第5章 上下文无关语言

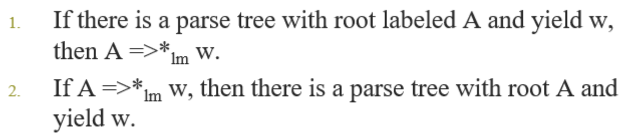
一些术语

终结符，非终结符，产生式，初始符号，产生式，推导，最左推导，最右推导，句型，句子。

语法分析树

语法分析树的产物：The concatenation of the labels of the leaves in left-to-right order

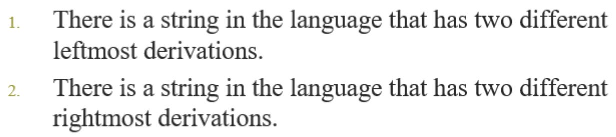
语法分析树，最左推导，最右推导，任意推导是等价的（以下结论对最右推导同）



1. If there is a parse tree with root labeled A and yield w, then A =>*$_{lm}$ w.
2. If A =>*$_{lm}$ w, then there is a parse tree with root A and yield w.

二义性

A CFG is *ambiguous* if there is a string in the language that is the yield of two or more parse trees

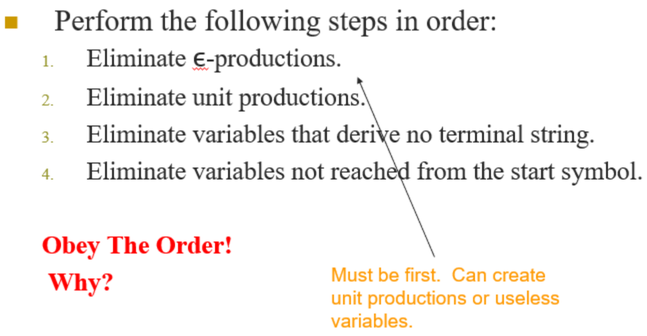


1. There is a string in the language that has two different leftmost derivations.
2. There is a string in the language that has two different rightmost derivations.

歧义是语法的性质，而不是语言的！

固有歧义性：有的语言是固有歧义的，即其不存在无歧义的文法。

乔姆斯基范式



- Perform the following steps in order:
  1. Eliminate ϵ-productions.
  2. Eliminate unit productions.
  3. Eliminate variables that derive no terminal string.
  4. Eliminate variables not reached from the start symbol.

Obey The Order!
Why?

Must be first. Can create unit productions or useless variables.

消除ε产生式

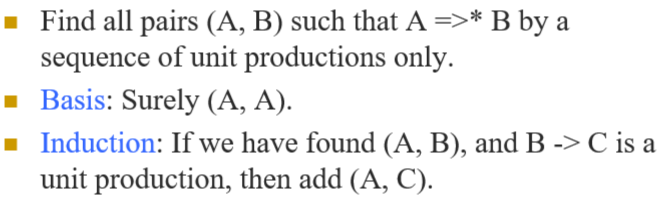
1）找到“可空符号”，即可以推导出ε的符号

算法：如果A->ε，则A是可空符号；如果存在产生式A->a，且a中全是可空符号，则A是可空符号。

2）将A -> X1...Xn 转换成一系列产生式：即对于右侧可空符号集的每一个子集，在右边产生式提前将其消去，得到一个新的产生式。

消除单元产生式

1）寻找单元对（注意递归中必须使用单元产生式）



- Find all pairs (A, B) such that A =>* B by a sequence of unit productions only.
- Basis: Surely (A, A).
- Induction: If we have found (A, B), and B -> C is a unit production, then add (A, C).

2）对于所有的非单元产生式B->a，加入A->a

消除无法推导出终结符串的非终结符

1）如果A->w，则A可推出终结符串

2）如果A->X1...Xn，且X1...Xn均可推出终结符串，则A可推出终结符串

消除从初始符号不可达的非终结符

1）S->a，则a中的非终结符可达

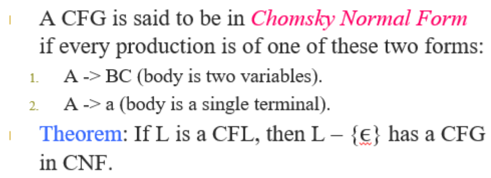
2）A->a且A可达，则a中的非终结符可达

删除顺序不能反，先删“下不去”，再删“过不来”

理由：先删不能推导出终结字符串的符号，得到剩下的符号都是可以推导出终结字符串的；再考虑可达性，选出从S可达的符号。

如果先考虑可达性，那么会有一些不能推导出终结字符串的符号，它们从S可达，并且其上的产生式导致某些本来不可达的符号也变得可达。之后再删除不能推导出终结字符串的符号，其上的产生式被删除，但此时可能会有部分不可达的符号留了下来

乔姆斯基范式



- A CFG is said to be in *Chomsky Normal Form* if every production is of one of these two forms:
  1. A -> BC (body is two variables).
  2. A -> a (body is a single terminal).
- Theorem: If L is a CFL, then L – {ϵ} has a CFG in CNF.