseur doit traiter plusieurs hypothèses possibles en parallèle. Cela se fait par une stratégie d'analyse syntaxique de type gauche-droite avec un traitement parallèle des alternatives, combinant une approche incrémentale, essentiellement ascendante avec un filtre descendant. En d'autres termes, pour chaque nouveau mot lu, l'analyseur cherche à le combiner avec les constituants déjà construits (contexte gauche), au moyen des opérations suivantes : i) Projection : un constituant maximal (XP) est projeté sur la base des traits inhérents du mot lu. ii) Attachement à gauche : le constituant étant à gauche du nouveau constituant est attaché comme sous-constituant du nouveau constituant. iii) Attachement à droite : le nouveau constituant gauche est attaché comme complément d'un constituant du contexte gauche.

4.2 Les règles proposées et le traitement des données

Dans cette section, nous présenterons notre proposition d'intégrer dans l'analyseur FipsGreek les structures du syntagme nominal grec que nous avons déjà présentées dans la section 3. Comme nous venons le voir, dans le formalisme du système Fips les constituants syntagmatiques sont attachés à gauche et à droite des catégories syntaxiques. En ce qui concerne l'analyse du syntagme nominal, nous adoptons l'hypothèse théorique portant sur l'analyse du DP (Abney, 1987) selon laquelle le syntagme nominal se compose de la projection d'une catégorie fonctionnelle D, remplie généralement par un déterminant lexical, et qui prend comme complément une projection lexicale de catégorie NP (Horrocks & Stavrou, 1987). Ainsi, nous obtenons le syntagme $[_{DP} [D [_{NP} [N]]]]$, illustré dans la figure

2(a), où la catégorie D peut être occupée par une catégorie lexicale autre que l'article, comme illustré dans 2(b), et elle peut sélectionner un complément NP ou un autre D (cf. 2c), d'après les règles de sélection déterminées dans notre grammaire :

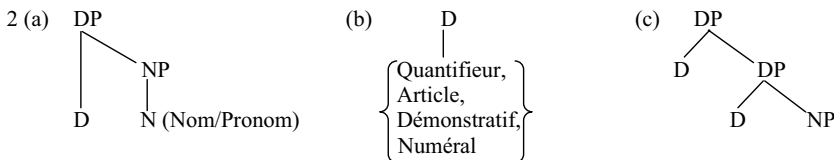


Figure 2(a-c): la structure du DP grec

En grec, comme en français, les éléments appartenant à la catégorie D sélectionnent un NP dont la tête doit être lexicalement réalisée. Par exemple, la structure du DP *to avtokínhto* 'la voiture' est rendue ainsi : [_{DP} *to* [_{NP} *avtokínhto*]]. Dans ce syntagme, la tête nominale *avtokínhto*, occupe la position N et sa projection maximale (NP) s'attache comme complément de D. En plus, l'analyseur, pendant le processus de la reconnaissance des mots, doit vérifier l'accord en genre, nombre et cas, réalisé entre le D et le NP (accord tête-compl) ; ce type d'accord est obligatoire entre tous les déterminants fléchis (articles, quantifieurs, numéraux) et le nom et fait partie intégrale des règles d'attachement des constituants. Dans notre analyse, chaque déterminant est transposé en tant que tête D sélectionnant soit un autre DP (ex. 9) soit un NP comme complément. En outre, dans le lexique chaque entrée lexicale est spécifiée pour tous les traits morphologiques, sélectionnels et sémantiques nécessaires afin d'être utilisés au cours de l'analyse pour autoriser les différents types d'attachements. Ainsi, une séquence de différents types de déterminants, comme celle de l'exemple (9), sera analysée sur la base des traits de sélection qui sont associés à chaque lexème. La structure de sortie du syntagme nominal en (9) est présentée en (10) :

- (9) *Όλες αυτές οι δέκα ετήσιες πολιτιστικές εκδηλώσεις του συλλόγου μας*
 Det_{quant} Det_{dem} Det_{def} Det_{num} Adj Adj activité Det_{def} association PR_{cl}_{gén}
 'Toutes ces dix annuelles activités culturelles de notre association'
- (10) [_{DP} *όλες* [_{DP} *αυτές* [_{DP} *οι* [_{DP} *δέκα* [_{NP} [_{AP} *ετήσιες*] [_{AP} *πολιτιστικές*] *εκδηλώσεις*] [_{DP} *του* [_{NP} *συλλόγου* [_{N(+CL)} *μας*]]]]]]]]]

Or, quelques déterminants ont un comportement distributionnel plus compliqué que d'autres. Cette catégorie contient les quantifieurs *όλος* 'tout' et *ολόκληρος* 'entier', ainsi que les démonstratifs *αυτός* 'celui-ci', *εκείνος* 'celui-là' et *αυτός* 'ce-ci' qui sont prénominaux en tant que déterminants, ou postnominaux, ayant un comportement adjectival (cf. *Όλος ο κόσμος* et **ο όλος κόσμος* 'tout le monde' vs. *Ο κόσμος όλος* 'le monde entier'). Pour traiter ces éléments, nous avons opté pour une double entrée lexicale : celle du déterminant sélectionnant un autre déterminant ou un nom, et celle de l'adjectif portant le trait lexical [+postnominal], ce qui leur permet d'occuper des positions postnominales². Quant aux pronoms clitiques, dans le cadre de notre analyse, ils constituent des pronoms, c'est-à-dire une sous-catégorie de N portant le trait lexical [+CL], pouvant s'attacher aux noms, adjectifs, numéraux

² Notons que les adjectifs comme *όλος*, étant des adjectifs non-intersectifs, ne sont pas compatibles avec la polydéfinitude (cf. *ο κόσμος (*ο) όλος*). Voir aussi note 4.

et quantifieurs³. La figure 3(a) présente la configuration du syntagme *το πρώτο καινούργιο αυτοκίνητό της* ‘sa première voiture neuve’ illustrant cette hypothèse ainsi que les possibilités du mouvement du clitique possessif et de son attachement à un adjectif (*καινούργιο της*) ou à un numéral (*πρώτο της*) en formant une chaîne clitique qui le lie à sa position initiale (*αυτοκίνητό της*). Plus précisément, Fips procède à la reconnaissance des mots du syntagme jusqu’à l’identification du nom *αυτοκίνητο* et à la projection de la tête clitique à sa droite. Dans le cas où le clitique s’est déplacé plus haut, l’analyseur crée une trace et en parallèle une coindexation du clitique avec sa position canonique. Quant à la forme nominale pleine du génitif possessif *της Ειρήνης*⁴ ‘d’Irène’ dans le syntagme *το αυτοκίνητο της Ειρήνης* ‘la voiture d’Irène’, donné dans 3(b), elle occupe la position du complément de la tête nominale *το αυτοκίνητο* ‘la voiture’. Son analyse est déterminée par une règle d’attachement du DP (+génitif) à droite du NP. La structure rendue par l’analyseur est la suivante : [DP *το* [NP *αυτοκίνητο* [DP *της* [NP *Ειρήνης*]]]]. La représentation arborescente de la structure en 3(b) illustre les deux possibilités de placement du DP possessif, soit à sa position canonique soit par déplacement dans le [Spec, DP] pour des raisons de focalisation.

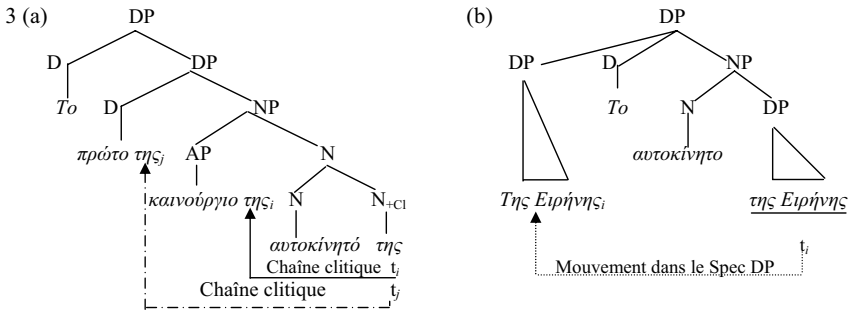


Figure 3(a-b) : Déplacement des clitiques et du DP possessif

Examinons maintenant le phénomène de la répétition du déterminant défini avec un modificateur adjectival prénominal (Det_{def}+Adj+Det_{def}+Nom) ou postnominal⁵ (Det_{def}+Nom+Det_{def}+Adj). Vu que dans les deux cas les adjectifs ont une lecture prédicative, nous proposons un traitement relativement similaire, c’est-à-dire que Fips les analyse comme des structures FP de type *small clause*. Plus précisément, lors de la reconnaissance d’une séquence Det_{def}+Adj qui doit être impérativement précédée ou suivie d’une autre séquence Det_{def}+Nom, après la vérification de l’accord entre les deux séquences, nous proposons qu’une projection fonctionnelle FP se crée. Dans cette projection de prédication FP, l’article projetant un DP occupe la position du sujet et l’adjectif constitue le prédicat. La figure 4(a-b) illustre mieux cette configuration : dans la représentation du syntagme nominal *το αυτοκίνητο*

³ Les éléments clitiques possessifs au génitif sont de la même nature (pronominale) que les éléments clitiques des phrases. A savoir, ils portent des traits de personne, ils ne font pas l’accord avec leur hôte et ils sont intransitifs.

⁴ L’élément *της* est le Det_{def} et non le pronom clitique dont la forme est identique à l’article défini.

⁵ La propagation de l’article défini n’est pas possible avec tous les adjectifs. En fait, les adjectifs dits non-intersectifs bloquent la polydéfinitude (Alexiadou & Wilder, 1998).

το μικρό ‘la voiture la petite’ la projection fonctionnelle s’attache à droite du NP (cf. 4(a)), contrairement à la représentation du syntagme το μικρό το αυτοκίνητο ‘la petite la voiture’ où le FP s’antépose en occupant la place du [Spec, DP] (cf. 4(b)).

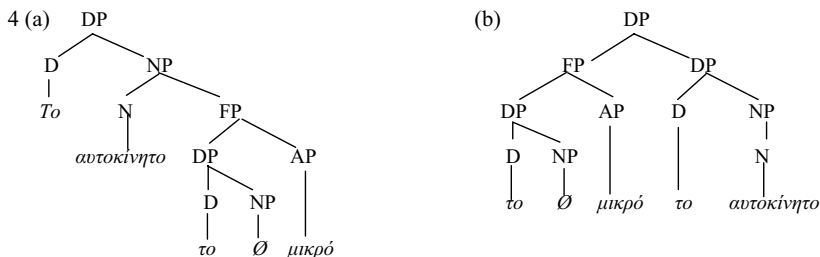


Figure 4(a-b) : Analyses proposées pour les structures polydéfinies

5 Evaluation

L’analyseur FipsGreek a été testé sur un ensemble de fichiers d’exemples des 1000 syntagmes nominaux manuellement extraits de textes en ligne. Ce petit corpus de test a été créé dans le but de nous permettre de contrôler chaque règle de grammaire relative à chaque structure syntaxique traitée. Notre lexique était déjà enrichi avec les nouveaux mots contenus dans ces fichiers. De ce fait, nous avons écarté le problème que les mots inconnus présentent dans l’analyse et nous avons obtenu le taux de 80% des syntagmes complètement analysées. Une analyse est dite complète si l’analyseur parvient à construire une structure arborescente des constituants couvrant l’intégralité de la phrase. Lorsqu’une analyse est incomplète l’analyseur retourne une séquence d’analyses partielles couvrant toute la phrase. Une évaluation manuelle des analyses complètes révèle que les analyses linguistiquement incorrectes représentent un très faible pourcentage de 4%. Les analyses incomplètes ou incorrectes sont dues à trois raisons principales. Premièrement, à l’ambiguïté lexicale, l’analyseur n’ayant pas identifié la bonne catégorie des mots dans les syntagmes. Deuxièmement, au fait que dans les syntagmes non analysés, il existe des fautes d’orthographe, des mots étrangers, des abréviations et des sigles que l’analyseur n’a pas pu deviner. Troisièmement, au manque de la couverture de règles de la grammaire dans le système, fait qui est la cause de la majorité d’échecs. A savoir, il y a encore diverses structures que nous ne traitons pas, telles que les appositions, les constructions elliptiques, les figures du discours stylistiques, les dépendances à longue distance (Horrocks & Stavrou, 1987), etc.

6 Conclusion

Cet article présente une approche au traitement automatique des structures syntaxiques du syntagme nominal en grec moderne qui repose sur le postulat théorique de l’analyse DP. Cette analyse permet la sélection non seulement d’un NP mais aussi d’un DP par un autre DP et, par conséquent, des couches successives des DP que nous rencontrons dans le contexte du syntagme nominal en grec moderne. Notre proposition de modélisation utilise l’avantage que présente l’analyse hiérarchique du DP plutôt qu’une analyse plus linéaire avec une structure NP simple, pour la formalisation et l’implémentation des structures nominales du grec dans le cadre de l’analyseur multilingue Fips. Ce dernier est conçu avec des procédures et des modules généraux et des modules spécifiques au grec. Ainsi, FipsGreek, s’intégrant dans ce

système multilingue procède à une analyse de surface enrichie, parallèle à celles effectuées pour les autres langues, ce qui conduit à l'amélioration de la procédure du parsing dans le système de la traduction. Par ailleurs, les résultats obtenus montrent que notre modèle constitue un outil adéquat pour le traitement automatique des spécificités que présente le grec. La présente étude peut être étendue et couvrir l'analyse syntaxique de la phrase en donnant des perspectives à la recherche menée dans le futur.

Remerciements

Cette recherche a bénéficié du financement du Fonds national suisse de la recherche scientifique (projet n° 101412-103999). L'auteur tient à remercier Eric Wehrli, et Christopher Laenzlinger en particulier, pour leurs discussions et commentaires fructueux.

Références

- ABNEY S. (1987). *The English Noun Phrase in its Sentential Aspect*. Cambridge, MA: MIT Press.
- ALEXIADOU A., WILDER C. (1998). Adjectival modification and multiple determiners. In Alexiadou A. et al. (éd.), *Possessors, Predicates and Movement in the Determiner Phrase*, 303-332. Benjamins.
- BALDZIS S.D., EUMERIDOU E., KOLALAS S. (2002). A Complete and Comprehensive System for Modern Greek Language Processing Proposed as a Modern Greek Language Call Method Developer. *Literary and Linguistic Computing* 17(4), 373-400.
- ANDROUTSOPOULOU A. (1995). The distribution of definite determiners and the syntax of Greek DPs. In *Proceedings of CLS*.
- DRAGGIOTIS A., GRIGORIADOU M., PHILOKYPPOU G. (1998). The DINOUS parser. *Natural Language Engineering* 4(2), 145-173.
- HOLTON D., MACKRIDGE P., PHILIPPAKI-WARBURTON I. (1997). *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- HORROCKS G., STAVROU M. (1987). Bounding theory and Greek syntax: evidence for wh- movement in NP. *Journal of Linguistics* 23, 79-108.
- KERMANIDIS K., SGARBAS K., FAKOTAKIS N., KOKKINAKIS G. (2001). A PC-PATR-Based Syntactic Description of Modern Greek. *Literary and Linguistic Computing* 15(3), 291-311.
- KOLLIAKOU D. (2004). Monadic definites and polydefinites: their form, meaning and use. *Journal of Linguistics* 40, 263-323.
- LEONI DE LEÓN A., MICHOU A. (2006). Traitement des clitics dans un environnement multilingue. Actes de la 13^{ème} conférence TALN 2006, 541-550.
- MARKOPOULOS G. (2001). *A Two-level Description of the Greek Noun Morphology with a Unification-based Word Grammar*. Ph.D. Dissertation: University of Athens.
- STAVROU M. (1999). The position and serialization of APs in the DP: evidence from Greek. In Alexiadou A. et al. (éd.), *Studies in Greek Syntax*, 201-225. Dordrecht: Kluwer.
- STAVROU M. (1996). Adjectives in Modern Greek. An instance of predication or an old issue revisited. *Journal of Linguistics* 32, 79-111.
- TRIANTOPOULOU T. (1997). A Description of the Modern Greek Noun Phrase Using the "Affix Grammars over a Finite Lattice" Formalism. *Literary and Linguistic Computing* 12(2), 119-133.
- RALLI A., GALIOTOU E. (1987). A morphological processor of Modern Greek. *Proceedings of the ACL European Chapter Meeting*. Denmark.
- WEHRLI E. (2004). Un modèle multilingue d'analyse syntaxique. In Auchlin A. et al. (éd.), *Structures et discours : Mélanges offerts à Eddy Roulet*, 311-329. Canada: Nota bene.