

文法と語の生成

言語は文法で定義されるということを形式化する

水谷 正大

Masahiro Mizutani

簡単な例

文 $\Rightarrow$ 主語 + 述語

主語 $\rightarrow$ 犬

述語 $\rightarrow$ 泳ぐ

主語 $\rightarrow$ 魚

述語 $\rightarrow$ 駆ける

正規文法

書き換え規則

導出例

文 $\Rightarrow$ 主語 + 述語 $\Rightarrow$ 犬 + 述語 $\Rightarrow$ 犬 + 駆ける

文 $\Rightarrow$ 主語 + 述語 $\Rightarrow$ 魚 + 述語 $\Rightarrow$ 魚 + 泳ぐ

文 $\Rightarrow$ 主語 + 述語 $\Rightarrow$ 犬 + 述語 $\Rightarrow$ 犬 + 泳ぐ

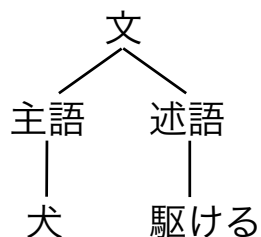
文 $\Rightarrow$ 主語 + 述語 $\Rightarrow$ 魚 + 述語 $\Rightarrow$ 魚 + 駆ける

非終端記号(non terminal symbols) 文、主語、述語

終端記号(terminal symbols) 犬、魚、泳ぐ、駆ける

Masahiro Mizutani

文 $\Rightarrow$ 主語 + 述語 $\Rightarrow$ 犬 + 述語 $\Rightarrow$ 犬 + 駆ける



導出木 (derivation tree)

Masahiro Mizutani

無限に導出できる例

書き換え規則

正規文法

(正確な日本語文法表記ではない)

文 $\Rightarrow$ 主語 + 述語 述語 $\Rightarrow$ 動詞句

主語 $\Rightarrow$ 名詞句

名詞句 $\Rightarrow$ 名詞 | 形容詞句 + 名詞

形容詞句 $\Rightarrow$ 形容詞 | 形容詞 + 形容詞句

動詞句 $\Rightarrow$ 動詞 | 副詞句 + 動詞

副詞句 $\Rightarrow$ 副詞 | 副詞 + 副詞句

記号“|” は or を表す

Masahiro Mizutani

導出例

名詞句 + 駆ける $\Rightarrow$ 犬駆ける

形容詞句 + 犬 $\Rightarrow$ 白い犬

形容詞句 + 犬

$\Rightarrow$ 白い + 形容詞句 + 犬

$\Rightarrow$ 白い + 美しい犬

$\Rightarrow$ 白い + (美しい + 形容詞句)犬

$\Rightarrow$ 白い美しい大きい犬

$\Rightarrow$ 白い美しい(大きい + 形容詞句) 犬

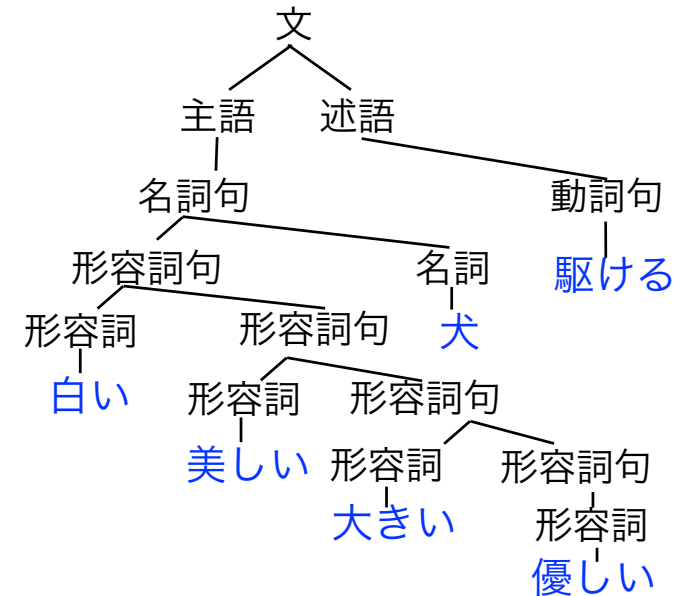
$\Rightarrow$ 白い美しい大きい優しい犬

$\Rightarrow$ 白い美しい大きい(優しい + 形容詞句) 犬

⋮

Masahiro Mizutani

「白い美しい大きい優しい犬駆ける」の導出木



Masahiro Mizutani

文法Gの形式的定義

$$G = (V_N, V_T, S, R)$$

V_N 非終端記号集合 文には現れないが、導出の際に中間的に使われる記号。自然言語では、品詞カテゴリに相当。

V_T 終端記号集合 文に現れる記号集合で この記号からの書き換え規則はない。自然言語では個々の単語に相当

S 開始記号 $S \in V_N$ で、文を導出するための最初の記号

R 書き換え規則集合 $\alpha \Rightarrow \beta$ の形の集合
 α, β は $V_N \cup V_T$ の有限長記号列

Masahiro Mizutani

導出(derivation)と言語

A^* 集合AのKleene閉包(closure)

集合Aの要素を0回ないし有限回繰り返した語

$$A^* = \{x_1 x_2 \dots x_k \mid k \geq 0, \forall x_i \in A\}$$

$\alpha, \beta, \gamma \in (V_N \cup V_T)^*$ とする

α に書き換え規則を1回だけ適用して β が得られるとき、導出

$$\alpha \Rightarrow \beta$$

書き換え規則を複数回適用して γ が得られるとき、導出

$$\alpha \stackrel{*}{\Rightarrow} \gamma$$

$\Rightarrow$ の閉包、つまり直接導出を0回または任意の有限回適応して α から γ が得られることを意味

文法が定める言語

ある文法で開始記号から導出される文の集合

Masahiro Mizutani

正規文法

書き換え規則が次の2種類に限られている文法

$$\begin{array}{ll} X \rightarrow aY & X, Y \text{ は任意の非終端記号} \\ X \rightarrow a & a \text{ は任意の終端記号} \end{array}$$

正規文法は、同一の終端非終端記号が繰り返し現れるような書き換え規則があれば無限の文を導出できる。
しかし、その構造は限定されている

Masahiro Mizutani

文脈自由文法の例

$$G = (V_N = \{S\}, V_T = \{0, 1\}, S, R = \{S \rightarrow 0S1 \mid \varepsilon\}) \quad \varepsilon \text{ は空語}$$

導出例

$$\begin{aligned} S &\Rightarrow \varepsilon \\ S &\Rightarrow 0S1 \Rightarrow 0\varepsilon 1 = 01 \\ S &\Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 0011 \\ S &\Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 000S111 \Rightarrow 000111 \\ &\vdots \\ S &\stackrel{*}{\Rightarrow} 0^n S 1^n = 0^n 1^n \end{aligned}$$

文法Gは文脈自由言語 $\{0^k 1^k \mid k \geq 0\}$ を定義する

Masahiro Mizutani

文脈自由文法

書き換え規則

$$\begin{array}{ll} A \rightarrow \alpha & A \text{ は任意の非終端記号} \\ & \alpha \in (V_N \cup V_T)^* \\ & \text{終端記号と非終端記号の任意の記号列} \end{array}$$

Chomskyの標準形

$$\begin{array}{ll} X \rightarrow YZ & X, Y, Z \in V_N \\ X \rightarrow a & a \in V_T \end{array}$$

文脈自由文法も無限の文を導出でき、その構造は正規言語よりも複雑になり得る

Masahiro Mizutani

統語解析(parsing/syntactic analysis)

導出とは逆に、ある文を入力として与え、何らかの文法（統語論）に基づいてその記号列の統語構造を構文木(parsing tree)として出力すること

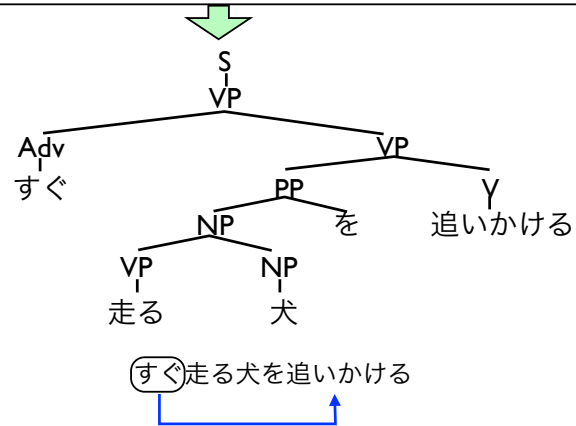
Q. すぐ走る犬を追いかける

Masahiro Mizutani

「すぐ走る犬を追いかける」を解析

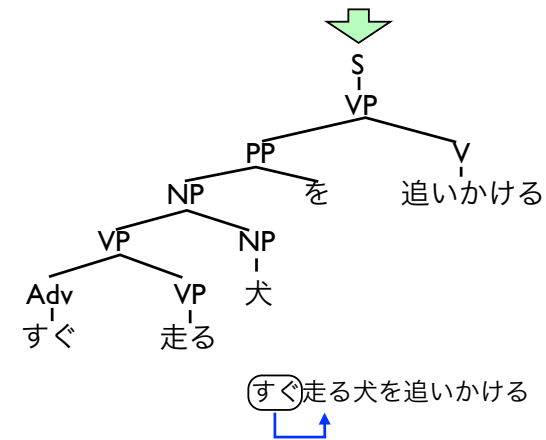
文脈自由文法

$S \rightarrow VP$	$VP \rightarrow Adv + VP$	$VP \rightarrow PP + V$
$PP \rightarrow NP + を$	$NP \rightarrow VP + NP$	$Adv \rightarrow \text{すぐ}$
$VP \rightarrow \text{走る}$	$NP \rightarrow \text{犬}$	$V \rightarrow \text{追いかける}$



Masahiro Mizutani

別の統語構造もある



言語の理解には統語論的知識以外にもさまざまな知識
が必要になる

Masahiro Mizutani