JAIST Repository
https://dspace.jaist.ac.jp/

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

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修 士 論 文

HPSG によるユークリッド『原論』の構文木の作成

指導教官　東条 敏 教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

仙田圭介

2002 年 2 月 15 日

Copyright © 2002 by Keisuke Senda

要 旨

本論文には，主辞駆動句構造文法（HPSG）に基づく古代ギリシア語に対する文法について述べられている．その目的は古典に関する研究を行うとき，テキストの表象のみでなく，そこから得られる他の情報を研究の一材料として使うことである．

古典の研究では，使用される資料は原本ではなく，写本である場合が多い．これらの真贋はテキストの表現について他の著作，著者を調べることで判断する．テキスト情報に加え，構文構造を知ることはその判断の一助となると考えられる．そのような古典の 1 つである古代ギリシア語で著されたユークリッド『原論』には命題を参照することについての問題が存在する．ユークリッド『原論』では，以前に現れた命題を参照するとき，命題に振られた番号ではなく，その命題の表現を繰り返す．このため，参照元を探すのは困難である．このような問題の解決に構文木の情報を使用することが可能であると考えた．

古代ギリシア語は名詞句は性・数・格，動詞が相・法・人称・数・時称を持つ，というように各単語の持つ情報が多い．このことから，構文解析に用いる文法として，単語に多くの情報を持たせることで文法規則数が少なくてすむという特徴をもつ HPSG が適していると考えた．

そこで本研究では，HPSG を用いてユークリッド『原論』の構文木を作成するシステムを実装した．実装には HPSG で使われる型付き素性構造を記述することのできる論理型プログラミング言語，LiLFeS を用いた．LiLFeS により，古代ギリシア語のための文法を記述し，語順の自由さ，遠隔依存などに代表される古代ギリシア語の文法のために HPSG の辞書記述，ID-schema を拡張し，その解析を可能とした．また，ギリシア語特有の表現である particle についても構文解析前に入力文に対し処理を行うことで，解析を可能とした．

以上のようにして作成した文法は 229 個の辞書記述と 10 個の ID-schema を持つ．これらを用いることで，ユークリッド『原論』 第 2 巻命題 2 から命題 5 までの全 76 文中，75 文について正解の木を含む解析木を生成することが可能となった．

目 次

1 はじめに 1

1.1 研究の背景と目的 1

1.2 本論文の構成 3

2 古代ギリシア語 4

2.1 古代ギリシア語の特徴 4

2.1.1 語の活用 4

2.1.2 語順の自由さ 5

2.1.3 語の省略 5

2.1.4 遠隔依存の問題 7

2.2 particle 8

2.2.1 particle の性質 8

2.2.2 particle の引き起こす省略 11

3 HPSG 概論 12

3.1 HPSG とは 12

3.1.1 型付き素性構造 12

3.1.2 型階層 15

3.1.3 単一化 17

3.1.4 辞書記述 17

3.1.5 ID-schema，principle 19

4 HPSG による古代ギリシア語のための構文解析システム 20

4.1 処理の流れ 20

4.2 ギリシア文字から ASCII 文字への変換 20
4.3 前処理 22
4.3.1 アクセント，気息記号，つづりの修正 22
4.3.2 コンマの修正 24
4.3.3 冠詞の後ろに続く語の修正 25
4.3.4 particle の位置の修正 26
4.3.5 省略された語の補完 27
4.4 構文解析 28
4.4.1 LiLFeS 28
4.4.2 素性 29
4.4.3 辞書記述 31
4.4.4 schema 32
4.5 解析例 37

5 考察 45
5.1 前処理について 46
5.2 辞書記述と ID-schema 47
5.3 省略された語の補完 47

6 おわりに 49
6.1 まとめ 49
6.2 今後の課題 50

A ギリシア語対応表 54

第 1 章

はじめに

1.1　研究の背景と目的

古代の書物から歴史を知ろうとする研究が様々な分野で行われている．これらの研究で使用される資料（古典）は碑文などの特殊な例でない限り，その時代に著された原本ではなく，後世に他の人の手によって書き写された写本である場合が多い [9]．しかし，写本では写字生の写し違いは避けられず，そこで書き写された文字が明瞭に書かれているとも限らない．また，意味的に不明瞭である点に注釈を加えるなど，より恣意的に変更を行う場合もある．このように様々な形に写本は派生していくが，互いが互いを補う形でより新たな写本が生まれる場合もあり，写本間の関係を系統立てるのみでなく，底本にどのような表現が用いられていたのかについても知ることは困難である．これらの推定はテキストの表現の異同を同著者の別の著作，別著者の著作などから判断する必要があるが，それらのテキストは膨大なものに上るうえ，不確かな写本を元に推定せざるを得ない部分が少なからず存在する．テキストの表象のみからではこれらの判断は困難である．文章を書くときには，同じ意味を表す文を書いてもその書き手によって用いる語，語順の並びなどは異なってくるが，テキストの表象のみでは二つの写本間に同じ文であるべき部分が異なっている個所が複数あったとしても，そこに何らかの関連を見つけ出すことは困難なためである．

一方，現在まで古典を保存するための電子化では，テキストの表象のみを保存することが主流である．語の検索の容易さなどでそれらの電子化されたテキストは活躍しているが，さらにそれに付随した情報を付け加えることで，電子化されたテキストはより利

便性の高い文書となると考えられる．その付加する情報の一つに構文構造があげられる．上で述べた写本間での差異についても，構文構造という情報があることで，二つの写本間の異なりの性質が詳細に見え，文同士，ひいては写本同士の関連を見出すことができるのではないかと考えられる．

以上のように古典にはテキストの電子化のみではなく，それに付随した理解を助けるような情報も加えるべきだと考えるが，今回対象とした古典，ユークリッド『原論』[2] においては構文構造はより重要な意味を持つと考えられる．

ユークリッド『原論』は古代ギリシア時代の数学者，ユークリッドの著作と考えられている．その内容は命題，定理とそれに対する証明が次々と述べられているものである．当然，命題の証明にはそれ以前に証明した命題が用いられることがある．現代ならば，命題の参照には各命題に割り振られた命題番号を用いることで,「命題３より…」などのような表現が使われる．しかし，ユークリッド『原論』が著された時代には命題番号を付ける習慣は存在しなかった．そのため，以前に証明した命題などを参照するときにはその証明した命題の文句を書き並べる事を行う．これは回りくどいだけでなく，参照元を捜し求めるのが困難であるという問題も発生させた．参照元とそれに対応する参照先が同じ文であるという保証はないため，単純なテキスト検索ではこの問題は解決できない．この命題参照の問題についても構文構造を用いることで解決が可能であると考える．

一方，構文構造を得るために用いられる文法の代表的なものに，正規文法や文脈自由文法に代表される句構造文法や単一化文法がある．このうち，句構造文法では文法的構造を解析できたとしても，次へのステップとして意味的構造を得ようとするのは困難である．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．古典の解析では意味情報を得ることは有用であるといえる．本研究ではこの単一化文法の一種である主辞駆動句構造文法（Head-driven Phrase Structure Grammar，以下，HPSG）[7, 8] に注目した．

HPSG は単語・句・文が多くの辞書情報をもつ一方で，構文解析を行い，構文木を生成するための規則は少ないという特長を持つ．ユークリッド『原論』が表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を用いてギリシア数学の古典，ユークリッド『原論』の構文木を作成するシステムを実装し，ユークリッド『原論』，ひいては古代ギリシア語における HPSG の有用性について考察する．

1.2 本論文の構成

本論文では2章において，ユークリッド『原論』が著されている古代ギリシア語についてその特徴と問題点について述べ，3章で HPSG について説明を行う．

4章では本研究で実装した HPSG に基づく ユークリッド『原論』のための文法と，構文解析に付随する処理について説明した後，構文解析器による解析例を示す．

最後に，5章において実装したシステムに関する考察を行う．

第 2 章

古代ギリシア語

ユークリッド『原論』は紀元前 300 年頃に古代ギリシアの数学者ユークリッドによって著された全 13 巻からなる数学文献であり，西欧では教科書として 19 世紀まで使われつづけた．その原本は古代ギリシア語で書かれているため，本章ではギリシア語とその特徴ににについて述べる．

以下で説明される文法，単語の意味・用法は [3, 10] などを参考にした．また，本論文中に現れるギリシア語については，それに対応する日本語を付録 A に示す．

2.1 古代ギリシア語の特徴

2.1.1 語の活用

古代ギリシア語の英語などと大きく異なる点として，語の活用の多さが挙げられる．

名詞，形容詞，冠詞などは性・数・格の情報を持ち，例えば，冠詞 + 名詞からなる名詞句では性・数・格が一致しなければならない．また，前置詞は格支配を行う．複数の格を支配できる前置詞も存在するが，そのときには支配している格によって前置詞の意味が異なる．

また，動詞は相，法，人称，数，時称に従い変化する．主語と動詞との間では人称と数が一致する必要がある．

2.1.2　語順の自由さ

直線 AB は，以下のどちらでも記述することができる．

εὐθεῖα ἡ AB

ἡ AB εὐθεῖα

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．

例えば,「AE は AZ，GE に等しい」という意味の文は以下のように書かれている．

(Ἴσον ἐστι) (τὸ AE) (τοῖς AZ, ΓE).

この文は上の括弧で括られたまとまりごとに VSO の順に並んでおり，図 2.1 のように構文木が作成される．

ギリシア語の場合，各名詞が格の情報をもつため，このような構文が可能となる．そのため，上の例文とまったく同じ意味の文として例えば以下のように書くことも可能である．

Τὸ AE ἴσον ἐστι τοῖς AZ, ΓE.

Τὸ AE τοῖς AZ, ΓE ἴσον ἐστι.

これらはそれぞれ，SVO，SOV の語順である．

2.1.3　語の省略

ユークリッド『原論』では直線 AB，正方形 AΔEB といった語が頻出する．これらの語が現れるとき，‘εὐθεῖα ἡ AB’，‘τὸ AΔEB τετράγωνον’ のように書かれない個所が少なからず存在する．

ギリシア語の冠詞には名詞と同じように性・数・格の情報が含まれている．そのため，名詞とそれに対応する冠詞，形容詞の間では性・数・格の一致がされなければならない．このことを利用して語の省略がなされる．

例えば，直線 AB では以下のようにかかれることが多い．

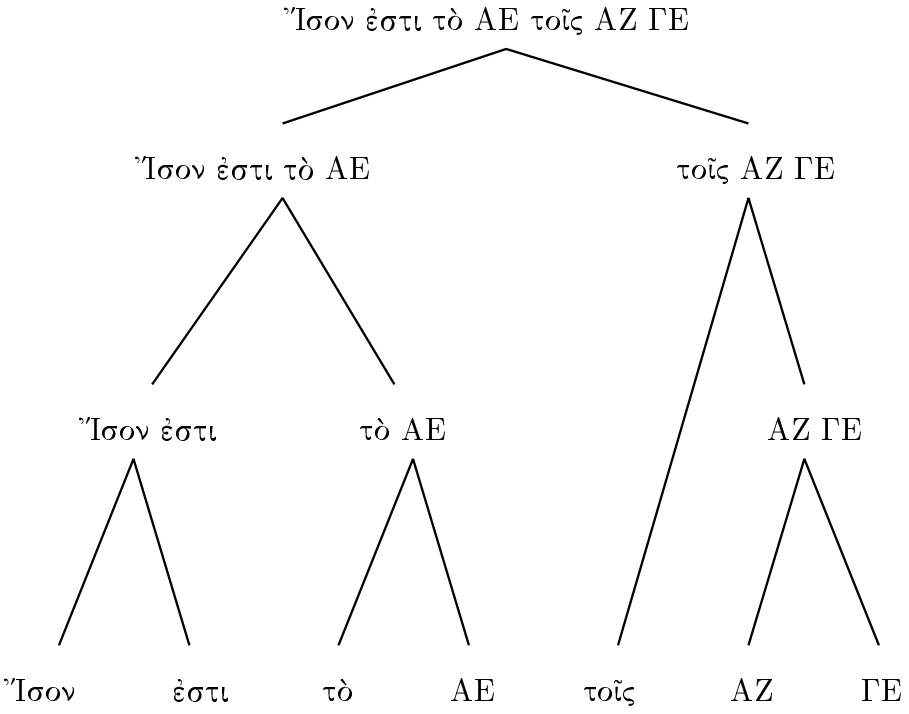

図 2.1: “Ἴσον ἐστι τὸ ΑΕ τοῖς ΑΖ, ΓΕ.” の解析

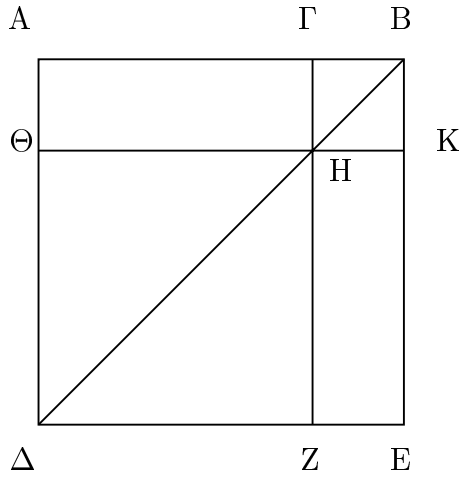

図 2.2: ユークリッド『原論』に表れる図の例（第 2 巻命題 4 より）

ἡ AB

上の句において，‘ἡ’ は女性・単数・主格の冠詞である．この句では名詞は存在しないが，この冠詞があるために省略された名詞は性・数・格の一致を必要とすることから，女性・単数・主格のものであることがわかる．また，冠詞に続く語は AB と 2 文字の記号で表されているため，線や四角形といったような 2 文字の記号で表現できるような幾何学の言葉と推定できる．

よって，女性名詞であり，2 文字の記号で表現できるような語は直線が存在するのみのため，直線を表す単数・主格の語の ‘εὐθεῖα’ が省略されているとわかる．

図 2.2 で表されるこのような語の例を表 2.1 に挙げる．

また，語の省略はこのような 1 つの句の中だけにとどまらず，1 文中の構成素の 1 つが省略されてしまう場合も存在する．これについては particle（小辞）と呼ばれる語が関連することが多いため，2.2 で説明する．

2.1.4 遠隔依存の問題

「(直線)AB, BΓ 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．

表 2.1: 省略される語の例（すべて単数・主格の場合）

名詞を省略されたときの表現	省略された名詞	意味
ἡ AB	εὐθεῖα	直線 AB
τὸ AΔEB	τετράγωνον (ὀρθογώνιον)	正方形（四角形）AΔEB
τὸ AE	τετράγωνον (ὀρθογώνιον)	正方形（四角形）AE 対角線で表されている
τὸ ἀπὸ τῆς AB	τετράγωνον	直線 AB 上の正方形
τὸ AΔB	τρίγωνον	三角形 AΔB
ἡ ὑπὸ AΔB	γωνία	角 AΔB

τὸ ὑπὸ τῶν AB, BΓ 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．これを解析した結果を図 2.3 に示す．

また，このような句も 2.1.3 に示すような省略が用いられ，以下のように記される場合もある．

τὸ ὑπὸ τῶν AB, BΓ

2.2 particle

2.2.1 particle の性質

古代ギリシア語には particle(小辞) と呼ばれる特殊な語がある．この語は副詞，もしくは接続詞のような働きを行うが，他の語の構成する文法構造にかかわらず，常に句の構成要素の 2 番目に現れるため，文の解析を行う上での妨げとなる．

以下の 2 つの文を考える．

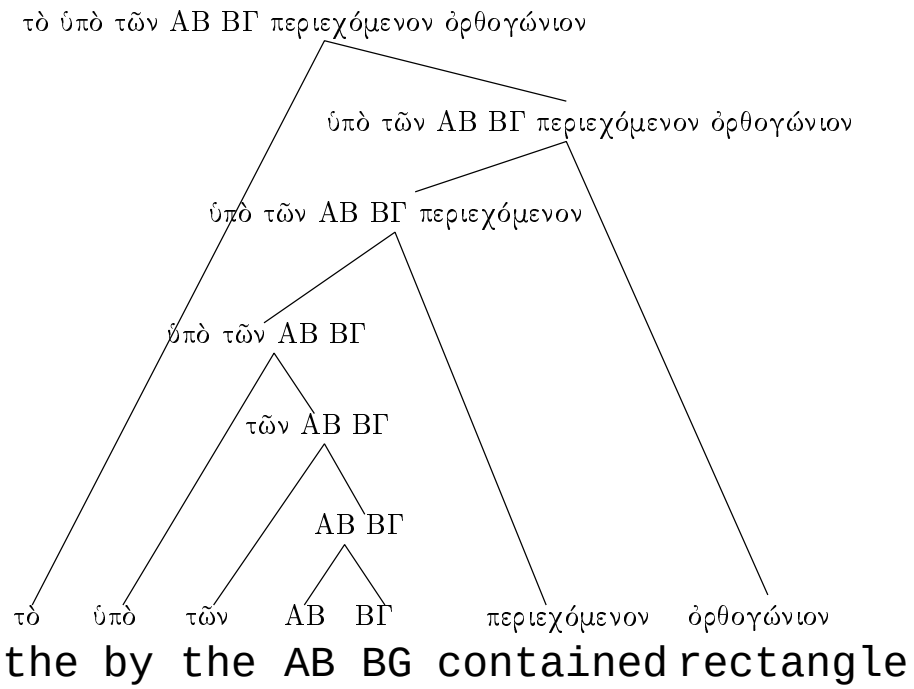

図 2.3: ‘τὸ ὑπὸ τῶν ΑΒ, ΒΓ περιεχόμενον ὀρθογώνιον’ の解析

Ἀναγεγράφθω ἀπὸ τῆς ΑΒ τετράγωνον τὸ ΑΔΕΒ.

（直線）AB 上に正方形 AΔEB が描かれたとする．

καὶ ἡ ὑπὸ ΓΗΒ γωνία τῇ ὑπὸ ΗΒΓ ἐστιν ἴση

そして角 ΓHB は（角）HBΓ に等しい

例えば，上の文は，図 2.4 のように解析できる．

しかし，実際には以下のようにそれぞれ「何となれば」,「したがって」という意味を表す particle（下線部）が本文中には存在する．

Ἀναγεγράφθω <u>γὰρ</u> ἀπὸ τῆς ΑΒ τετράγωνον τὸ ΑΔΕΒ

καὶ ἡ ὑπὸ ΓΗΒ <u>ἄρα</u> γωνία τῇ ὑπὸ ΗΒΓ ἐστιν ἴση

上の文のように，particle はすでに文法的に文として完成されている文に入り，文全体にかかる．ボトムアップに解析すると，図 2.5 のように particle により解析木の作成に支障をきたしてしまう．また，2 文目を見るとわかるように，particle は常に 2 番目の単語として現れるわけではなく，文によってその位置は不安定である．

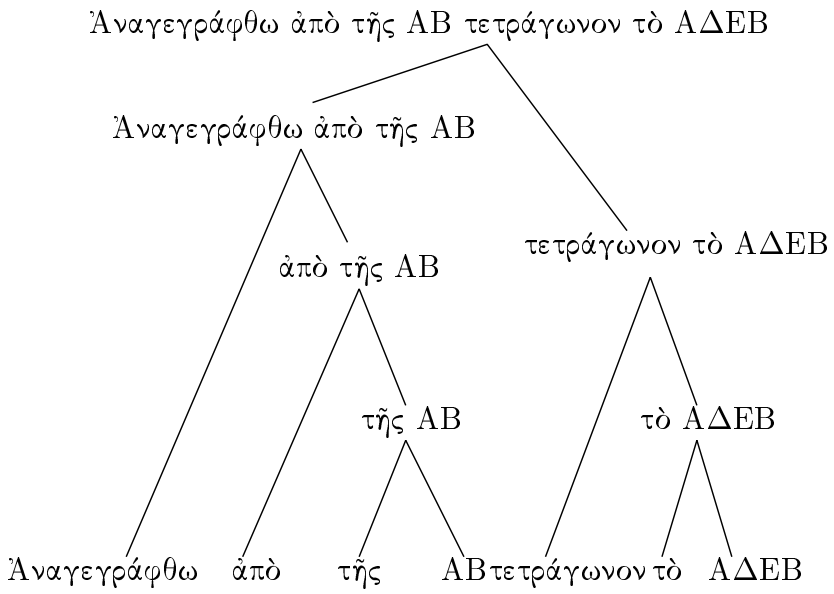

図 2.4: “Ἀναγεγράφθω ἀπὸ τῆς ΑΒ τετράγωνον τὸ ΑΔΕΒ.” の解析

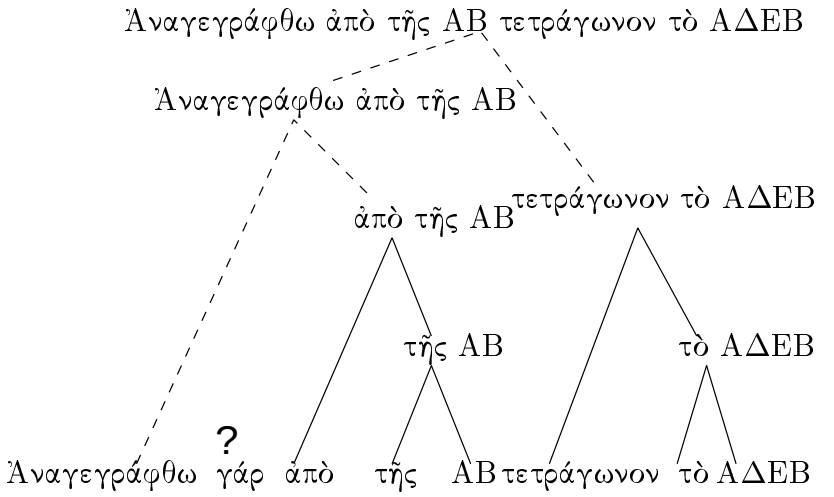

図 2.5: “Ἀναγεγράφθω γὰρ ἀπὸ τῆς ΑΒ τετράγωνον τὸ ΑΔΕΒ.” の解析

2.2.2 particle の引き起こす省略

また，前節の例でもあるように，particle は前の文からのつながりをあらわすときに使われることが多い．このようなときには前の文に現れた単語が省略されて文が表されることがある．

> καὶ ἐπεὶ ἴσον ἐστὶ τὸ ΑΗ τῷ ΗΕ, καί ἐστι τὸ ΑΗ τὸ ὑπὸ τῶν ΑΓ, ΓΒ·
>
> ἴση γὰρ ἡ ΗΓ τῇ ΓΒ.

上の文章を日本語で表すと，以下のようになる．

> そして,（長方形）AH は（長方形）HE に等しく，AH は AΓ, ΓB による長方形である．
>
> 何となれば，HΓ が ΓB に等しいからである．

上のギリシア語で表された文章の場合，2 文目には重要な文の構成素の一つである動詞が存在しないが，2 文目に「何となれば」という意味を持つ ‘γὰρ’ という particle が存在している．これが前の文を受けて 2 文目があることを示し，2 文目の省略されている動詞が 1 文目と同じ（ここでは ‘ἐστὶ’）であることがわかる．

第 3 章

HPSG 概論

3.1 HPSG とは

HPSG(Head-driven Phrase Structure Grammar)[7, 8] は，主辞が重要な役割を果たす単一化文法の一種であり，以下のように述べることができる．

- 文法は，ID-schema，principle と呼ばれる文法規則と辞書記述からなる．
- 文法の構成要素はすべて型付き素性構造 [1] と呼ばれるもので記述される．
- ルールには一般的な性質しか書かかず，詳細な情報は辞書記述に書くことで，ルールの数をできる限り減らす．
- 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．

3.1.1 型付き素性構造

型付き素性構造とは，有向グラフの各ノードに型と呼ばれるラベル，各アークに素性と呼ばれるラベルがふられているものである．例として *she* という単語の型付素性構造を図 3.1 に示す．この図では，型名をイタリックで表し，素性名を大文字で表している．

左端にある矢印がルートを表し，その型は *word* である．型には 2 つの素性 (PHON, SYNSEM) が存在し，PHON 素性の型は *nelist* である．

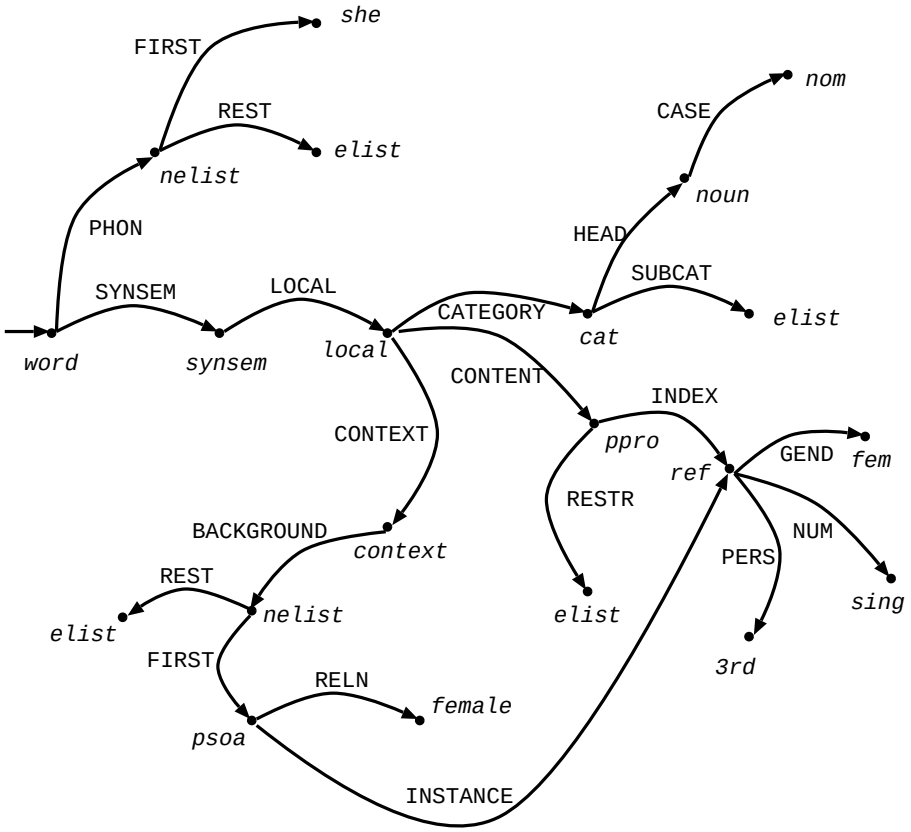

図 3.1: 型付き素性構造の例

$$
\begin{bmatrix}
\textit{word} & & &

$$
\begin{bmatrix}
\textit{word} & \\
\text{PHON} & \langle \textit{she} \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\textit{local} & \\
\text{CATEGORY} & \begin{bmatrix} \textit{cat} & \\ \text{HEAD} \mid \text{CASE} & \textit{nom} \\ \text{SUBCAT} & \langle\rangle \end{bmatrix} \\
\text{CONTENT} & \begin{bmatrix} \textit{ppro} & \\ \text{INDEX} \ \boxed{1} & \begin{bmatrix} \textit{ref} & \\ \text{PER} & \textit{3rd} \\ \text{NUM} & \textit{sing} \\ \text{GEND} & \textit{fem} \end{bmatrix} \\ \text{RESTR} & \langle\rangle \end{bmatrix} \\
\text{CONTEXT} \mid \text{BACKGROUND} & \left\langle \begin{bmatrix} \textit{psoa} & \\ \text{RELN} & \textit{female} \\ \text{INST} & \boxed{1} \end{bmatrix} \right\rangle
\end{bmatrix}
\end{bmatrix}
$$

図 3.3: AVM の省略表記の例

囲まれた数字がそれである．

また，有向グラフ上で *nelist*，*elist* で表されていたリストは，AVM では省略表記として ‘< >’ で表されている．省略表記としては他に，値が特に入っていないような素性はその素性を省略するというものがある．このような素性にはボトムと呼ばれる特殊な型が入っている．このため，ある型の素性が 1 つしかない場合は AVM においてはボックスを書かずに素性と素性の間に記号 “|” を用いる．これを用いて図 3.2 を書き換えたのが図 3.3 である．

3.1.2 型階層

型には階層構造が与えられ，この型階層では階層が上になるほど詳細な情報が表されている．この型階層から各型の最小上界を知ることができる．

最小上界とは，複数の型が与えられたとき，その複数の型より詳細で最も一般的な型のことをいう．例えば，図 3.4 において，“diamond” と “rectangle” の最小上界は “square

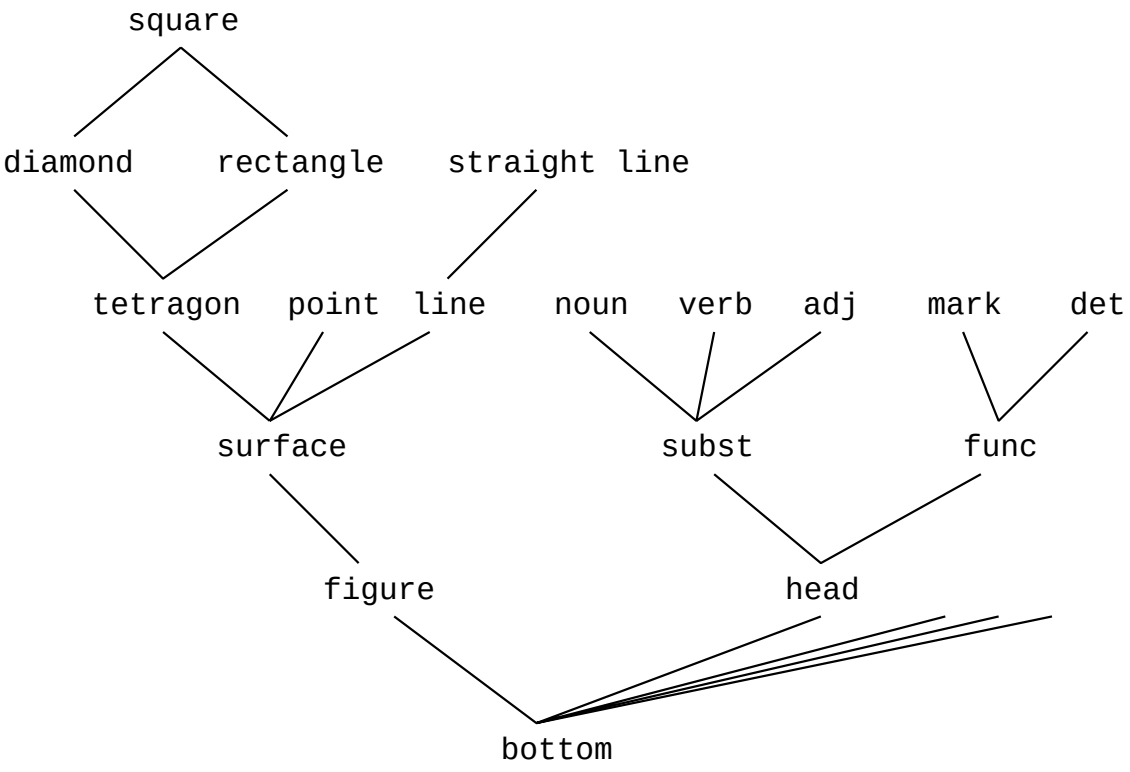
square
diamond
rectangle
straight line
tetragon
point
line
noun
verb
adj
mark
det
surface
subst
func
figure
head
bottom

図 3.4: 型階層の例

であり”，“head” と “subst” の最小上界は “subst” である．また，“surface” と “func” の最小上界は存在しないことになる．

3.1.3 単一化

HPSG では文の解析・生成は素性構造間の単一化によって行われる．

単一化とは 2 つの型の間の最小上界を取ることをいう．

型付き素性構造間で単一化を行うときには，素性構造間の同一の素性に対して再帰的に単一化を行う．このときマージされたノードの型は元の 2 つの素性構造が持つ型の最小上界となる．

型付き素性構造間で単一化を行おうとしたとき，どこか一つの素性で単一化を行うことができなければ，その単一化は成功しないことになる．

HPSG において，単一化が成功しないということは，ひいては文の解析・生成ができなくなることにつながる．

3.1.4 辞書記述

辞書記述とは，各単語，句，文に与えられるものであり，多くの情報が型付き素性構造の形で記述される．

例えば動詞ならば，主語はどのような名詞をとるのか，目的語はいくつとるのか，などといった情報や人称や数，時制についても書かれる．

例として,「描かれたとせよ」という意味を表す古代ギリシア語，‘ἀναγεγράφθω’ という語の辞書記述を図 3.5 に示す．

この語は命令法・受動相・現在完了・3 人称・単数の動詞（SYNSEM | LOCAL | CONTENT | INDEX 内）であり，主語として主格の名詞（句）を取り（SYNSEM | LOCAL | CAT | VAL | SUBJ 内），その他に ἀπό という語を含む前置詞句を取る（SYNSEM | LOCAL | CAT | VAL | COMPS 内）ということが書かれている．

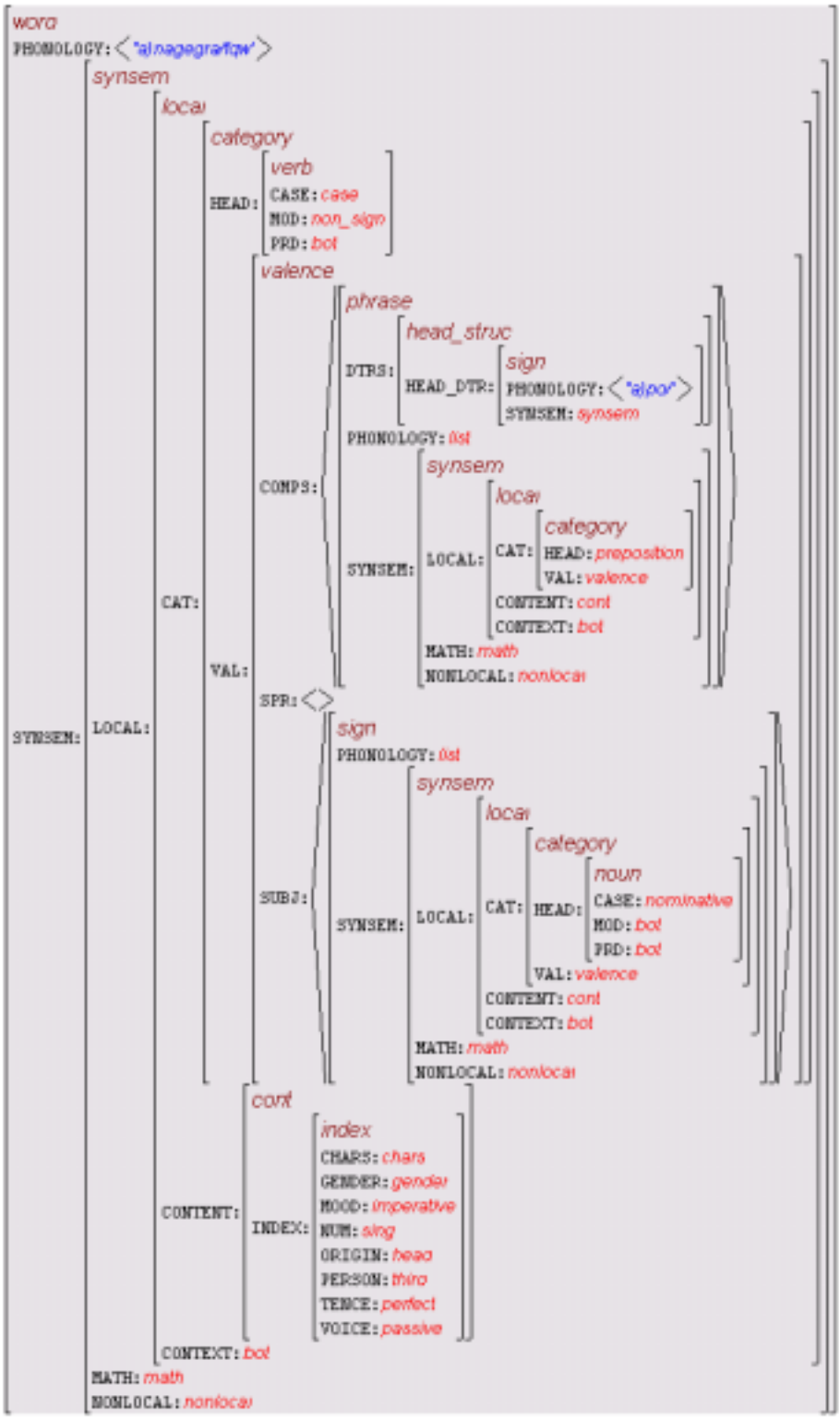

図 3.5: ‘ἀναγεγράφθω’ の辞書記述

3.1.5 ID-schema，principle

ID-schema，principle は文脈自由文法の書き換え規則にあたるものである．

ID-schema は子供となる 2 つの素性構造の単一化の位置を指定し，親の素性構造のうち，構文木を作るのにかかわる素性の形を決定する素性構造である．

principle はすべての親子間で守られなければならない制約のことをいう．例えば，主辞となる子の品詞は親に受け継がれ，親の品詞となるという principle (head feature principle) が存在する．

第 4 章

HPSG による古代ギリシア語のための構文解析システム

この章ではシステム全体を述べる．

4.1 処理の流れ

図 4.1 に全体の処理の流れを示す．
図にあるように，処理は大きく分けて以下の 3 つに分かれる．

- ギリシア文字から英字への変換
- 前処理
- LiLFeS による解析

次節よりそれぞれの処理について詳しく述べる．

4.2 ギリシア文字から ASCII 文字への変換

古代ギリシア語ではギリシア文字で表される．よって，これを ASCII 文字に変換する必要がある．この変換表を表 4.1 に示す．なお，この変換は The Thesaurus Linguae Graecae (TLG) [12] の Beta Code に従った．

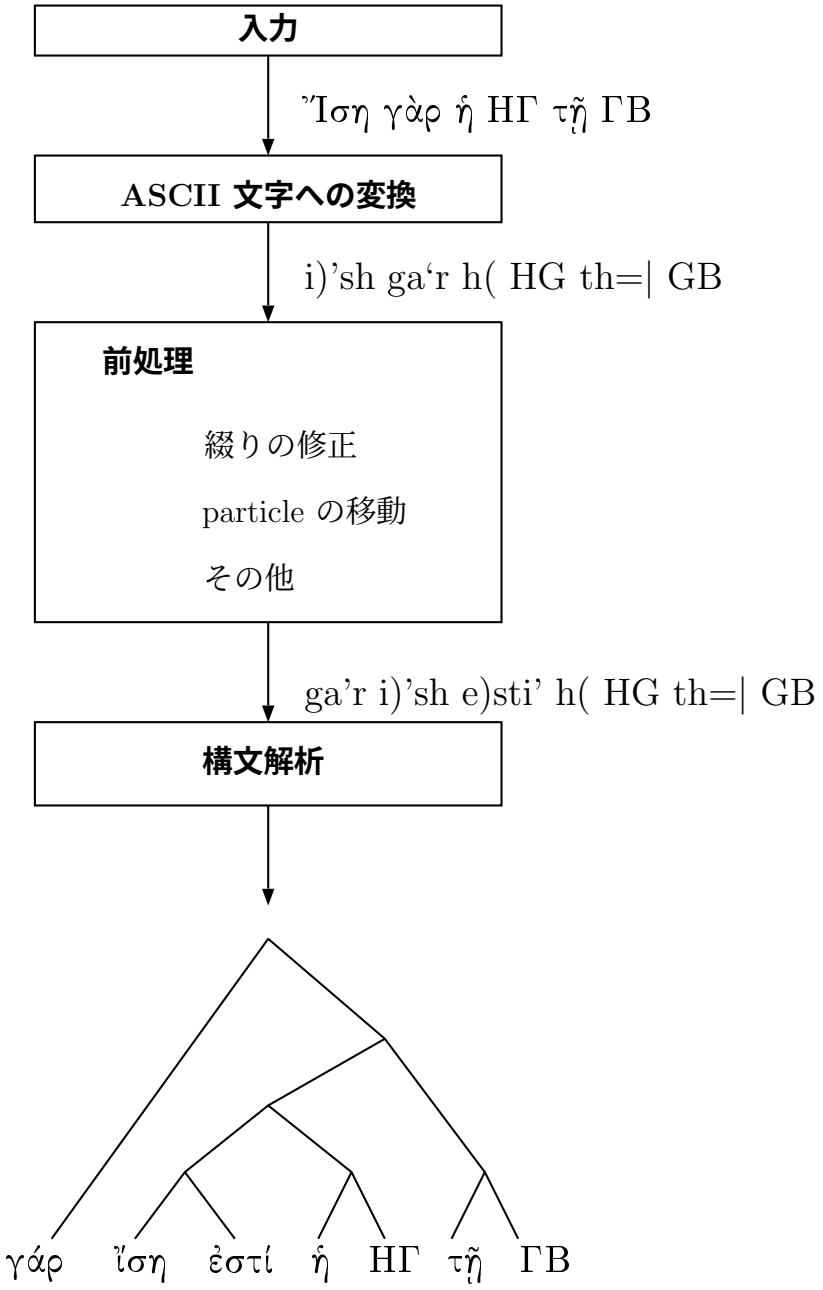

図 4.1: ユークリッド『原論』解析処理の流れ

なお，この ASCII 文字の作成時にユークリッド『原論』中に現れる図形を表す記号のみを大文字とし，その他の文字はすべて小文字表記に改めている．

4.3 前処理

ギリシア語には 2 章で挙げたような特徴が存在する．そのため，直接本文を解析すると，何らかの問題が生じることが多い．

そこで，構文解析を行う前に前処理を行う．なお，この前処理は特に断りがない限り，Perl を用いている．

前処理では以下のことを行う．

- アクセント，規則記号，つづりの修正
- ,（コンマ）の修正
- 冠詞の後ろに続く語の修正
- particle の位置の修正
- 省略された語の補完

4.3.1 アクセント，気息記号，つづりの修正

以下のような文を考える．

> ἀλλ’ ἡ ὑπὸ AΔB τῇ ὑπὸ ABΔ ἐστιν ἴση, ἐπεὶ καὶ πλευρὰ ἡ BA τῇ AΔ ἐστιν ἴση·
>
> しかし角 AΔB は角 ABΔ に等しい，なぜなら辺も BA が AΔ に等しいから．

ここで，一番最初の語 ‘ἀλλ’’ に注目する．‘ἀλλ’’ は英語の ‘but’ にあたる接続詞である．接続詞のため，活用は存在しない．

しかし，‘ἀλλ’’ はこのままの形では辞書に存在しない．

表 4.1: ギリシア文字–ASCII 文字対応表

ギリシア文字	ASCII 文字	読み	ギリシア文字	ASCII 文字	読み
Α α	a	alpha	Σ σ ς	s	sigma
Β β	b	beta	Τ τ	t	tau
Γ γ	g	gamma	Υ υ	u	upsilon
Δ δ	d	delta	Φ φ	f	phi
Ε ε	e	epsilon	Χ χ	x	chi
Ζ ζ	z	zeta	Ψ ψ	y	psi
Η η	h	eta	Ω ω	w	omega
Θ θ	q	theta	ἀ	a)	smooth breathing
Ι ι	i	iota	ἁ	a(	rough breathing
Κ κ	k	kappa	ά	a/	acute accent
Λ λ	l	lambda	ὰ	a\	grave accent
Μ μ	m	mu	ᾶ	a=	circumflex accent
Ν ν	n	nu	ᾳ	|	iota subscription
Ξ ξ	c	xi	,	,	comma
Ο ο	o	omicron	.	.	period
Π π	p	pi	·	;	raised dot
Ρ ρ	r	rho			

辞書に存在する形は ‘ἀλλά’ である．このように語のつづりが変わる理由は後ろの語が関係している．

この文中の ‘ἀλλ’’ の場合は後ろに続く語 ‘ἡ’ が母音[1]であるため，‘ἀλλά’ の一番最後の母音が取り除かれている．

この文中ではほかに ‘ὑπὸ’ は辞書には ‘ὑπό’ として掲載されている．この語は後ろにほかの語が続くために，アクセントのつき方がもとの形と異なっている．

これらの語は発音の仕方が異なってくるのみであり，文法的構造，意味的構造ともに変化しないため，前処理での修正を行う．このような語をすべてもとの形に変更させると，以下の文になる．

> ἀλλά ἡ ὑπό AΔB τῇ ὑπό ABΔ ἐστί ἴση, ἐπεί καί πλευρά ἡ BA τῇ AΔ ἐστί ἴση·

この語の修正は前もってつづりの異なる語の一覧が表されている辞書を用意しておき，その辞書とのパターンマッチングによって行っている．

4.3.2 コンマの修正

ユークリッド『原論』に現れるコンマには以下の 2 種類存在する．

- 文の区切りを表すコンマ
- 図形を示す記号の区切りを表すコンマ

このうち，文の区切りを表すコンマは構文木を作成するために重要な情報を提供するが，記号の区切りを表すコンマは読者の見易さのためにつけられているのみであり，文法上，大きな役割を果たしているわけではない．電子的に文を入力するときには記号間にはスペースが入り，コンマは意味をなさない．

したがって，図形を特定するために用いられる記号間の区切りを表すコンマは用いる必要はないと判断し，削除を行っている．

[1]ギリシア語の母音は α, ι, υ, ε, η, ο, ω の 7 つである

4.3.3 冠詞の後ろに続く語の修正

以下のような文を考える．

> λέγω, ὅτι τὸ ἀπὸ τῆς ΑΒ τετράγωνον ἴσον ἐστὶ τοῖς τε ἀπὸ τῶν ΑΓ, ΓΒ τετραγώνοις καὶ τῷ δὶς ὑπὸ τῶν ΑΓ, ΓΒ περιεχομένῳ ὀρθογωνίῳ.
>
> 私は言う，AB 上の正方形は AΓ, ΓB の上の正方形ども（の和）および AΓ, ΓB によって囲まれる長方形の 2 倍と（の和）に等しい．

この文には,「AΓ, ΓB の上の正方形ども」という句と「AΓ, ΓB によって囲まれる長方形の 2 倍」という句が存在している．これらはそれぞれギリシア語では以下のように書かれている．

> τοῖς τε ἀπὸ τῶν ΑΓ, ΓΒ τετραγώνοις
>
> τῷ δὶς ὑπὸ τῶν ΑΓ, ΓΒ περιεχομένῳ ὀρθογωνίῳ

それぞれの句において 2 番目に ‘τε’, ‘δὶς’ という句が現れる．それぞれこれらの句がなければ 2 章で示した正方形，長方形をあらわす句となる．すなわち，‘τε’, ‘δὶς’ はどちらも名詞句の中に通常の文法規則とは異なる規則で現れていることがわかる．

ここに現れる ‘τε’ は particle である．すなわち，particle の性質である,「句の構成要素の 2 番目に現れる」という規則にしたがっている．この語は英語の ‘and’ にあたる語 ‘καὶ’ という語と一組で英語の ‘both ... and ...’ の意味を表すため，先ほど挙げた句

> τοῖς ἀπὸ τῶν ΑΓ, ΓΒ τετραγώνοις

にかかることがわかる．

また，‘δὶς’ は「2 倍の」という意味の副詞だが，ここでは形容詞的に用いられており，同様に,

> τῷ ὑπὸ τῶν ΑΓ, ΓΒ περιεχομένῳ ὀρθογωνίῳ

にかかることがわかる．

そのため，この前処理の段階で ‘τε’, ‘δὶς’ の場所を各句の先頭に移動させる処理を行うこととした．

これにより，元の文は以下のように修正される．

λέγω, ὅτι τὸ ἀπὸ τῆς ΑΒ τετράγωνον ἴσον ἐστὶ τε τοῖς ἀπὸ τῶν ΑΓ, ΓΒ τετραγώνοις καὶ δὶς τῷ ὑπὸ τῶν ΑΓ, ΓΒ περιεχομένῳ ὀρθογωνίῳ.

この処理は冠詞の後ろに particle, 形容詞が現れたときにその前後の順序を入れ替えることで実現している．

4.3.4 particle の位置の修正

‘τε’ の一部と ‘τε’ 以外の particle は基本的に文として成立する句全体にかかるように用いられる．

したがって，これらの particle が現れたときにはコンマ，もしくはピリオドで区切られた句の中の先頭に移動させることで，この後に行われる解析を行いやすくしている．ただし，句の先頭が接続詞の場合，接続詞は後ろに続く句全体にかかるため，particle は接続詞の後ろに置かれる．

例えば，以下の文を考える．

ἀλλ’ ἡ μὲν ΓΒ τῇ ΗΚ ἐστιν ἴση , ἡ δὲ ΓΗ τῇ ΚΒ ·

しかし，まず ΓΒ は ΗΚ に等しく，また ΓΗ は ΚΒ に（等しい）．

上の文には，particle は ‘μὲν, δὲ’ の 2 つがあり，それぞれ,「まず」,「また」を表す．これらの語を移動させた結果は以下のようになる．

ἀλλ’ μὲν ἡ ΓΒ τῇ ΗΚ ἐστιν ἴση , δὲ ἡ ΓΗ τῇ ΚΒ ·

この文の先頭 ‘ἀλλ’’ は英語の ‘but’ にあたる語であり，接続詞である．そのため，particle ‘μὲν’ は文の先頭ではなく，‘ἀλλ’’ の後ろに移動している．

また，もう 1 つの particle ‘δὲ’ はコンマで区切られた句の先頭に移動している．

これにより，図 4.2 のような解析木を作成することが可能となる．

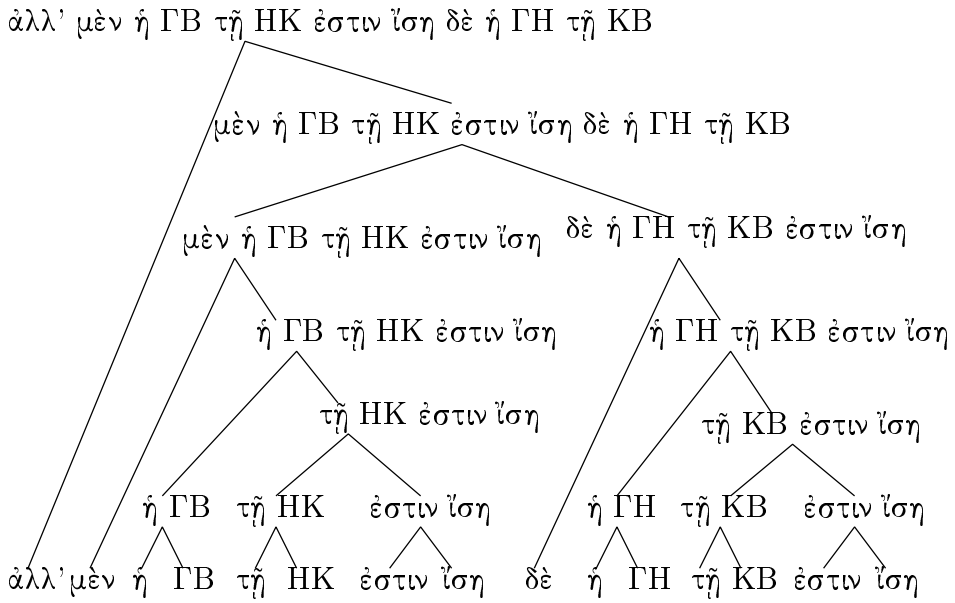

図 4.2: ‘ἀλλ' ἡ μὲν ΓΒ τῇ ΗΚ ἐστιν ἴση , ἡ δὲ ΓΗ τῇ ΚΒ ·’ の解析

4.3.5 省略された語の補完

τὸ ὑπὸ τῶν AB, BΓ **への補完**

「AB, BΓ によって囲まれた長方形」という意味を表す語‘τὸ ὑπὸ τῶν AB, BΓ’ は 2.1.4 で示したように,

τὸ ὑπὸ τῶν AB, BΓ περιεχόμενον ὀρθογώνιον

の省略形である．省略形でかかれた場合，品詞の並びは,

τὸ	ὑπὸ	τῶν	AB	BΓ
冠詞 –	前置詞 –	冠詞 –	記号（名詞） –	記号（名詞）

であり，句として成り立たない.

このような特殊な形で語が並んでいるのは，長方形をあらわすときと正方形を表すときのみなので，前処理において語の補完を行っている.

上の例の場合，先頭の語 ‘τό’ が中性冠詞であり,「〜によって」という意味を表す ‘ὑπό’ という前置詞の後に 2 本の直線を表す ‘τῶν AB, BΓ’ があるため,「囲まれた長方形」という語が省略されているとわかる.「囲まれた長方形」につく冠詞は先頭の句の先頭にあ

表 4.2: 補うべき語の対応（長方形 (ὀρθογώνιον) は中性名詞．単数の場合）

現れる省略形	省略されている語	格
τὸ ὑπὸ τῶν ΑΒ ΒΓ	περιεχόμενον ὀρθογώνιον	主格・対格
τοῦ ὑπὸ τῶν ΑΒ ΒΓ	περιεχομένου ὀρθογωνίου	属格
τῷ ὑπὸ τῶν ΑΒ ΒΓ	περιεχομένῳ ὀρθογωνίῳ	与格

る ‘τό’ のため，性・数・格の一致により，単数であり，主格もしくは対格の活用が必要とわかり，‘περιεχόμενον ὀρθογώνιον’ を補うことができる．

補うべき語の対応の例を表したものを表 4.2 に示す．

その他の補完

2.2 で説明したように，文中の動詞が省略されていることが存在する．この省略された動詞の特定は構文解析を行っていない前処理の段階では困難なため，手作業でこれらを補った．

また，同様に主語となる名詞句や冠詞が省略されることもあり，これらの推定も困難なため，手作業で補完を行っている．

4.4 構文解析

4.4.1 LiLFeS

本研究では構文解析を行うのに LiLFeS[4, 5] を用いた．LiLFeS は CYK 法を用いて高速にボトムアップ解析を行う HPSG パーサであり，型付き素性構造を記述できる論理型プログラミング言語の一種である．LiLFeS 上では現在までに日本語 [6]・英語 [11] の適用範囲の広い文法が書かれ，これらのうち，日本語文法では ID-schema を日本語文法のために改良することなどにより，EDR コーパスに対して約 92% の有効範囲を示し，70% を超える正解率を収めている．

本研究ではこの LiLFeS 上でギリシア語のための文法を作成することで，ユークリッ

ド『原論』の構文解析を行った.

4.4.2 素性

構文解析を行うために各単語に付与した主な素性について述べる.

PHONOLOGY 音韻を表す素性．単語のそのものを表す.

SYNSEM シンタックスとセマンティックスを表す素性.

HEAD SYNSEM | LOCAL | CATEGORY に入る素性．主辞に関する情報が格納される.

CASE HEAD に格納される素性．格の情報が表される.

MODIFY HEAD に格納される素性．その語の修飾する語の主辞の情報が格納される.

VALENCE SYNSEM | LOCAL | CATEGORY に入る素性．この素性に属する素性構造により，その語が取るべき主語，補語などが格納される.

SUBJECT VALENCE に格納される素性．主語として取るべき語の情報がリストで格納される.

COMPLEMENTS VALENCE に格納される素性．主語以外に補語として取るべき語の情報がリストで格納される.

SPECIFIER VALENCE に格納される素性．冠詞として取るべき語の情報がリストで格納される.

NUMBER SYNSEM | LOCAL | CONTENT | INDEX 以下に格納される素性．以下で述べる素性，PERSON, GENDER, MOOD, TENCE, VOICE についても同様．数の情報が表される．数に格納される型は単数，双数，複数の 3 種類である.

PERSON 人称の情報が表される．人称には 1 人称，2 人称，3 人称が存在する.

GENDER 性の情報が表される．性は男性，中性，女性が存在する.

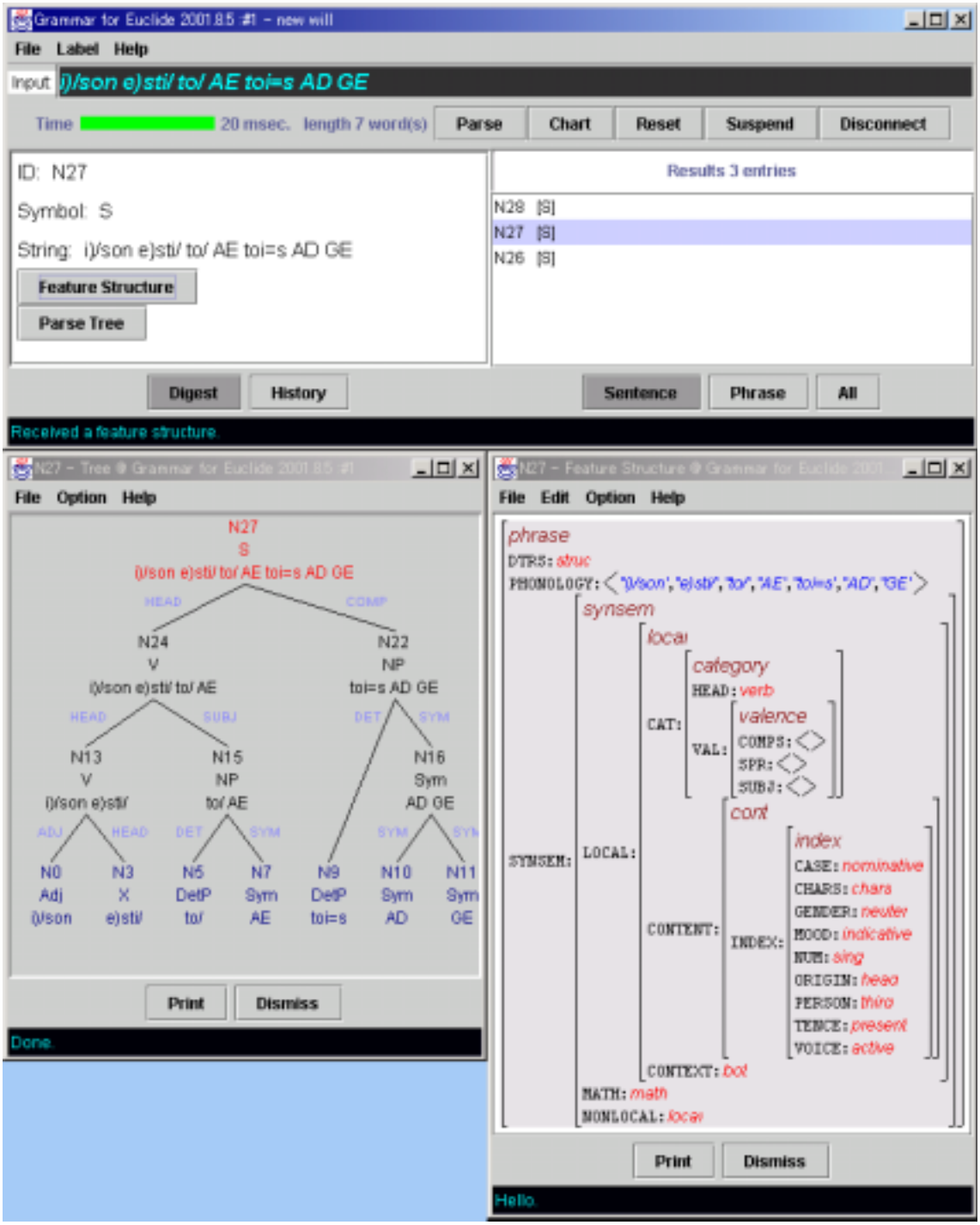

図 4.3: LiLFeS のフロントエンド GUI–Will の入出力画面

$$
\begin{bmatrix}
\text{PHON} & \langle\ \text{ἴσον ἐστι}\ \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \textit{verb} \\
\text{VAL} & \begin{bmatrix}
\text{SUBJ} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} \\ \text{CASE} & \textit{nom} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{COMP} & \left\langle \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \textit{noun} \\ \text{CASE} & \textit{dat} \end{bmatrix} \end{bmatrix} \right\rangle \\
\text{SPR} & \langle\ \rangle
\end{bmatrix}
\end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \begin{bmatrix}
\text{NUM} & \textit{sing} \\
\text{MOOD} & \textit{ind} \\
\text{VOI} & \textit{act} \\
\text{TEN} & \textit{pres} \\
\text{PER} & \textit{thir}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\text{PHON} & \langle\ \text{τοῖς AZ ΓE}\ \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \begin{bmatrix} \textit{sym} \\ \text{CASE} & \textit{dat} \end{bmatrix} \\
\text{VAL} & \begin{bmatrix}
\text{SUBJ} & \langle\ \rangle \\
\text{COMPS} & \langle\ \rangle \\
\text{SPR} & \langle\ \rangle
\end{bmatrix}
\end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \begin{bmatrix} \text{GEND} & \textit{neu} \\ \text{NUM} & \textit{plur} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

図 4.4: 素性構造の例

MOOD 法の情報が表される．法には直説法，接続法，希求法，命令法，不定法がある．

TENCE 時制の情報が表される．時制には現在，未来，未完了過去，アオリスト，現在完了，過去完了，未来完了の 7 種類が存在する．

VOICE 相の情報が表される．相には能動相，中動相，受動相がある．

SLASH SYNSEM | NONLOCAL | INHERITANCE 以下に格納される素性．語のギャップを表現するのに用いる．

4.4.3 辞書記述

動詞句 ‘ἴσον ἐστι’（等しい）と名詞句 ‘τοῖς AZ, ΓE’ （AZ, ΓE に），についての辞書記述の素性構造を図 4.4 に簡単に示す．

図において ‘ἴσον ἐστί’ は，SYNSEM | LOCAL | CATEGORY | VALENCE 素性に主語として主格の名詞を取ること，補語として与格の名詞を取ることが記述され，ほかに語の持つ情報として，直説法・能動相・現在・3 人称・単数であることが示されてい

る．また，下部に表されている ‘τοῖς AZ, ΓE’ では品詞が *symbol*，性・数・格がそれぞれ，中性・複数・与格であることが INDEX 素性，HEAD 素性中に書かれている．

4.4.4 schema

schema とは，ID-schema と principle を単一化させたもののことをいう．以下にこれらの schema を挙げる．なお，ここでは順序に関する規則（LP rule）についても触れる．

Head-subject schema

Head-subject schema は動詞句とその主語となる語（句）について規定する schema である．その素性構造を図 4.5 に示す．

図 4.5 より，補語となる語の素性構造（DAUGHTERS | SUBJECT_DAUGHTER）が構造共有により，動詞句（主辞．素性構造内の DAUGHTERS | HEAD_DAUGHTER 以下）の SYNSEM | LOCAL | CATEGORY | VALENCE | SUBJECT 内のリストと単一化されなければならない．すなわち，動詞句が主語として指定した語のみを補語としてとることができる．

さらに，SYNSEM | LOCAL | CONTENT | INDEX が構造共有されていることより動詞句と主語とが持つ人称・数の情報が単一化できなければならない．なお，この情報は同様に構造共有により親に受け継がれることになる．

その他に親には SYNSEM | LOCAL | CATEGORY | VALENCE 内の SUBJECT のリストが空で，同じ VALENCE 素性内の COMPLEMENTS，SPECIFIRE の各リストは主辞の情報が受け継がれる．

なお，Head-subject schema は英語などの場合，常に補語（主語）が左，主辞（動詞句）が右に位置するが，ギリシア語ではその語順は自由であるため，主辞が左，補語が右に現れてもよいとした．

Head-complement schema

Head-complement schema は動詞句と主語以外の補語との結合について規定する schema である．この schema の素性構造を図 4.6 に示す．

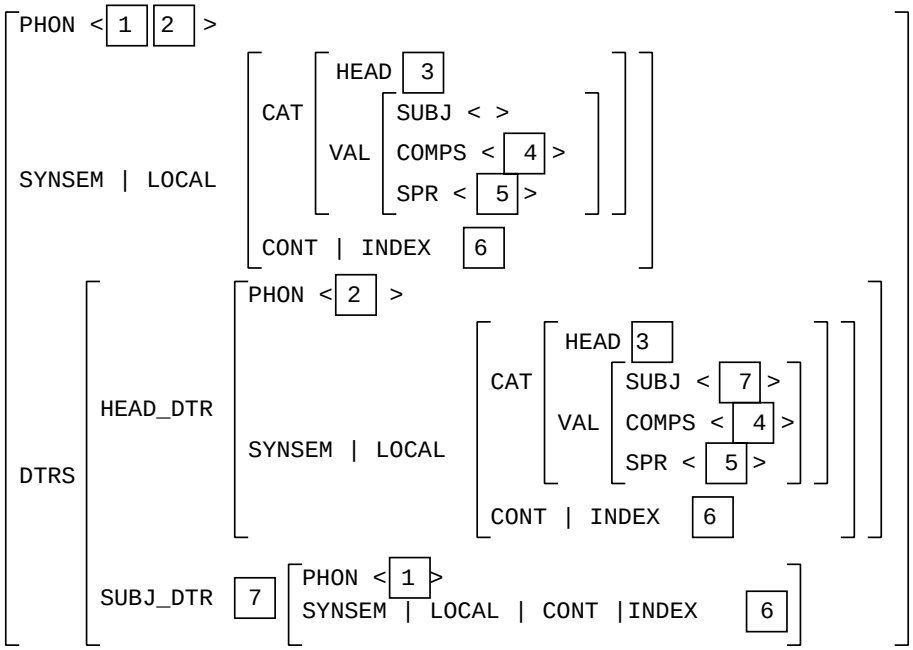

図 4.5: Head-subject schema

図より，主辞となる語（動詞句）の SYNSEM | LOCAL | CATEGORY | VALENCE | COMPLEMENTS の素性構造と補語となる語の素性構造が単一化できなければならない．すなわち，主辞が補語として指定する情報を持った語（句）でなければ単一化することはできない．

また，親の素性構造は SYNSEM | LOCAL | CATEGORY | HEAD，SYNSEM | LOCAL | CONTENT | INDEX の各素性は主辞の素性を受け継ぎ，VALENCE | COMPLEMENTS は空になるとわかる．

語順については Head-complement schema についても，自由なものとした．

Head-specifire schema

冠詞と名詞から名詞句を作る schema である．素性構造は図 4.7 とした．

対応する冠詞と名詞において性・数・格が一致しなければならない．性・数の一致はそれぞれの SYNSEM | LOCAL | CONTENT | INDEX 内の GENDER と NUMBER 素性について，格については SYNSEM | LOCAL | CATEGORY | HEAD | CASE に

$$
\begin{bmatrix}
\text{PHON} & \langle \boxed{1}\ \boxed{2} \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{4} \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{SPR} & \langle \boxed{5} \rangle \end{bmatrix} \end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \boxed{6}
\end{bmatrix} \\
\text{DTRS} & \begin{bmatrix}
\text{HEAD_DTR} & \begin{bmatrix}
\text{PHON} & \langle \boxed{1} \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle \boxed{4} \rangle \\ \text{COMPS} & \langle \boxed{7} \rangle \\ \text{SPR} & \langle \boxed{5} \rangle \end{bmatrix} \end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \boxed{6}
\end{bmatrix}
\end{bmatrix} \\
\text{COMP_DTR} & \boxed{7} \mid \text{PHON} \langle \boxed{2} \rangle
\end{bmatrix}
\end{bmatrix}
$$

図 4.6: Head-complement schema

ついて単一化が可能かどうかを見ることとなる．

語順は補語（冠詞）が先であり，主辞（名詞）が後であると固定している．

Head-compound schema

名詞の複合語を作る schema である．その素性構造を図 4.8 に示す．2 つの語の品詞が単一化でき，性・数・格が一致することが複合語の作られる条件となる．

図 4.8 からわかるように，親となる句は 2 つの語で単一化された品詞，性，数，格がそのまま受け継がれる．

Symbol-compound schema

以下のような句があるとする．

τὰ AB, BΓ

この句がユークリッド『原論』上に現れたときには,「(四角形）AB, BΓ」という意味になる．

```
[ PHON < [1] [2] >
  SYNSEM | LOCAL [ CAT [ HEAD [3]
                         VAL [ SUBJ < [4] >
                               COMPS < [5] >
                               SPR < > ] ] ]
                 CONT | INDEX [6] ]
  DTRS [ HEAD_DTR [ PHON < [2] >
                    SYNSEM | LOCAL [ CAT [ HEAD [3]
                                           VAL [ SUBJ < [4] >
                                                 COMPS < [5] >
                                                 SPR < [7] > ] ]
                                   CONT | INDEX [6] ] ]
         SPR_DTR [7] [ PHON < [1] >
                       SYNSEM | LOCAL | CONT |INDEX [6] ] ] ]
```

図 4.7: Head-specifire schema

```
[ PHON < [1] [2] >
  SYNSEM | LOCAL [ CAT [ HEAD [3]
                         VAL [ SUBJ < >
                               COMPS < >
                               SPR < > ] ]
                 CONT | INDEX | [4] ]
  DTRS [ HEAD_DTR [ PHON < [1] >
                    SYNSEM | LOCAL [ CAT [ HEAD [3]
                                           VAL [ SUBJ < >
                                                 COMPS < >
                                                 SPR < > ] ]
                                   CONT | INDEX | [4] ] ]
         COMPOUND_DTR [ PHON < [2] >
                        SYNSEM | LOCAL [ CAT [ HEAD [3]
                                               VAL [ SUBJ < >
                                                     COMPS < >
                                                     SPR < > ] ]
                                       CONT | INDEX | [4] ] ] ]
```

図 4.8: Head-compound schema

$$
\begin{bmatrix}
\text{PHON} & \langle \boxed{1}\,\boxed{2} \rangle \\
\text{SYNSEM} \mid \text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix} \text{HEAD} & \textit{symbol} \\ \text{VAL} & \begin{bmatrix} \text{SUBJ} & \langle\ \rangle \\ \text{COMPS} & \langle\ \rangle \\ \text{SPR} & \langle\ \rangle \end{bmatrix} \end{bmatrix} \\
\text{CONT} \mid \text{INDEX} \mid \text{NUMBER} & \textit{plural}
\end{bmatrix} \\
\text{DTRS} & \begin{bmatrix}
\text{HEAD_DTR} & \begin{bmatrix} \text{PHON} & \langle \boxed{1} \rangle \\ \text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{HEAD} & \textit{symbol} \end{bmatrix} \\
\text{SECOND_DTR} & \begin{bmatrix} \text{PHON} & \langle \boxed{2} \rangle \\ \text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{HEAD} & \textit{symbol} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

図 4.9: Symbol-compound schema

ここで現れる冠詞 ‘τά’ は中性・複数・主格である．ここで複数形が用いられるのは 2 つの四角形，‘AB’ と ‘BΓ’ をさしているためである．

‘AB, BΓ’ の 2 語からなる句を Head-compound schema を用いて生成した場合，これらの語の数は単数であるため，出来上がる句 ‘AB, BΓ’ も単数となる．しかし，このような状況で用いられる冠詞の数は複数である．そのため図 4.9 に示すこの schema を定義した．

Head-compound schema と異なるのは親に受け継がれる素性のうち，SYNSEM | LOCAL | CONTENT | INDEX | NUMBER が数が複数であることを表す *plural* になることのみである．

Head-adjunct schema

名詞に形容詞がかかるとき，副詞が動詞にかかるときなどに用いられる schema である．

図 4.10 に示すように補語 (ADJUNCT_DAUGHTER) の SYNSEM | LOCAL | CATEGORY | HEAD | MODIFIRE と主辞が構造共有する．すなわち，補語が主辞についての規定をしている．

また，2 つの語の SYNSEM | LOCAL | CONTENT | INDEX を構造共有することで，名詞と形容詞の間において性・数・格の一致をさせることができる．

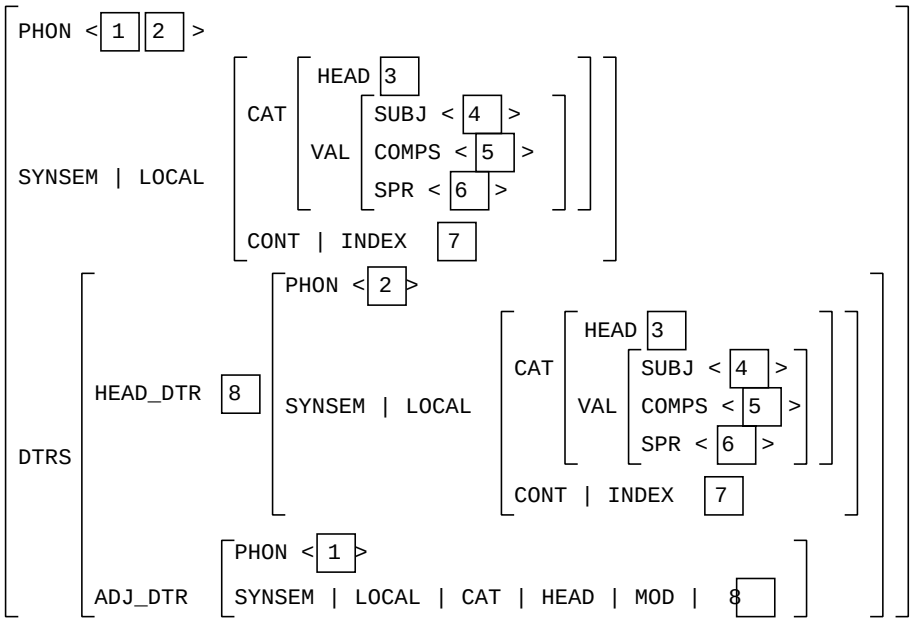

図 4.10: Head-adjunct schema

なお，Head-adjunct schema についても，Head-subject schema や Head-complement schema と同様に主辞，補語の語順についてはどちらも可能としている．

Head-filler schema

関係詞のために用いられる schema であり，図 4.11 のように表される．

SYNSEM | NONLOCAL | INHERITANCE | SLASH 以下の素性構造と補語 FILLER_DAUGHTER の素性構造が単一化される必要がある．すなわち，ギャップとして残された場所を埋める句が補語として取られる．

これにより，親の素性構造はそのギャップが埋められた以外はすべて主辞の素性構造を受け継ぐ．

4.5 解析例

以下に 2.1.4 に示した遠隔依存の問題を含む句,

τὸ ὑπὸ τῶν ΑΒ, ΒΓ περιεχόμενον ὀρθογώνιον

$$
\begin{bmatrix}
\text{PHON} & \langle \boxed{1}\ \boxed{2} \rangle \\
\text{SYNSEM} & \begin{bmatrix}
\text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \boxed{3} \\
\text{VAL} & \begin{bmatrix}
\text{SUBJ} & \langle \boxed{4} \rangle \\
\text{COMPS} & \langle \boxed{5} \rangle \\
\text{SPR} & \langle \boxed{6} \rangle
\end{bmatrix}
\end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \boxed{7}
\end{bmatrix} \\
\text{NONLOCAL} \mid \text{INHER} \mid \text{SLASH} & \langle\ \rangle
\end{bmatrix} \\
\text{DTRS} & \begin{bmatrix}
\text{HEAD_DTR} & \begin{bmatrix}
\text{PHON} & \langle \boxed{2} \rangle \\
\text{SYNSEM} & \begin{bmatrix}
\text{LOCAL} & \begin{bmatrix}
\text{CAT} & \begin{bmatrix}
\text{HEAD} & \boxed{3} \\
\text{VAL} & \begin{bmatrix}
\text{SUBJ} & \langle \boxed{4} \rangle \\
\text{COMPS} & \langle \boxed{5} \rangle \\
\text{SPR} & \langle \boxed{6} \rangle
\end{bmatrix}
\end{bmatrix} \\
\text{CONT} \mid \text{INDEX} & \boxed{7}
\end{bmatrix} \\
\text{NONLOCAL} \mid \text{INHER} \mid \text{SLASH} & \langle \boxed{8} \rangle
\end{bmatrix}
\end{bmatrix} \\
\text{FILLER_DTR}\ \boxed{8} & \begin{bmatrix}
\text{PHON} & \langle \boxed{1} \rangle \\
\text{SYNSEM} \mid \text{LOCAL} \mid \text{CAT} \mid \text{VAL} & \begin{bmatrix}
\text{SUBJ} & \langle\ \rangle \\
\text{COMPS} & \langle\ \rangle \\
\text{SPR} & \langle\ \rangle
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

図 4.11: Head-filler schema

の解析例を示す.

LiLFeS は句，文をボトムアップに解析していくので，それに習い，以下ではボトムアップに解析例を述べる.

前処理

句を構文解析する前に前処理を行う．アクセントの修正，コンマの修正が行われ，以下のような句になる.

τό ὑπό τῶν AB BΓ περιεχόμενον ὀρθογώνιον

AB BΓ

語 ‘AB’ の素性構造は図 4.12 に示すように表される．また，‘BΓ’ も同様に表すことができる．すなわち，品詞が *symbol* であり，性・格の情報は特に持たない.

この 2 語を結合するのに Symbol-compound schema を用いることで，句 ‘AB BΓ’ が生成される.

τῶν AB BΓ

次に女性・複数・属格の名詞句（SYNSEM | LOCAL | CATEGORY | HEAD の型は *symbol*）‘τῶν AB BΓ’ を作成することを考える.

Head-specifire schema と ‘τῶν’，‘AB BΓ’ の 2 つの語句を単一化させることで，図 4.13 に示すような素性構造が作られる.

ὑπὸ τῶν AB BΓ

Head-complement schema により，属格支配の前置詞 ‘ὑπό’ と属格の名詞句 ‘τῶν AB BΓ’ が結合され，前置詞句となる.

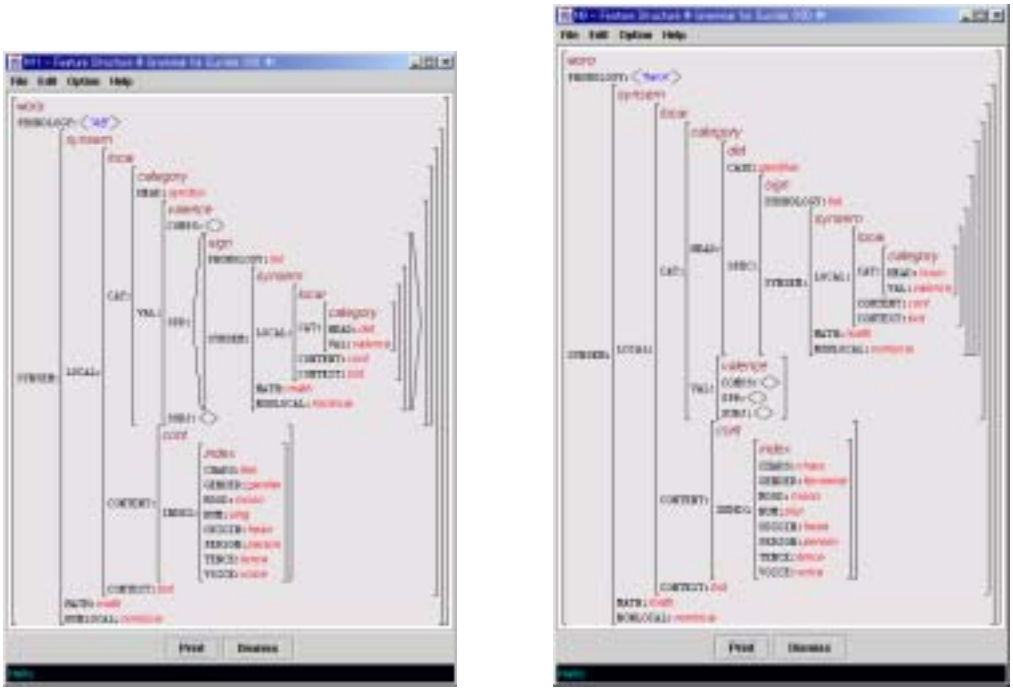

図 4.12: ‘AB, τῶν’ の素性構造

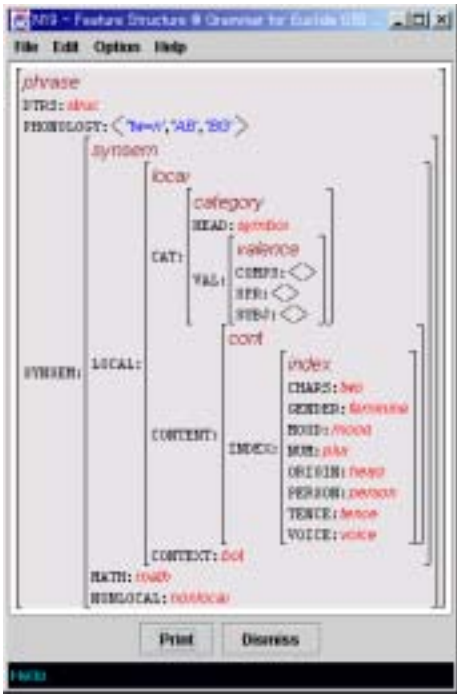

図 4.13: ‘τῶν AB ΒΓ’ の素性構造

ὑπὸ τῶν ΑΒ ΒΓ περιεχόμενον

‘περιεχόμενον’ は分詞である．つまり，動詞的な役割と形容詞的な役割を同時にはたす．ここでは，Head-complement schema を用いることで，図 4.14 のように，補語として ‘ὑπό τῶν AB BΓ’ を取り，結合可能であることを示す．

それぞれの辞書記述は図 4.14 の下部の 2 つの素性構造のように表される．これらを Head-complement schema（図 4.14 の中央部）と単一化させる．すなわち，‘περιεχόμενον’ は DAUGHTERS | HEAD_DAUGHTER 素性に入り，‘ὑπό τῶν AB BΓ’ は DAUGHTERS | COMPLEMENT_DAUGHTER 素性に入る．このとき，DAUGHTERS | HEAD_DAUGHTER | SYMSEM | LOCAL | VALENCE |COMPLEMENTS と COMPLEMENT_DAUGHTER 全体において構造共有が存在する．

よって，‘περιεχόμενον’ の SYMSEM | LOCAL | VALENCE | COMPLEMENTS 素性に含まれる型，HEAD | *preposition*，と ‘ὑπό τῶν AB BΓ’ の対応する素性が単一化可能かどうか確認する．

ともに HEAD の型は *preposition* のため，単一化可能である．

したがって，この 2 つの句は Head-complement schema を用いることにより結合することができ，その他の構造共有から親として新たに図 4.14 の上部に示す型付き素性構造を生成することができる．

ὑπὸ τῶν ΑΒ ΒΓ περιεχόμενον ὀρθογώνιον

分詞 ‘περιεχόμενον’ はここでは名詞 ‘ὀρθογώνιον’ にかかるために形容詞的な役割をはたしている．

従って，Head-modifire schema を用いることで，名詞句，‘ὑπὸ τῶν AB BΓ περιεχόμενον ὀρθογώνιον’ が生成される．

τὸ ὑπὸ τῶν ΑΒ ΒΓ περιεχόμενον ὀρθογώνιον

‘ὑπὸ τῶν AB BΓ περιεχόμενον ὀρθογώνιον’ が名詞句として成り立つためには冠詞が必要である．

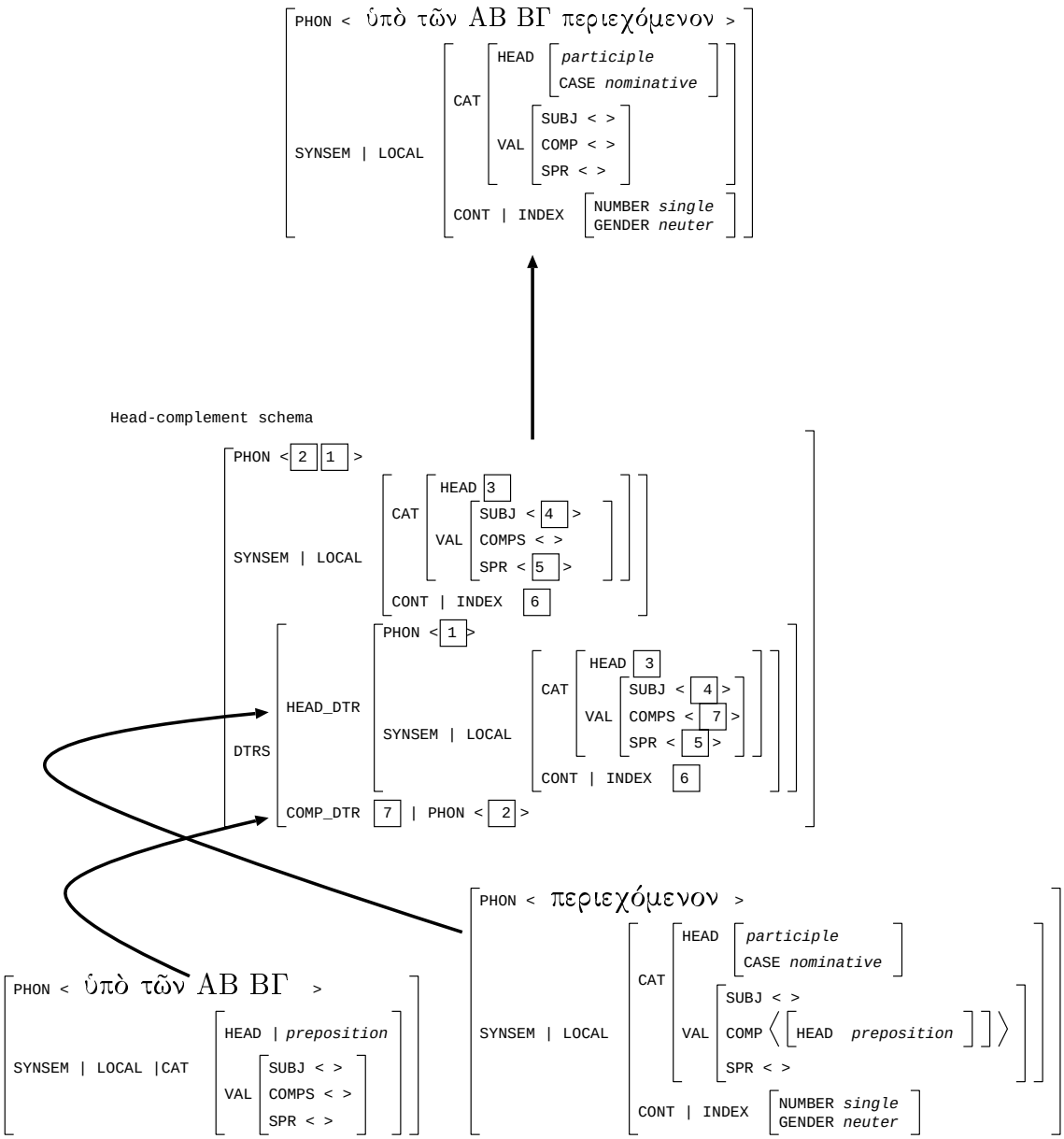

図 4.14: ‘ὑπὸ τῶν AB ΒΓ περιεχόμενον’ の解析

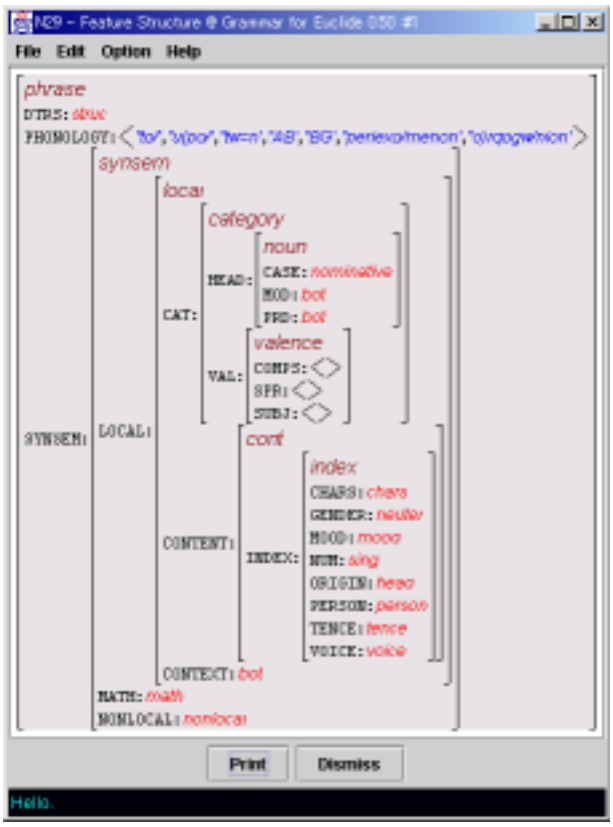

図 4.15: ‘τὸ ὑπὸ τῶν AB BΓ περιεχόμενον ὀρθογώνιον’ の素性構造

Head-specifire schema を用いることで，ここでの最終的な句，‘τὸ ὑπὸ τῶν AB BΓ περιεχόμενον ὀρθογώνιον’ が生成される．この句の素性構造を図 4.15 に示す．

また，以上のようにして LiLFeS 上で作られた解析木を図 4.16 に示す．

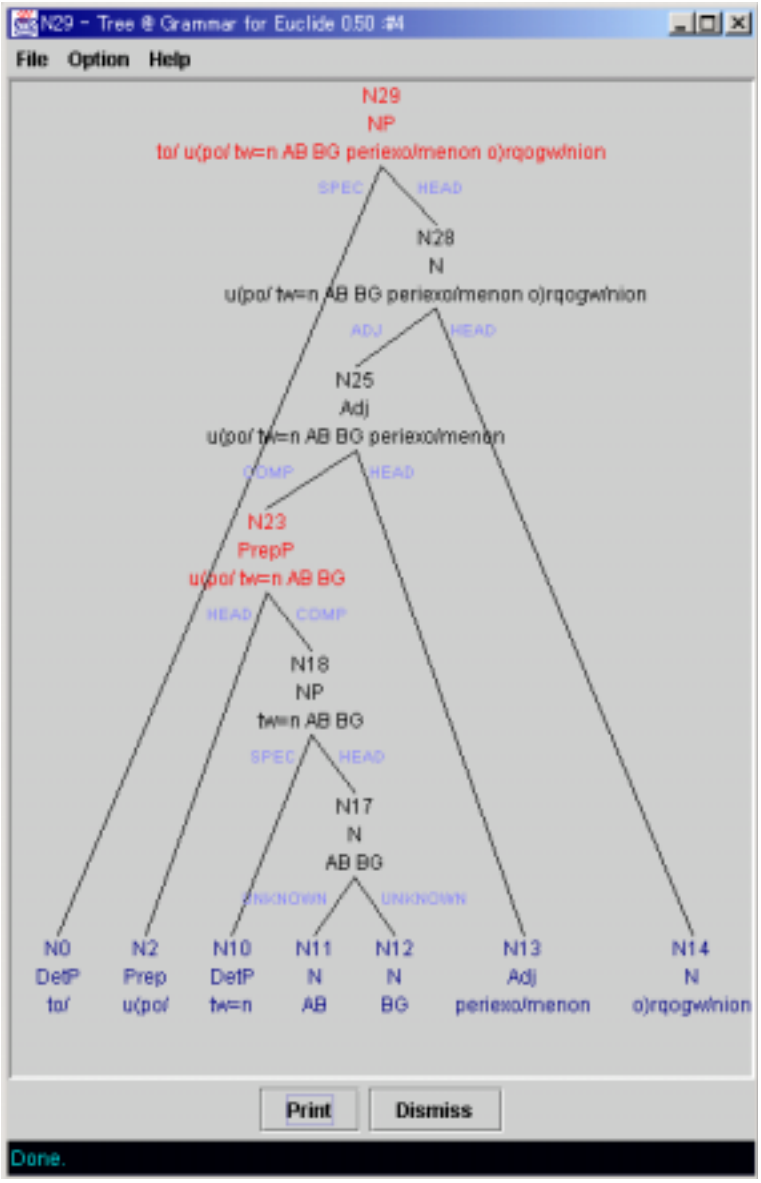

図 4.16: ‘τὸ ὑπὸ τῶν AB BΓ περιεχόμενον ὀρθογώνιον’ の解析木

第 5 章

考察

4 章で述べたような前処理を行い，辞書記述，schema を作成することで，ユークリッド『原論』第 2 巻 命題 2 から命題 5 までの構文解析を行うことができるようになった．

現在のシステムでは命題 2 から命題 5 までの全 76 文中，75 文まで正解の木を含む解析木を作成することが可能である．表 5.1 に作成した辞書記述数，ID-schema 数などを挙げる．

表 5.1: 解析に用いた辞書記述数その他

解析できた文の数	75
単語数（‘AB’ などの記号を含む）	1340
辞書記述数（記号を含まず）	165
辞書記述数（記号のみ）	64
ID-schema	10
前処理により語（句）を補った数	8
手作業により語（句）を補った数	12

5.1 前処理について

前処理では以下の事を行っていた．

- アクセント，規則記号，つづりの修正
- ,（コンマ）の修正
- 冠詞の後ろに続く語の修正
- particle の位置の修正
- 省略された語の補完

アクセントなどの修正は文法構造，意味構造ともに完全に同じ語を辞書記述に書く必要がなくなり，その量の減少に役立っている．特に，英語の be 動詞にあたる ‘ἐστί’ という語は他に，‘ἐστι, ἐστίν, ἐστὶ, ἐστιν, εἰσιν’ というように書くことがあり，修正の効果は大きい．

コンマの修正については文の切れ目と記号の区切りを混同することがなくなるという点で役立っている．

また，冠詞のすぐ後ろに続く語の位置の修正と particle の位置の修正により，すでにできあがってしまっている文法構造を破壊することがなくなった．また，位置を移動したことにより，例えば particle ならば，通常の副詞，接続詞などと同様に扱えるようになった．

省略された語の補完は不完全であり，自動的に語が補完されることは少ないが，完全な構文解析を行う上では不可欠な作業である．今後，現在は手で補っている語句についても自動的に補えるようにすることが重要であると考える．

これらの前処理はパターンマッチングによって行っているのみなので，機能は低いが，動作速度が速く，構文解析に必要となる文法構造には破綻をきたさない．従って，前処理として十分に機能しているといえる．

5.2 辞書記述とID-schema

辞書記述は5.1で述べたようなアクセントなどの修正を終えた後の語すべてについて，少なくとも1つ書かれている．これらの辞書記述はユークリッド『原論』の第2巻 命題2から命題5にかけては似たような議論が続くため，単語の多様さにおいて乏しい．しかし，particle をはじめとして古代ギリシア語の基本的な単語の種類については網羅されていることから，比較的容易に解析できる文を増すことができると思われる．

また，ID-schema は4.4.4で挙げた7種類に Head-subject schema，Head-complement schema，Head-adjunct schema の左右の語句が入れ替わっているもの3種類を加えて10としている．

5.3 省略された語の補完

手で補っている語句の内訳は動詞句が6，主語となる名詞句が4，冠詞が2である．

これらのうち，動詞句と名詞句については直前の文に現れている同じ役割をはたす構成素を補完することで文として成立するようになった．すなわち，構文情報を参照することでどの語が省略されているかを推定することができると考えられる．

対して冠詞の補完はそれらに比べ，困難であると感じた．例えば以下の文で下線の引かれている語 ‘τῶν’ は実際の文には現れていない．

> λέγω , ὅτι τὸ ὑπὸ τῶν AB, BΓ περιεχόμενον ὀρθογώνιον μετὰ τοῦ ὑπὸ <u>τῶν</u> BA AΓ περιεχομένου ὀρθογωνίου ἴσον ἐστὶ τῷ ἀπὸ τῆς AB τετραγώνῳ .
>
> 私は言う，AB, BΓ によって囲まれる長方形に BA, AΓ によって囲まれる長方形をあわせたものは，AB 上の正方形に等しい．

このうち，下線部が含まれている句「BA, AΓ によって囲まれる長方形」を取り出すと以下のようになる．

> τοῦ ὑπὸ <u>τῶν</u> BA, AΓ περιεχομένου ὀρθογωνίου

つまり,「(直線) BA, AΓ」を表すための冠詞が存在していない．このため，‘BA, AΓ’ がどのようなものを指しているのかを知るためにはその周りの句 ‘ τοῦ ὑπὸ , περιεχομένου

ὀρθογωνίου ’ から類推する必要がある．しかし，例えば，直前の句 ‘ὑπό’ は属格，与格，もしくは対格を支配する前置詞であるため，格の特定ができず，他に名詞や形容詞も現れていないため，性・数の特定も困難である．

結局，前置詞 ‘ὑπό’（属格支配ならば「〜によって」）と直後の分詞 ‘περιεχομένου’（囲まれて）の意味的情報から類推することしかできない．

第 6 章

おわりに

6.1 まとめ

HPSG に基づく古代ギリシア語についての文法を LiLFeS 上で実装することで，ユークリッド『原論』 第 2 巻 命題 2 から命題 5 までに現れる文，全 76 文中，75 文において正解を含む解析木を作成できることが可能となった.

この文法では古代ギリシア語のために辞書記述の拡張を行うことで，古代ギリシア語の特徴として見られる 1 語に対する情報量の多さから豊かな辞書記述を容易に表すことができた.

しかし，古代ギリシア語では 1 語に対する情報量の多さから語順が比較的自由であるため，辞書記述に従い，主に語順に関する id-schema の拡張を行った．これにより，遠隔依存の問題などが解消した．また，ユークリッド『原論』に特化した id-schema, Symbol-compound schema を作成することで，数学の議論によく現れる記号を柔軟に扱うことができるようになった.

また，構文解析に前処理を導入することで，他の文法規則にとらわれない，という特徴をもつギリシア語特有の語，particle の位置を文頭に移動させ，他の接続詞などと同じように扱えるようにした.

6.2 今後の課題

解析できなかった残りの 1 文については Head-filler schema について更なる考慮をすることで解析が可能と考えている．

また，現在手作業で行っている省略語の補完についても構文解析結果から足りない語の情報をフィードバックし，足りない語を埋めるなどして，自動で行えるようにすることを考える必要がある．

今後はユークリッド『原論』の他の命題の解析，他の文献の解析を可能にした後，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をできるように情報の有効な利用方法を提案することが目標となる．

謝辞

古代ギリシア語について 1 から教えていただいたことに加え，ユークリッド『原論』の日本語訳のご提供もしていただいた大阪府立大学 総合科学部 人間科学科 斎藤憲助教授に深く感謝いたします.

また，HPSG についてのさまざまな助言をいただき，LiLFeS という便利なツールを紹介いただいた，知識工学講座の鳥澤健太郎助教授に深く感謝いたします.

参考文献

[1] B. Carpenter. *The logic of typed feature structures.* Cambridge University Press, 1992.

[2] Euclides. *Euclidis Elementa.* Ed. I. L. Heiberg and E. S. Stamatis. Leipzig : Teubner, 1969.

[3] Liddell and Scott. *An Intermediate Greek – English Lexicon.* Oxford University Press, 1959.

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

[5] T. Makino, M. Yoshida, K. Torisawa, and J. Tsujii. LiLFeS – Towards a Practical HPSG Parser. In *COLING-ACL '98*, 1998.

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

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

[8] I. A. Sag and T. Wasow. *Syntactic Theory.* CSLI Publications, 1999.

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

[10] 田中美知太郎, 松平千秋. ギリシア語入門. 岩波全書, 1962.

[11] 建石由佳, 鳥澤健太郎, 牧野貴樹, 西田健二, 渕上正睦, 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.

[12] Thesaurus Linguae Graecae. http://www.tlg.uci.edu/~tlg/.

付 録 A

ギリシア語対応表

ギリシア語	読み	意味
ἀναγεγράφθω	anagegraftho	描かれたとせよ
ἀπὸ	apo	の上に
ἄρρά	ara	しかし
γὰρ	gar	何となれば
δέ	de	また (particle)
δίς	dis	2 倍の
ἐπεί	epei	であるから
ἐστι	esti	英語の be 動詞にあたる
εὐθεῖα	eytheia	直線
ἡ	he	冠詞 (女性・単数・主格)
ἴσον	ison	等しい
καί	kai	そして
καί ... τε ...	kai ... te ...	英語の ‘both ... and ...’ にあたる
λέγω ὅτι	rego hoti	英語の ‘I say that’ にあたる
μέν	men	まず (particle)
ὀρθογώνιον	orthogonion	長方形
περιεχόμενον	periekhomenon	囲まれる

ギリシア語	読み	意味
πλευρά	pureura	辺
τετράγωνον	tetragonon	正方形
τῇ	te	冠詞 (女性・単数・与格)
τῆς	tes	冠詞 (女性・単数・属格)
τὸ	to	冠詞 (中性・単数・主格)
τοῖς	tois	冠詞 (中性・複数・与格)
τῶν	ton	冠詞 (女性・複数・属格)
τῷ	to	冠詞 (中性・単数・与格)
ὑπὸ	hypo	によって