

Traitement des clitiques dans un environnement multilingue

Jorge Antonio Leoni de León, Athina Michou

Université de Genève, Département de linguistique
Laboratoire d'Analyse et de Technologie du Langage
{jorge.leonideleon ; athina.michou}@lettres.unige.ch

Résumé

Cet article décrit le traitement automatique des pronoms clitiques en espagnol et en grec moderne, deux langues de familles distinctes, dans le cadre de l'analyseur syntaxique FIPS multilingue, développé au Laboratoire d'Analyse et de Technologie de Langage (LATL). Nous abordons la distribution des pronoms clitiques, leurs similarités ainsi que leurs particularités par rapport à leur usage général. Ensuite nous présentons la méthode appliquée pour leur traitement, commune aux deux langues. Nous montrons que l'algorithme proposé peut facilement s'étendre à d'autres langues traitées par Fips qui partagent le phénomène de la cliticisation.

Mots-clés : analyseur syntaxique, clitiques, traitement multilingue.

Abstract

In this article we present an algorithm for the computational processing of clitic pronouns in Spanish and Modern Greek by the multilingual syntactic parser Fips that has been developed at the Laboratory For Language Analysis And Technology, LATL. We provide an overview of the distribution of clitic pronouns, their similarities and particular behaviour compared to general use. Then we present a common method of processing clitics for both languages. Finally, we show that the algorithm developed can be easily extended to other languages with clitics, when treated by Fips.

Keywords: syntactic parser, clitics, multilingual processing.

1. Introduction

À la suite du développement des analyseurs Fips pour le français et l'anglais (Laenzlinger et Éric Wehrli, 1991), le Laboratoire d'Analyse et de Technologie du Langage (LATL) s'est engagé dans une reformulation multilingue de cet analyseur (Éric Wehrli, 2004). Le système, inspiré des théories générativistes chomskyennes (Chomsky, 1995, chapitre 1 avec Howard Lasnik), (Haegeman, 1994), est élaboré sur les principes généraux de simplification maximale des structures syntaxiques, modularité des processus d'analyse et construction incrémentale des analyses avec traitement parallèle des alternatives. Il contient des projections syntagmatiques maximales et minimales (têtes) contenues dans le schéma ternaire de structure syntagmatique $[_{XP}Y X^o Z]$ (où Y et Z sont des listes de projections maximales, éventuellement vides), lequel permet plusieurs sortes d'opérations, comme le mouvement ou la création de chaînes (Éric Wehrli, 1997).

L'approche multilingue facilite l'ajout de nouvelles langues, grâce à un noyau grammatical chargé des calculs procéduraux communs à toutes les langues du système et à des extensions

spécifiques à une famille de langues. Cette approche diminue le temps de développement et simplifie drastiquement la maintenance informatique. Fips utilise une procédure uniforme qui gère la cliticisation pour les langues romanes¹. Notre objectif est de montrer qu'elle est non seulement facilement adaptable à d'autres langues non romanes, dans ce cas, le grec, mais qu'elle permet aussi d'aborder, d'une manière élégante et linguistiquement justifiée, des phénomènes tels que le redoublement clitique et la montée longue des clitiques en espagnol (voir section 2). Le traitement des clitiques grecs et espagnols est ainsi ajouté à l'analyseur Fips, dans lequel le français et l'italien sont déjà incorporés. Notre proposition sera accompagnée d'une démonstration de notre application.

2. Présentation de données

Un clitique est une unité fonctionnelle inaccentuée, prosodiquement dépendante. Par conséquent, il doit s'attacher à une unité lexicale lui servant d'hôte, avec laquelle il a des forts liens syntaxiques et sémantiques². On peut affirmer aussi que les clitiques sont des morphèmes dont le comportement n'est, à part entière, ni celui des unités lexicales réalisées dans la phrase³, ni celui des éléments purement flexionnels⁴. Ces morphèmes ont, d'un côté, les propriétés des éléments liés (aux arguments verbaux correspondants) et, de l'autre, ils se comportent comme des éléments syntaxiques autonomes (Kayne, 1975 ; Jaeggli, 1986) ; rien ne sépare un pronom clitique de son hôte à l'exception d'un autre pronom clitique. Il ne faut pas oublier que les pronoms clitiques portent les traits de cas, personne et nombre (et parfois genre). L'interprétation des traits peut varier dialectalement de manière importante à l'intérieur d'une même langue (sujet au delà de la portée de cet article). Les clitiques occupent des positions distinctes, associées aux têtes de certaines projections maximales.

Tant en espagnol qu'en grec, les pronoms clitiques peuvent se trouver en position préverbale (proclise) ou postverbale (enclise), selon les circonstances favorisant la cliticisation⁵ :

- | | |
|---|--|
| <p>(1) a. GR : Aftós ídhe tin kopéla.
Il a vu la fille.</p> <p>b. ES : Él vio a la niña.
Il a vu à la fille.
<i>Il a vu la fille.</i></p> | <p>(2) a. GR : Aftós tin ídhe.
Il l' a vu.</p> <p>b. ES : Él la vio.
Il l' a vu .
<i>Il l'a vue.</i></p> |
|---|--|

Les phrases (2a) et (2b) correspondent à la cliticisation de (1a) et (1b), respectivement. Ainsi, (2b) montre que le pronom « *la* » se substitue au nom « *la niña* » ; de même, en (2a), le pronom « *tin* » remplace le nom « *tin kopéla* ». Mais le grec et l'espagnol ont d'autres similitudes, hormis la proclise sur les verbes à l'indicatif ou au subjonctif. Ils se caractérisent aussi par la présence du phénomène du redoublement clitique :

¹ Le phénomène de la cliticisation a souvent constitué l'objet d'analyses linguistiques théoriques et d'implémentations informatiques dans un autre cadre générativiste, celui de HPSG (Alexopoulou et Kolliakou, 2002 ; Kordoni et Neu, 2005).

² Cependant, dans le cas de la montée longue, les liens sémantiques sont relâchés.

³ Comme un verbe ou une préposition.

⁴ Par exemple, les terminaisons verbales indiquant personne et nombre : {*-ons*} dans « *mangeons* » et {*-ez*} dans « *mangez* ».

⁵ L'acronyme GR indique les phrases en grec ; ES, celles en espagnol.

- (3) a. GR : Aftós tin ídhe tin kopéla. (4) a. ES : Ellos quieren darte la carta (a ti).
 Il l' a vu la fille Ils veulent donner-te la lettre (à toi)
Il l'a vue Ils veulent donner la lettre à toi.
- b. ES : Yo lo vi a él. b. ES : Ellos quieren dártela.
 Je l'ai vu lui. Ils veulent donner-te-la.
 Je l'ai vu. *Ils veulent te la donner.*

Le redoublement clitique consiste en la présence simultanée du clitique (datif ou accusatif, par exemple) et de l'élément auquel il réfère. Dans (3b), le clitique accusatif espagnol « *lo* » réfère à un syntagme manifeste dans la phrase –« *a él* »–; tout comme le grec « *tin* » par rapport à « *tin kopéla* ». Le datif en espagnol est aussi concerné, ce qui est illustré en (4a) par la relation existante entre « *te* » et « *a ti* ».

Cependant, entre les deux langues nous trouvons des différences importantes. Par exemple, en espagnol, les pronoms clitiques sont restreints au verbe ; tandis qu'en grec leur usage s'étend aux prépositions adverbiales (5) et au nom (6) :⁶

- (5) GR : brostá mou
 devant moi^{Clitique génitif}
Devant moi

- (6) GR : to spíti mas
 la maison^{Nom} notre^{Clitique génitif}
notre maison

Les pronoms clitiques, dépourvus d'indépendance morphosyntaxique et d'accent, sont obligés de former une unité soit avec le verbe qui les suit (2), soit avec le verbe qui les précède (4b). Aussi dans l'enclise, il existe des différences remarquables entre les deux langues, car bien qu'en grec l'enclise soit possible avec les verbes à l'impératif⁷ ou au gérondif⁸, les enclitiques sont toujours orthographiquement séparés par des espaces, contrairement à l'espagnol, où ils sont soudés au verbe (l'enclise en espagnol requiert que le verbe soit à l'infinitif, au gérondif ou à l'impératif) :

- (7) a. GR : Dhós' to mou ! (8) a. ES : ¡Dámelo !
 Donne le à moi^{Génitif} ! Donne-moi-le
Donne-le-moi ! Donne-le-moi !
- b. GR : Dhós' mou to ! b. ES : * ¡Dálome !
 Donne à moi^{Génitif} le Donne-le-moi
Donne-le-moi !

D'ailleurs, par rapport au grec, l'espagnol se caractérise par la montée longue des clitiques :

⁶ Les clitiques possessifs peuvent s'attacher, en tant qu'enclitiques « flottants », soit à la tête nominale, comme en (i) soit à un modifieur adjectival (ii) ou numéral (iii), sans changer le sens de la phrase (Holton *et al.*, 2000) : (i) o prótos megálos athlitís **mas**, (ii) o prótos megálos **mas** athlitís, (iii) o prótos **mas** megálos athlitís (« *notre premier grand athlète* »).

⁷ Le grec moderne ne dispose pas d'infinitif équivalent à l'infinitif français.

⁸ Les objets indirects en grec sont exprimés soit par le cas morphologique du génitif, soit par une construction périphrastique. L'ordre des séquences enclitiques grecques est abordé en (13).

- (9) a. Yo tengo que comprarlo. (10) a. Ellos quieren dártela (a ti).
 Je ai que acheter-le. b. Ellos te la quieren dar (a ti).
Je dois l'acheter.
- b. Yo lo tengo que comprar.
 Je le ai que acheter.
Je dois l'acheter.

Dans (9b) le clitique « *lo* » passe d'une enclise dans la phrase enchâssée infinitive « *comprar* » à une proclise sur le verbe principal (« *tengo* »), déplacement aussi effectué dans (10) – reprise de l'exemple (4) – par les deux clitiques « *te la* ». Ce mouvement est considéré comme montée longue (« *clitic climbing* » en anglais).

Tant en espagnol, qu'en grec, les clitiques se regroupent en séquences ordonnées. Ainsi, en espagnol nous avons le filtre (11) (Perlmutter, 1970), valable autant pour la proclise que pour l'enclise :

(11) Séquences de clitiques en espagnol :

$$se \left\{ \begin{array}{l} te_{[II,sg,\{Acc,Dat\}]} \\ OS_{[II,pl,\{Acc,Dat\}]} \end{array} \right\} \left\{ \begin{array}{l} me_{[I,sg,\{Acc,Dat\}]} \\ NOS_{[I,pl,\{Acc,Dat\}]} \end{array} \right\} \left\{ \begin{array}{l} lo_{[III,sg,Acc]} \\ la_{[III,sg,Acc]} \\ le_{[III,pl,Dat]} \\ los_{[III,pl,Acc]} \\ las_{[III,pl,Acc]} \\ les_{[III,pl,Dat]} \end{array} \right\}$$

Où : I,II,III=Personne sg=Singulier pl=Pluriel Acc=Accusatif Dat=Datif

Ce filtre permet la reconnaissance des séquences grammaticales dans les groupes clitiques (par exemple, « *me los* », « *nos lo* », « *me la* », « *te nos* », etc.), tout en écartant les suites agrammaticales (comme « *te os* » et « *nos me* »). Autrement dit, les clitiques, pour se réaliser dans la phrase, doivent appartenir à des classes clitiques différentes. Ainsi, grâce au filtre (11), il est possible de prédire l'agrammaticalité des séquences [*Datif, Accusatif*] « *le(s) lo(s)* » et « *le(s) la(s)* », lesquelles, en effet, n'ont pas lieu en espagnol. Le verbe « *dar* » (« *donner* ») peut être accompagné d'un de ces clitiques ; lorsque les deux devraient se réaliser ensemble, le pronom « *se* » se substitue aux pronoms « *le* » et « *les* » : « *se lo di* » (« *je le lui ai donné* »), « *se los di* » (« *je le leur ai donné* »). Pour expliquer ce phénomène, la règle morphophonologique du « *faux se* » (Perlmutter, 1970) a été proposée ; cependant nous considérons qu'il est aussi possible de considérer que le pronom SE comporte un trait [*Datif*] pour ces séquences. D'une autre part, en ce qui concerne le grec, les clitiques respectent aussi un ordre rigoureux dans la proclise :

(12) Séquences des pronoms proclitiques en grec :

$$\left. \begin{array}{l} \text{mou}_{[I,sg,G\acute{e}n]} \\ \text{sou}_{[II,sg,G\acute{e}n]} \\ \text{tou}_{[III,sg,G\acute{e}n]} \\ \text{t}\bar{\text{S}}_{[III,sg,G\acute{e}n]} \\ \text{mas}_{[I,pl,G\acute{e}n]} \\ \text{sas}_{[II,pl,G\acute{e}n]} \\ \text{tous}_{[III,pl,G\acute{e}n]} \end{array} \right\} \left. \begin{array}{l} \text{ton}_{[III,sg,Acc]} \\ \text{tin}_{[III,sg,Acc]} \\ \text{to}_{[III,sg,Acc]} \\ \text{tous}_{[III,pl,Acc]} \\ \text{tis}_{[III,pl,Acc]} \\ \text{tes}_{[III,pl,Acc]} \\ \text{ta}_{[III,pl,Acc]} \end{array} \right\} V$$

Où : I,II,III=Personne sg=Singulier pl=Pluriel Acc=Accusatif
Gén=Génitif NÉG=Négation V=Verbe

En revanche, tel que l'exemple (7) le montre, les enclitiques grecs sont interchangeables :

(13) Séquences des pronoms enclitiques en grec :

$$V_{[imperatif]} \left\{ \begin{array}{l} \text{Gén} \quad \text{Acc} \\ \text{Acc} \quad \text{Gén} \end{array} \right\}$$

Les séquences (11) et (12) peuvent être synthétisées encore plus. Ainsi, de (11) il découle qu'en espagnol les clitiques se regroupent selon le trait de personne, dans un filtre permettant d'exclure les séquences agrammaticales : SE II I III. D'ailleurs, en grec, le filtre (12) est basé non sur le trait de personne, mais sur le cas morphologique : *Gén Acc*. Les exemples (5) et (6) ont montré la possibilité de cliticisation sur les prépositions adverbiales et sur le nom en grec : Adv {*Gén*} ou Nom {*Gén*}. Le filtre pour cette catégorie n'est pas nécessaire, car il n'existe qu'un seul clitique au cas génitif.

Les prépositions adverbiales en grec permettant l'enclise interdisent l'usage des syntagmes nominaux lexicaux au génitif. Ces adverbes se comportent de manière similaire aux verbes : un argument doit être exprimé soit sous forme de clitique au génitif, soit sous forme de syntagme nominal périphrastique. Par conséquent, dans le traitement automatique on peut utiliser les mêmes règles d'attachement de clitiques pour les verbes que pour les adverbes, comme nous le verrons par la suite.

3. Théorie des clitiques

Les particularités des clitiques (incorporation à un élément hôte, interprétation comme argument, accord avec le participe, etc.) sont abordées de trois manières en linguistique formelle. La première approche, dite lexicale, postule que les clitiques appartiennent au verbe (par exemple) et qu'ils sont générés avec eux : *Pronom_{clitique} + Verbe = Verbe* (Miller et Sag, 1997). Outre une surcharge des procédures lexicales, la théorie rencontre le problème d'interprétation d'un clitique qui se trouve sur la tête qui ne le sélectionne pas (montée longue des clitiques).⁹ La deuxième approche, que nous appellerons *génération basique* (Jaeggli, 1986), suggère que les clitiques sont générés à la base du verbe, dont ils entretiennent une relation de liage avec

⁹ Bien que le cas avec les verbes modaux ait été traité de manière satisfaisante à partir de la notion de prédicats complexes, il reste en suspens la question du clitique français « en » interprété à partir d'un sujet ou d'un objet direct : « la porte en a été ouverte », « il en a lu le premier chapitre ».

les positions d'argument. Le redoublement clitique joue en faveur de cette proposition, car bien que les positions d'arguments soient occupées, les clitiques apparaissent coindexés avec elles. La phrase espagnole (4a) illustre ce phénomène : le syntagme prépositionnel « *a ti* », argument au datif du verbe « *dar* », redouble le clitique « *te* ». Ainsi, il semblerait que le clitique n'a pas son origine dans la position d'argument, car elle est remplie par un syntagme. Toutefois, la génération basique se heurte au phénomène de la montée longue de clitiques, car si le clitique était généré à la base du verbe, dont il lie les arguments, il ne devrait pas apparaître sur le verbe principal avec lequel le clitique n'entretient pas de rapports thématiques. Par conséquent, la théorie par génération basique du verbe se heurte aux mêmes problèmes que l'analyse lexicale.

La troisième approche formelle, dite du *mouvement clitique* (Kayne, 1975), considère les clitiques comme des pronoms générés dans la position d'argument, qui doivent se déplacer vers un hôte (dans notre cas, un verbe). L'avantage par rapport à la *génération basique* est que, pour la *montée longue*, elle permet d'établir le rapport entre le clitique sur le verbe de la phrase principale et les arguments du verbe infinitif enchâssé. Évidemment, la *montée longue* n'est possible que sous certaines conditions dépendantes du type de verbe (modal) permettant cette opération dans la phrase principale et le point de départ du clitique dans une phrase enchâssée (où le verbe est à l'infinitif ou au gérondif). Le redoublement clitique est rendu possible grâce à la préposition (« *a* »), qui est un marqueur de cas, datif en (4a) (ce qui n'est pas le cas en grec). Dans le cadre de l'approche par mouvement, il est proposé que le clitique et le complément redoublé forment un syntagme nominal complexe d'où le clitique est extrait pour se déplacer sur son hôte.

4. Fips multilingue

Fips multilingue est un analyseur syntaxique en développement basé sur un modèle procédural d'analyse syntaxique profonde pour le français et l'anglais (Éric Wehrli, 2004). Cet analyseur recourt à une approche 'objets' qui, d'une part, permet de traiter les différences observées dans les langues et, d'autre part, simplifie l'ajout de nouvelles langues¹⁰. L'orientation objet permet notamment une meilleure modularité de traitement. La grammaire est formulée selon le cadre théorique générativiste –(Chomsky, 1995, chapitre 1 avec Howard Lasnik), (Haegeman, 1994)– qui repose sur la description des propriétés abstraites universelles des langues naturelles et permet leur codification et leur représentation de manière uniforme pour des familles des langues. Plus précisément, l'analyseur est défini sur la base des objets suivants, qui sont de type dynamique propre à chaque langue : a) les unités lexicales correspondant aux entrées de la base de données lexicale (*LexicalItem*) ; b) les projections syntaxiques correspondant aux structures syntaxiques créées par l'analyseur (*Projection*) ; et c) les items d'analyse correspondant aux analyses plausibles faites par l'analyseur (*Item*).

Un certain nombre de méthodes sont associées aux objets susmentionnés, dont les plus importantes sont Project (« projeter »), Merge (« assembler ») et Move (« déplacer »). L'opération de projection crée un constituant syntaxique (qui dispose d'une tête lexicale), qu'on essaye d'attacher (à gauche ou à droite), à d'autres structures projetées, par le biais de la méthode d'assemblage. L'opération de mouvement consiste à associer un élément déplacé à une catégorie vide ; cette dernière, étant en position d'argument ou d'ajout, est liée à l'élément extraposé par le biais d'une chaîne. Par exemple, pour la phrase « *Le chat l'a vue* », où l'objet direct est réalisé par le clitique « *l'* » (« *la* »), l'analyseur retournera la structure suivante :

¹⁰ Les langues sur lesquelles on travaille sont, par ordre de développement : l'anglais, le français, l'allemand, l'italien, l'espagnol, le grec et le roumain.

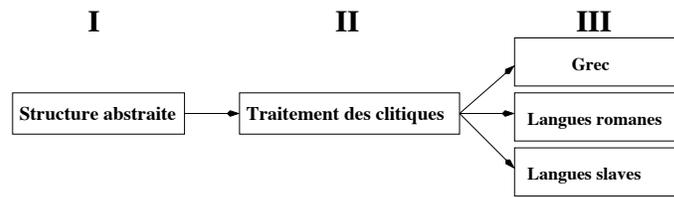


Figure 1. Schéma multilingue à trois niveaux

$[_{TP}[_{DP} le[_{NP} chat]] [_{VP}vue [_{DP} e]]]$.¹¹ Un exemple de cette opération concerne les pronoms clitiques des langues romanes. L'approche utilisée fait appel à une structure de données qui prend la forme d'extension des structures de données de base et une procédure d'interprétation qui prend la forme de méthode associée à cette structure de données spécifique. La figure 1 illustre la hiérarchie à 3 niveaux déjà appliquée au français, l'ajout des structures et des processus propres au traitement des pronoms clitiques espagnols et grecs prend place au deuxième niveau. On se rend compte que cette approche peut s'étendre facilement au traitement des pronoms clitiques des langues slaves pour lesquelles une approche purement lexicale aurait échoué¹².

La stratégie de Fips est dirigée par les données ; les mots d'une phrase sont lus itérativement de gauche à droite. Chaque mot est recherché dans le lexique. Pour chaque lecture trouvée dans le lexique, une projection syntaxique est créée (projection maximale), que l'on tente de combiner avec les structures précédemment construites (plus spécifiquement : les structures maximalelement développées dans le contexte gauche immédiat de la nouvelle projection). Deux types d'attachement sont considérés : l'attachement à gauche et l'attachement à droite. On parle d'attachement à gauche lorsque la structure précédemment créée est attachée comme sous-constituant gauche de la nouvelle projection. Un exemple d'attachement à gauche serait, en français, l'attachement d'un groupe nominal (sujet) comme sous-constituant gauche d'une projection *TP* (meta-projetée sur la base d'un verbe conjugué). L'attachement à droite, beaucoup plus fréquent dans les langues naturelles, correspond au cas de figure où la nouvelle projection peut s'attacher comme sous-constituant droit d'une structure sur sa gauche. Ainsi, pour le fragment « *le chat* », lorsque l'analyseur lit le mot « *chat* », il crée une projection *NP* qui s'attache comme sous-constituant droit de la projection *DP* (dont la tête est « *le* ») immédiatement sur sa gauche.

5. Le traitement des clitiques

Le traitement des clitiques dans Fips compte deux étapes : (i) l'attachement et (ii) l'interprétation. Une fois identifié, un clitique est attaché à la tête de la structure verbale et simultanément inséré dans une structure temporaire. Cette structure a deux fonctions : d'une part de conserver les clitiques jusqu'au moment où le verbe principal de la phrase est lu, et, d'autre part, vérifier la bonne formation des séquences des clitiques. En espagnol et en grec les clitiques correspondent à des arguments ou à des ajouts et ils sont interprétés par rapport au verbe principal de la phrase. Une trace est laissée dans la position d'argument du verbe, laquelle est liée par le clitique formant ainsi une chaîne. La figure 2 illustre cette méthodologie.

Plus concrètement, dans la proclise, Fips procède à la création de projections des têtes clitiques

¹¹ La projection *TP* représente le syntagme de l'inflexion, qui contient le sujet et le prédicat ; le *DP* est le syntagme du déterminant, qui contient les formes nominales.

¹² Les clitiques des langues slaves se caractérisent par leur placement en seconde position dans la phrase et non sur une catégorie déterminée.

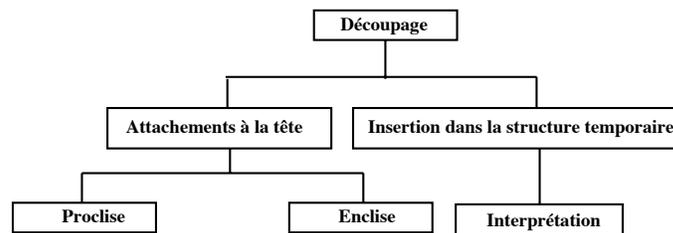


Figure 2. Séquences des procédures pour le traitement des clitiques

au fur et à mesure que les clitiques sont identifiés. Lorsque le verbe est reconnu, le système fait appel à la structure temporaire¹³ et tente de d'interpréter tous les clitiques par rapport au verbe. La procédure pour les enclitiques est similaire, mais elle commence par l'identification du verbe et, dans le cas de l'espagnol, par le découpage des enclitiques. Le tableau 5 illustre un exemple de traitement de l'enclise en grec et en espagnol, exemples (7a) et (8a) respectivement.

	Espagnol	Grec
i	La séquence de caractères « <i>dámelo</i> » est découpée, l'accent est enlevé : [da][me][lo]	Dans la séquence « <i>dhós'to mou</i> », le verbe « <i>dhós'</i> » est identifié.
ii	L'impératif « <i>da</i> » donne lieu à la projection [TP[VP da]].	L'impératif « <i>dhós'</i> » donne lieu à la projection [TP <i>dhós'</i> [VP]].
iii		Le système lit le premier pronom clitique (« <i>to</i> ») de la chaîne de caractères.
iv	Une tête enclitique [<i>Clitique</i> me] est créée pour le premier pronom clitique et elle est associée à la projection verbale.	Une tête enclitique [<i>Clitique</i> to] est créée pour le premier pronom clitique et elle est associée à la projection verbale.
v		Le système lit le deuxième pronom clitique (« <i>mou</i> ») de la chaîne d'entrée.
vi	Une tête enclitique [<i>Clitique</i> lo] est créée pour le deuxième pronom clitique et elle est associée à la projection verbale.	Une tête enclitique [<i>Clitique</i> mou] est créée pour le deuxième pronom et elle est associée à la projection verbale.
vii	Les clitiques, filtrés par la structure temporaire, sont interprétés par rapport au verbe principal.	
viii	Des coindexations sont réalisées entre les clitiques et les arguments.	
ix	Succès, les clitiques occupent les positions attendues dans la structure temporaire et ils sont interprétables comme arguments du verbe.	

Tableau 1. Algorithme de traitement de clitiques (enclise)

Le tableau 5 illustre le traitement du redoublement de l'objet direct en grec et de l'objet indirect en espagnol. Les exemples concernés sont « *Aftós tin ídhe tin kopéla* » (3a) pour le grec et, pour l'espagnol, « *Yo le pegué a él* » (« *Je l'ai frappé* »)¹⁴. Dans le traitement du redoublement, la réalisation d'un pronom datif relâche les contraintes sur la présence d'un syntagme dans la position d'argument (auquel le clitique est lié), à condition que ce syntagme soit prépositionnel et sa tête soit la préposition redoublante « *a* ». En grec, à la différence de l'espagnol, le redouble-

¹³ Définie à partir des filtres sur les clitiques (11) pour l'espagnol et (12) pour le grec.

¹⁴ Il faut préciser que le verbe espagnol « *pegar* » se construit avec un pronom datif pour le sens de « *frapper quelqu'un* » ; alors que le pronom accusatif impliquerait le sens de « *coller* ».

ment est réalisé sans l'intermédiaire d'une préposition. Le redoublement clitique en grec n'a pas d'autres contraintes (comme une intonation spéciale, par exemple). Donc, un traitement parallèle à l'espagnol est possible, mais sans la nécessité de prendre en considération la condition de la présence d'une préposition.

	Espagnol	Grec
i	Le pronom clitique « <i>le</i> » est lu et identifié.	Le pronom clitique « <i>tin</i> » est lu et identifié.
ii	Une tête proclitique [<i>Clitique</i> <i>le</i>] est projetée et elle est associée à tête de la projection verbale.	Une tête proclitique [<i>Clitique</i> <i>tin</i>] est projetée et elle est associée à la tête de la projection verbale.
iii	Les clitiques, filtrés par la structure temporaire, sont interprétés par rapport au verbe principal.	
iv	Le système reconnaît un syntagme prépositionnel [<i>PP a...</i>] et il tente de l'attacher au verbe. Alors, une coindexation est réalisée entre le clitique et l'argument.	Le système reconnaît un syntagme déterminant [<i>DP tin kopela</i>] et il tente de l'attacher au verbe. Alors, une coindexation est réalisée entre le clitique et l'argument.
v	Succès. La phrase est reconnue comme grammaticale.	

Tableau 2. *Algorithme de traitement du redoublement (proclise)*

Par rapport au point (iv) du tableau 5, nous devons signaler que l'interprétation implique le placement d'une trace après le verbe. Le fait qu'un syntagme soit reconnu ensuite et qu'il soit possible de l'attacher au verbe force l'interprétation comme redoublement. Alors, la trace est substituée par le syntagme (*PP* en espagnol et *DP* en grec).

Comme déjà mentionné, la montée longue –« *ellos te la quieren dar* » (10) – est possible en espagnol avec des verbes modaux et avec le verbe « *tener* » (à condition que celui-ci soit suivi de la conjonction « *que* »). Par conséquent, la montée longue requiert que la structure d'arguments du verbe principal soit saturée et que l'argument du verbe infinitif enchâssé soit compatible avec le clitique monté. Pour le traitement des séquences [*Datif, Accusatif*] de la troisième personne (« *le(s) lo(s)* », « *le(s) la(s)* »), tel que nous l'avons signalé lors de la discussion à propos du « faux SE » (voir section 2), le clitique SE doit posséder dans son entrée lexicale le trait datif, de sorte que le filtre sur les clitiques puisse agir correctement et rende possible la présence d'un *SE_{datif}* suivi d'un clitique avec le trait de troisième personne, ce qui est correct en espagnol : « *se lo(s)* », « *se la(s)* ».

6. Conclusions

Dans cet article, nous avons présenté une méthodologie de traitement automatique d'occurrences de pronoms clitiques en espagnol et en grec dans le cadre de l'analyseur syntaxique Fips. Nous avons montré comment les similitudes entre les langues peuvent être exploitées en TAL afin d'améliorer le temps de développement et d'uniformiser les analyses dans un système multilingue. L'analyse tire avantage d'un algorithme de traitement des pronoms clitiques reposant sur l'analyse par création de chaînes (mouvement). Les différences, bien qu'importantes, sont aussi partiellement adaptables à notre stratégie générale, ce qui découle de l'avantage reposant sur la modularité des procédures (cliticisation à travers les langues).

Remerciements

Cette recherche est soutenue par le Fonds national suisse de la recherche scientifique. Nous tenons à remercier le directeur du projet en cours Eric Wehrli, nos collègues Christopher Laenzlinger, Gabriele Musillo, Luka Nerima, Gabriela Soare et Eszter Varga pour leur aide et commentaires précieux ainsi que les collaborateurs du LATL pour leur soutien pendant la durée de ce travail.

Références

- ALEXOPOULOU T. et KOLLIAKOU D. (2002). « On Linkhood, Topicalization and Clitic Left Dislocation ». In *Journal of Linguistic*, 38 (2), 193-245.
- CHOMSKY N. (1995). *The Minimalist Program*. MIT Press, Cambridge.
- HAEGEMAN L. (1994). *Introduction to Government and Binding Theory*. Basil Blackwell, Oxford.
- HOLTON D., MACKRIDGE P. et PHILIPPAKI-WARBURTON I. (2000). *Grammatikí tis Ellinikis Glóssas*. Patakis, Athènes.
- JAEGGLI O. (1986). In H. Borer (éd.), *Syntax and Semantics*, chapter Three Issues in the Theory of Clitics : Case, Double NPs, and Extraction, p. 15-42. Academic Press, Inc. : Orlando.
- KAYNE R. (1975). *French Syntax*. MIT Press, Cambridge, Massachusetts.
- KORDONI V. et NEU J. (2005). « Deep Analysis of Modern Greek ». In K.-Y. S. et al. (éd.), *IJCNLP 2004* : Springer-Verlag. LNAI 3248, Berlin Heidelberg, p. 674–683.
- LAENZLINGER C. et ÉRIC WEHRLI (1991). « FIPS : Un Analyseur interactif pour le français ». In *TA Informations*, 32 (2), 35-49.
- MILLER P. H. et SAG I. (1997). « French Clitic Mouvement without Clitics or Mouvement ». In *Natural Language and Linguistic Theory*, 15, 573-639.
- PERLMUTTER D. (1970). *Deep and Surface Constraints in Syntax*. Holt, Rinehart and Wilson, New York.
- ÉRIC WEHRLI (1997). *L'analyse syntaxique des langues naturelles*. Masson, Paris.
- ÉRIC WEHRLI (2004). *Un modèle multilingue d'analyse syntaxique*, In A. Auchlin, M. burger, L. Filliettaz, A. Grobet, J. Moeschler, L. Perrin et C. R. et Louis de Saussure (éds.), *Structures et discours : Melanges offerts à Eddy Roulet*, p. 311–332. Langue et pratiques discursives. Éditions Nota bene : Canada.