

# 自然言語処理論 I

## 3.文法1(文脈自由文法と句構造文法)

1

## 文脈自由文法(復習)

- 書き換え規則の集合
  - $A \rightarrow B C D$
  - 左辺は非終端記号1個
  - 右辺は任意の長さの記号列
- 文法Gは4つ組 $\langle V_N, V_T, P, \sigma \rangle$ で定義

2

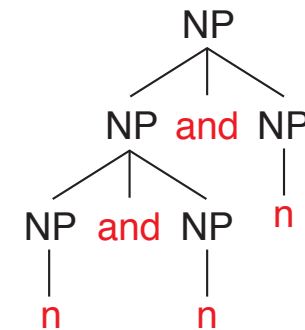
## 導出(derivation)

- 書き換え規則の適用過程を表す
  - 文法
    - ◆  $NP \rightarrow NP \text{ and } NP$
    - ◆  $NP \rightarrow n$
  - 導出
    - $NP \Rightarrow NP \text{ and } NP$
    - $\Rightarrow NP \text{ and } n$
    - $\Rightarrow NP \text{ and } NP \text{ and } n$
    - $\Rightarrow n \text{ and } NP \text{ and } n$
    - $\Rightarrow n \text{ and } n \text{ and } n$

3

## 導出木(derivation tree)

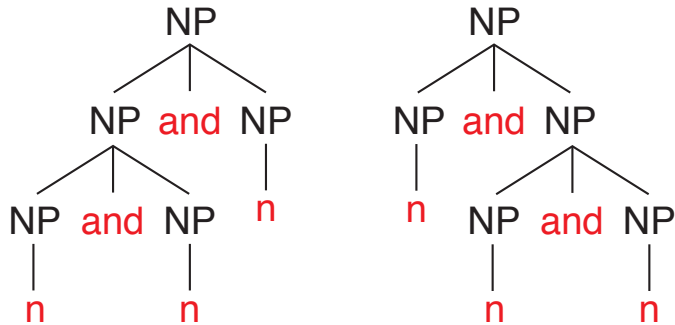
- 導出を木構造で表現したもの



4

## 導出木

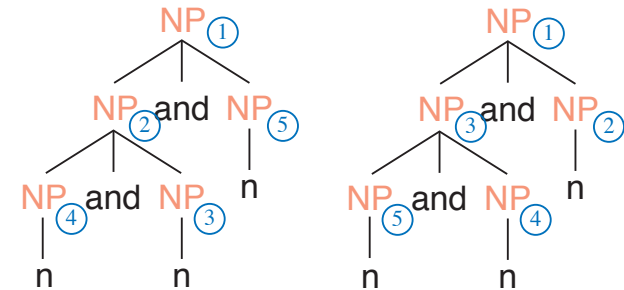
- 一つの終端記号列に対し、導出や導出木は一般に複数存在する



5

## 導出と導出木

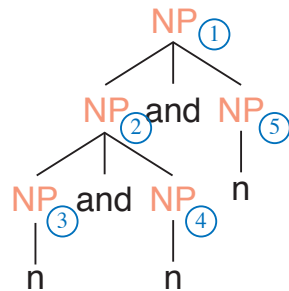
- 1つの導出木に対する導出は複数存在
  - 非終端記号を展開する順序は複数ある



6

## 最左導出

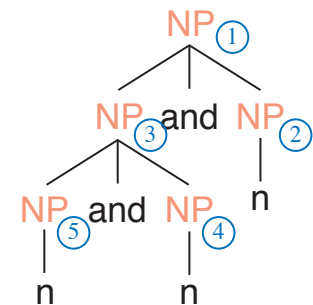
- left most derivation
  - 記号列の最も左にある非終端記号を常に先に展開する導出
  - NP ⇒ NP and NP
    - ⇒ NP and NP and NP
    - ⇒ n and NP and NP
    - ⇒ n and n and NP
    - ⇒ n and n and n



7

## 最右導出

- right most derivation
  - 記号列の最も右にある非終端記号を常に先に展開する導出
  - NP ⇒ NP and NP
    - ⇒ NP and n
    - ⇒ NP and NP and n
    - ⇒ NP and n and n
    - ⇒ n and n and n



8

## ここまでのまとめ

- 1つの文字列に対して、導出木は一般に複数ある
  - 曖昧性(ambiguity)を持つ
- 1つの導出木に対して、導出は一般に複数ある
- 1つの導出木に対して、最左導出(最右導出)は1つしかない

9

## 文脈自由文法の標準形

- チョムスキー標準形
  - Chomsky Normal Form
  - 書き換え規則は以下の2通りのいずれか
    - ◆  $A \rightarrow BC$
    - ◆  $A \rightarrow a$   
(A,B,Cは非終端記号, aは終端記号)
  - 全てのCFGはチョムスキー標準形に変換できる

10

## チョムスキー標準形への変換

- 変換手続
  - 右辺に終端記号が2つ以上あるとき、終端記号を非終端記号に置き換える
    - ◆ (変換前)  $A \rightarrow bcD$
    - ◆ (変換後)  $A \rightarrow X_b X_c D \quad X_b \rightarrow b \quad X_c \rightarrow c$
  - 右辺に非終端記号が2つ以上あるとき、新しい非終端記号を導入して右辺長を2にする
    - ◆ (変換前)  $A \rightarrow BCDE$
    - ◆ (変換後)  $A \rightarrow Y_1 E \quad Y_1 \rightarrow Y_2 D \quad Y_2 \rightarrow BC$
  - 新しく導入する非終端記号( $X_i, Y_i$ )は、他とは違う記号にすることに注意

11

## チョムスキー標準形への変換

- 例

$S \rightarrow ABCD$		$S \rightarrow Y_1 D$	$D \rightarrow X_d D$
$A \rightarrow a$		$Y_1 \rightarrow Y_2 C$	$D \rightarrow d$
$B \rightarrow b$		$Y_2 \rightarrow AB$	$X_c \rightarrow c$
$B \rightarrow CD$	⇒	$A \rightarrow a$	$X_d \rightarrow d$
$C \rightarrow cD$		$B \rightarrow b$	
$D \rightarrow dD$		$B \rightarrow CD$	
$D \rightarrow d$		$C \rightarrow X_c D$	

12

# 句構造文法

- 自然言語の構造を表す文脈自由文法
- 英語の句構造文法の例

S → NP VP	VP → v	pron → I
NP → pron	VP → v NP	det → a
NP → det n	VP → VP PP	v → broke
NP → NP PP	PP → prep NP	prep → with
		n → desk   drawer

記号の意味

S(sentence), NP(noun phrase), VP(verb phrase),  
PP(prepositional phrase), n(noun), v(verb), pron(pronoun),  
prep(preposition), det(determiner)

# 構文木(parse tree)

- 句構造文法による導出木

