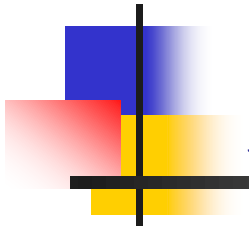


第二讲 语言知识的形式化表示

http://ccl.pku.edu.cn/doubtfire/

提纲

- 1 自然语言现象举例
- 2 关于自然语言的知识
- 3 知识的表示
  - 3.1 有限状态自动机（正则表达式）
  - 3.2 上下文无关文法
  - 3.3 特征结构与合一运算
- 4 小结

1 自然语言现象举例

例1

A. It is unlikely that Lee will be elected.

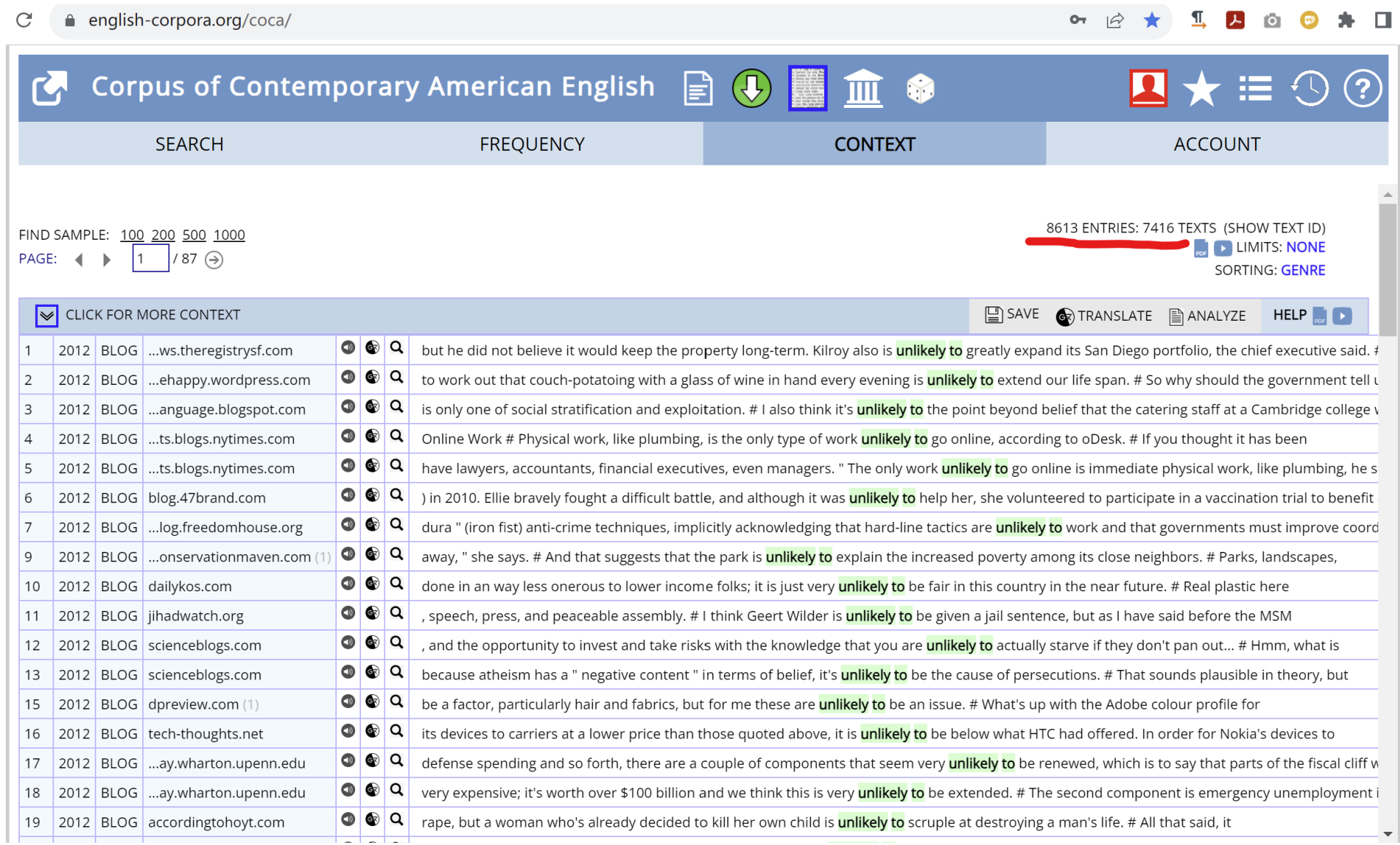
→ A'. Lee is unlikely to be elected.

B. It is improbable that Lee will be elected.

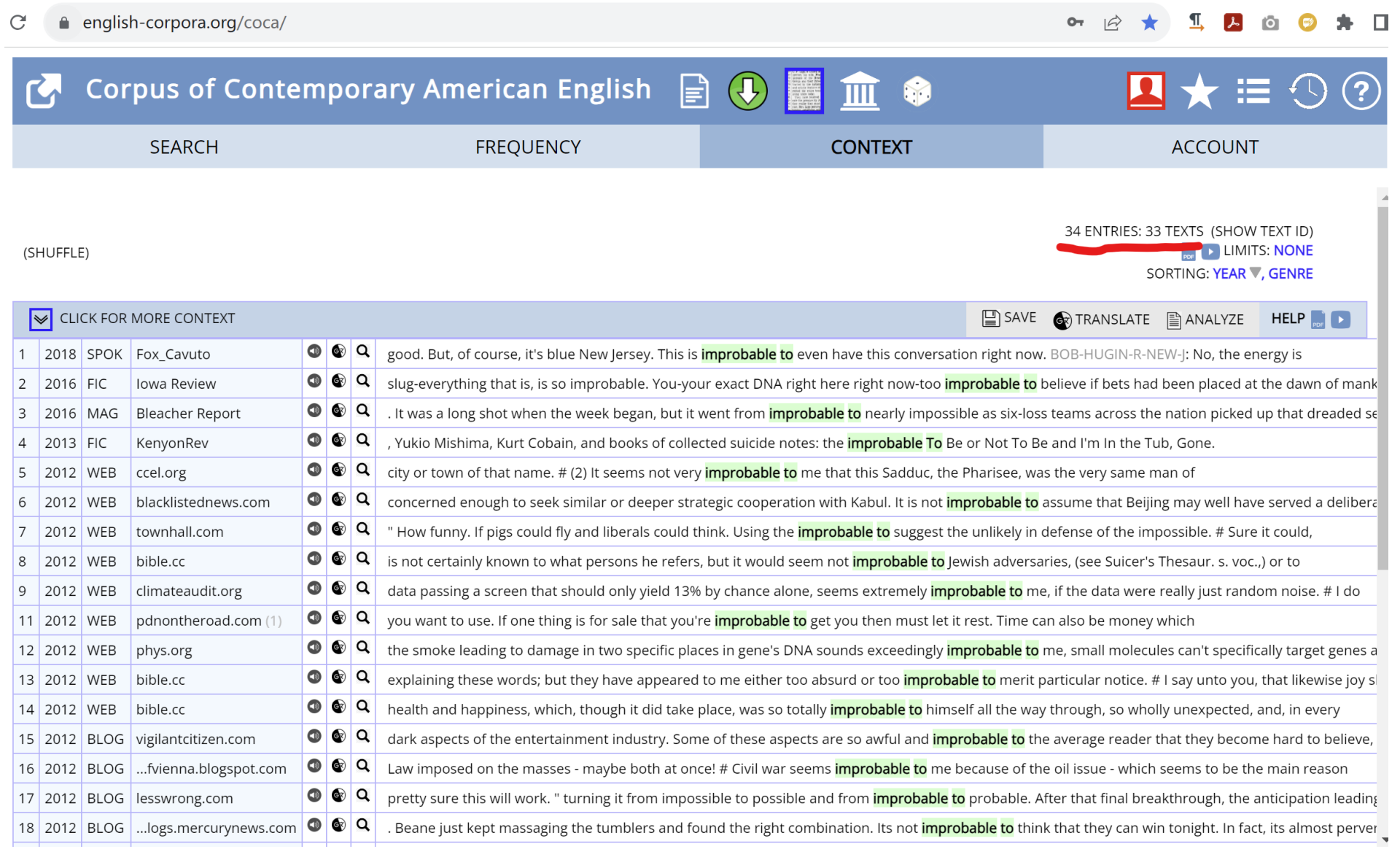
→ B'. * Lee is improbable to be elected.

Ivan A.Sag & Thomas Wasow, 1999, Syntactic Theory: A Formal Introduction, CSLI Publications

https://www.english-corpora.org/coca/



https://www.english-corpora.org/coca/



自然语言现象举例（续）

例2a

C. 张三爱好下围棋

→ C’. 下围棋是张三的爱好

D. 张三喜欢下围棋

→ D’. *下围棋是张三的喜欢

自然语言现象举例（续）

例2b

C. 读过 那本书 的 学生 不多

→ C’. 那本书 读过 的 学生 不多

D. 参观 故宫 的 学生 回来了

→ D’. * 故宫 参观 的 学生 回来了

自然语言现象举例（续）

例3

E. 这件事容易办 ⟶ E’. 办这件事容易

|| ||

F. 这件事好办 ⟶ F’. 办这件事好

自然语言现象举例（续）

例4

G. 马文才害死了梁山伯

H. 梁山伯被马文才害死了

I. G / *H ，欺骗了祝英台。

自然语言现象举例（续）

例5

J. 桌子上有两本书
K. 有两本书在桌子上

L. J ？K，一盏台灯，一个笔筒，还有一摞练习本

2 关于自然语言的知识

对于自然语言，人具有以下三个层面的能力：

- 人们一般可以判断一个表达形式是否属于一种语言，比如上面例1和例2中，人们能够判断出句子A'不属于英语（即不被说英语者接受），句子C'不属于汉语（即不被说汉语者接受）；
- 对于一种语言中的两个表达形式，人们一般可以判断二者之间是否具有某种关系，比如同义关系，两个表达式所对应的命题之间的逻辑蕴含关系，等等。像上面例3，人们能够判断句子E跟E‘是同义关系，但F跟F’不是同义关系。
- 对于一种语言中两个同义的表达式，人们一般可以判断在特定场合下使用哪一个表达式更好，比如上面例4和例5。

语言知识的分层

- 句法知识 ★ ★ ★ ★ ★
- 语义知识 ★ ★ ★
- 语篇知识 ★

语音知识（参见附录）

语言知识

（1） “ X 的 Y ”结构形式：dzjg

（2） X＝名词、形容词、动词；

Y＝名词

（3） “爱好” ＝> 动词 | 名词

“喜欢” ＝> 动词

“张三” ＝> 名词

语言知识（续）

（1）好$_1$，好$_2$，容易 ＝> 形容词 ?

（2）A + B : zzjg （A=形容词，B＝动词）

（3）B + A : zwjg

（4）zzjg(a,b) ⇔ zwjg(b,a) $a \in A; b \in B$

（5）“好$_1$”只能进入 zzjg （＝“easy”）

“好$_2$”只能进入 zwjg（＝“good”）

“容易” 可以进入zzjg, zwjg

自然语言现象举例（续）

例3

E. 这件事容易办 ⟶ E’. 办这件事容易

E″. 这件事**很容易**办 → 办这件事**很容易**

‖ ⫲

F. 这件事好办 ⟶̸ F’. 办这件事好

F″ 这件事**很好**办 —/→ 办这件事**很好**

通过最小对比“观察”语言现象

1 a 很好解决　　b 很容易解决

2 a 很难解决　　b * 很困难解决

3 a 不好解决　　b 不容易解决

4 a 不难解决　　b * 不困难解决

通过结构变换“观察”语言现象

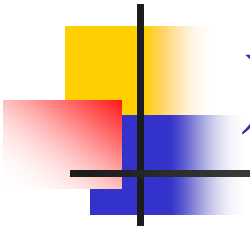
1 a 阿Q是有资格拿奖学金的　　b 拿奖学金阿Q是有资格的

2 a 阿Q是有时间谈恋爱的　　b 谈恋爱阿Q是有时间的

3 a 阿Q是有大人物撑腰的　　b * 撑腰阿Q是有大人物的

4 a 阿Q有钱　　b 阿Q很有钱

5 a 阿Q有碎银子　　b * 阿Q很有碎银子



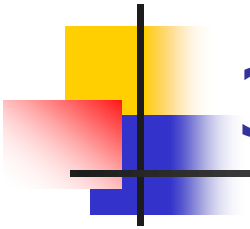
对语言知识的认识

代数学（理性主义）的定义方法

- 确定性定义方法
- 语言是由规则所定义的句子的集合

统计学（经验主义）的定义方法

- 不确定性定义方法
- 语言就是一个概率分布，又称为语言模型
- 语言中的每一个句子都有自己的出现概率

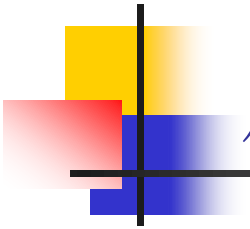


3 知识的表示

- 用自然语言来描述关于自然语言的知识
- 用形式语言来描述关于自然语言的知识

对象语言（Object Language）

元语言（Meta Language）



从自然语言到形式语言

- 避免混淆，“动词”不是动词
- 避免罗嗦，“从前有个山，山上有个庙……”
- 可计算，结构化的数据

形式语言（Formal Language）的一些例子

- $2+5 = 7$
- $2H_2 + O_2 = 2H_2O$
- P & Q （P: 董永是放牛郎；Q:董永喜欢七仙女）
- IS_COWBOY (x) & IS_Vega(y) & LOVE (x, y)

计算机语言

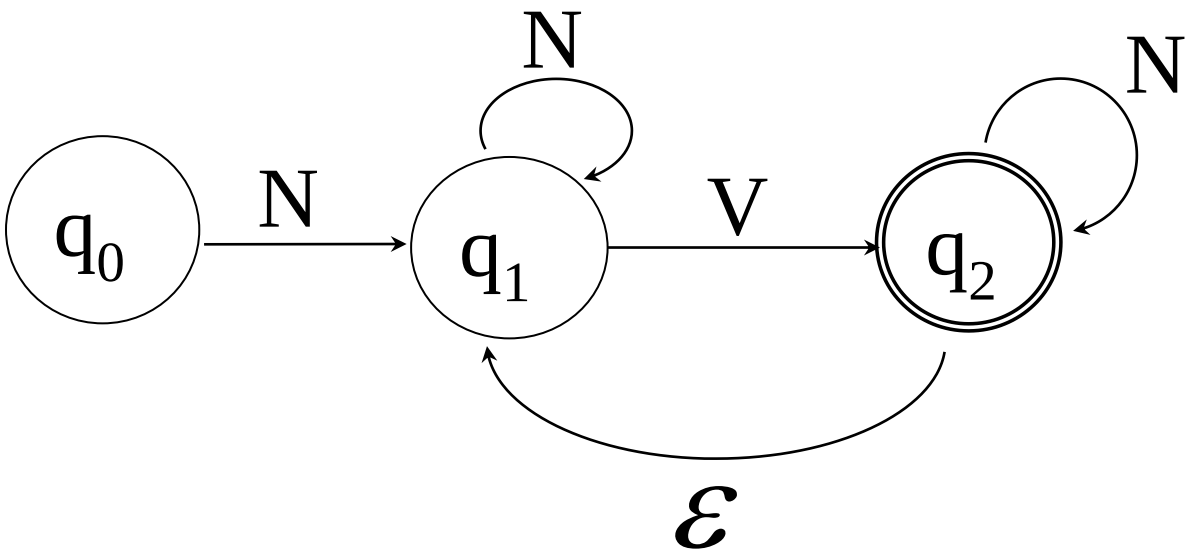
```
# include "stdio.h"
main ( )
{
   printf("\n\t hello, world");
   return 0;
}
```

一个C语言例子

如何描述（严格定义）一个语言

- 枚举 — 列举性定义
  - 给出语言中的所有句子
  - 对于含无限多个句子的语言不合适
- 文法 — 描述性定义
  - 给出生成语言中所有句子的方法
  - 当且仅当能够用该方法产生的句子才属于该语言
- 自动机
  - 给出识别该语言中句子的机械方法 — 过程性定义

3.1 有限状态自动机



有限状态自动机（Finite State Automata）

状态转移表（state transition table）

弧(输入) / 状态 转移	N	V	ε
q_0	q_1		
q_1	q_1	q_2	
q_2	q_2		q_1

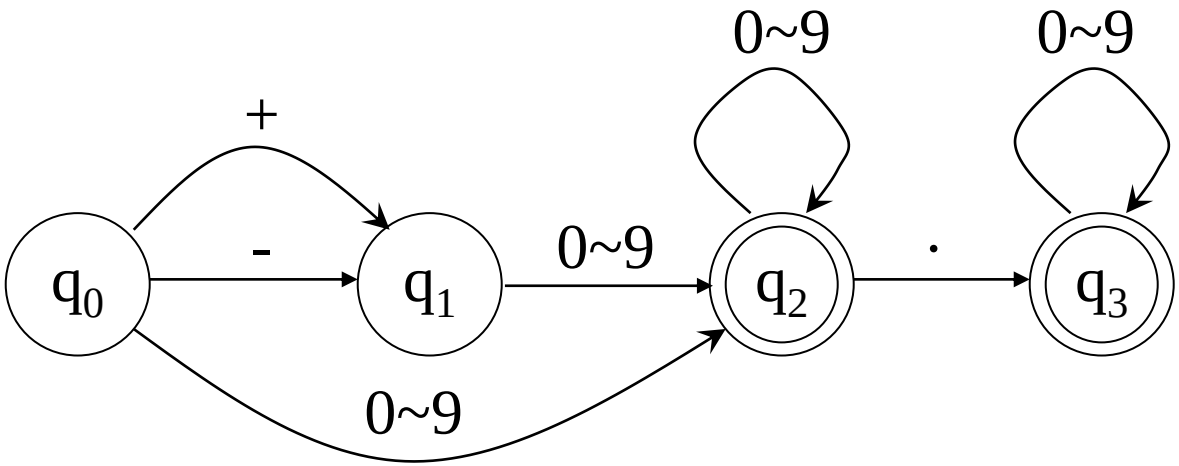
状态转移过程示例

字符串	状态转移过程
N V	$q_0 \rightarrow q_1 \rightarrow q_2$
N V N	$q_0 \rightarrow q_1 \rightarrow q_2 \rightarrow q_2$
N V N V N	$q_0 \rightarrow q_1 \rightarrow q_2 \rightarrow q_2 \rightarrow q_1 \rightarrow q_2 \rightarrow q_2$
……	……

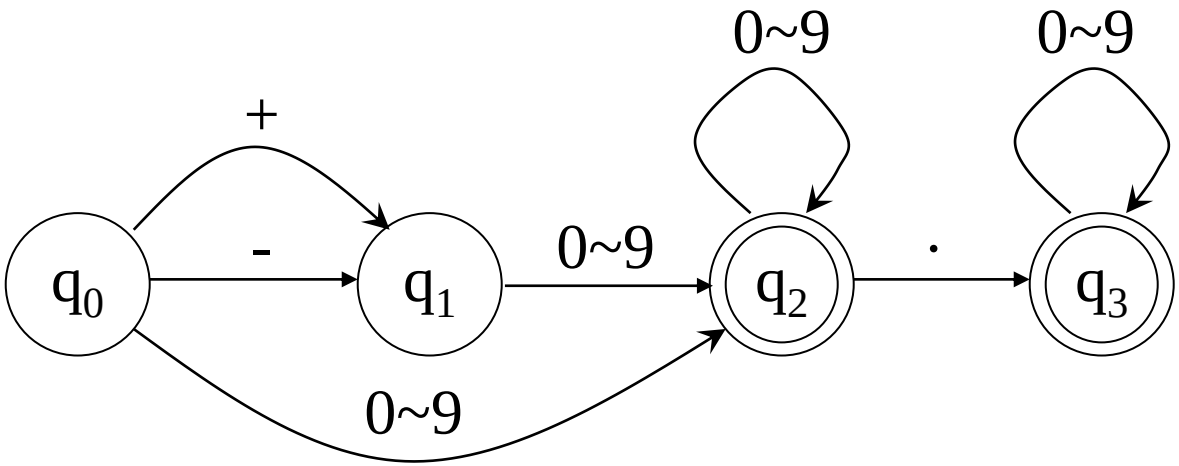
有限状态自动机(FSA)的形式定义

- 一个有限状态自动机M是一个五元组：$(Q, \Sigma, q_0, F, \delta)$
  - 有限个状态组成的状态集: Q
  - 有限字母组成的字母表: Σ
  - 开始状态 $q_0 \in Q$
  - 终止状态的集合 $F \subseteq Q$
  - 状态转移函数 $\delta(q,i)$: Q xΣ→ Q

识别一个十进制实数的自动机



正则表达式（Regular Expression）



(\+|-)? [0-9]+(\.[0-9]*)?

Jurafsky & Martin(2000), 第2章

Eliza 中的正则表达式操作

- 人： My boyfriend made me come here.
- Eliza: Your Boyfriend made you come here.
- 人： He says I'm depressed much of time.
- Eliza: I am sorry to hear you are depressed.

正则表达式替换：

1） /(*)my(*)me(*)/ => / \1your\2you\3 /

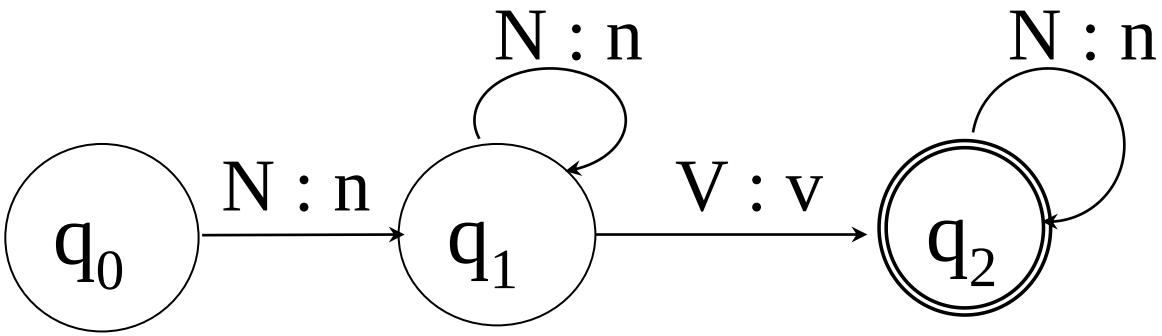
2） / (*)I'm (depressed) (*)/ => / I am sorry to hear you are \2\3./

可在Word中用“替换”功能实现替换

Eliza的操作就是做一连串的正则表达式的匹配和替换。Jurafsky & Martin(2000), 第2章，pp32-33.

弧上带输出的FSTN：Transducer

有限状态转录机



董永喜欢七仙女 — Dong_Yong loves Qi_Xiannv

董永七仙女喜欢 — Dong_Yong Qi_Xiannv loves

课堂练习

- 构造一个有限状态转移网络，可以接受汉语的重叠形式AABB，ABB，ABAB，…

课堂练习

- 构造识别名词词组的FSA

三本书

语法书

阿Q的书

阿Q的三本书

阿Q的三本语法书

阿Q的三本汉语语法书

阿Q的三本古代汉语语法书

阿Q和他的三本汉语语法书

……

从FSA到上下文无关文法（CFG）

- FSA：无法描述自然语言的层次结构特性

听说服装设计很吃香 —— 听说那套服装设计得很有品位

听说孩子丢了 —— 听说孩子丢了一只鞋

听说北京队大败 —— 听说北京队大败上海队

3.2 上下文无关文法

- **符号**
- **字母表：**有限个任意符号组成的非空集合Σ
  - 例1：所有汉字组成的集合构成一个字母表。
  - 例2：汉语中所有的词也构成一个字母表。
  - 例3：字母*a*, *b*, *c*也组成一个字母表。
- **字符串：**由字母表Σ上的字符组成的长度有限的序列
  - 若字母表$\Sigma=\{a, b\}$，则*a*, *b*, *ab*, *aba*, *aabb* 等等都是字母表上的字符串。

语言的形式定义

语言：是字母表上的字符串的任意集合。

例1. 若$\Sigma = \{a, b\}$，则定义在Σ上的语言可以是

$L1=\{ab, ba\}$

$L2=\{ab, abab, ababab, \ldots\}$

形式文法

形式文法：一个形式文法G由四个部分组成，可记作

$G=\{V_N, V_T, S, P\}$，其中：

V_N：称为文法G的非终结符号字母表，V_N不出现在G所表示的语言集合的句子中；

V_T：称为文法G的终结符号字母表，G所表示的语言的句子由V_T中的元素组成，

$V_N \cap V_T = \phi$；

S：代表句子符号，$S \in V_N$。

P：代表一组式子组成的集合，P中的式子具有如下形式：

$$\alpha \rightarrow \beta$$

形式文法（续）

产生式规则（production rule）

重写规则（rewriting rule）

$$\alpha \to \beta$$

产生式需要满足下面的条件：

1）α可以是V_N和V_T上的任意字符串，不能是空字符；

2）β可以是V_N和V_T上的任意字符串，可以是空字符；

3）P中至少有一个产生式中的α得由S来充当；

上下文无关文法

- 对产生式规则 $\alpha \to \beta$ 做如下约定：

$$|\alpha| = 1 \quad \alpha \in V_N \quad \beta \in (V_N \cup V_T)^*$$

这样的形式文法就是“上下文无关文法”。

一个上下文无关文法的例子

设文法$G_0 = (V_N, V_T, S, P)$，其中

$V_N = \{S, NP, VP, N, V\}$,

$V_T = \{$喜欢, 知道, 董永, 七仙女$\}$,

P 中产生式如下：

1. $S \rightarrow NP\ VP$
2. $VP \rightarrow VP\ NP$
3. $VP \rightarrow VP\ S$
4. $VP \rightarrow V$
5. $NP \rightarrow N$
6. $N \rightarrow$ 董永
7. $N \rightarrow$ 七仙女
8. $V \rightarrow$ 喜欢
9. $V \rightarrow$ 知道

直接推导、推导、句型、句子、语言

直接推导： S => NP VP

推导：S=> NP VP => NP V => N V

上式可以简写为：S $\overset{*}{\Longrightarrow}$ N V

句型：NP VP，NP V，N V，… 是G_0的句型

句子：仅含终结符号的句型，N V

语言：给定一个文法G_0，该文法所产生的所有句子组成的集合，称为***该文法所定义的语言***

G_0所描述的语言L_0

S1: 董永喜欢七仙女

S2: 董永知道董永喜欢七仙女

S3: 七仙女知道董永

S4: 七仙女喜欢董永知道董永

S5: 七仙女喜欢董永董永董永七仙女

……

不属于L_0的字符串

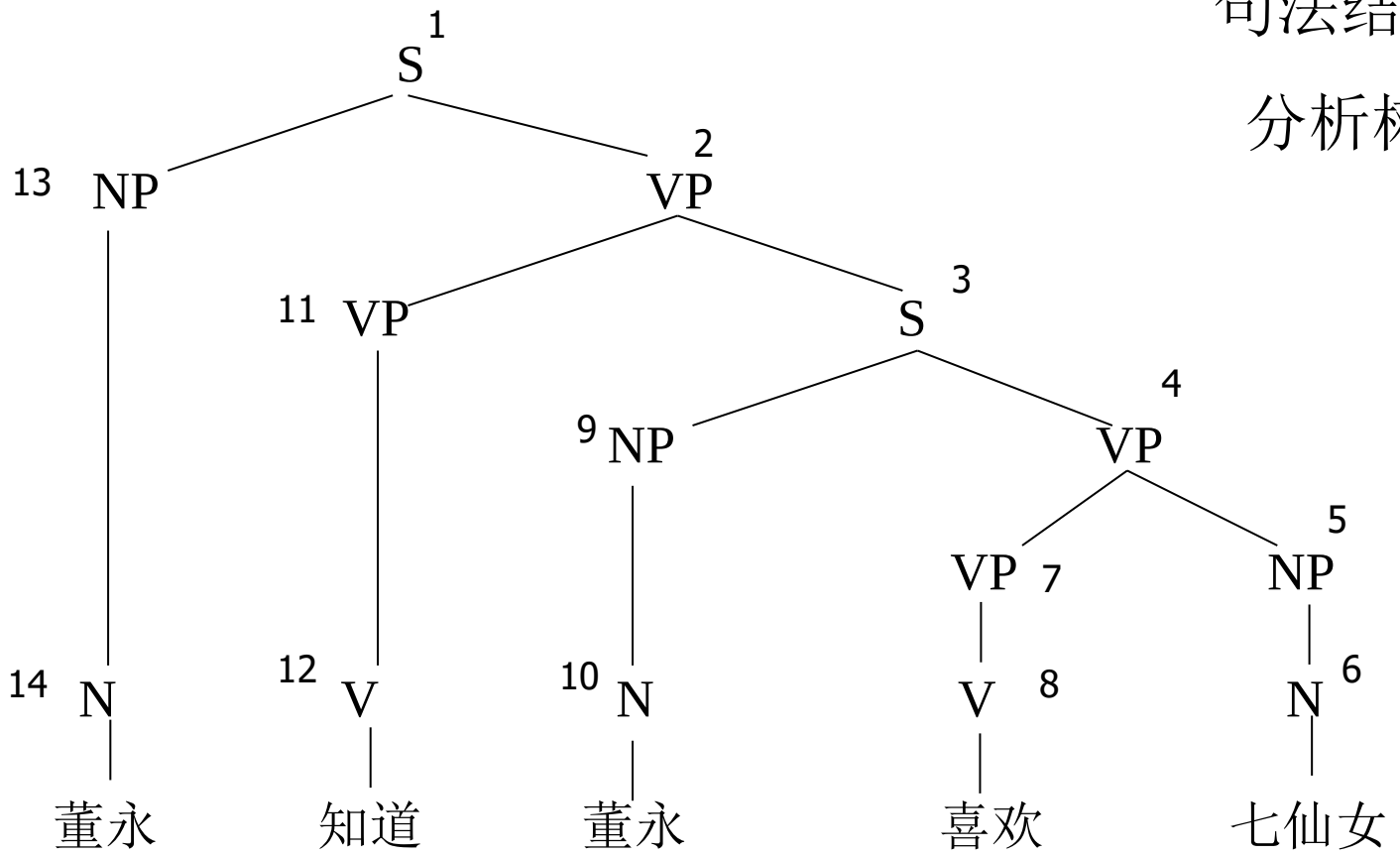
S1': 知道喜欢知道七仙女

S2': 董永董永七仙女知道喜欢

S3': 七仙女董永喜欢

……

句子结构的树形描述


句法结构
分析树
1
S
13
NP
2
VP
11
VP
3
S
9
NP
4
VP
VP
7
5
NP
14
N
12
V
10
N
V
8
N
6
董永
知道
董永
喜欢
七仙女

文法的三个作用

- 生成：产生语言L中所有的句子；
- 判定：一个字符串（String）是否属于语言L；
- 分析：得到L中句子的结构树；

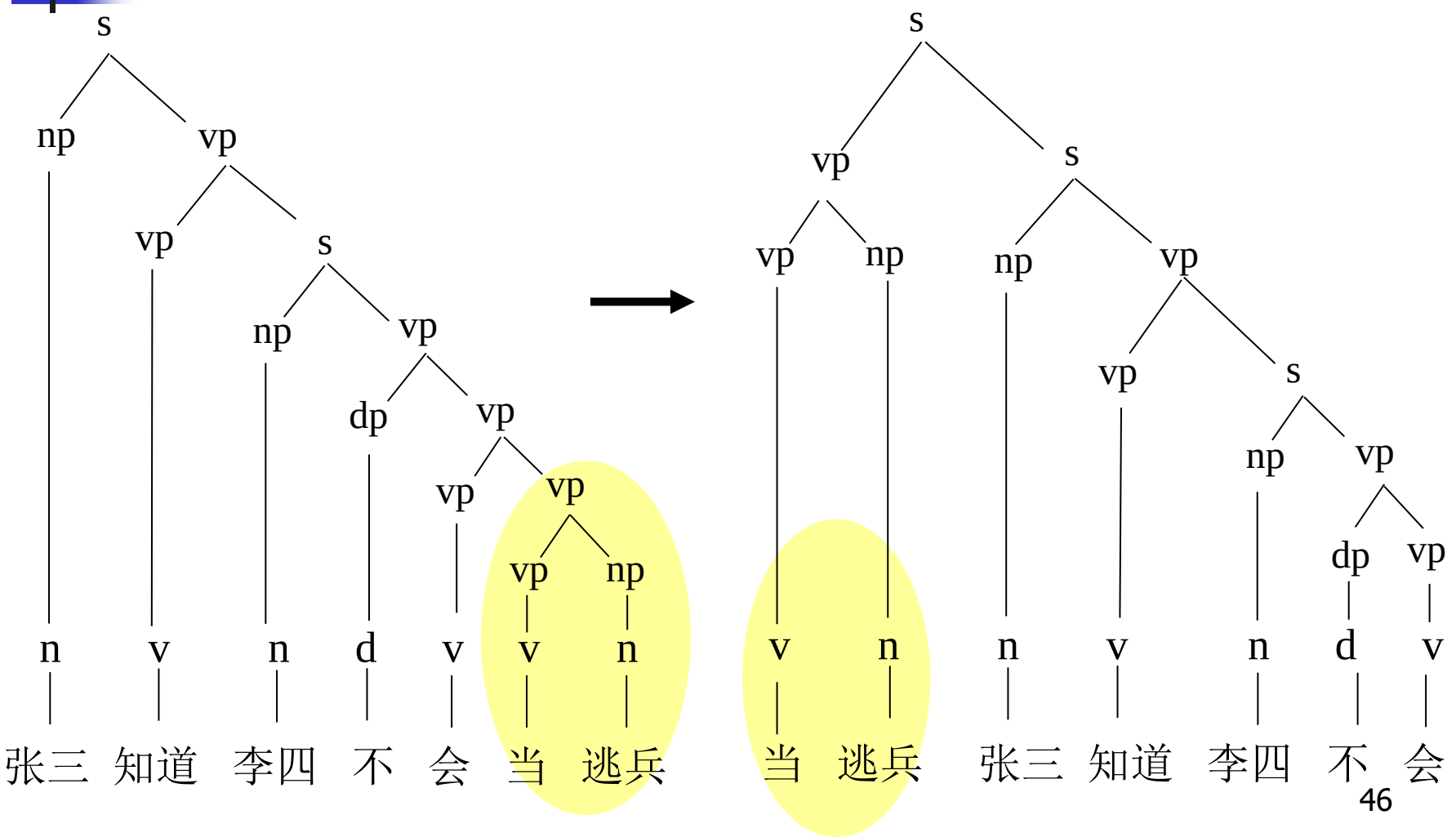
分析出句子的结构是进行自然语言信息处理的基础

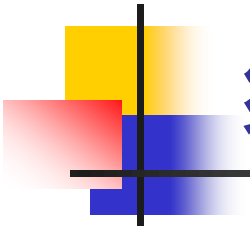
比如移位变换，就必须建立在结构分析的基础上

张三知道李四不会当逃兵 -> 当逃兵，张三知道李四不会

-> * 知道李四，张三不会当逃兵

句法结构分析的效用：控制转换





练习

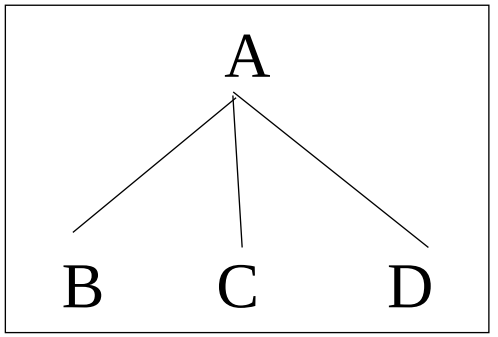
对于语言L={ab, aabb, aaabbb, …, a^nb^n, …}
n是自然数。

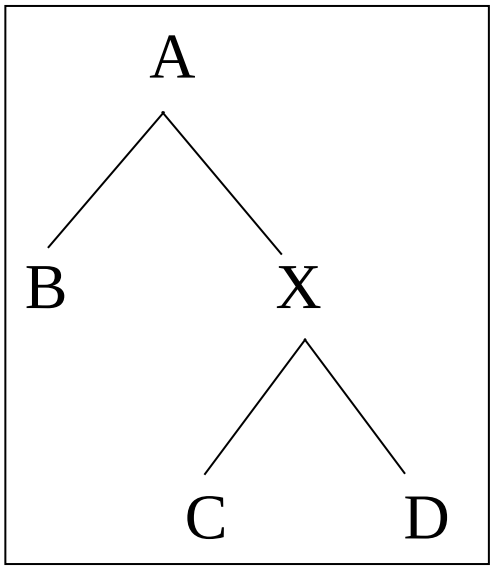
（1）请写出L的上下文无关文法；

（2）要求产生式右部不能超过两个符号；

乔姆斯基范式（Chomsky Normal Form）

- A → B C
- A → a





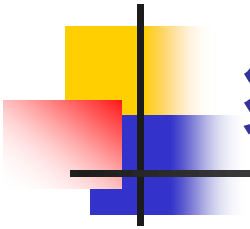
引入一个新的非终结符X

1 A → B C D

一个三分支规则可以转换为两个二分支规则

1 A → B X

2 X → C D



练习

1 写出汉语表示自然数的词的CFG

2 用你写的CFG，画出下列数字的分析树：

一亿零三百万　　三万六千五百八十一

基于简单范畴的文法的缺陷

- 范畴划分有不同的颗粒度（granularity）

  例如英语句子的构成规则：

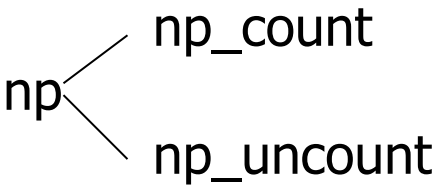
  S → NP VP

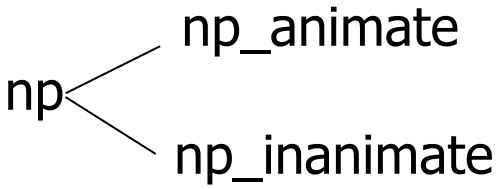
  如果考虑到英语主谓语单复数的搭配，就要将NP和VP分成NPsingular和NPplural和VPsingular和VPplural，并其将规则改写成：

  S → NPsingular VPsingular

  S → NPplural VPplural

- 范畴划分有不同的角度（perspective）





3.3 特征结构与合一运算

- 引入特征结构弥补简单范畴的不足

  - 特征结构（Feature Structure）
    复杂特征集（Complex Feature Set）
  - 特征结构定义为“特征”的集合
  - 所谓“特征”，是一个由“属性”和“值”组成的二元组，“属性”也称为“特征名”，“值”也称为“特征值”
  - 在特征结构中，要求所有的“特征”的“属性”互不相同
  - 空特征结构：不含任何特征的特征结构 记作：[ ]

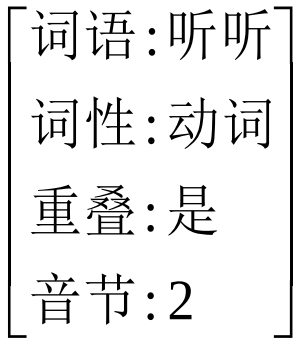
$$\begin{pmatrix} attribute_1=value_1 \\ attribute_2=value_2 \\ \cdots \\ \cdots \\ \cdots \\ attribute_n=value_n \end{pmatrix}$$

特征结构的嵌套与共享

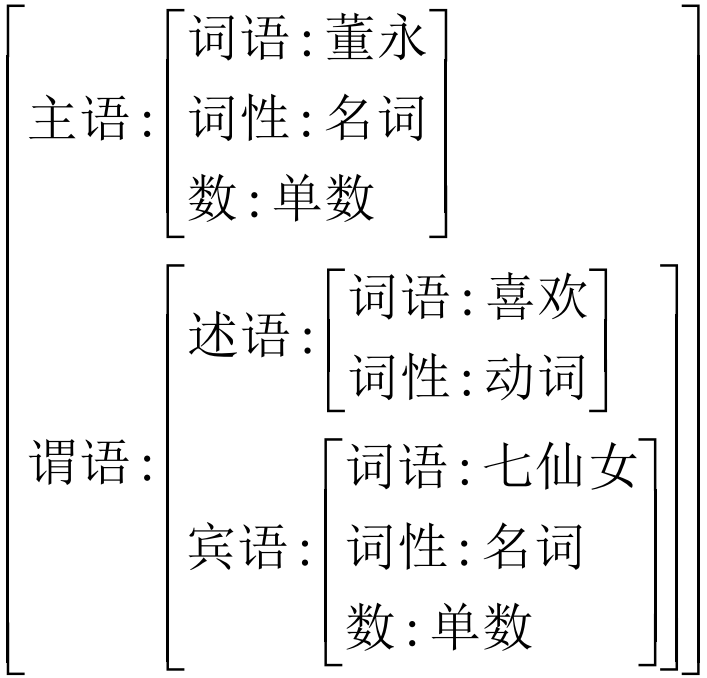
1）“特征值”可以是一个字符串值或数值等简单类型，也可以是另一个特征结构，这就是所谓的特征结构的“嵌套”；
为了区别于特征结构形式的特征值，我们把简单的字符串形式的特征值称为原子（atom）

2）两个特征可以共享一个值，这是所谓的特征值的“共享”（也称为“重入”/ Reentrance）。

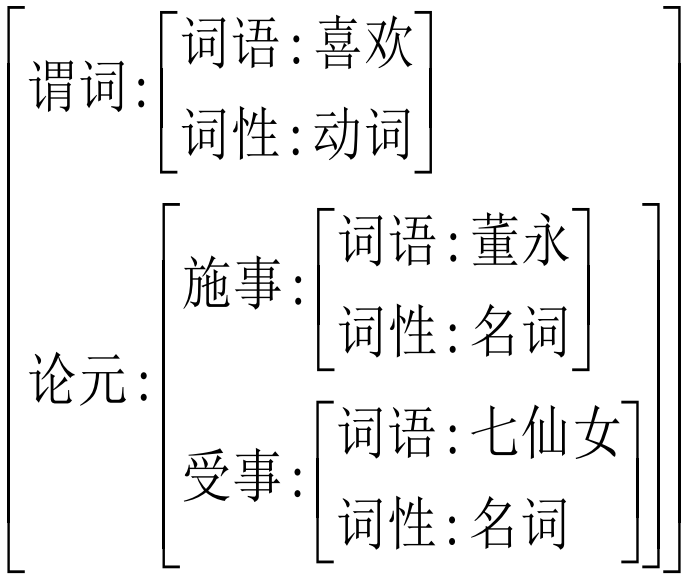
特征结构示例（框式表示法）



a. 简单特征结构



b. 复杂特征结构（嵌套）



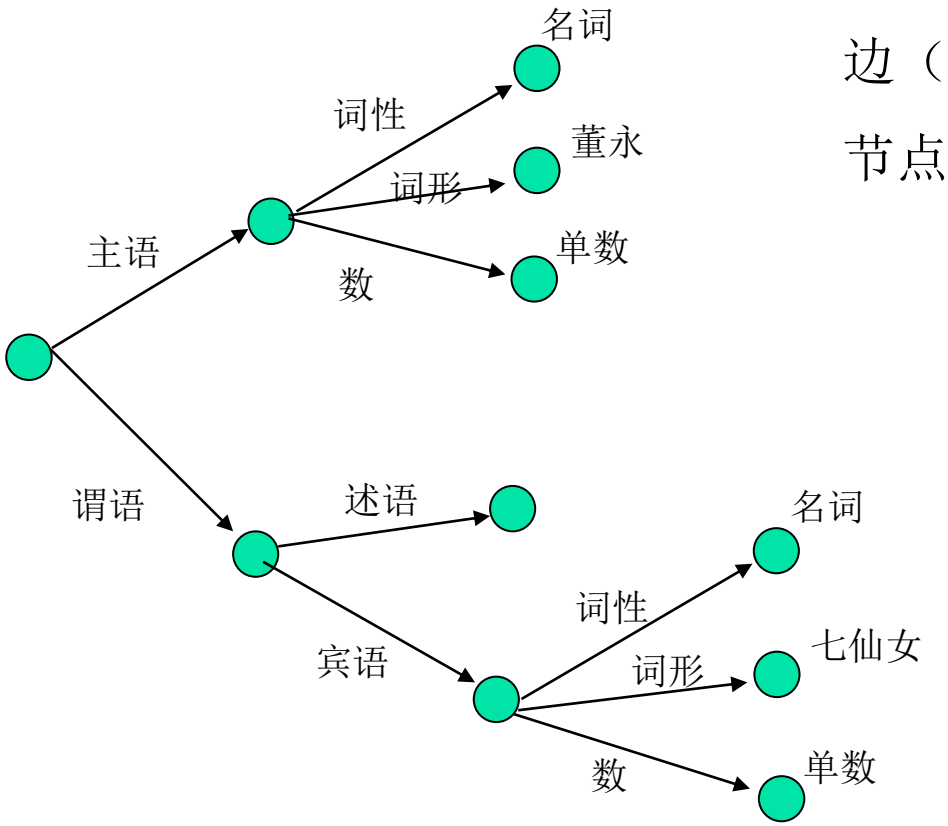
c. 复杂特征结构

特征结构的表表示法

((主语: (词语:董永)(词性:名词)(数:单数))
(谓语: (述语:(词语:喜欢)(词性:动词))
(宾语:(词语:七仙女)(词性:名词)(数:单数))))

有向无环图（**D**irected **A**cyclic **G**raph）

特征结构的图表示法



边（edge）表示特征

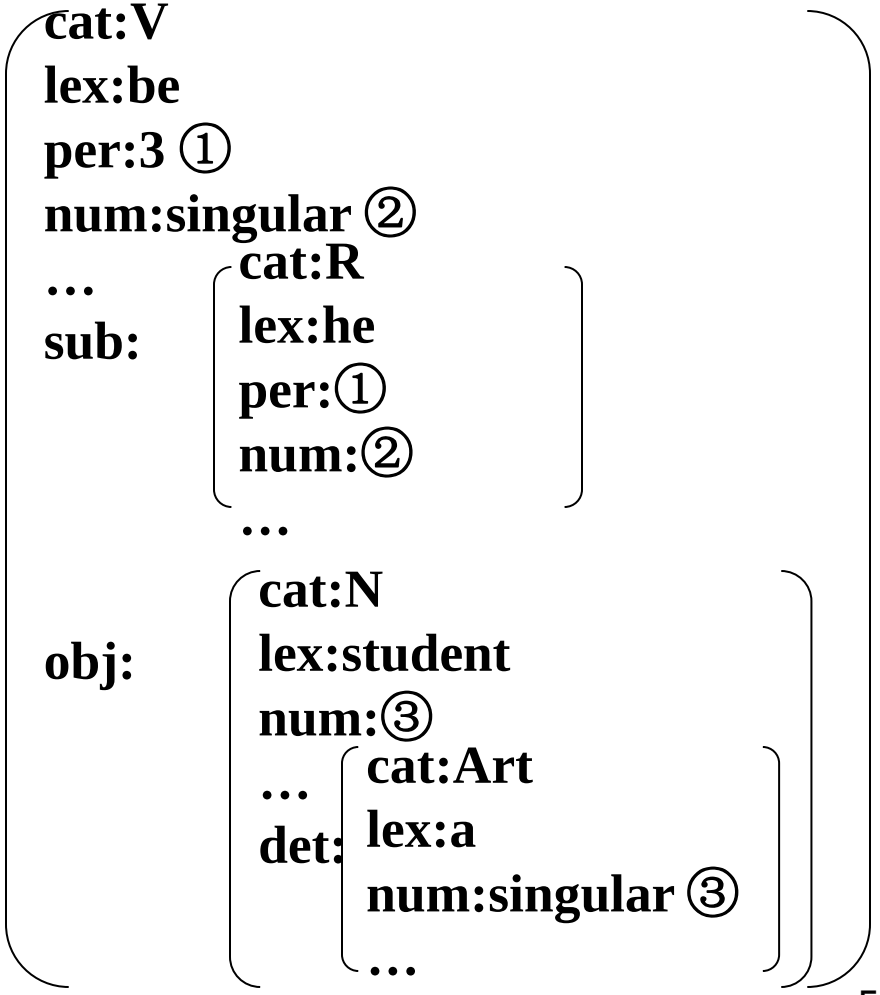
节点（node）表示特征值

两个特征结构的值共享

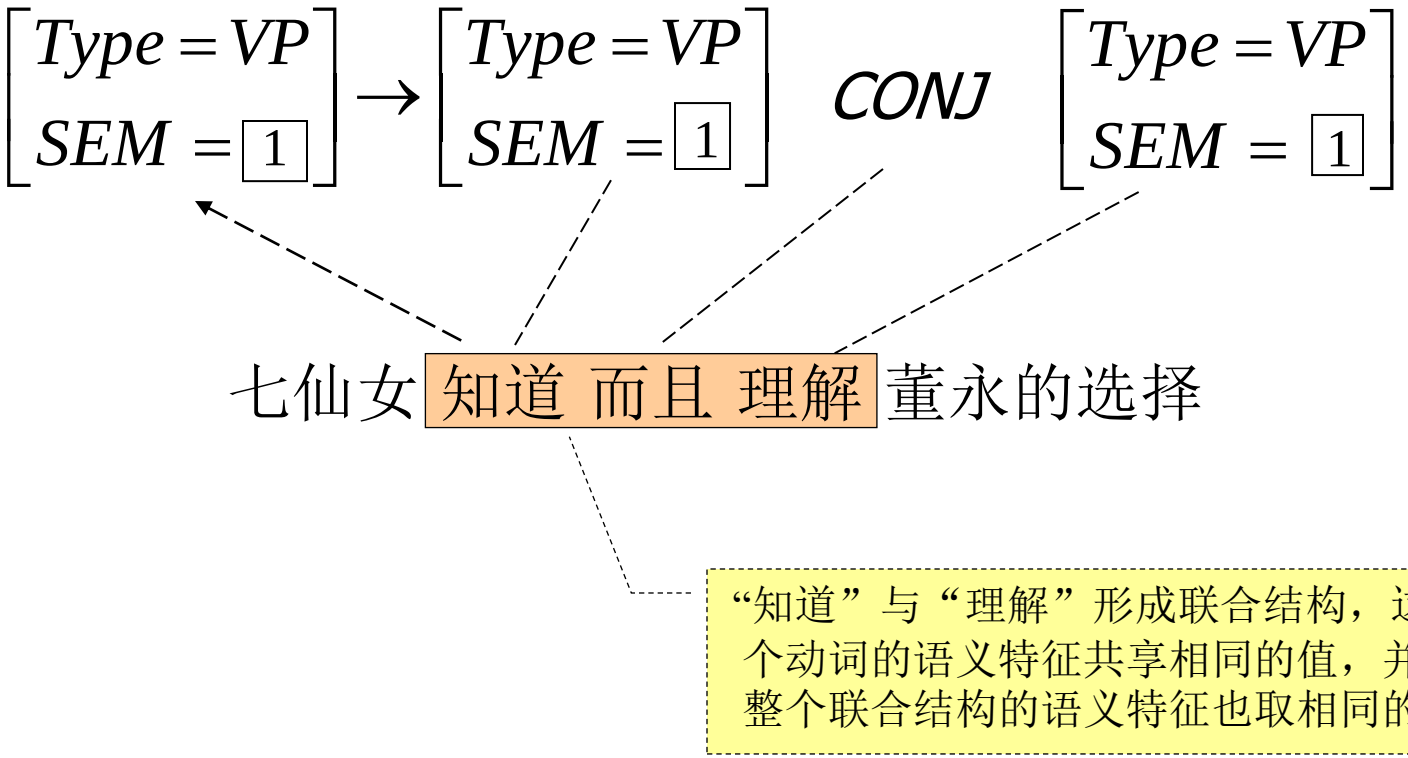
- 例子：

  He is a student.

- 在特征结构表示中，一般用数字表示重入的特征结构
- 在重入的多个特征结构中，只需在一处说明其特征值



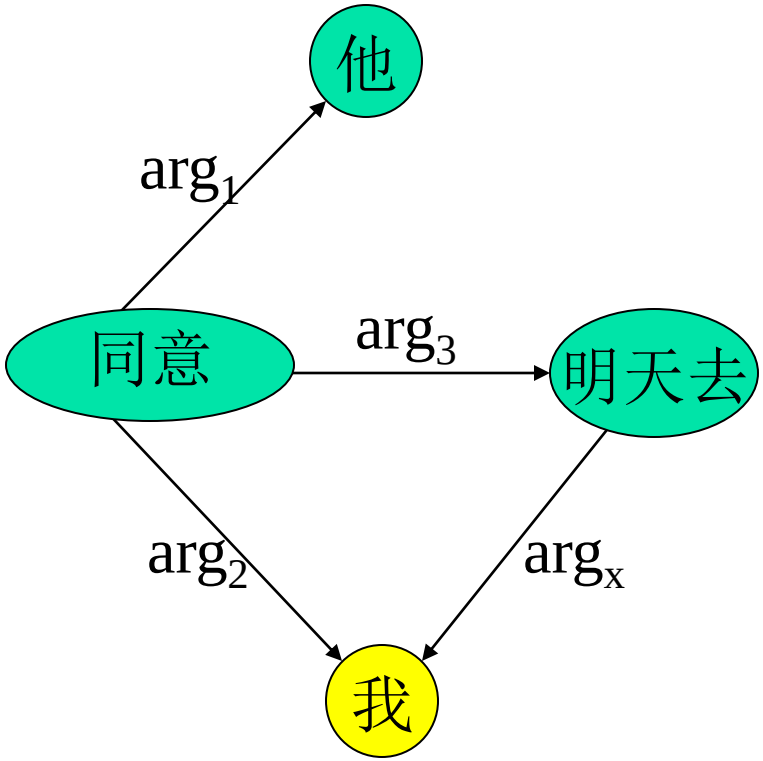
两个特征结构的值共享（续）

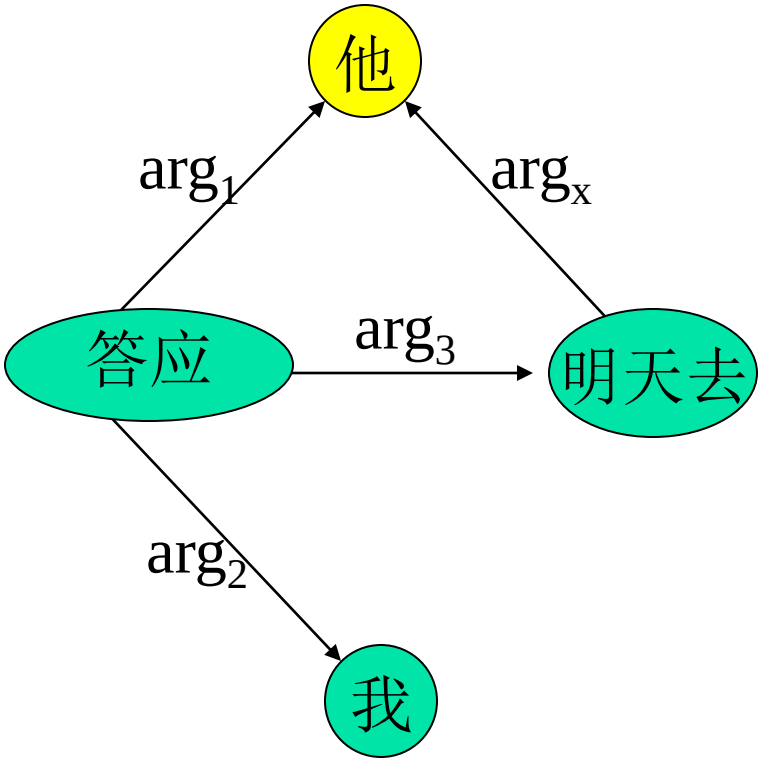


运用特征结构表征词语差异

- 他**同意**去 vs. 他**答应**去
- 他**同意**明天去 vs. 他**答应**明天去
- 他**同意**我明天去 vs. 他**答应**我明天去

特征结构的共享（有向图表示）

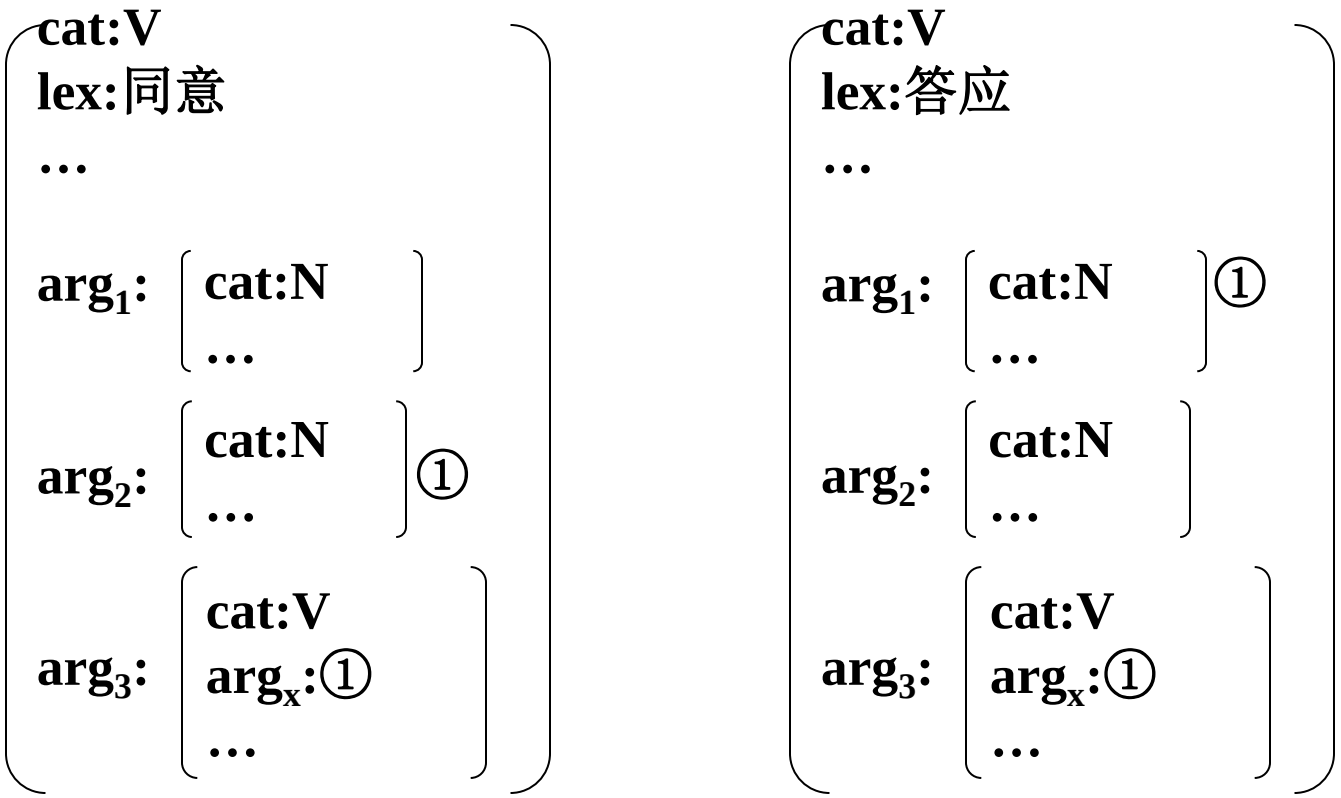




他同意我明天去　　vs.　　他答应我明天去

两个特征结构的值共享（续）

- “同意”和“答应”的区别



特征结构间的包孕关系subsumption

- 特征结构F1包孕F2，记作$F1 \dot{\subseteq} F2$，当且仅当
  （1）若特征 $f \in F1$，则 $f \in F2$，并且
  （2）若f的值是特征结构，则 $value_{F1}(f) \dot{\subseteq} value_{F2}(f)$
  （3）若f的值是简单原子，则 $value_{F1}(f) = value_{F2}(f)$
- 空特征结构包孕任何特征结构

特征结构包孕关系举例

$$[\textit{Number}\quad \text{SG}] \dot{\subseteq} \begin{bmatrix} \textit{Number} & \text{SG} \\ \text{PERSON} & 3 \end{bmatrix}$$

$$[\textit{Agree}\ [\textit{Number}\quad \text{SG}]] \dot{\subseteq} \begin{bmatrix} \textit{CAT} & \text{NP} \\ \text{Agree} & \begin{bmatrix} \textit{Number} & \text{SG} \\ \text{PERSON} & 3 \end{bmatrix} \end{bmatrix}$$

$$[\ ] \dot{\subseteq} \begin{bmatrix} \textit{Number} & \text{SG} \\ \text{PERSON} & 3 \end{bmatrix}$$

$$[\textit{Number}\quad \text{SG}] \dot{\not\subseteq} [\textit{Number}\quad \text{PL}]$$

特征结构的合一运算

- 合一运算（Unification）：将两个独立的特征结构F1，F2组合为一个新的特征结构F3，满足条件：$F1 \dot{\subseteq} F3$ 并且 $F2 \dot{\subseteq} F3$
- 合一的含义是：对两个特征结构进行类似于集合求**并**的一种运算，从而可以在“小”的特征结构基础上形成“大”的特征结构，这种运算非常适于刻划“小”的语言单位发展为“大”的语言单位的过程中的信息增加，即F3中包含了F1，F2所包含的信息（合一运算的单调性monotonic）

合一成功、失败、合一结果为空

- 合一操作有成功或失败两种可能：
  - 合一成功，则原来的两个独立的特征结构成为同一个特征结构；
  - 合一失败，维持原状。

- 注意：合一失败和合一结果为空是不同的
  - 合一失败，两个特征结构之间不发生共享；
  - 合一结果为空，表示合一成功，两个特征结构共享，变成同一个特征结构，只是这个特征结构是空特征结构；
  - 只有两个空特征结构合一，结果才是空。

合一实例（一）

$$A = \begin{bmatrix} 结构:述宾 \\ 功能:述语 \\ 词性:动词 \\ 及物:是 \end{bmatrix} \qquad B = \begin{bmatrix} 词语:咳嗽 \\ 词性:动词 \\ 及物:否 \end{bmatrix}$$

$$A \cup B = \phi$$

合一失败

合一实例（二）

令 A = $\begin{bmatrix} \text{施事}:C \\ \text{谓词}:\text{知道} \end{bmatrix}$ B= $\begin{bmatrix} \text{词语}:\text{董永} \\ \text{语义类}:\text{人} \end{bmatrix}$，其中 C = [语义类:人]

则将 C 和 B 合一后，特征结构A变为：

$$\begin{bmatrix} \text{施事}:\begin{bmatrix} \text{语义类}:\text{人} \\ \text{词语}:\text{董永} \end{bmatrix} \\ \text{谓词}:\text{知道} \end{bmatrix}$$

合一成功

合一实例（三）

$$E = \begin{bmatrix} \text{Agree:} & [\text{Number: Singular}]\ ① \\ \text{Subject:} & [\text{Agree:}\ ①] \end{bmatrix}$$

$$F = \begin{bmatrix} \text{Subject:} & [\text{Agree:} & [\text{Person: 3}]] \end{bmatrix}$$

$$E \bar{\cup} F = \begin{bmatrix} \text{Agree:} & [\text{Number: Singular}\\ \text{Person: 3}]\ ① \\ \text{Subject:} & [\text{Agree:}\ ①] \end{bmatrix}$$

合一成功

合一运算的性质

- 交换律：$A \bar{\cup} B = B \bar{\cup} A$
- 结合律：$A \bar{\cup} (B \bar{\cup} C) = (A \bar{\cup} B) \bar{\cup} C$
  - 合一运算的执行顺序与结果无关（order independent）
  - 合一运算的结合律使得特征结构真正成为一种“描述性”知识表示方法，而不是“过程性”的表示方法
  - “描述性”知识表示方法的含义在于，对于一个变量的约束和赋值是等同的，我们可以在对一个变量赋值之前就给出对它的约束，而不必等到对这个变量赋值之后才对它进行约束
  - 比如，我们可以在词典中指出，汉语动词“同意”的arg_3的arg_1必须和“同意”的arg_2合一，虽然这时我们并不知道在具体的句子中 “同意”的各个arg是什么
  - 特征结构的“描述性”特点有利于在词典中给出词语的个性化描述

合一运算的两个基本作用

（1）检查两个特征结构所包含的信息是否相容，这可以作为语言成分组合时的测试手段（比如前面例一）

（2）当两个相容的特征结构组合成更大的特征结构时，信息增加（比如上页例二、三）

合一对于语言知识表示和处理的意义

◆ 由于句法和语义分析都可以用“合一”来作为基本运算，不仅句子的合法性可以通过语义手段来判断，而且，还可以把句子的句法结构和语义表示用合一运算这种方式更加自然的衔接起来。

◆ 对不同的复杂特征集进行合一运算，其结果同运算所进行的先后次序无关，不论合一从那个方向开始，也不论是先合一还是后合一，合一的结果都是相同的。合一运算的这种无序性非常便于并行处理，而且还使我们有可能自由地选择分析算法和自然语言描述的语法理论。

小结

有限状态自动机

上下文无关文法

特征结构（合一运算）

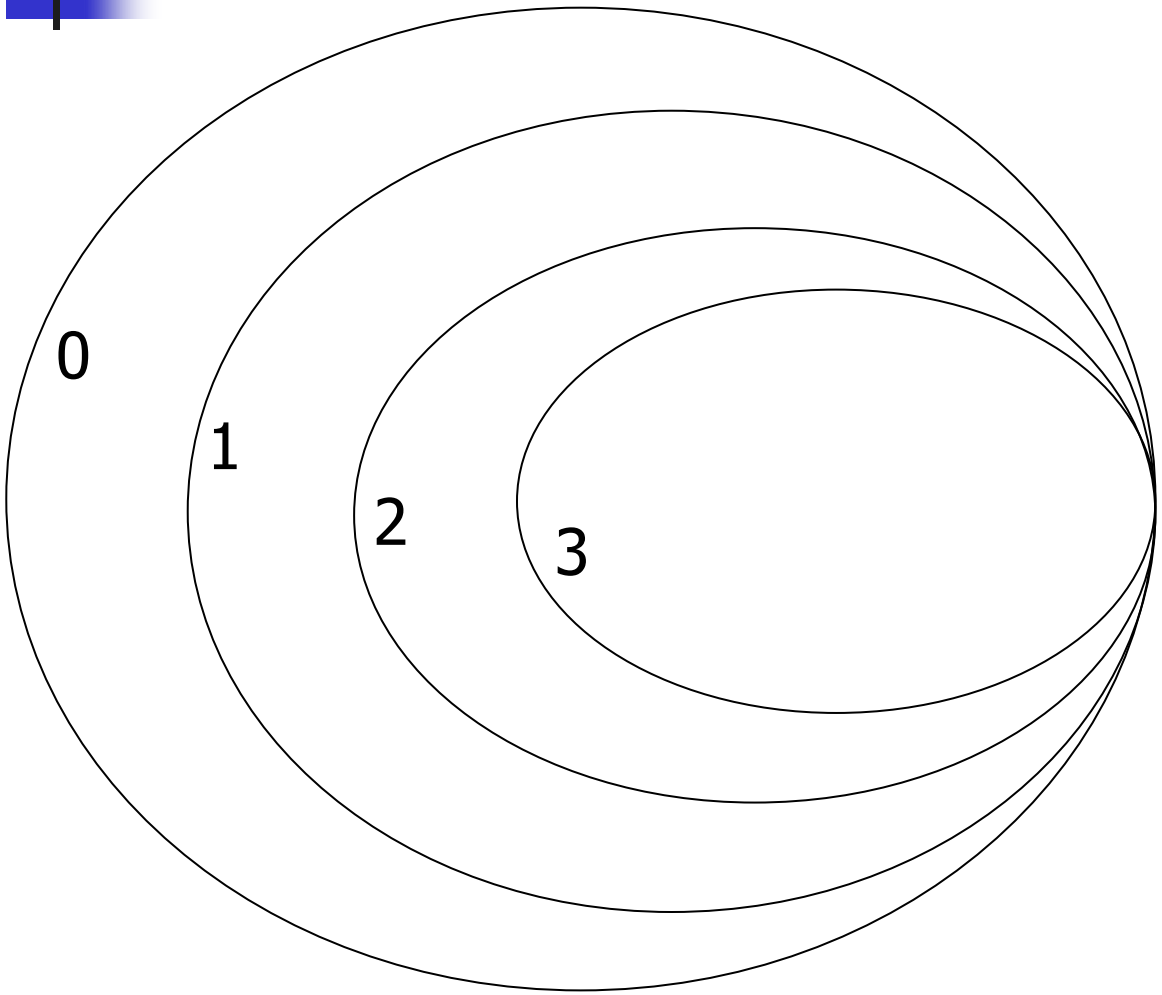
线性序列（Linear Structure）
树结构（Tree Structure）
图结构（Graph Structure）

- ✓ 形式表示方法仅仅是工具，它本身并不增加知识，只是让知识以严密、清晰的方式呈现出来。
- ✓ 以形式化的方式来表述自然语言知识一方面便于计算，另一方面也有助于发现语言学问题。

进一步阅读文献

- 冯志伟等译（2005）《自然语言处理综论》第1章，第10.3.2，第13章。
- 杜淑敏 等（1990）《编译程序设计原理》，北京大学出版社
- 何成武（1990）《自动机理论及其应用》，科学出版社
- 王兵山、吴兵编（1988）《形式语言》，国防科技大学出版社
- Noam Chomsky (1959) On certain formal properties of grammars, Information and Control, 1959, Vol.2, pages 137-167
- 陆汝钤（2000）《人工智能》（上册），科学出版社
- 陆汝钤（1993）《数学·计算·逻辑》，湖南教育出版社
- 沙新时 等（1993）《基于合一语法的通用句法分析器：设计与实施》，载《中文信息学报》1993年第2期。

附录1 Chomsky Hierarchy



G_0：无限制重写文法

G_1：上下文相关文法

G_2：上下文无关文法

G_3：正则文法

L_0：递归可枚举语言

L_1：上下文相关语言

L_2：上下文无关语言

L_3：正则语言

PSG, CSG, CFG, RG

- PSG: 无限制
- CSG: $|\alpha| \le |\beta|$
- CFG: $|\alpha| = 1 \quad \alpha \in V_N \quad \beta \in (V_N \cup V_T)^*$
- RG: $|\alpha| = 1 \quad \alpha \in V_N \quad \beta = tN \quad t \in V_T, N \in V_N$

$$\text{or} \qquad \beta = Nt$$

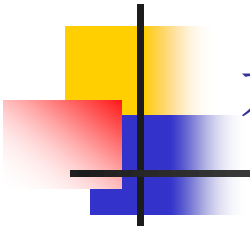
附录2 文法、自动机和语言

	语法	自动机	语言	复杂度
0型	无限制短语结构文法	图灵机	递归可枚举语言	半可判定
1型	上下文有关文法	线性有界自动机	上下文有关语言	NP完全
2型	上下文无关文法	下推自动机	上下文无关语言	多项式
3型	正则文法	有限自动机	正则语言	线性

各型文法的判定难度

- PSG：半可判定
  对于一个属于G_{type0}的句子L，总可以在确定步内判断出“是”；但对于一个不属于G_{type0}的句子L’，不存在一个算法，可以在确定步内判断出“否”。
- CSG：可判定，复杂度：NP完全
- CFG：可判定，复杂度：多项式
- RG：可判定，复杂度：线性

参阅：陆汝钤（1993）《数学·计算·逻辑》，湖南教育出版社。第六、八章。



有关形式文法的问题

文法的二义性（Ambiguity）

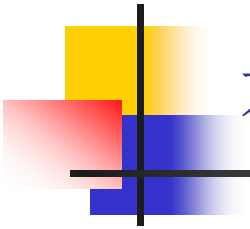
各型文法的描述能力

不同文法之间的等价性（强等价、弱等价）

文法的机器学习问题

对于CFG，不存在一个确定的算法，可以在给定句子集合L基础上，学到L的CFG文法（这个数学结果是对Chomsky关于儿童语言能力先天说的一个支持）

……



有关自动机的问题

自动机的分类

各类自动机的形式定义

各类自动机与各型形式文法的对应关系

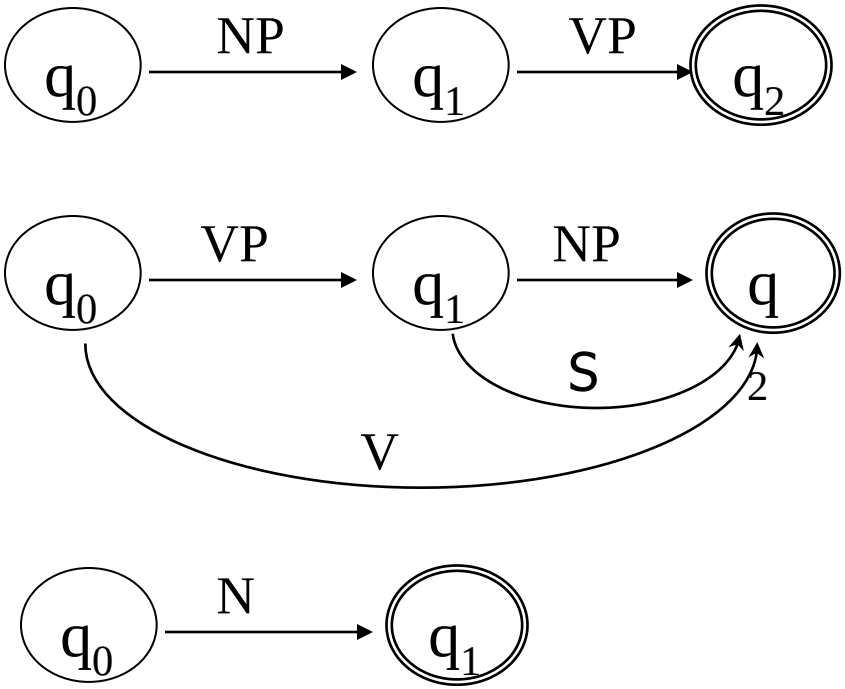
……

附录3 递归转移网络（Recursive Transition Network）

1. ***S -> NP VP*** → S

2. ***VP -> VP NP***
3. ***VP -> VP S***
4. ***VP -> V*** → VP

5. ***NP -> N*** → NP



ATN（Augmented Transition Network）：
增加了条件测试与寄存器的RTN

附录4 跟语音相关的语言知识（1）

	单音节	双音节
单音节	暗 喜	*暗 高兴(喜欢)
双音节	*暗暗 喜	暗暗 高兴(喜欢)

	单音节	双音节
连	**连 查**了五家公司	* ***连 调查***了五家公司
一连	**一连 查**了五家公司	**一连 调查**了五家公司

附录4 跟语音相关的语言知识（2）

	例子	意义
轻音	张三　一天　就　挣三十块钱，大家都很羡慕他	“就”**前**量少
重音	张三　一天　**就**　挣三十块钱，大家都很同情他	“就”**后**量少

	例子	意义
轻音	他刚结婚，　又 正在度蜜月，怎么就遭遇了这样的不幸呢？	**追加**
重音	张三　**又**　来了	**重复**