Essex Graduate Student Papers in Language & Linguistics 11 (2009), 1 – 20

Implementing a fragment of Modern Greek Grammar, using the Xerox Linguistics Environment (XLE)

Kakia Chatsiou[1]

Abstract

This paper presents a computational grammar of a fragment of Modern Greek, following the principles of the *Lexical Functional Grammar* (LFG*) Parallel Grammar* (ParGram) *Project* (P.A.R.C, 2008) a collaborative effort among researchers in industrial and academic institutions whose objective is to build wide coverage deep-parsing grammars for a wide variety of languages. The grammar is built using the *Xerox Linguistics Environment* (XLE) parser (P.A.R.C, 2009) and at the moment covers the syntax of basic clause and word order phenomena in Modern Greek, and the syntax of *pu*-Restrictive Relative Clauses, with particular focus on the distribution of the gap/resumptive relativisation strategy.

In our paper, we present a brief overview of the XLE system, and the *Parallel Grammar (ParGram)* initiative. We present the fragment of Modern Greek, focusing on the coverage and the main assumptions underlying the current version of the grammar. We conclude by evaluating our grammar and discussing areas in need of immediate improvement to be dealt with in future versions as well as some future development directions.

1 Introduction

This paper presents a computational grammar of a fragment of Modern Greek, built following the principles of the *Lexical Functional Grammar* (LFG*) Parallel Grammar* (ParGram) *Project* (P.A.R.C, 2008), a collaborative effort among researchers in industrial and academic institutions around the world whose aime is to produce wide coverage deep-parsing grammars for various languages. The Modern Greek grammar is built manually using the *Xerox Linguistics Environment* (XLE) parser (P.A.R.C, 2009a) and the current version covers the syntax of basic clause and word order phenomena, and thesyntax of *pu*-Restrictive Relative Clauses, with particular focus on the distribution of the gap/resumptive relativisation strategy.

The paper is organised as follows: in section 2, we present a brief overview of the XLE system and an overview of the *Parallel Grammar (ParGram)* initiative. Section 3 presents

[1] Department of Language and Linguistics, University of Essex, UK, email: achats@essex.ac.uk

the grammar fragment of Modern Greek, illustrating the coverage and the main assumptions underlying the current version as well as a brief evaluation of the system. Finally, in section 4 we present areas of possible improvement to for future versions as well as future development directions.

2 About the Xerox Linguistics Environment (XLE)

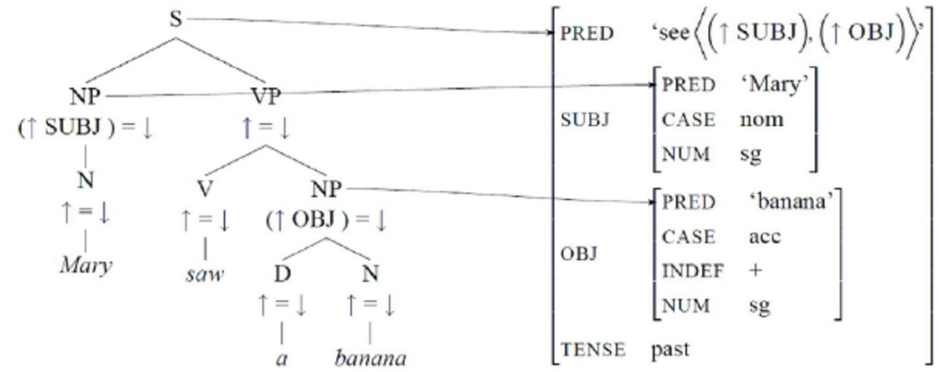
The *Xerox Linguistics Environment (XLE)* is a platform for implementing Lexical Functional Grammars (LFG). LFG is a theory of grammar initially set forth in Bresnan & Kaplan (1982): It is a *lexical* theory, since the lexicon plays an active role when accounting for linguistic phenomena and *functional* since it uses primitive grammatical functions like SUBJ(ect), OBJ(ect) and OBL(ique) to account for the grammatical role of each element in the sentence.

The basic mechanism behind the LFG formalism is the existence of different levels of representation. In their paper, Bresnan & Kaplan (1982) defined two levels of representation: the *constituent structure (c-structure)* and the *functional structure (f-structure)*. The c-structure is where dominance and precedence relations between constituents are expressed and is represented via a phrase structure tree (Dalrymple, 2001:92). The *f-structure*, is where grammatical relations are represented. It is reserved for encoding more abstract syntactic notions such as grammatical functions, case, agreement and generally "everything apart from categorical status, linearization and dominance" (Asudeh, 2004:38) and is represented by an attribute value matrix (AVM)[2]. Examples of

[2] The validity of the f-structure representation is ensured by complying to a number of well-formedness conditions: the *consistency/uniqueness condition* which ensures that "each attribute in each f-structure will have at most one value" (Dalrymple, 2001: 39); the *completeness condition* which ensures that all governable elements (such as SUBJ, OBJ and so on) are realised and that if one of the elements of the argument list is missing, the f-structure will be *incomplete* and will be ruled out as ill-formed, (Dalrymple, 2001:37) ; finally the *coherence condition* which certifies that there are no additional governable elements in the f-structure and that the presence of an extra governable grammatical function in the f-structure results in its being ruled out as *incoherent* (Dalrymple, 2001:39). The properties of f-structures will not concern us here in detail; for further information see among others Bresnan (2001), Dalrymple (2001), Falk (2001) and Sells (1985:44-46). It is worth noting that c-structures may vary among languages, but their corresponding f-structures are quite similar. This is an observation at the core of all ParGram projects, which aim at the creation of 'parallel' grammars keeping the f-structure as similar as possible across languages, and using the c-structure to depict the structural differences across languages.

the c-structure of the sentence *Mary saw a banana* and its corresponding f-structure are shown in (1) below:

(1)

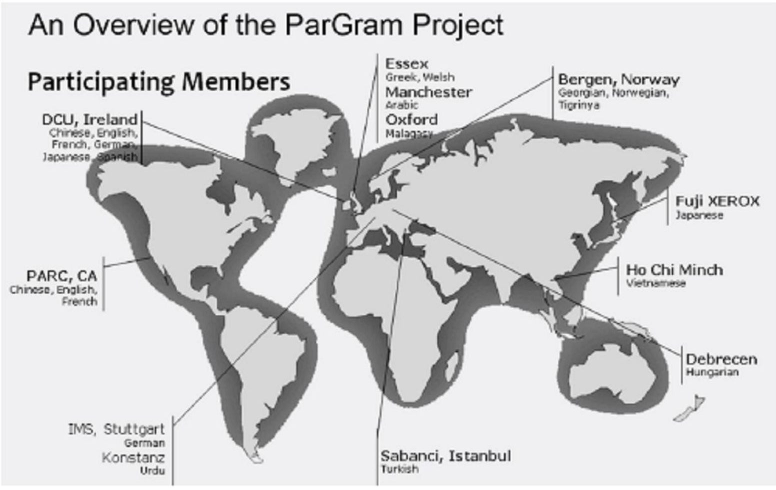


The c-structure is linked to the f-structure by means of the φ-projection, represented as a set of f-structure annotations on the c-structure nodes. F-structure information is passed on to the mother node using the $\uparrow$ from the daughter node ($\downarrow$). For example, the $(\uparrow \text{SUBJ}) = \downarrow$ notation under the subject NP node, indicates that the f-structure of the current node will be part of the SUBJ f-structure of the mother S node. A detailed presentation of the LFG formalism and the way the mapping from the c- to the f-structure works, goes beyond the scope of this paper; the reader is thus referred to Bresnan (2001), Dalrymple (2001) and Falk (2001) among others for an introduction to the theory and its latest developments.

XLE is a platform for developing such grammars developed at *Palo Alto Research Center (PARC).* It is implemented in *C* and is available under Unix, Linux and MacOS operating systems. XLE includes a parser, a generator, and a finite state morphological analyser and it can be used both for parsing and generation of natural languages. It includes tools for other grammar development activities, such as performance analysis and test-suites and has built-in debugging, grammar maintenance and finite state tools to facilitate the job of the grammar developers.

XLE has been used for a range of Natural Language Applications ranging from Machine translation, using the *Transfer* System (P.A.R.C, 2007), to *Computer Assisted Language Learning* (Butt and King, 2007). XLE has been used by researchers involved in the *Parallel Grammar (ParGram) project*, with academic and industrial participating members from across the world[3]. Figure 1 shows the participating members' locations, as well as the languages they have been working on (as of January 2009).

Figure 1: ParGram Participating Sites



XLE has also been used as the core technology employed in a novel search engine[4] which aims at improving the way we find information by enabling the user to form queries using natural language. It is currently under development at *Powerset*, a company recently acquired by Microsoft. Powerset is using *FreeBase* as its semantic knowledge database, and its technology is currently being used to improve the searching experience in M*icrosoft's Bing* (www.bing.com) search engine.

---

[3] Some of the project's objectives include building broad coverage grammars which will parse and generate a wide range of a language's phenomena, and providing linguistically motivated analyses for the phenomena under consideration. All grammars are guided by a common set of linguistic principles and a commonly agreed-upon set of grammatical analyses and features as well as a similar treatment of core cross-linguistic phenomena. Finally, with respect to the methods used in grammar engineering, all members apply a common set of methods and evaluation strategies and at the same time try to achieve a balance between efficiency, performance, readability and maintainability across grammars.

[4] Powerset's search engine (www.powerset.com) aims at improving users' searching experience of Wikipedia by allowing themto type full questions/sentences in the search box as well as keywords. On the results page, the user gets a summary of the search results compiled from different articles.

Let us now have a closer look at the XLE implementation of the Modern Greek grammar fragment.

2 XLE Implementation of a fragment of Modern Greek

The current version of the fragment is a preliminary effort to develop a large-scale LFG Computational grammar for Modern Greek. Being built following the principles underlying similar Parallel Grammar projects, it shares the objectives and principles outlined above, aiming at being *parallel* to similar projects for other languages as well as balancing maintainability and achieving large coverage. The current main focus is on the syntactic rules and thus the lexicon is kept as minimal as possible. Future versions are expected to focus on its expansion employing the use of the XLE's built-in Finite State Morphological analyser.

2.1 Some assumptions

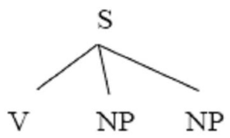
One of the main assumptions underlying the current version of the fragment concern Modern Greek constituent order. Contrary to the standard view proposed in the literature[5], we assume just for the current fragment that all possible word orders (such as VSO, SVO, OSV and OVS) in declarative main clauses are equally acceptable and grammatical. This is rather simplifying things, since the degree of acceptability of the different word orders varies across speakers; such a simplification was necessary since the main focus of the implementation lied on the implementation of *pu*-Restrictive Relative Clauses. Future versions will certainlly refine the grammar to account for these differences.

---

[5] Although there seems to be an overall agreement in the literature concerning VSO as the *basic constituent order of subordinate clauses* (Tzartzanos, 1963:276; Lascaratou, 1998:161; Mackridge, 1985:237) and the rather fixed constituent order within a nominal phrase (Markantonatou, 1992:255-256; Lascaratou, 1998:63), there seems to be great controversy with regards to constituent order in declarative sentences. As Holton et al (1997:426) point out, due to its rich morphological marking system, Modern Greek demonstrates a relative freedom in the way constituents are ordered within an independent clause, as seen in example (3) where each constituent order will produce well-formed (but not equally acceptable for all speakers) sentences.

Following recent proposals by some scholars (Alexopoulou, 1999; Tsiplakou, 1998; Tzanidaki, 1996 among others), who have argued against a configurational account for Modern Greek, based on evidence from the similar status of subject and object (Tzanidaki, 1996), the absence of dummy subjects (Alexopoulou, 1999:7) and the availability of VP ellipsis (Alexopoulou, 1999; Tsiplakou, 1998), we represent Modern Greek word order *non-configurationally*, similarly to the representation in (2):

(2)



Our grammar fragment presently focuses on building rules of the syntax. We have also not accounted for the morphology of the lexical items in the lexicon section in the current version, but instead, we have introduced a separate lexical entry for each different form according to case, gender, number and person.

2.2 Fragment Coverage

In this section we present the current grammar fragment coverage. Our grammar accounts for basic word order phenomena, basic agreement patterns (like subject-verb agreement and internal DP agreement), basic subcategorization frames and account for the pro-drop character of the language. To these, we added the LFG analysis of *pu*-Restrictive Relatives and the distribution of the gap/resumptive strategy in local and long distance dependences presented in Chatsiou (in preparation). Sections 2.2.1 and 2.2.2 include a discussion of the phenomena implemented in the fragment and how we went about implementing them in XLE.

2.2.1 Phenomena treated in the c-structure

Our fragment accounts for all possible word orders of declarative clauses, as illustrated in example (3):

(3) a. VSO
taise i yineka ton papagalo
fed$_{.3SG}$ the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$ the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$
b. SVO
i yineka taise ton papagalo
the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$ fed$_{.3SG}$ the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$
c. OSV
ton papagalo i yineka taise
the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$ the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$ fed$_{.3SG}$
d. OVS
ton papagalo taise i yineka
the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$ fed$_{.3SG}$ the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$
e. VOS
taise ton papagalo i yineka
fed$_{.3SG}$ the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$ the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$
'The woman fed the parrot.'
f. SOV
i yineka ton papagalo taise
the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$ the$_{.MSG.ACC}$ parrot$_{.MSG.ACC}$ fed$_{.3SG}$
'The woman fed the parrot.'

All these c-structures share the same f-structure, shown in (4) below[6]:

(4) f-structure of *'The woman fed the parrot.'*

```
 [PRED   'feed<[1:woman], [7:parrot]>'                  ]
 [SUBJ   [PRED 'woman'                               ]  ]
 [      1[CASE nom, DEF +, GEND f, NUM sg, PERS 3]   ]  ]
 [OBJ    [PRED 'parrot'                              ]  ]
 [      7[CASE acc, DEF +, GEND m, NUM sg, PERS 3]   ]  ]
5[TENSE past                                            ]
```

[6] The reader familiar with LFG, might find that this f-structure looks a bit different from standard LFG notation (also compare with the f-structure in (1)). XLE's output f-structure has `[1:woman], [7:parrot]` where one would expect $(\uparrow \text{SUBJ})(\uparrow \text{OBJ})$. This is just a convention; `[1:woman]` points to the f-structure of the *woman* predicate, and is re-entrant with the SUBJect's f-structure (both have 1 as their index).

This is implemented using the *shuffle operator* (P.A.R.C., 2009b) which 'shuffles' the elements on the right-hand side of the S rule. The syntax of this operator is illustrated in (6)[7]:

(5) `S --> DP1; V; DP2.`

(6) `S --> [ DP1 ], [ V ] , [ DP2 ].`

Modern Greek is a pro-drop language, as shown in (7):

(7) petai
fly.3SG
'S/he flies.'

This is achieved quite straightforwardly, by making the subject-DP optional in the c-structure rules and by adding an optional equation on the lexical entry of the verbs that assigns a PRED value to the SUBJ f-structure in case this is not present otherwise, as in (8)[8].

(8) *(the optional subject DP in the S rule)*
`S --> [ ( DP : (^ SUBJ ) = ! ) ] , ...`

(The lexical entry of a pro-drop verb)

---

[7] The S rule in (5) illustrates the ordinary XLE syntax for writing phrase structure rules and succeeds for any string of elements containing a DP1, followed by a V and a DP2 in that order. The S rule in (6), however, succeeds for any string of elements, provided that it contains a DP1, a V and a DP2 in any order. This is indicated by including the elements we wish to 'shuffle' in square brackets (`[]`) and separating them with a comma (`,`) as opposed to separating them with a semicolon (`;`), as shown in (5). Thus, the rule in (6) can be satisfied by any of the following orders:

(i) DP1 - V - DP2
DP1 - DP2 - V
V - DP1 - DP2
V - DP2 - DP1
DP2 - V - DP1
DP2 - DP1 - V

[8] Again here, XLE notation slightly deviates from the standard LFG one: ^ corresponds to the ↑ arrow; ! corresponds to the ↓ arrow. Note that the way we denote optionality of constituents in rules, marked with round brackets ( ) is different from denoting optionality of the f-structure annotations, which is marked with curly brackets { }. The same curly brackets denote disjunction when they appear in a rule, as in (13). Finally $ stands for the ∈ (element) notation.

```
petai V *   (^ PRED)='fly<(^SUBJ)>'
            (^ SUBJ NUM) = SG
            (^ PERS) = 3
            {(^ SUBJ PRED) = 'pro'}
            (^ TENSE) = present.
```

The current fragment also includes an implementation of the analysis of Modern Greek *pu*-Restrictive Relative Clauses presented in Chatsiou (in preparation), where we put forward an LFG analysis of the treatment of the distribution of the gap/resumptive relativisation strategy in Modern Greek Restrictive, Non-restrictive and Free Relative Clauses.

A detailed examination of the characteristics of relative clauses goes beyond the scope of this paper; we will however briefly refer to some of the most important characteristics that are of interest to the implementation[9]. One of them is the internal constituent order of *pu*-RRCs. In particular, contrary to the controversy that the same issue has raised for independent declarative clauses (Tzartzanos, 1963; Siewierska et al, 1998; Philippaki-Warburton, 1985; Tsimpli, 1996; Holton et al, 1997; Alexopoulou, 1999), it is generally agreed in the literature that the 'basic' or underlying constituent order of relative clauses is relatively fixed (Tzartzanos, 1963:276; Mackridge, 1985:237). As shown in (9) and illustrated in examples (10) and (11), *pu*-Restrictive Relative clauses are introduced by a complementizer or a relative pronoun, followed by a resumptive pronoun[10], followed by the verb of the relative clause, and by zero or more instances of any nominal or adverbial elements in any order.

(9) *complementizer/relative pronoun* + (resumptive pronoun) + V + XP*

(10) o papagalos pu edose o andras tis yinekas
the$_{\text{.MSG.NOM}}$ parrot$_{\text{.MSG.NOM}}$ that gave$_{\text{.3SG}}$ the$_{\text{.MSG.NOM}}$ man$_{\text{.MSG.NOM}}$
the$_{\text{.FSG.GEN}}$ woman$_{\text{.FSG.GEN}}$
'The parrot that the man gave to the woman.'

[9] The reader is referred to Alexiadou & Anagnostopoulou (2000), Alexopoulou (2006, 2008), Alexopoulou & Keller (2007), Ingria (1981), Joseph (1980), Klein (1993), Varlokosta (1997, 1998, 1999), Varlokosta & Kotzoglou (2003) for alternative analyses of *pu*-Restrictive Relative Clauses in frameworks other than LFG, and to Asudeh (2004), Dalrymple (2001), Falk (2002) for LFG analyses of Restrictive Relative clauses in other languages.

[10] Resumptive pronouns in pu-RRCs can be obligatorily present, optional or obligatorily absent, depending on the context. See ex. (25) for a table presenting the distribution of resumption in *pu*-RRCs.

(11) o papagalos pu edose tis yinekas o andras
the$_{\text{.MSG.NOM}}$ parrot$_{\text{.MSG.NOM}}$ that gave$_{\text{.3SG}}$ the$_{\text{.FSG.GEN}}$ woman$_{\text{.FSG.GEN}}$ the$_{\text{.MSG.NOM}}$ man$_{\text{.MSG.NOM}}$
'The parrot that the man gave to the woman.'

The elements following the verb may occur in any order[11]; the complementizer, the resumptive pronoun and the verb, however, should occur in that order. We capture these two different behaviours by using the declarative clause S in the C' rule, in which all elements are 'shuffled' using the *shuffle operator* (`[]`) for the elements to appear in free word order after the V. The complementizer, the resumptive, the verb and the antecedent DP appear in fixed order. This is why they appear outside the shuffling operator as illustrated in (12) [lines 3-9].

(12)
```
1    DP --> D; N'.
2    N' --> N; (CP: !$ (^ ADJUNCT)).
3    CP --> {e: {(ADJUNCT $ ^)
4                (^ COMPFORM)= pu
                       ⋮
7               |(^ COMPFORM)=oti}
8           |DP};
9           C'.
10   C' --> C; S.
11   S --> [(DP: (^ SUBJ)=!)],
12         [(V; CP:(^ COMP)=!)],
13         [(NP: (^ SUBJ|OBJ|OBJ2)=! (! PRONTYPE)=c rp);
14          V],
15         [(DP: (^ OBJ)=!)],
16         [{(DP: (^ OBJ2)=!)
17          |(PP: (^ OBL)=! (! CASE)=acc (!PFORM)=cse)}].
```

2.2.2 Phenomena treated in the f-structure

The fragment accounts for some basic subcategorization frames (transitive, intransitive and ditransitive verbs including the realisation of indirect objects as either a genitive DP or an accusative PP$_{se}$) as illustrated in examples (13) to (16):

[11] As previously explained, each of the possible orders differs in terms of their degree of markedness; we will however assume here that all these orders are equivalent.

(13) o papagalos petai
the.$_{\text{MSG.NOM}}$ parrot.$_{\text{MSG.NOM}}$ fly.$_{\text{3SG}}$
'The parrot flies.' (*intransitive*)

(14) i andres taisan tus papagalus
the.$_{\text{MPL.NOM}}$ men.$_{\text{MPL.NOM}}$ fed.$_{\text{3PL}}$ the.$_{\text{MPL.ACC}}$ parrot.$_{\text{MPL.ACC}}$
'The men fed the parrots.' (*transitive)*

(15) edose i yineka ton papagalo ston andra
gave.$_{\text{3SG}}$ the.$_{\text{FSG.NOM}}$ woman.$_{\text{FSG.NOM}}$ the.$_{\text{MSG.ACC}}$ parrot.$_{\text{MSG.ACC}}$ to.the.$_{\text{MSG.ACC}}$ andra.$_{\text{MSG.ACC}}$
'The woman gave the parrot to the man.' (*ditransitive with PP$_{se}$*)

(16) i yineka edose ton papagalo tu andra
the.$_{\text{FSG.NOM}}$ woman.$_{\text{FSG.NOM}}$ gave.$_{\text{3SG}}$ the.$_{\text{MSG.ACC}}$ parrot.$_{\text{MSG.ACC}}$ the.$_{\text{MSG.GEN}}$ man.$_{\text{MSG.GEN}}$
'The woman gave the man the parrot.' (*ditransitive with NP$_{gen}$*)

The VP rule below summarizes the four subcategorization frames:

(17)

```
VP --> [  V: ^=! ],
       [( DP: (^ OBJ)=! )],
      [({ DP: (^ OBJ2)=! (! CASE )=gen
        |  PP:  (^  OBL)=!  (!  CASE)=acc  (!PFORM)=c
    se})].
```

Both the DP$_{gen}$ and the PP$_{se}$ are alternative manifestations of the indirect object, but they are assigned a different grammatical function: the genitive DP is an OBJ2 and the PP introduced by the *se* particle is an OBLique.

Examples like (18) are successfully ruled out by application of the *coherence condition* (Dalrymple, 2001:39) using information from the lexical entry of the verb *petai* (flies) (cf. ex. (9)), which ensures that there are no additional governable elements in the f-structure and that the presence of an extra governable grammatical function (in this case the extra OBJ) in the f-structure results in its being ruled out as *incoherent*:

(18) * petai o papagalos tin yineka
flies.$_{3SG}$ the.$_{MSG.NOM}$ parrot.$_{MSG.NOM}$ the.$_{FSG.ACC}$ woman.$_{FSG.ACC}$
'*The parrot flies the woman.'

The grammar successfully assigns the appropriate case to nominal elements depending on the requirements of the verb[12], as illustrated in examples (19) and (20), successfully ruling out examples like (21) by application of the *consistency/uniqueness condition* which ensures that "each attribute in each f-structure will have at most one value" (Dalrymple, 2001:39):

(19) o andras taise ton papagalo
the.$_{MSG.NOM}$ man.$_{MSG.NOM}$ fed.$_{3SG}$ the.$_{MSG.ACC}$ parrot.$_{MSG.ACC}$
'The man fed the parrot.'

(20) ton papagalo taise o andras
the.$_{MSG.ACC}$ parrot.$_{MSG.ACC}$ fed.$_{3SG}$ the.$_{MSG.NOM}$ man.$_{MSG.NOM}$
'The man fed the parrot.'

(21) * o papagalos taise o andras
the.$_{MSG.NOM}$ parrot.$_{MSG.NOM}$ fed.$_{3SG}$ the.$_{MSG.NOM}$ man.$_{MSG.NOM}$
'The man fed the parrot.' (*intended meaning*)

This is accounted for lexically, on the template for each verb frame, as in the example below:

(21a) *Lexical entry for edose:*

```
edose       V *    @(DTR gave)
                   @subj-3sg
                   (^ TENSE)=present.
```

(21b) *Templates*

```
DTR(P) =
       { (^ PRED)='P<(^ SUBJ)(^ OBJ)(^ OBJ2)>'
        |(^ PRED)='P<(^ SUBJ)(^ OBJ)(^ OBL)>'}
       { (^ SUBJ PRED)='pro'}
         (^ SUBJ CASE) = nom
         (^ OBJ CASE)= acc.
```

[12] Usually – but not always – nominative for subjects, accusative for objects or objects of the PP$_{se}$, genitive for indirect objects.

On the DP level, our grammar accounts for number, case and gender agreement within a DP or a PP, as in example (22), successfully ruling out ungrammatical examples like (23). This is achieved again by application of the *consistency/uniqueness condition* as illustrated in the f-structure in (23), where the f-structure is ruled out as ungrammatical, since there are more than one values for the same feature (NUM) in a given f-structure:

(22) tis yinekas
the.$_{\text{FSG.GEN}}$ woman.$_{\text{FSG.GEN}}$
'of the woman'

(23) *tis yineka
the.$_{\text{FSG.GEN}}$ woman.$_{\text{FSG.ACC}}$
'of the woman'

$$
{}_{1}\begin{bmatrix} \text{PRED 'woman'} \\ \text{NUM} \quad \left[= \begin{array}{|c|} \hline \text{pl} \\ \hline \text{sg} \\ \hline \end{array} \right] \\ \text{CASE nom, GEND f, PERS 3} \end{bmatrix}
$$

With respect to the implementation of the analysis of the gap/resumptive strategy in local and long distance dependencies in *pu*-RRCs, we opted to account for the fact that the resumptive pronoun has the same form as the unstressed monosyllabic clitic (weak form) of the personal pronoun and the definite article in the lexicon. As shown in (24), this is treated using a disjunction (indicated by the ; notation) over the two types of lexical categories that *tis* can be assigned to: it can either be a D (definite article), a resumptive pronoun (NP) or alternatively a clitic.

(24)
```
tis   D *   (^ DEF)=+  (^ GEND)=f
            {(^ NUM)=sg (^ CASE)=gen
            |(^ NUM)=pl (^ CASE)=acc};
      NP *  {(^ PRED)='pro'} (^ PERS)=3 (^ NUM)=sg
            (^ GEND)=f (^ CASE)=gen (^ PRONTYPE)=rp;
      NP *  (^ PRED)='pro' (^ PERS)=3 (^ NUM)=sg
            (^ GEND)=f (^ CASE)=gen (^ PRONTYPE)=clitic.
```

The distribution of resumption in *pu*-RRCs is shown in Figure 3 and is accounted for in the f-structure by a series of f-structure equations in the CP rule in (25).

Figure 3. The distribution of the gap/resumption strategy in *pu*-RRCs

Grammatical function of resumptive/gap in the pu-RRC	Antecedent's Grammatical Function in main clause					
	Subj		Obj		Obj2	
	Loc	LD	Loc	LD	Loc	LD
Subj	gap	gap	gap	gap	gap	gap
Obj	gap/	gap/	gap/	gap/	gap/	gap/
Obj2	rp	rp	rp	rp	rp	rp
OoP	rp	rp	rp	rp	rp	rp

In the S rule (which is the same rule for declarative clauses, an optional resumptive pronoun is allowed to occur before the verb (only), be defining an optional NP in lines 13-14: `(NP: (^ SUBJ|OBJ|OBJ2)=! (! PRONTYPE)=c rp)`. The correct assignment of the gap or the resumptive strategy is accounted for by a disjunction on two equations: `(^ TOPIC)= (^ {COMP* SUBJ|COMP* OBJ})` (line 5) accounts for the distribution of the gap strategy, and `(^ {COMP OBJ| COMP* {OBJ2|ADJUNCT $ OBJ| GFPOSS}} PRONTYPE)=c rp` (line 6) does the same for the distribution of the resumption strategy. `(^ COMPFORM)= pu` (line 2) checks that the *pu*-RRC is introduced by a *pu* complementizer, and `(ADJUNCT $ ^)` (line 1) together with the N' rule (line 2) ensure that the *pu*-RRC is going to be an adjunct to its antecedent nominal element.

(25)

```
1    DP --> D; N'.
2    N' --> N; (CP: !$ (^ ADJUNCT)).
3    CP --> {e: {(ADJUNCT $ ^)
4                (^ COMPFORM)= pu
5                 {(^ TOPIC)= (^ {COMP* SUBJ|COMP* OBJ})
6             |(^ {COMP OBJ| COMP* {OBJ2|ADJUNCT $ OBJ| GFPOSS}}
               PRONTYPE)=c rp}
7                |(^ COMPFORM)=oti}
8            |DP};
9            C'.
10   C' --> C; S.
11   S --> [(DP: (^ SUBJ)=!)],
12         [(V; CP:(^ COMP)=!)],
13         [(NP: (^ SUBJ|OBJ|OBJ2)=! (! PRONTYPE)=c rp);
14          V],
15         [(DP: (^ OBJ)=!)],
16         [{(DP: (^ OBJ2)=!)
17          |(PP: (^ OBL)=! (! CASE)=acc (!PFORM)=cse)}].
```

The grammar also parses grammatical examples as in (26) where the pu-RRC is embedded within one or more *oti* complement clauses. This is achieved by placing a disjunction on the pu-RRC f-structure information (`(^ COMPFORM)=oti`, line 7) and simply allowing for an optional *oti* complement CP in the S rule (`[V; CP:(^ COMP)=!]`,, line 12). Of course we also need to add the appropriate lexical entry in the lexicon of a verb that subcategorises for *oti* complement clauses as in (27) below:

(26) i yineka pu o Petros ipe oti taise ton papagalo.
the.$_{\text{FSG.NOM}}$ woman.$_{\text{FSG.NOM}}$ that the.$_{\text{MSG.NOM}}$ Peter said.$_{\text{3SG}}$ that fed.$_{\text{3SG}}$ the.$_{\text{MSG.ACC}}$ parrot.$_{\text{MSG.ACC}}$
'The woman Peter said she fed the parrot.'

(27)
```
ipe V * (^ PRED) = 'said<(^ SUBJ)(^ COMP)>'
        {(^ SUBJ PRED) = 'pro'}
        (^ SUBJ CASE) = nom (^ PERS) = 3    (^ NUM) = sg.
```

Finally, the system accepts optional marking of punctuation at the end of a parsed sentence – period (.) and questionmark (?) – and assigns the appropriate clause type (declarative or interrogative respectively) in the f-structure, as in (28) and (29):

(28) o andras taise ton papagalo.
the.$_{\text{MSG.NOM}}$ man.$_{\text{MSG.NOM}}$ fed.$_{\text{3SG}}$ the.$_{\text{MSG.ACC}}$ parrot.$_{\text{MSG.ACC}}$
'The man fed the parrot.'

```
 [PRED  'feed<[1:man], [7:parrot]>'                       ]
 [SUBJ   [PRED 'man'                                  ]   ]
 [     1 [CASE nom, DEF +, GEND m, NUM sg, PERS 3     ]   ]
 [OBJ    [PRED 'parrot'                               ]   ]
 [     7 [CASE acc, DEF +, GEND m, NUM sg, PERS 3     ]   ]
5[CLAUSETYPE declarative, TENSE past                      ]
```

(29) o andras taise ton papagalo?
the.$_{\text{MSG.NOM}}$ man.$_{\text{MSG.NOM}}$ fed.$_{\text{3SG}}$ the.$_{\text{MSG.ACC}}$ parrot.$_{\text{MSG.ACC}}$
'Did the man feed the parrot?'

```
 [PRED  'feed<[1:man], [7:parrot]>'                       ]
 [SUBJ   [PRED 'man'                                  ]   ]
 [     1 [CASE nom, DEF +, GEND m, NUM sg, PERS 3     ]   ]
 [OBJ    [PRED 'parrot'                               ]   ]
 [     7 [CASE acc, DEF +, GEND m, NUM sg, PERS 3     ]   ]
5[CLAUSETYPE interrogative, TENSE past                    ]
```

2.3 Evaluation

The XLE system comes with a built-in set of test-suite tools that assist grammar developers in checking their grammar progress and detect any bugs and areas of improvement. For the purposes of evaluating our fragment, we built two testfiles testing the coverage of our grammar as described in section 2.2. Testsuite1 `demo-gre-v.0.9-basic_testfile.tfl` contains test items testing the basic declarative word order, subject-verb agreement, agreement within the DP, the pro-drop character of the language, some basic subcategorization frames for verbs, and optional punctuation. On the other hand, Testsuite2 `demo-gre-v.0.9-rrcs_testfile.ftl` contains test items relevant to the coverage of *pu-RRCs* with focus on the distribution of the gap/resumptive strategy in local and long-distance dependencies.

Out of a total 176 items, 108 grammatical test items had 1 parse, 67 ungrammatical test items had 0 parses and 1 item had 2 parses. Although the accuracy of the system might appear too artificial and constructed, it is worth noting that the current version of Modern Greek Grammar is a fragment. As such, it covers a restricted range of phenomena and it is only natural that the test items have been built to suit the phenomena under investigation. So why is it useful or interesting to build a fragment of a grammar in the first place if both the set of phenomena is limited and the testsuites are especially built to match them?

Mainly because it allows us to implement smaller pieces of grammar and test that they are robust and efficient and that they produce the expected output before attempting to incorporate them in a larger grammar. Another advantage is that simultaneous development of complex phenomena in the same grammar may influence both the accuracy of description of the phenomenon as well as the effectiveness of the system. Our choice of implementing a fragment of Modern Greek grammar was due not only to the above advantages but also to the fact that since this was our first attempt to build a computational grammar using the XLE platform, we were also interested in understanding the process of building a grammar and we intended to use this fragment as a starting point for future larger-scale implementations of Modern Greek.

3 Conclusions and Future Development Directions

This paper presented a computational grammar fragment for Modern Greek, built following the principles of the LFG ParGram Project and included among others a basic grammar covering simple word order phenomena, simple agreement phenomena as well as an implementation of an LFG account of the gap/resumption strategy in *pu* Restrictive RCs. It goes without saying that the current fragment of Modern Greek grammar is at its preliminary stage and it is only natural that there are a lot of phenomena not yet been accounted for. It is expected that future versions will build upon the current fragment of Modern Greek grammar to account for the syntax and semantics of *opios*-Restrictive, of Non-Restrictive and Free Relative Clauses, examples of which are shown in (30), (31) and (32) respectively:

(30) i yineka tin opia vrike o andras ine sto nosokomio
the$_{.FSG.NOM}$ woman$_{.FSG.NOM}$ the$_{.FSG.ACC}$ who$_{.FSG.ACC}$ found$_{.3SG}$ the$_{.MSG.NOM}$ man$_{.MSG.NOM}$ is$_{.3SG}$ to.the$_{.NSG.ACC}$ hospital$_{.NSG.ACC}$
'The woman whom the man found is at the hospital.' (*opios-Restrictive RC*)

(31) i Kiki, pu tin agapai o Stelios, ine arosti
the$_{.FSG.NOM}$ Kiki that her$_{.FSG.ACC}$ love$_{.3SG}$ the$_{.MSG.NOM}$ Stelios, is$_{.3SG}$ ill$_{.FSG.NOM}$
'Kiki, that Stelios loves, is ill.' (*pu-Non-Restrictive RC)*

(32) opjos irthe efige
whoever$_{.MSG.NOM}$ came$_{.3SG}$ left$_{.3SG}$
'Whoever came, left.' (*Free Relative RCi)*

Another area of improvement of the current version concerns the incorporation of the use of Discourse Functions, where appropriate, to account for the different degrees of markedness and acceptability of the different word orders. Word orders like SOV, VOS, OVS and OSV are usually taken as alternatives to the two basic word orders (SVO and VSO). Their first element is usually taken to be a topicalised/focused element (marked with small capital font in the examples below). An example of an SVO and its corresponding OSV order is given in (33) and (34):

(33) SVO

i yineka taise ton papagalo

$the_{.FSG.NOM}$ $woman_{.FSG.NOM}$ $fed_{.3SG}$ $the_{.MSG.ACC}$ $parrot_{.MSG.ACC}$

'The woman fed the parrot.'

(34) OSV

TON PAPAGALO i yineka taise

$the_{.MSG.ACC}$ $parrot_{.MSG.ACC}$ $the_{.FSG.NOM}$ $woman_{.FSG.NOM}$ $fed_{.3SG}$

'It was the parrot that the woman fed.'

We also intend to enrich our lexicon as appropriate to reflect the phenomena under investigation, as well as expanding out grammar to cover other constructions such as coordination, examples of which are shown in (35) and (36):

(35) i Kiki vrike ton papagalo ke o Stelios ti filise.

$the_{.FSG.NOM}$ Kiki $found_{.3SG}$ $the_{.MSG.NOM}$ $parrot_{.MSG.NOM}$ and $the_{.MSG.NOM}$ Stelios $her_{.FSG.ACC}$ $kissed_{.3SG}$

'Kiki found the parrot and Stelios kissed her.'

(36) i yineka ke o andras agapun ton papagalo

$the_{.FSG.NOM}$ $woman_{.FSG.NOM}$ and $the_{.MSG.NOM}$ $man_{.MSG.NOM}$ $love_{.3PL}$ $the_{.MSG.ACC}$ $parrot_{.MSG.ACC}$

'The woman and the man love the parrot .'

4 Acknowledgements

I would like to thank the audiences of the *2006 ParGram Meeting*, Oxford, the audiences of the *5th Language and Computation Day*, Essex, as well as the audiences of *LangUE 2007* postgraduate conference and *FLATLANDS* 2008 at Essex. I am also grateful to Louisa Sadler and two anonymous reviewers for comments and feedback on previous versions of this paper. I am also indebted to the ESRC for funding parts of my study with a 1+3 Quota Studentship No. PTA-2004-031-00112, support which is gratefully acknowledged here. Of course, all remaining errors are my own.

5 References

Alexiadou, A. & Anagnostopoulou, E. (2000). "Asymmetries in the Distribution of clitics. The case of Greek restrictive Relatives". In Beukema, f. & den Dikken, M. (eds). *Clitic Phenomena in European Languages*. Amsterdam and Philadelphia: John Benjamins, 47-70.

Alexopoulou, D. (1999). *The Syntax of Discourse functions in Greek: a non-configurational approach*. PhD Thesis, University of Edinburgh.

Alexopoulou, D. (2006). “Resumption in Relative Clauses”. *Natural Language and Linguistic Theory*, **24**, (1), 57-111.

Alexopoulou, D. (2008) “Truly intrusive: resumptive pronominals in questions and relative clauses”. In Grohmann, K. & Tsimpli, I.M. (eds). *Exploring the Left periphery*. Special issue, *Lingua*.

Alexopoulou, D & Keller, F. (2007). “Locality, Cyclicity and Resumption: at the interface between grammar and the human parser”. *Language*, **83.1**, 110-160.

Asudeh, A. (2004). *Resumption as resource management*. PhD Thesis, University of Stanford.

Bresnan, J. & Kaplan, R. (1982). ‘Lexical functional Grammar: a formal system for grammatical representation’. In Bresnan, J. (ed). (1982). *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press, 173-281.

Bresnan, J. (2001). *Lexical Functional Syntax*. Oxford: Blackwell.

Butt, M. & King, T.. (2007). ‘XLE and XFR: A Grammar Development Platform with a Parser/Generator and Rewrite System’. *International Conference on Natural Language Processing (ICON) Tutorial*. Available at http://www2.parc.com/isl/members/thking/icon-tutorial-2007.ppt [accessed 16.01.09]

Chatsiou, K. (in preparation). *The syntax and semantics of Modern Greek Relative Clauses with particular focus on resumption: an LFG approach (provisional title)*. PhD Thesis, University of Essex.

Dalrymple, M. (2001). *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Falk, Y. (2001). *Lexical Functional Grammar: an introduction to parallel constraint-based syntax*. Stanford, CA: CSLI Publications.

Falk, Y. (2002). “Resumptive Pronouns in LFG”. In Butt, M. & Holloway-King, T. *Proceedings of the LFG02 Conference*. CSLI online publications.

Holton, D., Mackridge, P. & Philippaki-Warburton, I. (1997). *Greek: a comprehensive grammar of the modern language*. London & NY: Routledge.

Holton, D., Mackridge, P. & Philippaki-Warburton, I.. (1997). *Greek: a comprehensive grammar of the Modern Greek Language*. London and New York: Routledge.

Ingria, R. (1981). *Sentential Complementation in Modern Greek*. PhD Dissertation, Massachussetts Institute of Technology.

Joseph, B. (1980). “Recovery of Information in Relative clauses: evidence from Greek and Hebrew”. *Journal of Linguistics*, **10**, 237-244.

Klein, E. (1993). *Toward Second Language Acquisition. A study of Null-Prep*. Dodrecht: Kluwer Academic Publishers.

Lascaratou, C. (1998). ‘Basic characteristics of Modern Greek’. In Siewierska, A. (ed). *Constituent order in the languages of Europe*. Berlin & NY: Mouton de Gruyter, 151-171.

Mackridge, P. (1985). *The Modern Greek language: a descriptive analysis of standard Modern Greek*. Oxford: Clarendon Press.

Markantonatou, S. (1992). *The syntax of Modern Greek NPs with a Deverbial Nominal Head*. PhD Thesis, University of Essex.

P.A.R.C. (2007). *The Packed Rewriting System (Transfer)* .Available at http://www2.parc.com/isl/groups/nltt/xle/doc/transfer-manual.html .[accessed 16.01.09]

P.A.R.C. (2008). *Parallel Grammar and Parallel Semantics Projects [online]*. Available at http://www2.parc.com/isl/groups/nltt/pargram/default.html [accessed 16.01.2008].

P.A.R.C. (2009a). *XLE [online]*. Available at http://www2.parc.com/isl/groups/nltt/xle/ [accessed 16.01.2009].

P.A.R.C. (2009b). *Shuffle: Grammatical Notations [online]*. Available at http://www2.parc.com/isl/groups/nltt/xle/doc/notations.html#N1.14 [accessed 16.01.2009].

Philippaki-Warburton, I. (1985). 'On the theory of Empty Categories: The missing subject and the clitic pronouns in Modern Greek'. In *Studies in Greek Linguistics,* Thessaloniki: Kiriakidis, 131-153.

Sells, P. (1985). *Lectures on contemporary syntactic Theories: an introduction to Government and Binding Theory, Generalised Phrase Structure Grammar and Lexical Functional Grammar.* Stanford: CSLI.

Siewierska, A., Rijkoff, K. & Dik, B. (1998). 'Word order variables in the languages of Europe'. In Siewierska, A. (ed). *Constituent order in the languages of Europe*. Berlin & NY: Mouton de Gruyter, 783-812.

Tsimpli, I.-M. (1996). Focus, Negation and Quantification in Greek. *DGFS, Workshop on Focus*, University of Goettingen.

Tsiplakou, S. (1998). *Focus in Greek: Its structure and interpretation*. PhD Thesis, University of London..

Tzanidaki, D. (1996). *The Syntax and Pragmatics of Subject and Object Position in Modern Greek*. PhD Thesis, UCL.

Tzartzanos, A. (1963). *Neoelliniki Syntaxis [Modern Greek Syntax] vol. B.* Athens: OEDB.

Varlokosta, S. (1997). *The structure and Acquisition of Relative Clauses in Modern Greek.* Ms. University of Maryland.

Varlokosta, S. (1998). "Η κατάκτηση των αναφορικών περιοριστικών προτάσεων (The acquisition of Restrictive Relative Clauses)". In *Studies in Greek Linguistics.* Thessalokini: Kiriakidis, **18**, 98-108.

Varlokosta, S. (1999). "The structure and the acquisition of prepositional relatives in Modern Greek: wh-movement or pronominal linkage?". In Joseph, B., Philippaki-Warburton, I. (eds). *Themes in Greek Linguistics II.* Amsterdam: John Benjamins.

Varlokosta, S. & Kotzoglou, G. (2003). "The distribution of the clitic pronoun in Restrictive Relative Clauses of Modern Greek: a new approach". In Theophanopoulou-Kontou, D. Lascaratou, C, Sifianou, M, Georgiafentis, M. & Spiropoulos, B. (eds). *Contemporary trends in Greek Linguistics*. Athens: Patakis, 169-191.