课时2 高级语言及其语法描述

考点	重要程度	占分	题型
1.上下文无关文法的定义	★★★	3~5	简答题、填空题
2.推导、语言的概念	★★★★	3~5	简答题、填空题
3.文法二义性的判定	★★★	3~5	简答题
4.文法的分类及应用	★★★★	8~10	简答题

2.1 上下文无关文法

1.文法： 描述语言的语法结构的形式规则

一个上下文无关文法G是一个四元式

$G=(V_T, V_N, S, P)$，其中

V_T： 终结符集合(非空)

V_N： 非终结符集合(非空)，且$V_T \cap V_N=\varnothing$

S： 文法的开始符号，$S \in V_N$

P： 产生式集合(有限)，每个产生式形式为

$P \rightarrow \alpha$， $P \in V_N$， $\alpha \in (V_T \cup V_N)^*$

开始符号S至少必须在某个产生式的左部出现一次。

【题1】定义只含+，*的算术表达式的文法

G=<{i，+，*，(，)}，{E}，E，P>，其中，P由下列产生式组成：

$E \rightarrow i$

$E \rightarrow E+E$

$E \rightarrow E*E$

$E \rightarrow (E)$

视频讲解更清晰

2.几点规定：

“$\rightarrow$” 也可以用 “::=" 表示， 这种表示称为**巴科斯范式(BNF)**

$$P \rightarrow \alpha_1$$

$$P \rightarrow \alpha_2 \quad \text{可缩写为 } P \rightarrow \alpha_1|\alpha_2|\ldots|\alpha_n$$

$$\ldots$$

$$P \rightarrow \alpha_n$$

其中， “|”读成 “或” ，称为P的一个候选式。

表示一个文法时，通常只给出开始符号和产生式，

如上例，可表示为：$G(E)$： $E \rightarrow i \mid E+E \mid E*E \mid (E)$

2.2 推导、语言的概念

一、程序语言

程序语言由两方面定义：

1. **语法：** 一组规则，用它可以形成和产生一个合式(well-formed)的程序。

词法规则： 单词符号的形成规则。单词符号是语言中具有独立意义的最基本结构。

一般包括：常数、标识符、基本字、算符、界符等。

描述工具： 有限自动机

语法规则： 语法单位的形成规则。规定了如何从单词符号形成语法单位。

语法单位通常包括：表达式、语句、分程序、过程、函数、程序等。

描述工具： 上下文无关文法

2. **语义:** 一组规则，用它可以定义一个程序的意义。

目前大多数编译程序使用基于属性文法的语法制导翻译方法来分析语义。

二、程序语言的基本功能和层次结构

程序语言的基本功能：描述数据和对数据的运算。

所谓程序，本质上说是描述一定数据的处理过程。

程序

|

子程序或分程序、过程、函数

|

语句

|

表达式

|

数据引用　算符　函数调用

三、程序语言的语法描述

1.概念

考虑一个有穷字母表$\sum$字符集

其中每一个元素称为一个字符（符号：是语言中最基本的不可再分的单位）

$\sum$上的字(也叫字符串、符号串) 是指由$\sum$中的字符所构成的一个有穷序列。

不包含任何字符的序列称为空字（空串），记为ε。

用$\sum^*$表示$\sum$上的所有字的全体，包含空字ε。

如: 设$\sum=\{a，b\}$，则$\sum^*=\{\varepsilon,a,b,aa,ab,ba,bb,aaa,\ldots\}$

$\sum^*$的子集U和V的连接（积）定义为$UV=\{\alpha\beta \mid \alpha\in U \ \&\ \beta\in V\}$

例：

设：$U=\{a, aa\}$ $V=\{b, bb\}$
那么：$UV=\{ab, abb, aab, aabb\}$

Σ^*的子集U和V的连接（积）定义为$UV=\{\alpha\beta \mid \alpha\in U \ \&\ \beta\in V\}$

V自身的n次积记为$V^n = VV\ldots V$

规定$V^0=\{\varepsilon\}$，令$V^*=V^0 \cup V^1 \cup V^2 \cup V^3 \cup \ldots$ 称V^*是V的闭包;

记$V^+=VV^*$，称V^+是V的正规闭包（即去除空字）。

例：设$U=\{a, aa\}$

那么：$U^* = \{\varepsilon, a, aa, aaa, aaaa, \ldots\}$

$U^+ = \{a, aa, aaa, aaaa, \ldots\}$

四、推导

1.定义：称$\alpha A\beta$直接推出$\alpha\gamma\beta$，即

$$\alpha A\beta \Rightarrow \alpha\gamma\beta$$

仅当$A \to \gamma$是一个产生式，

且α，$\beta \in (V_T \cup V_N)^*$。

2.如果$\alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \alpha_n$，则我们称这个序列是从$\alpha_1$到$\alpha_n$的一个推导。若存在一个从$\alpha_1$到$\alpha_n$的推导，则称$\alpha_1$可以推导出$\alpha_n$

如：对文法G(E)：$E \to i \mid E+E \mid E*E \mid (E)$

$$E \Rightarrow (E) \Rightarrow (E+E) \Rightarrow (i+E) \Rightarrow (i+i)$$

通常，用 $\alpha_1 \overset{+}{\Rightarrow} \alpha_n$ 表示：从α_1出发，经过一步或若干步，可以推出α_n。

用 $\alpha_1 \overset{*}{\Rightarrow} \alpha_n$表示：从$\alpha_1$出发，经过0步或若干步，可以推出$\alpha_n$。

所以：$\alpha \overset{*}{\Rightarrow} \beta$ 即 $\alpha = \beta$ 或 $\alpha \overset{+}{\Rightarrow} \beta$

定义：假定G是一个文法，S 是它的开始符号。如果 $S \overset{*}{\Rightarrow} \alpha$ ，则α称是一个句型。仅含终结符号的句型是一个句子。文法G所产生的句子的全体是一个语言，将它记为 L(G)。

$$L(G) = \{\alpha \mid S \overset{+}{\Rightarrow} \alpha,\ \alpha \in V_T^*\}$$

【题1】 (i*i+i)是文法 G(E)： $E \rightarrow i \mid E+E \mid E*E \mid (E)$的一个句子。

证明：

$E \Rightarrow (E) \Rightarrow (E+E) \Rightarrow (E*E+E) \Rightarrow (i*E+E) \Rightarrow (i*i+E) \Rightarrow (i*i+i)$

E，(E)，(E*E+E)，…，(i*i+i)是句型。

【题2】文法$G_1(A)$：$A \to c|Ab$，$G_1(A)$的语言

$L(G_1)=\{c，cb，cbb，\ldots\}$ 以c开头，后继若干个b

【题3】文法$G_2(S)$：

$$S \to AB$$

$$A \to aA|a$$

$$B \to bB|b$$

$G_2(S)$的语言?

$L(G_2)=\{a^m b^n|m，n>0\}$

【题4】 给出产生语言为$\{a^n b^n | n \geq 1\}$的文法。

$G_3(S)$:

$S \to aSb$

$S \to ab$

【题5】 给出产生语言为$\{a^m b^n | 1 \leq n \leq m \leq 2n\}$的文法。

$G_4(S)$:

$S \to aSb \mid aaSb$

$S \to ab \mid aab$

从一个句型到另一个句型的推导往往不唯一。

$$E+E \Rightarrow i+E \Rightarrow i+i \quad E+E \Rightarrow E+i \Rightarrow i+i$$

最左推导： 任何一步$\alpha \Rightarrow \beta$都是对α中的最左非终结符进行替换。

最右推导： 任何一步$\alpha \Rightarrow \beta$都是对α中的最右非终结符进行替换。

2.3 语法二义性的判断

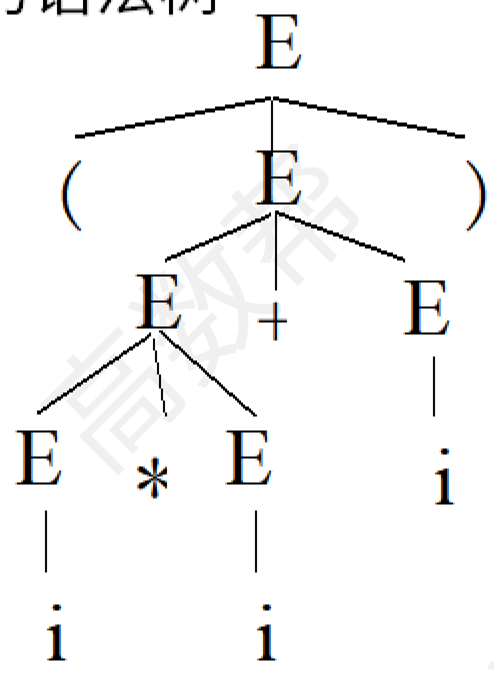
一、语法树

用一张图表示一个句型的推导，称为语法树。

G(E)：$E \to i \mid E+E \mid E*E \mid (E)$

$(i*i+i)$

$(i*i+i)$的语法树



$$E \Rightarrow (E) \Rightarrow (E+E) \Rightarrow (E*E+E) \Rightarrow (i*E+E) \Rightarrow (i*i+E) \Rightarrow (i*i+i)$$

$$E \Rightarrow (E) \Rightarrow (E+E) \Rightarrow (E+i) \Rightarrow (E*E+i) \Rightarrow (E*i+i) \Rightarrow (i*i+i)$$

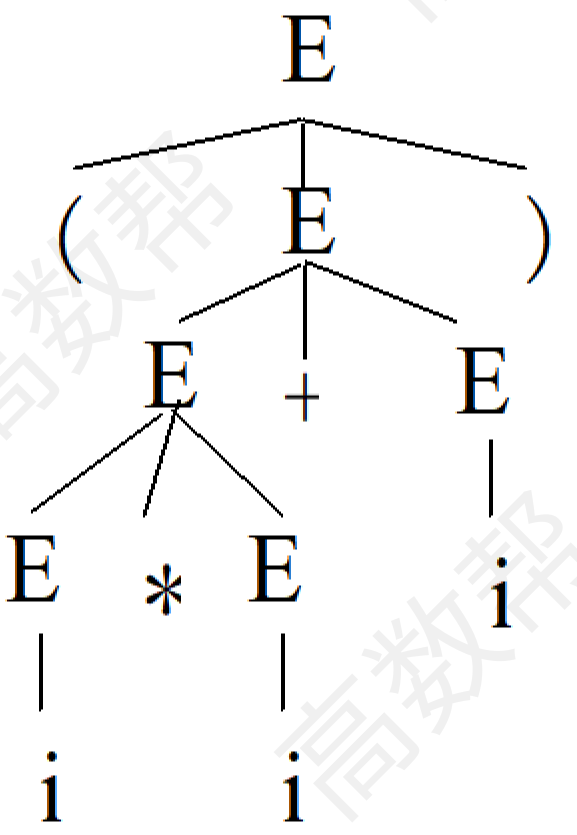
一棵语法树是不同推导过程的共性抽象。

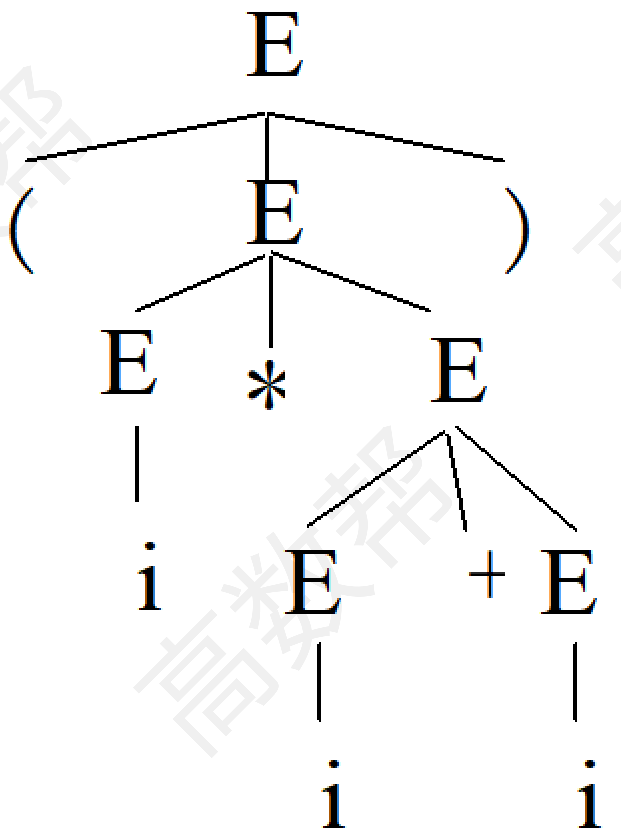
如果使用最左(右)推导，则一个最左(右)推导与语法树一一对应。

一个句型是否只对应唯一一棵语法树？

G(E)：E →i | E+E | E*E | (E)

(i*i+i)





二、二义性

1.定义： 如果一个文法存在某个句子对应两棵不同的语法树，则说这个文法是二义的。

说明：对于文法G，若一个句子存在多个不同的最左推导（或最右推导），那么这个文法是二义的。

$G(E)$： $E \rightarrow i|E+E|E*E|(E)$ 是二义文法。

2.语言的二义性： 一个语言是二义性的，如果对它不存在无二义性的文法。

可能存在G和G'，一个为二义的，一个为无二义的。但$L(G)=L(G')$

3.二义性问题是不可判定问题，即不存在一个算法，它能在有限步骤内，确切地判定一个文法是否是二义的。

4.可以找到一组无二义文法的充分条件。

二义文法：

G(E)：$E \rightarrow i|E+E|E*E|(E)$

无二义文法：

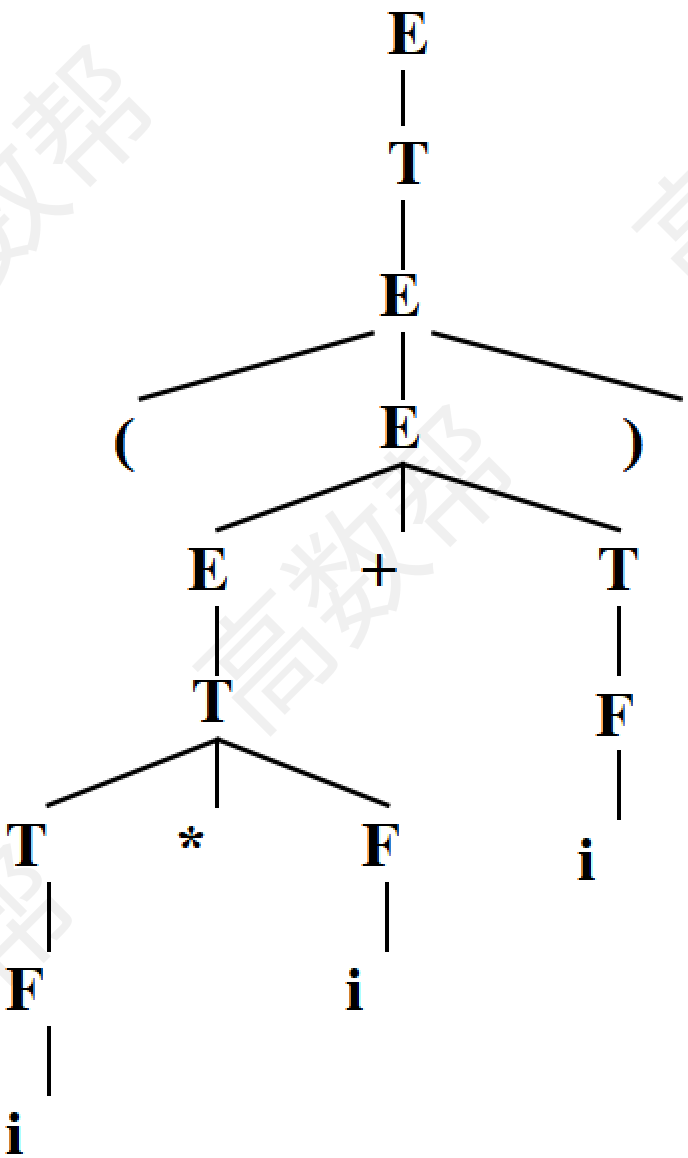
G(E)：$E \rightarrow T \mid E+T$

$T \rightarrow F \mid T*F$

$F \rightarrow (E) \mid i$

考虑句子(i*i+i)

$$
\begin{aligned}
E &\Rightarrow T \\
&\Rightarrow F \\
&\Rightarrow (E) \\
&\Rightarrow (E+T) \\
&\Rightarrow (T+T) \\
&\Rightarrow (T*F+T) \\
&\Rightarrow (F*F+T) \\
&\Rightarrow (i*F+T) \\
&\Rightarrow (i*i+T) \\
&\Rightarrow (i*i+F) \\
&\Rightarrow (i*i+i)
\end{aligned}
$$



描述程序设计语言时，对于上下文无关文法的限制：

1 不含P→P形式的产生式

2 每个非终结符P必须有用处

即： $S \overset{*}{\Rightarrow} \alpha P \beta$

$P \overset{*}{\Rightarrow} r \qquad r \in V_T^*$

2.4 文法的分类及应用

Chomsky于1956年建立形式语言体系，他把文法分成**四种类型**：0，1，2，3型。

与上下文无关文法一样，它们都由四部分组成，但对产生式的限制有所不同。

0型(短语文法，图灵机)：

产生式形如： $\alpha \rightarrow \beta$

其中：$\alpha \in (V_T \cup V_N)^*$且至少含有一个非终结符；$\beta \in (V_T \cup V_N)^*$

1型(上下文有关文法，线性界限自动机)：

产生式形如： $\alpha \rightarrow \beta$

其中：$|\alpha| \leq |\beta|$，仅 $S \rightarrow \varepsilon$ 例外。

2型(上下文无关文法，非确定下推自动机)：

产生式形如： $A \rightarrow \beta$

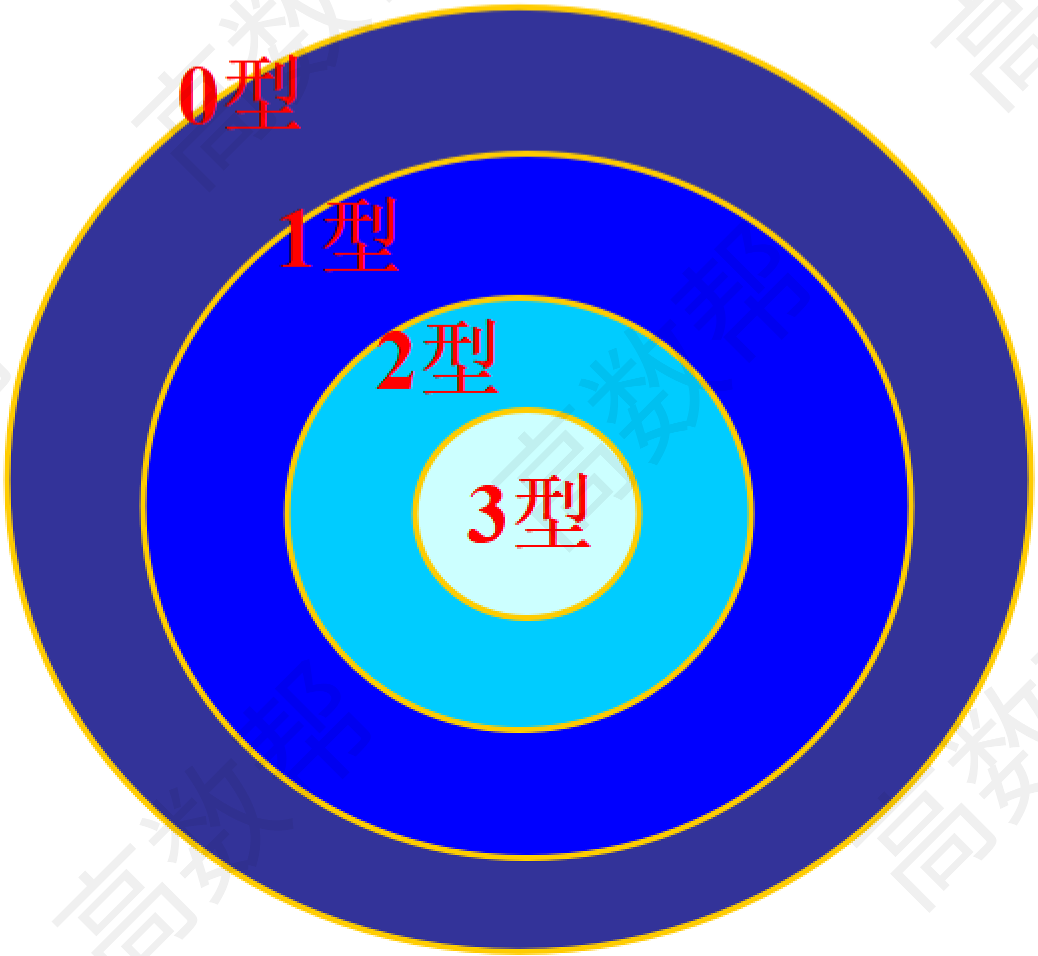
其中：$A \in V_N$；$\beta \in (V_T \cup V_N)^*$。

3型(正规文法，有限自动机)：

产生式形如：$A \rightarrow \alpha B$ 或 $A \rightarrow \alpha$　其中：$\alpha \in V_T^*$；A，$B \in V_N$　右线性文法

产生式形如：$A \rightarrow B\alpha$ 或 $A \rightarrow \alpha$　其中：$\alpha \in V_T^*$；A，$B \in V_N$　左线性文法

四种类型描述能力比较



【题6】 $L=\{a^n b^n | n\geq 1\}$ 不能由正规文法产生，但可由上下文无关文法产生：

G(S)： $S \rightarrow aSb|\ ab$

$L=\{a^n b^n c^n | n\geq 1\}$不能由上下文无关文法产生，但可由上下文有关文法产生：

G(S)： $S \rightarrow aSBC|\ aBC$

$CB \rightarrow BC$

$aB \rightarrow ab$

$bB \rightarrow bb$

$bC \rightarrow bc$

$cC \rightarrow cc$

视频讲解更清晰

【题7】指出语言类型并给出语法规则

$L1 = \{a^i b^i b^i | i>0\}$　上下文有关文法，见上页

$L2 = \{a^i b^i c^j | i,j>0\}$

$L3 = \{a^i b^j c^k | i,j,k>0\}$

$L4 = \{a^i b^j | i>j>0\}$

L2：有自嵌套（$a^i b^i$）结构，为2型语言，因此是上下文无关文法

$G[S]:S \rightarrow AC$

$A \rightarrow aAb \mid ab$

$C \rightarrow Cc \mid c$

$L3 = \{a^i b^j c^k | i,j,k>0\}$

3型文法（正规文法）

G[S]：右线性：$S \rightarrow aS \mid aB$

$B \rightarrow bB \mid bC$

$C \rightarrow cC \mid c$

左线性：$S \rightarrow Sc \mid Bc$

$B \rightarrow Bb \mid Ab$

$A \rightarrow Aa \mid a$

$L4 = \{a^i b^j | i>j>0\}$

上下文有关文法

令$i-j=k(k>0)$,则$i=j+k$

$a^i b^j = a^{j+k} b^j = a^j a^k b^j$

G[S]:

$S \rightarrow aAb \mid aSb$

$A \rightarrow a \mid aA$