

生成規則・生成文法

生成規則を与えることでも

言語を定めることが出来る

→ **生成文法 (generative grammar)**

生成規則による“文法に適っている”語の生成

- 初期変数を書く
- 今ある文字列中の或る変数を
生成規則のどれかで書換える
- 変数がなくなったら終わり

文脈自由文法 (context-free grammar)

文脈が全て空列 ε

即ち、規則が全て $A \rightarrow w$ ($A \in V$) の形

文脈自由文法の形式的定義

- V : 有限集合 (変数の集合)
- Σ : 有限集合 (終端記号の集合)
ここに $V \cap \Sigma = \emptyset$
- R : 有限集合 $\subset V \times (V \cup \Sigma)^*$ (規則の集合)
- $S \in V$: 開始変数

$(A, w) \in R$ が生成規則 $A \rightarrow w$ を表す

例：言語 $A = \{a^n b^n \mid n \geq 0\}$ は
正規言語ではないが文脈自由言語である：

- $S \rightarrow aSb \mid \varepsilon$

従って、

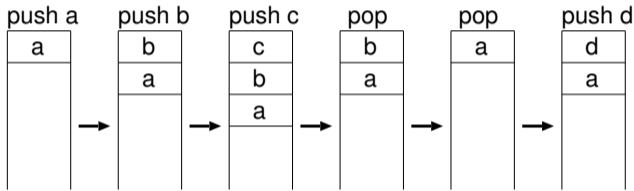
文脈自由言語は正規言語より真に広い!!

さて、正規言語を計算するモデルが
有限オートマトンであった

文脈自由言語を計算するモデル
… プッシュダウンオートマトン

プッシュダウンオートマトン

(非決定性) 有限オートマトンに
プッシュダウンスタックを取り付けたもの



無限 (非有界) の情報を保持できるが、
読み書きは先頭だけ

… LIFO (Last In First Out)

プッシュダウンオートマトンの形式的定義

$$M = (Q, \Sigma, \Gamma, \delta, s, F)$$

- Q : 有限集合 … 状態の集合
- Σ : 有限集合 … **alphabet**
- Γ : 有限集合 … **stack alphabet**
 $\Sigma_\varepsilon := \Sigma \cup \{\varepsilon\}$, $\Gamma_\varepsilon := \Gamma \cup \{\varepsilon\}$ と置く
- $\delta : Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \longrightarrow \mathcal{P}(Q \times \Gamma_\varepsilon)$
: 遷移関数 (非決定的) … 可能な遷移先全体
- $s \in Q$ … 初期状態
- $F \subset Q$ … 受理状態の集合

$$\delta : Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \longrightarrow \mathcal{P}(Q \times \Gamma_\varepsilon)$$

- $(r, y) \in \delta(q, a, x)$ とは、
「入力 a を読んだとき、
状態 q でスタックの先頭が x なら、
スタックの先頭を y に書換えて、
状態 r に移って良い」
ということ (pop; push y)
- $x = y$ は書き換え無し
- $x = \varepsilon$ は **push** のみ
- $y = \varepsilon$ は **pop** のみ
- $a = \varepsilon$ は入力を読まずに遷移

スタックマシン

このように

記憶場所としてプッシュダウンスタックを備えた
計算モデルや仮想機械・処理系を

一般に**スタックマシン**という

例：

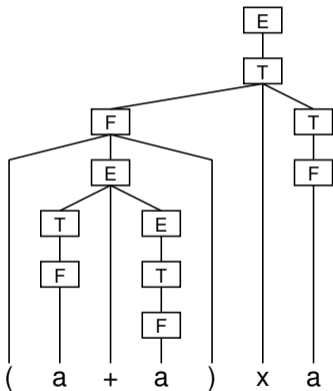
- 逆ポーランド電卓
- **PostScript**

構文解析木

生成規則の適用過程を木で表したもの

$$G = (V, \Sigma, R, E)$$

- $V = \{E, T, F\}$
- $\Sigma = \{a, +, \times, (,)\}$
- R :
 - ★ $E \longrightarrow T \mid T + E$
 - ★ $T \longrightarrow F \mid F \times T$
 - ★ $F \longrightarrow a \mid (E)$



定理 :

L : 正規言語



L が或る有限オートマトンで認識される

定理 :

L : 文脈自由言語



L が或るプッシュダウンオートマトンで
認識される

本質的な違いは？

文脈自由言語は再帰 (**recursion**) を記述できる

文脈自由言語と再帰

- $S \rightarrow aSb \mid \varepsilon$

```
S(){
    either
        "";
    or
        { "a"; S(); "b"; }
}

main(){
    S();
}
```

再帰：関数 $S()$ の中で、自分自身を呼び出す

計算機での関数呼出・再帰の実現

関数呼出は原理的には次の仕組みで行なっている

- 現在の実行番地 (戻る場所) を覚えておく
- 関数を実行する
- 関数を実行し終わったら、
覚えていた実行番地に戻って呼出側の実行再開

再帰呼出では呼出す度に覚えておく番地が増える

→ スタックに積んで覚えておく
(関数呼出の際に番地を **push**、戻ったら **pop**)

正規言語における再帰

正規表現： $(aa)^*$

- $S \rightarrow aaS \mid \varepsilon$

```
S(){
  either
  "";
  or
  { "aa"; S(); }
}

main(){
  S();
}
```

→ 末尾再帰の除去

```
main(){
  loop {
    "aa";
  }
}
```

繰返して記述可能
(再帰は不要)

正規言語・文脈自由言語と再帰

- 正規言語は繰返しを記述できる
 - 文脈自由言語は再帰を記述できる
 - 再帰の実装にはスタックを要す
 - 正規言語の生成規則は次の形に出来る
 - ★ $X \longrightarrow xY$ ($X, Y \in V, x \in \Sigma$)
 - ★ $X \longrightarrow x$ ($X \in V, x \in \Sigma_\epsilon$)
- 特に、末尾再帰であり再帰の除去可能

文脈自由言語の Pumping Lemma

文脈自由言語 A に対し、

$\exists n \in \mathbb{N} :$

$\forall w \in A, |w| \geq n :$

$\exists u, v, x, y, z \in \Sigma^* : w = uvxyz$

(1) $vy \neq \varepsilon$ (即ち $v \neq \varepsilon$ または $y \neq \varepsilon$)

(2) $|vxy| \leq n$

(3) $\forall k \geq 0 : uv^kxy^kz \in A$

文脈自由言語の例

回文全体の成す言語は文脈自由

- $S \rightarrow aSa|bSb|a|b|\epsilon$

問：回文全体の成す言語を認識する

プッシュダウンオートマトンを構成せよ

文脈自由言語の例

回文全体の成す言語は文脈自由

- $S \rightarrow aSa | bSb | a | b | \epsilon$

問：回文全体の成す言語を認識する

プッシュダウンオートマトンを構成せよ

文脈自由言語の例

回文全体の成す言語は文脈自由

- $S \rightarrow aSa|bSb|a|b|\epsilon$

問：回文全体の成す言語を認識する

プッシュダウンオートマトンを構成せよ

プッシュダウンオートマトンでは

認識できない言語の例

同じ文字列 2 回の繰返しから成る文字列全体

$$A = \{ww \mid w \in \Sigma^*\}$$

入力を読み直せないのが弱点

→ より強力な計算モデルが必要

プッシュダウンオートマトンでは

認識できない言語の例

同じ文字列 2 回の繰返しから成る文字列全体

$$A = \{ww \mid w \in \Sigma^*\}$$

入力を読み直せないのが弱点

→ より強力な計算モデルが必要

プッシュダウンオートマトンでは

認識できない言語の例

同じ文字列 2 回の繰返しから成る文字列全体

$$A = \{ww \mid w \in \Sigma^*\}$$

入力を読み直せないのが弱点

→ より強力な計算モデルが必要

一つの方法としては、

入力を覚えておくために

プッシュダウンスタックをもう一つ

使えることにする

実際これで真により強い計算モデルが得られる

しかし、通常はこれと同等な

次のような計算モデルを考える

… チューリングマシン

一つの方法としては、

入力を覚えておくために

プッシュダウンスタックをもう一つ

使えることにする

実際これで真により強い計算モデルが得られる

しかし、通常はこれと同等な

次のような計算モデルを考える

… チューリングマシン

チューリングマシン

- 有限個の内部状態を持つ
- 入力データはテープ上に一区画一文字ずつ書き込まれて与えられる
- データを読み書きするヘッドがテープ上を動く
- 遷移関数は次の形：
内部状態とヘッドが今いる場所の文字とによって、その場所の文字を書き換え、次の内部状態に移り、ヘッドを左か右かに動かす
- 受理状態または拒否状態に達したら停止するが、停止しないこともある