Creating multi-purpose linguistic resources for Modern Greek: a deep Modern Greek Grammar

Valia Kordoni*, Julia Neu*

*Department of Computational Linguistics, Saarland University
P.O. Box 15 11 50, D-66041 Saarbrücken, Germany
{kordoni, june}@coli.uni-sb.de

Abstract

Here we describe the development of a re-usable, multi-purpose linguistic resource for Modern Greek: a deep computational Modern Greek Grammar. The grammar is written in the HPSG formalism and is being developed in a multilingual context with MRS semantics, contributing to an open-source repository of software and linguistic resources with wide usage in education, research, and application building.

1. Introduction

In this paper we describe the development of a large grammar fragment of Modern Greek in a multilingual context. The grammar is couched in the theoretical framework of HPSG (Head-Driven Phrase Structure Grammar; (Pollard and Sag, 1994)) and benefits from an organization of semantics based on MRS (Minimal Recursion Semantics; (Copestake et al., 1999); (Copestake et al., 2001)).

MRS, a framework for computational semantics, in which the meaning of expressions is represented as a flat bag of Elementary Predications (EPs), combines naturally with typed feature structures, like the ones used in HPSG, and allows for structures which are underspecified for scopal information and can be compared across languages. HPSG itself is also suitable for multilingual grammar development, since the analyses in the extensive literature written in it can be shared across languages, but also parametrized accordingly, and its characteristic type hierarchy enables the writing of grammars that are easy to extend. Moreover, there are by now many useful open-source tools for writing, testing, and efficiently processing grammars written in HPSG and MRS: the LKB system for grammar development (Copestake, 2002), [incr tsdb()] for testing grammars and tracking changes (Oepen and Carroll, 2000), and PET, a very efficient HPSG parser for processing (Callmeier, 2000).

The tool we use for the development, i.e., the writing and the testing of the Modern Greek grammar is the Grammar Matrix (Bender et al., 2002), an open source tool designed for the rapid development of multilingual broad coverage grammars couched in HPSG and MRS and based on LKB.

As described thoroughly in (Bender et al., 2002), some of the very important and useful components of the Grammar Matrix include types defining the basic feature geometry necessary for starting off with the development of a grammar fragment in a given language couched in HPSG, types associated with MRS, a number of rules which reflect the general principles of HPSG (e.g., the Head Feature Principle, the Nonlocal Feature Principle, and so forth), constructional types, which correspond to the Immediate Dominance (ID) Schemata of HPSG (e.g., the head subject schema, the head complement schema, and so forth),

and to more specific constructions, such as relative clauses (see (Sag, 1997)), as well as configuration and parameter files for the LKB grammar engineering environment (see (Copestake, 2002)).

In the following we focus on some detailed examples from the deep Modern Greek grammar we have been developing since January 2003 using the Grammar Matrix, as part of the DELPHIN Collaboration (Deep Linguistic Processing with HPSG: An International Collaboration; for more see http://www.delph-in.net/), which currently involves research groups from DFKI in Saarbrücken, Saarland University, Stanford University, Tokyo University, the University of Sussex, Cambridge University, and the University of Trondheim, and whose main current research takes place in three areas: (i) robustness, disambiguation and specificity of HPSG processing, (ii) the application of HPSG processing to Information Extraction, and (iii) Multilingual Grammar Engineering, aiming mainly at the further promotion of the central role that robust and deep processing of natural language in a multilingual context based on HPSG plays nowadays in human language technology.

Our aim here is twofold: to present the deep computational grammar of Modern Greek, and to show the practical support we have drawn from the Grammar Matrix platform, which has enabled us to focus on the implementation of very demanding Modern Greek data right from the beginning of the project.

2. Modern Greek HPSG Syntax

The fundamental notion of an HPSG grammar is the sign, a complex feature structure which conveys phonological, syntactic, semantic, pragmatic and discourse information at the same time. The attribute-value matrix of a sign in the Modern Greek HPSG grammar is somewhat similar to a sign in the LinGO English Resource Grammar (ERG; (Flickinger, 2000)), as well as the HPSG Japanese Grammar (see (Siegel and Bender, 2002)), with information about the orthographical realization of the lexical sign in STEM, syntactic and semantic information in SYNSEM, nonlocal information in NONLOC, head information that percolates up the tree structure via HEAD and subcategorization information in VAL(ENCE) (whose values are SUBJ, COMPS, SPEC, and SPR). These features, which

are part of the sign geometry, as well as a large number of types which inherit from sign itself, are already defined in the Grammar Matrix.

For the implementation of the Modern Greek HPSG grammar we only needed to define further the parts of speech which are relevant to Modern Greek along with their specific head features. To these belong determiners, nouns, pronouns, affixes, prepositions and verbs, all of which inherit directly or indirectly from (the type) sign in the Grammar Matrix.

As far as the subcategorization patterns of verbal predicates in Modern Greek are concerned, it has turned out that for the best part these can be accounted for in the implementation of the Modern Greek grammar by relying on the material that the Grammar Matrix already provides for.

The grammar implementation is based on a system of types. Since the Modern Greek grammar is being developed for use in applications, it treats a wide range of constructions in Modern Greek, including valence alternations, cliticization phenomena, word order phenomena, subordinate clauses, unbounded dependency constructions (UDCs), raising and control, relative clause constructions, constructions headed by passives, and politeness constructions, among others. Due to space limitations, though, only some of these phenomena can be described here. Thus, for purposes of exemplification we focus here on some examples from cliticization phenomena in Modern Greek. For a more detailed description of all the phenomena covered in the Modern Greek grammar see (Neu and Kordoni, 2003).
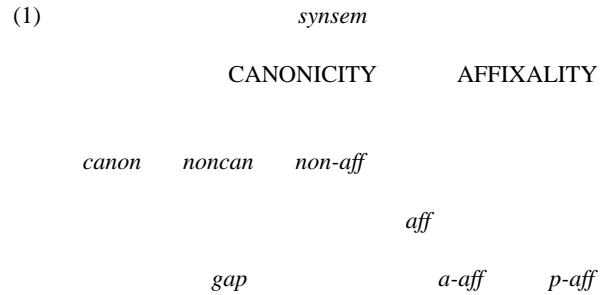
2.1. Cliticization Phenomena

Central to the efficient deep processing of Modern Greek is the implementation of clitics and clitic doubling phenomena (cf., (Kordoni, 2001)).

In general, Modern Greek distinguishes mainly between genitive and accusative clitics. Clitics in Modern Greek share a significant number of properties with what have been argued to be *pronominal affixes* in some Romance languages, such as French, and Italian (cf., (Miller and Sag, 1997), (Monachesi, 1996)). That is, they cannot be topicalized, they cannot be substituted by full pronouns, they cannot be coordinated, and they cannot be modified.

The implementation of Modern Greek pronominal affixes in the deep computational Modern Greek grammar described here draws on mainstream and well-established theoretical HPSG proposals, according to which in HPSG words may come along with an argument structure (ARG-ST/DEPS-ST), which is an attribute that determines the combinatorial potential of a word, including specific subcategorization restrictions (cf., (Manning and Sag, 1999)). The members of ARG-ST may be of sort *canonical* or of sort *noncanonical* (i.e., *gaps* or *affixes*; see also (1)). As shown in (1), *canon(ical-synsem)*: "is the type associated with all signs; *noncan(onical-synsem)* corresponds to an ARG-ST position that is not realized as a local syntactic dependent of the head. The latter subtype is in turn divided into the subtype *aff(ixal-synsem)* and *gap(-synsem)*. It is the presence of elements of type *aff* on a verb's ARG-ST list that triggers the morphological realization of the corresponding pronominal affixes. The type *non-aff* provides a

cross-cutting classification, subsuming all types of *synsem* other than *aff*" (Miller and Sag, 1997).

(1)   *synsem*

              CANONICITY   AFFIXALITY

       canon   *noncan*   *non-aff*

                                                aff

              gap   *a-aff*   *p-aff*

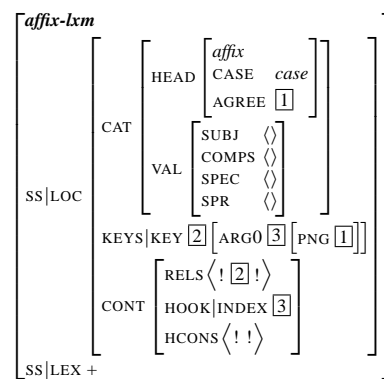In the Modern Greek grammar pronominal affixes are defined as in Figure 1.

Figure 1: Type for affixes in the Modern Greek grammar

For the implementation of pronominal affixes in clitic doubling constructions in Modern Greek:

(2)   O   Petros   to   fonazei   to   koritsi.
      the Peter.N cl.A call.3S   the girl.A
      "Peter is calling the girl".

we have introduced a new rule in the grammar, the *clitic doubling rule*.

The *clitic doubling rule*[1] inherits from the type head-final of the Grammar Matrix, and it enables a verb (the head) and an adjacent affix appearing on the left of the verb to combine (see Figure 2).

The verb comes along with a DEPS-ST list, whose lone element is token-identical to the synsem of the affix. After combining with the affix, the DEPS-ST list of the verb is empty. The COMPS list of the verb remains unchanged; it just gets fully copied to the projection V' (see Figure 2), since at this stage the verb still needs to combine with a direct object (i.e., the "clitic-doubled" NP *to koritsi* in example (2)), and potentially with more (optional) arguments. Finally, the values of the verb's SUBJ, SPR and SPEC are also copied to the projection V' (see Figure 2).

In order to restrict the agreement possibilities of the affix, we restrict the agreement features (person, number, and gender denoted as the value of an attribute AGREE) and

---

[1]Due to space limitations we do not show the *clitic doubling rule* here.

```
                        S
          NP                      VP
     o Petros          V'                  NP
                (head-clit-doubl-rule)

              AFF                 V      to koritsi
               to              fonazei
```
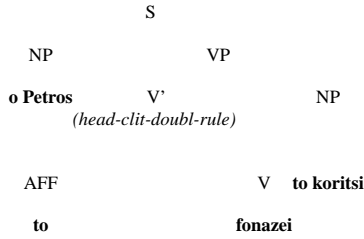
Figure 2: Tree representation for the sentence "*O Petros to fonazei to koritsi*"

CASE of the affix accordingly, so that the affix be coindexed with the first element of the COMPS list of the verb (i.e., coindexed with the direct object NP).

The information carried by the affix in the semantic representation of Modern Greek clitic doubling constructions in the Modern Greek grammar is already represented in the definition of transitive verbs. The value of the ARG0 of the affix is token-identical to the value of the corresponding AFFIX feature of the verb.[2]

Thus, in the semantic representation of clitic doubling constructions in Modern Greek like the one in example (2), the verb is represented bearing a "slot" for an affix whose value is shared with the ARG0 value of the affix itself; that is, in the semantic representation the affix appears as a direct dependent of the verb (for more see next Section).

3.   MRS Semantics for Modern Greek

The deep computational grammar of Modern Greek, some important aspects of which we describe here, benefits from an organization of semantics based on MRS (Minimal Recursion Semantics; (Copestake et al., 1999); (Copestake et al., 2001)). Types associated with MRS are already included in the Grammar Matrix tool we use for the development of the Modern Greek grammar.

As already mentioned in Section 1., MRS, a framework for computational semantics, in which the meaning of expressions is represented as a flat bag of Elementary Predications (EPs) encoded as values of a RELS attribute, combines naturally with typed feature structures, like the ones used in HPSG, and allows for structures which are underspecified for scopal information and easily comparable across languages, thus being very appropriate for grammars, like the one we describe here, which are developed in a multilingual context, as part of an international initiative for multilingual deep linguistic processing based on HPSG (cf., Delphin Collaboration in Section 1.). Due to space limitations, here we discuss only the semantics related to the syntactic phenomena we have presented in Section 2.1.

3.1.   The implementation of the semantics of Clitic Doubling Constructions in Modern Greek

The analysis of Modern Greek clitic doubling constructions (see example (2) and Figure 2 in Section 2.1. above)
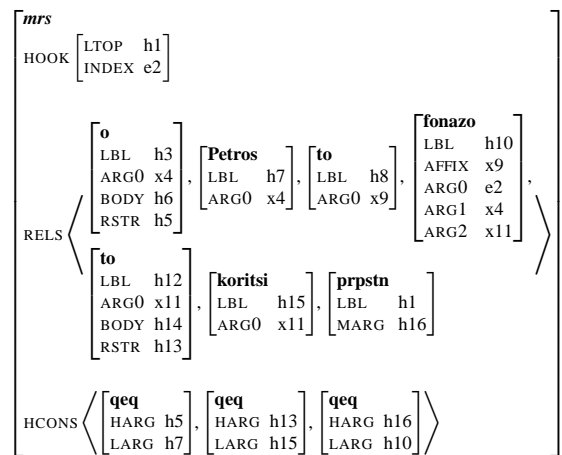
has led us to innovations in the correspondence between semantic and syntactic structures in comparison to the Grammar Matrix.

In the MRS semantic representation of the sentence in (2) of Section 2.1. that we are showing in (3), *h1* is the ltop (local top) handle of the sentence, which appears again as handle of **prpstn** (proposition; this conveys that the sentence is a proposition, rather than a question, for instance). Each noun gets an instantiation variable x_n (see *x4*, *x11*, and so forth in (3)), which is bound to the respective determiners of the nouns. Each entry in the MRS semantic representation in (3) gets a handle which may be part of a QEQ-constraint of the HCONS list.[3] The verbal head in (3) comes along with three arguments: ARG0, ARG1 and ARG2. Except ARG0, which denotes the event variable of the verb itself, these arguments are denoted by variables which are shared with those of the corresponding noun entries in the RELS. The HCONS list includes the QEQ-constraints which hold between the nouns and their determiners, as well as between the MARG of the **prpstn** (*h16*) and the handle of the verb (*h10*).

The verb in (3) also allows for an affix. Note that the affix entry bears a variable (*x9*), which is identical to the variable denoting the affix dependent of the verb. The handle of the affix, though, does not contribute to the QEQ-constraints. The affix has been combined with the verb by means of the clitic doubling rule (see Figure 2), and not by means of the head complement rule, a fact which is mirrored in the way this combination of the affix and the verb is denoted in the semantic representation of the sentence in (2) shown in (3): the affix is not a "real argument" of the verb; it is part of the morphosyntactic features of the verb.

(3)   Semantic representation in MRS of the sentence in (2):

$$
\begin{bmatrix}
\textbf{mrs} \\
\text{HOOK} \begin{bmatrix} \text{LTOP} & h1 \\ \text{INDEX} & e2 \end{bmatrix} \\[2em]
\text{RELS} \Big\langle
\begin{bmatrix} \textbf{o} \\ \text{LBL} & h3 \\ \text{ARG0} & x4 \\ \text{BODY} & h6 \\ \text{RSTR} & h5 \end{bmatrix},
\begin{bmatrix} \textbf{Petros} \\ \text{LBL} & h7 \\ \text{ARG0} & x4 \end{bmatrix},
\begin{bmatrix} \textbf{to} \\ \text{LBL} & h8 \\ \text{ARG0} & x9 \end{bmatrix},
\begin{bmatrix} \textbf{fonazo} \\ \text{LBL} & h10 \\ \text{AFFIX} & x9 \\ \text{ARG0} & e2 \\ \text{ARG1} & x4 \\ \text{ARG2} & x11 \end{bmatrix}, \\
\begin{bmatrix} \textbf{to} \\ \text{LBL} & h12 \\ \text{ARG0} & x11 \\ \text{BODY} & h14 \\ \text{RSTR} & h13 \end{bmatrix},
\begin{bmatrix} \textbf{koritsi} \\ \text{LBL} & h15 \\ \text{ARG0} & x11 \end{bmatrix},
\begin{bmatrix} \textbf{prpstn} \\ \text{LBL} & h1 \\ \text{MARG} & h16 \end{bmatrix} \Big\rangle \\[2em]
\text{HCONS} \Big\langle
\begin{bmatrix} \textbf{qeq} \\ \text{HARG} & h5 \\ \text{LARG} & h7 \end{bmatrix},
\begin{bmatrix} \textbf{qeq} \\ \text{HARG} & h13 \\ \text{LARG} & h15 \end{bmatrix},
\begin{bmatrix} \textbf{qeq} \\ \text{HARG} & h16 \\ \text{LARG} & h10 \end{bmatrix} \Big\rangle
\end{bmatrix}
$$

4.   Conclusion and Outlook

In this paper we have presented some important aspects of the deep computational grammar of Modern Greek we have been developing using the Grammar Matrix (Bender et al., 2002).

Our aim in this paper has been twofold: to present some interesting aspects of the HSPG grammar of Modern Greek, as well as to indicate which aspects of the Grammar Matrix

---

[2]The AFFIX feature of the verb has been introduced as part of a new relation *new-arg12-rel*, which inherits from the relation *arg12-rel* of the Grammar Matrix.

[3]For more on QEQ-constraints and HCONS lists see (Copestake et al., 1999) and (Copestake et al., 2001).

as broad coverage grammar development platform are helpful and which have needed further development in the case of Modern Greek.

In future work, the next immediate steps in the efficient deep analysis and processing of Modern Greek include the connection of the Modern Greek grammar system to a morphological analysis system, the incorporation of use of default entries for words unknown to the Modern Greek HPSG lexicon, the construction of a treebank, and the application to the grammar of stochastic disambiguation methods, like, for instance, the ones developed for the ERG by the Redwoods project at Stanford University (cf., (Oepen et al., 2002)), in an effort to treat ambiguity, one of the most important performance issues for broad coverage grammars.

As a final remark, it should be mentioned that the Modern Greek HPSG grammar is accompanied by a detailed documentation (Neu and Kordoni, 2003), which is revised by the grammar writers every time new phenomena are incorporated in the grammar and which aims at ensuring reusability of the grammar itself.

5. References

Bender, Emily, Dan Flickinger, and Stephan Oepen, 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In John Carroll, Nelleke Oostdijk, and Richard Sutcliffe (eds.), *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics. Taipei, Taiwan*.

Callmeier, Ulrich, 2000. Pet – a platform for experimentation with efficient HPSG processing techniques. *Journal of Natural Language Engineering 6(1): Special Issue on Efficient Processing with HPSG: Methods, Systems, Evaluation*:99–108.

Copestake, Ann, 2002. *Implementing Typed Feature Structure Grammars*. CSLI Lecture Notes, Number 110. Standord: CSLI Publications.

Copestake, Ann, Dan Flickinger, Ivan A. Sag, and Carl J. Pollard, 1999. Minimal Recursion Semantics: An Introduction. Ms., Stanford University.

Copestake, Ann, Alex Lascarides, and Dan Flickinger, 2001. An Algebra for Semantic Construction in Constraint-based Grammars. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL 2001), Toulouse, France*.

Flickinger, Dan, 2000. On building a more efficient grammar by exploiting types. *Journal of Natural Language Engineering 6(1): Special Issue on Efficient Processing with HPSG: Methods, Systems, Evaluation*:15–28.

Kordoni, Valia, 2001. *Psych Verb Constructions in Modern Greek: a semantic analysis in the Hierarchical Lexicon*. Ph.D. thesis, University of Essex, Colchester, UK.

Manning, Christopher D. and Ivan A. Sag, 1999. Dissociations between Argument Structure and Grammatical Relations. In Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol (eds.), *Lexical And Constructional Aspects of Linguistic Explanation*. Stanford, Calif.: CSLI Publications, pages 63–78.

Miller, Philippe H. and Ivan A. Sag, 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory*, 15:573–639.

Monachesi, Paola, 1996. *The syntax of Italian clitics*. Ph.D. thesis, University of Tilburg, Tilburg.

Neu, Julia and Valia Kordoni, 2003. Implementation of a Modern Greek Grammar Fragment in HPSG. Technical report, Department of Computational Linguistics, University of Saarland.

Oepen, Stephan and John Carroll, 2000. Performance Profiling for Parser Engineering. *Journal of Natural Language Engineering 6(1): Special Issue on Efficient Processing with HPSG: Methods, Systems, Evaluation*:81–97.

Oepen, Stephan, Kristina Toutanova, Stuart Shieber, Chris Manning, Dan Flickinger, and Thorsten Brants, 2002. The LinGO Redwoods Treebank. Motivation and Preliminary Applications. In *Proceedings of 19th International Conference on Computational Linguistics, Coling 2002. Taipei, Taiwan*.

Pollard, Carl and Ivan A. Sag, 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Sag, Ivan A., 1997. English Relative Clause Constructions. *Journal of Linguistics*, 33.2:431–484.

Siegel, Melanie and Emily M. Bender, 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization, Coling 2002 Post-Conference Workshop*. Taipei, Taiwan.