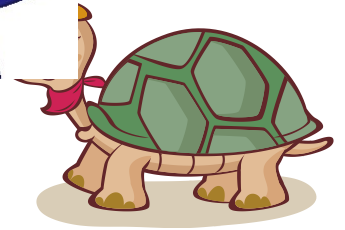


山場だよ！



言語プロセッサ —第7回目—

構文解析(続き)

東京工科大学
コンピュータサイエンス学部
亀田 弘之

今日の内容

- 構文解析

- LL(1)文法(復習)

- First集合(復習)

- Follow集合(復習)



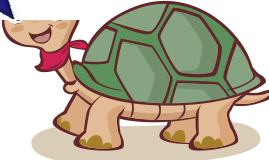
今日でマスターしてしまおう!

- 構文解析表による構文解析法

- 予測的構文解析のモデル

- 構文解析表の作り方

- 動作 など



今日は気合
が必要だよ

LL(1)文法

最終確認事項

- LL(1)文法のイメージ:

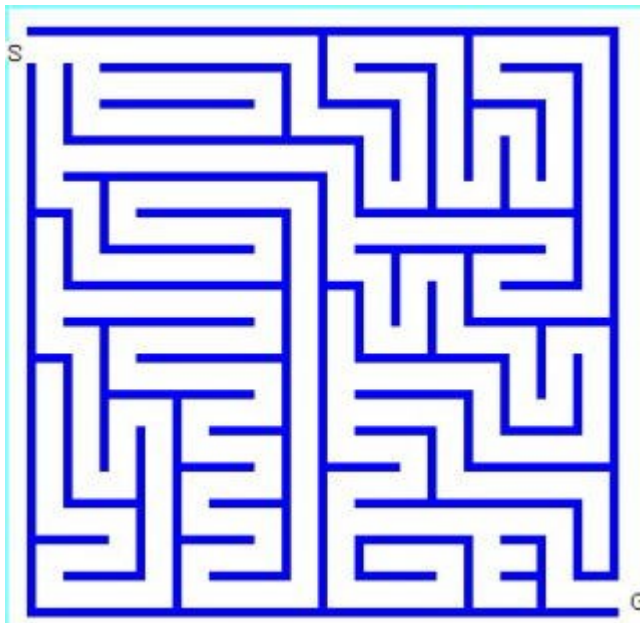
$$A \rightarrow \alpha \mid \beta$$

という規則で、 α か β のどちらの書き換えを選
ぶかを決めるとき、入力の先頭記号1個を見
ることにより、バックトラックが起きないような
選択が可能な文法。

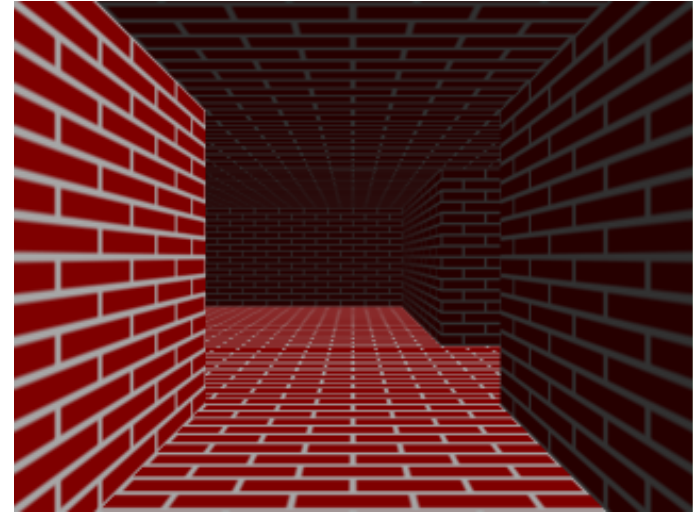
つまり、適応すべき文法規則を、1文字先
読みすれば決定できるということ。

高速処理には大切な性質です。

(参考)迷路の例



(出典) <http://tappe.jp/>



(出典) <http://www.yagamifestival.com/11th/group/2010/10/post-43.html>



形式文法(復習)

確認事項(常識)

文法 $G=(N, V, S, P)$,

ただし、

- N : 非終端記号の集合(構文構造記述用語集)
- V : 終端記号の集合(語彙)
- S : 開始記号
- P : 書換え規則の集合

LL(1)文法の条件

再確認事項

1. 文法における制限

- 左再帰性の除去
(左再帰の書き換え規則を変形により削除する。
無限ループの回避を目的としている。)
- 括りだし(factoring)
(バックトラックの回避を目的としている。)

2. 構文解析方法

- Top down
- 再帰呼び出し
- **1文字先読み**

1文字先読み十分性の条件は？

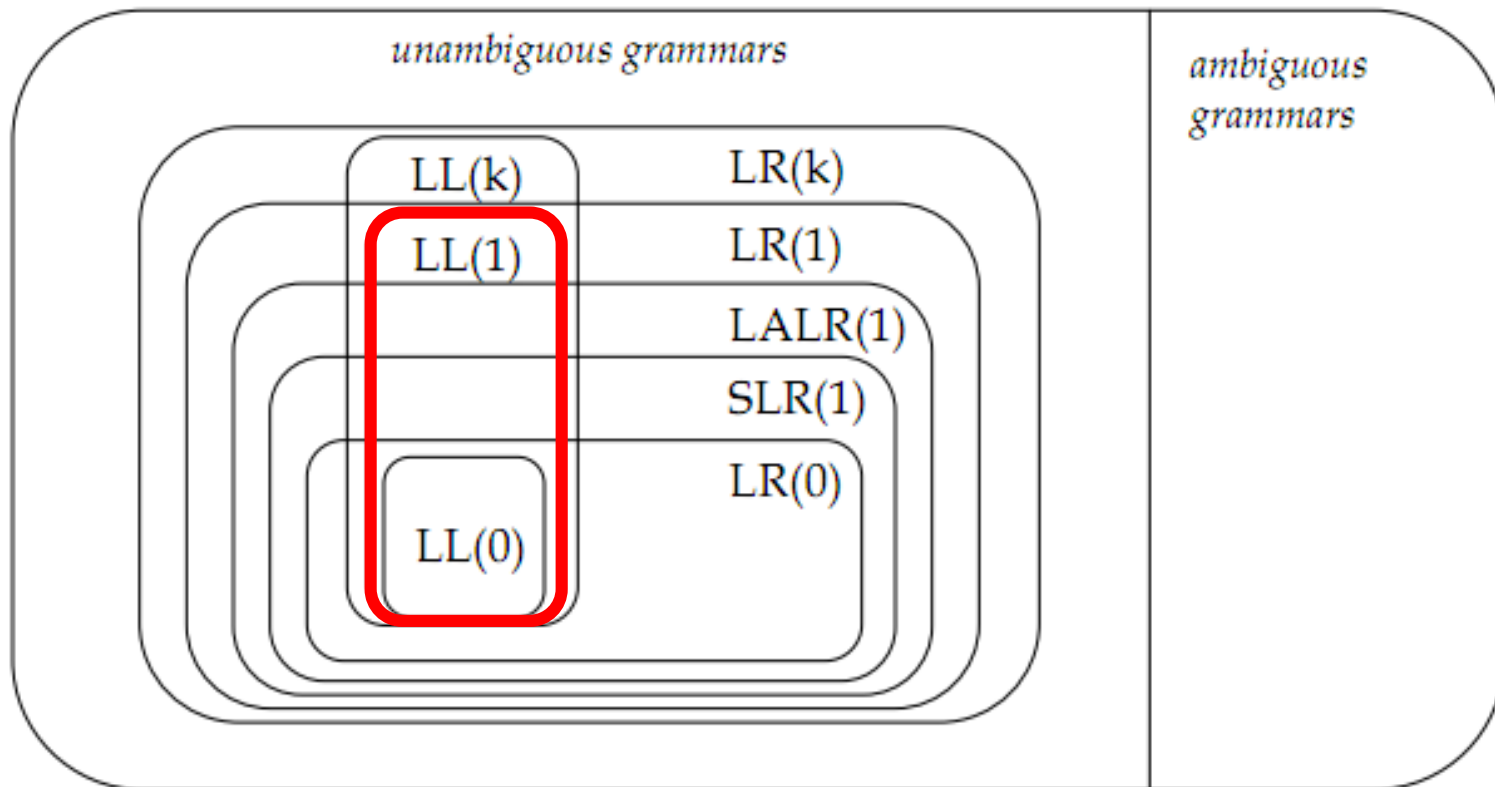
- **First集合とFollow集合**を用いて作成される **構文解析表**において、表内のどのセルにも動作が多重に定義されていないこと。
(詳細は後述します。)

確認事項

(参考情報) LL(1)の位置づけ

LL(1) versus LR(k)

A picture is worth a thousand words:



First集合

最終確認事項

【定義】

$$\text{First}(\alpha) = \{ a \mid a \in V_t, \alpha \stackrel{*}{\Rightarrow} a\dots \}$$

ただし、 $\alpha \stackrel{*}{\Rightarrow} \epsilon$ ならば、 $\epsilon \in \text{First}(\alpha)$

【意味】

文字列 α から導出される文字列の先頭(最左端)に現れる終端記号の集合。

例

構文規則

- $S \rightarrow \text{if} (E) S \text{ else } S$
| $\text{while} (E) S$
| $\{ S \}$

Firstを求める様子

- $\text{First}(S)$:
 - $S \rightarrow \text{if} (\dots \text{より、}$
 $\text{if} \in \text{First}(S)$
 - $S \rightarrow \text{while} (E \dots \text{より、}$
 $\text{while} \in \text{First}(S)$
 - $S \rightarrow \{ S \} \text{より、}$
 $\{ \in \text{First}(S)$

First集合

最終確認事項

【First集合を求めるアルゴリズム】

以下を、どのFirst集合にも新たに追加するものがなくなるまで繰り返す。

1. $\text{First}(\varepsilon) = \{\varepsilon\}$
2. $\text{First}(a\alpha) = \{a\}$ if $a \in V_t$ (V_t は終端記号の集合)
3. if($\text{First}(Y) \not\supseteq \varepsilon$) Y は空文字列にならない。
 $\text{First}(Y\alpha) = \text{First}(Y)$
else
 $\text{First}(Y\alpha) = (\text{First}(Y) - \{\varepsilon\}) \cup \text{First}(\alpha)$
4. if($X \rightarrow \alpha$) $\text{First}(X) = \text{First}(X) \cup \text{First}(\alpha)$

練習問題

- ① First(abcd)
- ② First(b α)

練習問題

① $S \rightarrow S S + \mid S S * \mid a$

- First集合を求めよ。

② $S \rightarrow 0 S 1 \mid 0 1$

- First集合を求めよ。

Follow集合

最終確認事項

【定義】

$$\text{Follow}(X) = \{ a \mid a \in V_t, S \stackrel{*}{\Rightarrow} \dots Xa \dots \}$$

【意味】

Xの直後に現れる 終端記号の集合

Follow集合

最終確認事項

【Follow集合を求めるアルゴリズム】

以下を、どのFollow集合にも新たに追加するものがなくなるまで繰り返す。

1. Follow(S) に \$ を加える。
2. 規則 $A \rightarrow \alpha B \beta$ ($B \in N$) に対して、
 - (ア) First(β)をFollow(B)に加える。
ただし、 $\epsilon \in \text{First}(\beta)$ のときは ϵ は加えない。
 - (イ) $\epsilon \in \text{First}(\beta)$ または $\beta = \epsilon$ ならば、Follow(A) を Follow(B) に加える。

First集合とFollow集合

再確認事項

【定義】

1. $\text{First}(\alpha) = \{a \mid a \in V_t, \alpha \stackrel{*}{\Rightarrow} a\dots\}$

ただし、 $\alpha \stackrel{*}{\Rightarrow} \varepsilon$ ならば、 $\varepsilon \in \text{First}(\alpha)$

2. $\text{Follow}(X) = \{a \mid a \in V_t, S \stackrel{*}{\Rightarrow} \dots Xa\dots\}$

演習 (Let's challenge!)

**First集合とFollow集合を
求めてみよう！**



確認問題

文法 G_1 について答えよ。

- ① 終端記号はどれか？
- ② 非終端記号はどれか？
- ③ 開始記号はどれか？
- ④ First集合を求めよ。
- ⑤ Follow集合を求めよ。
- ⑥ 構文解析表を求めよ。
- ⑦ G_1 はLL(1)か？

G_1 :

$S \rightarrow aBd$

$B \rightarrow bC$

$C \rightarrow c|\epsilon$

本番演習 (*Let's challenge!*)

- 次の文法を考える。FirstとFollowを求めよ。

$$\begin{cases} E \rightarrow E + T \mid T \\ T \rightarrow T * F \mid F \\ F \rightarrow (E) \mid \text{id} \end{cases}$$

**First集合とFollow集合を
求めてみよう！**



まず、何をしますか？

1. 左再帰性がないかどうかを確認する。
 - ・どうやって確認する？
2. 左再帰性があれば、それを除去する。
 - ・どうやって除去する？

First集合とFollow集合

【例】

文法 $G=(V,N,E,P)$

$$P = \left\{ \begin{array}{l} E \rightarrow TE', \\ E' \rightarrow +TE' \mid \varepsilon \\ T \rightarrow FT' \\ T' \rightarrow *FT' \mid \varepsilon \\ F \rightarrow (E) \mid id \end{array} \right\}$$

教科書p.86より

- $\text{First}(E) = \text{First}(T) = \text{First}(F) = \{ (, \text{id} \}$
- $\text{First}(E') = \{ +, \varepsilon \}$
- $\text{First}(T') = \{ *, \varepsilon \}$
- $\text{Follow}(E) = \text{Follow}(E') = \{), \$ \}$
- $\text{Follow}(T) = \text{Follow}(T') = \{ +,), \$ \}$
- $\text{Follow}(F) = \{ +, *,), \$ \}$

皆さんこうになりましたか？

これをもとに、 構文解析へと進みましょう

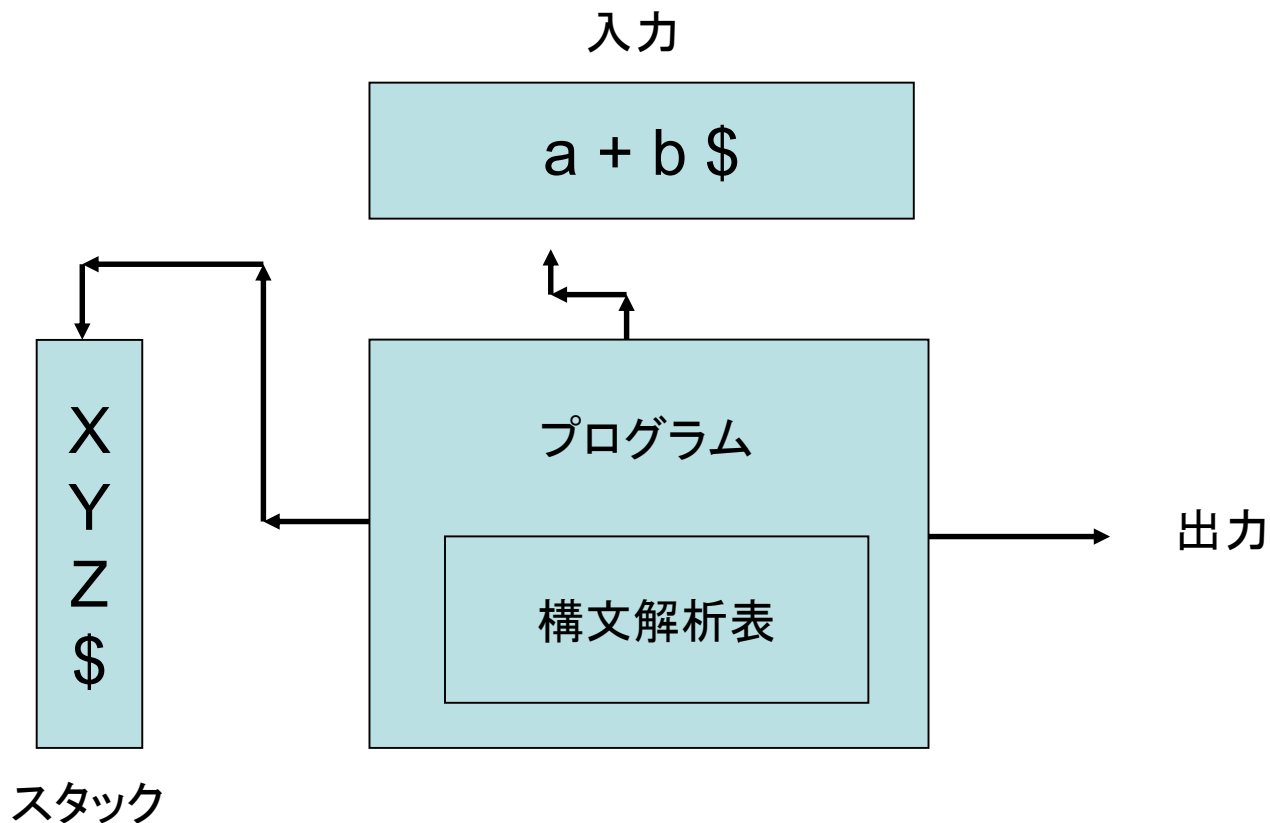
話題

- 構文解析表
- 予測的構文解析
- 実際のトレース など

構文解析表による構文解析法

- 予測的構文解析のモデル
- 構文解析表の作り方
- 構文解析のアルゴリズム

予測的構文解析のモデル



構文解析表の作り方

- 入力: 文法G
- 出力: 構文解析表M
- 手順: (次のページ参照)

構文解析表作成手順

1. 文法の各規則 $A \rightarrow \alpha$ に対して、ステップ2と3を行う。
2. 各終端記号 $a \in \text{First}(\alpha)$ に対して、 $M[A, a]$ に $A \rightarrow \alpha$ を記入する。
3. $\epsilon \in \text{First}(\alpha)$ ならば、各終端記号 $b \in \text{Follow}(A)$ に対して、 $M[A, b]$ に $A \rightarrow \alpha$ を記入する。
 $\epsilon \in \text{First}(\alpha)$ かつ $\$ \in \text{Follow}(A)$ ならば、 $M[A, \$]$ に $A \rightarrow \alpha$ を記入する。
4. M の未記入欄にerrorを記入する。

構文解析のアルゴリズム

1. $X = a = \$$ ならば、“構文解析成功” を出力し停止。
2. $X = a \neq \$$ ならば、スタックから X をpopし、入力ポインタを1つ進める。
3. $A \in V_t$ ならば、 $M[X, a]$ を調べる。
 $M[X, a] = \{ X \rightarrow ABC \}$ ならば、 C, B, A の順にスタックにpushし、 $X \rightarrow ABC$ を実行する。
 $M[X, a] = \text{error}$ ならば、停止。

First & Follow

- (もう一度自力でやってみよう)



- $\text{First}(E) = \text{First}(T) = \text{First}(F) = \{ (, \text{id} \}$
- $\text{First}(E') = \{ +, \varepsilon \}$
- $\text{First}(T') = \{ *, \varepsilon \}$
- $\text{Follow}(E) = \text{Follow}(E') = \{), \$ \}$
- $\text{Follow}(T) = \text{Follow}(T') = \{ +,), \$ \}$
- $\text{Follow}(F) = \{ +, *,), \$ \}$

こうでしたよね！

構文解析表を作成する

- (各自で作成できるようにしよう！)

構文解析表

	id	+	*	()	\$
E						
E'						
T						
T'						
F						

構文解析表の作り方

- 入力: 文法G
- 出力: 構文解析表M
- 手順: (次のページ参照)

構文解析表作成手順

1. 文法の各規則 $A \rightarrow \alpha$ に対して、ステップ2と3を行う。
2. 各終端記号 $a \in \text{First}(\alpha)$ に対して、 $M[A, a]$ に $A \rightarrow \alpha$ を記入する。
3. $\epsilon \in \text{First}(\alpha)$ ならば、各終端記号 $b \in \text{Follow}(A)$ に対して、 $M[A, b]$ に $A \rightarrow \alpha$ を記入する。
 $\epsilon \in \text{First}(\alpha)$ かつ $\$ \in \text{Follow}(A)$ ならば、 $M[A, \$]$ に $A \rightarrow \alpha$ を記入する。
4. M の未記入欄にerrorを記入する。

構文解析表

	id	+	*	()	\$
E	$E \rightarrow T E'$			$E \rightarrow T E'$		
E'		$E' \rightarrow + T E'$			$E' \rightarrow \epsilon$	$E' \rightarrow \epsilon$
T	$T \rightarrow F T'$			$T \rightarrow F T'$		
T'		$T' \rightarrow \epsilon$	$T' \rightarrow * F T'$		$T' \rightarrow \epsilon$	$T' \rightarrow \epsilon$
F	$F \rightarrow \text{id}$			$F \rightarrow (E)$		

構文解析器の動作手順

- 予測的構文解析系は、入力、スタック、構文解析表および出力からなる。
- 入力には構文解析対象の文字列があり、末尾に記号 \$ が添えられている。
- スタックには底を示す記号 \$ が入っている。初期状態では開始記号が入っている。
- 構文解析表は配列 $M[A,a]$ として記述する。 A は非終端記号、 a は終端記号か \$。

1. スタックから記号を1つポップする。それを X とする。また現在の入力記号の先頭を a とする。
2. $X=a=\$$ ならば、処理は正常終了。
3. $X=a\neq\$$ ならば、スタックから X をポップし、入力ポインタを1つ進める。
4. X が非終端記号ならば、 $M[A,a]$ を調べる。

5. $M[A, a] = \{ X \rightarrow UVW \}$ ならば、スタックの X を WVU (U が最上位)に書き換える。出力としてこの書き換え規則を印字する。
6. $M[A, a] = \text{error}$ ならば回復ルーチンを起動する。

動作のトレース

- 入力: $id + id * id$

落ち着いてトレースしよう！



動作のトレース

<u>スタック</u>	<u>入力</u>	<u>出力</u>
\$E	id+id*id	
\$E'T	id+id*id	$E \rightarrow TE'$
\$E'T'F	id+id*id	$T \rightarrow FT'$
\$E'T'id	id+id*id	$F \rightarrow id$
\$E'T'	+id*id	
\$E'	+id*id	$T' \rightarrow \varepsilon$ (以下省略)

以上で構文解析の話しの 概要はほぼ終わりです。

- 引き続き練習を積み重ねてください。
 - 用語を知ること
 - 用語の意味を知り、自分で説明できるように。
 - なぜそのような用語が定義されているのかを考えて、自分の言葉で理解する。
 - 左再帰削除ができるように
 - Factoringができるように
 - First集合やFollow集合を自分で求められるように
 - 構文解析表が作れるように
 - 構文解析処理をトレースできるように

参考情報

1. 構文解析まで終われば、後は少し楽になります。
2. 構文解析は解析の中でも難関部分で、今日でも多くの研究がなされています。
3. 構文解析の次は意味解析
(解析の最終段階)。
4. それ以後は合成の段階になります。
5. 来週は**ここまでの総合演習**をやります。

Let's call it a day!
今日はここまで！



寒くなったので健康には十分注意しましょう。

復習(確認)問題

文法

- $E \rightarrow E + T \mid T$
- $T \rightarrow T * F \mid F$
- $F \rightarrow (E) \mid \text{id}$

左再帰性を除去しなさい。

(直接)左再帰性除去法

1. A に関する左再帰性を取り除くためにまず,
 $A \rightarrow A\alpha_1 | A\alpha_2 | \dots | A\alpha_m | \beta_1 | \beta_2 | \dots | \beta_n$
のようにまとめる.

2. 上記の形を基に以下のように変形する.

$$A \rightarrow \beta_1 A' | \beta_2 A' | \dots | \beta_n A'$$

$$A' \rightarrow \alpha_1 A' | \alpha_2 A' | \dots | \alpha_m A' | \varepsilon$$

パターンを
覚えよう!



新たな文法規則

- $E \rightarrow T E'$
- $E' \rightarrow + T E' \mid \varepsilon$
- $T \rightarrow F T'$
- $T' \rightarrow * F T' \mid \varepsilon$
- $F \rightarrow (E) \mid id$

自由課題

- 上記のアルゴリズムは任意の文法に対しても適用できるが、文法によってはMの欄に対して複数の規則が書き込まれることがある。

【例】

$$P = \{ S \rightarrow i C t S S' \mid a, S' \rightarrow e S \mid \varepsilon, \\ C \rightarrow b \}$$

($M[S', \varepsilon]$ を求めてみよ。)

- LL(1)文法はこのようなことが起きない文法。

確認問題（自宅で復習してみよう）

文法 G_1 について答えよ。

- ① 終端記号はどれか？
- ② 非終端記号はどれか？
- ③ 開始記号はどれか？
- ④ First集合を求めよ。
- ⑤ Follow集合を求めよ。
- ⑥ 構文解析表を求めよ
- ⑦ G_1 はLL(1)か？
- ⑧ 入力 abd の構文解析動作を書き下しなさい。

G_1 :

$S \rightarrow aBd$

$B \rightarrow bC$

$C \rightarrow c|\varepsilon$