Title	HPSG
Author(s)	,
Citation	
Issue Date	2003-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1668
Rights	
Description	Supervisor: , ,

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Extended the HPSG Grammar for Ancient Greek

Ken'ichiro Nakajima (110091)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 14, 2003

Keywords: HPSG, Ancient Greek, LiLFeS, Syntactic Parsing.

In this paper, we describe Head-driven Phrase Structure Grammar for the Ancient Greek we developed.

The purpose is to enable us to treat not only the these literature's surface information but also the syntactic one.

On many literature called by Ancient Greek classics, we have already lost originals, and so we are retaining only copied manuscripts at the present day. Because many manuscripts of these books have been produced by transcribers until now, they were added an individual interpretation or were mis-copied by transcribers. Thus, they have the problems that how was the original. In addition, even the book said to have been written by a specific author at present, it is not known whether it is really right or not.

The way to solve these problems is to retrieve as much information from sentences as possible, and then to research them. In order to do the above, we consider to retrieve syntactic information from the literatures of Ancient Greek.

Ancient Greek which we worked has some features– some adjectives rule nouns by the agureement of case. When an adjective or an article modify a noun phrase, they require the agreement of gender, number, and case. Although Ancient Greek is regarded to be the SOV language, the place after a verb is important for the word to modify or to compensate the subject and the complement of verb phrase. Therefore, it is said that the VO language. For these reasons, when the adjective rules the noun

by the agreement of case, the noun is not necessary to be juxtaposed by the adjective. Not only the subject of a sentence but also other elements which was understood by readers or listeners can be omitted. There are the modification relations which are crossing in a sentence.

On the judgment of grammatical interpretation, however, grammar difiners cannot avoid subjective decision. At this point, we recognize the appropriateness of Euclid's Elements. Because logical structure of Euclid Elements is more clear than other epics', like the Homer Iliad, Odyssey, and just the mathematical proofs are written through all books. We, therefore, can rather objectively decide syntactic information.

In consequence, we developed the grammar which can parse Euclid Elements, after we have analyzed Ancient Greek based on Head-driven Phrase Structure Grammar(HPSG).

HPSG, which is one of unification grammars for sentence parsing, consists of many lexical entries. These entries can have rich information as features. In addition, a phrase is formed by unification these informations of each word with a few grammatical rules, called by ID-schemata and principles. In Ancient Greek, each word has many attributes, like gender, number, and case, and a phrase is constructed by their agreements. We handle these attributes as features. As a result, we acquire the syntactic information from the surface one.

There is already the Ancient Greek grammar based on HPSG proposed by Senda. This grammar which targets 75 sentences of 76 sentences in Book 2 in Elements has high accuracy. This grammar, however, has the defect which can be applied to only a certain kinds of sentences. This is due to the strictness of this grammar. Our purpose is to design a Ancient Greek grammar which is capable of wide application than Senda's.

In some of the rules of other language grammars which is based on HPSG has been proposed, a phrase consists of the pair of the complement phrase or the modifier phrase with the head phrase. We analyzed the characteristic which Ancient Greek grammar has; parts of some classes are free from word order, but other classes isn't. We improved these rules based on our analysis. In addition, we developed some rules, one of the rules allows to make a phrase structure accomplish, taking an adjunct phrase from the outside. Another rule allows to make the verb phrase can omit the subject

feature when it already has the complement feature even if it requires that subject feature which must be saturated in a dictionary. Zero copula rule allows to change to a verb phrase which already fills complement feature from the word which can be the complement of a verb. Furthermore, we considered how to handle the differences and similarities between the noun type and the adjective type, how to think the behavior of the conjunction type, how to select which is head word between the article or noun type or symbol type. Then we redefined the type hierarchy based on it. Including these all, our grammar has 16 schemata and 1354 lexical entries.

We could get totally 79.029% coverage which we applied in application on Book 7 and 8 of Euclid Elements. This part has 16.9 words per one sentence. There are 8 sentences in Book 7 and 7 sentences in Book 8 which cannot be parsed, for they are too ambiguous in modification relations. These sentences did not have enough syntactic information. It says about each book, Book 7 is 85.884% and Book 8 is 69.439%.

Incidentally, when we experiment without abjunct from behind rule and subject ellipsis rule, we can get the coverage of 69.711% on Book 7, 49.576% on Book 8. Simultaneously, when without subject ellipsis rule, we can get 78.843% on Book 7, and 65.042% on Book 8. According to the above, the rules which we suggested have some degree of positive significance for parsing of Ancient Greek classics. Our grammar was implemented on LiLFeS. A parser which we used is naive CKY-style parser on LiLFeS. Thus our purpose is accomplished.

On the other hands, this experiments showed that the parsing time was comparatively long. We consider that increase of the parsing time is related to the ambiguity of modification relations.

In this work, we used only the surface information. Meanwhile, we didn't use the semantic and pragmatic information. As a result, we leave the problem about the ambiguity of modification relations. To reduce the local ambiguity contributes to exploit the reduction of the ambiguity of a whole of sentence.