授業の進め方

- 演習の日も講義を行う
- 授業の前半に講義を行い、後半に演習を行う
- 演習の解答を記入する A4 の紙 (ルーズリーフ不可) を持参すること
- 成績は演習の点数、中間試験、期末試験にて決める
- 教科書 「コンピュータサイエンスのための言語理論入門」, R. Smith 著, 吉田 敬一 他訳, 共立出版, 1986, 2300 円
- 参考書「オートマトン, 言語理論, 計算論 I [第 2 版]」J. Hopcroft, R. Motowani, J. Ullman 著, 野崎 昭弘, 高橋 正子, 町田 元, 山崎 秀記 訳, サイエンス社, 2002, 2800 円

10 月の授業の予定
10/1(金)
10/6(水)
10/13(水) 休講
10/15(金) 休講
10/20(水)
10/27(水)
10/29(金)

この科目は何の役に立つのか

日本語、英語のような人間が自然に使っている言語を自然言語と呼び、C 言語、HTML のようなコンピュータのために人間が作った言語を形式言語と呼ぶ。

形式言語の解釈プログラムを作るための基礎理論を学ぶ

例: C 言語における変数の宣言
「int a;」「int a=0;」「int a,b;」「int a=1.0,b;」
様々な変数の宣言の形式が許されるが、様々な宣言を正しく解釈するプログラムを手作業で作るのは面倒だし、ミスも出やすい。

許される宣言の形式を簡潔に記述し、その記述から自動的に解釈プログラムを生成するプログラムがあれば便利である。そのようなプログラムは UNIX では lex, yacc という名前で入っている。

そのような形式言語の記述の仕方と、形式言語の解釈をするための計算のモデルを学ぶ。

記号と言葉の紹介

言語の記号が属する集合をアルファベットと呼び、通常 Σ で表す。
例： $\Sigma = \{a, b\}$.
Σ^*: Σ の要素を 0 個以上有限個並べた列からなる集合

以下 $x, y \in \Sigma^*$ とする

ϵ	:	長さ 0 の列
$\lvert x \rvert$	:	列 x の長さ
xy	:	列 x の後に列 y を連接して (つなげて) 得られる列

任意の列 $x \in \Sigma^*$ に対して

$$x = x\epsilon = \epsilon x.$$

アルファベット Σ 上の言語 L とは Σ^* の空でない部分集合のことである。

記号の紹介２ — 言語に関する演算

L_1, L_2 を言語とする。 L_1 と L_2 の連接 L_1L_2 とは

$$L_1L_2 = \{xy \mid x \in L_1, y \in L_2\}$$

例: $L_1 = \{001, 10, 111\}$, $L_2 = \{\epsilon, 001\}$ としたとき
$L_1L_2 = \{001, 10, 111, 001001, 10001, 111001\}$

L を言語とする.

$$\begin{aligned}
L^0 &= \{\epsilon\}, \\
L^1 &= L, \\
L^2 &= LL, \\
L^3 &= LLL, \\
L^n &= LL \cdots L (L \text{ を } n \text{ 回連接})
\end{aligned}$$

言語のクリーネ閉包

$$L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \cdots$$

を L のクリーネ (kleene) 閉包と呼ぶ。

$$L^+ = L^1 \cup L^2 \cup L^3 \cup \cdots$$

も同様に定義されるが特に名前は無い。

演習問題 1 L^* が空列 ϵ を含まないことはあるか？

演習問題 2 L^+ が空列 ϵ を含むのは、言語 L がどのような条件を満たすときか？

- 演習問題を解く時間を後でまとめて取る
- 演習問題は後で掲示するので、今ノートに写す必要は無い

言語の定義の仕方

一般には言語 L は無限集合である
例: すべての文法的に正しいホームページ (HTML ファイル) の集合

無限集合なので要素を列挙することによって定義することはできない。ではどうすればよいか？

変数 a, b, c と演算子 $+, -, \times, /$ および括弧からなる式の集合を定義することを考える。式を以下のように定義してみる。

項は式である、式＋項は式である、式－項は式である、
係数は項である、項×係数は項である、項／係数は項である、
a は係数である、b は係数である、c は係数である、(式) は係数である。

$a + b$ は係数＋係数で、項＋項で、式＋項で、式である、となるので、$a + b$ が式であることがわかる。

上記のようなやり方で式の集合を定義できる。しかし記述が繁雑なので、もう少し簡単に書いてみる。

数式の定義の簡略化

式を「E」(expression の略)、項を「T'」(term の略)、係数を「F」(factor の略) と書く。

「項 (T’) は式 (E) である」などを「 $E \to T'$」などと書く。

そうすると上記の数式の定義は
$E \to T'$, $E \to E + T'$, $E \to E - T'$,
$T' \to F$, $T' \to T' \times F$, $T' \to T'/F$,
$F \to a$, $F \to b$, $F \to c$, $F \to (E)$
とすっきり書ける。

数式の定義の簡略化２

$E \to T'$, $E \to E + T'$, $E \to E - T'$,
$T' \to F$, $T' \to T' \times F$, $T' \to T'/F$,
$F \to a$, $F \to b$, $F \to c$, $F \to (E)$

上記の規則は記号の書き換え規則と見做すことができる。今考えている式の集合は記号列の中に E, T', F が含まれなくなるまで、記号 E を書き換えて得られる記号列の集合と考えられる。

上記の定義で E, T', F は言語を定義するために導入した記号でアルファベットの要素ではない。このような記号を非終端記号と呼ぶ。

$a, b, c, +, -, \times, /, (,)$ は言語を構成する記号で「終端記号」と呼ぶ。

上記の定義で「式」とは E を書き換えて得られる記号列と定義した。このように書き換えの最初になる記号を開始記号と呼ぶ。

<u>句構造文法</u>

前 OHP のようなものを形式的に厳密に定義したものが句構造文法である。句構造文法は以下の４つ組 (N, T, P, S) である。
N: 非終端記号の有限集合
T: 終端記号の有限集合
$N \cap T = \emptyset$ でなければならない
P: 生成規則の集合。一つの生成規則は $\alpha \in (N \cup T)^+$, $\beta \in (N \cup T)^*$ を用いて $\alpha \to \beta$ という形で書かれる。
S: 開始記号

普通終端記号はアルファベット小文字で、非終端記号はアルファベット大文字で表す。

演習問題3 $N = \{E, T', F\}, T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $\epsilon \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。

演習問題4 $N = \{E, T', F\}, T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $aa \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。

生成規則の省略記法

$F \to a,\ F \to b,\ F \to c,\ F \to (E)$ のように左辺が同じ生成規則を繰り返して書くのは面倒である。そこで

$$F \to a|b|c|(E)$$

のように、右辺を縦線で区切ってひとつにまとめて書くことがある。

導出の記号

$G = (N, T, P, S)$ を任意の句構造文法とし、生成規則 $\alpha \to \beta$ を含むとし、$\gamma_1 \alpha \gamma_2 \in (N \cup T)^+$ とする。

記号：$\gamma_1 \alpha \gamma_2 \underset{G}{\Longrightarrow} \gamma_1 \beta \gamma_2$

意味：$\gamma_1 \alpha \gamma_2$ を G の生成規則を用いて書き換えることにより $\gamma_1 \beta \gamma_2$ を得ることができる

読み方： $\gamma_1 \alpha \gamma_2$ は $\gamma_1 \beta \gamma_2$ を生成する、または、$\gamma_1 \beta \gamma_2$ は $\gamma_1 \alpha \gamma_2$ より導出される。

導出の記号 2

$\alpha_1 \underset{G}{\Longrightarrow} \alpha_2,\ \alpha_2 \underset{G}{\Longrightarrow} \alpha_3,\ \ldots,\ \alpha_{n-1} \underset{G}{\Longrightarrow} \alpha_n$ をまとめて

$$\alpha_1 \underset{G}{\Longrightarrow} \alpha_2 \underset{G}{\Longrightarrow} \cdots \underset{G}{\Longrightarrow} \alpha_n$$

と書く

$\alpha_n \neq \alpha_1$ のときに α_n は生成規則を 1 回以上用いて α_1 から導出されるが、この関係を

$$\alpha_1 \overset{+}{\underset{G}{\Longrightarrow}} \alpha_n$$

と書く

「$\alpha_1 = \alpha_n$ または $\alpha_1 \overset{+}{\underset{G}{\Longrightarrow}} \alpha_n$」であることを

$$\alpha_1 \overset{*}{\underset{G}{\Longrightarrow}} \alpha_n$$

と書く。

文法 G が明らかなときに二重矢印の下の G は略す。

文、文型

句構造文法 $G = (N, T, P, S)$ を考える。

G の文型 (sentential form): $S \overset{*}{\underset{G}{\Longrightarrow}} \alpha$ である $(N \cup T)^*$ の要素

G の文 (sentence): $S \overset{*}{\underset{G}{\Longrightarrow}} \alpha$ である T^* の要素

演習問題5 式の文法 $G = (N, T, P, E)$, $N = \{E, T', F\}$, $T = \{a, b, c, +, -, \times, /, (,)\}$,

$$
\begin{aligned}
P &= \{E \to T' | E + T' | E - T', \\
&\quad T' \to F | T' \times F | T'/F, \\
&\quad F \to a|b|c|(E)\}
\end{aligned}
$$

に対して、G の文を一つ挙げ、その文の導出を $\alpha_1 \Longrightarrow \alpha_2 \Longrightarrow \cdots \Longrightarrow \alpha_n$ という形式で書け。なるべく周りの人とは異なる文を挙げること。

演習問題6 問題5の文法の文型を一つ挙げよ。なるべく周りの人とは異なる文を挙げること。

$\underline{L(G)}$

句構造文法 $G = (N, T, P, S)$ を考える。
集合 $L(G)$ を開始記号 S から導出される T^* に属する記号列の集合とする、言い替えれば

$$L(G) = \{\alpha \in T^* \mid S \overset{*}{\underset{G}{\Longrightarrow}} \alpha\}$$

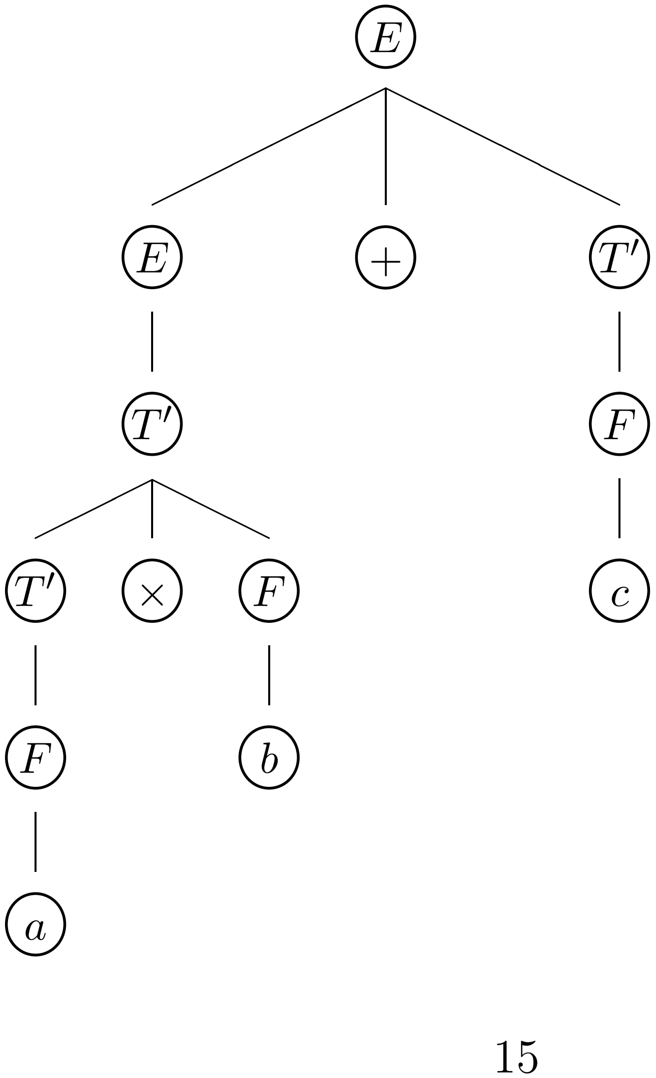
要するに $L(G)$ は文法 G によって定義される言語である。

異なる文法 G_1, G_2 から同じ言語が生成される、つまり、$L(G_1) = L(G_2)$ であることが有り得る。このとき G_1 と G_2 は等価であると言う。

導出の視覚的な表現法 — 導出木

式の文法 $G = (N, T, P, E)$, $N = \{E, T', F\}$, $T = \{a, b, c, +, -, \times, /, (,)\}$, $P = \{E \to T'|E + T'|E - T', T' \to F|T' \times F|T'/F, F \to a|b|c|(E)\}$ を考える

$E \Longrightarrow E + T' \Longrightarrow T' + T' \Longrightarrow T' \times F + T' \Longrightarrow F \times F + T' \Longrightarrow a \times F + T' \Longrightarrow a \times b + T' \Longrightarrow a \times b + F \Longrightarrow a \times b + c$ という導出を以下のように書ける。



上記のような表記法を導出木と呼ぶ。

演習問題 7 問題 5 で解答した導出に対応する導出木を書け

文法の曖昧さ

言語 G とその文 α が与えられたときに、α の導出木は一般的に１通りに定まらない。

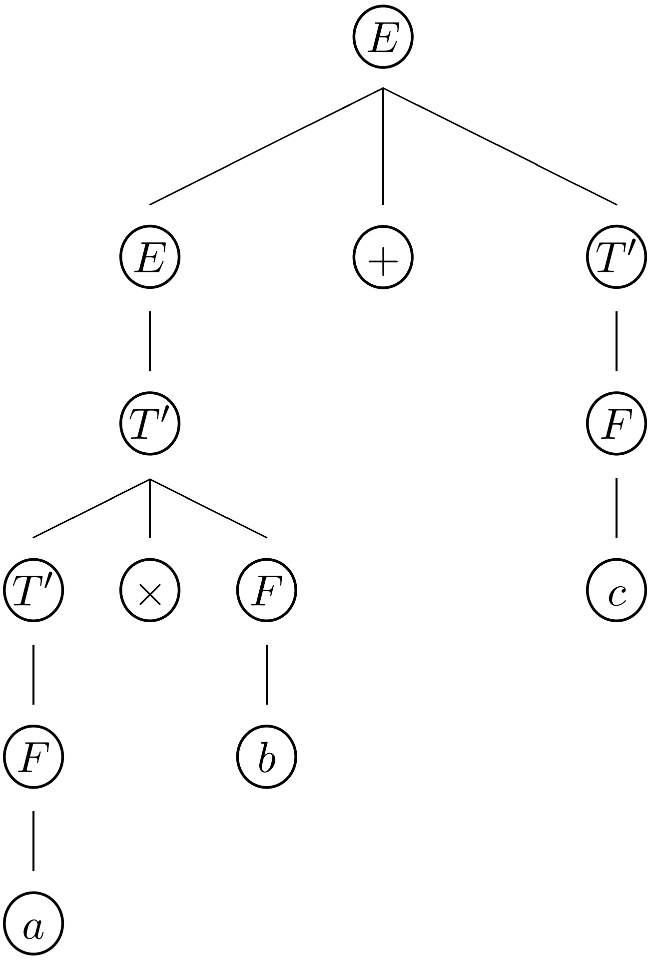
ある言語 G がすべての文 $\alpha \in L(G)$ に対して α の導出木が１通りに定まるときに、言語 G は曖昧ではないと言う。

例：式の文法は曖昧ではない (説明略)。

導出木はどう役に立つのか？

文法を適切に定義してある場合、導出木から記号列の意味を抽出することができる。従って導出木の構成は解釈プログラムの最初の重要な段階である。

例で出している式の文法は、すべての式に対し１通りの導出木を与え、なおかつその導出木は加減乗除と括弧の優先順位を考慮した木になる。(教科書を持っている人は 34 ページの図 2.7 を見よ)



これから何をやるのか

句構造文法は一般的すぎるので、ある列 α が $L(G)$ に所属するのかどうか判定する効率の良い方法が知られていない。また所属するとしても効率的に導出木を作る方法も知られていない。従って今の所句構造文法に実用的な意味は無い。

句構造文法に制限を加えることにより、$\alpha \in L(G)$ が効率的に判定することや導出木の生成などが可能になる。そのような制限された文法をこれから学んで行く。

演習問題一覧

答案に学籍番号、名前、ふりがなを書いて下さい。
演習問題 1 ある言語 L について L^* が空列 ϵ を含まないことはあるか？
演習問題 2 L^+ が空列 ϵ を含むのは、言語 L がどのような条件を満たすときか？
演習問題 3 $N = \{E, T', F\}$, $T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $\epsilon \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。
演習問題 4 $N = \{E, T', F\}$, $T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $aa \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。
演習問題 5 式の文法 $G = (N, T, P, E)$, $N = \{E, T', F\}$, $T = \{a, b, c, +, -, \times, /, (,)\}$, $P = \{E \to T'|E + T'|E - T', T' \to F|T' \times F|T'/F, F \to a|b|c|(E)\}$ に対して、G の文を一つ挙げ、その文の導出を $\alpha_1 \Longrightarrow \alpha_2 \Longrightarrow \cdots \Longrightarrow \alpha_n$ という形式で書け。なるべく周りの人とは異なる文を挙げること。また講義でやった例は避けること。
演習問題 6 問題 5 の文法の文型を一つ挙げよ。なるべく周りの人とは異なる文型を挙げること。
演習問題 7 問題 5 で解答した導出に対応する導出木を書け
演習問題 8 アルファベット $\Sigma = \{0, 1\}$ 上の言語 $L = \{1, 01\}^*$ を生成する句構造文法を書け
演習問題 9 授業でわかりにくかった所や要望をなるべく多く書いて下さい

演習問題解答

演習問題 1 ある言語 L について L^* が空列 ϵ を含まないことはあるか？
答え：ない。

$$L^* = L^0 \cup L^1 \cup \cdots, \quad L^0 = \{\epsilon\}$$

なので。
演習問題 2 L^+ が空列 ϵ を含むのは、言語 L がどのような条件を満たすときか？
答え： L が空列 ϵ を含むとき
演習問題3 $N = \{E, T', F\}, T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $\epsilon \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。
答え：有り得ない。なぜならば、生成規則の左辺は $(N \cup T)^+$ の要素で、$(N \cup T)^+$ は空列を含まないから。
演習問題4 $N = \{E, T', F\}, T = \{a, b, c, +, -, \times, /, (,)\}$ としたときに $aa \to a$ という生成規則を持つ句構造文法は有り得るか？理由を付けて答えよ。
答え：有り得る。左辺が終端記号だけからなる列であるのが奇妙な感じがするが、定義上ありえる。しかし実際には左辺が終端記号だけからなる生成規則を含む文法を扱うことは稀である。

演習問題5 式の文法 $G=(N,T,P,E), N=\{E,T',F\}$, $T=\{a,b,c,+,-,\times,/,(,)\}, P=\{E\to T'|E{+}T'|E{-}T', T'\to F|T'\times F|T'/F, F\to a|b|c|(E)\}$ に対して、G の文を一つ挙げ、その文の導出を $\alpha_1 \Longrightarrow \alpha_2 \Longrightarrow \cdots \Longrightarrow \alpha_n$ という形式で書け。なるべく周りの人とは異なる文を挙げること。
答え：講義で例を一つ挙げたので省略
演習問題6 問題5の文法の文型を一つ挙げよ。なるべく周りの人とは異なる文型を挙げること。
答え： $E+T'$ など。非終端記号を含む列を書かないと文になってしまうので不正解。
演習問題7 問題5で解答した導出に対応する導出木を書け
授業でやったので解答略
演習問題8 アルファベット $\Sigma=\{0,1\}$ 上の言語 $L=\{1,01\}^*$ を生成する句構造文法を書け
答え：正解はたくさん有る。例えば

$$G=(\{S\},\{0,1\},\{S\to\epsilon|SS|1|01\},S)$$

など