

## HPSGによるユークリッド『原論』の構文木の作成

仙田 圭介

北陸先端科学技術大学院大学  
情報科学研究科

k-senda@jaist.ac.jp

斎藤 憲

大阪府立大学 総合科学部  
人間科学科

ksaito@hs.cias.osakafu-u.ac.jp

東条 敏

北陸先端科学技術大学院大学  
情報科学研究科

tojo@jaist.ac.jp

## 1 はじめに

古代の文書・著作は碑文などの例外を別にすれば、当時の原本は残っておらず、後世に書き写された写本でのみ伝わる。これらの資料の真贋の決定、著者の同定、編纂物における原著者の同定や、後世の加筆部分の確定などの作業において、テキストの表現・文体の解析は重要な方法である[4]。テキストの表象に加え、構文構造を知ることはこの判断の一助になる。構文構造は構文木で表現し、その情報を利用することが出来れば便利である。

ギリシア数学の古典、ユークリッド『原論』では、過去に取り上げられた命題を参照するとき、その命題番号ではなく、命題の表現を繰り返すため、参照元を探すことが困難であるという問題が存在する。これらの問題の解決についても構文木の情報を使用することが可能であると考えられる。

一方、言語の解析に対して用いられる文法の代表的なものに、句構造文法や単一化文法がある。句構造文法では、文法的構造を解析できたとしても、意味的構造を得るのは困難である。それに対して、単一化文法では文法的な解析と同時に意味的な解析を行う。意味情報を得ようとするときにこれは有用な手法であると言える。本研究では、この単一化文法の一つである主辞駆動句構造文法 (Head-driven Phrase Structure Grammar, 以下、HPSG)[3] に注目した。

HPSG は、ID-schema, principle と呼ばれる文法規則と辞書記述からなり、これらの文法の構成要素はすべて型付き素性構造と呼ばれるもので記述される。そして、その文法規則には一般的な性質しか書かず、詳細な情報は辞書記述で表す。また、文の解析は ID-schema, principle に基づく型付き素性構造間の単一化によって行われることとなる。

このため、HPSG は単語・句・文が多くの辞書情報を持つ一方で、構文木を生成するための規則は少ないという特徴をもつ。ユークリッド『原論』が著されている古代ギリシア語は動詞変化などの活用のため、素性が多く、異なる素性で似たようなたくさんの文法規則

則を書かなければならない。

HPSG ならば、辞書に書き加えるという容易な手段でこれに対応することができる。また、ユークリッド原論に現れる単語の種類は 300 から 500 と少なく、この点でも HPSG を用いることが適していると考えられる。そこで本研究では、HPSG を用いてギリシア語のための文法を LiLFeS[1] 上に作成し、ユークリッド『原論』第2巻、命題2から命題5までの文について正解の構文木を作成するシステムを実現した。

## 2 古代ギリシア語の特徴と文法の構成

文法を型付き素性構造を扱うことのできるプログラミング言語 LiLFeS[1] を用いて実装した。LiLFeS 上では日本語 [2]・英語 [5] の文法が実装されている。そのため、文法を実装する上で問題となる主な点を古代ギリシア語の特徴に絞って以下に挙げる。なお、以下に現れるギリシア語とそれに対応する日本語については付録に添付した。

## 2.1 語順の自由さ

「直線 AB」は、'εὐθεῖα ἢ AB', 'ἢ AB εὐθεῖα' のどちらでも記述することができる。

また、名詞などが格情報を持っているため、一文の中での語順も自由である。たとえば、英語などに見られる SVO のような語順だけでなく、VSO, VOS などのような表現も可能である。

「AE は AZ, GE に等しい」という意味の文は以下のように書ける。

例 2.1 (ἴσον ἐστί) (τὸ AE) (τοῖς AZ, ΓΕ).

(等しい)(AEは)(AZ, ΓΖに)。

この文は VSO の順に並んでいるが、以下のように書くこともできる。

例 2.2 (Τὸ ΑΕ) (ἴσον ἐστὶ) (τοῖς ΑΖ, ΒΕ).

上は SVO の語順であるが、その他の語順で表されたとしても、同様に文として成立する。

このため、主語と動詞の関係を規定する schema (ID-schema と principle を単一化したもの) である Head-subject schema は主語、動詞の語順について問わないこととした。同様の理由で、動詞とその主語以外の補語について規定する Head-complement schema についても語順は問わないこととした。

## 2.2 語の省略

「直線 AB」は、直線を表す語「εὐθεῖα」を用いず、「 $\eta$  AB」のみでも「直線 AB」の意味で通る。

名詞、冠詞は性・数・格の情報を持つため、名詞が省略されたとしても、冠詞から性・数・格の情報が得られ、頻繁に用いられる表現では、これらの情報から省略された名詞を推定できるためである。

「直線 AB」といったときの AB に代表されるようなある特定の図形を表現するために用いている記号については *symbol* という名の品詞を新たに定義した。

品詞についての型階層において、*symbol* は *noun* の特殊なものとした。これにより、冠詞+記号という並びの句を Head-specifier schema により特殊な名詞句の一つとして扱うことができる。

また、実は「 $\eta$  εὐθεῖα」(直線)という表現自体が、「真っ直ぐな」という意味の形容詞の女性形(および冠詞)に過ぎず、「 $\eta$  εὐθεῖα γραμμῆ」の省略形である。「直線 AB」にあたる表現ではさらに「εὐθεῖα」までもが省略可能である。

## 2.3 遠隔依存の問題

「(直線)AB, ΒΓ によって囲まれる長方形」はギリシア語で書くと以下のように書くことができる。

例 2.3 τὸ ὑπὸ τῶν ΑΒ, ΒΓ περιεχόμενον ὀρθογώνιον

この文の各単語をそのまま英単語に置き換えると次のような句ができる。

the by the AB, BG contained rectangle

このとき、最後に位置する長方形をあらわす名詞 *rectangle* につく冠詞 *the* はこの句の最初に現れる *the* である。なお、同様の語順はドイツ語に見られる。

2.1 で示したように、Head-complement schema は動詞と補語の順序は問わないとしている。これにより、上に示した句は図 1 に示すような解析木を作成することができた。

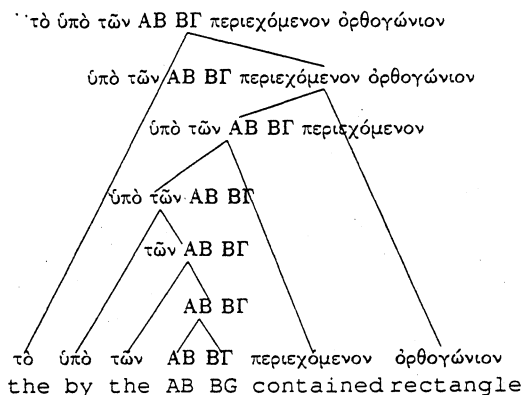


図 1: 'τὸ ὑπὸ τῶν ΑΒ, ΒΓ περιεχόμενον ὀρθογώνιον' の解析木

## 2.4 particle

古代ギリシア語には *particle*(小辞)と呼ばれる特殊な一群の語があり、接続詞や副詞の機能を持つ。この語は他の語の構成する文法構造にかかわらず、常に文の構成要素の 2 番目に現れるため、構文解析をする上での妨げとなる。

以下の文において、下線の引かれている語が *particle* と呼ばれるものである。

例 2.4 Ἄναγεγράφθω γάρ ἀπὸ τῆς ΑΒ τετραγώνων τὸ ΑΔΕΒ.

何となれば、(直線) AB 上に正方形 ΑΔΕΒ が描かれたとする。

この文に現れる *particle*, 'γάρ' は理由を表し、日本語では「何となれば」と訳す。このように、*particle* は接続詞的な役割を果たし、文・句全体に掛かることが多い。そこで、LiLFeS を用いた構文解析を行う前処理として、*particle* の移動をしている。

前処理により、「γάρ」などの *particle* が現れたときには現れた文・句の先頭に移動させることにより構文解析が容易になった。*particle* の移動による文法構造の破綻は見られていない。

また、対になって用いられる *particle*, 'μέν ... δέ ...' (一方... 他方...) においては、対になる後半部分の動詞などが省略されることがある。

このように、*particle* が用いられたことなどからくる語の省略については、どのような語が省略されたかを推定するのが困難であるため、当面は手で補うこととした。





図 4: 'ἴσον ἐστὶ τοῖς ΑΖ ΓΕ' の素性構造

## 4 おわりに

構文解析前の前処理と合わせて、ユークリッド『原論』に見られる語順の自由さや語の省略、対応する冠詞と名詞の間が離れる問題、particle に置ける諸問題などについて対応できるような文法を実装した。

これにより、ユークリッド『原論』第2巻 命題2 から命題5 (76 文) 中、75 文について、正解を含む解析木を作成することが可能となった。なお、作成した辞書記述は 144 個 (図形を特定するための“AB”のような記号類は除く)、ID-schema は 13 個である。

残りの 1 文については SLASH 素性を改良することにより可能になると考えている。

今後は現在手作業で行っている省略語の補完を自動的に行えるようにすること、ユークリッド『原論』の他の命題の解析、他の文献の解析を可能にすることで、実際の古典研究においてこれらの解析結果を用いることができるようにすることなどが目標となる。

## 謝辞

HPSG についてのさまざまな助言をいただき、LiL-FeS という便利なツールを紹介いただいた、北陸先端科学技術大学院大学 鳥澤健太郎助教授に深く感謝いたします。

## 参考文献

[1] T. Makino, K. Torisawa, and J. Tsujii. LiLFeS - Practical Programming Language For Typed Feature Structures. In *Natural Language Pacific Rim Symposium '97*, 1997.

[2] Y. Mitsuishi, K. Torisawa, and J. Tsujii. HPSG-Style Underspecified Japanese Grammar with Wide Coverage. In *COLING-ACL '98*, pp. 876-880, 1998.

[3] C. Pollard and I. A. Sag. *Head-driven Phrase Structure Grammar*. The University of Chicago Press, 1994.

[4] 斎藤憲. ユークリッド『原論』の成立. 東京大学出版会, 1997.

[5] 建石由佳, 鳥澤健太郎, 牧野貴樹, 西田健二, 洲上正睦, 辻井潤一. LTAG 文法からの変換による HPSG 文法の作成. 情報処理学会 研究報告 NL-122, pp. 119-126, 1997.

## 付録 — ギリシア語対応表

ギリシア語	品詞	意味 (性・数・格・ 法・相・時制・人称)
ἀναγεγράφω	動詞	(命令法・受動相・現在完了 3人称・単数) 描かれたとせよ
ἀπό	前置詞	(属格支配) ... の上に
γάρ	小辞	何となれば
ἐστὶ	動詞	(直説法・能動相・現在 3人称・単数) 英語の be 動詞にあたる
εὐθεία	形容詞	(女性・単数・主格) 真っ直ぐな (名詞化して) 直線
ἡ	冠詞	(女性・単数・主格)
ἴσον	形容詞	(中性・単数・主格) 等しい
ὀρθογώνιον	名詞	(中性・単数・主格) 長方形
περιεχόμενον	分詞	(受動相・現在 中性・単数・主格) 囲まれる
τετράγωνον	名詞	(中性・単数・主格) 正方形
τῆς	冠詞	(女性・単数・属格)
τό	冠詞	(中性・単数・主格)
τοῖς	冠詞	(中性・複数・与格)
τῶν	冠詞	(女性・複数・属格)
ὑπό	前置詞	(属格支配) ... によって (受動の動作主を示す)