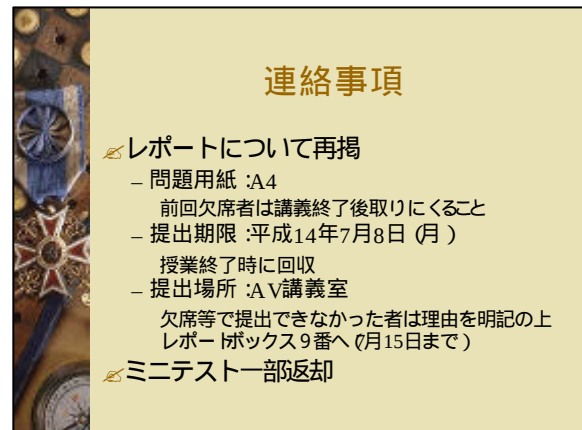




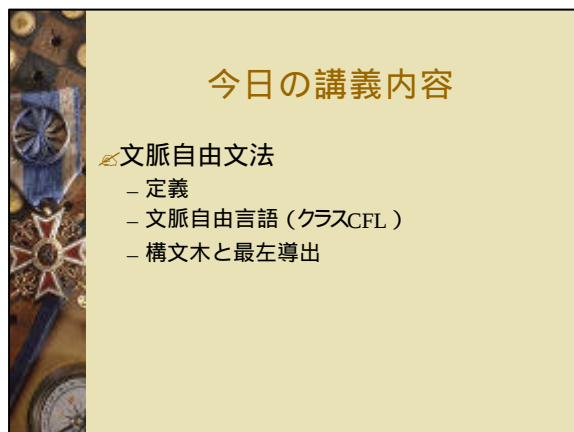
計算の理論 I 文脈自由文法

月曜3校時
大月美佳



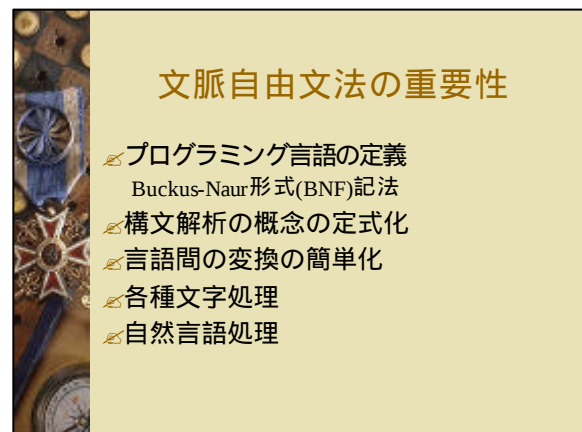
連絡事項

- レポートについて再掲
 - 問題用紙 :A4
 - 前回欠席者は講義終了後取りにくること
 - 提出期限 :平成14年7月8日 (月)
 - 授業終了時に回収
 - 提出場所 :A \講義室
 - 欠席等で提出できなかった者は理由を明記の上レポートボックス9番へ(7月15日まで)
- ミニテスト一部返却



今日の講義内容

- 文脈自由文法
 - 定義
 - 文脈自由言語 (クラスCFL)
 - 構文木と最左導出



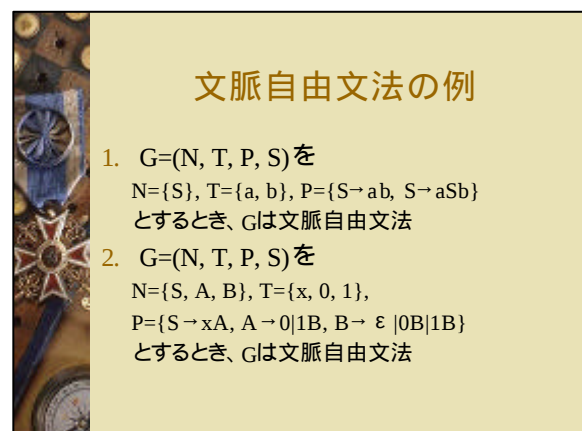
文脈自由文法的重要性

- プログラミング言語の定義
 - Buckus-Naur形式(BNF)記法
- 構文解析の概念の定式化
- 言語間の変換の簡単化
- 各種文字処理
- 自然言語処理



文脈自由文法の定義

- 文脈自由文法(context-free grammar)
 - 非終端アルファベット N (有限集合)
 - 要素 $x \in N$ 非終端記号(non-terminal symbol)
 - 終端アルファベット T (有限集合)
 - 要素 $x \in T$ 終端記号(terminal symbol)
 - 開始記号(start symbol) $S \in N$
 - 生成規則(production) $P \subseteq N \times (N \cup T)^*$
 - $A \rightarrow \alpha$ と書く ($\alpha = \varepsilon$ のとき ε 生成規則)
 - $A \rightarrow \alpha_1 \dots \alpha_n \mid \alpha_n = A \rightarrow \alpha_1 \dots \alpha_n, A \rightarrow \alpha_n$
- $G = (N, T, P, S)$



文脈自由文法の例

- $G = (N, T, P, S)$ を
 $N = \{S\}, T = \{a, b\}, P = \{S \rightarrow ab, S \rightarrow aSb\}$
とすると、 G は文脈自由文法
- $G = (N, T, P, S)$ を
 $N = \{S, A, B\}, T = \{x, 0, 1\},$
 $P = \{S \rightarrow xA, A \rightarrow 0|1B, B \rightarrow \varepsilon | 0B|1B\}$
とすると、 G は文脈自由文法

G

- ✎ $G=(N, T, P, S)$ に対して、 $V=N \cup T$ とする。
- ✎ u, v が次の(1), (2)を満たすとき $u \xrightarrow{G} v$ と書くことにする。
 (1) $u=xAy, v=x\alpha y$ ($x, y, \alpha \in V^*, A \in N$)。
 (2) $A \rightarrow \alpha$ は G の生成規則。

導出 (derive)

- ✎ $\xrightarrow{G}^*$
 G の反射的かつ推移的閉包
- ✎ V^* の要素の列 $w_0, \dots, w_n$ が
 $w_0 \xrightarrow{G} w_1 \xrightarrow{G} \dots \xrightarrow{G} w_n$
 となっているとき、
 w_0 から w_n が導出されるといい、
 $w_0 \xrightarrow{G^*} w_n$ または $w_0 \xrightarrow{n} w_n$
 と書く。 $(G$ は省略可: $\xrightarrow{\quad}, \xrightarrow{*}, \xrightarrow{n})$

文脈自由言語 (context-free language)

- ✎ G の生成する語 w
 開始記号 S より
 終端アルファベット T 上の記号列 w が
 G の生成規則によって導出されるとき
- ✎ T 上の言語 $L \subseteq T^*$ に対して
 1. G が L を生成する: $L=L(G)$ となるとき。
 2. 文脈自由言語: L を生成する文脈自由文法が存在する。

文脈自由言語の例 その1

- G を例1の文脈自由文法としたとき、
 $S \rightarrow aSb \mid aaSbb \mid \dots \mid a^{n-1}Sb^{n-1} \mid a^nb^n$
 となり、
 $L(G)=\{a^nb^n \mid n \geq 1\}$
 である。
- G を例2の文脈自由文法としたとき、
 $S \rightarrow xAx \mid x0$
 $S \rightarrow xAx \mid x1B \mid x10B \mid x10$
 となり、
 $L(G)=\{xu \mid u=1v, v \in \{0, 1\}^* \text{ または } u=0\}$

文脈自由言語の例 その2

- $G=(N, T, P, S)$ を
 $N=\{S\}, T=\{(), \mid\}, P=\{S \rightarrow SS, S \rightarrow (S), S \rightarrow ()\}$
 とすると、言語 $L(G)$ はかっこの入れ子構造
- $G=(N, T, P, S)$ を
 $N=\{S\}, T=\{0, 1, \Phi, +, *, \mid\}, \{\},$
 $P=\{S \rightarrow \Phi \mid 01(S+S) \mid SS \mid S^*\}$
 とすると、言語 $L(G)$ は $\{0, 1\}$ 上の正規表現全体

文脈自由文法のクラス

- ✎ 文脈自由文法のクラスを **CFL** と書く
- ✎ **CFL**
 $=\{L \mid L=L(G) \text{ となる文脈自由文法 } G \text{ がある}\}$

グラフ (1.2節 pp. 2 ~ 5)

どんなときに使うのか？

- 状態遷移図
- 文法木
- 各種アルゴリズムのデータ構造
- その他いろいろ

グラフの定義

グラフ $G=(V, E)$

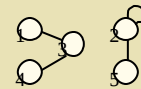
V : 有限個の頂点(vertex, node)の集合

E : 頂点の対 $((v_1, v_2)$ と表記)で示される辺(edge)の集合

例 (図1.1 p.3)

$V=\{1, 2, 3, 4, 5\}$

$E=\{(n, m)|n+m=4 \text{ または, } n+m=7\}$



道

道 (path)、路

- グラフのある頂点の列 $v_1, v_2, \dots, v_k$ ($k \geq 1$) が道であるというのは
- $(v_1, v_2), (v_2, v_3), \dots, (v_{k-1}, v_k)$ がいずれも辺であるということ

閉路 (cycle) : $v_1=v_k$ のとき

道の例 (図1.1 p.3)

- 1, 3, 4
- 2
- 2, 5



有向グラフ

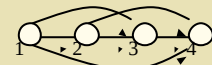
有向グラフ (directed graph, digraph) G

- $G=(V, E)$
- E の要素が有向辺(arc)
- $v \rightarrow w$: v から w へ向かう有向辺

前者 (predecessor) 後者 (successor)

例 (図1.2 p.3)

$(\{1, 2, 3, 4\}, \{i \rightarrow j | i < j\})$



有向グラフの道

有向グラフの道 (path)

$v_1 \rightarrow v_{i+1} \ (1 \leq i < k)$ が有向辺であるような頂点の列 $v_1, v_2, \dots, v_k$ (ただし, $k \geq 1$)

例 (図1.2 p.3)

$1 \rightarrow 2 \rightarrow 3 \rightarrow 4$: 1 から 4 への道



木

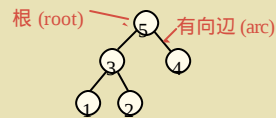
木 (tree, ordered directed tree)

次の性質を持つ有効グラフ

1. 前者を持たず、各頂点への道が必ず存在する根 (root) と呼ばれる頂点を持つ
2. 根以外の頂点はそれぞれただ一つ前者を持つ
3. 各頂点の後者は左から右へ一列に順序つけられている

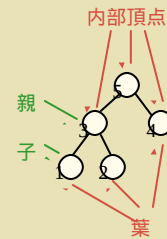
木の書き方

- 根を上、各有向辺を下に向けて書く
- 有向辺の矢印は書く必要がない
- 頂点は (なんらかの) 順序に従って左から右に書く

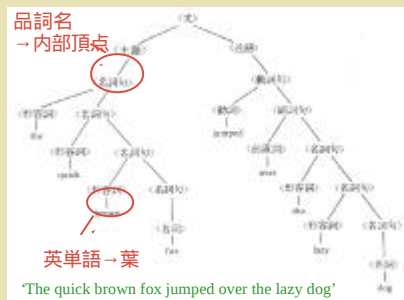


木特有の用語

- 親 (parent, father?): 前者
- 子 (child, son?): 後者
- 葉 (leaf): 子を持たない頂点
- 内部 (interior) 頂点: 葉でない頂点
- 先祖 (ancestor) と子孫 (descendant)
 - 頂点 v_i から v_j への道があるとき (v_i : 先祖, v_j : 子孫)
 - 各頂点は自分自身の先祖かつ子孫



構文木



構文木 (parse tree)

- 文脈自由文法 $G=(N, T, P, S)$ に対して、構文木を帰納的に定義する。
- ここで、 $V=N \cup T$ とする。

構文木の定義(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



は構文木である。

構文木の定義(2)

- $A \rightarrow \epsilon$ に対して、



は構文木である。

構文木の定義(3)

(3) $t_1, \dots, t_n$ を根のラベルが、 $A_1, \dots, A_n$ 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

```
graph TD; A((A)) --- A1((A1)); A --- An((An)); A1 --- t1([t1]); An --- tn([tn]);
```

は構文木である。

A tree diagram illustrating a hierarchical structure. At the top is a root node labeled A . Below A are two nodes, A_1 and A_n , connected by an ellipsis $\dots$. Each node A_i is connected to a triangle labeled t_i .



構文木の定義(4)

(4) 上の(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。

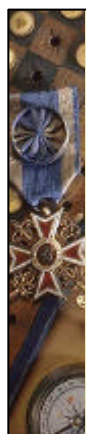
構文木の例

$G=(N, T, P, S)$ を
 $N=\{S\}$, $T=\{a, b\}$, $P=\{S \rightarrow aSb, S \rightarrow aSb\}$
としたとき、構文木は右図。

```
graph TD; S1[S] --- a1[a]; S1 --- S2[S]; S1 --- b1[b]; S2 --- a2[a]; S2 --- S3[S]; S2 --- b2[b]; S3 --- a3[a]; S3 --- S4[S]; S3 --- b3[b]; S4 --- a4[a]; S4 --- S5[S]; S4 --- b4[b]; S5 --- a5[a]; S5 --- S6[S]; S5 --- b5[b]; S6 --- a6[a]; S6 --- S7[S]; S6 --- b6[b]; S7 --- a7[a]; S7 --- S8[S]; S7 --- b7[b]; S8 --- a8[a]; S8 --- S9[S]; S8 --- b8[b]; S9 --- a9[a]; S9 --- S10[S]; S9 --- b9[b]; S10 --- a10[a]; S10 --- S11[S]; S10 --- b10[b]; S11 --- a11[a]; S11 --- S12[S]; S11 --- b11[b]; S12 --- a12[a]; S12 --- S13[S]; S12 --- b12[b]; S13 --- a13[a]; S13 --- S14[S]; S13 --- b13[b]; S14 --- a14[a]; S14 --- S15[S]; S14 --- b14[b]; S15 --- a15[a]; S15 --- S16[S]; S15 --- b15[b]; S16 --- a16[a]; S16 --- S17[S]; S16 --- b16[b]; S17 --- a17[a]; S17 --- S18[S]; S17 --- b17[b]; S18 --- a18[a]; S18 --- S19[S]; S18 --- b18[b]; S19 --- a19[a]; S19 --- S20[S]; S19 --- b19[b]; S20 --- a20[a]; S20 --- S21[S]; S20 --- b20[b]; S21 --- a21[a]; S21 --- S22[S]; S21 --- b21[b]; S22 --- a22[a]; S22 --- S23[S]; S22 --- b22[b]; S23 --- a23[a]; S23 --- S24[S]; S23 --- b24[b]; S24 --- a24[a]; S24 --- S25[S]; S24 --- b25[b]; S25 --- a25[a]; S25 --- S26[S]; S25 --- b26[b]; S26 --- a26[a]; S26 --- S27[S]; S26 --- b27[b]; S27 --- a27[a]; S27 --- S28[S]; S27 --- b28[b]; S28 --- a28[a]; S28 --- S29[S]; S28 --- b29[b]; S29 --- a29[a]; S29 --- S30[S]; S29 --- b30[b]; S30 --- a30[a]; S30 --- S31[S]; S30 --- b31[b]; S31 --- a31[a]; S31 --- S32[S]; S31 --- b32[b]; S32 --- a32[a]; S32 --- S33[S]; S32 --- b33[b]; S33 --- a33[a]; S33 --- S34[S]; S33 --- b34[b]; S34 --- a34[a]; S34 --- S35[S]; S34 --- b35[b]; S35 --- a35[a]; S35 --- S36[S]; S35 --- b36[b]; S36 --- a36[a]; S36 --- S37[S]; S36 --- b37[b]; S37 --- a37[a]; S37 --- S38[S]; S37 --- b38[b]; S38 --- a38[a]; S38 --- S39[S]; S38 --- b39[b]; S39 --- a39[a]; S39 --- S40[S]; S39 --- b40[b]; S40 --- a40[a]; S40 --- S41[S]; S40 --- b41[b]; S41 --- a41[a]; S41 --- S42[S]; S41 --- b42[b]; S42 --- a42[a]; S42 --- S43[S]; S42 --- b43[b]; S43 --- a43[a]; S43 --- S44[S]; S43 --- b44[b]; S44 --- a44[a]; S44 --- S45[S]; S44 --- b45[b]; S45 --- a45[a]; S45 --- S46[S]; S45 --- b46[b]; S46 --- a46[a]; S46 --- S47[S]; S46 --- b47[b]; S47 --- a47[a]; S47 --- S48[S]; S47 --- b48[b]; S48 --- a48[a]; S48 --- S49[S]; S48 --- b49[b]; S49 --- a49[a]; S49 --- S50[S]; S49 --- b50[b]; S50 --- a50[a]; S50 --- S51[S]; S50 --- b51[b]; S51 --- a51[a]; S51 --- S52[S]; S51 --- b52[b]; S52 --- a52[a]; S52 --- S53[S]; S52 --- b53[b]; S53 --- a53[a]; S53 --- S54[S]; S53 --- b54[b]; S54 --- a54[a]; S54 --- S55[S]; S54 --- b55[b]; S55 --- a55[a]; S55 --- S56[S]; S55 --- b56[b]; S56 --- a56[a]; S56 --- S57[S]; S56 --- b57[b]; S57 --- a57[a]; S57 --- S58[S]; S57 --- b58[b]; S58 --- a58[a]; S58 --- S59[S]; S58 --- b59[b]; S59 --- a59[a]; S59 --- S60[S]; S59 --- b60[b]; S60 --- a60[a]; S60 --- S61[S]; S60 --- b61[b]; S61 --- a61[a]; S61 --- S62[S]; S61 --- b62[b]; S62 --- a62[a]; S62 --- S63[S]; S62 --- b63[b]; S63 --- a63[a]; S63 --- S64[S]; S63 --- b64[b]; S64 --- a64[a]; S64 --- S65[S]; S64 --- b65[b]; S65 --- a65[a]; S65 --- S66[S]; S65 --- b66[b]; S66 --- a66[a]; S66 --- S67[S]; S66 --- b67[b]; S67 --- a67[a]; S67 --- S68[S]; S67 --- b68[b]; S68 --- a68[a]; S68 --- S69[S]; S68 --- b69[b]; S69 --- a69[a]; S69 --- S70[S]; S69 --- b70[b]; S70 --- a70[a]; S70 --- S71[S]; S70 --- b71[b]; S71 --- a71[a]; S71 --- S72[S]; S71 --- b72[b]; S72 --- a72[a]; S72 --- S73[S]; S72 --- b73[b]; S73 --- a73[a]; S73 --- S74[S]; S73 --- b74[b]; S74 --- a74[a]; S74 --- S75[S]; S74 --- b75[b]; S75 --- a75[a]; S75 --- S76[S]; S75 --- b76[b]; S76 --- a76[a]; S76 --- S77[S]; S76 --- b77[b]; S77 --- a77[a]; S77 --- S78[S]; S77 --- b78[b]; S78 --- a78[a]; S78 --- S79[S]; S78 --- b79[b]; S79 --- a79[a]; S79 --- S80[S]; S79 --- b80[b]; S80 --- a80[a]; S80 --- S81[S]; S80 --- b81[b]; S81 --- a81[a]; S81 --- S82[S]; S81 --- b82[b]; S82 --- a82[a]; S82 --- S83[S]; S82 --- b83[b]; S83 --- a83[a]; S83 --- S84[S]; S83 --- b84[b]; S84 --- a84[a]; S84 --- S85[S]; S84 --- b85[b]; S85 --- a85[a]; S85 --- S86[S]; S85 --- b86[b]; S86 --- a86[a]; S86 --- S87[S]; S86 --- b87[b]; S87 --- a87[a]; S87 --- S88[S]; S87 --- b88[b]; S88 --- a88[a]; S88 --- S89[S]; S88 --- b89[b]; S89 --- a89[a]; S89 --- S90[S]; S89 --- b90[b]; S90 --- a90[a]; S90 --- S91[S]; S90 --- b91[b]; S91 --- a91[a]; S91 --- S92[S]; S91 --- b92[b]; S92 --- a92[a]; S92 --- S93[S]; S92 --- b93[b]; S93 --- a93[a]; S93 --- S94[S]; S93 --- b94[b]; S94 --- a94[a]; S94 --- S95[S]; S94 --- b95[b]; S95 --- a95[a]; S95 --- S96[S]; S95 --- b96[b]; S96 --- a96[a]; S96 --- S97[S]; S96 --- b97[b]; S97 --- a97[a]; S97 --- S98[S]; S97 --- b98[b]; S98 --- a98[a]; S98 --- S99[S]; S98 --- b99[b]; S99 --- a99[a]; S99 --- S100[S]; S99 --- b100[b]; S100 --- a100[a]; S100 --- S101[S]; S100 --- b101[b]; S101 --- a101[a]; S101 --- S102[S]; S101 --- b102[b]; S102 --- a102[a]; S102 --- S103[S]; S102 --- b103[b]; S103 --- a103[a]; S103 --- S104[S]; S103 --- b104[b]; S104 --- a104[a]; S104 --- S105[S]; S104 --- b105[b]; S105 --- a105[a]; S105 --- S106[S]; S105 --- b106[b]; S106 --- a106[a]; S106 --- S107[S]; S106 --- b107[b]; S107 --- a107[a]; S107 --- S108[S]; S107 --- b108[b]; S108 --- a108[a]; S108 --- S109[S]; S108 --- b109[b]; S109 --- a109[a]; S109 --- S110[S]; S109 --- b110[b]; S110 --- a110[a]; S110 --- S111[S]; S110 --- b111[b]; S111 --- a111[a]; S111 --- S112[S]; S111 --- b112[b]; S112 --- a112[a]; S112 --- S113[S]; S112 --- b113[b]; S113 --- a113[a]; S113 --- S114[S]; S113 --- b114[b]; S114 --- a114[a]; S114 --- S115[S]; S114 --- b115[b]; S115 --- a115[a]; S115 --- S116[S]; S115 --- b116[b]; S116 --- a116[a]; S116 --- S117[S]; S116 --- b117[b]; S117 --- a117[a]; S117 --- S118[S]; S117 --- b118[b]; S118 --- a118[a]; S118 --- S119[S]; S118 --- b119[b]; S119 --- a119[a]; S119 --- S120[S]; S119 --- b120[b]; S120 --- a120[a]; S120 --- S121[S]; S120 --- b121[b]; S121 --- a121[a]; S121 --- S122[S]; S121 --- b122[b]; S122 --- a122[a]; S122 --- S123[S]; S122 --- b123[b]; S123 --- a123[a]; S123 --- S124[S]; S123 --- b124[b]; S124 --- a124[a]; S124 --- S125[S]; S124 --- b125[b]; S125 --- a125[a]; S125 --- S126[S]; S125 --- b126[b]; S126 --- a126[a]; S126 --- S127[S]; S126 --- b127[b]; S127 --- a127[a]; S127 --- S128[S]; S127 --- b128[b]; S128 --- a128[a]; S128 --- S129[S]; S128 --- b129[b]; S129 --- a129[a]; S129 --- S130[S]; S129 --- b130[b]; S130 --- a130[a]; S130 --- S131[S]; S130 --- b131[b]; S131 --- a131[a]; S131 --- S132[S]; S131 --- b132[b]; S132 --- a132[a]; S132 --- S133[S]; S132 --- b133[b]; S133 --- a133[a]; S133 --- S134[S]; S133 --- b134[b]; S134 --- a134[a]; S134 --- S135[S]; S134 --- b135[b]; S135 --- a135[a]; S135 --- S136[S]; S135 --- b136[b]; S136 --- a136[a]; S136 --- S137[S]; S136 --- b137[b]; S137 --- a137[a]; S137 --- S138[S]; S137 --- b138[b]; S138 --- a138[a]; S138 --- S139[S]; S138 --- b139[b]; S139 --- a139[a]; S139 --- S14
```

又右图。

a a a b b b



最左導出

(leftmost derivation)

➤ 導出 u^*v 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、その導出は最左導出である。

- $G = (N, T, P, S)$ を
 $N = \{S\}$, $T = \{\}$, $\{\}$, $P = \{S \rightarrow SS, S \rightarrow (S), S \rightarrow ()\}$
 とすると
 $S \quad S() \quad ((SS)) \quad (((()S)) \quad (((()())$
 は最左導出。

[illegible]

走査 (traverse)

((()))

補題

➤ $G=(N, T, P, S)$ を文脈自由文法とする。導出 $u \rightarrow v$ があるならば、 u から v への最左導出がある。

証明

$u = w_1 A_1 w_2 \dots w_k A_k w_{k+1} (A_i \in N, w_j \in T^*)$ とする。

このとき、 v の分解 $v = w_1 v_1 w_2 \dots w_k v_k w_{k+1}$ があって、 $A_i \rightarrow v_i (1 \leq i \leq k)$ となる。

そこで、 $A_1 \rightarrow v_1$ に対する構文木を取り、その左側順走査を考えれば、 A_i から v_i への最左導出が得られる。

それらの最左導出を、非終端記号の順に適用していけば求める最左導出が得られる。

5

最右導出

(rightmost derivation)

- 最左導出と反対に右から置き換えていく導出

$G=(N, T, P, S)$ を

$N=\{S\}, T=\{(), ()\}, P=\{S \rightarrow SS, S \rightarrow (S), S \rightarrow ()\}$

とすると

$S \ (S) \ ((S)) \ (((S))) \ ((((S))))$

は最左導出。

曖昧性

- 曖昧

ある語に二つ以上の構文木があるとき

– 二つ以上の最左導出をもつ

または

– 二つ以上の最右導出をもつ

とき

- 本質的に曖昧 → 4.7 (p.128)

今日のミニテスト

- ミニテスト

教科書・資料を見ても良い

- 資料、ミニテストがない人は前へ

- 交換採点を行い、提出したら帰ってよし

- 類題

演習問題4.8 (p.135)