Analysis of *The Elements* by HPSG

Satoshi Tojo
School of Information Science
Japan Advanced Institute of Science and Technology
Ishikawa 923-1292, Japan
tojo@jaist.ac.jp

Ken Saito
Department of Human Sciences
College of Integrated Arts and Sciences, Osaka Prefecture University
Osaka 599-8531, Japan
ksaito@be.cias.osakafu-u.ac.jp

Abstract

In this study, we analyse *The Elements* of Euclid by Head-driven phrase structure grammar (HPSG). In ancient Greek, phrase structures depend mainly upon the agreement of features (gender, case, person, etc.) so that there is much liberty in word order; this is the reason why we adopt a unification-based grammar. In addition to ordinary grammar rules, we add several rules which concerns ellipsis and crossed-dependency. We show that the grammar could cover over 79% for 1154 sentences of Books 7 and 8 of *The Elements*.

1 Introduction

The Elements, the colossal collection of fundamental propositions in mathematics, was compiled from the fifth century B.C., and the existent version is attributed to Euclid, active presumably around 300 B.C. However, the text in our possession is based on several medieval manuscripts after ninth century. By that time the text had been copied and edited so many times, so that it is now difficult to distinguish the genuine contents originated by the author from those which had been added in the process of compilation in later years [4, 8, 1].

Our objective of this study is two-fold; one is to propose a grammar for ancient Greek, the word order of which are rather free though their phrase structures strongly depend upon the agreements of gender, number, and case. The other is to analyze *The Elements*, that is, to find the dependency of mathematical concepts, or of co-occurrence of them, that may contribute to the restoration of the original edition.

Ancient Greek has many variations in eras and in regions. We treat the dialect of Attika that was spoken in Athenes in the fourth or fifth century B.C. [2, 6]. Any lexical item in a sentence may be arbitrarily omitted when it is obvious in the context. For example, Book 7 Proposition 2:

μετρείτω, καὶ ἔστω ὁ Z
(measure, and let be the Z)

needs to be compensated with missing words as follows [10]: "Let such a number measure them, and let it be Z."

In this paper, the italicized label in a feature structure represents its type. Also, we arbitrarily omit accent symbols as well as guttural sound marker in the figures.

2 Principles and Schemata

In this section, we design a fundamental grammar of ancient Greek in HPSG. Although we have consulted the modern Greek one proposed by Kolliakou [5], we need to modify it in many aspects because it is very different in case, mood, phonology, and so on.

Besides such conventional HPSG principles as Head feature principle, Valence[1] principle, Phonology principle, and Gap principle, we add **Particle principle**; if a daughter includes a particle it is reversely inherited to her mother (cf. Section 3).

As for schemata, we employ the followings.

Head Complement Schema (1) regulates how a word takes a complement.

$$\begin{bmatrix} phrase \\ \text{COMPS} \quad \langle\ \rangle \end{bmatrix} \rightarrow \text{H} \begin{bmatrix} word \\ \text{COMPS} \quad \langle \boxed{1} \rangle \end{bmatrix} \boxed{1} \tag{1}$$

Here COMPS is a complement feature; because some verbs prepose complements, we need to implement two different schemata for pre-/post-positions of $\boxed{1}$ in (1).

Head Subject Schema (2) combines a phrase, the head of which is a verb, with a subject word. In ancient Greek, a subject may be located between the complement and the verb itself. Thus, we need to implement two schemata, in accordance with the location of subjects.

$$\begin{bmatrix} phrase \\ \text{SUBJ} \quad \langle\ \rangle \end{bmatrix} \rightarrow \boxed{1}[\quad] \text{H} \begin{bmatrix} word \\ \text{SUBJ} \quad \langle \boxed{1} \rangle \end{bmatrix} \tag{2}$$

Head Modifier Schema (3) concerns the modification; if MOD (modifier) feature in VALENCE can be unified with HEAD of a phrase, the phrase is modified. In Figure 1 (Book 8, preposition 24),[2] we show an example of modification; the numeric δύο (two) modifies the noun ἀριθμοί (numbers), and the prepositional phrase headed by πρός (to) modifies the noun λόγον (proportion), respectively.[3] In general, modifiers are placed in different positions, dependent on if they are in the restrictive use or in the predicative use. In case of the restrictive use as in οι (ἐλάχιστοι ἀριθμοί (the smallest number), the modifier is placed between the determiner and the noun; in which case the value of SPR (specifier) of the modified phrase, which concerns the agreement, is not filled. In case of predicative modification, the modifier is placed outside of the preposition and the noun, as οἱ ἀριθμοὶ ἐλάχιστοι (the number smallest) or ἐλάχιστοι οἱ ἀριθμοί (smallest the number); in which case SPR is filled. The preposition does not always appear and thus this does not matter whether SPR is filled or not, though the modifier itself must be filled when it modifies another phrase.

$$\begin{bmatrix} phrase \\ \text{COMPS} \quad \langle\ \rangle \\ \text{MODS} \quad \langle\ \rangle \end{bmatrix} \rightarrow \boxed{1}[\quad] \text{H} \begin{bmatrix} word \\ \text{COMPS} \quad \langle\ \rangle \\ \text{MODS} \quad \boxed{1} \end{bmatrix} \tag{3}$$

[1] Although the term *valence feature* is used only for those that are obligatory in sentences [9] such as complements for verbs, we extend this feature to include optional modifiers in this study.

[2] The main verb of this sentence ἔχωσιν (have) is in the conjunctive; thus, the whole structure is not a sentence but a clause headed by ἐάν (if).

[3] In an English translation of [10], the prepositional phrase by πρός (to) modifies ἔχωσιν (have). However, because the meanings are same, we accept this result of analysis avoiding unnecessary complication of grammar rules.

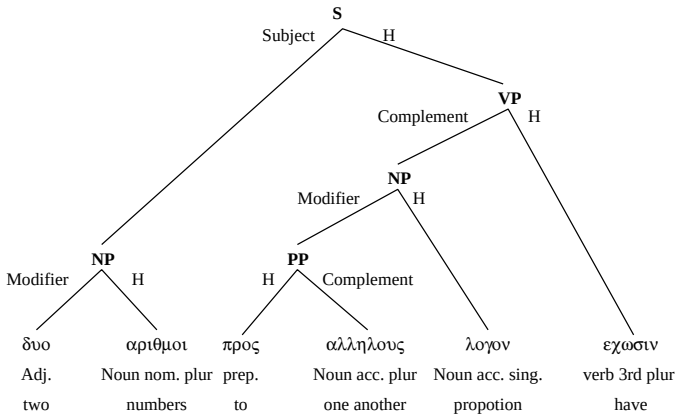

Figure 1: A modification by an adjective

It is still arguable as to which is the head between the noun and the determiner; some grammars regulate that determiners dominate nouns, and others are converse. Although both of these have their own advantages and disadvantages, we regard that nouns are heads, that is, nouns dominate determiners, contrary to Creider [3]. This facilitates the analysis of such named geometric objects as *circle O*, *line segment AB*, and so on. In mathematical literature of ancient Greek, these objects are often represented by determiners; for example, 'a feminine determiner + two characters' represents a line segment, 'a neuter determiner + a character' a point, 'a neuter determiner + ὑπό + two characters, two characters', a rectangle, and so on. These phrases are considered to form noun phrases directly from nouns, instead of forming DP (determiner phrase). Therefore, we avoid the useless distinction of DP and NP, and render the nouns heads.

Nouns and adjectives must agree in Gender, Number, and Case are included in SPR feature. We define *decl* type, that is a declinable part of speech in the type hierarchy, which subsumes nouns and adjectives. A determiner together with a word of type *decl* form a phrase with **Head Specifier Schema** (4).

$$\begin{bmatrix} phrase \\ \text{SPR} \quad \langle\ \rangle \end{bmatrix} \rightarrow \boxed{1}\begin{bmatrix}\ \end{bmatrix}\ \text{H}\begin{bmatrix} decl \\ \text{SPR} \quad \langle \boxed{1} \rangle \end{bmatrix} \tag{4}$$

In this section, we adjusted those fundamental schemata used in HPSG to ancient Greek. However these schemata could not accommodate some specific problems of ancient Greek. In the following section, we augment schemata to resolve these problems.

3 Gaps and Dependency

The problems of ancient Greek mainly lie on (i) gaps caused by ellipsis of lexical items, and (ii) crossed dependencies. In this section, we discuss these problems.

Nominalization Determiners of ancient Greek nominalize not only adjectives but also adverbs and other phrases. If they inflect in accordance with gender, number, and case, they could be clues to find dependencies. As we cannot assume that any kind of parts

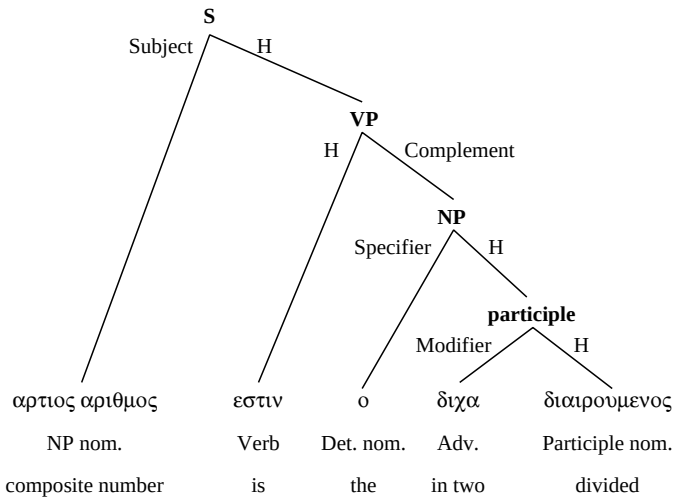

Figure 2: Nominalization

of speech are nominalized, we restrict that only those under *decl* (declinable) can take determiners. In Figure 2 (Book 7, Definition 6), διαιρούμενος (divided) is a participle and functions as a verb, so that it can modified by an adverbial phrase δίχα (in two). Furthermore, a participle has features of an adjective, it can be nominalized, headed by a determiner. In this case, the corresponding determiner becomes ὁ (the), considering its gender, number, and case. Thus, those lexical items classified into *decl* take determiners in accordance with SPR feature, by Head Specifier Schema (4).

Demonstrative determiners Determiners in ancient Greek also play roles of demonstratives. In the following sentence (Book 7, Proposition 2):

> λειφθήσεταί τις ἄρα ἀριθμός, ὃς μετρήσει τὸν πρὸ ἑαυτοῦ
> (therefore some number will be left which will measure the one before itself),

there is only a prepositional phrase πρὸἑαυτοῦ (before itself) and is no noun after the determiner τόν. However, we can guess from the agreement of gender and number, together with context (that is, Euclidean mutual division), the determiner of masculine single causative τόν corresponds to the antecedent of the clause ἀριθμός (number).

Indefinite pronoun Indefinite pronouns are used either pronouns or adjectives. When it is used as a pronoun it does not inflect. However when it is used as an adjective, it requires agreements. The following sentence (Book 7, Proposition 3):

> τοὺς Δ, Γ ἀριθμοὺς ἄρα ἀριθμός τις μετρήσει
> (therefore some number will measure the numbers D, G),

the noun ἀριθμός (number) is modified by τὶς (some), and gender, number, case are agreed in them. Therefore, in the lexicon τὶς includes the information concerning the agreement.

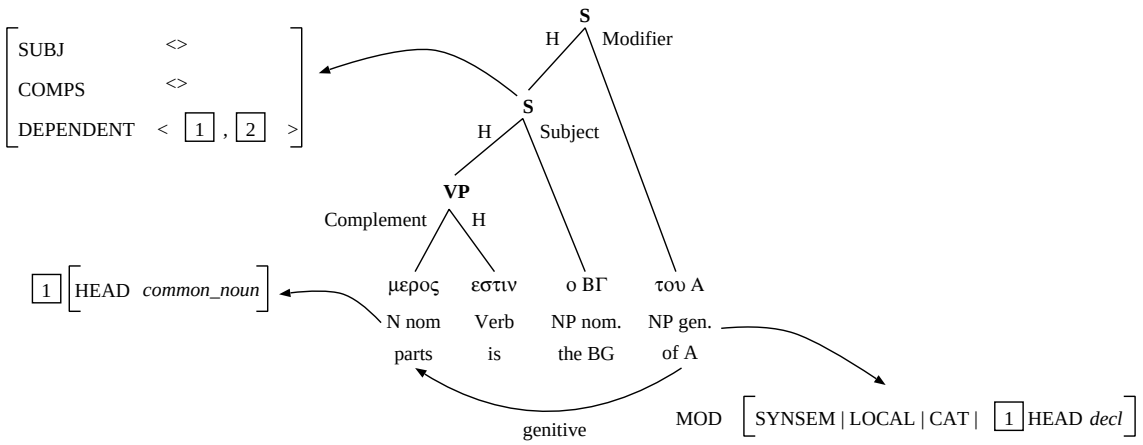

Figure 3: Backward modification

Postposition Generally speaking, dependencies rarely cross in natural language. However, in ancient Greek, a complementary phrase may appear at the tail of a sentence, to explain a preceding phrase in the sentence, and this results in crossed dependency. See the example in Figure 3 (Book 7, Proposition 4). In order to solve this, we need to provide us with SLASH feature. In Head Complement Schema and Head Subject Schema, when VAL feature is filled, the HEAD feature of the filler is retained in VAL|DEPENDENT feature. This head in this feature may have dependency relations with a modifier at the tail of a sentence. This mechanism is implemented in the following **Post-Adjective Schema** (5).

$$\left[\ \text{DEPEND}\ \ \langle \boxed{2} \rangle\ \right] \rightarrow \text{H}\left[\begin{array}{l}\text{HEAD}\ \mathit{verb}\\ \text{DEPEND}\ \langle \boxed{1}, \boxed{2} \rangle\end{array}\right]\left[\begin{array}{l}\text{HEAD}\ \boxed{1}\\ \text{VAL}\left[\begin{array}{l}\text{SUBJ}\ \langle\ \rangle\\ \text{COMPS}\ \langle\ \rangle\end{array}\right]\end{array}\right] \tag{5}$$

Ellipsis of subjects Subjects are omitted often when the verb is in the first person or in the second person. We define **Subject Ellipsis Schema** (6) as follows.

$$\left[\begin{array}{ll}\mathit{phrase}\\ \text{SUBJ} & \langle\ \rangle\\ \text{COMPS} & \langle\ \rangle\\ \text{MOD} & \langle\ \rangle\end{array}\right] \rightarrow \left[\begin{array}{ll}\mathit{phrase}\\ \text{HEAD} & \mathit{verb}\\ \text{SUBJ} & \langle \mathit{sign} \rangle\\ \text{COMPS} & \langle\ \rangle\\ \text{MOD} & \langle\ \rangle\end{array}\right] \tag{6}$$

Zero copula Copulative verbs are also often omitted, and such sentences are called zero-copula. In particular, εἰμί (be) requires nominative case both for subject and for complement. In order to admit this phenomenon, we need to revise the ordinary subject-

complement connection as **Zero Copula Schema** (7).

$$\left[\begin{array}{l}\text{HEAD}\ \left[\begin{array}{l}verb\\ \text{CASE}\ nom\end{array}\right]\\ \text{VAL}\ \left[\begin{array}{l}\text{SUBJ}\ \langle\ \left[\begin{array}{l}\text{HEAD}\ \left[\begin{array}{l}decl\\ \text{CASE}\ nom\end{array}\right]\\ \text{VAL}\ \left[\begin{array}{ll}\text{SUBJ} & \langle\ \rangle\\ \text{COMPS} & \langle\ \rangle\end{array}\right]\end{array}\right]\ \rangle\\ \text{COMPS}\ \langle\ \rangle\end{array}\right]\end{array}\right] \rightarrow \left[\begin{array}{l}\text{HEAD}\ \left[\begin{array}{l}decl\\ \text{CASE}\ nom\end{array}\right]\\ \text{VAL}\ \left[\begin{array}{ll}\text{SUBJ} & \langle\ \rangle\\ \text{COMPS} & \langle\ \rangle\end{array}\right]\end{array}\right] \tag{7}$$

Case agreement by adjectives An adjectives requires a certain case for those to be modified. In addition, an adjective dominates the dative case when it represents degree or time. In the following sentence (Book 7, Proposition 21):

> εἰ γὰρ μή, ἔσονταί τινες τῶν A, B ἐλάσσονες ἀριθμοὶ ἐν τῷ αὐτῷ λόγῳ ὄντες τοῖς A, B
> (For, if not, there will be some numbers less than A, B (being) in the same ratio with A, B),

the final pronoun αὐτῷ (same) is used as an adjective. Thus, this takes the dative: τοῖς A, B (with A, B). Therefore, the tree structure becomes as Figure 4. In addition, comparative adjectives make their targets of comparison the genitive, and these targets can be rendered as complements of adjectives. Such kind of complements may appear afar beyond clauses. In order to process such *long distance dependency*, we define **Slash Schema** (8) with SLASH feature. With this schema, an adjective copies COMPS to SLASH while HEAD retains as it is.

$$\left[\begin{array}{ll}phrase & \\ \text{COMPS} & \langle\ \rangle\\ \text{SLASH} & \langle\boxed{1}\rangle\end{array}\right] \rightarrow \left[\begin{array}{ll}phrase & \\ \text{HEAD} & adjective\\ \text{COMPS} & \langle\boxed{1}\rangle\\ \text{SLASH} & \langle\ \rangle\end{array}\right] \tag{8}$$

Distinction of nouns and adjectives In ancient Greek, ajdectives may be substituted for nouns, or modified nouns are omitted when they are obvious. Although modern European languages take nouns for subjects, ancient Greek takes an adjective as a subject preceded by a determiner as we mentioned before; even more confusing, an adjective can be a subject as it is without a determiner. The similar phenomenon occurs not only for adjectives but also for participles. Thus, in this study, we put *decl*, that is a super category for adjectives and nouns, in the place where ordinary other languages require nouns. The following is an example of such a lexical item ἐστιν (is).

$$\left\langle \text{ἐστιν}, \left[\begin{array}{ll}word & \\ \text{HEAD} & common\ verb\\ \text{SUBJ} & \langle\left[\begin{array}{ll}\text{HEAD} & decl\\ \text{CASE} & nominative\end{array}\right]\rangle\end{array}\right]\right\rangle$$

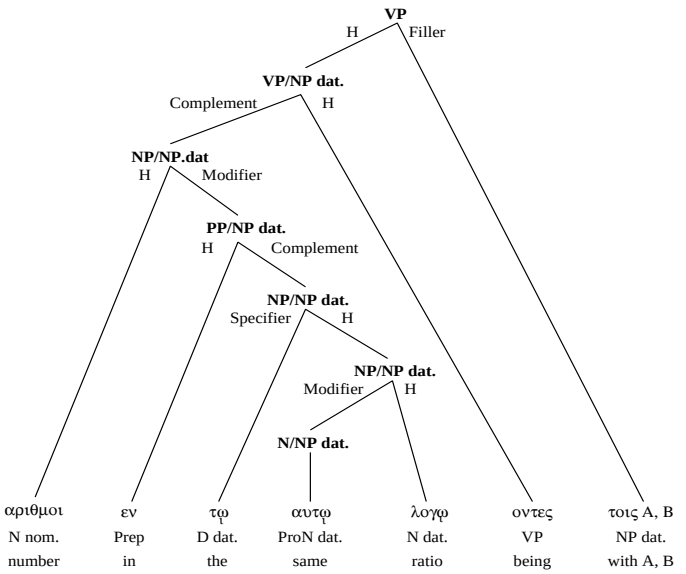

Figure 4: An adjective with the dative complement

Particle Particles in ancient Greek appear at the second position of sentences, and with this feature these words should be treated differently from other words. Because the particle may work as conjunctives, keeping the second position, we defined **Particle schema** (9), in which particles leave only PHONOLOGY feature. We will mention the problem of particles later in Section 5.

$$\begin{bmatrix} phrase \\ \text{PARTICLE} \quad \langle \boxed{1} \rangle \end{bmatrix} \rightarrow \text{H} \begin{bmatrix} phrase \\ \text{PARTICLE} \quad \langle\ \rangle \end{bmatrix} \begin{bmatrix} word \\ \text{PHONO} \quad \langle \boxed{1} \rangle \\ \text{HEAD} \quad particle \end{bmatrix} \tag{9}$$

4 Parsing Result

In this experiment, we employ *LiLFeS* parser [7] for the efficient process of feature agreement. The number of words which appear in *The Elements* seems between 300 and 500, though we registered all the inflections as lexical items in this study, that comes to 1,352. We applied the grammar, including 16 schemata, to the Book 7 and 8 of *The Elements* and evaluated the coverage and the efficiency. As to the coverage we counted the ratio of parsed sentences that acquired the beginning symbol of the parsing tree, and as to the efficiency we measured the mean time of parsing. Among those sentences, 8 sentences in Book 7, and 7 sentences in Book 8 could not be analyzed for ambiguity. The experiment is done with a processor athron 1800+, memory 500MB.

Both of computation time and coverage are very different between two Books; 85.9% of 673 sentences in Book 7 are parsed in 3.531 seconds per sentence on average while so are 69.4% of 481 sentences in Book 8 in 6.665 seconds. This is because Book 8 includes many adjectives which dominate cases and results in ambiguity.

Post-Adjective Schema (5) and Subject Ellipsis Schema (6) surely augment possible interpretations. In order to evaluate the efficiency of these two schemata, we compared the coverage subtracting these schemata, first only (5) and then both of (5) and (6), from

the full grammar. As a result, we found that the ratio is impaired about by 10%, and thus these schemata were ensured to play important roles in the analysis.

5 Remaining Problems

In this section, we summarize several difficult problems. First, the *absolute genitive*, that is a modifier by a genitive participle, cannot be treated properly in our grammar; if we admit this in MOD feature, the parser would produce more useless trees. Conjunction by particles is another cumbersome problem. δέ together with μέν builds a parallel structure though verbs are often omitted; thus, only cased nouns are remained in one clause, causing confusion to the parser. Another, no less difficult problem is that relative clauses may precede the main clauses. In which case, the parser can hardly identify the antecedent.

In this study, we analyzed the superficial syntax of single sentences with no consideration of the context. Thus, those sentences which are strongly dependent on the context are still hard to be analyzed. If we were to improve the coverage of analysis, we would need to develop some mechanism to inherit information from preceding sentences.

Acknowledgment

The authors would like to thank two superb implementers Ken'ichiro Nakajima and Keisuke Senda for the Greek HPSG parser.

References

[1] Acerbi, F. Drowning by multiples: Remarks on the fifth book of Euclid's elements, with special emphasis on Prop.8. *Archive for History of Exact Sciences*, 57:175–242, 2003.

[2] Crane, G. The Perseus Digital Library, `http://www.perseus.tufts.edu/`.

[3] Creider, C. and Hudson, R. Case Agreement in Ancient Greek:implications for a theory of convert elements. In Kensei Sugiyama, editor, *Word Grammar*. forthcoming.

[4] Knorr, W. R. The wrong text of euclid: On heiberg's text and its alternatives. *Centaurus*, 38:208–276, 1996.

[5] Kolliakou, D. A composition approach to modern greek 'weak form' possessives. In J-P. Koening G. Webelhuth and A. Kathol, editors, *Lexical and Constructional Aspects of Linguistic Explanation*, pages 31–46. 1998.

[6] *Liddel and Scott's an Intermediate Greek-English Lexicon*. Oxford University, 1959.

[7] Makino, T., Torisawa, K., and Tsujii, J. LiLFeS-towards a practical HPSG parser. In *COLING-ACL'98*, 1998.

[8] Rommevaux, S., Djebbar, A., and Vitrac, B. Remarques sur l'histoire du texte des éléments d'euclide. *Archive for History of Exact Sciences*, 55:221–295, 2001.

[9] Sag, I. A. and Wasow, T. *Syntactic Theory*. The University of Chicago Press, 1999.

[10] Heath, T. L. *Euclid the Thirteen Books of the Elements*. Cambridge University Press, 1925.