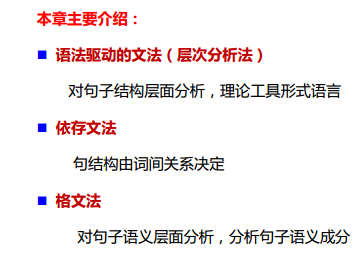
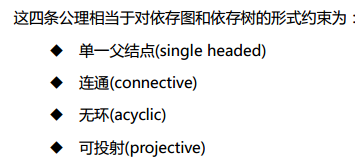
# 第二章 乔布斯基形式文法





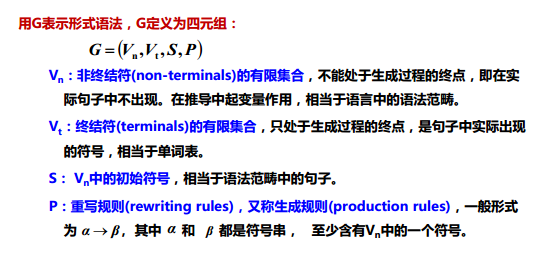


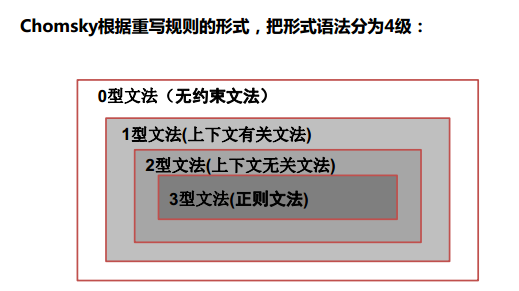


格文法：施事格，受事格，工具格

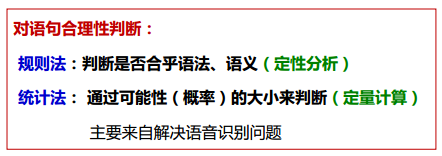


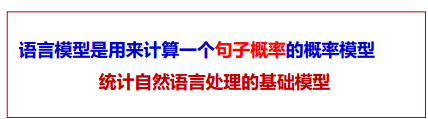
**下为乔布斯基形式文法：**





# 第五章 语言模型



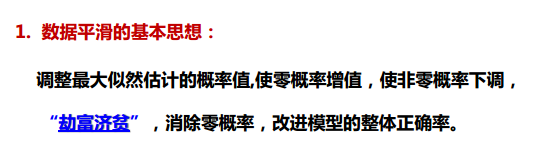




参数估计（模型训练）：获得模型中所有的条件概率（模型参数）

用最大似然估计计算参数

**语言模型对于训练文本的类型、主题和风格等都十分敏感**

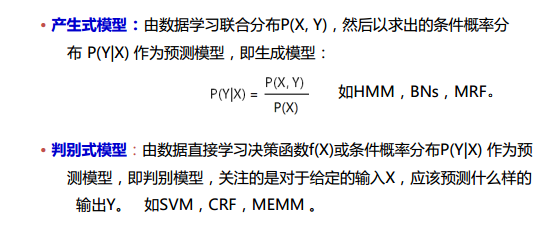


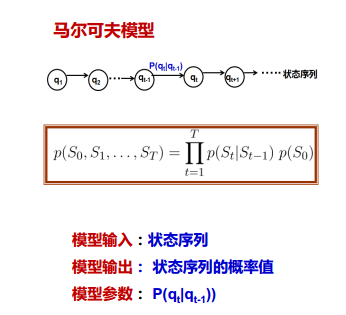


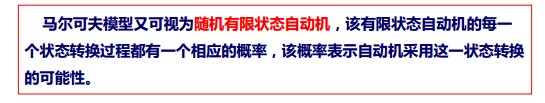
评价：实用方法和理论方法（困惑度）

改进：基于缓存的语言模型、基于混合方法的语言模型

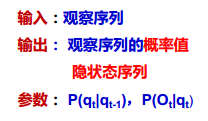
# 第六章 概率图模型

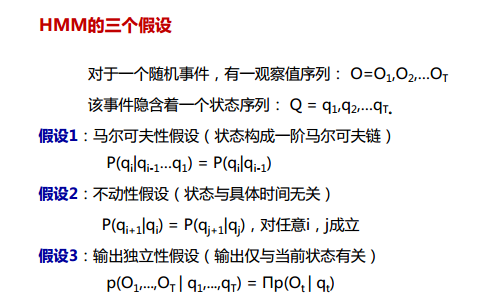




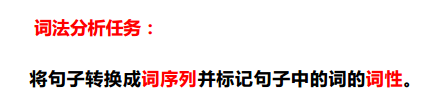


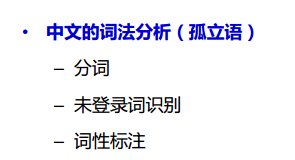
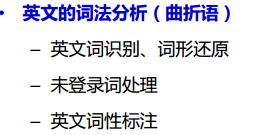






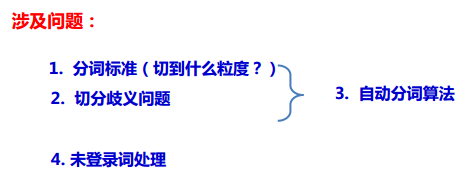
# 第七章 词法分析







中文自动分词

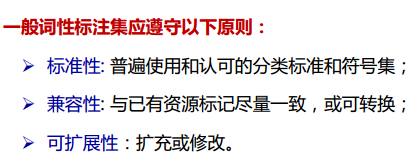


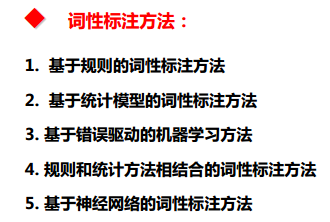
分词基本原则：切分、合并原则

切分歧义：交集型歧义、组合型歧义

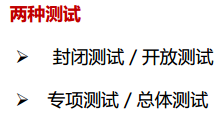
自动分词算法：基于规则的方法[正向、逆向最大匹配]、基于统计的方法[全切分、最少分词法]







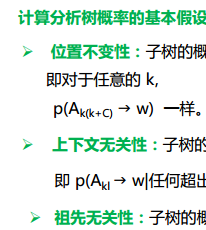
词法分析评价：P\R\F



# 第八章 句法分析

**完全句法分析算法**

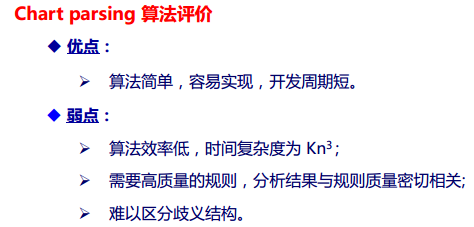
**PCFG(概率无关上下文文法)**



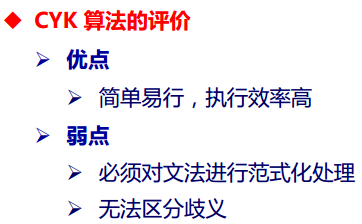
 **1. 线图分析法（Chart）[节点和边组成]**

**活动边集(ActiveArc)** 、**线图Chart（非活动边）、**

**代理表(待处理表)(Agenda)、输入缓冲区**

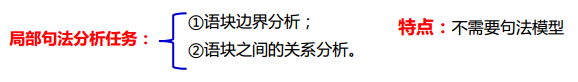
 **2. CYK 分析算法**

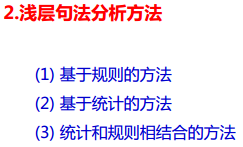
**(CYK) 算法思想：通过构造识别矩阵进行分析**



**3. 完全句法分析评估**

**P、R、F、词性标注准确率、交叉括号数和交叉准确率**

**浅层句法分析方法（局部句法分析）**

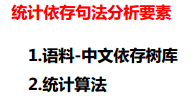


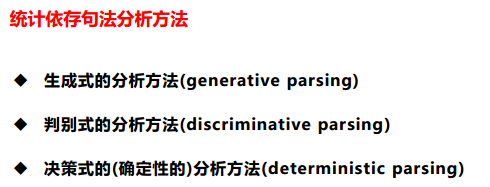


**基于最大熵的组块分析：定义组块、标签集、训练语料处理、最大熵建模、语料和特征模板训练模型、应用模型求解**

**依存句法分析**







**依存评估**



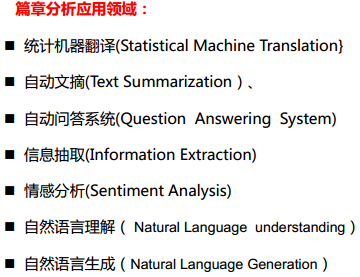


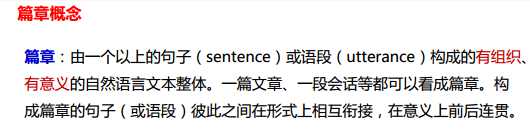






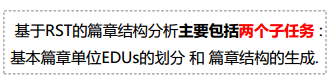
# 第十章 篇章分析

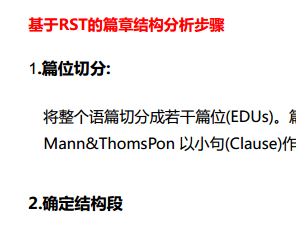




篇章结构：微观、宏观篇章结构







RST特点：语篇核心、关联词语和关系命题、从属连接

