

Title	HPSGによるユークリッド『原論』の構文木の作成
Author(s)	仙田, 圭介
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1538
Rights	
Description	Supervisor:東条 敏, 情報科学研究科, 修士

HPSG によるユークリッド『原論』の構文木の作成

仙田 圭介 (010062)

北陸先端科学技術大学院大学 情報科学研究科

2002年2月15日

キーワード: HPSG, 構文解析, LiLFeS, 古代ギリシア語.

この論文では, 我々が開発した主辞駆動句構造文法 (HPSG) による古代ギリシア語のための文法について述べる. その目的は古典の研究に役立てることである.

古典の研究では, 対象となるテキストの原本を使うことはまれで, 写本を用いることが多いが, これらの真贋を判断することは難しい. 判断のためには対象テキストだけでなく, 同著者の別著作や他の文献を参照する必要がある. そのため, 扱うテキストは膨大な数に上り, 写本・原本間の違いも増える. また, 写本は写字生が写し間違いを起こすため, その表現はあいまいになる. 加えて, 改訂が行われると, 原本が書き換えられるかもしれない. そのため, 重要な情報が重要でない情報によって隠されてしまう.

また, 古典の電子化による保存はテキストのみを保存することが主流である. これらのテキストは検索の容易さなどから活躍しているが, それに付随した情報を加えることで, より利便性の高い文書となることがきる. 構文構造はそのような情報の一つであり, 写本間の違いを知るのに使えると考えられる.

そして, 古代ギリシアの数学者, ユークリッドによって紀元前 300 年ころに著されたユークリッド『原論』はこれらの古典の 1 つである. ユークリッド『原論』には上で述べた問題以外に命題の参照に関する問題がある. ユークリッド『原論』は以前に現れた命題を参照するのに命題の番号を用いることをせず, 代わりにその命題の表現を繰り返す. そのため, 参照された命題を探することは難しい. 我々は構文木の情報を使うことでこの問題を解決することができると考えた. これらのことからユークリッド『原論』の構文解析を行うこととした.

一方, 構文解析のための代表的な文法の一つに正規文法や文脈自由文法に代表される句構造文法がある. しかし, 句構造文法は意味情報を得ることが困難であるため, 単一化文法を用いた. 単一化文法は構文構造と意味構造を同時に得ることができる. 我々の目的は古典に情報を与えることであるため, 意味情報を加えることは有用である. そして, 我々は単一化文法の中でも HPSG に注目した. HPSG の特徴は語彙エントリが多くの情報を持ち, ID-schema と principle と呼ばれる文法規則が少ないことである. この特徴は多くの語の活用を持つギリシア語に向いている. 例えば, 名詞は性・数・格, 動詞は相・法・

人称・数そして時制を持つ．CFG を用いたならば，これらの活用のために多くの書き換え規則を書かなければならない．HPSG ならば語彙エントリを加えることで修正できる．これは書き換え規則を加えることより簡単であると言える．なぜなら，書き換え規則を加えることは他の文に誤りを引き起こすかもしれない．さらに，ユークリッド『原論』に現れる語の種類は 300 から 500 と少ない．この点においても我々は HPSG がより向いていると考える．

そこで，本研究ではユークリッド『原論』のための構文木を作成するシステムを実装し，このシステムの有用性を考察した．このシステムは LiLFeS で実装されている．HPSG の語彙エントリと文法規則は型付き素性構造と呼ばれるもので表現されているが，LiLFeS はその型付き素性構造を扱えるプログラミング言語である．LiLFeS は型付き素性構造の効率的な操作のために開発されており，その上では適用範囲の広い英語と日本語文法のための HPSG パーサが存在する．

最初に，我々はギリシア語のために素性構造を拡張した．この拡張は主にギリシア語の活用のためである．こうして，ユークリッド『原論』のすべての語は型付き素性構造で記述された．次に，それに基づき ID-schema を拡張した．ギリシア語では語が格情報を持つため，その順序は比較的自由である．そのため，素性構造に格情報を入れることが解析を可能にさせる．

しかし，これらのみではユークリッド『原論』の解析には不完全である．なぜなら，ギリシア語には particle と呼ばれる語がある．これは接続詞や副詞のように働くが，語の構成素の 2 番目に現れるという変わった振る舞いをするため，文の構造を壊してしまう．この問題を構文解析で解決するのは困難である．そこで，我々は前処理を行うこととした．結果として，前処理は particle を移動させることで，文の解析が可能となった．前処理で行った他の処理を以下に述べる：綴り，コンマなどの修正，冠詞と他の語との順序の修正，そして省略語の補完である．なお，補完の一部は手作業で行った．

上で述べたように，我々は前処理と古代ギリシア語文法を実装した．前処理は様々な修正と補完で解析を補っている．そして，古代ギリシア語のために語彙エントリ，ID-schema を拡張した．ID-schema の拡張により，古代ギリシア語の基本的な文法が解析できるようになった．なお，語彙エントリ数は 229 個，ID-schema 数は 10 個である．その結果，この文法はユークリッド『原論』第 2 巻命題 2 から命題 5 までの 76 文中，75 文において，正解の木を含む解析木を生成することができた．