

HPSGを用いた古典ギリシア語文法の拡張

中嶋 健一郎
北陸先端科学技術大学院大学
情報科学研究科
k-nakaji@jaist.ac.jp

齊藤 憲
大阪府立大学 総合科学部
人間科学研究科
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

本論文には、主辞駆動句構造文法 (HPSG)[1][5][2] に基づく古典ギリシア語文法について述べる。

ギリシア古典と呼ばれる文献の多くは原本が既に失われており、現在には写本だけが伝わっている。これまでで写字生などによって写本が繰り返されてきたという過程で、写し間違いや個人の解釈を追加したりなどの理由により、どこまでが原本なのか不明となってしまうという問題がある [6]。こういった事柄に対して現代人の我々が著者の言い分を正確に汲み取るために行えることは、それらの文献から少しでも多くの情報を取得し、それを用いて、できるだけ古代ギリシア人の考え方を客観的に研究することである。古典ギリシア語の構文情報を取り出すことは、こうした研究の一助となることができると考える。

そこで本研究では HPSG を用いて、古典ギリシア語を分析し、ユークリッドの『原論』を解析できるような文法を作成し実装した。その結果ユークリッド『原論』7章及び8章の合計 66 個の命題、1154 文に対して正解を含む構文木を 79.029% の割合で得ることができた。

2 HPSG について

本研究で用いた HPSG とは

- 文法の構成要素は全て、型付き素性構造と呼ばれるもので記述される
- 辞書記述に重点を置き、それらを結びつける規則は少ない
- 部分的な文法であっても解析が可能である [4]

- 文の解析は、ID-schema, principle と呼ばれる規則に基づいた型付き素性構造間の単一化によって行なわれる

という特徴を持つ構文解析のための文法枠組である。古典ギリシア語では性・数・格などの違いによって多数の属性を持ち、それらの結び付きによって文ができているという特徴を持つ。そのため、これらの属性を素性として、結び付きを HPSG における単一化によって文法を記述し、解析を行なうことで、表層情報から構文情報を取り出すことができる。

3 古典ギリシア語の特徴

古典ギリシア語は、文字の通り古代ギリシアで使われたものである。しかし使われていた期間が紀元前 1300 年頃から西暦 400 年頃までと非常に長期にわたり、その間には多くの方言が発生した。ここではその中でもユークリッドやプラトンの時代の標準語であったアッティカ方言に関していくつかの特徴を以下に挙げる。

3.1 語順の自由さ

文法規則の中で語順の制約を持つものは、いくつか存在するのみであり、語順に関する自由度はかなり高い。制約を持つものの例としては、大抵の前置詞はそれによって支配される語に先行して現れるなどといったものがある。またそうした規則すらいくつかの前置詞では順序が逆転したりすることも起こりうる。前節で示したような格変化の多様性のため、語順によって主語や補語がどれかということを示す必要性が薄く、それよりも話者の注目するものを先へ先へと持つてく

るという性質が見て取れる。

言語としての分類は、印欧語系として考えられており SOV 型言語であるといえる。しかし、その時点における話題の中でより重要なものが存在する場合、先行しているはずの主語や補語が、繰り下がる形で動詞の後ろに下がってしまうこともある。また動詞そのものが重要であると話者が考えている場合、動詞が先行し、主語や補語はそれに続く形で表される。そのみならず、通常の形で SOV と続いた後、動詞の後に主語や補語を修飾する語が続く場合も往々にして存在する。そのため、SOV 型であると同時に SVO 型と言うこともできる。

また、冠詞と名詞の間や、形容詞とそれによって修飾されるもの間には性・数・格の一致が、関係代名詞とそれによって修飾されるもの間には性・数の一致が求められる。これによって、語と語の間が著しく離れるようなことや、係り受けに交差があっても、修飾関係が聞き手に理解できると考えられている。

3.2 自明なものの省略

古典ギリシア語では、その時に話題になっているものや直前に会話に出現したものなど自明であるとされるものは省略されることがありうる。ここでいう自明なものとは主語、動詞他にもさまざまな要素が含まれる。例として、

μετρεῖτω, καὶ ἐστω ὁ Z
V, conj. V NP-nom.

という文がある。これは英訳されたものを参考にすると

“Let such a number measure them, and let
it be Z”

という意味の文である。しかし、実際には measure, and is the Z としか書かれていない。この文の主語は直前の文と同一であり、補語は直前の文で語っていたものを指している。本来この文は直前の文の追加的な説明をしているはずなのである。しかしどこにもそのような痕跡は存在しない。また、動詞が第 1 人称、第 2 人称の時には、語の形から自明であるので主語をはっきりと述べることは少ない。このような省略現象は主語だけでなく他の要素にも同様に起こりうる。

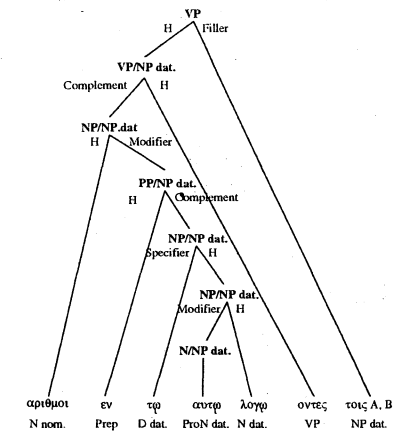


図 1 形容詞的性質を持つ語が補語をとる例

3.3 形容詞による格支配

この基本的な性質とは別に、一部の形容詞は英語における with 節によって修飾されるように、形容詞が、程度や時を表すための与格を支配することがある。これは基本の格とは別に本来は具格、処格だったものを与格、奪格だったものを属格が代用するようになった経緯からだと考えられる。例えば、次のような文がある。

εἰ γὰρ μὴ, ἔσσονται τινες τῶν Α, Β ἐλασσόνες
ἀριθμοὶ ἐν τῷ αὐτῷ λόγῳ ὄντες τοὺς Α, Β

この文において、最後に出現する強意代名詞 αὐτῷ は形容詞的に使われており、『〜と同じ』と意味となっている。そのため補語として与格の τοὺς Α, Β を取る。よって図 1 のような木構造が正しいと考えられる。他にも、比較表現の場合に比較対象を属格で表現することも行われる。

3.4 後置

一般に自然言語では係り受け関係の交差は起こりえないといわれているが、動詞の後ろには、それまでに出現した句に関する補足事項を述べる場所であるという特性のため古典ギリシア語では交差が観察できる。例として、図 2 のような係り受けが見られる。

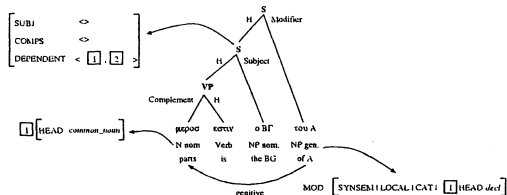


図 2 後方からの修飾

4 HPSG による古典ギリシア語文法

本研究では LiLFeS[3] を用いて HPSG に基づいた文法を記述した。我々が定義したものからいくつかを説明する。

4.1 語彙階層

HPSG では品詞のタイプなどによって語彙階層を定義することで、親タイプに制約を加えた場合、その子タイプにも同様の制約を課することができる。古典ギリシア語では、形容詞と名詞の区別があいまいなことや、冠詞が部分的に句を支配することがあるという特徴を持つため、当然英語文法などとは語彙階層は異なり、同様に各素性の型階層も異なる。

4.2 辞書記述

古典ギリシア語では、性・数・格が重要な情報であり、文中の同格関係、修飾関係をこれらの一致から得ることができる。上記の語彙階層における制約に従い各々の語の素性に値を与える。例として、形容詞の素性構造は次のようなものになる。

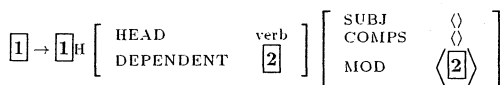
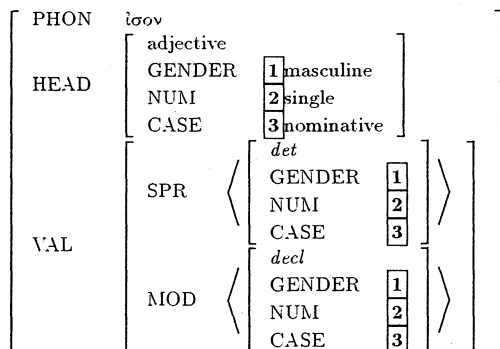


図 3 主語削除規則

四角で囲まれた数字はタグと呼ばれ、同じ数字の場所は同一のものをさしている。よって、この語は男性・単数・主格の語であり、性・数・格の一致する格変化するものを修飾し、性数格の一致する冠詞を取ることが記述されている。

4.3 ID-schema

主語と動詞句から文を作ることと規定する head subject schema, 補語を取ることができる句が補語と結びついて句を作ることと規定する head complement schema, 名詞が冠詞を支配し句を作ることと規定する head specifier schema, 修飾関係を規定する head modifier schema, 文の空所を埋めるための head filler schema のような他言語に見られるような規則から、主語削除規則や、後方からの修飾句規則 (図 3) など合計 16 個のスキーマを定義した。

図 3 のスキーマでは、飽和している動詞句であるなら、主語または補語として取ったものに対して、後方から修飾句がかかることができることを規定している。

4.4 実験結果

上記のスキーマと辞書記述 1352 を持つこの文法を、ユークリッド『原論』の 7 巻と 8 巻に対するカヴァレッジと解析時間を用いて評価した。『原論』の一文当たりの平均語長は 16.9 語となっている。出現する文のうち、表層情報からは曖昧性が高くなりすぎ、解析を行なうことのできなかった文が 7 章に 8 文、8 章に 7 文存在する。これらは失敗したものとした。カヴァレッジについては解析に成功する文の中で開始記号を得ることができる文の割合を用いて評価し、解析時間については解析に成功する文の一文当たりの平均解析時間を用いて評価する。実験は、athron 1800+ dual, memory 1.5GB 上で行なった。実験結果を表 1, 図 4 に示す。

表 1 本研究の文法による解析結果

	開始記号を得ることができた文の割合 (%)	解析時間 (sec)
Book 7	85.884	3.531
Book 8	69.439	6.665

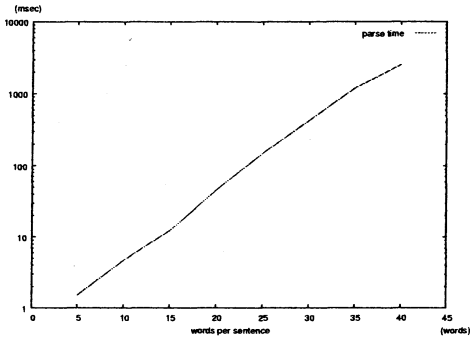


図 4 一文当たりの解析時間

解析時間、カヴァレッジともに、7章と8章の間に非常に差が見られる。7章に関してはどちらもかなりよい結果を得ることができたが、それに対して、8章では開始記号を得ることができない文が目立った。

5 考察

解析時間については、平均語長が 16.9 語ということもあり比較的長めの結果となった。しかし、一文当たり 30 語程度までならば実用範囲だと考えている。8章には格支配を行なうことがある形容詞が多数出現し、結果としてあいまい性を増大させたため、解析時間が7章と比較して倍増した。

カヴァレッジについては、文法規則を追加、改良することで更に広くすることができると思う。開始記号を得ることができないにも関わらず解析に成功してしまう主な原因は、後方からの同格関係、修飾関係の制約がやさしすぎることである。特に、同格関係では、表層的には格の一致以外に参考にできる情報は存在しないことが問題である。

また、本研究では、主観をできるだけ取り除くために、意味情報をできるだけ用いず表層情報のみを用い

て解析を行うような文法を記述した。そのため、あいまい性という点において問題が残る。

6 まとめと今後の課題

本研究では、HPSGに基づき古典ギリシア語文法を作成した。これによりユークリッド『原論』の7巻、8巻に対し正解を含む構文木を 79.029%の割合で得ることができた。

今後の課題として、この文法に語の持つ意味情報、文脈にある論理性などから、正しい係り受けや、主語と補語の区別を判断すること、辞書記述の追加改良を更に加えることなどがある。これらにより部分的なあいまい性を減少させることができ、結果的に文全体の曖昧性を減少させることができると考えられる。

参考文献

- [1] Carl Pollard, Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, 94.
- [2] Ivan A. Sag, Thomas Wasow. *Syntactic Theory*. The University of Chicago Press, 1999.
- [3] Takaki Makino, Kentaro Torisawa, and Jun'ichi Tsujii. LiLFeS-towards a practical HPSG parser. In *COLING-ACL '98*, 1998.
- [4] Yutaka Mitsuishi, Kentaro Torisawa, and Jun'ichi Tsujii. HPSG-style underspecified japanese grammar with wide coverage. In *COLING-ACL '98*, 1998.
- [5] 郡司隆男. 制約にもとづく統語論と意味論. 産業図書, 1994.
- [6] 齊藤憲. ユークリッド『原論』の成立. 東京大学出版会, 1997.