

情報工学実験III

計算機C講座 村上仁一

課題

自然言語処理パーザの試作

自然言語

人間と人間のインターフェース
日本語、英語

人工言語

人間とコンピュータのインターフェース
C、Fortran, Basic

自然言語処理

自然言語でコンピュータとインターフェース

例

仮名漢字変換 機械翻訳 対話処理
検索 その他

言語学

音韻論

形態論

統語論

意味論

文系の研究
古代ギリシア文明から

言葉の意味とは？

右

左

北

東

(意味を定義することは困難)

ギリシア

ソフィスト
詭弁に対抗

弁証法

文の意味が理解可能
正文

意味が理解不可能
非文

正文 と 非文 の判断

仮想機械の存在（汎用計算機？）

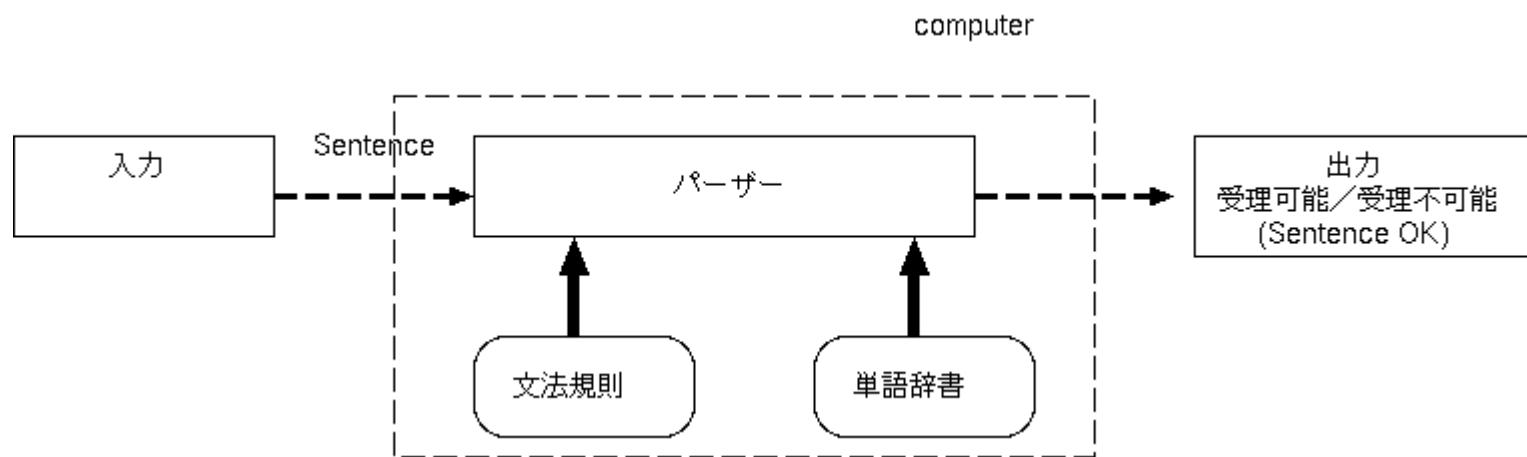
受理可能	正文
受理不可能	非文

仮想機械

= パーサー + 文法 + 単語辞書

文法

正文, 非文を受理可能, 受理不可能で
判断できる機械の規則



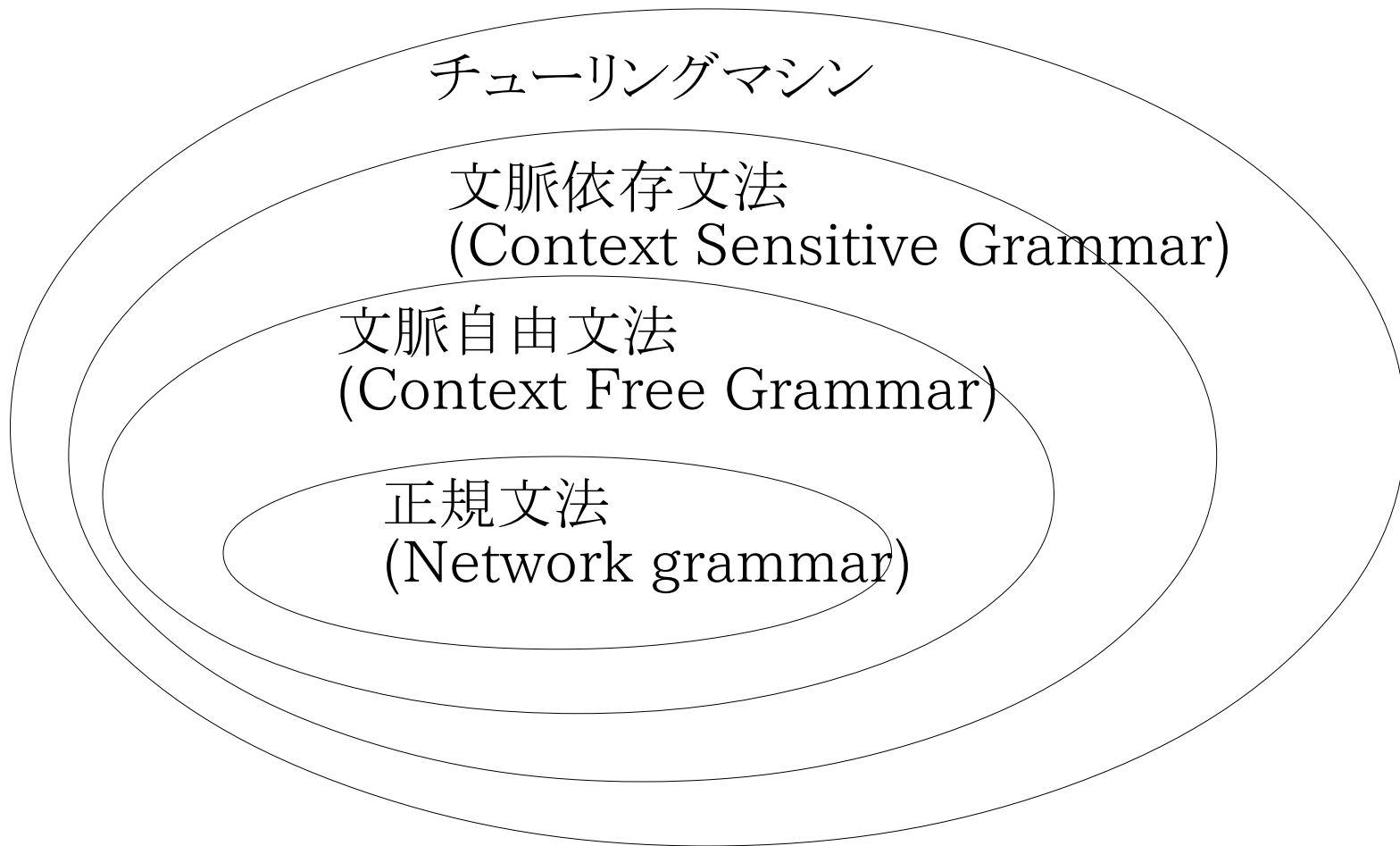
言語パーザー

自然言語処理の基本

自然言語処理パーザー

プログラム + 文法規則 + 単語辞書

受理

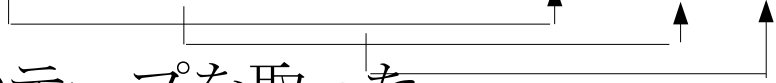


チョムスキーの言語階層

チューリングマシン

文脈依存文法

一郎、次郎、三郎は、赤、青、黄
のテープを取った。



(Context Sensitive Grammar)

文脈自由文法

東京にいる人^に会い^にいった。



(Context Free Grammar) 複文

正規文法

(Network grammar)

私は 山に いった。
単文

正規文法

有限状態オートマトン

非終端記号 A, B, C, \dots (品詞)

終端記号 a, b, c, \dots (単語)

$A \rightarrow B$

$B \rightarrow b$

自然言語処理の基本

形態素解析、構文解析

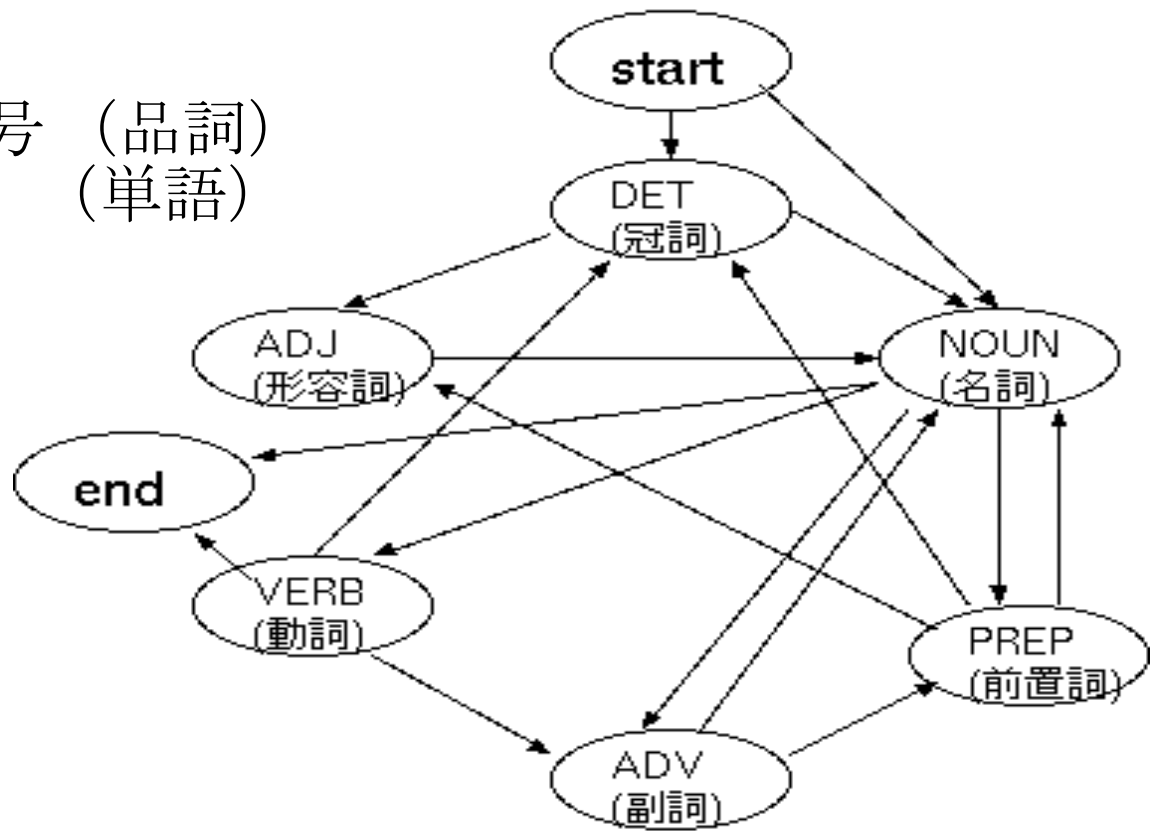
形態素解析

	the	child	runs	quickly	to	the	large	house	
start									end
	DET	NOUN	VERB	ADV	PREP	DET	ADJ	NOUN	
初期状態	冠詞	名詞	動詞	副詞	前置詞	冠詞	形容詞	名詞	最終状態

A → B
B → a

A, B, C 非終端記号 (品詞)
a 終端記号 (単語)

文法規則

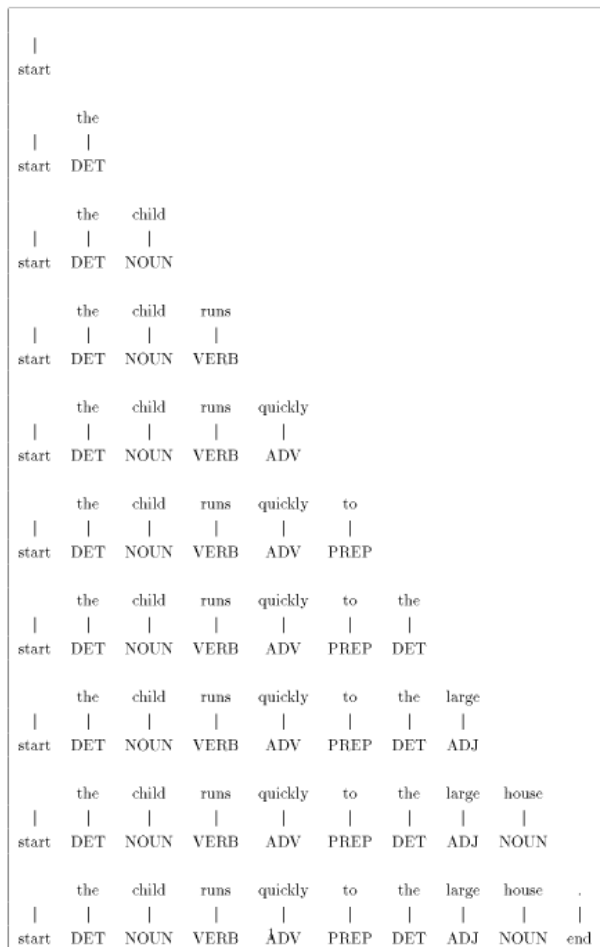


有限状態オートマトン

ネットワーク文法

文法規則		単語辞書	
start	→ DET (冠詞)		
start	→ NOUN (名詞)		
DET(冠詞)	→ ADJ(形容詞)		
DET(冠詞)	→ NOUN(名詞)		
ADJ(形容詞)	→ NOUN(名詞)		
NOUN(名詞)	→ PREP(前置詞)	the	→ DET(冠詞)
PREP(前置詞)	→ NOUN (名詞)	child	→ NOUN(名詞)
PREP(前置詞)	→ ADJ(形容詞)	runs	→ VERB(動詞)
PREP(前置詞)	→ DET(冠詞)	quickly	→ ADV(副詞)
NOUN(名詞)	→ VERB(動詞)	to	→ PREP(前置詞)
VERB(動詞)	→ ADV(副詞)	the	→ DET(冠詞)
VERB(動詞)	→ DET(冠詞)	large	→ ADJ(形容詞)
ADV(副詞)	→ PREP(前置詞)	house	→ NOUN(名詞)
NOUN(名詞)	→ ADV(副詞)		
ADV(副詞)	→ NOUN(名詞)		
VERB(動詞)	→ end		
NOUN(名詞)	→ end		

表 1: NetWork 文法の動作



network文法の解析

実験 1 周目

1.1 ネットワーク文法の実験1

人間の手において以下の文を形態素解析せよ。

"I saw a girl with a telescope"

1.2 ネットワーク文法の仕様書

ネットワーク文法を解析するプログラムの仕様書を作成せよ。

仕様書

他人がプログラムや仕様書のメンテナンスができるように記述する文章

外部仕様書

プログラム全体の動作を記述した文章

入力データ 使用するデータベース 出力データの概要

プログラムの各モジュールの記述

内部仕様書

各モジュールの詳細の記述

各モジュールの役割、入力データ、

使用するデータベース、および出力データの詳細

(詳細なformatも含む.)

実験のメモ

1)editor

emacsを利用すること.

2)EUCコード

プログラムには, 日本語のコメントを書くこと.

なおEUC(UTF)コードで書くこと.

(EUC,SJIS,JIS,ASCIIの違いを述べよ)

あ JIS 0x2422 EUC 0xA4A2

\$ ASCII 0x24 " ASCII 0x22

3)make

プログラムは, 分割して書くこと.

また分割コンパイルが可能なように makeを使うこと.

4)プログラムと辞書の分離

文法規則と単語辞書はfileとして作成すること.

Cのprogramming

strcpy, strcmp, strcat, fopen
fgets, fscanf, fprintf

make, gdb

shell

実験 2 周目

1.1 ネットワーク文法の実験 2

プログラムにおいて以下の文を形態素解析せよ。

"The child runs quickly to the large house."

1.2 ネットワーク文法の実験 2

プログラムにおいて以下の文を形態素解析せよ。

"I saw a girl with a telescope"

文脈自由文法 (CFG)

PushDownオートマトン

非終端記号 A,B,C... (品詞)

終端記号 a,b,c.. (単語)

$A \rightarrow B + C$

$B \rightarrow b$

$C \rightarrow c$

構文解析

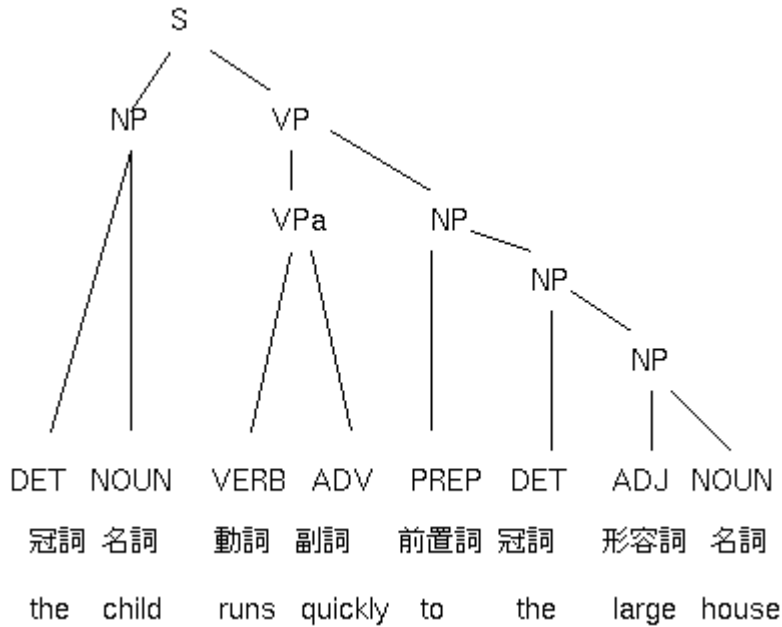
文脈自由文法

$A \rightarrow B + C$

$B \rightarrow a$

A, B, C 非終端記号
(品詞)

a 終端記号
(単語)



文脈自由文法

文法規則	単語辞書
SENTENCE(文) → NP(名詞句) + VP(動詞句)	
NP(名詞句) → DET(冠詞) + NOUN(名詞句)	the → DET(冠詞)
NP(名詞句) → DET(冠詞) + NP(名詞句)	child → NOUN(名詞)
NP(名詞句) → PREP(前置詞) + NP(名詞句)	runs → VERB(動詞)
NP(名詞句) → ADJ(形容詞) + NOUN(名詞)	quickly → ADV(副詞)
VP(動詞句) → VERB(動詞) + NP(名詞句)	to → PREP(前置詞)
VP(動詞句) → VP1(動詞句) + NP(名詞句)	the → DET(冠詞)
VP(動詞句) → VERB(動詞) + ADV(副詞)	large → ADJ(形容詞)
VP(動詞句) → VERB(動詞)	house → NOUN(名詞)
VPa(動詞句) → VERB(動詞) + ADV(副詞)	
VPa(動詞句) → VERB(動詞) + NP(名詞句)	

(S (NP (DET ``the") (NOUN ``child")) (VP (VPa (VERB
``runs")(ADV ``quickly")) NP (PREP ``to") (NP (DET
``the")(NP (ADJ ``large") (NOUN ``house"))))))

S 式

構文解析方法

Top Down 法

CYK 法

Early 法

LR 法

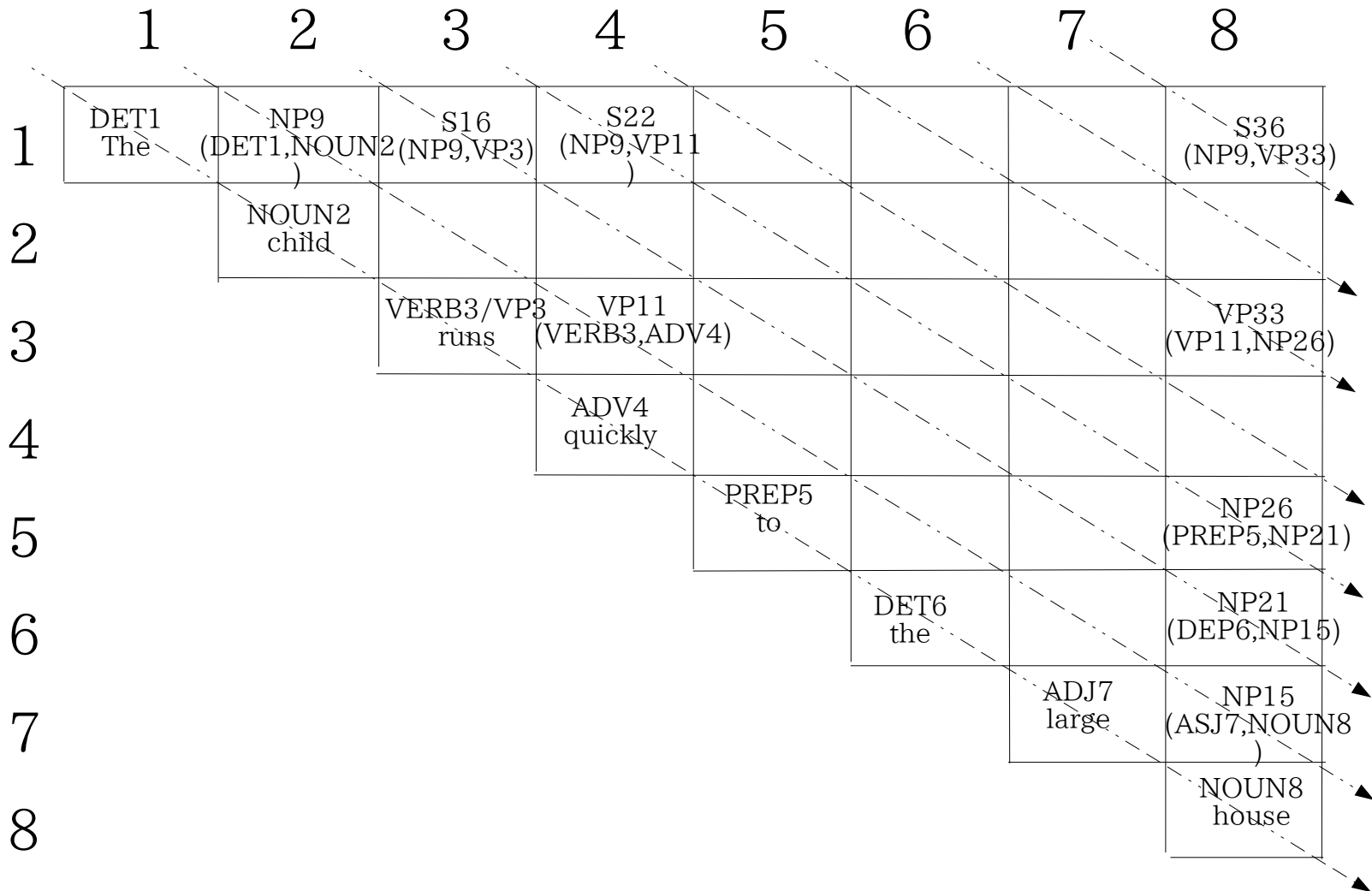
Top Down 法のフロー

1	S_1	\rightarrow	$NP_1 + VP_1$	
2	NP_1	\rightarrow	$DET_2 + NOUN_2$	1のNPを処理する.
3	DET_2	\rightarrow	the	
4	$NOUN_2$	\rightarrow	child	1のNPが受理されたのでVPを処理する.
5	VP_1	\rightarrow	$VERB_3 + NP_3$	
6	$VERB_3$	\rightarrow	runs	
7	NP_3	\rightarrow	$DET_7 + NOUN_7$	
8	DET_7	\rightarrow	quickly	
9	NP_3	\rightarrow	$DET_7 + NP_7$	
10	DET_7	\rightarrow	quickly	
11	NP_3	\rightarrow	$PREP_{11} + NP_{11}$	
12	$PREP_{11}$	\rightarrow	quickly	
13	NP_3	\rightarrow	$ADJ_{13} + NOUN_{13}$	
14	ADJ_{13}	\rightarrow	quickly	
15	VP_1	\rightarrow	$VP_{15} + NP_{15}$	
16	VP_{15}	\rightarrow	$VERB_{16} + ADV_{16}$	
17	$VERB_{16}$	\rightarrow	runs	
18	ADV_{16}	\rightarrow	quickly	
19	NP_{15}	\rightarrow	$DET_{19} + NOUN_{19}$	
20	DET_{19}	\rightarrow	to	
21	NP_{15}	\rightarrow	$DET_{21} + NP_{21}$	
22	DET_{19}	\rightarrow	to	
23	NP_{15}	\rightarrow	$PREP_{23} + NP_{23}$	
24	$PREP_{23}$	\rightarrow	to	24が受理
25	NP_{23}	\rightarrow	$DET_{25} + NOUN_{25}$	
26	DET_{25}	\rightarrow	the	
27	$NOUN_{25}$	\rightarrow	large	26が受理されず.
28	NP_{23}	\rightarrow	$DET_{28} + NP_{28}$	
29	DET_{28}	\rightarrow	the	25が受理
30	NP_{28}	\rightarrow	$DET_{30} + NOUN_{30}$	
31	DET_{30}	\rightarrow	large	
32	NP_{28}	\rightarrow	$DET_{32} + NP_{32}$	
33	DET_{32}	\rightarrow	large	
34	NP_{32}	\rightarrow	$PREP_{34} + NP_{34}$	
35	$PREP_{34}$	\rightarrow	large	
36	NP_{32}	\rightarrow	$ADJ_{36} + NOUN_{36}$	
37	ADJ_{36}	\rightarrow	large	
38	$NOUN_{36}$	\rightarrow	house	

Top Down 法 ¹ (ひたすらルールを適応)

Top-Downパーサでは記述してはいけない文法例
(左再帰規則)

VP(動詞句) → VP(動詞句) + NP(名詞句)



CYK法(一部)

CYKの文法規則

$A_{i,j}$ には $A_{i,k} A_{k+1,j}$ の書き換え規則

例

$$\begin{aligned} A_{3,7} &= A_{3,3} A_{4,7} \\ &= A_{3,4} A_{5,7} \\ &= A_{3,5} A_{6,7} \\ &= A_{3,6} A_{7,7} \end{aligned}$$

実験 3 周目

3.1 文脈自由文法の実験1 (TopDown)

人間の手にて以下を文をTopDownで構文解析せよ。

3.2 文脈自由文法の実験2 (CYK)

人間の手にて以下を文をCYKで構文解析せよ。

3.3 CYK法のプログラムの使用書を作成せよ。

実験 4,5 日目

4: 文脈自由文法の実験

プログラム (CYK) において以下の文を構文解析せよ。

"The child runs quickly to the large house."

5: 文脈自由文法の実験

プログラム (CYK) において以下の文を構文解析せよ。

"I saw a girl with a telescope"

実験 6,7 日目

レポート書き

6) レポート提出

7) レポート再提出

レポート

注意点: 自分が身についた技術を他人に報告
＜他人が読む文章＞
＜自己満足にならないこと＞

1) texで記述. (章だてに便利)

2) 章構成

他人が読むとき, 非常に便利.
基本 ”はじめに” ”本文” ”まとめ”

(例)

a)はじめに, b)実験の目的, c)実験の条件 d)実験の結果,
e)考察, f)実験の感想 g)まとめ, h)参考文献

(b-dは”はじめに” ”本文” ”まとめ”で繰り返す.)
(特に重要 ”はじめに”, ”まとめ”)

日本語の文章を書く上で参考になる本

日本語の作文技術 本田勝一 朝日文庫 ISBN4-02-260808-0 540円
実戦・日本語の作文技術 本田勝一 朝日文庫

参考:論文

“はじめに” “本文” ”まとめ”

1)過去の研究 2)問題点 3)仮説
7)考察 8)今後の課題

4)実験 5)実験の詳細 6)結果

3) 図はOpenOfficeにて

4) 付録

A) ネットワーク文法

0) hand trace

1) Shell

2) makefile

3) ソースプログラム

4) 実行結果 解析途中の結果

(正常終了, 異常終了ともに)

B) 文脈自由文法

以下, 同様

5) 提出期限

次回の別の実験が始まる週の水曜日の13時

印刷物 1504の前

メール murakami@eecs.tottori-u.ac.jp

考察の内容

2) I saw a girl with a telescope には大きくわけて2つの意味(解釈)がある.

a) この解釈を述べよ.

b) 2つの解釈をするために加える文法規則は何か?

c) 2種類の構文解析結果

4) Time flies like an arrow と Fruit flies like an apple の

a) 翻訳の違い

b) 構文解析の違い

c) 文法規則と単語をどのようにすればいいか?

d) プログラムで1位に選択するにはどうすればいいか?

6) Time flies like an arrow をコンピュータで解釈した場合の多義数