



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

自然语言理解与机器翻译课程

# 句法解析

李辰

2024年9月

# 课程提纲

一、句法解析

二、语义关系抽取



# 3/3

概率基础



语言模型



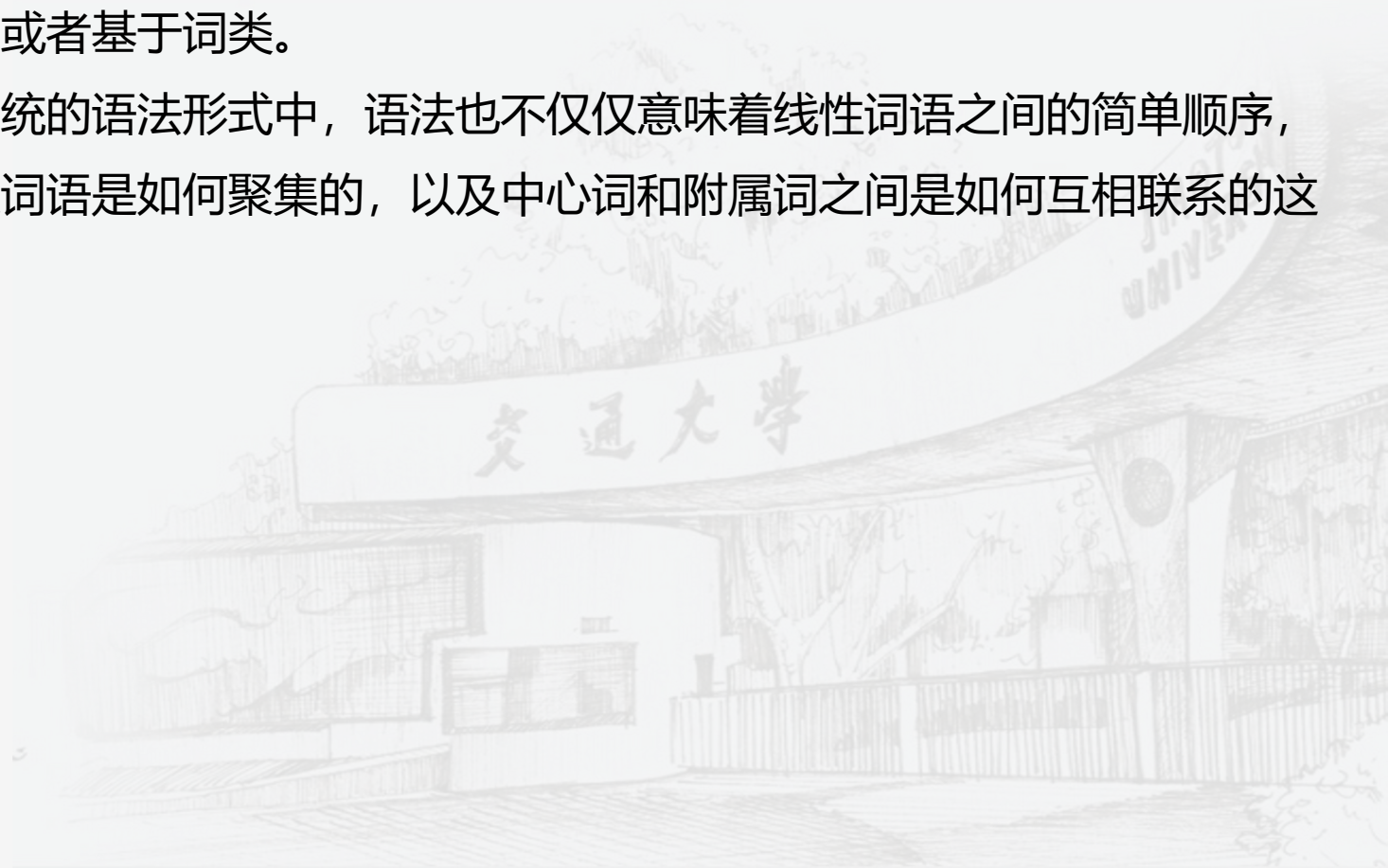
NLU任务



# 句法解析

## 什么是语法？

- 人们说话的方式都存在一些结构和规则，语言学中的语法分析的目标就是努力分离出这些语法结构。
- 我们一般所说的语法实际上就是指描述词语排列的方法，而排列的方式或者直接基于词，或者基于词类。
- 即使在最传统的语法形式中，语法也不仅仅意味着线性词语之间的简单顺序，它必须描述词语是如何聚集的，以及中心词和附属词之间是如何互相联系的这一类问题。





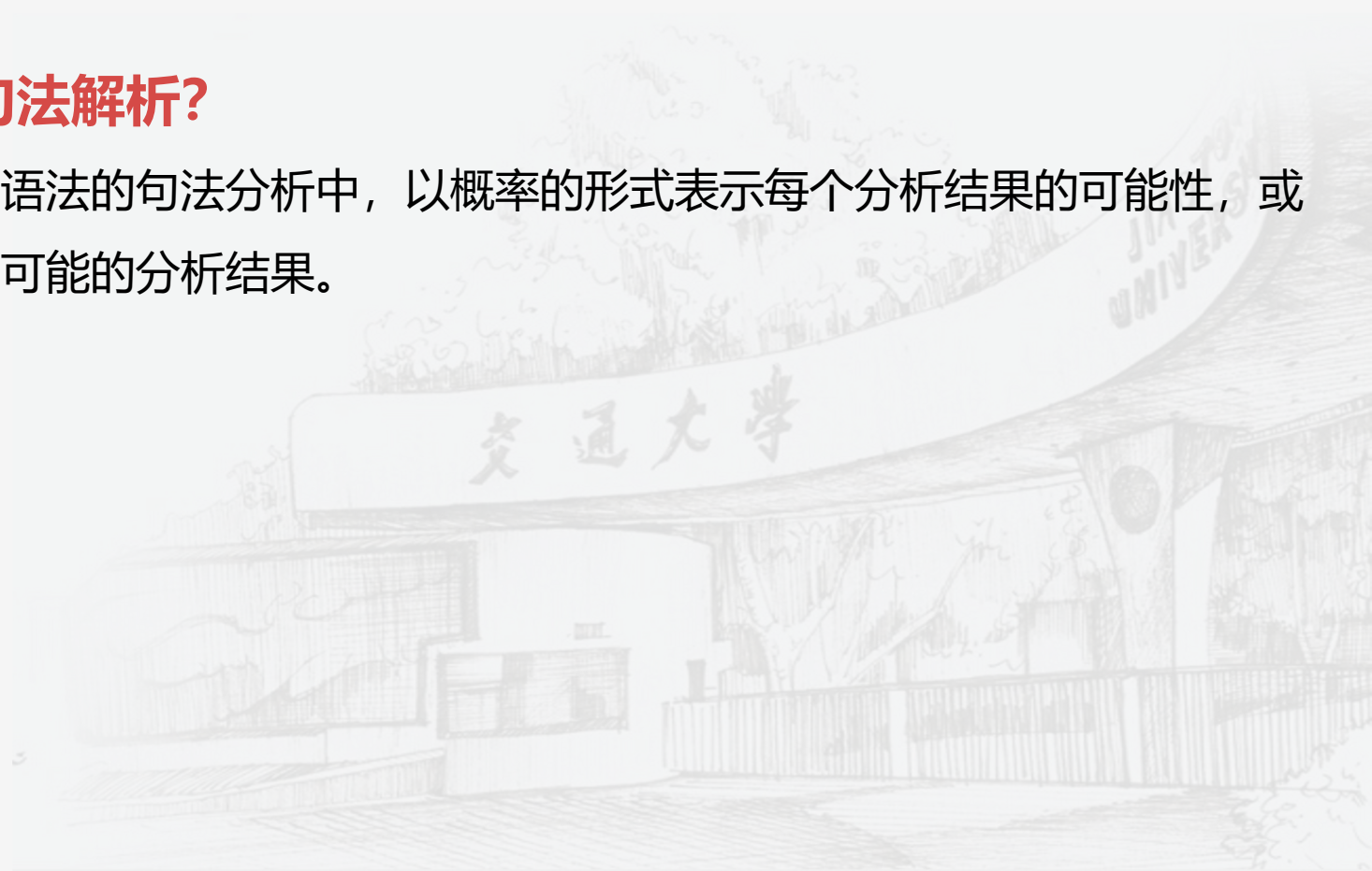
# 句法解析

## 什么是句法解析？

- 句法分析的思想是能够根据某种语法给出一个句子的句法分析树。

## 什么是概率句法解析？

- 在基于某种语法的句法分析中，以概率的形式表示每个分析结果的可能性，或者只返回最可能的分析结果。



# 句法解析

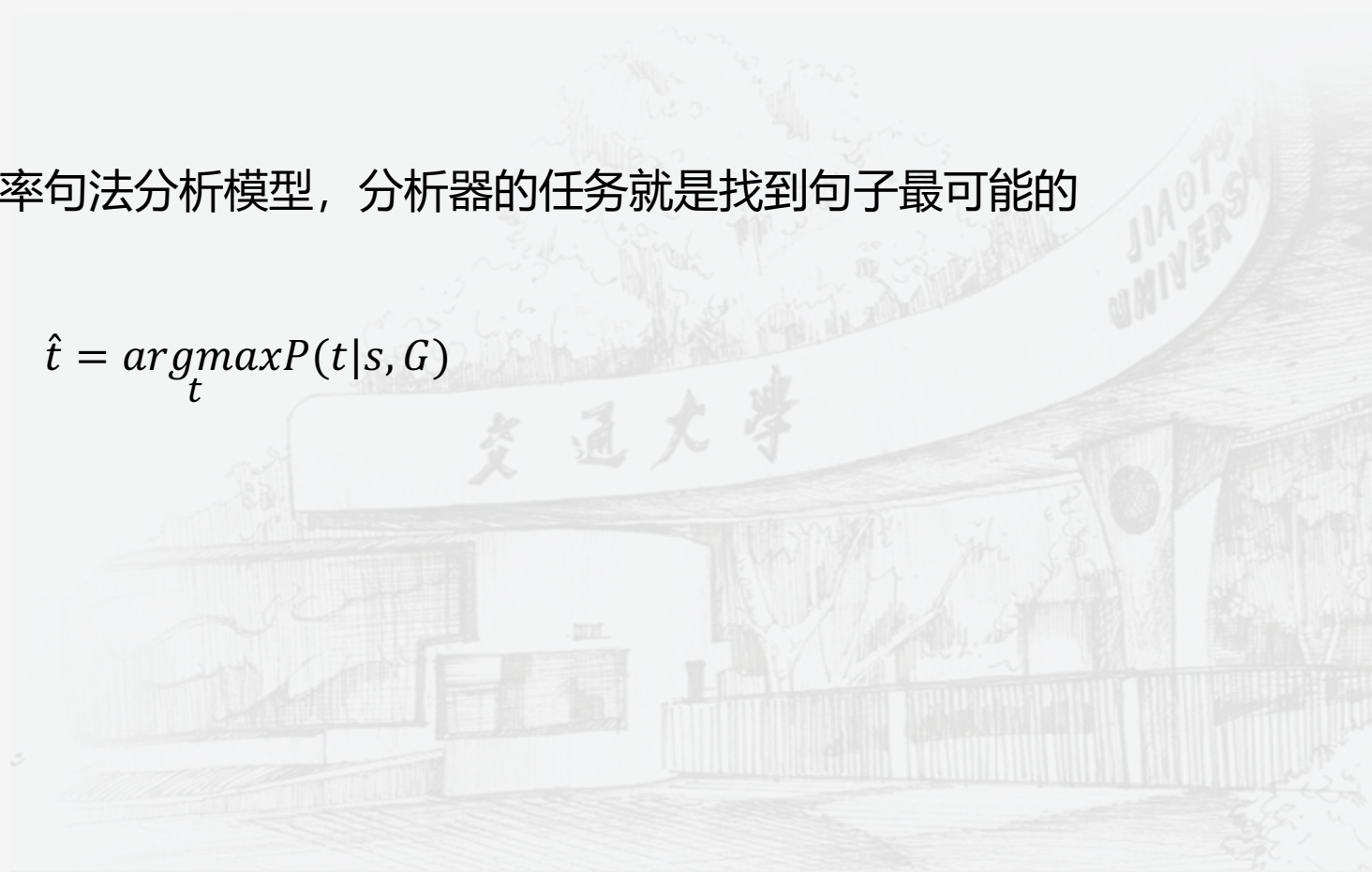
## 概率句法分析模型

- 对一个句子  $s$ ，基于语法  $G$ ，给出分析树  $t$  的概率：

$$P(t|s, G) \quad \text{其中} \quad \sum_t P(t|s, G) = 1$$

- 给定一个概率句法分析模型，分析器的任务就是找到句子最可能的分析结果  $\hat{t}$ ：

$$\hat{t} = \underset{t}{\operatorname{argmax}} P(t|s, G)$$



# 句法解析与语言模型

## 句法分析模型与语言模型的关系

- 可以直接由一个语言模型构造一个句法分析模型。整个句子的概率如下：

$$\begin{aligned} P(s) &= \sum_t P(s, t) \\ &= \sum_{\{t: \text{yield}(t)=s\}} P(t) \end{aligned}$$

- 那么最好的句法分析结果是：

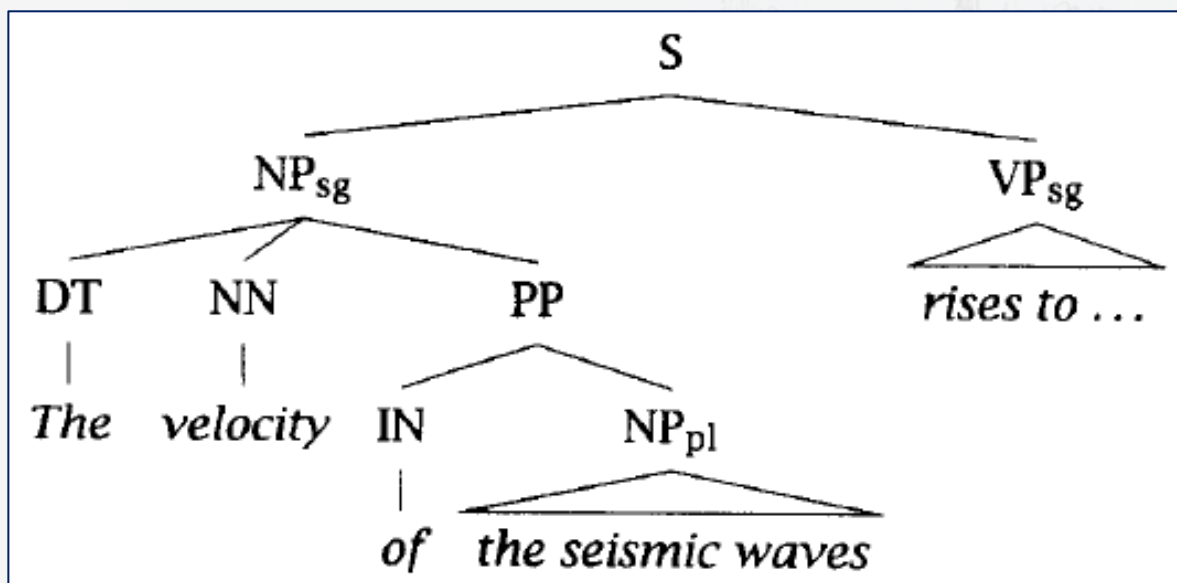
$$\hat{t} = \underset{t}{\operatorname{argmax}} P(t|s) = \underset{t}{\operatorname{argmax}} \frac{P(t, s)}{P(s)} = \underset{t}{\operatorname{argmax}} P(t, s)$$

- 还没有有效的方法可以把任意的句法分析模型转化成语言模型。

# 概率上下文无关文法

## 概念

- 语言有着复杂的递归结构，基于树的模型可以体现这一点。
- 概率上下文无关文法 (Probabilistic Context Free Grammar, 简称PCFG)
- 为规则增添了概率的简单CFG，指明了不同重写规则的可能性大小。

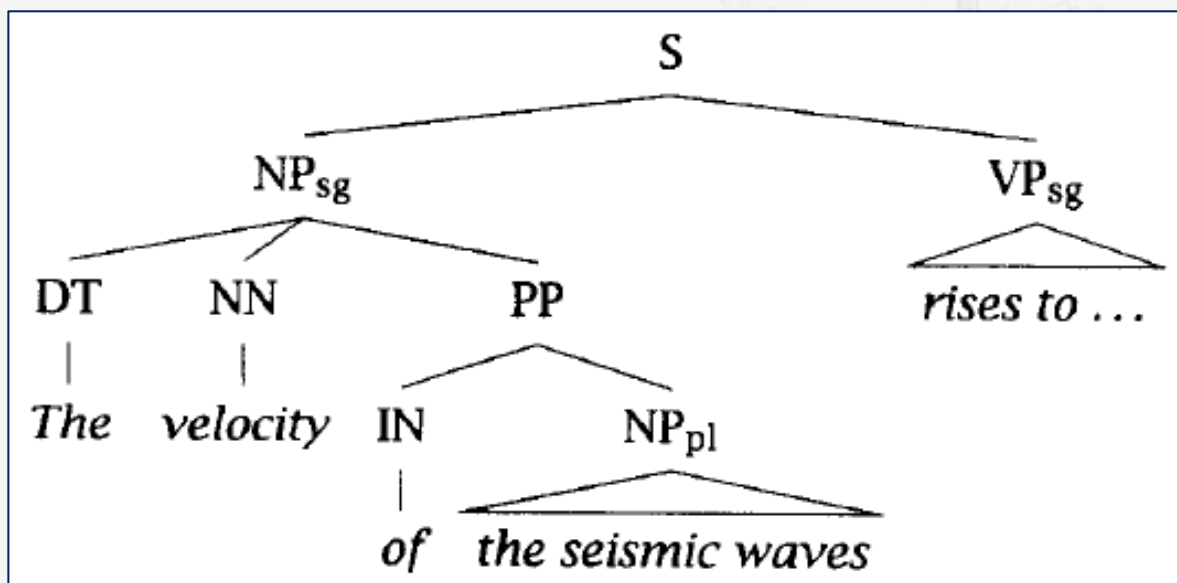




# 概率上下文无关文法

## 模型的假设

- 位置不变性：一颗子树的概率并不依赖于词串中它所支配的词语所处位置；
- 上下文无关性：子树的概率与不被子树支配的词无关；
- 祖先无关性：子树的概率和子树之外的祖先节点无关。



# 概率上下文无关文法

## 三个基本问题

- 问题1：词串概率计算

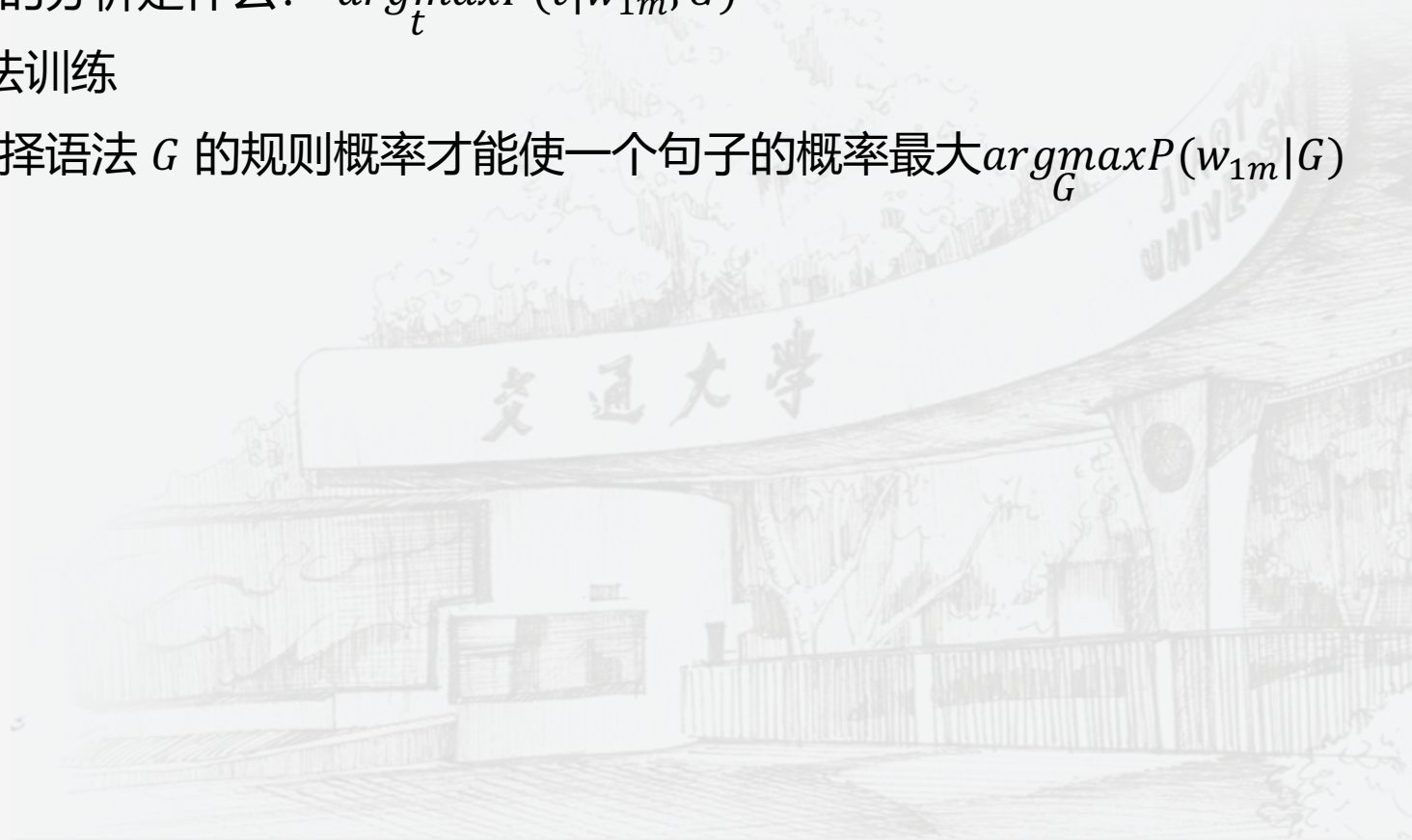
对应于一个语法  $G$ ，句子  $w_{1m}$  的概率  $P(w_{1m}|G)$  是什么？

- 问题2：最佳句法分析

句子最可能的分析是什么： $\operatorname{argmax}_t P(t|w_{1m}, G)$

- 问题3：文法训练

我们怎么选择语法  $G$  的规则概率才能使一个句子的概率最大  $\operatorname{argmax}_G P(w_{1m}|G)$



# 概率上下文无关文法

## 三个基本问题

- 为简单起见，我们只考虑满足Chomsky规范形式语法的例子，这种语法只含有特定形式的一元和二元规则：

$$N^i \rightarrow N^j N^k$$

$$N^i \rightarrow w^j$$

- 则有：

$P(N^j \rightarrow N^r N^s | G)$       如果有  $n$  个非终结节点，则有一个  $n^3$  的参数矩阵

$P(N^j \rightarrow w^k | G)$       如果有  $V$  个终结节点，则有  $nV$  个参数

for  $j = 1, \dots, n,$

$$\sum_{r,s} P(N^j \rightarrow N^r N^s) + \sum_k P(N^j \rightarrow w^k) = 1$$

# 概率上下文无关文法

## 问题1：一个句子的概率

$$\begin{aligned} P(w_{1m}) &= \sum_t P(w_{1m}, t) \\ &= \sum_{\{t: \text{yield}(t) = w_{1m}\}} P(t) \end{aligned}$$

## 一棵分析树的概率

- 把构造它的各个子树的规则概率相乘即可。

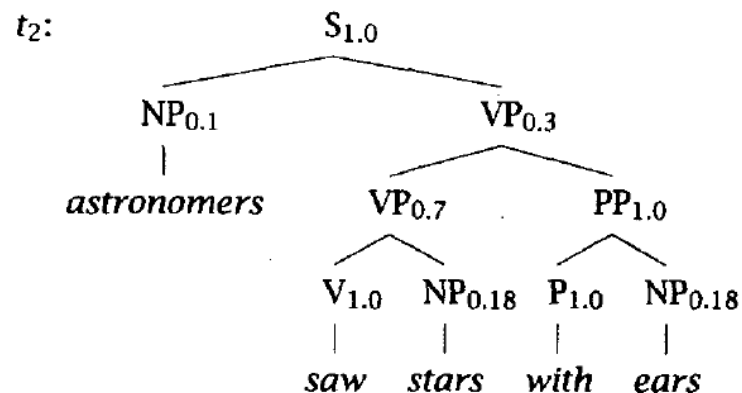
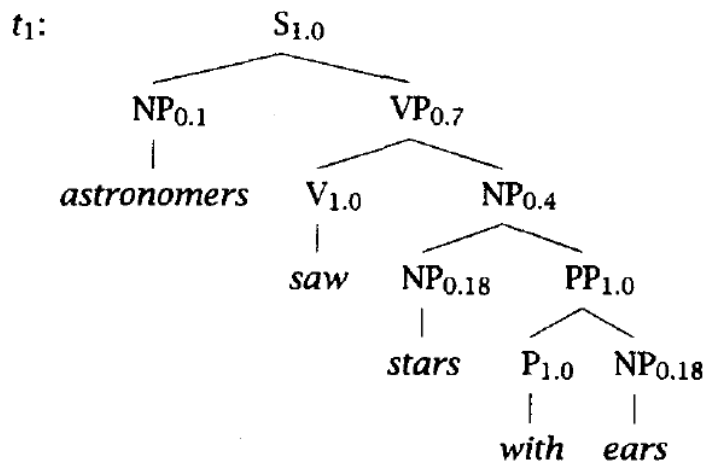


# 概率上下文无关文法

## 问题1：一个句子的概率

- 例1：假定有下表所示的语法，句子astronomers saw stars with ears则有两个概率句法分析树。

$S \rightarrow NP VP$	1.0	$NP \rightarrow NP PP$	0.4
$PP \rightarrow P NP$	1.0	$NP \rightarrow \textit{astronomers}$	0.1
$VP \rightarrow V NP$	0.7	$NP \rightarrow \textit{ears}$	0.18
$VP \rightarrow VP PP$	0.3	$NP \rightarrow \textit{saw}$	0.04
$P \rightarrow \textit{with}$	1.0	$NP \rightarrow \textit{stars}$	0.18
$V \rightarrow \textit{saw}$	1.0	$NP \rightarrow \textit{telescopes}$	0.1



# 概率上下文无关文法

## 问题1：一个句子的概率

- 例1：假定有下表所示的语法，句子astronomers saw stars with ears则有两个概率句法分析树。

$$\begin{aligned}P(t_1) &= 1.0 \times 0.1 \times 0.7 \times 1.0 \times 0.4 \times 0.18 \times 1.0 \times 1.0 \times 0.18 \\ &= 0.000\ 907\ 2\end{aligned}$$

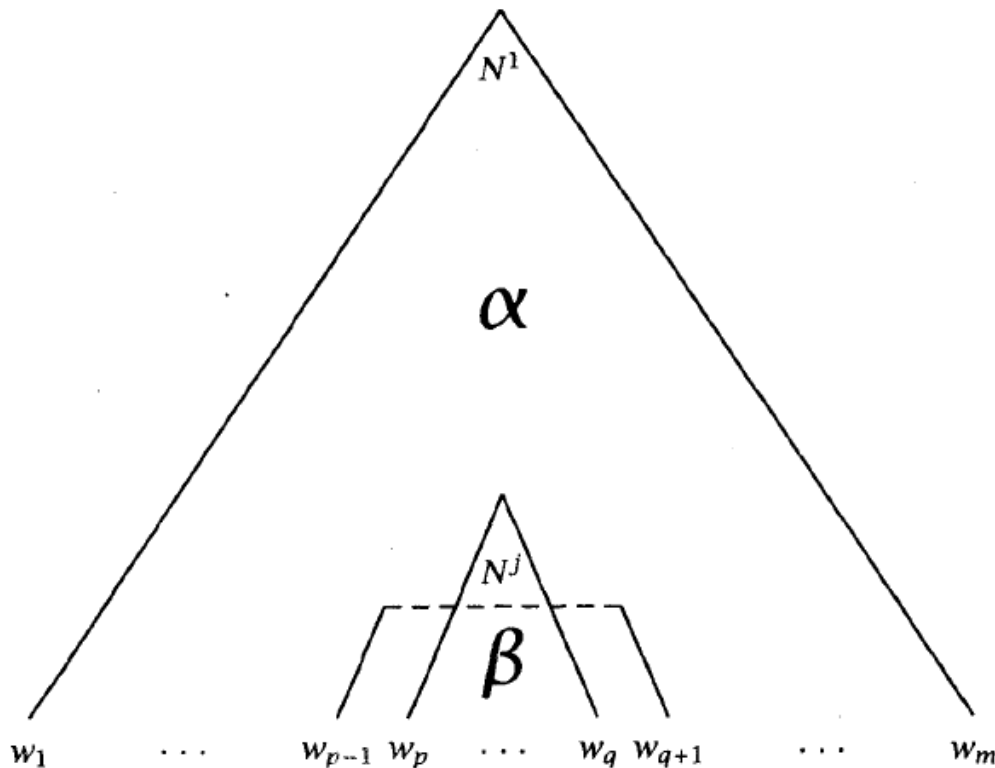
$$\begin{aligned}P(t_2) &= 1.0 \times 0.1 \times 0.3 \times 0.7 \times 1.0 \times 0.18 \times 1.0 \times 1.0 \times 0.18 \\ &= 0.000\ 680\ 4\end{aligned}$$

$$P(w_{15}) = P(t_1) + P(t_2) = 0.001\ 587\ 6$$

# 概率上下文无关文法

## 内部概率与外部概率

- 内部概率:  $\beta_j(p, q) = P(w_{pq} | N_{pq}^j, G)$
- 外部概率:  $\alpha_j(p, q) = P(w_{1(p-1)}, N_{pq}^j, w_{(q+1)m} | G)$



## 问题2：最佳句法分析

- 可以通过使用内部算法寻找总体最优来构建一个Viterbi风格的算法，这个算法能够找到句子最可能的分析，同时记下给出这个总体最优的规则。
- 我们希望找到以某个非终结符为根，并覆盖某个特定子词串的部分分析树的最高概率。我们依然采用累积器的方法：

$\delta_i(p, q)$  = 一棵子树  $N_{pq}^i$  分析的最高内部概率



## 问题3：文法训练

- 训练思路
  - 基本思路是语法学习或语法归纳。
  - 但这只是在限定的意义下，即我们假定语法由一定数目的终结符和非终结符以及预先给定的开始符构成，同时假设规则集合也是提前给定的。
  - 通常我们假定所有的重写规则都存在，或者假定一些语法中预定义的结构。
  - 训练这个语法的过程很简单，即只需要在这个体系结构内找到分配给不同语法规则的最优概率。



## 问题3：文法训练

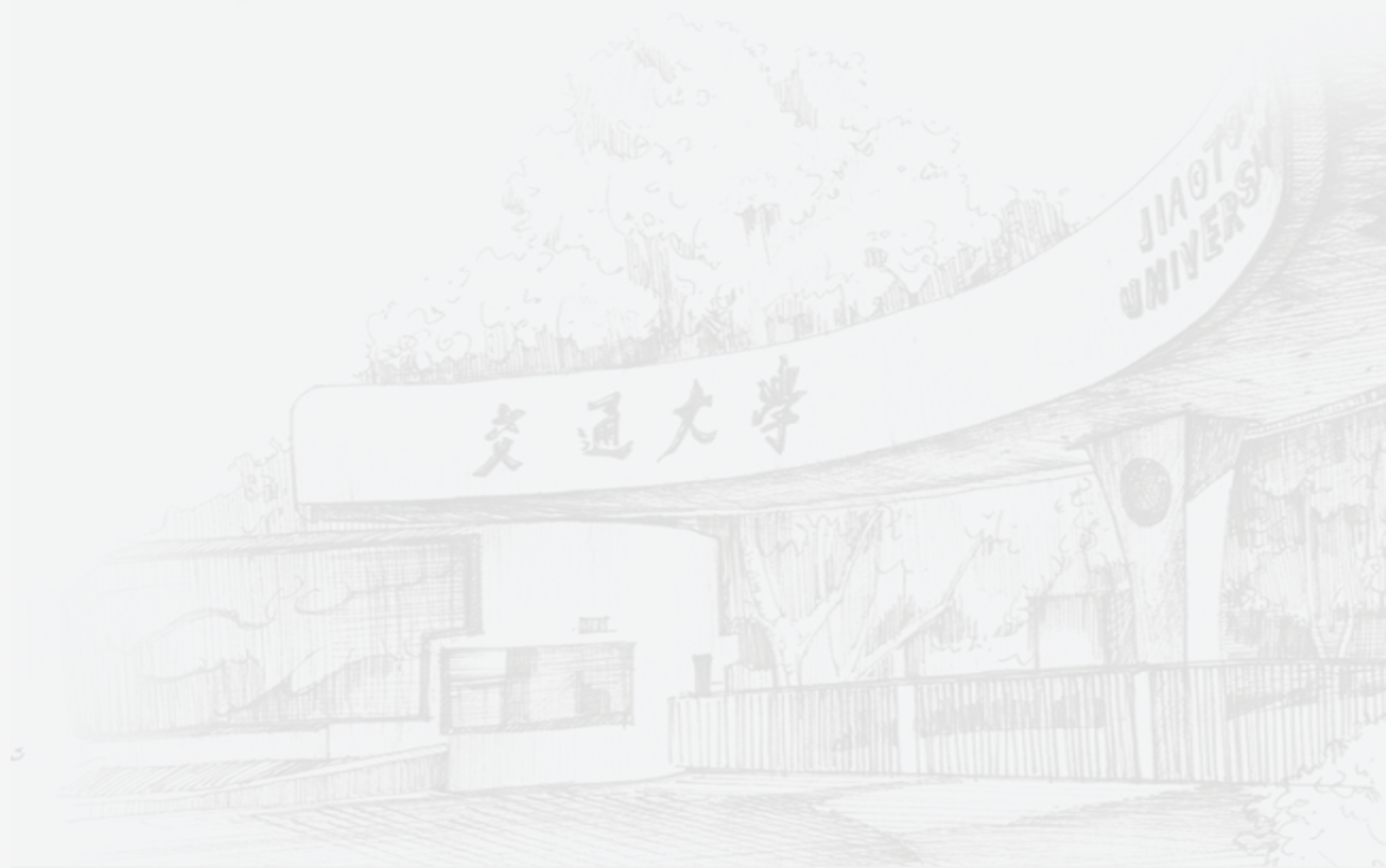
- 训练算法
  - 参照HMM的训练方法，构造一个EM训练算法，即内部-外部算法，它允许我们在未标注的句子训练PCFG的参数。
  - 基本假设是：一个定义良好的词法可能会生成训练集中的所有句子，因此需要找到最大化训练数据似然性的语法。
  - 我们首先基于单个句子进行训练，然后通过假定句子之间的独立性扩展到一个含有大量句子的大规模真实训练语料库上。



# 依存句法

## 概念

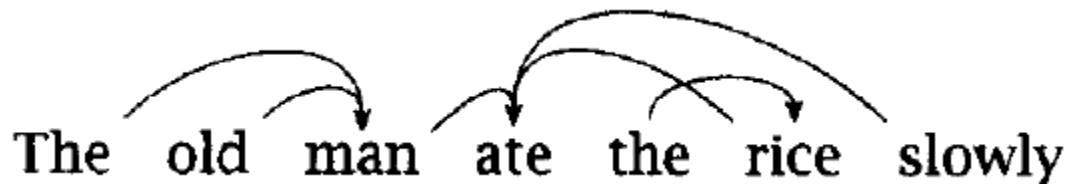
- 采用词之间的依存关系来描述语言结构，这样的—个框架称为依存句法。
- 在依存句法中，某个词是句子的中心词，其它的词或者依赖于这个词，或者依赖于那些通过—系列依存关系与中心词有联系的其它词。



# 依存句法

## 依存关系的表示

- 依存关系通常由带箭头的曲线表示，如：



## 依存句法的意义

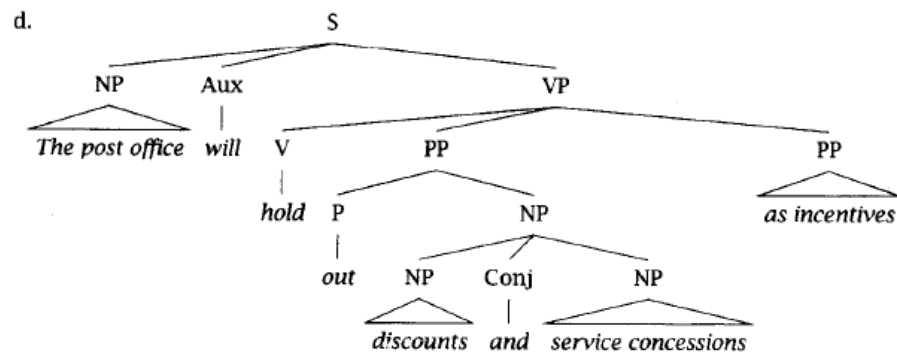
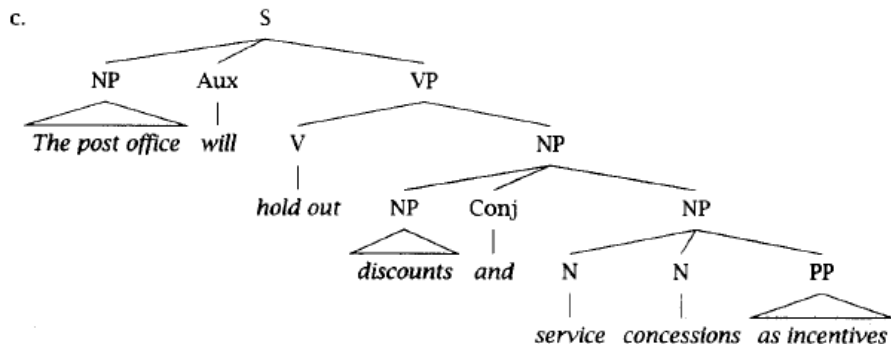
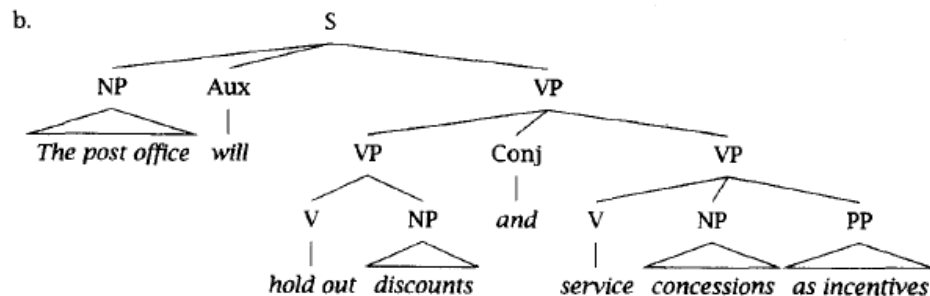
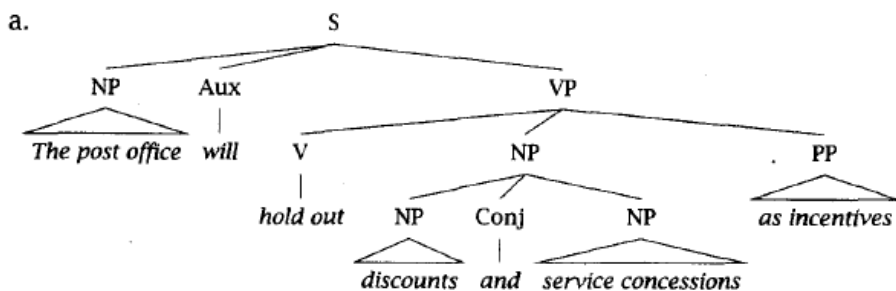
- 在统计NLP中依照依存关系来思考是很有用的，此外，了解短语结构和依存模型之间的关系也很有价值。



# 句法消歧分析

## 歧义的例子

- The post office will hold out discounts and service concessions as incentives.



# 句法消歧分析

## 歧义的处理

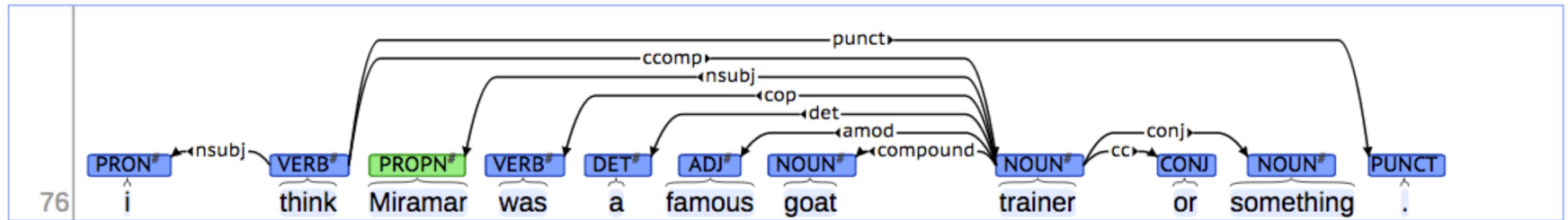
- 在传统的句法分析器中，作为分析结果的句法分析树通常被送到下游的语义和世界知识模型中，并由它们做出选择。
- 一个统计句法分析器通常在运行的过程中使用各种扩展的词语符号和类别搭配来代替语义和世界知识，以进行消歧，并给出一个比率来表示各种不同的分析作为实际分析的可能性。



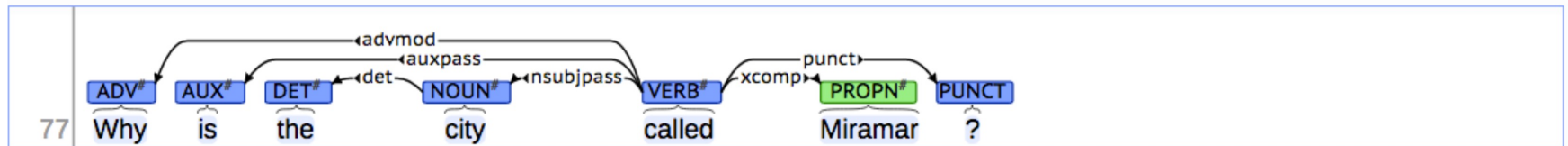
# 树库

[Universal Dependencies: <http://universaldependencies.org/> ;  
cf. Marcus et al. 1993, The Penn Treebank, Computational Linguistics]

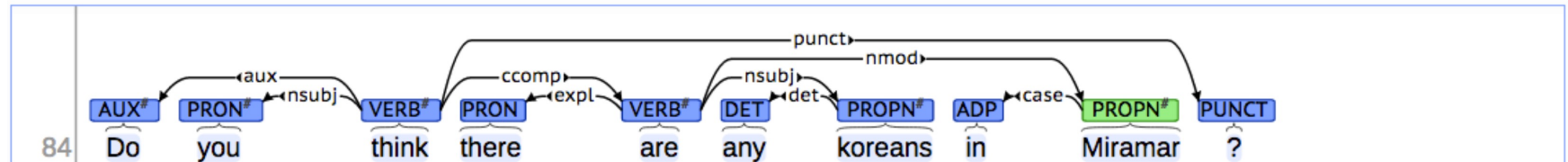
[context] [conllu]



[context] [conllu]



[context] [conllu]



# 树库

## 概念

- 纯粹的语法归纳方法不会产生人们所需的句法分析树，对于这个问题，一个显而易见的解决办法是给学习工具提供一些分析树的例子。这些分析树的一个集合称为树库（Treebank）。
- 这些经过正确句法分析的句子集合对于构建一个统计句法分析器是很有用的，因此很多研究单位都研制了树库，目前为止，从规模和可读性上来讲应用最为广泛的树库是Penn树库。





# 树库

## 例子 —— Penn treebank

表 12.1 Penn 树库中短语类别的缩写。左边的一列是常见的类别。这些分类还包括了一些少见的古怪类别

S	简单从句(句子)	CONJP	多词联合短语
SBAR	带补语的 S'从句	FRAG	片段
SBARQ	<i>wh</i> -question S'从句	INTJ	感叹词
SQ	倒装 <i>Yes/No</i> 问题 S'从句	LST	表符号
SINV	倒装陈述 S'从句	NAC	非组成成分聚集
ADJP	形容词短语	NX	名词短语中的名词性成分
ADVP	副词短语	PRN	插入语
NP	名词短语	PRT	小品词
PP	介词短语	RRC	约简关系从句
QP	量词短语(名词短语内部)	UCP	不协调短语
VP	动词短语	X	不知道或不确定
WHNP	<i>wh</i> -名词短语	WHADJP	<i>wh</i> -形容词短语
WHPP	<i>wh</i> -介词短语	WHADVP	<i>wh</i> -副词短语

# 树库

## 例子 —— Penn treebank

```
( (S (NP-SBJ The move)
    (VP followed
      (NP (NP a round)
        (PP of
          (NP (NP similar increases)
            (PP by
              (NP other lenders))
            (PP against
              (NP Arizona real estate loans))))))
    ,
    (S-ADV (NP-SBJ *)
      (VP reflecting
        (NP (NP a continuing decline)
          (PP-LOC in
            (NP that market))))))
  .))
```



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

# Q & A

