语言和文法

短语结构文法

- 基本定义：一个字母表 Σ 是一个非空有限的集合，这个集合中的元素称为符号. 由 Σ 中的符号组成的有限长度的字符串称为 Σ 上的词. 不包含任何符号的字符串称为空串，记为 λ. Σ 上所有词的集合记为 Σ^*，Σ^* 的任何子集称为 Σ 上的一个语言.

- 短语结构文法的定义：一个短语结构文法是一个四元组 $G=(\Sigma,T,S,P)$ 其中 Σ 是一个字母表，$T\subset\Sigma$ 是所有的终止符集， $S\in\Sigma$ 是起始符，P 是产生式集合，集合 $\Sigma-T$ 即为 N，N 中的元素成为非终止符，P 中的每个产生的左端必须至少包含一个非终止符。
  Note: 字母表中包含终止符和非终止符，自然包含起始符。**起始符只有一个。**

- 派生的定义：Σ 上的两个字符串 $\omega_0=lz_0r$ 和 $\omega_1=lz_1r$，若存在产生式 $z_0\to z_1$，则称 ω_0 可直接派生 ω_1，记作 $\omega_0\Rightarrow\omega_1$，若由一系列产生式可将 ω_0 转化为 ω_1，则称ω_0 可派生 ω_1，记作 $\omega_0\overset{*}{\Longrightarrow}\omega_1$。由 ω_0 转化成 ω_1 过程中的每一步直接派生得到的序列称为派生。

短语结构文法的类型

根据产生式的类型可以将短语结构文法分类：

- **0型文法：** 产生式没有限制

- **1型文法：** 只有两种类型的产生式：$\omega_0=lAr\to\omega_1=lar$，其中 l 和 r 是由 **0个或者多个** 终结符或者非终结符构成的串，a 是终结符或非终结符构成的 **非空串** 。或者 $S\to\lambda$，但是 S 不能出现在任何其他产生式的右边。

- **2型文法：** 只有形如 $\omega_0\to\omega_1$ 的产生式，其中 ω_0 是一个单个的非终结符的符号。

- **3型文法：** 只有形如 $\omega_0\to\omega_1$ 的产生式，同时满足 $\omega_0=A$ 且 $\omega_1=aB$ 或者 $\omega_1=a$，其中 A 和 B 是非终结符，a 是终结符，或者满足 $\omega_0=S,\ \omega_1=\lambda$

1型文法称为上下文有关文法，2型文法称为上下文无关文法，3型文法称为正则文法。上下文有关文法是非缩约的（单调的，非收缩的）。
Note: 非缩约是指字符串在派生过程中只能不断变长，而不能缩短。

派生树

适用于上下文无关语法生成的语言。用数的根节点表示起始符，内部节点表示派生过程中产生的非终止符，叶节点表示终止符。一个非终止符对应的产生式右端有多少符号，这个节点就有多少子节点。

巴科斯-诺尔范式

用来表示2型文法的另外一种方式，将左端都是同一个非终结符的产生式合并为一个式子，如产生式 $A\to Aa,\ A\to a,\ A\to AB$ 可以合并为 $\langle A\rangle ::= \langle A\rangle a|a|\langle A\rangle\langle B\rangle$