To appear in the *Proceedings of 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

National Centre for Scientific Research "Demokritos", Athens

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.

Introduction

This paper presents a large-scale computational grammar of Greek, 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 personalized, i.e. reflect the interest of the user as well as certain educational goals, and are both textual and spoken («Ion Androutsopoulos et al. 2001»[2]). M-PIRO builds upon the ILEX natural language generation system («Mick O'Donnell et al, 2001»[3]) and extends ILEX's technology by incorporating improved

[1] More information about M-PIRO is available at: http://www.ltg.ed.ac.uk/mpiro/.

[2] Ion Androutsopoulos et al, 2001, Generating Multilingual Personalized Descriptions of Museum Exhibits–The M-PIRO Project, *29th Conference on Computer Applications and Quantitative Methods in Archaeology*, Gotland.

[3] Mick O'Donnell et al, 2001, ILEX: An Architecture for a Dynamic Hypertext Generation System, *Natural Language Engineerin*g, Cambridge, 7(3):225–250.

multilingual capabilities, high-quality speech output, authoring facilities, extended user modelling mechanisms, as well as a more modular core generation engine.

The M-PIRO generation system consists of four modules: *content selection*, which selects which facts to convey, *document planning*, which determines the ordering of the facts and the rhetorical relations that hold between them, *micro-planning*, which specifies in abstract terms how each fact should be expressed in each language, and *surface realization*, which produces the surface form of each sentence. 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 functional grammars, which are able to treat the syntactic, semantic and even pragmatic aspects of language in one integrated model («Michael Halliday 1985»[4]).

This paper discusses MPIRO's systemic approach, the general organization of the grammar, its coverage, and future goals in very abstract terms. Moreover, given that space limitations do not allow for a thorough discussion, it serves as a link to an extended version of this paper, where the grammar is presented in much greater detail[5].

Grammar Overview

Systemic grammars are primarily concerned with the functions of language and with how these functions are mapped into surface forms by making a series of choices that determine the syntactic characteristics of the sentence being constructed. The linguistic options are called *features*[6] and are represented by *system networks*, with each *system* being a point of choice (Figure 1 shows the top system of the grammar). Systemic grammars have the advantage of allowing resource sharing between different languages in multilingual generation («John Bateman et al. 1991»[7]), and, hence, guarantee more

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

[5] The extended version of this paper is available at: *http://www.iit.demokritos.gr/~adimit/papers/5thcilg.pdf.*

[6] 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.

[7] John Bateman et al, 1991, The Re-use of Linguistic Resources across Languages in Multilingual Generation Components, *Proceedings of IJCAI'91*, Sydney, Australia.

economical resources and reduced development time. Following this approach, the Greek grammar was not built from scratch; it was constructed by taking the English grammar used in M-PIRO («Mick O'Donnell 1994»[8]) as a starting point and making the appropriate modifications and additions in order to develop the Greek generation component.

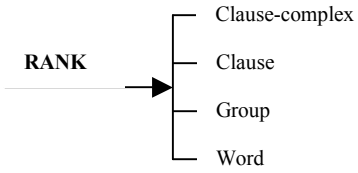

Figure 1: The top system of the grammar, RANK.

Micro-planning specification
(defexpression made-of
:language :greek
:arg1 exhibit6
:arg2 silver
:verb make-verb
:voice passive
:aspect perfect
:arg2-prep of)

Figure 2: Micro-planning specification of the sentence "Αυτό το νόμισμα έχει φτιαχτεί από ασήμι".

The input to the surface realization component is a list of micro-planning specifications (Figure 2); the goal is to transform such specifications into natural language sentences. One of the central properties of systemic grammars is the *rank scale*, which defines the types of linguistic units used in the grammar. The rank scale is often represented as the initial system of the grammar (Figure 1). The Greek grammar is divided into four ranks: the *clause-complex* rank, which describes the structure of multi-clausal units, the *clause* rank, which describes the clause structure, the *group* rank, which describes the structure of noun phrases (NPs) and prepositional phrases (PPs), and the *word* rank, which describes the grammatical, syntactic and semantic features of words. At the current state of the Greek grammar, the last three of these ranks have been developed. Regarding the clause complex rank, the corresponding part of the English grammar is currently used. 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. Furthermore, apart from the grammar, there are two complementary resources: the *lexicon* and the *morphological component*. The lexicon provides the lexical items

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

necessary for the realization of words and the morphological component is responsible for the morphological generation.

Linguistic Coverage

The clause rank of the Greek grammar currently covers declarative main clauses and relative clauses. The linguistic phenomena that have been accounted for up to now include the constituent order (the S-V-O and O-V-S orders are currently supported), clause aspect, pronominal subjects, case assignment, pronoun-antecedent agreement etc. Further work is under way to account for clitics, more order variations and other types of clauses (subjunctive, imperative and non-finite clauses).

The group rank currently covers a wide range of NPs (common, proper, pronominal, classifying and generic NPs) handling phenomena such as agreement, proximal deixis, proper deixis, as well as the basic ordering of Greek NPs, as discussed in «David Holton et al. 1997»[9]. Elliptical and adjectival NPs, as well as other possible order variations will be supported in future versions of the grammar.

The word rank focuses on the morphological realization of words, which is carried out in conjunction with the morphological component. The grammar currently supports the morphological generation of nouns, verbs, determiners and pronouns, while future versions will include an account for the morphology of adjectives.

Conclusions

In this paper we presented a large-scale computational grammar of Greek, used in the M-PIRO project. This grammar is couched in the framework of Systemic Functional Linguistics, which is widely used in generation. 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, further order variations in clauses and NPs, other NP types, morphological treatment of adjectives. Our long-term goal is to produce a wide coverage computational grammar of Greek that can be exploited in several natural language generation applications.

[9] David Holton et al., 1998, *Γραμματική της Ελληνικής Γλώσσας*, Αθήνα, Εκδόσεις Πατάκη, σελ. 332.