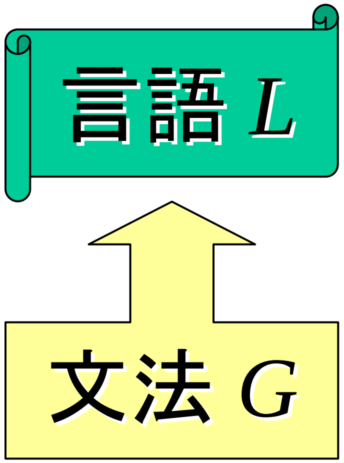
5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4. 文法と言語の曖昧さ

CFLで言えば、<u>複数の構文木を持つ</u>



言語 = 文法的に正しい語の集合
- 各語は、文法による<u>構造</u>を持つ

<u>文法が曖昧</u>
⇔ある語が文法上‘正しい構造’を複数持つ

言語によっては文法の曖昧さを除くことができる。しかし、本質的に曖昧な言語もある。

★CFL L で、「$L(G)=L$ を満たすどんな CFG G」も曖昧な文法になるものがある。

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4. 文法と言語の曖昧さ

– いくつかの違ったレベルの曖昧さ

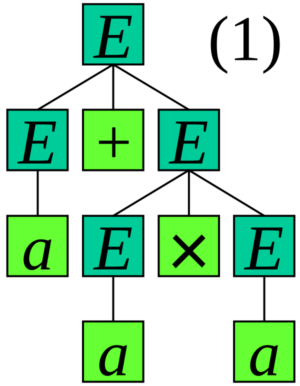
1. 文法の構成を工夫すれば回避できる
   その言語に対して、上手に構成してやれば回避可能

2. 文法に付加的なルールを想定すれば回避できる
   文法＋αで回避可能

3. 本質的に曖昧
   言語 L を表現するどんな文法も曖昧になってしまう
   ⇒言語 L は本質的に曖昧である、と言う。

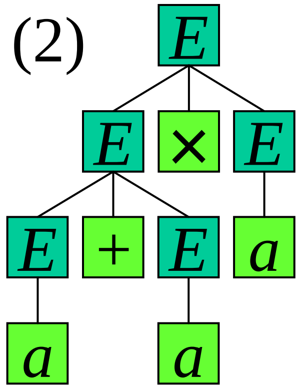
5. 文脈自由文法と言語(3): (テキスト5.4)

(1)



5.4.1. 曖昧な文法

例) $G_1=(\{E\},\{+,\times,a\}, E\to E+E \mid E\times E \mid a, E)$ (2)

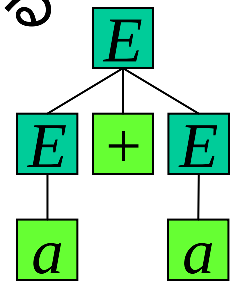


- 文形式 $a+a\times a$ の本質的に違う導出

(1) $E \Rightarrow E+E \Rightarrow E+E\times E \overset{*}{\Rightarrow} a+a\times a$

(2) $E \Rightarrow E\times E \Rightarrow E+E\times E \overset{*}{\Rightarrow} a+a\times a$

注) 導出が違っていても、本質的に同じ構造もある



$E \Rightarrow E+E \Rightarrow a+E \Rightarrow a+a$

$E \Rightarrow E+E \Rightarrow E+a \Rightarrow a+a$

★<u>本質的に違う導出</u>＝<u>構文木の形</u>が違う

5. 文脈自由文法と言語(3): (テキスト5.4)

5.4.1. 曖昧な文法

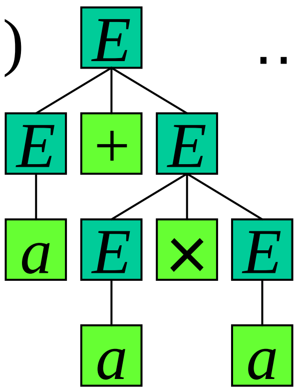
例) $G_1=(\{E\},\{+,\times,a\}, E\rightarrow E+E \mid E\times E \mid a, E)$

- 文形式 $a+a\times a$ の本質的に違う導出

$(1) E \Rightarrow E+E \Rightarrow E+E\times E \overset{*}{\Rightarrow} a+a\times a$

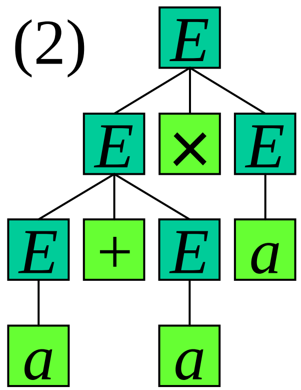
$(2)\ E \Rightarrow E\times E \Rightarrow E+E\times E \overset{*}{\Rightarrow} a+a\times a$

(1)



…式 $a+(a\times a)$ に対応

(2)



…式 $(a+a)\times a$ に対応

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4.1. 曖昧な文法

$G_1=(\{E\},\{+,\times,a\}, E\rightarrow E+E \mid E\times E \mid a, E)$ の曖昧さ

1. 演算子[+]と[×]の優先順位が表現できない

   (1)$E \Rightarrow E+E \Rightarrow E+E\times E \overset{*}{\Rightarrow} a+\underline{a\times a}$ ←○

   (2)$E \Rightarrow E\times E \Rightarrow E+E\times E \overset{*}{\Rightarrow} \underline{a+a}\times a$

2. 同じ演算子内の順番が不定

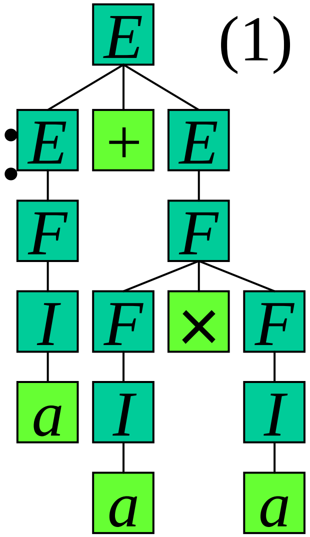
   (1)$E \Rightarrow E+E \Rightarrow E+E+E \overset{*}{\Rightarrow} a+\underline{a+a}$

   (2)$E \Rightarrow E+E \Rightarrow E+E+E \overset{*}{\Rightarrow} \underline{a+a}+a$ ←○

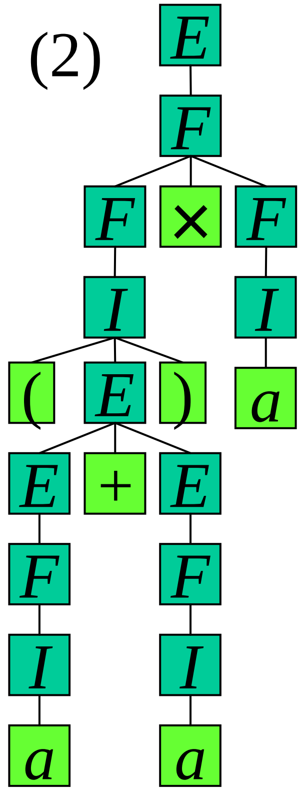
3. (導出の曖昧さ)

5. 文脈自由文法と言語(3):
(テキスト5.4)

(1)



(2)



5.4.2. 文法の曖昧さの除去

1. 演算子[+]と[×]の優先順位を表現する

「式」をもっと細かく定義しなおす;

1. 因数(I): 識別子 a または括弧で囲まれた式
2. 項(F): 因数の積、つまり因数を×でつないだもの
3. 式(E): 項の和、つまり項を + でつないだもの

$$G_2=(\{I,F,E\},\{+,\times,a,(,)\}, A, E) \quad A:\begin{cases} I \to a \mid (E) \\ F \to F \times F \mid I \\ E \to E + E \mid F \end{cases}$$

$E \Rightarrow F \Rightarrow I$の順でしか展開できない

$$(1)\ E \Rightarrow E+E \Rightarrow E+F \overset{*}{\Rightarrow} F+I \times I \overset{*}{\Rightarrow} a+a \times a$$

$$(2)\ E \Rightarrow F \Rightarrow F \times F \overset{*}{\Rightarrow} (E+E) \times F \overset{*}{\Rightarrow} (a+a) \times a$$

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4.2. 文法の曖昧さの除去

2. 同じ演算子内の順番を決める
   [左を優先するための変数]を入れる: 例えば

   $E \Rightarrow E+E \mid \alpha$

   は以下の変形をする。

   $F \Rightarrow \alpha$

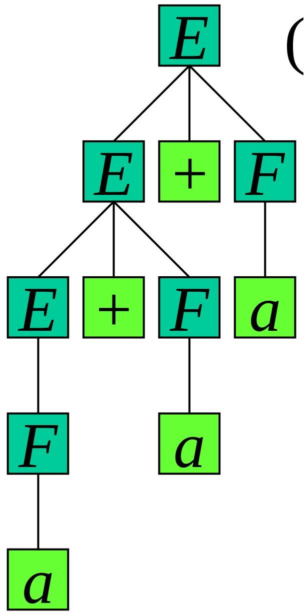
   $E \Rightarrow E+F \mid F$

左にしか伸展できない

G_3=({F,E},{+,×,a}, A, E)

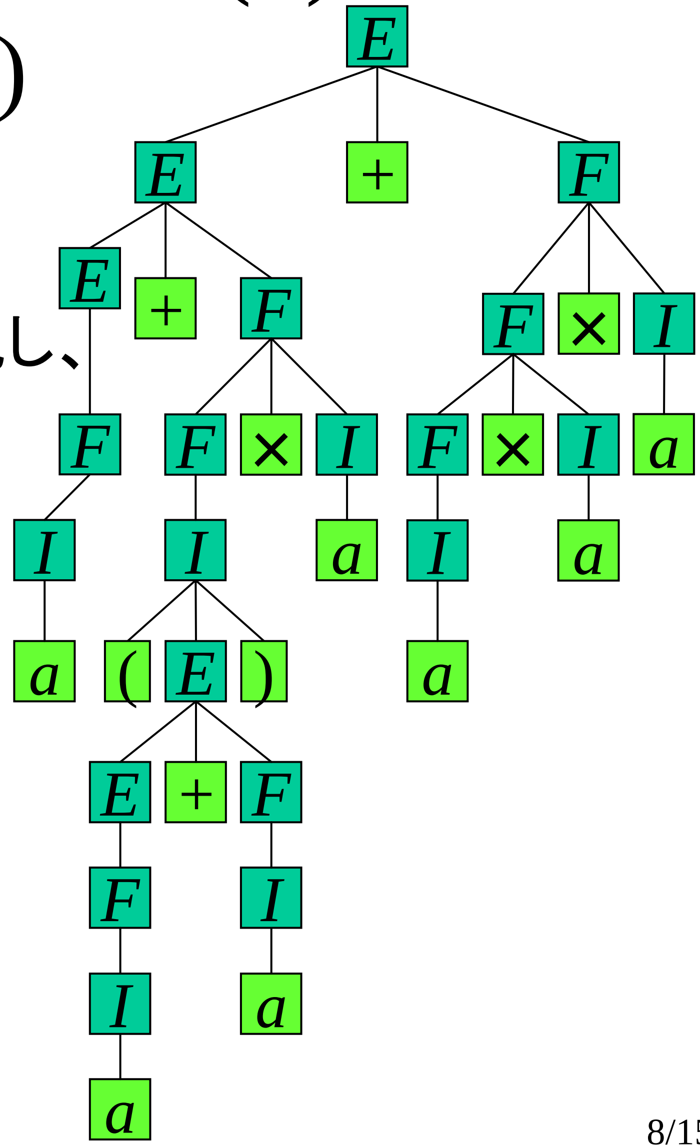
$$A:\begin{cases} F \to a \\ E \to E+F \mid E\times F \mid F \end{cases}$$

(1) $E \Rightarrow E+F \Rightarrow E+F+F \Rightarrow F+F+F \overset{*}{\Rightarrow} a+a+a$



(1)

5. 文脈自由文法と言語(3):
(テキスト5.4)



5.4.2. 文法の曖昧さの除去

1. 演算子[+]と[×]の優先順位を表現し、
2. 左優先を表現する

G_4=({I,F,E},{+,×,a,(,)}, A, E)

$$A:\begin{cases} I \rightarrow a \mid (E) \\ F \rightarrow F \times I \mid I \\ E \rightarrow E + F \mid F \end{cases}$$

例) $E \overset{*}{\Rightarrow} a+(a+a)\times a+a\times a\times a$

に対する構文木は右のものだけ

5.4.3. 曖昧さを表現する手段としての最左導出

3. 導出に関する曖昧さをなくす

   …最左導出によって、構文木の「たどり方」は一意的に決まる。

[定理] 語の構文木の個数と最左導出の個数は同じ

1,2 の対策によって、与えられた「式」の構文木は一意的に決まる。

3 の対策によって、導出の順番に関する曖昧さはなくなる。

実用上、多くのCFLは上記の方法で曖昧でないCFGを構築することができる。

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4. 文法と言語の曖昧さ

5.4.4. 本質的な曖昧さ

言語 L が本質的に曖昧 ⇔ L を表現する任意の文法が曖昧

- ✓ CFLは本質的に曖昧な言語を含む
- ✓ 与えられた言語が本質的に曖昧かどうかは、計算によって判定することはできない

⇒ここでは本質的に曖昧な言語の例を示すにとどめる

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4. 文法と言語の曖昧さ

5.4.4. 本質的な曖昧さ

言語 $L = \{ a^n b^n c^m d^m \mid n \geqq 1, m \geqq 1\}$

$\cup \{ a^n b^m c^m d^n \mid n \geqq 1, m \geqq 1\}$

L は、$aa^*bb^*cc^*dd^*$ の部分集合で、

- a と b が同数かつ c と d が同数、または
- a と d が同数かつ b と c が同数

であるような語の集合

5. 文脈自由文法と言語(3): (テキスト5.4)

5.4. 文法と言語の曖昧さ

5.4.4. 本質的な曖昧さ

言語 $L=\{a^n b^n c^m d^m\} \cup \{a^n b^m c^m d^n\}$ を表現する文法の例:

$G=(\{S,A,B,C,D\},\{a,b,c,d\},X,S)$

$$
X: \begin{cases}
S \rightarrow AB \mid C \\
A \rightarrow aAb \mid ab \\
B \rightarrow cBd \mid cd \\
C \rightarrow aCd \mid aDd \\
D \rightarrow bDc \mid bc
\end{cases}
$$

前者/後者の別

語 $a^k b^k c^k d^k$ は???

5. 文脈自由文法と言語(3):
(テキスト5.4)

5.4. 文法と言語の曖昧さ

5.4.4. 本質的な曖昧さ

[定理] 言語 $L=\{a^n b^n c^m d^m\} \cup \{a^n b^m c^m d^n\}$ は本質的に曖昧である

[証明のアイデア]

L を表現するどんな文法も語 $a^k b^k c^k d^k$ に対しては複数の構文木を持つことを示す。$a^n b^n c^m d^m$ を生成する規則と、$a^n b^m c^m d^n$ を生成する規則の両方とも $a^k b^k c^k d^k$ を生成せざるをえない。

5.4. 文法と言語の曖昧さ

5.4.4. 本質的な曖昧さ

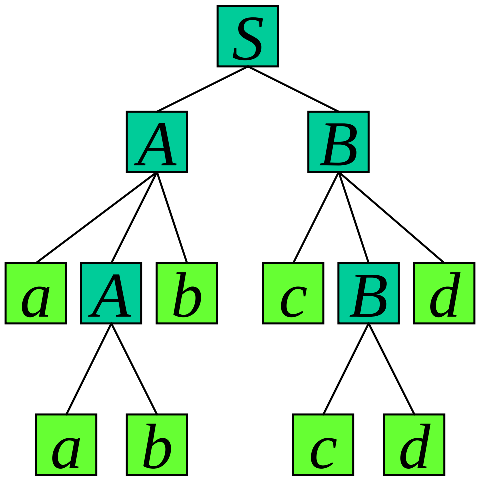
言語 $L=\{a^n b^n c^m d^m\} \cup \{a^n b^m c^m d^n\}$ を表現する文法の例:

$G=(\{S,A,B,C,D\},\{a,b,c,d\},X,S)$

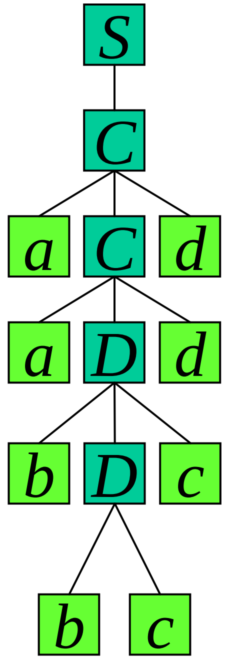
$$X:\begin{cases} S \to AB \mid C \\ A \to aAb \mid ab \\ B \to cBd \mid cd \\ C \to aCd \mid aDd \\ D \to bDc \mid bc \end{cases}$$

語 *aabbccdd* の2つの構文木

$a^n b^n c^m d^m$ の要素と見たとき



$a^n b^m c^m d^n$ の要素と見たとき



5. 文脈自由文法と言語(3): 演習問題(7)

1. 言語 L を以下に定義する。

$$L = \{a^n b^n c^m \mid n > 0, m > 0\} \cup \{a^n b^m c^n \mid n > 0, m > 0\}$$

   1. Lを生成する文脈自由文法 G を構成せよ。
   2. 1.で構成した G が曖昧であることを示せ。