計算理論　第12回
第7章：
文脈自由言語の性質 (1/3)

基礎工学部情報科学科
中川　博之

本日の概要

- 第7章：文脈自由言語の性質
  - テキスト：p.283～
  - 7.1 文脈自由言語の標準形

- 重要概念
  - CFGの単純化，Chomsky標準形（次週説明）

7.1 文脈自由言語の標準形

文脈自由文法の等価変換

- 文法を単純化する3つの変換手法
  - 無用な記号の除去
  - ε-規則の除去
  - 単位規則の除去
- Chomsky標準形（次週説明）への変換のために用いる

無用な記号の除去

- 有用，無用とは？

- 文法G=(V, T, P, S)の有用な記号 X (∈V∪T)

  → $\exists w \in L(G)$, $S \overset{*}{\Rightarrow} \alpha X \beta \overset{*}{\Rightarrow} w$ のような導出が存在する

  – 文字列wを生成する際に出現する（文字列wを生成するのに貢献している）

無用な記号

- 文法G=(V, T, P, S)の無用な記号X (∈V∪T)

  → ∀w∈L(G)において
  $S \overset{*}{\Rightarrow} \alpha X \beta \overset{*}{\Rightarrow} w$ のような導出が存在しない
  - どの文字列wを生成する際にも出現しない（文字列生成に全く貢献していない）

- 無用な記号は全く使われないので，そのような記号を除去しても生成される言語は変わらない

記号Xが有用である2つの条件

- Xは生成的(generating) : $X \overset{*}{\Rightarrow} w$ ($\in T^*$)
  - Xから終端記号列が得られる
  - wに含まれる終端記号自身も生成的
  - 生成的でなければ，語の生成に役立っていない
- Xは到達可能(reachable) : $S \overset{*}{\Rightarrow} \alpha X \beta$
  - 開始記号からたどり着ける
  - 到達可能でなければ，Xの出番が無い
- Xは有用 : 上記2条件がともに満たされる
- Xは無用 : それ以外の場合

有用な記号の求め方

- １．生成的でない記号を除去
- ２．到達可能でない記号を除去

例7.1

- 文法G=(V, T, P, S)
  - S→AB |a
  - A→b
- V={S, A, B}, T={a, b}
  - a, b: 自身を生成する
  - S : aを生成，A : bを生成
  - Bは生成的でない　　　→ Bを除去する

例7.1

- 変数集合VからBを除去
  - Bを含む生成規則(S→AB)も除去
- 新たな文法G’=(V’, T, P’, S)
  - S→a
  - A→b

AはSから到達可能でない！

- 変数集合VからAも除去
  - 生成規則(A→b)も除去
- 新たな文法G’’=(V’’, T, P’’, S)
  - S→a

有用な記号の求め方(逆順)

- 先ほどと逆の手順
- 1’. 到達可能でない記号を除去
- 2’. 生成的でない記号を除去

上手くいかない場合あり

例7.1 (逆順)

- 文法G=(V, T, P, S)
  - S→AB |a
  - A→b
- 1’．到達可能でない記号を除去

  → すべての記号が到達可能（除去しない）
- 2’．生成的でない記号を除去
  - Bが生成的でないので，Bと規則S→AB を除去
- 得られた文法（規則）
  - S→ a
  - A→ b

不要なものが残っている

定理7.2

- CFG G=(V, T, P, S) ただし，L(G)≠ϕ
- $G_1=(V_1,T_1,P_1,S)$を次の手順で得られる文法とする
  - (1) Gから生成的でない記号とそれを含む生成規則を除去．得られた文法をG_2とする
    - L(G) ≠ϕなのでSは除去されない
  - (2) G_2から到達可能でない記号とそれを含む生成規則を除去
- こうして得られたG_1は無用な記号を持たず，

$$L(G_1)=L(G)$$

生成的　到達可能

$G \rightarrow G_2 \rightarrow G_1$

定理7.2の証明

- 以下の2つを示す必要あり
  - G_1に残った記号はいずれも有用
  - $L(G_1)=L(G)$
- 記号X ($\in V_1 \cup T_1$)が最終的に残ったとすると
  - 第1ステップで除去されていないので生成的
  - つまり，$\exists w \in T^*$: $X \overset{*}{\underset{G}{\Rightarrow}} w$
  - この導出は，現れる記号がすべて生成的なので G_2の導出でもある．つまり，$X \overset{*}{\underset{G_2}{\Rightarrow}} w$

生成的　到達可能

$G \rightarrow G_2 \rightarrow G_1$

定理7.2の証明:有用性

- 記号Xは第2ステップでも除去されていないのでG_2のSから到達可能
  - つまり，$S \overset{*}{\underset{G_2}{\Rightarrow}} \alpha X \beta$ となるα, βが存在
  - この導出中に現れる記号もすべてG_2で到達可能であるので，上記導出はG_1の導出でもある．つまり，$S \overset{*}{\underset{G_1}{\Rightarrow}} \alpha X \beta$

XはG_1において到達可能

生成的　到達可能

$G \rightarrow G_2 \rightarrow G_1$

定理7.2の証明：有用性

- $\alpha X\beta$に現れる記号はすべてG_1で到達可能
  - しかも，$V_2 \cup T_2$に属するのでG_2で生成的
- $\alpha X\beta \overset{*}{\underset{G_2}{\Rightarrow}} xwy$ に含まれる記号はすべてG_2で到達可能．つまり，この導出はG_1の導出でもある．

$\rightarrow \alpha X\beta \overset{*}{\underset{G_1}{\Rightarrow}} xwy$　　XはG_1において生成的

- 前スライドの結論と併せて

$$S \overset{*}{\underset{G_1}{\Rightarrow}} \alpha X\beta \overset{*}{\underset{G_1}{\Rightarrow}} xwy$$

- つまり，すべてのX ($\in V_1 \cup T_1$)がG1で有用
  - 無用な記号を持たない

生成的　到達可能

$G \rightarrow G_2 \rightarrow G_1$

定理7.2の証明：$L(G_1)=L(G)$

- $L(G_1) \subseteq L(G)$
  - G_1はGから記号と生成規則を除去したものなので，所属する語は増えない
- $L(G_1) \supseteq L(G)$
  - $w \in L(G)$なら，$S \overset{*}{\underset{G}{\Rightarrow}} w$
  - この導出中に現れる記号はいずれも生成的かつ到達可能なのでG_1にも含まれるため，この導出はG_1の導出でもある．つまり，$S \overset{*}{\underset{G_1}{\Rightarrow}} w$
  - 従って，$w \in L(G_1)$

7.1.2 生成的記号と到達可能記号の計算

生成的記号の計算方法

- 文法G=(V, T, P, S)の生成的記号集合Z(⊆V∪T)の計算手順
- 基礎:
  - Tの各要素をZに加える
    - Tの各要素は自分自身を生成する
- 帰納:
  - 生成規則A→αにおいて，αの全記号がZに属すとき，AをZに加える
    - A→εの場合も加える
  - これをZに追加できる記号が無くなるまで繰り返す

例7.3

- （再掲）例7.1の文法G=(V, T, P, S)

S→AB | a, A→b

- 基礎：a,bはそれ自身で生成的 → Z={a, b}
- 帰納：
  - A→bにおいてbは生成的なので，ZにAを追加
  - S→aにおいてaは生成的なので，ZにSを追加

- 結論：生成記号の集合 Z={a, b, A, S}

到達可能記号の計算方法

- 文法G=(V, T, P, S)の到達可能記号集合Z’(⊆V∪T)の計算手順

- 基礎：
  - SをZ’に加える
    - Sは開始記号であり，ステップ0で到達可能
- 帰納：
  - 記号AがZ’に属していれば，生成規則A→αの右辺αの各記号をZ’に加える
  - これをZ’に追加できる記号が無くなるまで繰り返す

例7.5

- （再掲）例7.1の文法G=(V, T, P, S)

S→AB | a, A→b

- 基礎：Sは到達可能　Z’={S}
- 帰納：
  - S∈Z’かつS→AB|a よりZ’={S, A, B, a}
  - A∈Z’かつA→b よりZ’={S, A, B, a, b}
- 結論：到達可能記号の集合 Z’={S, A, B, a, b}

ミニレポート: 12-1

- テキスト p.299　問7.1.1:
  - 次の生成規則により定義される文法と等価で無用な記号を持たない文法を求めよ.
  - $S \to AB \mid CA$
  - $A \to a$
  - $B \to BC \mid AB$
  - $C \to aB \mid b$

7.1.3 ε-規則の除去

ε-規則

- A→ε の形の生成規則のこと
  - 文法の設計上で非常に便利
- ε-規則は必須ではない
  - ε-規則を含まない等価な文法が存在
  - ただし言語がεを含まない場合
- 任意の文脈自由言語Lに対し，L-{ε} を生成するε-規則なしのCFGが存在

ε-規則の除去

- 方針：「消去可能」な変数を発見する
- 変数Aが消去可能： $A \overset{*}{\Rightarrow} \varepsilon$
  - 変数Aを文法から除去するわけではない
- 以下の場合は注意
  - 生成規則：B→CAD
  - 変数Aは消去可能（$A \overset{*}{\Rightarrow} \varepsilon$）
- この場合，生成規則から単純にAを削除してはいけない
  - Aがε以外を導出する可能性もあるため

ε-規則除去の基本方針

- 生成規則：B→CAD
- 変数Aは消去可能（$A \overset{*}{\Rightarrow} \varepsilon$）

[方針] 2つの規則を作る

- 1) 生成規則B→CD を追加
  - Aからεへの導出の場合に相当
- 2) 生成規則B→CAD を追加（残す）
  - Aからε以外の導出の場合に相当
- 生成規則 A→ε は除去

では，どのように消去可能な変数を見つける？

消去可能変数の決定法

- 基礎：$A \to \varepsilon$がGの規則ならば，Aは消去可能
- 帰納：$B \to C_1 C_2 \ldots C_k$がGの規則で，各C_iがすべて消去可能ならば，Bは消去可能
  - 帰納段階では，右辺が変数のみの生成規則を調べればよい

例7.8:ε-規則の除去

- S → AB
- A → aAA | ε
- B → bBB | ε

- 消去可能な変数を求める
  - A, B: ε-規則を持つため消去可能
  - S: 生成規則S→ABにて，A, Bが消去可能であるため，Sも消去可能

例7.8:ε-規則の除去

- S→AB に注目
  - 変数A，Bともに消去可能
  - 新たな文法G’ に規則 S→AB | A | B を追加
- A→aAA に注目
  - 変数Aは消去可能
  - 文法G’に規則 A→aAA | aA | a を追加
- A→εに注目
  - ε-規則なので，G’には加えない
- B→bBB | ε も同様

文法G’ の生成規則

- S → AB | A | B
- A → aAA | aA | a
- B → bBB | bB | b

まとめ：ε-規則の除去方法

- CFG G=(V, T, P, S)からε-規則を除去したG1= (V, T, P_1, S)の構成方法
  1. Gの消去可能変数を求める
  2. Gの生成規則 A→ε：
     P_1に追加しない
  3. Gの生成規則 A→α (αは消去可能変数を含まない)：
     P_1に追加
  4. Gの生成規則 A→α（|α|≥1）：
     右辺の消去可能変数を（含む/含まない）の組み合わせすべてに対する生成規則をP_1に追加.
     ただし A→εは追加しない

7.1.4 単位規則の除去

単位規則 (unit production)

- A→B (A, B : 変数) の形の生成規則
  - 文法設計の際には有益

- ただし，単位規則は冗長
  - CFG Lは単位規則を含まない文法で構築可能

- 単位規則を含むCFGと等価な，
  単位規則を含まないCFGが存在

単位規則除去の基本方針

- 単位規則を含む文法Gから，含まない文法G’を構成
- $A \Rightarrow B_1 \Rightarrow B_2 \Rightarrow \ldots \Rightarrow B_k \Rightarrow \alpha$ (各B_iは変数)のとき，生成規則$A \rightarrow \alpha$をG’に追加
  - 単位規則なしでAから直接αを導出
- 単位対(unit pair) (A,B)
  - 単位規則だけを使って$A \overset{*}{\Rightarrow} B$となるもの
  - ただし，A,Bは変数
- 単位規則の除去は，単位対集合に基づいて実施

単位対集合の構成方法

- 基礎: (A, A)は単位対
  - A $\overset{*}{\Rightarrow}$ A は長さ0の導出であるため
- 帰納: (A, B)が単位対かつB→Cが単位規則なら，(A, C)は単位対
  - A $\overset{*}{\Rightarrow}$ Bと規則B→Cにより，A $\overset{*}{\Rightarrow}$ Cであるため

例7.10

- 例5.27の数式の文法
  - E → T | E+T
  - T → F | T*F
  - F → I | (E)
  - I → a|b|Ia|Ib|I0|I1

単位対は？

- 基礎:
  - (E, E), (T, T), (F, F), (I, I)
- 帰納:
  - (E, T): (E, E) とE→Tより
  - (T, F): (T, T)とT→Fより
  - (F, I): (F, F)とF→Iより
  - (E, F): (E, T)とT→Fより
  - (T, I): (T, F)とF→Iより
  - (E, I): (E, F)とF→Iより

単位規則の除去方法

- 方針：A⇒B⇒αのとき，生成規則A→αを追加
  - 生成規則A→Bを除去

- CFG G=(V, T, P, S)から，単位規則を除去したG1=(V, T, P_1, S)の構成法
  1. Gのすべての単位対を求める
  2. すべての単位対(A, B)とGの非単位規則B→αに対し，規則A→αをP_1に追加
     - 単位規則をこのタイミングで除去

例7.12：単位規則の除去

- 文法
  - E → T | E+T
  - T → F | T*F
  - F → I | (E)
  - I → a|b|Ia|Ib|I0|I1

単位対	追加する規則
(E, E)	E→E+T
(T, T)	T→T*F
(F, F)	F→(E)
(I, I)	I → a|b|Ia|Ib|I0|I1
(E, T)	E→T*F
(E, F)	E→(E)
(E, I)	E→a|b|Ia|Ib|I0|I1
(T, F)	T→(E)
(T, I)	T→a|b|Ia|Ib|I0|I1
(F, I)	F→a|b|Ia|Ib|I0|I1

例7.12:単位規則の除去

- 左辺の変数ごとにまとめると
  - E → E+T|T*F|(E)|a|b|Ia|Ib|I0|I1
  - T → T*F|(E)|a|b|Ia|Ib|I0|I1
  - F → (E)|a|b|Ia|Ib|I0|I1
  - I → a|b|Ia|Ib|I0|I1

文法単純化の順序

- 安全な単純化は以下の順序で行う
  (1) ε-規則の除去
  (2) 単位規則の除去
  (3) 無用な記号の除去

※この順序で実施しないと完全に除去できない場合がある

定理7.14：文法の単純化

- G：ε以外の列を少なくとも一つ生成するCFGに対して，$L(G_1)=L(G)-\{\varepsilon\}$ となる，
  - ε-規則を持たない
  - 単位規則を持たない
  - 無用な記号を持たない

  ような文法G_1が存在
- 証明（略）
  - 前スライドのような順序で作ったG_1が該当

ミニレポート

ミニレポート: 12-2

- テキストp299 問7.1.2 a)～c): ※ d) は不要

次の生成規則

- S → ASB|ε
- A → aAS | a
- B → SbS | A | bb

により定義される文法に対して,

a) ε-規則を除去せよ.

b) 単位規則を除去せよ.

c) 無用な記号があるか. あれば除去せよ.