A shorter form of this paper was presented at the *5th International Conference on Greek Linguistics*, Sorbonne, France, September 2001.

A LARGE-SCALE SYSTEMIC FUNCTIONAL GRAMMAR OF GREEK

Aggeliki Dimitromanolaki, Ion Androutsopoulos and Vangelis Karkaletsis

Software and Knowledge Engineering Laboratory
Institute of Informatics and Telecommunications
National Centre for Scientific Research "Demokritos"
GR-153 10 Aghia Paraskevi, Athens, Greece
e-mail: {adimit, ionandr, vangelis}@iit.demokritos.gr

Abstract

This paper presents a large-scale computational grammar of Greek, couched in the framework of Systemic Functional Linguistics. The grammar is being developed in the context of M-PIRO, a multilingual natural language generation project, where personalized descriptions of museum exhibits are generated automatically from a single database source. Although the grammar is still under development, it already provides a wide coverage of the Greek syntax and morphology. Our long-term goal is to produce a wide-coverage computational grammar of Greek suited to generation applications.

1 Introduction

In the last few decades, the evolution in computational linguistics and the development of real-world Natural Language Processing (NLP) applications has increased the need for large-scale computational grammars. *Natural Language Generation* (NLG) [Reiter and Dale 2000], the process of producing texts from some underlying nonlinguistic representation of information, is such an application. This paper presents a large-scale computational grammar of Greek, couched in the framework of Systemic Functional Linguistics [Halliday 1985]. The grammar is being developed in the context of M-PIRO, a multilingual natural language generation project, where descriptions of museum exhibits are generated automatically in three languages (English, Greek and Italian) from a single database source[1]; the descriptions are personalised, i.e. reflect the interest of the user as well as certain educational goals, and are both textual and spoken [Androutsopoulos et al. 2001]. M-PIRO builds upon the ILEX natural language generation system[2] [Oberlander et al. 1998] and extends ILEX's technology by incorporating improved multilingual capabilities, high-quality speech output, authoring facilities, extended user modelling mechanisms, as well as a more modular core generation engine.

NLG systems are usually decomposed into more specific modules [Reiter 1994; Reiter and Dale 1997]. The M-PIRO generation system consists of the following modules (Figure 1):

[1] M-PIRO is a project of the Information Societies Programme of the European Union, running from February 2000 to January 2003. The project's consortium consists of the University of Edinburgh (UK, co-ordinator), ITC-irst (Italy), NCSR "Demokritos" (Greece), the University of Athens (Greece), the Foundation of the Hellenic World (Greece), and System Simulated Ltd (UK). More information about the project is available from: http://www.ltg.ed.ac.uk/mpiro/.

[2] See http://www.cstr.ed.ac.uk/cgi-bin/ilex.cgi.

- Content selection. The system selects, from a database, which facts to convey. Facts represent either relationships between the entities of the domain (e.g., X made Y) or attributes of entities (e.g., name, dimensions). An example fact is shown in the left column of Table 1.
- Document planning. The structure of the text, i.e. the ordering of the facts and the rhetorical relations that hold between them [Mann and Thompson 1988], is decided.
- Micro-planning. It specifies in abstract terms how each fact should be expressed in each language; for example, which verb to use, in what voice and tense. An example micro-planning specification is shown in the right column of Table 1[3]. Micro-planning also includes processing steps that determine which facts can be aggregated in a single sentence (e.g., "This vase dates from approximately 550 BC and was found in Attica" rather than "This vase dates from approximately 550 BC. It was found in Attica."), and what type of referring expression should be generated for each entity (e.g., "Doryphorus", "this statue", or "it" may be more or less appropriate in the context of previous sentences).
- Surface realization. This module uses a grammar of the target language to generate a text that is syntactically, morphologically and orthographically correct.

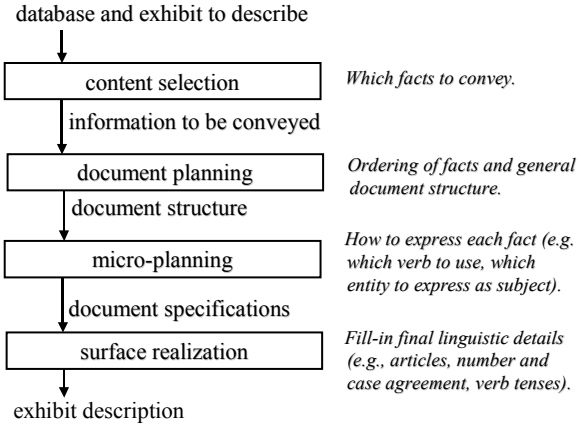


Figure 1: Stages of natural language generation in M-PIRO

Fact: made-of(exhibit6,silver)	**Micro-planning specification**
Predicate: made-of Arg1: exhibit6 Arg2: silver	(defexpression made-of :language :greek :arg1 exhibit6 :arg2 silver :verb make-verb :voice passive :aspect perfect :arg2-prep of)

Table 1: Example fact and micro-planning specification

For the last stage of the generation process a large-scale computational grammar of each target language is required. The most popular approach to surface realization is that of using systemic grammars [Halliday 1985; Teich 1999]. These grammars are primarily concerned with the functions of language and with how these functions are mapped into surface forms by making a series of increasingly fine-grained choices that determine the syntactic characteristics of the sentence being constructed. The linguistic options, which are available at certain specifiable

[3] M-PIRO's facts and micro-planning specifications follow Ilex's notation [Mellish et al. 1998].

contexts, are called *features*[4]. Features are represented by *system networks*, with each *system* being a point of choice (Figure 2).

Systemic grammars have been successfully employed in the surface realization task of NLG as, for example, in the projects PENMAN [Mann and Matthiessen 1983], TECHDOC [Rosner and Stede 1994], KOMET [Teich 1992] and ILEX [Oberlander et al. 1998]. The prominent role of systemic grammars in surface realization is due to the fact that such grammars can treat the syntactic, semantic and even pragmatic aspects of language in one integrated model [Teich 1999]. Systemic grammars have also the advantage of allowing resource sharing between different languages in multilingual generation [Bateman et al. 1991]. More specifically, resources for several languages can be combined so that commonalities are shared rather than re-represented. This advantage is very important to NLG, since the resources become more economical and the development time is significantly reduced. This functionality has been successfully deployed in KPML [Bateman 1997], a multilingual generator that provides large-scale realization resources and a basic engine for using such resources for surface realization.

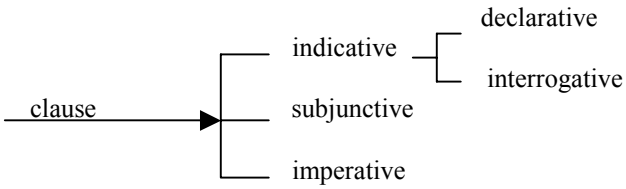


Figure 2: An example system network for the mood of a clause

The approach taken in M-PIRO is similar to that of KPML [Bateman 1997]. The Greek grammar was not built from scratch; it was constructed by taking the English grammar used in Ilex [O'Donnell 1994; O'Donnell 1996] as a starting point and making the appropriate modifications and additions in order to develop the Greek surface generation component. In this way, we preserve grammar commonalities across different languages, therefore allowing for faster resource development and easier maintainance of the system.

The remainder of this paper is organized as follows. Section 2 describes the general organization of the grammar. Section 3 presents the systemic-based description of clauses, focusing on the problems of constituent order and pronominal subjects. Section 4 presents the approach that we have adopted for dealing with the structure and the properties of Greek noun phrases and Section 5 is devoted to the words and their morphology. Section 6 concludes and discusses future work plans.

2 Greek Systemic Grammar: General Organization

The input to the surface realization component is a list of micro-planning specifications as the one shown in Table 1. The goal is to transform such a list into a cohesive and coherent Greek text. One of the central properties of systemic grammars is the *rank scale*, which defines the types of linguistic units used in the grammar and serves two purposes. First, it provides an effective description of the paradigmatic relations pertaining to the grammatical units of a language, going on the assumption that every grammatical unit has a set of grammatical features that is disjoint from the set of features of the other units. Second, it constitutes a hypothesis about syntagmatic organization in terms of constituency relations[5]. The rank scale is often represented as the initial system of the grammar (Figure 3).

Following the English grammar, the Greek grammar of M-PIRO is divided into four ranks (Figure 3). The highest rank of the grammar is the *clause-complex* rank, which describes the

[4] The term *feature* is not used as in Unification formalisms, where feature refers to an attribute and its value. The Systemic usage is closer to the value.

[5] In this sense, the rank scale is comparable to the bar levels in X-bar syntax; it incorporates, however, very different claims about the actual shape of the constituent structure.

structure of multi-clausal units, the next lower one is the *clause* rank, which describes the clause structure, the next is the *group* rank, which describes the structure of noun phrases (NPs) and prepositional phrases (PPs), and the lowest one is the *word* rank, which describes the grammatical, syntactic and semantic features of words. The surface realization takes place in a top-down manner, moving from the highest rank of the grammar to the lowest one. At the current state of the Greek grammar, the last three of these ranks have been developed and will be presented in Sections 3-5. Regarding the clause complex rank, the corresponding part of the English grammar is currently used. We are investigating if this is sufficient for Greek. It is worth mentioning that the Greek grammar uses 182 systems of which 62 are new, i.e. they were not inherited from its English ancestor. These new systems handle Greek-specific phenomena, which are not present in English. At the clause rank, such phenomena include the constituent order, the clause aspect, the pronominal subjects, the case assignment, the pronoun-antecedent agreement etc. At the group rank new features were introduced to deal with phenomena such as agreement and proximal deixis, while the new systems of the word rank mainly involve information about the morphological realization of words. Table 2 presents the number of systems used at each rank of both grammars.

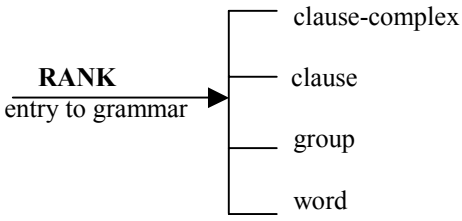


Figure 3: The top system of the grammar, RANK

RANK	ENGLISH	GREEK
Clause complex	7	7
Clause	37	54
Group	41	65
Word	35	56

Table 2: Number of systems at each rank in English and Greek grammar.

Furthermore, apart from the grammar, there are two complementary resources: the *lexicon* and the *morphological component*. The lexicon provides the lexical items necessary for the realization of words, and their features. The features must be consistent with the features defined in the word rank of the grammar. The morphological component is responsible for the morphological generation; it accepts as input the lexical item retrieved from the lexicon and the grammatical (e.g. noun declension) and syntactic features (e.g. accusative singular) of the word to be generated and returns a fully inflected word.

3 Greek Clauses

The clause rank first defines the structure of the clause, i.e. the constituents that make up a clause and their possible orders. According to this definition, the clause consists of the following elements, of which only *Subj* and *Pred* are obligatory and the rest of them are optional. These elements have corresponding slots that are stored in the grammar and carry the information required for the surface realization:

- Adjunct1: A modifier (adverb or prepositional phrase) that precedes the subject. E.g., "*Σήμερα* ο Δορυφόρος βρίσκεται στο ... " ("*today* Doryphorus is located in the …").
- Subj: The subject of the clause, which is always a noun phrase.
- Neg: The negative particle.
- Perf: An auxiliary verb, e.g. "έχω" ("have") in the perfect tenses.

- Pred: The main verb of the clause.
- Advmod: An adverbial modifier typically placed before Pred or between Pred and Obj. E.g. "Οι κούροι απεικόνιζαν *συνήθως* νέους άνδρες" ("the statues of kouroi *usually* represented young men").
- Obj: A general slot representing not only the object of the clause, but also the subject complement in copula clauses or a prepositional phrase attached to the verb.
- Adjunct2: A modifier (adverb or prepositional phrase) that follows the object (or the predicate in case there is no object). E.g. "Προς το τέλος της αρχαϊκής περιόδου, τα νομίσματα κάνουν την εμφάνισή τους *στις εμπορικές συναλλαγές*" ("towards the end of the archaic period, coins were used *for transactions*").

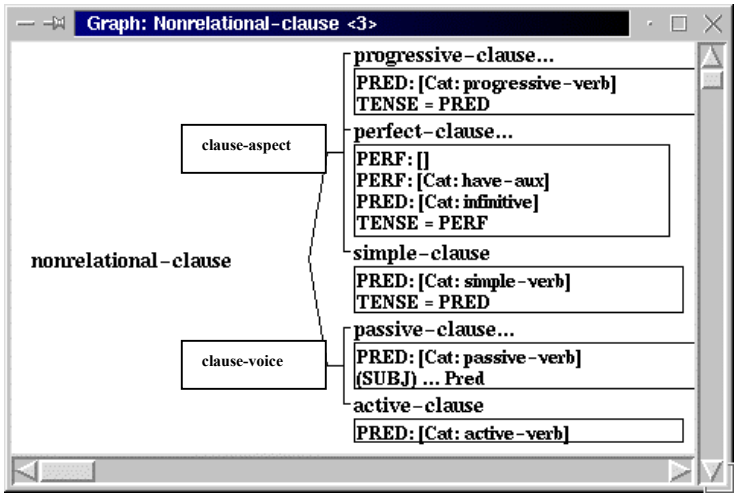


Figure 4: The system network for clause aspect and clause voice

Once the constituents of the clause have been defined, various properties of the clause form (clause type, transitivity, mood, tense, aspect etc.) are controlled by specifying features of the clause unit, which are organized into systems. Figure 4 shows a system network consisting of two systems, *clause aspect* and *clause voice*[6]. *Nonrelational clause* is the *entry condition* of these systems, i.e. the context in which the choice of aspect or voice must be made. Regarding aspect, *progressive clause*, *perfect clause* and *simple clause* are the three features (options) among which a choice must be made. Respectively, the features for voice are *active-clause* and *passive-clause*. The box below each feature contains the *realization statements* of the feature, which encode the structural consequences of each choice, functioning as a link between the paradigmatic (options) and the syntagmatic (forms) axis of the language. Features may also have *selection constraints*, i.e. preconditions for the selection of a feature, which are not shown here for reasons of simplicity.

To make all these clearer, let us consider how the micro-planning specification shown in Table 1 will be processed by the clause rank of the grammar. Some additional links help the system do the mapping between the expressions found in the micro-planning specification and the grammar. Arg1 and arg2 represent the *Subj* and *Obj* slots of the clause structure respectively. According to the specification, the main verb of the clause (*Pred*) must be a verb with the id *make-verb*. Following Ilex's grammar, we distinguish between *relational clauses* (whose main verb is a copula) and *nonrelational clauses* (any other verb). This means that the clause of our example will have the feature *nonrelational-clause*. The next slot of the micro-planning specification (*:voice passive*) leads to the selection of the feature *passive-clause* at the system *clause voice* (Figure 4). The realization statement of this feature (*Cat: passive-verb*) adds the feature *passive-verb* to the *Pred* slot of the clause structure. Similarly, the feature *perfect-clause* is selected at the

[6] The tool that allows the graphical representation of systems and system networks was developed by Mick O'Donnell and is part of the WAG system [O'Donnell 1994].

clause aspect system (Figure 4), which adds the feature *have-aux* to the *Perf* slot and the feature *infinitive* to the *Pred* slot. The last realization statement of *perfect-clause* (*TENSE = PERF*) determines that Perf must also carry the information about the basic tense of the clause (present, past or future). Since the micro-planning specification does not provide any information about the tense of the clause, the default tense, i.e. present, is assumed (default values are determined in the additional links connecting the grammar with the micro-planning specification). Consequently, the feature *present-clause*, which belongs to a system called *clause-tense* not shown here, is selected. So, *Perf* will also have the feature *present-verb*. Number and person features are selected by two systems, called *clause-number* and *clause-person* respectively, depending on the number and person of the subject; in this example the features *singular-verb* and *thirdperson* are added to the *Perf* slot. The last slot of the micro-planning specification informs us that we have to use a preposition with the id *of* in the *Obj* slot of the clause. So, the clause structure of this micro-planning is given in Figure 5. The features *nominative-np* and *accusative-np* of the Subj and Obj slot respectively are forced by the realization statements of the *clause* feature not shown here.

Subj: [exhibit6] *nominative-np*
Perf: *have-aux present-verb singular-verb thirdperson*
Pred: *make-verb passive-verb infinitive*
Obj: *of* + [silver] *accusative-np*

Figure 5: An example output of the clause rank

CURRENT COVERAGE	FUTURE GOALS
Passive clauses	Clitics
Simple, perfect and progressive aspect	Subjunctive and imperative clauses
Clause tenses	Non-finite clauses
S-V-O and O-V-S word order	Other possible word orders
Relative clauses (agreement, case assignment etc.)	Other subordinate clauses
Pro-drop	
Copula clauses	

Table 3: Phenomena currently covered at clause rank and future goals

The clause rank of the Greek grammar currently covers declarative main clauses and relative clauses. Table 3 presents the basic linguistic phenomena that have been accounted for up to now and also the issues that we plan to address in the future. The remainder of this section is devoted to two issues that have received particular attention in Greek linguistics: *constituent order* and *pro-drop clauses*.

3.1 Constituent Order

The problem of Greek word order has been extensively studied [Horrocks 1982; Philippaki-Warburton 1982; Philippaki-Warburton 1985; Tsimpli 1990]. It is generally agreed that the order of the basic constituents (subject, verb, object) of a main clause is very flexible and that all possible combinations yield grammatical clauses. The difference between the various constituent orders of the same clause is not syntactic; it is rather pragmatic, since the ordering of the constituents depends on discourse reasons, i.e. which constituent is the topic of the discourse, which constituent we want to emphasize etc. In generation, a clause is required not only to be grammatical, but also to be free of false implicatures so as to contribute to the overall discourse coherence. The systemic-based Greek grammar handles this problem as follows.

Our methodology was to account first for the orders that are frequent in descriptive texts such as the texts produced by M-PIRO. A corpus analysis indicated that the main clauses found in descriptive texts have always one of the two NPs, either the subject or the object at the first position; hence, the order is either *S(subject)-V(verb)-O(object)* or *O-V-S*. This is easily explained

by the fact that each sentence in these texts contains a *topic*, the entity to be described, and some *new information* about that entity; and it is generally taken for granted in linguistics that the topic is always placed at the initial position of the sentence. Furthermore, it has been argued that the subject is the most natural topic of the clause [Holton et al. 1997] and this is also evidenced by the dominance of the S-V-O constituent order in our corpus. Hence, in our grammar the S-V-O order is assumed to be the default (basic) constituent order. See, for example, the following sentence:

1. "Αυτό το πορτρέτο απεικονίζει το Μέγα Αλέξανδρο."
   this the portrait depicts the Great Alexander
   "This portrait depicts Alexander the Great."

However, in some cases the topic is required to be the syntactic object of the clause. Consider, for instance, the following sentences, which constitute different realizations of the fact *painter-of(exhibit1, painter-of-kleofrades)*. One of these sentences will be generated next to the text fragment given in Figure 6.

2. "Ο ζωγράφος του Κλεοφράδους ζωγράφισε αυτόν τον αμφορέα."
   the painter the-gen Kleofrades painted this the amphora
   "The painter of Kleofrades painted this amphora."
3. "Αυτός ο αμφορέας ζωγραφίστηκε από το ζωγράφο του Κλεοφράδους."
   this the amphora was-painted by the painter the-gen Kleofrades
   "This amphora was painted by the painter of Kleofrades."
4. "Αυτόν τον αμφορέα (τον) ζωγράφισε ο ζωγράφος του Κλεοφράδους."
   this the amphora (it) painted the painter the-gen Kleofrades
   "This amphora was painted by the painter of Kleofrades."

> Αυτό το έκθεμα είναι ένας αμφορέας. Σήμερα βρίσκεται στο Μουσείο Martin von Wagner του Πανεπιστημίου του Wurzburg, που είναι στη Γερμανία. ?
> "This exhibit is an amphora. Today it is located in the Martin von Wagner Museum, University of Wurzburg, which is in Germany. ?"

Figure 6: Text fragment generated by M-PIRO

Sentence 2, S-V-O, would be natural if the subject "the painter of Kleofrades" had already been introduced in the discourse and was functioning as the topic of the utterance. In this case, however, "the painter of Kleofrades" has not been mentioned before; the text is about the amphora, which is also the topic of the current utterance, and this is the first mention of the painter. On the other hand, although 3 is both grammatical and consistent with the discourse parameters, the passive syntax does not look very natural in Greek and most speakers seem to prefer 4, i.e. the O-V-S order. In order to account for this case, we defined a system (Figure 7) that differentiates between two types of transitive clauses, one that has the subject and one that has the object as its topic (*object topicalisation*). Following Ilex [O'Donnell et al. 1998], the M-PIRO system keeps track of pragmatic information concerning the referring environment (e.g. previously mentioned entities, current topic, previous topic, current focus, previous focus). Such information is then passed to the grammar (Figure 7).

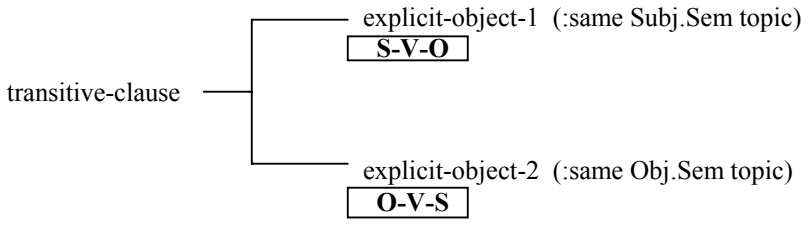


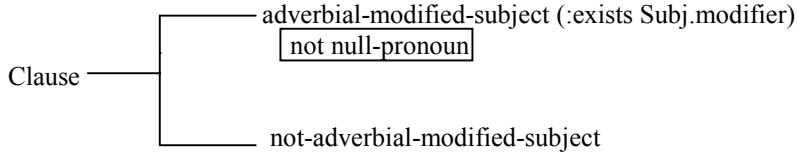
Figure 7: The system handling the constituent order of transitive clauses

In the cases of object topicalization, the presence of the clitic pronoun (e.g. "τον" in sentence 4) is obligatory [Warburton-Philippaki 1977]. For the time being, we do not provide a proper

account of clitics, but we are planning to handle this phenomenon in future versions of the grammar. Future versions will also allow for other word order variations in main clauses as well as for appropriate ordering in subordinate clauses.

3.2 Pro-drop

As opposed to English, where the pronoun is always overt, Greek allows both overt and null pronouns in the subject position [Dimitriadis 1995; Dimitriadis 1996; Joseph 1994; Philippaki-Warburton 1987]. In order to have as fluent texts as possible, we had to study the conditions of use of each of these pronominal subjects and provide an account for them in the Greek grammar.



Αυτός ο στατήρας σήμερα βρίσκεται στο Νομισματικό Μουσείο της Αθήνας. *Δημιουργήθηκε κατά τη διάρκεια της κλασικής περιόδου και έχει φτιαχτεί από ασήμι*.

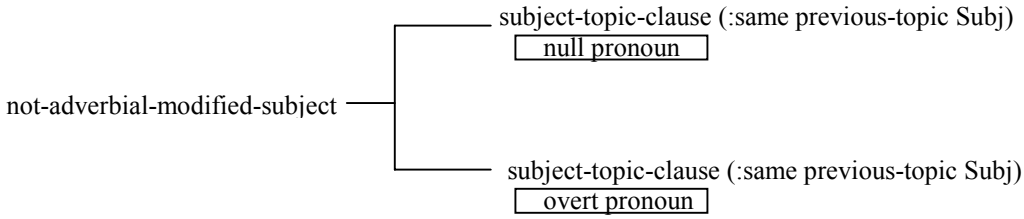
"Today this stater is located in the Numismatic Museum of Athens. It was created during the archaic period and is made of silver."

Αυτό το τετράδραχμο προέρχεται από την Αθήνα. *Και αυτό δημιουργήθηκε κατά τη διάρκεια της αρχαϊκής περιόδου*.

"This tetradrachm originates from Athens. It was also created during the archaic period."

Figure 8: Pronominal subjects depending on subject modifiers

Our first observation, resulted from the corpus analysis, was that null subject pronouns are excluded when the subject of the clause is modified by an adverb, e.g. "μόνο αυτό", "αυτό μονάχα" ("only this") or the emphatic "και" ("and"). In our grammar such clauses cannot have a null pronoun as subject. Figure 8 presents the system handling this phenomenon and two text fragments containing null and overt subject pronouns respectively.



Αυτό το αυτοκρατορικό πορτρέτο χρονολογείται στον πρώτο αιώνα μ.Χ. και σήμερα βρίσκεται στο Μουσείο Αγοράς της Αθήνας. *Δημιουργήθηκε κατά τη διάρκεια της ρωμαϊκής περιόδου*.

"This imperial portrait dates from the first century AD and today is located in the Archaeological Museum of Corinth. It was created during the roman period."

Ο ζωγράφος του Κλεοφράδους ζωγράφισε επίσης ένα αγγείο με σκηνές από το «Ιλίου Πέρσις». *Αυτό είναι ένα από τα πιο ενδιαφέροντα έργα του*.

"The painter of Kleofrades also painted a vase with scenes from Iliou Persis. This is one of his most interesting works."

Figure 9: Pronominal subjects depending on discourse factors

The second observation concerns the remaining clauses, i.e. clauses that do not contain modified subjects. When the entity expressed as the syntactic subject of the current clause is the topic of the previous utterance, a null pronoun is preferred. On the other hand, if this entity is not the topic of the previous utterance, an overt pronoun is needed in order to assure correct interpretation of the sentence and coherence of the text. This approach resembles the centering-

based approach of [Dimitriadis 1996][7]. In Figure 9 we summarize the system handling this phenomenon and provide two text fragments as examples of the correct realization of pronouns.

4 Greek Noun Phrases

The realization of noun phrases (NPs) is guided by the group rank of the grammar. The group rank also guides the realization of prepositional phrases (PPs), but space limitations do not allow discussing them here. Considering NPs, a basic distinction that is made is between *nominal* (containing a noun) and *pronominal* (containing a pronoun instead of the noun) NPs. Nominal NPs are further subdivided into *common* and *proper* NPs. As in the case of clauses, we first defined the structure of each of these NP types. The elements of a common NP are presented below (most of them are optional with only *Noun* being obligatory). These elements have corresponding slots that are stored in the grammar and carry the information required for the surface realization of NPs:

- Quantifier: Words expressing quantity, e.g. "όλοι" ("all"), "πολλοί" ("many"), "λίγοι" ("few").
- Prox-deictic: The demonstratives, "αυτός, αυτή, αυτό" ("this") etc., when they function as determiners and not as pronouns.
- Deictic: The definite and indefinite article and any other determiners that do not fall to another category (quantifier, proximal-deictic, numeral).
- Numeral: The cardinal numbers and numerals.
- Adjectives: Any number of adjectives is allowed.
- Noun: The noun of the phrase.
- Classifier: An NP functioning as a modifier of the current NP, e.g. possessive pronoun, genitive-np.

As pointed out in [Holton et al. 1997], when all these elements occur in the same NP, the most natural word order is as follows:

Quantifier - Prox-deictic – Deictic – Numeral – Adjectives – Noun – Classifier

Consider, for example, the following sentence:

"Όλοι εκείνοι οι πέντε μεγάλοι ζωγράφοι μας"
all those the five big painters our
"All those five great painters of ours"

In contrast, when a smaller number of elements is used, the linear order may vary depending on semantic or pragmatic factors, e.g. "αυτή η υδρία", "η υδρία αυτή" ("this hydria"). We have only accounted for the order presented above, which, we believe, is the basic (the most neutral in terms of context) word order of Greek NPs, but our future goals include a treatment of all possible order variations.

Subj: [exhibit6 ***tetradrachm-noun***]
nonpossessive-deixis proximal-deixis close-proximal-deixis singular-np neuter-np
nominative-np

Figure 10: An example input to the group rank

As in clauses, various aspects of the Greek NP form (definiteness, proximity, possessiveness, number, gender and case agreement between the elements etc.) are controlled by specifying features of the NP unit. Some NP features are shown in Figures 11-13. Let us now go back to the example considered in Section 3. The input to the group rank is an NP or a PP, such as the *Subj* or *Obj*, respectively, shown in Figure 5, and also some information provided by the *NP planner*, a different module of the system that decides how to refer to entities, i.e. which type of referring expression must be used in a given context [O'Donnell et al. 1998]. The information provided by

[7] As Dimitriadis [Dimitriadis 1996] points out "the Cb represents the immediate center of attention, and therefore constitutes centering theory's version of the notion of *topic*".

the NP planner is represented by features of the group rank. Figure 10 summarizes the information that constitutes the input to the group rank in the case of our example (the features in bold italics are provided by the NP planner).

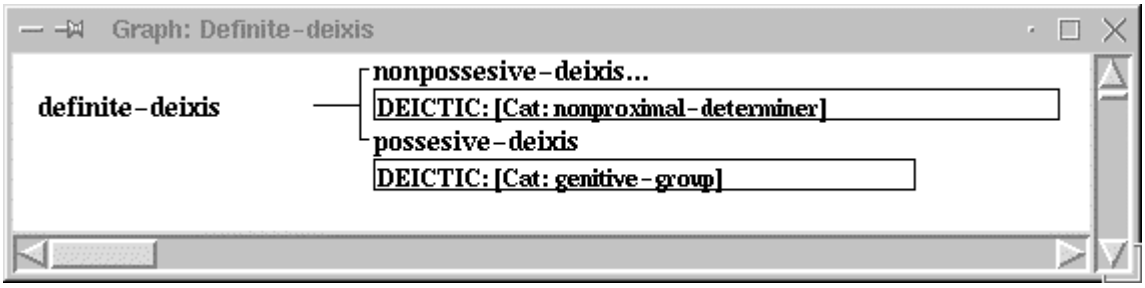


Figure 11: The system *deixis-possessiveness*

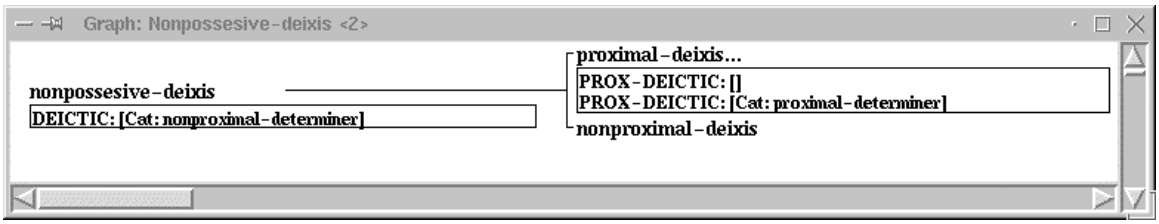


Figure 12: The system *deixis-proximity*

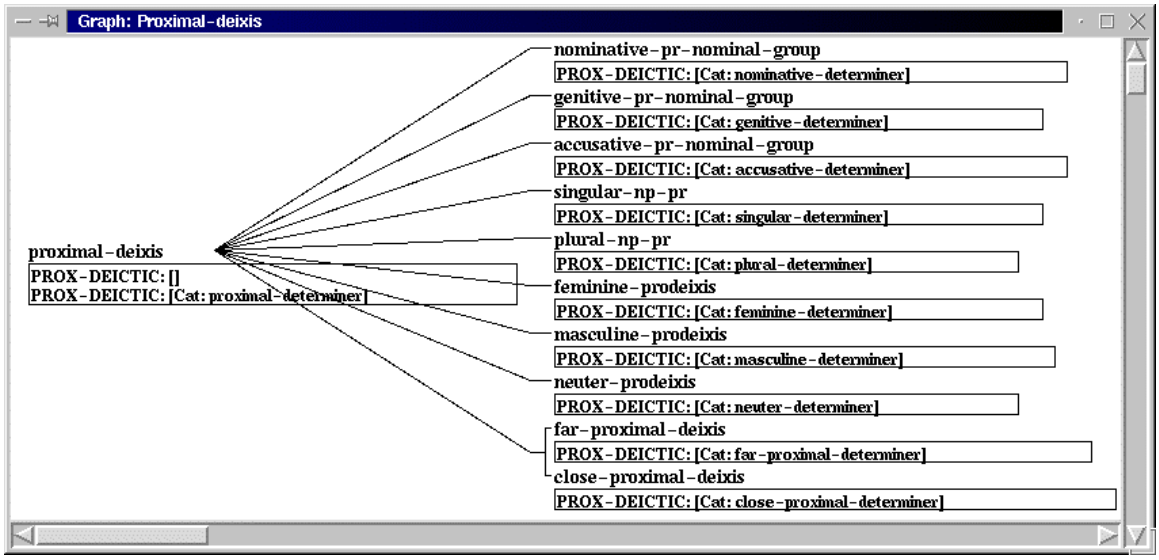


Figure 13: The system network of *proximal-deixis*

As shown in Figure 11, the selection of *nonpossessive-deixis* means that the feature *nonproximal-determiner* must be added to the *Deictic* slot of the NP structure. Similarly, the features *proximal-deixis* and *close-proximal-deixis* require the *Prox-deictic* slot to have the features *proximal-determiner* (Figure 12) and *close-proximal-determiner* (Figure 13), respectively. Subsequently, a decision on the number and gender of this NP must be made. These features also derive from the NP planner. Since the NP is marked as *singular-np* and *neuter-np*, the corresponding features, *singular-np-pr* and *neuter-prodeixis*, are selected in the system network shown in Figure 13 (the selection of these features is guided by their selection constraints not shown here). Similarly, the feature *nominative-np*, provided by the clause rank, leads to the selection of *nominative-pr-nominal-group* (Figure 13). Therefore, *Prox-deictic* is enriched with the features *singular-determiner*, *neuter-determiner*, *nominative-determiner*. In a similar way, the corresponding number, gender and case features are selected for the *Deictic* and the *Noun* slot of the NP structure. The final output of the group rank is given in Figure 14.

Prox-deictic: *proximal-determiner close-proximal-determiner singular-determiner neuter-determiner nominative-determiner*
Deictic: *nonproximal-determiner singular-determiner neuter-determiner nominative-determiner*
Noun: *tetradrachm-noun singular-noun neuter-noun nominative-noun*

Figure 14: An example output of the group rank

CURRENT COVERAGE	FUTURE GOALS
Ordering	Possible ordering variations
Determiners in proper NPs	NPs containing adjectives
Determiner-noun agreement	Elliptical NPs
Proximal deixis	
Pronoun-antecedent agreement	
Classifiers	
Generic NPs	

Table 4: Phenomena currently covered at the group rank and future goals

The group rank of the Greek grammar currently covers a wide range of NPs. Table 4 summarizes the basic linguistic phenomena that have been accounted for up to now and also the issues that we plan to address in the future.

5 Words and Morphology

The word rank is the part of the grammar, which, in conjunction with the lexicon and the morphological component, controls the realization of words. The *morphological*, *syntactic* and *semantic features* of words need to be specified in order to guide morphological and syntactic realization. Because of space limitations, we present only the first category of features, the morphological features, here.

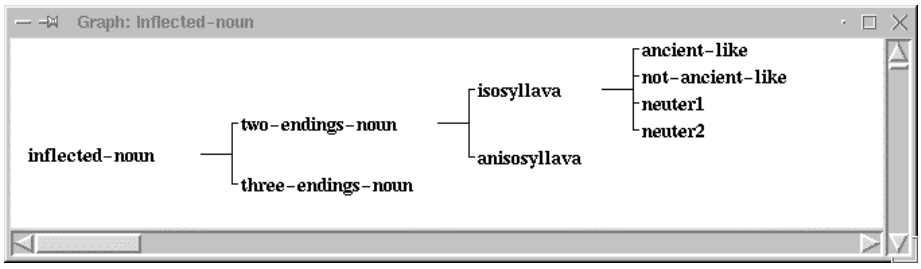


Figure 15: The system network of noun declension

Figure 15 shows the system network describing the declensions of nouns. We distinguish two main morphological categories of nouns: *two ending nouns* (*δικατάληκτα*) and *three ending nouns* (*τρικατάληκτα*). Two ending nouns are subdivided into *ισοσύλλαβα* (same number of syllables) and *ανισοσύλλαβα* (different number of syllables) [Klairis and Babiniotis 1998]. We further divided *ισοσύλλαβα* into four subcategories: masculine and feminine nouns ending in *–ες/-ων* in plural (e.g. "δραχμές", "δραχμών"), masculine and feminine nouns ending in *–εις/-εων* (e.g. "αμφορείς", "αμφορέων") in plural, neuter nouns ending in *–ο* or *–ι* in the nominative singular (e.g. "αγγείο", "σπίτι") and neuter nouns ending in *–ος* in the nominative singular (e.g. "λάθος"). Such declension features together with the gender features (masculine, feminine, neuter) are included in the lexicon entries (Figure 16) and in the morphological rules of the grammar (Figure 17) and are used to generate the inflection types of the words; note that the case and number features are already known to the system from previous stages of processing (Figure 14).

```
(def-lexical-item
:name tetradrachm-noun
:spelling "τετράδραχμο"
:grammatical-features (noun common-noun count-noun inflected-noun neuter-noun neuter1)
:concept (tetradrachm))
```

Figure 16: Lexical entry for "*tetrradrachm-noun*"

```
If features = singular-noun inflected-noun neuter-noun neuter1 (not genitive-noun)
then add suffix ""
```

Figure 17: An example morphological rule

In a similar way, we have accounted for the morphology of verbs. Figure 18 shows the system network describing the conjugations of verbs. The features shown there are combined with features already known to the system, e.g. voice, tense, number and person (Figure 5), for the generation of the surface verb forms. The grammar also supports the morphological generation of determiners and pronouns. Future versions of the grammar will include an account for the morphological generation of adjectives.

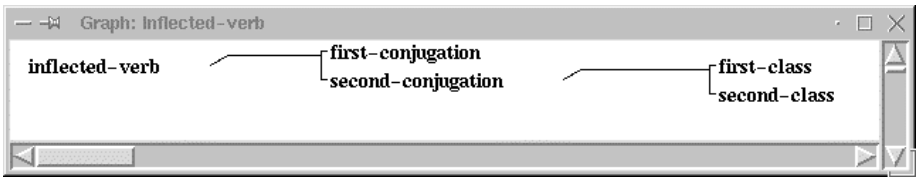


Figure 18: The system network of verb conjugations

As an example, consider the realization of the *Noun* slot of Figure 14. The system looks at the Greek lexicon and tries to find a lexical item with the id "*tetradrachm-noun*" (Figure 16). It retrieves the base form (marked as *spelling*) of the word ("*τετράδραχμο*") and the information related to its declension (*neuter1*), which is included in the *grammatical features* of the word. Then, it consults the morphological component trying to find what changes in the ending of words having the declension feature *neuter1* and also the features *singular-noun*, *neuter-noun*, *nominative-noun*, provided by the group rank of the grammar (Figure 14). The retrieved rule (Figure 17) informs us that no ending (*suffix*) is needed, so the word "τετράδραχμο" is generated as it is.

6 Conclusions and Future Work

In this paper we presented a large-scale computational grammar of Greek, used in the M-PIRO natural language generation system for the surface realization of Greek texts. This grammar is couched in the framework of Systemic Functional Linguistics, which is widely used in generation; this framework allows resource sharing in multilingual generation and, hence, guarantees more economical resources and reduced development time.

Although the grammar is still under development, it already provides a wide coverage of the Greek language. Further work is under way in order to extend the grammar; more specifically, our future goals are summarized as follows:

- Treatment of clitics.
- Subordinate clauses, apart from the relative clauses that are already dealt with.
- Further order variations in clauses and NPs.
- Other NP types (e.g. elliptical NPs, adjectival NPs).

- Morphological treatment of adjectives.

Our long-term goal is to produce a full coverage computational grammar of Greek that can be exploited in several natural language generation applications.

Acknowledgements

The authors would like to thank Mick O'Donnell, former member of the M-PIRO team, for his contribution to the integration of Greek resources into the M-PIRO multilingual system.

References

[Androutsopoulos et al. 2001] Androutsopoulos I., Kokkinaki V., Dimitromanolaki A., Calder J., Oberlander J. and Not E. Generating Multilingual Personalized Descriptions of Museum Exhibits – The M-PIRO Project. Presented at the *29th Conference on Computer Applications and Quantitative Methods in Archaeology*, Gotland, 2001.

[Bateman et al. 1991] Bateman J., Matthiessen C., Nanri C. and Zeng, L. The Re-use of Linguistic Resources across Languages in Multilingual Generation Components. In *Proceedings of IJCAI'91*, Sydney, Australia. 1991.

[Bateman 1997] Bateman J. Enabling technology for multilingual natural language generation: the KPML development environment. *Natural Language Engineering* 3(1): 15-55, 1997.

[Dimitriadis 1995] Dimitriadis A. When Pro-Drop Languages Don't: On Overt Pronominal Subjects in Greek. In *Penn Working Papers in Linguistics 2:2 (Proceedings of the 19th Penn Linguistics Colloquium)*, 1995.

[Dimitriadis 1996] Dimitriadis A. When Pro-Drop Languages Don't: Overt Pronominal Subjects and Pragmatic Inference. In *Proceedings of CLS 32*, 1996.

[Halliday 1985] Halliday M. *An Introduction to Functional Grammar*. Edward Arnold, London, 1985.

[Holton et al. 1997] Holton, D., Mackridge, P. and Philippaki-Warburton, I. *Greek. A Comprehensive Grammar of the Modern Language*. Routledge, London, 1997.

[Horrocks 1982] Horrocks, G. The Order of Constituents in Modern Greek. In Gazdar, Klein and Pullum (eds.), *Order, Concord and Constituency*, pp. 95-111. Foris Publications, Dordrecht, 1982.

[Joseph 1994] Joseph B.D. On Weak Subjects and Pro-Drop in Greek. In I. Philippaki-Warburton, K. Nicolaidis, & M. Sifianou (eds) *Themes in Greek Linguistics (Papers from the First International Conference on Greek Linguistics, Reading, September 1993)*. Amsterdam: John Benjamins Publishers, 1994, pp. 21-32.

[Klairis and Babiniotis 1998] Klairis C. and Babiniotis G. *Γραμματική της Νέας Ελληνικής Ι: Το Όνομα*. Ελληνικά Γράμματα, 1998.

[Mann and Matthiessen 1983] Mann W. C. and Matthiessen C. M. Nigel: A systemic grammar for text generation. *Technical Report ISI/RR-85-105*, Information Sciences Institute, February 1983.

[Mann and Thompson 1988] Mann W. C. and Thompson S. "Rhetorical Structure Theory: Towards a Functional Theory of Text Organization". *Text*, 3:243–281, 1988.

[Mellish et al. 1998] Mellish C., O'Donnell M., Oberlander J. and Knott A. "An Architecture for Opportunistic Text Generation". In *Proceedings of the 9th International Workshop on Natural Language Generation*, Niagara-on-the-Lake, Ontario, Canada, 1998.

[Oberlander et al. 1998] Oberlander J., O'Donnell M., Knott A. and Mellish C. Conversation in the museum: experiments in dynamic hypermedia with the intelligent labelling explorer. *New Review of Hypermedia and Multimedia*, 4, 11—32, 1998.

[O'Donnell 1994] O'Donnell M. *Sentence Analysis and Generation – a Systemic Perspective*. PhD thesis, Linguistics Dept., University of Sydney, 1994.

[O'Donnell 1996] O'Donnell M. Input specification in the WAG sentence generation system. In *Proceedings of the 8th International Workshop on Natural Language Generatio*n, Herstmonceux Castle, UK, June 1996.

[O'Donnell et al. 1998] O'Donnell M., Cheng H. and Hitzeman J. Integrating Referring and Informing in NP Planning. In *Proceedings of the Coling-ACL '98 Workshop on the Computational Treatment of Nominals*, August 16, 1998, Universite de Montreal, Canada. pp. 46-55.

[Philippaki-Warburton 1982] Philippaki-Warburton, I. Προβλήματα σχετικά με τη σειρά των όρων στις ελληνικές προτάσεις. *Γλωσσολογία*, 1: 99-107, 1982.

[Philippaki-Warburton 1985] Philippaki-Warburton, I. Word Order in Modern Greek. *Transanctions of the Philological Society*, pp. 113-143, 1985.

[Philippaki-Warburton 1987] Philippaki-Warburton, I. The theory of empty categories and the pro-drop parameter in Modern Greek. *Journal of Linguistics*, 23: 289-318, 1987.

[Reiter 1994] Reiter, E. Has a Consensus NL Generation Architecture Appeared, and is it Psycholinguistically Plausible? In *Proceedings of the 7th International Workshop on Natural Language Generation (INLGW-1994)*, pages 163-170, 1994.

[Reiter and Dale 1997] Reiter, E. and Dale, R. Building Applied Natural Language Generation Systems. *Natural Language Engineering* 3:57-87, 1997.

[Reiter and Dale 2000] Reiter, E. and Dale, R. *Building Applied Natural Language Generation Systems*. Cambridge University Press, 2000.

[Rosner and Stede 1994] Rosner D. and Stede M. Generating multilingual documents from a knowledge base: the TECHDOC project. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*. Volume I, pages 339-346. Kyoto, Japan, 1994.

[Teich 1992] Teich, E. KOMET: Grammar documentation. *Technical report*, GMD, Darmstadt, Germany, 1992.

[Teich 1999] Teich, E. *Systemic Functional Grammar in Natural Language Generation: Linguistic Description and Computational Representation*. Casssell Academic Publishers, London, 1999.

[Tsimpli 1990] Tsimpli, I. The Clause Structure and Word Order of Modern Greek. In J. Harris (ed.), *UCL Working Papers in Linguistics*, 2: 226-255, 1990.

[Warburton-Philippaki 1977] Modern Greek clitic pronouns and the "surface structure constraints" hypothesis. *Journal of Linguistics*, 13:259-281, 1977.