



HAL
open science

Computer-aided learning of transitive non-Locative constructions with a concrete direct object in Modern Greek

Kyriaki Ioannidou, Eleni Tziafa, Rania Voskaki

► **To cite this version:**

Kyriaki Ioannidou, Eleni Tziafa, Rania Voskaki. Computer-aided learning of transitive non-Locative constructions with a concrete direct object in Modern Greek. International Technology, Education and Development Conference, Mar 2013, Valencia, Spain. pp.946-954. hal-01219741

HAL Id: hal-01219741

<https://hal.science/hal-01219741>

Submitted on 10 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMPUTER-AIDED LEARNING OF TRANSITIVE NON-LOCATIVE CONSTRUCTIONS WITH A CONCRETE DIRECT OBJECT IN MODERN GREEK

Kyriaki Ioannidou, Eleni Tziafa, Rania Voskaki

Aristotle University of Thessaloniki (GREECE)
kiroanni@frl.auth.gr, etziafa@lit.auth.gr, rvoskaki@hotmail.com

Abstract

Formalised linguistic data integrated in Natural Language Processing (NLP) tools can prove useful in enhancing learning of Modern Greek as a foreign/second language. Our research focuses on transitive non-locative constructions with a concrete direct object (CDO). For example:

(a) *Η νοσοκόμα κλείδωσε την πόρτα*
(The nurse locked the door)

Our proposal focuses on learning CDO by means of Unitex [1], a corpus processing system based on automata oriented technology. For this purpose, we have elaborated a lexicon-grammar table that includes verbs used in simple sentences containing CDO in Modern Greek [2]. Lexicon-grammar is a syntactic model [3], limited to the description of elementary sentences of a natural language. The description is given in the form of a binary matrix. In our research, this means that a verb has or not a certain property. Moreover, we created a parameterised finite state automaton that constitutes the symbolic representation of a lexicon-grammar table, enriched by finite state automata (FSA) that represent Greek noun phrases [4]. By noun phrases we mean the noun (noun substantive, pronoun or any other nominal) as a head of the noun phrase, along with all its modifiers [5] (p. 243); [6] (pp.127-128, 153-154). For example:

(b) *Η νοσοκόμα κλείδωσε την πόρτα του δωματίου του ασθενή*
(The nurse locked the door of the room of the patient) (Literal translation)
(The nurse locked the patient's room)

We applied these linguistic resources on various authentic corpora (in different domains, such as science, economics, culture, etc.), in the context of Content and Language Integrated Learning (CLIL). We came up with a significant number of concordances, allowing the presentation of CDO within a bilateral context. The concordances are of great utility, since they provide the possibility to study the use of a given construction in large-scale corpora. Therefore, they contribute to heuristic methods of teaching. It is worth mentioning that the proposed method, dealing with CDO learning, is recommended as a complement to learners of levels B2, C1, C2 of the Common European Framework of Reference for Languages (CEFR), set forth by the Council of Europe. Additionally, we have taken into account sample tests, past papers and guides to the examinations for the Certificate of Attainment in Modern Greek, organised by the Centre for the Greek Language.

The present research work is part of the research activities of the Laboratory of Translation & Language Processing in Aristotle University of Thessaloniki. Part of this research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning"- "Investing in knowledge society" of the National Strategic Reference Framework (NSRF) 2007-2013. Research Funding Program: Heraclitus II. Investing in knowledge society through the European Social Fund¹.

Keywords: Computer-aided learning, Modern Greek as a foreign/second language, concordances, NLP.

1 INTRODUCTION

The paper proposes a method of locating transitive non-locative constructions with a concrete direct object (CDO) as an aid to learners of Modern Greek, using Unitex [1], a corpus processing system based on automata oriented technology. For this purpose, we constructed the following linguistic resources: (a) a lexicon-grammar table describing Modern Greek verbs used in CDO simple sentences [2]; (b) a parameterised finite state automaton representing the syntactic and semantic properties included in the lexicon-grammar table; (c) a set of finite state automata (FSA), representing Greek noun phrases [4]. We applied those formalised linguistic resources on authentic electronic texts, including journalistic and educational discourse. We present the obtained results along with the utility of formalised linguistic resources, integrated in Natural Language Processing (NLP) tools, in the framework of enhancing learning of Modern Greek as a foreign/second language. More precisely, the proposed method for CDO learning could serve as a complement to learners for levels B2, C1, C2 of the Common European Framework of Reference for Languages (CEFR), set forth by the Council of Europe. Moreover, we show how the corpus used meets the educational needs of Modern Greek learners, in the context of Content and Language Integrated Learning (CLIL). The above-mentioned matters will be successively presented.

The current research has evolved around the Modern Greek language processing and it is integrated into a program for the development and the gradual completion of the Greek version of Unitex. Unitex is a platform developed for the analysis of texts in natural languages, using linguistic resources and tools. This work pertains to the field of NLP and is a product of the Laboratory of Translation and Language Processing (LTLP), at the School of French Language and Literature, in the Aristotle University of Thessaloniki. LTLP is a member of RELEX², a network of laboratories in the domain of Computational Linguistics.

2 LINGUISTIC DATA FORMALISATION

In order to formalise the linguistic data, we were based on the syntactic framework of the Transformational Grammar defined by Z. S. Harris [7]. The Harrisian theory initiated the Natural Languages formal analysis on the basis of a mathematical model. We have also followed the Lexicon-Grammar methodology framework developed by M. Gross [3] and elaborated at the Laboratory of Automatic Documentation and Linguistics (LADL). According to the Lexicon-Grammar methodology elementary sentences are taken as the minimal unit for the study of meaning and syntax. The predicate is the core of an elementary sentence; in our study, this core is the verb. Each verb has a specific set of arguments (i.e. subject and complements), depending on its meaning. Hence, a lexicon-grammar table incorporates syntactic properties, both distributional and transformational, systematically described. The systematic description consists of matrices whose rows are verbs; columns are syntactic properties into which verbs may enter or not. Matrices are binary: a plus sign appears at the intersection of a given row and a given column, when the verb in the row enters the structure represented in the given column; a minus sign appears in the opposite case.

2.1 Lexicon-grammar table for transitive non-locative constructions with a concrete direct object in Modern Greek

The first step was to list the Modern Greek verbs used in CDO, in order to describe their syntactic properties in a lexicon-grammar table. We came up with 300 verbs, presenting the following basic structure: $N_0 V N1_{conc}$ ³[2]. For example:

- (1a) *Οι οπαδοί ξήλωσαν τις κερκίδες*
(The sports fans banged up **the bleachers**)

In our research, for the term 'concrete object', we mainly adopted the definition given in traditional grammars, considering as 'concrete nouns' a noun category referring to physical world, in opposition

² <http://infolingu.univ-mlv.fr/Relex/introduction.html>

³ N_0 stands for the subject, V for the verb and $N1_{conc}$ for the concrete direct object. In the examples (1a) and (2a) we have marked in bold the CDO.

to 'abstract nouns' (i.e. 1a). However, that criterion is not always pertinent [8]. Therefore, we also considered as 'concrete objects', nouns such as *wound*, illustrated in the following example:

(2a) *Το φάρμακο επούλωσε την πληγή*
 (The medicine healed **the** wound)

Furthermore, we analysed and described the listed verbs, by indicating for each verb its distributional and transformational properties, as shown in the following lexicon-grammar table.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1																									
77	-	+	-	-	δονώ	+	+	+	+	+	+	+	+	δόνηση	-	-	-	-	-	-	-	-	-	-	-
78	-	+	-	-	εγγίζω	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
79	+	-	-	-	εγκαινιάζω	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
80	+	-	-	-	εγκαινιάζω	+	+	+	+	+	+	+	+	εγκαινία	-	+	+	+	+	-	-	-	-	-	-
81	+	-	-	-	εκπονώ	+	+	+	+	+	+	+	+	εκπόνηση	-	-	-	-	-	-	-	-	-	-	-
82	+	+	-	-	ελέγχω	+	+	+	+	+	+	+	+	έλεγχος	-	+	+	+	+	-	-	-	-	-	-
83	+	-	-	-	έμπορος	+	+	+	+	+	+	+	+	εμπορία	-	-	-	-	-	-	-	-	-	-	-
84	+	-	-	-	εξαρθρώνω	+	+	+	+	+	+	+	+	εξάρθρωση	-	+	+	+	+	-	-	-	-	-	-
85	+	+	-	-	επτεξεργάζομαι	+	+	+	+	+	+	+	+	επτεξεργασία	-	+	+	+	+	-	-	-	-	-	-
86	+	-	-	-	επισκευστής	+	+	+	+	+	+	+	+	επισκευή	-	+	+	+	+	-	-	-	-	-	-
87	+	-	-	-	επισκευαστής	+	+	+	+	+	+	+	+	επισκευή	-	+	+	+	+	-	-	-	-	-	-
88	-	+	-	-	επουλώνω	+	+	+	+	+	+	+	+	επούλωση	-	-	-	-	-	-	-	-	-	-	-

Figure 1: Extract of the lexicon-grammar table for CDO in Modern Greek

2.2 Parameterised finite state automaton

In order to incorporate the description provided by the lexicon-grammar table in Unifex, we had to elaborate a parameterized FSA. Parameterized FSA are meta-graphs that allow the automatic generation of a set of graphs on the basis of a lexicon-grammar table. A FSA that describes the possible constructions is constructed manually. That FSA refers to the columns of the lexicon-grammar table in the form of parameters or variables. Afterwards, for each line of the table a copy of this graph is constructed, where the variables are replaced with the contents of the cell at the intersection of line and the column that corresponds to the variable. If a cell of the table contains the + sign, the corresponding variable is replaced by <E>. If the cell contains the - sign, the box containing the corresponding variable is removed, interrupting the paths through that box. In all other cases the variable is replaced by the contents of the cell.

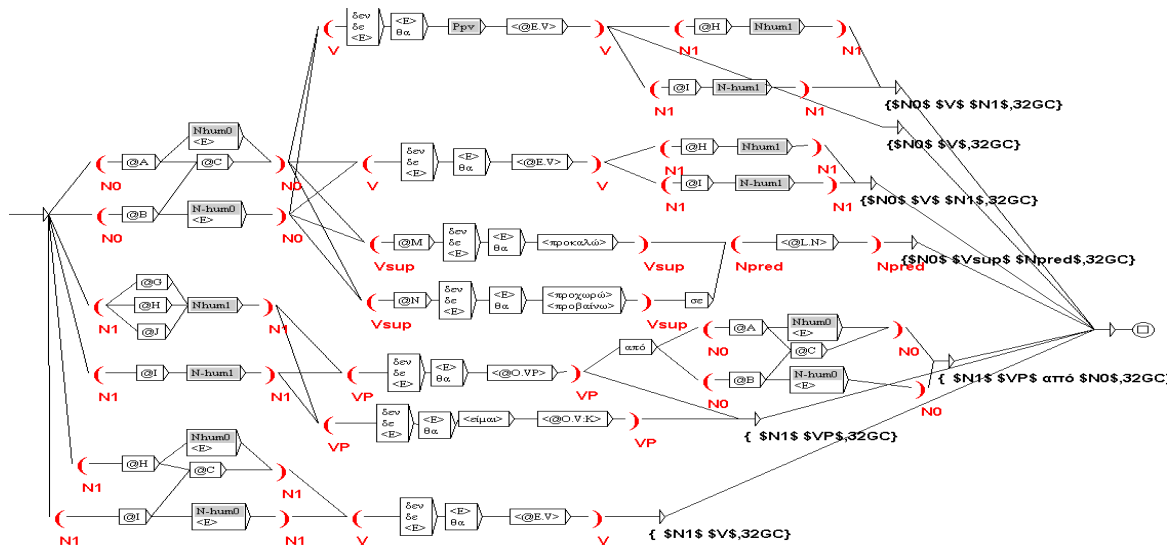


Figure 2: Extract of the parameterised automaton for CDO in Modern Greek

2.3 Finite state automata for Greek noun phrases

The parameterised FSA presented above has been enriched with a set of FSA describing complex noun phrases in object position. Within the scope of our research, by noun phrases we mean the noun (noun substantive, pronoun or any other nominal) as a lexical head of the noun phrase, along with its modifiers [5] (pp. 243); [6] (pp.127-128, 153-154)⁴. For example:

- (1b) *Οι οπαδοί ξήλωσαν τις (λόες τις) κερκίδες και τα κιγκλιδώματα*
[The sports fans banged up **the (/all the) bleachers and the railing**]
- (2b) *Το φάρμακο επούλωσε την πληγή του ασθενή*
(The medicine healed **the wound of the patient**) (Literal translation)
(The medicine healed **the patient's wound**)

Therefore, on the basis of the CDO simple phrases cited in section 2.1, we present in the above-mentioned phrases possible complex noun phrases in object position (marked in bold) that are successfully recognized after the application of the elaborated set of FSA.

In NLP, the task of noun phrase recognition by means of grammars is referred to as *noun phrase chunking*. *Chunking*, also known as *light parsing*, is a method for parsing natural language sentences into partial syntactic structures. This kind of analysis identifies the constituents of a sentence (noun groups, verbs, verb groups, etc.), but does not specify neither their internal structure nor their role in the main sentence.

However, we propose the use of parameterized FSA combined with chunking graphs, in order to take advantage of every type of FSA. On the one hand, the use of parameterised FSA allows us to recognize certain syntactic terms, direct objects in our study, as they are linked to a lexicon-grammar table. On the other hand, chunking automata allow the recognition of a more complex structure of syntactic terms.

More specifically, in the chunking grammars created, we described structures both of base noun phrases and of maximal-length noun phrases [12], [13], [14]. Base noun phrases are non-recursive noun phrases, that is, nouns (pronouns or other nominals) that are determined by determiners, adverbs or other words; however, they are not determined by other noun phrases. As Ramshaw & Marcus [12] explain, "these chunks basically select noun phrases that contain no nested noun phrases". For example:

- (3) *Αποστέρωσε το μικρό μπουκάλι*
(He sterilised the small bottle)

On the contrary, maximal-length noun phrases are base noun phrases, as described above, which are modified by another base noun phrase. As far as maximal-length noun phrases are concerned, the structures included in our descriptions are the following:

a. The first noun phrase in object position (*the house*) is in accusative case, while the second noun phrase that follows (*of their employer*) is in genitive case:

- (4) *Οι εργάτες έκαψαν το σπίτι του εργοδότη τους*
(The workers burnt **the house of their employer**) (Literal translation)
(The workers burnt **their employer's house**)

b. The first (*the doll*) and the second noun phrase (*its clothes*) in object position are linked by coordination:

⁴ For term definitions concerning the English language: [9] (pp. 771); [10] (pp. 332); [11] (pp. 233, 328).

- (5) Πλένω **την κούκλα και τα ρούχα της**
 (I am washing **the doll and its clothes**)

In our research, we have not taken into account the following cases:

a. Base noun phrases modified by prepositional phrases; according to Abney [15]: “noun phrases constitute chunks, but prepositional phrases are considered to belong to a level above the chunk level and are not marked as chunks”. For example:

- (6) Μετράω **τα σπίτια στο λόφο**
 (I am counting **the houses on the hill**)

b. Base noun phrases modified by relative clauses; the recognition of such phrases should be achieved during deep parsing, as they should be syntactically analysed. For example:

- (7) Πληκτρολόγησε **το κείμενο που της έδωσαν**
 (She typed **the text that they gave her**)

We constructed and used 382 FSA graphs in total [4], in order to recognize all possible structures of noun phrases as direct objects. These graphs are recalled by two super-graphs, uniting all graphs concerning nominative and accusative case respectively. The two super-graphs are integrated in the parameterised FSA in the appropriate positions, as described in section 2.2 in this paper. The high number of graphs is due to the presence of cases in Modern Greek. More specifically, in order to avoid morphological ambiguities, we gave a detailed description as far as morphological information is concerned, so as to ensure agreement in case, number and genre for all constituents of the noun phrase. In Fig. 3, we can see a FSA recognizing all possible combinations of adjectives and adverbs, when determining a noun in accusative case.

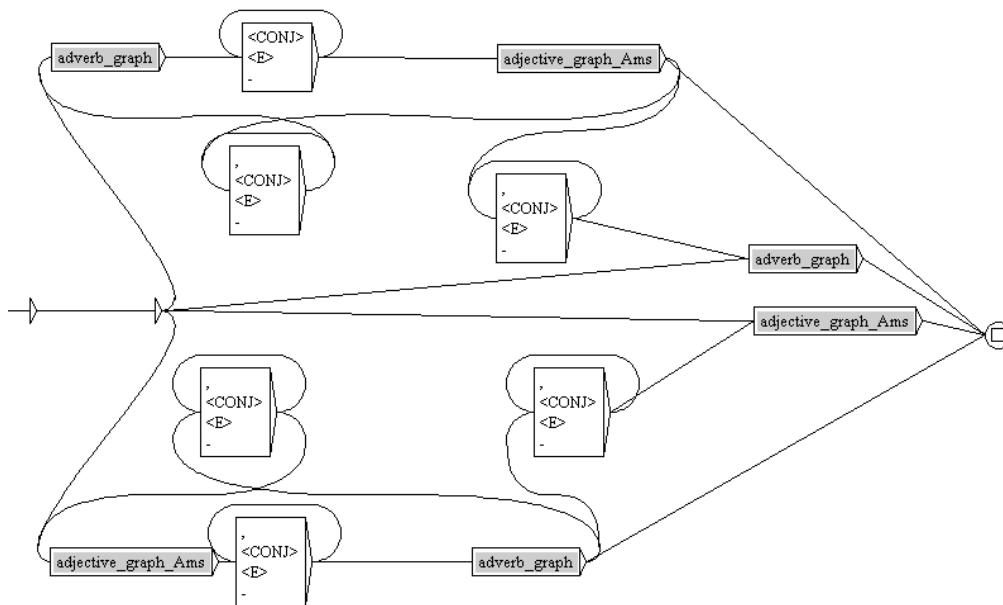


Figure 3: Finite state automaton recognizing all possible combinations of adjectives and adverbs, when determining a noun in accusative case

3 IMPLEMENTATION AND RESULTS

Our next step was to apply the enriched FSA (consisting of the parameterized FSA, presented in section 2.2 and the two super-graphs described in section 2.3) on an authentic corpus. We have to note that Modern Greek is a less resourced language; moreover, the corpora available as full texts are quite a few. The current research is based on the Modern Greek Texts Corpus, which is maintained by the Portal for the Greek Language⁵. It comprises of 7,000,000 words and it consists of three sub-corpora: a text corpus compiled by the newspaper *Ta Nea*, a text corpus compiled by the newspaper *Makedonia* and a text corpus formed by the school books of secondary education, as provided by the Pedagogical Institute. The texts included belong either to journalistic or educational discourse. The journalistic corpus comprises of more than 5,000,000 words (3,000,000 derived from the newspaper *Makedonia* and 2,000,000 words by the newspaper *Ta Nea*). The texts are organised in thematic units and text types. The pedagogical corpus comprises of 2,000,000 words and the texts are derived of students' and teachers' books in secondary education. The texts are also organised in thematic units and text types⁶.

The small size of the corpus is justified as for the study of grammatical structures a larger sample of texts is not considered necessary [16] (pp. 91); [17] (pp. 32). Moreover, smaller text corpora may provide adequate information when the frequency of the phenomenon under study, such as noun phrases, is relatively high [18].

Initially, we preprocessed the corpus, by means of Unitex. The corpus preprocess included the following operations: nominalization of separators, identification of tokens, normalization of non-ambiguous forms, splitting into sentences and application of electronic dictionaries. Then, we applied the enriched parameterized FSA. The corpus processing systems search in texts concrete morphosyntactic structures, which are treated and presented in the form of concordance tables. Concordancers generate a list, including all the words of a given text, chosen on the basis of the dictionary, retrieved alphabetically, and presenting each word within the bilateral context. Concordances are of great utility, because they allow studying the use of a given word in its context. Moreover, they provide information about the sort of context, narrow or broad, where the word occurs.

ει, ό,τι μαθαίναμε να το βλέπαμε και να [αγγίζαμε όλες τις επιφάνειες των σχημάτων](#) . Σύμφωνα με π
μφωνα με πρόσφατες μετρήσεις, η ανεργία [άγγιξε μεγάλο ποσοστό των φτωχών Ελλήνων](#) . Ο Μάρτης κάθε
κάθε μέρα πήγαινε σαν κύριος στο βαρέλι [άνοιγε την κάνουλά του](#) και έπινε το κρασάκι του. Το παι
του και έπινε το κρασάκι του. [Το παιδί άνοιξε](#) το πρωί το παράθυρο για να ξεμυρίσει το δωμάτιο
τιο Έφτιαχνα μόνος μου τον καφέ μου και [έτρωγα το πρωινό και το μεσημεριανό](#) που μου έφερνε η γυ
γητό, οι γείτονες έπιασαν τον κλέφτη κι [έφαγε το ξύλο της χρονιάς του!](#)

Figure 4: Extract of concordances of Modern Greek CDO

In the above extract of concordances, there are six CDO sentences, for three verbs: the first two contain the verb *touch*; the next two refer to the verb *open* and the two last illustrate the verb *eat*. We will comment on each one these results. For the first pair (8 & 9) of sentences, number 8 recognizes a CDO construction. Nevertheless, in sentence 9 the verb *αγγίζω* (*touch*) has a different use from the equivalent verb described in our lexicon-grammar table. Whereas the verb *touch* in our study means the physical contact, in this case, the verb *touch* when preceding a number means *reach*.

(8) *Αγγίζαμε όλες τις επιφάνειες των σχημάτων*

(We touched **all the surfaces of the shapes**)

(Literal translation)

⁵ http://www.greek-language.gr/greekLang/modern_greek/index.html

⁶ For the description of thematic units and text types see section 4.1.

(We touched **all the shapes surfaces**)

(9) *Η ανεργία άγγιξε **μεγάλο ποσοστό των φτωχών Ελλήνων***

(The unemployment touched **a high percentage of poor Greeks**)

As for the second pair of sentences (10 & 11), the first one is right. On the contrary, the second sentence corresponds to the structure “subject – verb – modifier – CDO”. However, the part “subject-verb” is incorrectly recognized as “CDO – verb”, due to the presence of a modifier (*in the morning*), which is not recognized neither by the parameterised FSA nor by the chunking FSA.

(10) *Άνοιγε **την κάνουλά του***

(He opened **the faucet**)

(11) *Το παιδί άνοιξε το πρωί **το παράθυρο***

(The child opened in the morning **the window**) (Literal translation)

(In the morning, the child opened **the window**)

As it concerns the last pair of sentences (12 & 13), the first one recognizes the CDO construction with a noun phrase that contains two nouns linked by coordination. The second sentence is a frozen expression that should be described in the appropriate lexicon-grammar table.

(12) *Έτρωγα **το πρωινό και το μεσημεριανό***

(I ate **the breakfast and the lunch**) (Literal translation)

(I ate **breakfast and lunch**)

(13) *Το παιδί έφαγε **το ξύλο της χρονιάς του***

(The child ate **the wood of his year**) (Literal translation)

(The child was beaten up)

In order to assess the quality of the elaborated FSA (both the parameterised FSA and the super-graphs), two basic evaluation measures should be used: precision and recall. In the context of the present study, precision could be defined as the ratio of CDO over the total number of CDO retrieved by the corpus processing system. Recall could be the ratio of CDO retrieved the corpus processing system over the total number of CDO in the corpus. For the precision percentage, it would be essential to calculate the CDO retrieved from the text, while for the recall percentage it would be required to locate the exact number of correct CDO included in the text. For this purpose, a syntactically tagged corpus is necessary. However, there are no tagged corpora available in Modern Greek, at least not freely available. Therefore, it has not been possible, in the framework of this study, to assess the quality of the FSA we created by recall and precision. Certainly, it is among the perspectives that we will endeavour to fulfill through future research.

4 ENHANCING LEARNING OF MODERN GREEK AS A FOREIGN SECOND LANGUAGE

In the framework of Information Communications Technology in education (ICT), formalised linguistic data integrated in NLP tools can prove useful in enhancing learning of Modern Greek as a

foreign/second language. Awareness is very important during foreign language learning, but awareness without input is not sufficient [19]. Therefore, the use of NLP tools is more than necessary.

More precisely, we propose learning of CDO by means of Unitex⁷, because it is an open-source system, operating for various natural languages; among others for Modern Greek. Linguistic resources (i.e. electronic dictionaries) as well as basic grammars in FSA form (i.e. sentence segmentation) are integrated in the system. Currently, the open-source version contains 30% of the elaborated (by LTLP) electronic dictionaries, a sample journalistic corpus and the basic grammars. However, we intend to incorporate more linguistic data in the open-source version.

The advantage of the corpus processing system is that the user has the possibility to add any corpus available, any corpus that fulfils the educational needs of his foreign language learning class. The user should only preprocess the added corpus, convert it in unicode format and apply the electronic dictionaries. This procedure provides a normalised corpus, segmented in sentences, with recognized lexical units. Every electronic dictionary elaborated to date presents a small percentage of unrecognized lexical units (less than 10% in journalistic corpus). By applying linguistic data, formalised as FSA, the user comes up with a significant number of concordances. The concordances are of great utility, since they provide the possibility to study the use of a given construction in authentic and/or large-scale corpora. Therefore, they contribute to heuristic methods of teaching [2] (pp.212-222).

4.1 Content and Language Integrated Learning

In the current research we used a journalistic and an educational discourse corpus. Both of them are classified in thematic units and text types [20]. More precisely, the journalistic corpus contains the following thematic units: (i) economics, (ii) book presentation, (iii) visual arts, (iv) arts and culture. In addition, it consists of the text types as follows: (i) sports news, (ii) reportage, (iii) gastronomy, (iv) interview, (v) advertisements, (vi) science news, and (vii) artistic reviews. As for the educational corpus it mainly comprises of various thematic units, such as: (i) biology, (ii) history, (iii) translated Ancient Greek texts, (iv) physics, (v) chemistry, (vi) economics theory, (vii) Modern Greek learning, (viii) religion and (ix) mathematics. In the text types' classification, it is divided in: (i) biographies, (ii) curriculum vitae, (iii) short stories, (iv) stories, (v) forecast, (vi) dialogs, (vii) interviews, (viii) advertisements, (ix) research and (x) e-mail.

The classification shows that it is characterised by a significant variety of different thematic fields, which principally pertain to curricular subjects. Therefore, it can contribute to the cross-cultural awareness of the Modern Greek learner through the medium of a language other than that normally used; we are especially referring to the educational corpus. Moreover, it can be effective, as a language learning complement, for the literacy of all levels of students, from teenagers (often more familiar with terminology included in curricular subjects) to adults. As it was recently set forth by researchers, in charge of the design of the exams for the Attainment in Modern Greek [21], CLIL is also taken into consideration while preparing exams.

4.2 Common European Framework of Reference for Languages

Teaching and/or learning of CDO in Modern Greek by means of Unitex and the formalised linguistic data described in the present study could serve as a complement to learners of B2, C1, C2 levels of the CEFR, set forth by the Council of Europe. In order to obtain a Certificate of Foreign Language Attainment, the candidate is examined, as known, on a set of skills. One of them is the examination of grammar accuracy. As for the C2 level, the use of complex structures is strongly required. For C1 level, the equivalent requirements are almost the same, allowing a few structural errors. Finally, in B2 level circumstantial structural oversight is allowed [22] (pp. 171). CDO in Modern Greek present a high degree of complexity, mainly in the case of complex noun phrases in object position, as they are presented in 2.3 of the present paper.

We have also taken into account sample tests, past papers and the guide [23] to the examinations for the Certificate of Attainment in Modern Greek, organised by the Centre for the Greek Language. More precisely, we studied past papers used for B2, C1 and C2 levels the last five years. Activities requiring the right order of a given text have been observed in those recent past papers. Those types of activity involve awareness of the language structures.

7 <http://www-igm.univ-mlv.fr/~unitex/>

5 CONCLUSIONS AND PERSPECTIVES

In this paper, we have described the procedure of recognition of CDO in Modern Greek by means of a corpus processing system based on automata oriented technology (Unitex). A lexicon-grammar table, a parameterised FSA and a set of FSA for Greek noun phrases have been utilised to track CDO in authentic corpora. We have also focused on the utility of the obtained concordances in enhancing learning of Modern Greek as a foreign/second language for learners of levels B2, C1, C2 of the CEFR.

Perspectives that we will endeavour to fulfil through future research are as follows: enrichment of the existing corpora; improvement of the FSA, in order to eliminate, as much as possible, ambiguities; tagging syntactically part of the corpus, so as to be able to evaluate the quality of the obtained results, by recall and precision. In addition, expanding recognition of other types of direct objects within the transitive non locatives constructions in Modern Greek –such as ‘human object’, ‘body part object’, ‘obligatory plural object’– would be useful in computer-aided learning of Modern Greek as a foreign/second language.

REFERENCES

- [1] Paumier, S. (2003). *De la reconnaissance de formes linguistiques à l'analyse syntaxique*. Thèse de doctorat, Paris, Université de Marne-la-Vallée.
- [2] Voskaki, O. (2011). *Le lexique-grammaire des verbes du grec moderne : Les constructions transitives non locatives à un complément d'objet direct*. Thèse de doctorat, Paris, Université Paris-Est.
- [3] Gross, M. (1975). *Méthodes en syntaxe. Régime des constructions complétives*. Paris : Hermann.
- [4] Ioannidou, K. (to appear). *Noun phrases in Modern Greek: Morphological disambiguation during automatic text analysis with applications to translation*. [in Greek]. Thesis, Thessaloniki, Aristotle University of Thessaloniki.
- [5] Holton, D., Mackridge, P. & Filippaki-Warburton, E. (2002). *Grammar of Greek language* [in Greek]. Athens: Pataki.
- [6] Clairis, Ch. & Babiniotis, G. (2009). *Grammar of Modern Greek. Structural-Functional-Communicative* [in Greek]. Athens: Ellinika grammata.
- [7] Harris, Z. S. (1968). *Mathematical Structures of Language*, New York, Wiley
- [8] Flaux, N. & Van de Velde, D. (2000). *Les noms en français, esquisse de classement*. Paris: Ophrys.
- [9] Sinclair, J. M. (1993). *Collins English Dictionary and Thesaurus*. Glasgow: HarperCollins Publishers.
- [10] Bussmann, H. (1996). *Routledge Dictionary of Language and Linguistics*. London & New York: Routledge.
- [11] Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*.
- [12] Ramshaw, L. A. & Marcus, M. P. (1995). "Text Chunking using Transformation-Based Learning". *ACL Third Workshop on Very Large Corpora* , pp. 82-94.
- [13] Tjong Kim Sang, E. F. (2000). "Noun Phrase Recognition by System Combination". *Proceedings of the 1st North American chapter of the Association for Computational Linguistics Conference* (pp. 50-55). Stroudsburg, PA: Association for Computational Linguistics.
- [14] Bai, X.-M., Li, J.-J., Kim, D.-I. & Lee, J.-H. (2006). "Identification of Maximal-Length Noun Phrases Based on Expanded Chunks and Classified Punctuations in Chinese". *21st International Conference on the Computer Processing of Oriental Languages 2006* (pp. 268-276). Berlin: Springer-Verlag.
- [15] Abney, S. (1996). *Chunk stylebook*. Technical report, SfS, University of Tübingen.
- [16] Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- [17] Givón, T. (1995). *Functionalism and grammar*. Amsterdam: John Benjamins.
- [18] McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London and New York: Routledge.

- [19] Schmidt, R. (1995). "Consciousness and foreign language: A tutorial on the role of attention and awareness in learning". In R. Schmidt (ed.), *Attention and awareness in foreign language learning*, Honolulu: University of Hawaii Press.
- [20] Politis, P. (2006). *Enhancement and extension of electronic corpora in the Portal for the Greek Language*. [In Greek], <http://goo.gl/Tm7hA>. Date of access 24 January 2013.
- [21] Dimitrakopoulou, M., Karakirgiou, M., Kokkinidou, A., Pranentsi, K. & Roussoulioti, Th. (2012). "Content and Language Integrated Learning and assessment of Modern Greek as second/foreign language: tends and perspectives". [Oral presentation], *15th International Conference on Applied Linguistics, 23-25 November 2012*. Thessaloniki, Greece.
- [22] Eustathiadis, S. & Tsagkalidis, A. (2008). *Common European Framework of Reference for Languages. Language learning, language teaching, assessment*. [Trans into Greek], Thessaloniki: Centre for the Greek Language.
- [23] Eustathiadis, S., Antonopoulou, N., Manavi, D. & Voyatzidou, S. (2001). *Certificate of Attainment in Modern Greek. An extensive examination program*. [In Greek], Thessaloniki: Centre for the Greek Language.