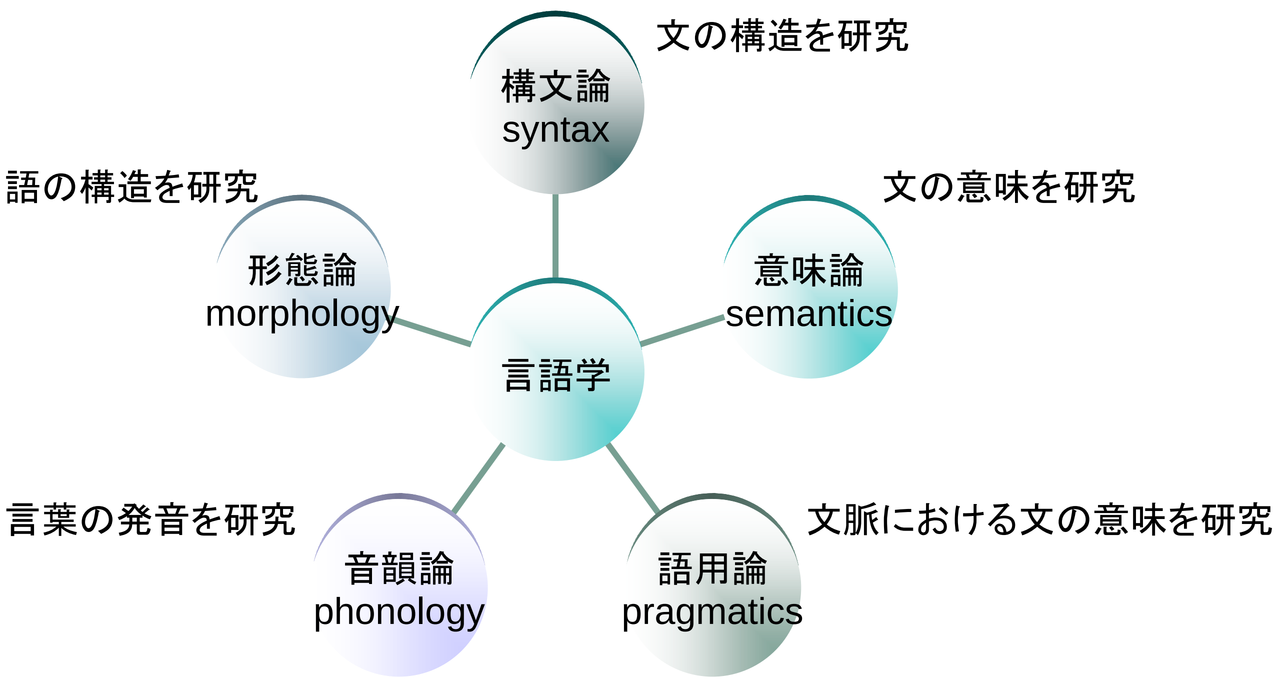
第3章
「文脈自由言語」

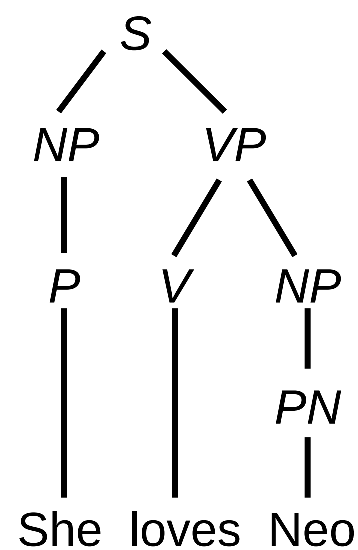
3.1 文脈自由文法
3.2 プッシュダウンオートマトン
3.3 文脈自由文法とプッシュダウンオートマトン
3.4 正規文法と有限オートマトン
3.5 文脈自由文法の性質
3.6 決定性プッシュダウンオートマトン

形式言語と自然言語

1950年代に米国の言語学者N. Chomskyが形式言語理論を提唱。形式文法とそれによって定義される言語を自然言語の数学的モデルとして研究。



She loves Neo



※ 上図写真:http://web.media.mit.edu/~nitin/mideast/chomsky.html

3.1 文脈自由文法

定義

- 文脈自由文法とは4つ組
  $$G = (N, \Sigma, P, S)$$
  によって定義される。
- N: 空でない有限集合。Nの要素を非終端記号という。
  Σ: 空でない有限集合。Σ の要素を終端記号という。
  S: $S \in N$で開始記号という。
  P: P は $N \times (N \cup \Sigma)^*$ の有限部分集合。
  - P の要素 (A, α) は生成規則とよばれ、$A \rightarrow \alpha$ と書かれる。
  - $\alpha = \varepsilon$ のとき ε 生成規則という。
  - 非終端記号A を生成規則の左側にもつ P の要素を
    $A \rightarrow \alpha_1, \ldots, A \rightarrow \alpha_n$ とするとき、これらをまとめて
    $A \rightarrow \alpha_1 | \ldots | \alpha_n$ と書くこともある。

文脈自由文法の例

- 例3.1 (例3.3)
  - $G=(N, \Sigma, P, S)$
    - $N = \{S\}$
    - $\Sigma = \{a, b\}$
    - $P = \{S \to ab, S \to aSb\}$
- 例3.2 (例3.4)
  - $G=(N, \Sigma, P, S)$
    - $N = \{S, A, B\}$
    - $\Sigma = \{x, 0, 1\}$
    - $P = \{S \to xA, A \to 0|1B, B \to \varepsilon|0B|1B\}$

語の導出

導出の定義

- $V = N \cup \Sigma$ とする。
  記号列 $u, v \in V^*$ が次を満たすとき $u \Rightarrow_G v$ とかく。
  G は省略することもある
  (1) $u = xAy$, $v = x\alpha y$　$(x, y, \alpha \in V^*, A \in N)$
  (2) $A \to \alpha$ は G の生成規則
- $\Rightarrow_G$ の反射的推移閉包を $\overset{*}{\Rightarrow}_G$ とかく。
- 導出の長さが n の場合、$\overset{n}{\Rightarrow}_G$ とかく。
- V^* の要素の列 $w_0, \ldots, w_n$ について、
  $w_0 \Rightarrow_G w_1 \Rightarrow_G \cdots \Rightarrow_G w_n$ のとき、w_0 から w_n が**導出される**という。

Gが生成する語、言語

- 開始記号Sより $w \in \Sigma^*$ が G の生成規則によって導出されるとき、w は G の生成する語とよばれる。また G の生成する言語を
$$L(G) = \{ w \in \Sigma^* \mid S \overset{*}{\Rightarrow}_G w \}$$
と書く。
- 言語 $L \subseteq \Sigma^*$ に対して $L = L(G)$ となるとき、G は L を生成するという。
- Σ 上の言語 $L \subseteq S^*$ に対して、L を生成する文脈自由文法が存在するとき、L は文脈自由言語とよばれる。

文脈自由言語とCFL

CFL の定義

- 文脈自由言語のクラスをCFLと書くことにする。
すなわち、

集合の集合（族）

CFL =
　{ L | L = L(G) となる文脈自由文法 G がある}

と定義する。

導出の例

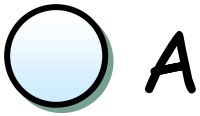
- 例3.5
  - $N = \{S, A, B\}$, $\Sigma = \{x, 0, 1, +, \cdot,), (\}$
  - $P = \{S \rightarrow (S + S) \mid S \cdot S \mid xA,$
    $A \rightarrow \varepsilon \mid 1B,\ B \rightarrow \varepsilon \mid 0B \mid 1B \mid 0 \mid 1\}$
- 例3.6
  - $N = \{S\}$, $\Sigma = \{), (\}$
  - $P = \{S \rightarrow SS,\ S \rightarrow (S),\ S \rightarrow ()\}$
- 例3.7
  - $N = \{S\}$, $\Sigma = \{0, 1, \phi, +, \cdot, *,), (\}$
  - $P = \{S \rightarrow \phi \mid 0 \mid 1 \mid (S+S) \mid (S \cdot S) \mid S^*\}$

構文木 導出の過程をVisualizeする道具

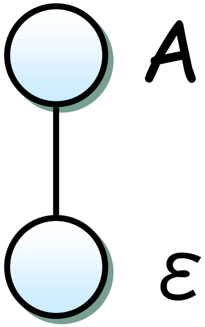
定義

- 構文木は次のように帰納的に定義される。ここで、V = N∪S とする。

(1) 各$A \in V$に対して、記号Aをラベルとする1つの頂点のみからなる木



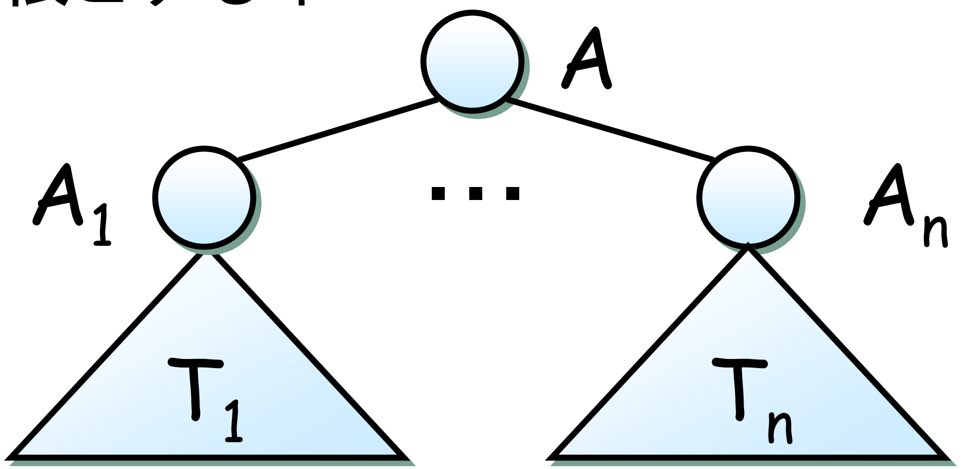
(2) 生成規則 $A \rightarrow \varepsilon$ に対して



構文木(2)

定義つづき

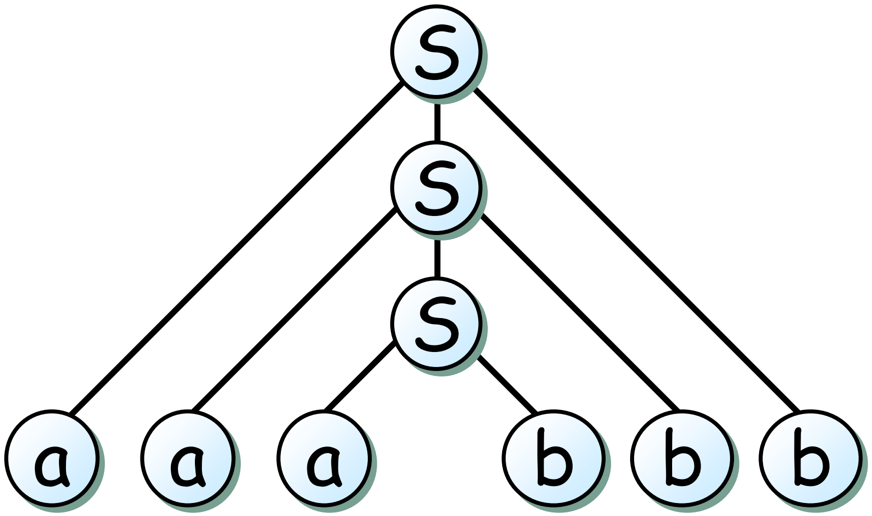
(3) $T_1, \dots, T_n$ を根のラベルが $A_1, \dots, A_n \in V$ である構文木とする。
G の生成規則 $A \rightarrow A_1 \dots A_n$ に対して、
Aを根とする木



(4) 上記(1)～(3)の規則を使って定義される有限の木のみを構文木という。

構文木の例

- 例3.8 (例3.1の導出木)



L(G) はどのような言語か？

最左導出

- 最左導出と左側順走査
  - 導出の各ステップにおいて一番左にある非終端記号が置き換えられているとき、最左導出という。
  - 最左導出 $A \overset{*}{\Rightarrow} u$ に対する構文木の走査は左側順走査となる。

- 例3.9 (例3.6の最左導出)
  - 練習: 語 (())(()()) を最左導出する過程は？

補題3.1

補題3.1

- $G=(N,S,P,S)$を文脈自由文法とする。
  導出 $u \overset{*}{\Rightarrow} v$ があるならば、
  u から v への最左導出がある。

証明

- $u = w_1A_1w_2 \dots w_kA_kw_{k+1}$ $(A_i \in N,\ w_j \in \Sigma^*)$ とする。

$$
\begin{array}{l}
u = w_1 A_1 w_2 A_2 w_3 \dots w_k A_k w_{k+1} \\
\qquad\quad \downarrow ① \quad \downarrow ② \quad \cdots \quad \downarrow ⓚ \\
v = w_1 v_1 w_2 v_2 w_3 \dots w_k v_k w_{k+1}
\end{array}
$$

$A_i \overset{*}{\Rightarrow} v_i \ (1 \leqq i \leqq k)$

左側順走査による
最左導出とする