



第三讲 汉语的句法规则系统

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



提 纲

一 汉语形式语法系统的构建

- 1 现代汉语的基本语法范畴
- 2 理论语言学的描述方式：**X-bar**结构理论
- 3 计算语言学的描述方式：上下文无关文法与特征结构合一描述的结合

二 汉语句法结构歧义类型及歧义消解举例

- 1 从人的角度看句法结构歧义
- 2 从计算机的角度看句法结构歧义
 - 2.1 外显型歧义与内含型歧义
 - 2.2 真歧义、准歧义、伪歧义
 - 2.3 句法结构歧义的统计分析
- 3 歧义消解举例



1 现代汉语的基本语法范畴

语言成分的两个性质：

(1) 组合性：两个小单位可以组合成更大单位；
syntagmatic relation

(2) 聚合性：两个单位之间可以相互替换；
paradigmatic relation



语言成分组合性的例子 (I)

多
很多

× 多 学生
很多 学生

很多 高 个子 学生

× 高 个子 很多 学生

个子
高 个子

× 个子 学生
高 个子 学生



语言成分组合性的例子 (II)

悠久 历史

——

历史 悠久

* 悠久 不 历史

——

历史 不 悠久

两千年的悠久历史

——

* 两千年的历史悠久



语言成分聚合性的例子

请指出下面例子在汉语中是能说的句子还是不能说的句子

- | | | |
|-----------------|----|-----------------|
| 1a. 他平常在北大打球 | —— | 1b. 他常常在北大打球 |
| 2a. 他平常都在北大打球 | —— | × 2b. 他常常都在北大打球 |
| × 3a. 他都平常在北大打球 | —— | × 3b. 他都常常在北大打球 |

语言成分聚合性的例子（续）

- 跟“平常”、“常常”类似的词分别有哪些？

平常

4a. 他过去在北大打球

5a. 他昨天在北大打球

6a. 他暑假在北大打球

.....

常常

4b. 他曾经在北大打球

5b. 他也 在北大打球

6b. 他刚刚在北大打球

.....



建立基本语法范畴的思路

1. 给出语言中基本组合关系的清单（**结构分类**）；
2. 观察词和短语进入基本组合关系中不同位置的能力，对词和短语进行**功能分类**；

<http://ccl.pku.edu.cn:8080/pos/> 汉语词类归属测试（调查）网页

<http://ccl.pku.edu.cn/course/xdhyjs/question.asp> 现代汉语语法基础知识网页

现代汉语的基本结构关系

结构类型	句法结构成分（位置）	实例
主谓结构	主语 + 谓语	老张 去 机器 很重
述宾结构	述语1 + 宾语	修理 桌子 学习 语法
述补结构	述语2 + 补语	看 清楚 站 稳
定中结构	定语 + 中心语1	木头 桌子 汉语 语法
状中结构	状语 + 中心语2	赶快 出发 非常 了解
连谓结构	前谓 + 后谓	走路 去 回家 休息
联合结构	前项 + 后项	长江 黄河 唱歌 跳舞

现代汉语的基本结构关系（续）

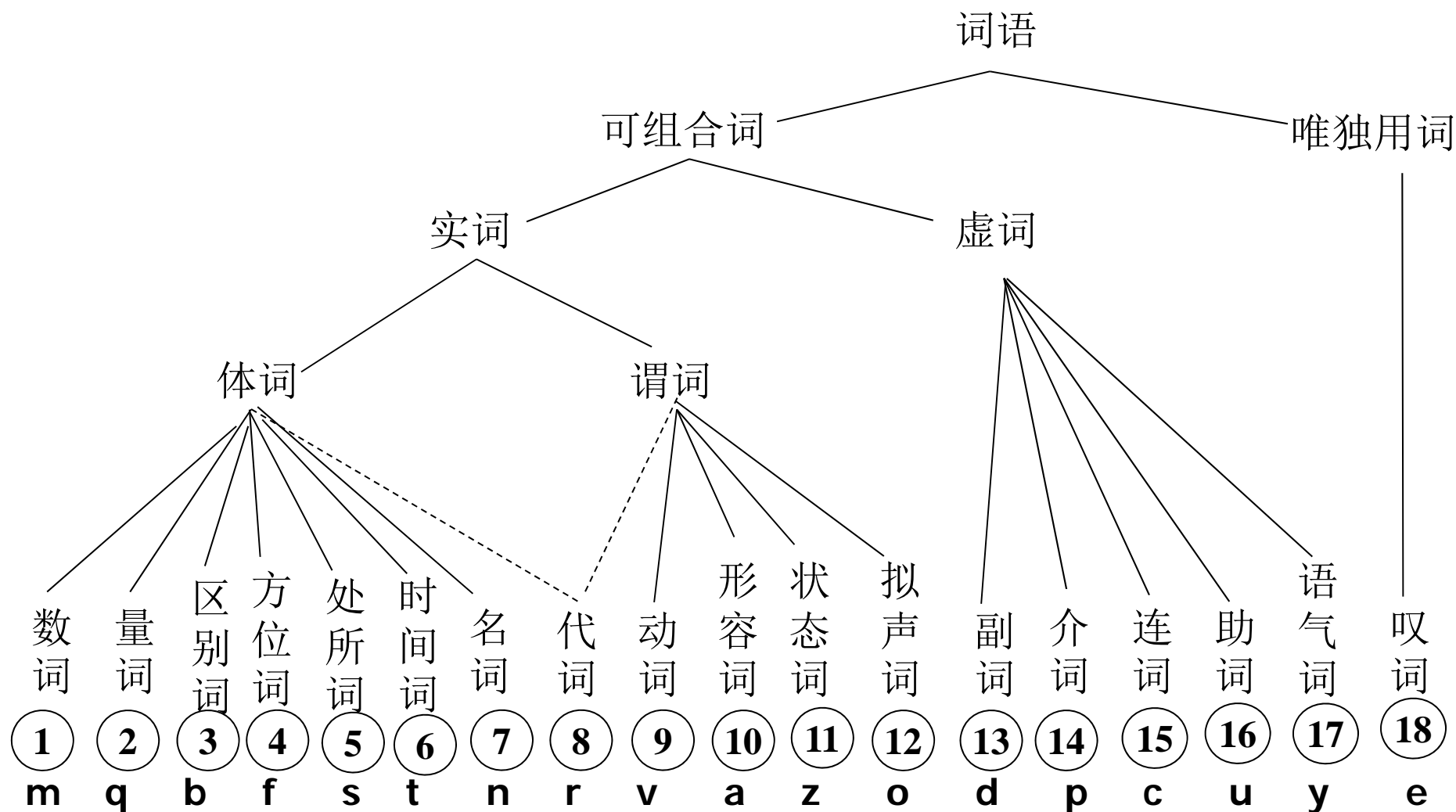
结构类型	句法结构成分（位置）	实例
“的”字结构	X + 的	听话 的 张三 的
“地”字结构	X + 地	悄悄 地 高兴 地
“所”字结构	所 + X X + 所 + Y	所 提 (条件) 学校 所 需要 (的)
介宾结构	介词 + 宾语	把 大家 向 窗外
方位结构	时间 处所 + 方向 相对位置	春节 以前 桌子 上
数量结构	数词 + 量词	三十二 本 两 批



根据词语占据的结构位置划分词类

- 数词： 可枚举
- 量词： 可枚举
- 名词： 数+量+ _____
- 动词： 主+ _____ _____+宾 _____+补
- 形容词： 主+ _____ ~~_____~~+宾 很 _____
-

现代汉语的词类系统





词类区别举例：形容词与状态词

白 — 雪白

香 — 喷香/香喷喷

1. 很 _____

2. 张三的脸 比 李四的脸 _____

3. 这碗饭 比 那碗饭 _____



汉语的短语类及示例

短语	实 例
副词性短语dp	飞快地；轻松而愉快地；波浪式地；
介宾短语pp	关于专家系统；从桌子上；被我们；在后面；比这里；从北京到那里； 除他之外；
处所短语sp	报纸上；我前面；我们班里；
时间短语tp	一个秋天的早晨；下星期一；吃饭前；
数量短语qp	两百张；三十岁；三场；多少斤；
数词短语mp	六七百；三万两千零五十；四又二分之一；五点三二；大多数；不少；几；



汉语的短语类及示例（续）

短语	实 例
小句dj	爱夸张事实的孩子往往喜欢喜剧；三点钟全体集合；今天星期一；他二十来岁；长两米；重三斤；
名词性短语np	粒子碰撞噪声检测仪；计算机在国外应用的现状；世界名牌服装；新问题；自己的； 桌椅门窗；理想与现实；支持总统的群众；给孩子们；服装设计；两国之间的合作；几十年的努力； 他们两位；录象带两百盘；最善良的一个；三斤重；两米宽；
动词性短语vp	把杂志放进抽屉里；进行多方面的经济结构的调整；从暴风雪中救出了一群羊；来了；请客人吃饭；去外婆家玩；烧毁证物并袭击警察；跑得我累死了；
形容词性短语ap	很不高兴；冷得发抖；比他们房间冷得多；干干净净的；通红通红的；亮了；干净不了三天；不礼貌而且不诚实；长三米；小两岁；



建立语法范畴时碰到的问题

1. 从典型到非典型：元素对集合的隶属度

例1：词类划分与具体词语的词类归属

例2：“所”字结构的功能分类

2. 从大类到小类：知识颗粒度的粗细

例3：np的内部划分

例4：跟“把”字结构组合的vp



类的典型成员和非典型成员

例1 词类划分与具体词语的词类归属

A. 电脑 风衣 风水 文笔 战争 看法

B. 般配 一样 矮小 安全 耐烦 方便

A 组中不大典型的名词是 (1) 风水 文笔 (2) 战争 (3) 看法

B 组中不大典型的形容词是 (1) 一样 耐烦 (2) 方便



类的典型成员和非典型成员

例2 “X 所 Y”结构应属什么短语类？

他所盖的房子

他所买的衣服

他所发明的机器

他所治疗的病人

他所制作的工艺品

.....

今天所讲的内容

刑警所说的电影

他所审查的论文

他所参观的展览会

他所传达的指示

.....

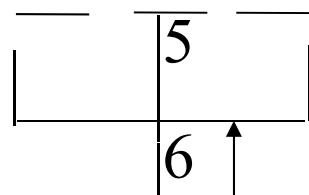
“所字结构”的层次分析

他 所 写 的 文 章

1 2

3 4

7 8



1-2 定中结构

3-4 “的”字结构

5-6 “所”字结构

7-8 主谓结构

他 所 写 的 文 章

1 2

3 4

5 6

7 8

1-2 定中结构

3-4 “的”字结构

5-6 主谓结构

7-8 “所”字结构



“所”字结构层次分析的问题

1. “X 所 Y 的 Z”中，“X 所 Y”是二分，还是三分？
2. “X 所 Y”结构整体的性质是怎样的？是体词性结构？谓词性结构？



“(X) + 所 + Y”示例

1. 所使用的案例还是很早的
2. 所采用的主要原料
3. 所要学习的内容还有很多
4. 所不同的只是垒球的规则比棒球复杂一些
5. 所藏宝石、软体动物和海洋生物标本尤为名贵
6. 所生子女属于母亲一方
7. 她找到了所要找的花

“所 + Y”前面可以没有成分

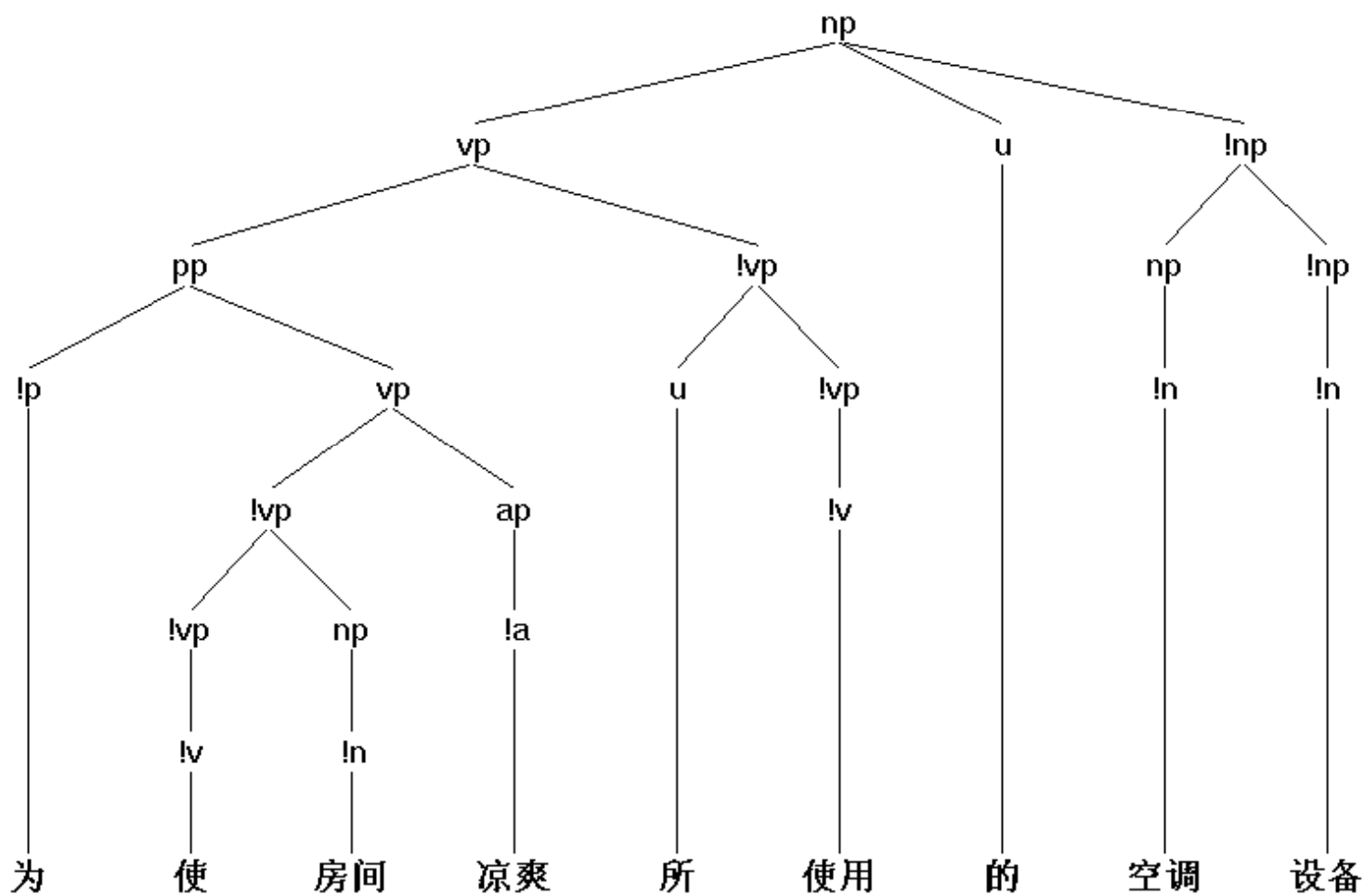


“所” 字结构的context

8. 全靠了他 **卖血** **所** **换得** 的钱, 才...
9. **为** **使** **房间** **凉爽** **所** **使用** 的空调设备
10. 毛泽东 **在** **这次会议** **上** **所** **作** 的报告
11. 即将 **由** **这次** **停火** **所** **带** **来** 的新形势
12. 他 **亲口** **所** **说** 的话

“所 + Y” 前面可以是vp, pp, dp等成分

“所”字结构的分析示例





从大类到小类

例3 np 的内部分类

1a 皮凉鞋

b 皮鞋和凉鞋

2a 网络文学

b 网络和文学

3a 新型轿车

b 新型的轿车

4a 皮手套

c 皮的手套



定中结构np — 联合结构np

	数量 + _____	vp + _____
np_dz 定中结构	三双 <u>皮凉鞋</u>	看 <u>网络文学</u>
np_lh 联合结构	*三双 <u>皮鞋和凉鞋</u>	*看 <u>网络和文学</u> 喜欢 <u>网络和文学</u>



粘合定中结构np — 组合定中结构np

	vp + _____	
np_dz_nh 粘合定中结构	进口 <u>新型轿车</u>	定中 述宾
np_dz_zh 组合定中结构	进口 <u>新型的轿车</u>	* 定中 述宾



从大类到小类

例4 跟“把”字结构组合的vp有何特殊性？

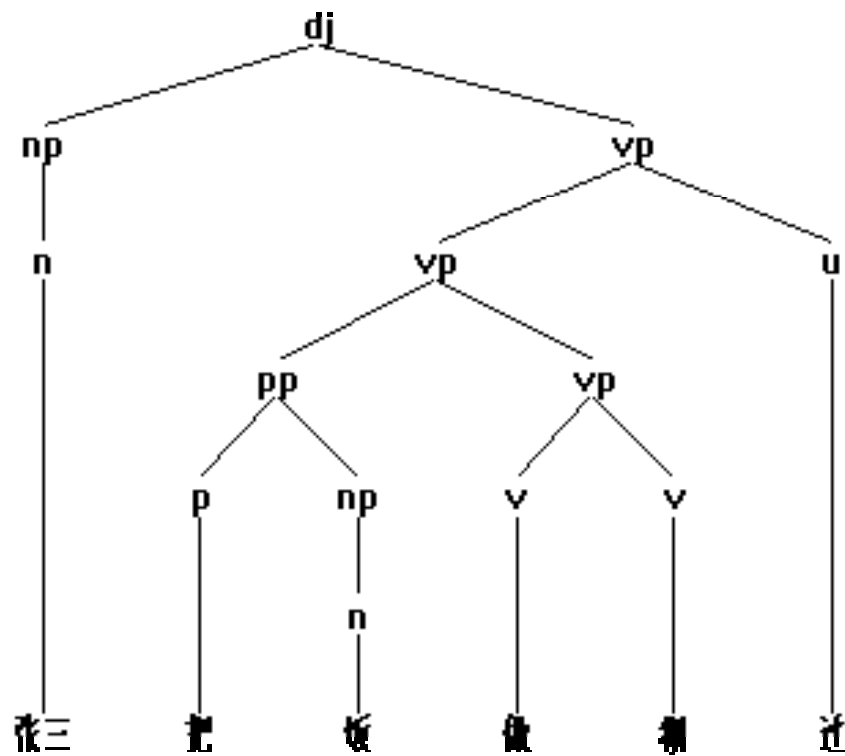
- a. 张三把饭做了
- b. 张三把饭做过
- c. 张三把饭做糊过

(1) 这些句子能说不能说？

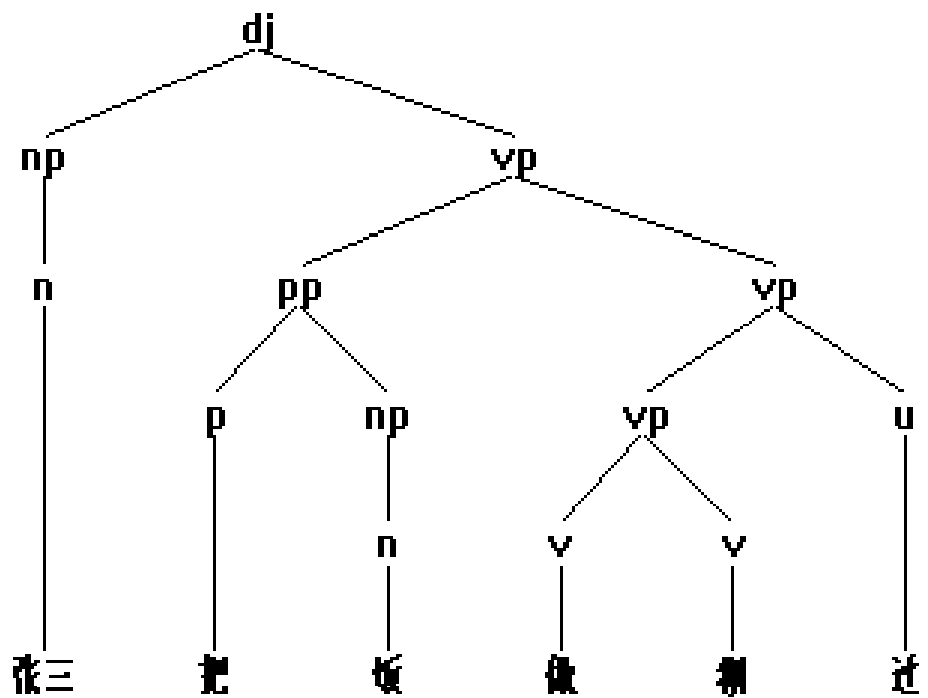
(2) 能说的句子，句法结构树怎么画？

“张三把饭做糊过”的句法结构？

甲



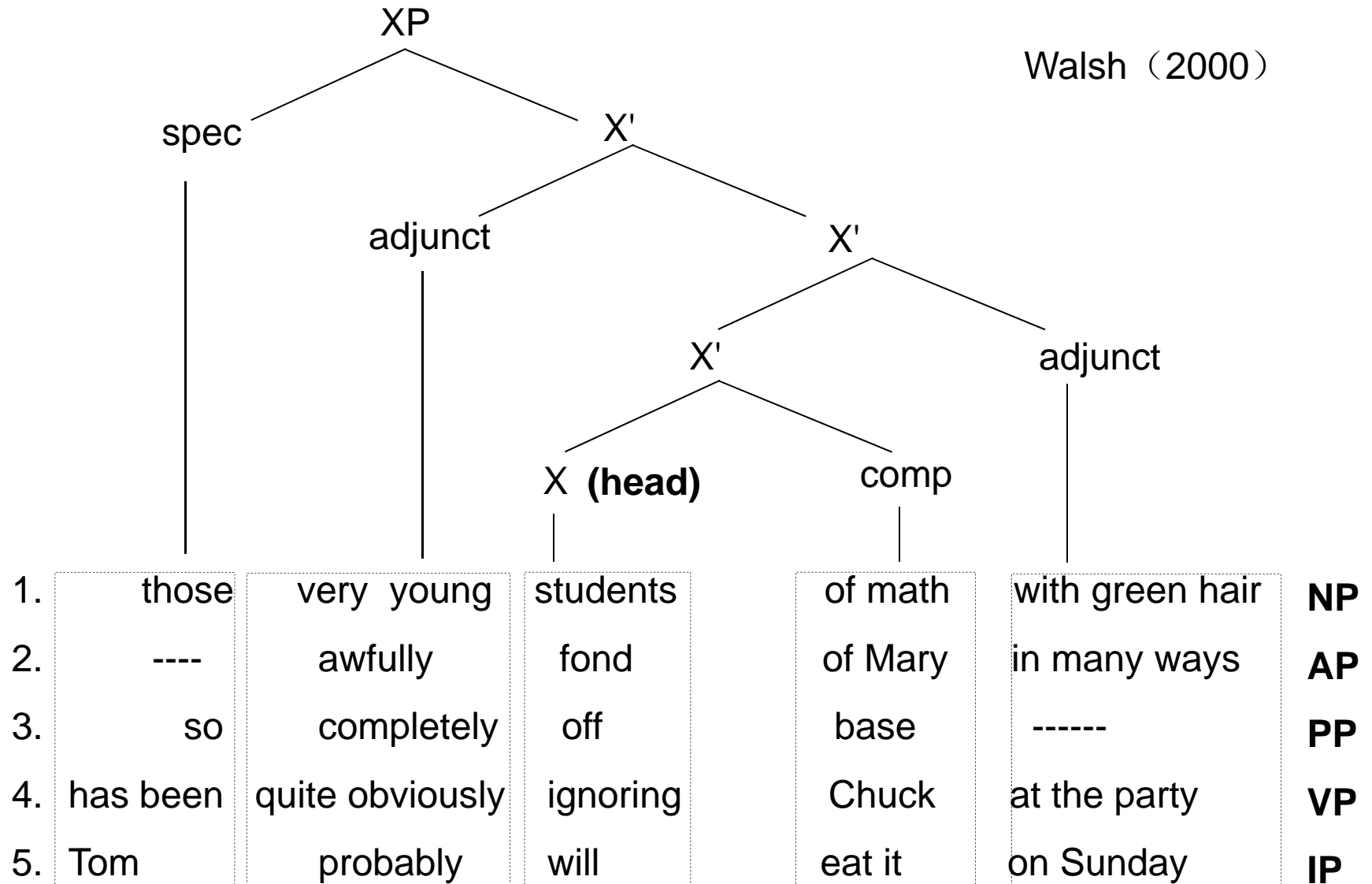
乙



1. 数学教员外号“杨半本”，他讲代数、几何，从来没有**把**一本书讲完**过**
2. 这跟作家的想象力没有任何关系，再有想象力，也不可能**把你没见过**，没听**过**，一点儿不懂不知道不熟悉不了解的东西写得栩栩如生
3. 那真能干，就没说**把饭做糊过**，不干不稀，特解饱。
4. 几个人**把稿子略略翻过**，不由目瞪口呆，顿时觉得自己矮了一挫
5. 我到一团这几个小时，你从未**把一支烟抽完过**，可又是一支接一支地抽。
6. 看上去的效果像是一个粗心的工人马马虎虎地**把小孔修补过**，但修得非常不精致。
7. 阿基夫的贵族们**把它重新装潢过**，膨大的饰布极尽奢华之能事。
8. 待到**把晚饭吃过**，爬出舱外一看，呀，好一轮圆月！

2 X-bar分析法假设的结构模式

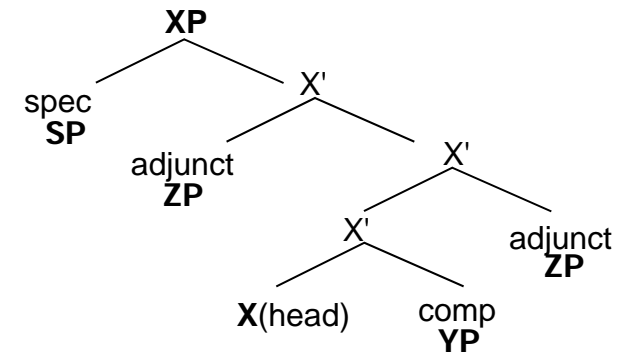
Walsh (2000)



X-bar 结构规则

1. $XP \rightarrow \text{Spec } X'$
2. $X' \rightarrow \text{adjunct } X'$
3. $X' \rightarrow X' \text{ adjunct}$
4. $X' \rightarrow X \text{ Comp}$
5. $X' \rightarrow X$

1. $XP \rightarrow \text{SP } X'$
2. $X' \rightarrow \text{ZP } X'$
3. $X' \rightarrow X' \text{ ZP}$
4. $X' \rightarrow X \text{ YP}$
5. $X' \rightarrow X$



- Chomsky, Noam (1970). Remarks on nominalization. In: R. Jacobs and P. Rosenbaum (eds.) Reading in English Transformational Grammar, 184-221. Waltham: Ginn.
- Jackendoff, Ray (1977). X-bar-Syntax: A Study of Phrase Structure. Linguistic Inquiry Monograph 2. Cambridge, MA: MIT Press.
- **Thomas Walsh, 2000, A Short Introduction to X-bar syntax and transformations, 2nd edition, Parlay Press**
- **Richard K. Larson, 2010, Grammar As Science, MIT Press, 2010, Unit 23, pp.343-354.**
- 何元建, X标杆理论与汉语短语结构, 《国外语言学》1995年第2期。



区分X周围不同语法范畴的依据

- 替换测试 (substitution test)
- 并列测试 (conjunction test)
- 移位测试 (movement test)

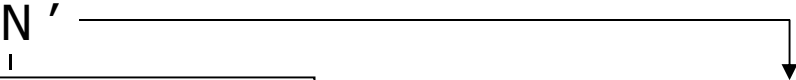


替换测试 (substitution test)

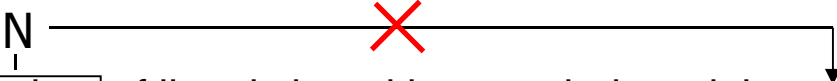
区分 N' 和 N

1 the diligent student of linguistics with green hair

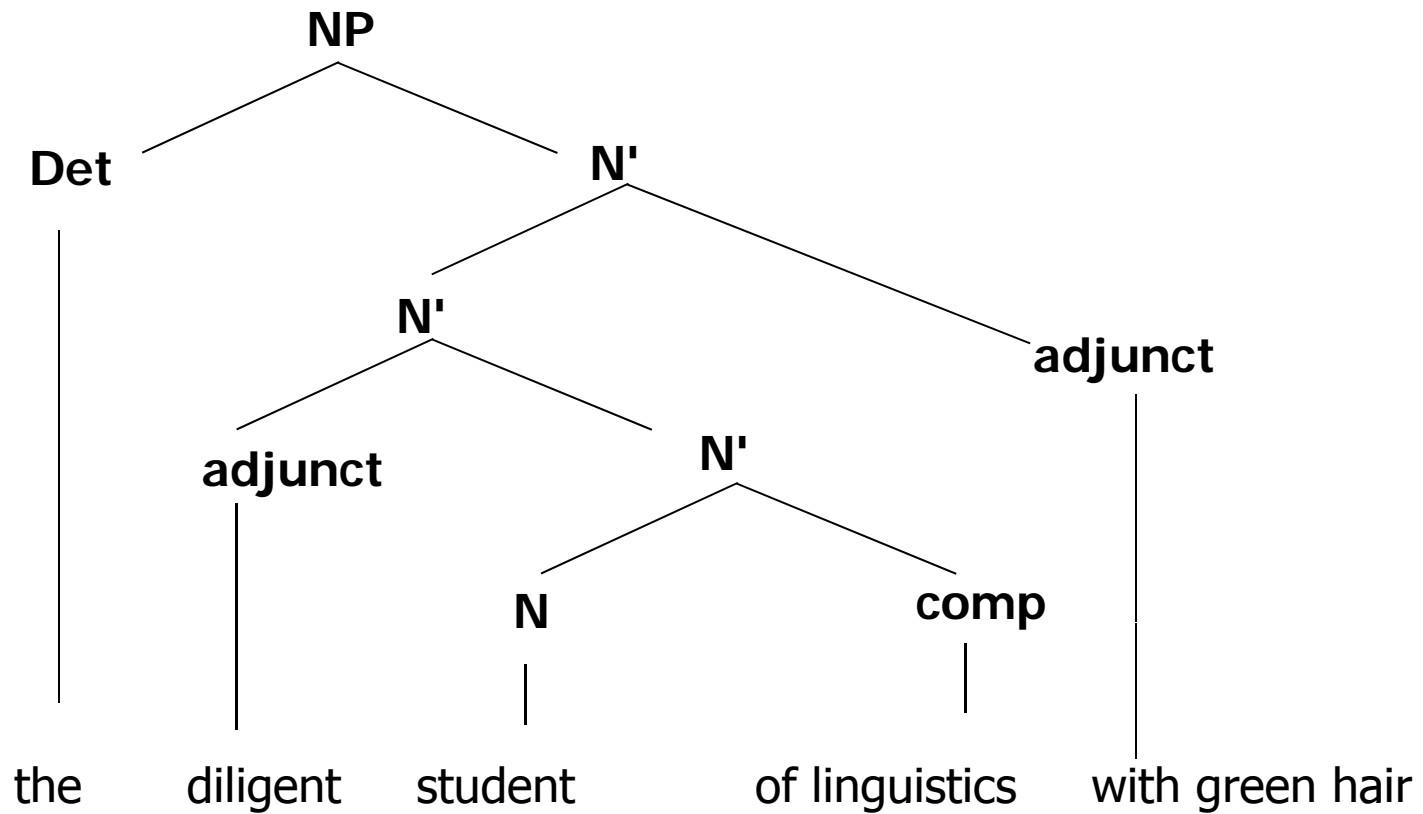
1a. I know the diligent student of linguistics with green hair and the **one** with blue hair.



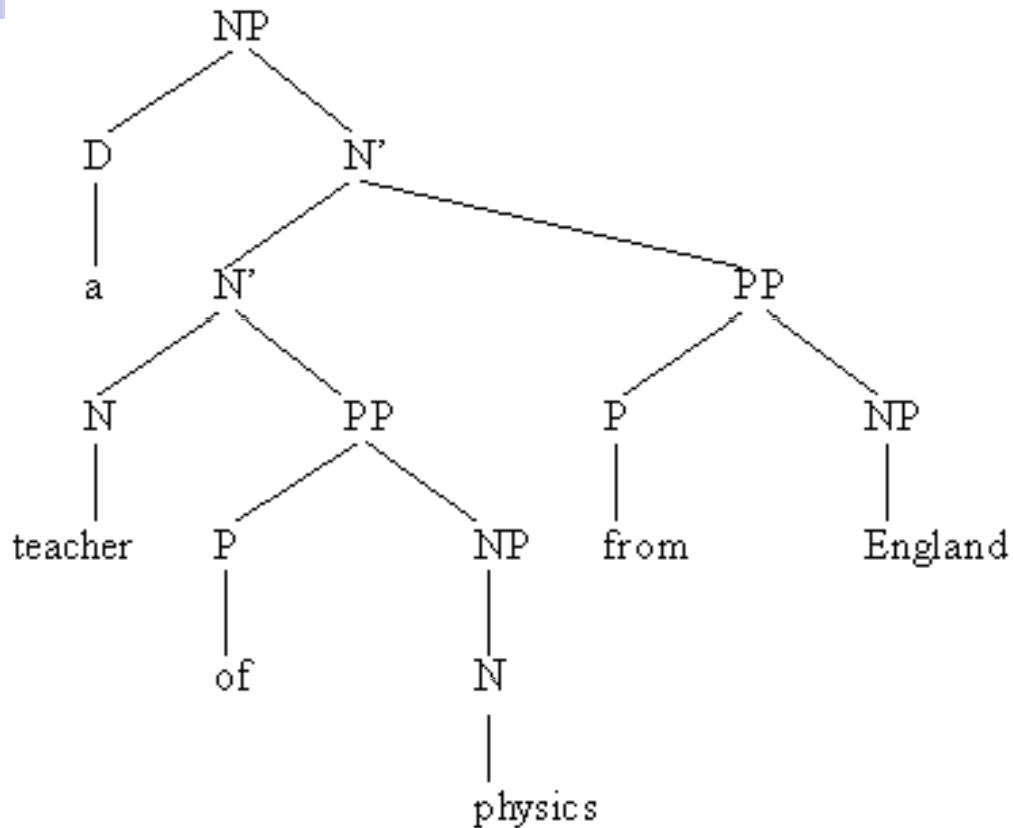
1b. * I know the diligent student of linguistics with green hair and the **one** of physics.



替换测试 (substitution test)



区分 N' 和 N



N'可以由one 替换，
N不能

I know the **teacher of physics** from England and the **one** from France.

*I know the **teacher** of physics from England and the **one** of chemistry.



并列测试 (conjunction test)

1. The teachers of physics and of chemistry.
2. The teachers from Turkey and from Spain.
3. * The teachers from Turkey and of physics.
4. * The teachers of physics and from Turkey.

区分 comp 和 adjunct 成分



移位测试 (movement test)

1 **diligently** study linguistics **with passion**

1a. study linguistics **with passion** **diligently**

1b. * study linguistics **diligently** **with passion**

2a A teacher **of physics** **from England**

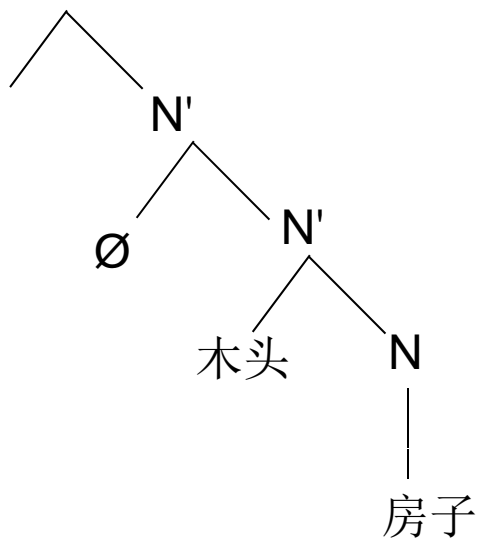
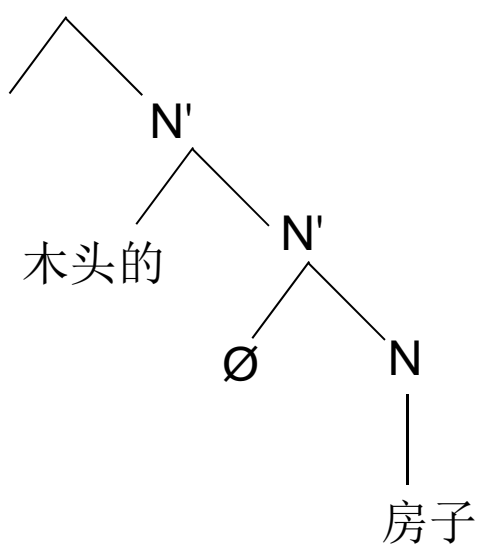
2b * A teacher **from England** **of physics**

comp 比 adjunct 更靠近 head

汉语中的例子（替换测试）

我 喜欢 木头 的 房子
我 喜欢 木头 房子
她 只 看 重要 的 文章
她 只 看 重要 文章

(房子) 我 喜欢 木头 的 那 栋
* (房子) 我 喜欢 木头 那 栋
(文章) 她 只 看 重要 的 那些
* (文章) 她 只 看 重要 那些



N' 可被指代词替换, N 不能



汉语中的例子（并列测试）

1. 物理系和化学系老师
2. 土耳其的和西班牙的老师
3. * 土耳其的和物理系的老师
4. * 物理系的和土耳其的老师
5. 新的 和 旧的 教科书
6. 新旧 教科书
7. * 新的 和 旧 教科书
8. * 旧 和 新的 教科书



汉语中的例子（移位测试）

adjunct complement head
一块儿 把 家里 收拾 干净

? 把家里 一块儿 收拾干净

adjunct head complement adjunct
马上 放了一些调料 在锅里

马上 在锅里 放了一些调料

? 在锅里 马上 放了一些调料

adjunct adjunct comp head
怪异的 绿头发 足球 运动员

* 怪异的 足球 绿头发 运动员

* 绿头发 足球 怪异的 运动员

* 绿头发 怪异的 足球 运动员

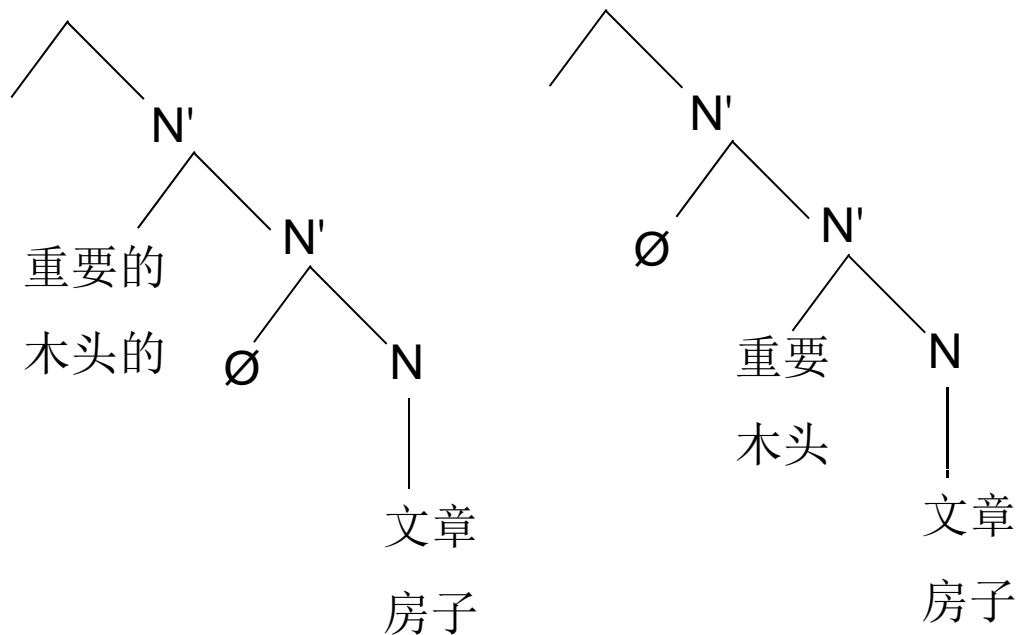
* 足球 怪异的 绿头发 运动员

* 足球 绿头发 怪异的 运动员

汉语中的例子（移位测试）

我 喜欢 木头的 房子
 我 喜欢 木头 房子
 她 只看 重要的 文章
 她 只看 重要 文章

房子 我 喜欢 木头的
 * 房子 我 喜欢 木头
 文章 她 只看 重要的
 * 文章 她 只看 重要

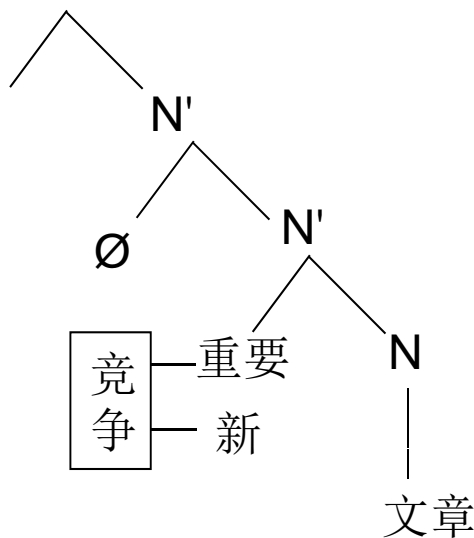
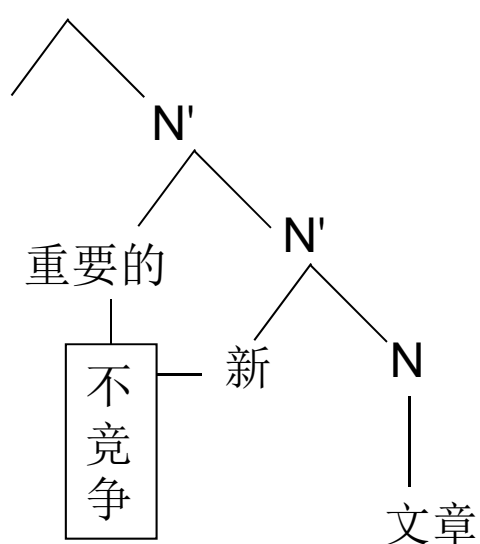


N' 可移位, N 不能移位

汉语中的例子（同位竞争测试）

新 文章
 重要 文章
 ? 重要 新 文章
 ? 新 重要 文章

重要的 新 文章
 新的 重要 文章
 重要的 新的 文章
 新的 重要的 文章



竞争测试可说明head周围的成分存在地位上的“不平等”，因而可以也应该区分为不同的范畴。

complement只有一个位置，modifier (adjunct) 可以有多个。



小结

- **X-bar**理论模型试图将不同的短语纳入同样的结构模式中，进行统一的描写；
- 结构共性为：
 - 一个组合中不同的语言成分地位有主次之分；
 - 不同的次要成分跟主要成分之间的相对位置有远近差异；
 - **comp**离主要成分（**head**）最近，**adjunct**居中，**spec**最远；
 - **comp**跟**head**的关系紧密，因而跟**comp**为兄弟节点的**X**节点不能被替换；**adjunct**跟**head**的关系松散，跟**adjunct**为兄弟节点的**X'**节点可以被（指代词）替换；



3 上下文无关文法与特征结构合一描述结合

- 产生式规则：
描述一个语言的基本范畴及其组合模式；
- 基于特征结构的合一约束：
 - ①描述基本范畴之间发生组合关系的条件；
 - ②描述组合后整体的功能特征；

描述内容：内部构成 + 外部功能

规则描述内容 实例	内部构成	外部功能
一件衣服	qp + np , 定中结构, qp是定语, np是中心语;	名词性短语 (np) , √主谓结构的主语, √述宾结构的宾语, √定中结构的中心语; ×述补结构的补语, ×定中结构的定语, ×状中结构的状语和中心语;



“一件衣服”相关的规则示例

&& {R1} np -> qp !np

①—— \$.内部结构=定中,\$.定语=%qp,\$.中心语=%np,\$.zuodingyu=否,...,

②—— %np.数量名=是,...,

③—— IF %qp.量词子类=个体 THEN %np.个体量词=%qp.原形 ENDIF,...

&& {R2} qp -> mp !q

\$.内部结构=数量,\$.定语=%mp,\$.中心语=%q,\$.zuodingyu=是,...,

&& {R3} np -> !n

\$.内部结构=单词

- ① 说明np整体的特征，包括内部结构，定语，中心语，整体的功能分布特征等；
- ② 说明对中心语np的约束条件（独立条件）
- ③ 说明对定语mp与中心语np之间的相互约束条件（相关条件）

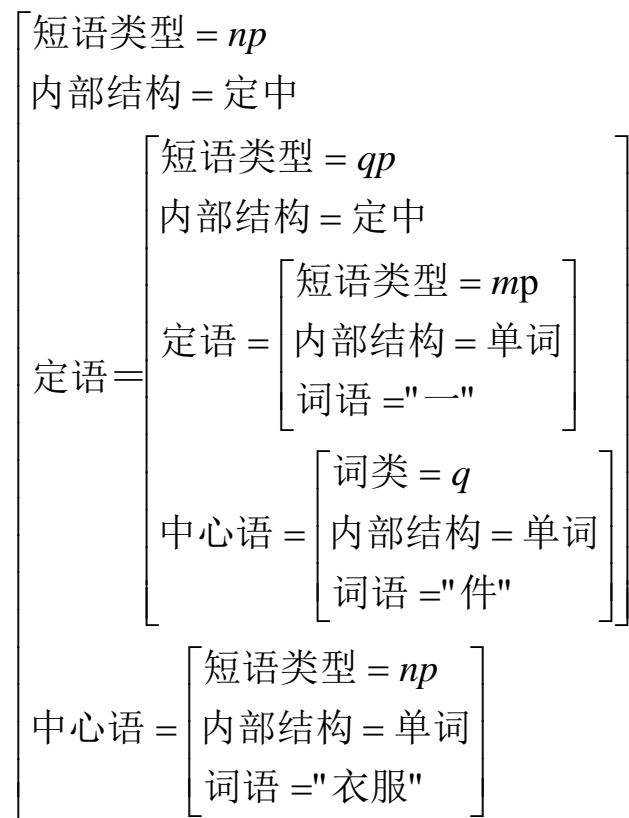
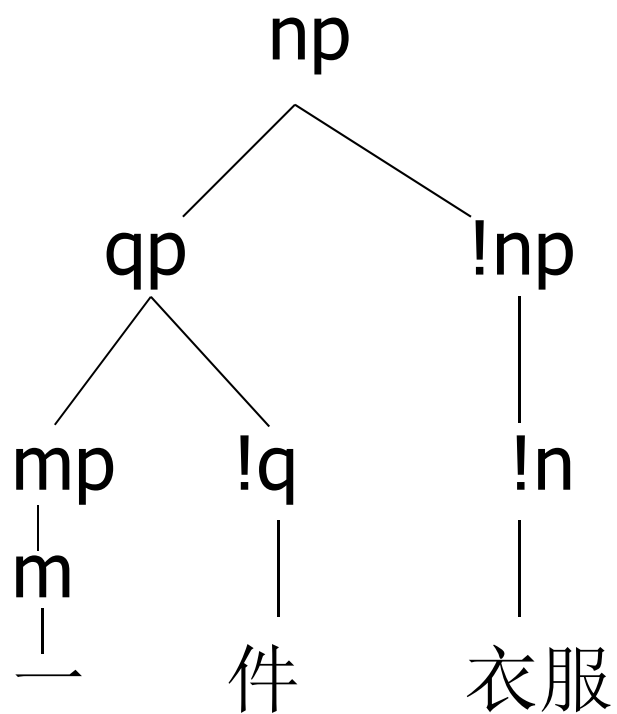


在词典中描述关于词的规则

词语	特征结构
.....
一	[词性:m,数词子类:基数]
件	[词性:q,量词子类:个体,表数:数]
衣服	[词性:n,数量名:是,个体量词:件 套,...]
心胸	[词性:n,数量名:否,...]
.....

俞士汶等,《现代汉语语法信息词典详解》(第2版),北京:清华大学出版社,2003年2月

“一件衣服”的分析结果





规则的应用示例

1. 她买了一件衣服
2. 董永拿走了七仙女的一件衣服
3. * 一个心胸
4. * 一个衣服
5. * np[[一件衣服] [领子]]
6. * np[[一件] [衣服领子]]



规则中的特征选取：注意覆盖范围不同

A np → qp !np :: \$.zuodingyu=否 /* “一件衣服” 不能再作定语 */
np → np !np :: %np. zuodingyu=是, %%np. zuozhongxinyu=是

B np → qp !np :: \$.内部结构=组合定中
np → np !np :: IF %np.内部结构=组合定中,%%np.内部结构=单词 FALSE

就特定的语言对象而言，往往可以用不同的约束条件，达到同样的约束效果。但是，所谓同样的效果一般是针对特定的范围而言的。在实际应用中，不同约束条件选取的特征不同，其覆盖范围往往也是不同的。比如上面的例子中**A**例的条件约束覆盖面就比**B**例更广；

规则中要注意描述功能的变异/差异

甲：
从无到有

- | | | | |
|----|------------------|----|-----------------|
| 1. | * 很 <u>有</u> | —— | 很 <u>有理想</u> |
| 2. | * <u>渠道</u> 拓展市场 | —— | <u>多渠道</u> 拓展市场 |
| 3. | * 有 <u>胖</u> | —— | 有 <u>那么胖</u> |
| 4. | * 有 <u>高</u> | —— | 有 <u>三米高</u> |
| 5. | * 达到 <u>人</u> | —— | 达到 <u>三万人</u> |

乙：
从有到无

- | | | | |
|----|------------------|----|---------------------|
| 1. | 比李四 <u>好</u> | —— | * 比李四 <u>很好</u> |
| 2. | 比李四 <u>高</u> | —— | * 比李四 <u>不高</u> |
| 3. | <u>很客气地</u> 说 | —— | * <u>最客气地</u> 说 |
| 4. | 你找他 <u>什么事</u> | —— | * 我找他 <u>这件事</u> |
| 5. | 你 <u>找</u> 他 什么事 | —— | * 你 <u>找过</u> 他 什么事 |

例1: 状中式ap的内部差异

很好
更好
不好
更不好

ap → dp !ap :: \$.内部结构=状中,...

ap → pp !ap :: \$.内部结构=状中,...

这些状中结构的功能特征并不完全一样。

相同的特征: 作谓语

不同的特征: 在pp(比)之后的位置

(张三) 比李四 <u>好</u>	——	比李四 很好	×
	——	比李四 更好	
	——	比李四 不好	×
	——	比李四 更不好	



例2：述宾式vp的内部差异

1a. 相信 上帝

→ 1b. 曾经 相信上帝

→ 1c. 相信上帝 的人

vp → dp vp :: \$.内部结构=状中,...

np → vp u<的> np :: \$.内部结构=定中,...

- 述宾式vp可以在dp之后出现，构成状中结构
- 述宾式vp可以在“的 np”之前出现，构成定中结构

2a. 相信 不 相信 上帝

→ 2b. 曾经 相信不相信上帝

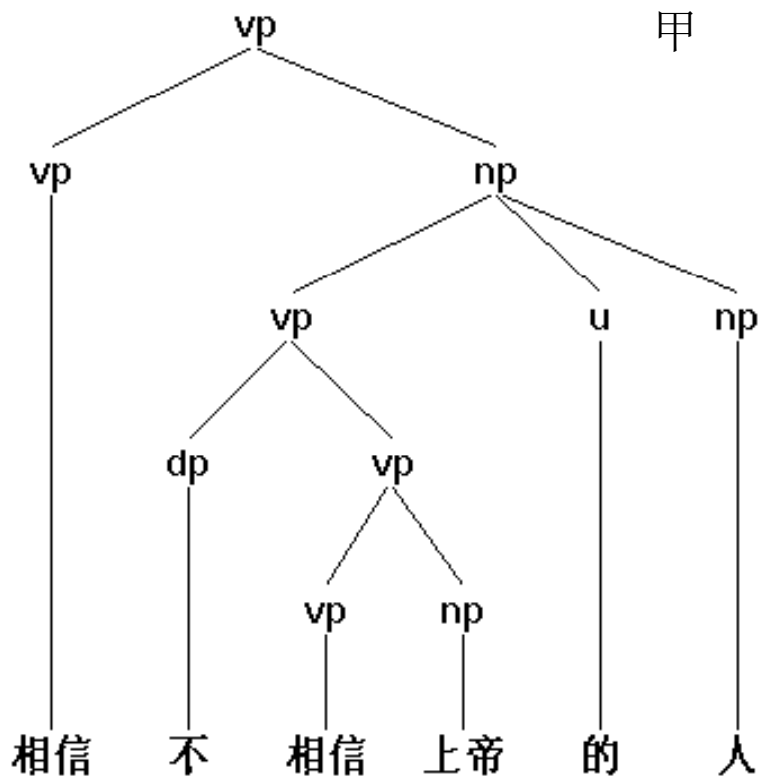
→ 2c. 相信不相信上帝 的人

3a. 曾经 相信不相信上帝 的人

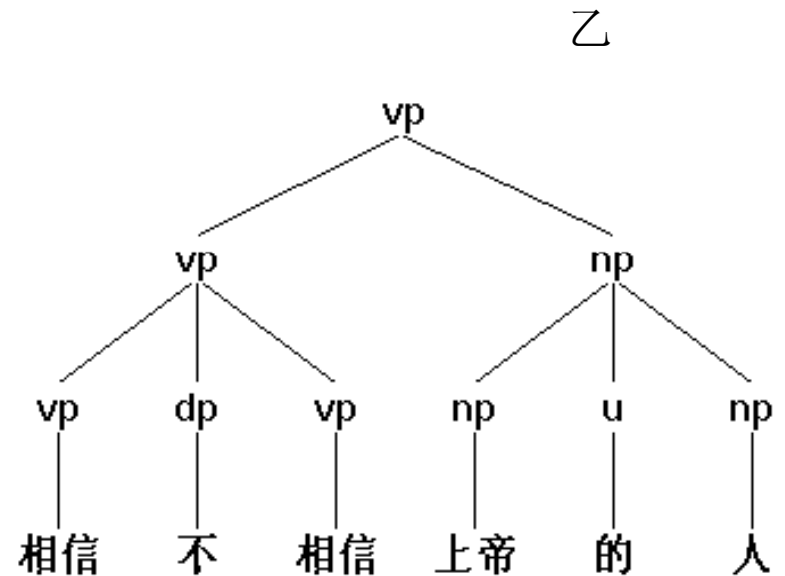
例2a的功能跟例1a的述宾式vp有显著差异。例2a的结构类型判别：述宾式？正反并列式？如果为后者，可增加约束：正反并列式不能作定语。

例2c和例3a中“相信不相信上帝”都不再是一个述宾式vp。

述宾式vp的内部差异

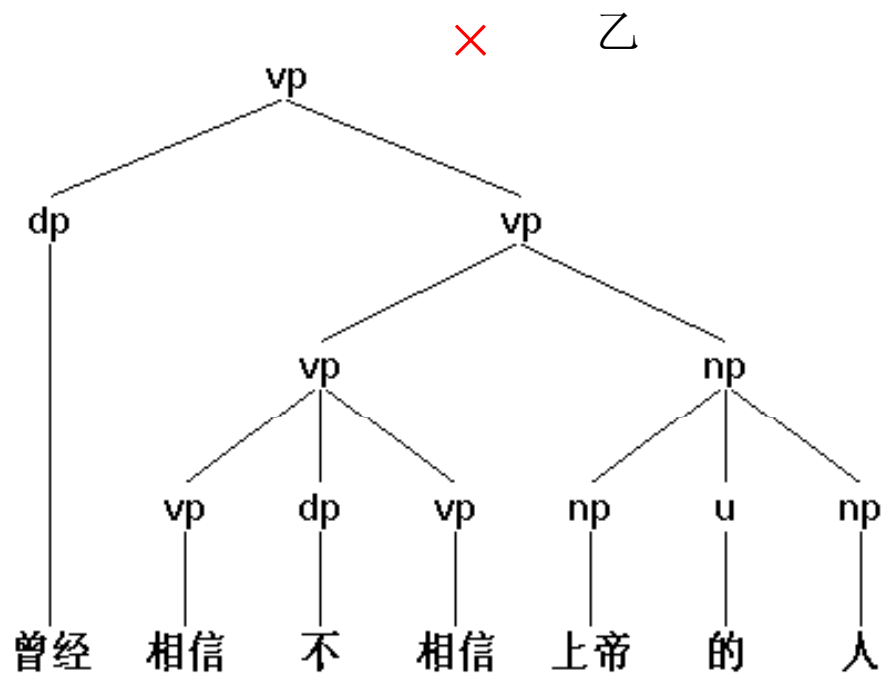
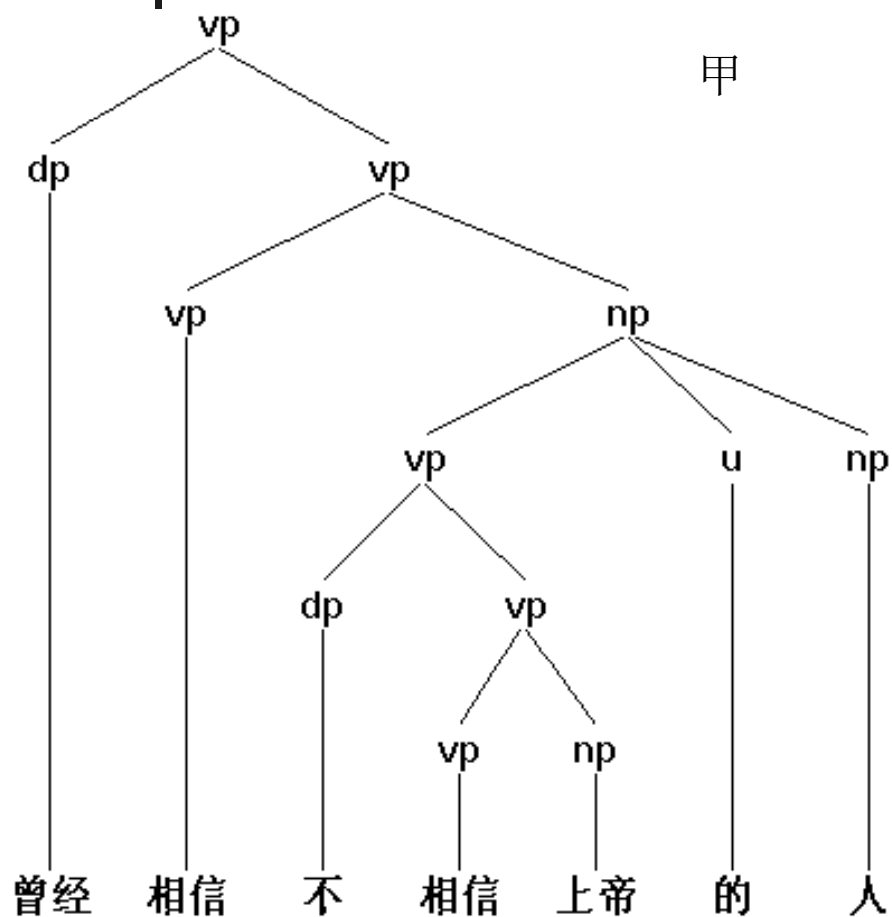


2c. 相信不相信上帝 的人



述宾式vp的内部差异

3a. 曾经 相信不相信上帝 的人



例3: “ap+的” 格式的内部差异

- | | | |
|------------------------|---|---------------|
| 1. 红de 给您 | → | np → ap de |
| 2. 认真 de 回答 | → | |
| 3. <u>很认真</u> de 回答了问题 | → | ap → ap de |
| 4. <u>最认真</u> de 回答了问题 | → | |
| 5. <u>遗憾</u> de 通知您 | → | np → ap de np |
| 6. <u>多么美妙</u> de 前景 | → | |

至少有三类不同的“ap+的”，其功能特征分别为：

- (A) 可以作主/宾语，不能作状语（例1，4）
- (B) 不能作主/宾语，可以作状语（例3，5）
- (C) 不能作主/宾语，只能作定语（例6）

例2 兼A、B
两类

例4 “把 NP VP u<过>”的规则

vp → pp<把> !vp :: \$.内部结构=状中,\$.状语=%pp,\$.中心语=%vp,
%vp.内部结构=~单词,

IF %vp.内部结构=附加 THEN %vp.附加语.原形=~过 ENDIF

C1

...

IF %vp.内部结构=附加,%vp.附加语.原形=过,
%vp.中心语.内部结构=单词 FALSE

C3

vp → vp u<了|着|过> :: \$.内部结构=附加,\$.中心语=%vp,\$.附加语=%u,...

IF %u.原形=过, %vp.内部结构=状中 THEN
%vp.状语.原形=~把 ENDIF

C2

层次构造: a 把 NP VP + 过

b 把 NP + VP 过

a ? b? (把饭做糊过)

C1

C3

+ C2

~~C1~~

选 a 只需要C1一项约束, 选b不需要C1约束, 但需要C2, C3两项约束



小结

任何一个为自然语言构造的形式文法系统 G ，都必须回答：

- 1) 为了描述一个自然语言 L ，需要多少个基本范畴 C_i ？
- 2) 需要在 $\{C_i\}$ 上定义多少种结构关系 R_i ？
- 3) 任给两个范畴 C_m, C_n ，它们之间是否有关系 R_x ？
- 4) C_m 与 C_n 之间形成关系 R_x ，需要满足什么条件？
- 5) C_m 与 C_n 组成的范畴 $C'_{(m,n)}$ ，具有什么样的性质？具体要考虑：
哪些性质是从 C_m, C_n 继承来的？
哪些是 C_m, C_n 原本没有，而在 $C'_{(m,n)}$ 中新增的功能属性？
哪些功能属性是 C_m, C_n 原本有，而在 $C'_{(m,n)}$ 中反而没有的？



思考题

1. 请为下面例子写出结构规则及合一约束条件。
 - (1) 一个心胸比较狭隘
 - (2) 一个心胸宽广的人
 - (3) 一件心胸宽广的人才会穿的衣服
2. 如果把汉语中“所”字结构的规则表述为： $vp \rightarrow u\langle\text{所}\rangle vp$ 那么，“所”后的vp要满足哪些约束条件？“所”字结构这种vp的整体功能应如何描述？（或者说，“所”字结构vp的分布跟其他类型的vp有哪些差异？）



思考题

3. 汉语的np短语包含哪些组合类型？请写出规则及约束条件。

np → !np 的 :: \$.内部结构=的字,...

np → !vp 的 :: \$.内部结构=的字,...

np → ap !np :: \$.内部结构=粘合定中,...

np → qp !np :: \$.内部结构=组合定中,...

np → np !np :: IF %np.内部结构=单词,%np.内部结构=单词 THEN
\$.内部结构=粘合定中 ELSE \$.内部结构=组合定中
ENDIF,...

np → !np c np :: \$.内部结构=联合,...

.....

附录1：一个汉语规则描述语言的Backus-Naur范式

<规则> ::= && <规则名> <产生式> [:: <约束>]

<规则名> ::= "{" <标识符> "}"

<产生式> ::= <短语类标记> → <短语结构>

<短语结构> ::= <结构成分> { <结构成分> }

<结构成分> ::= [<中心成分标识>] <语类标记>

 | [<中心成分标识>]<语类标记> "<"<终结符>">"

<中心成分标识> ::= !

<语类标记> ::= <词类标记> | <短语类标记>

<词类标记> ::= n|v|a|t|s|.....

<短语类标记> ::= np|vp|ap|tp|.....

```

<约束> ::= <约束式> {,<约束式>}
<约束式> ::= <合一等式>
           | <条件测试句>
<条件测试句> ::= IF <合一等式> {,<合一等式>} TRUE
           | IF <合一等式> {,<合一等式>} FALSE
           | IF <合一等式> {,<合一等式>} THEN <合一等式> {,<合一等式>} ENDIF
           | IF <合一等式> {,<合一等式>} ELSE <合一等式> {,<合一等式>} ENDIF
           | IF <合一等式> {,<合一等式>} THEN <合一等式> {,<合一等式>} ELSE
               <合一等式> {,<合一等式>} ENDIF
<合一等式> ::= <内部变量>=<原子>
           | <内部变量>=<内部变量>
           | <外部变量>=<外部变量>
           | <外部变量>=<True>
           | <外部变量>=<False>
           | <外部变量>=<None>
           | <外部变量>==<外部变量> <属性项目> {<属性项目>}
           | <外部变量>! =<外部变量> <属性项目> {<属性项目>}
<外部变量> ::= <根结点标记>
           | <位置标记><语类标记>
           | <外部变量>.<外部属性>
.....

```



附录2: “现代汉语语法信息词典”简介

北大计算语言所 + 北大中文系联合研制

1. 词典概要
2. 知识表达范式: 词类 + 属性描述
3. 词典规模

俞士汶等(1998, 2003)《现代汉语语法信息词典详解》(第1, 2版), 清华大学出版社、广西科学技术出版社1998年版。

概要：面向语言工程的词语观

(1) 类型：不拘一格

词/词组（短语），语素/单纯词，成语

(2) 词条：严格筛选

“义项”与“语法功能”相结合的原则

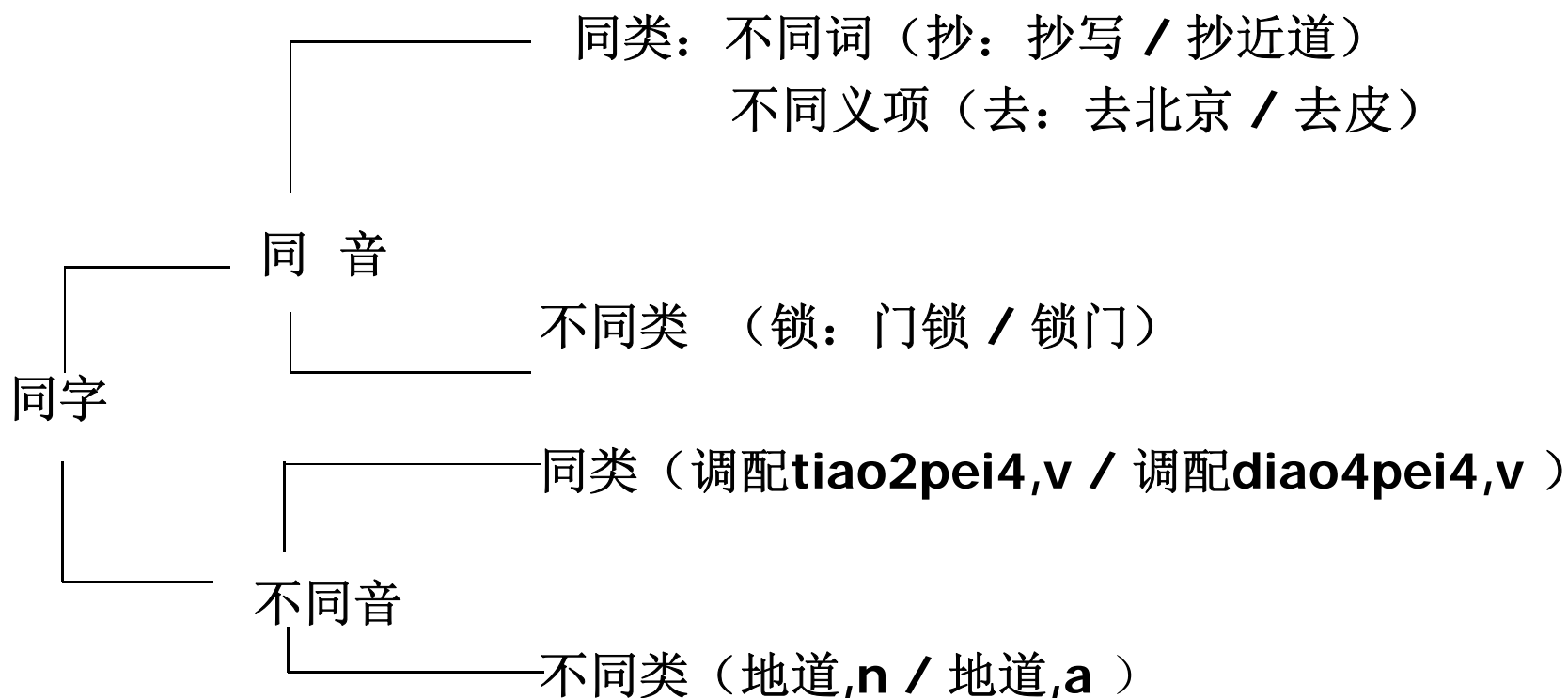
抄：不同词（抄写/抄近道）

去：不同义项（去北京 / 去皮）

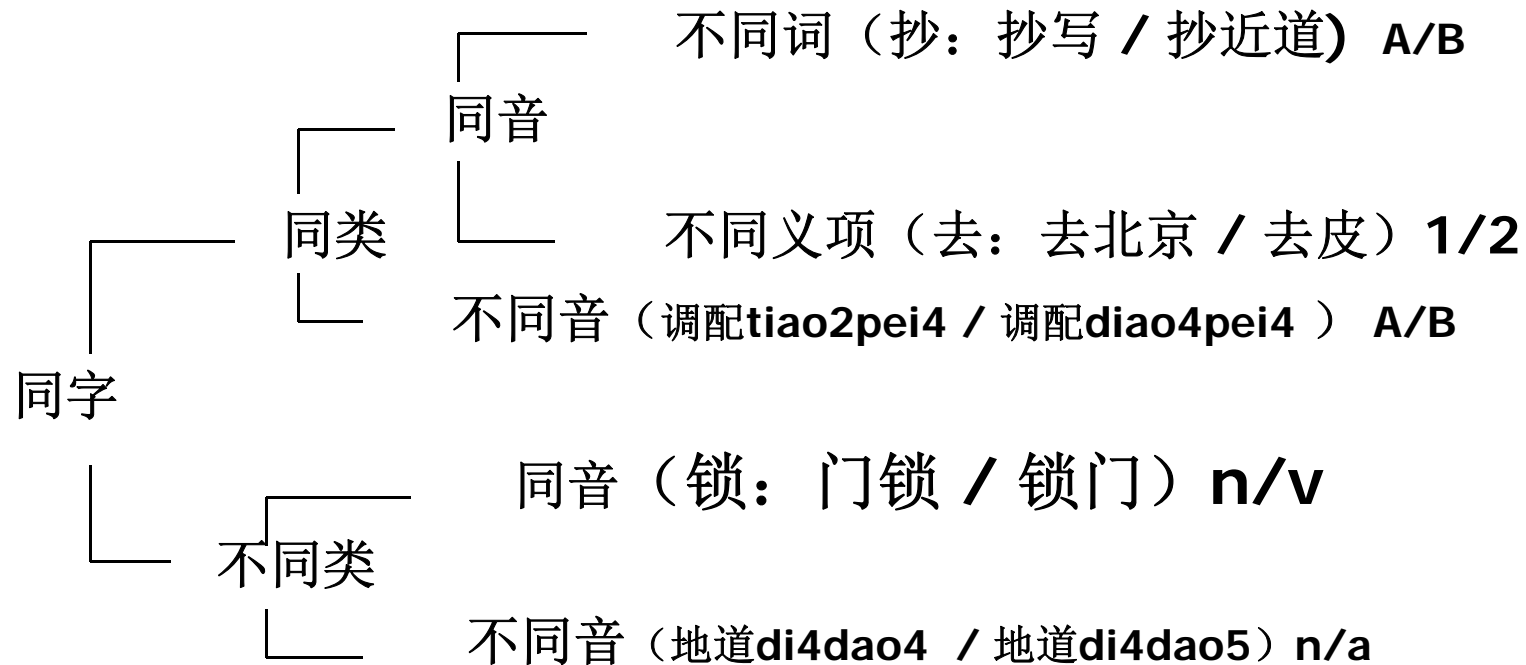
保管：两个义项，两个动词

材料：三个义项，一个词

同形词语的处理策略 (面向人的划分)



同形词语的处理策略 (面向机器的划分)



“词语” + “词类” + “同形” 作为数据库的主关键字 (Primary Key)



词语的归类

- 在本词典开发之前，众多的语法学家建立过各自的汉语词类体系，但都是以典型的例词说明各自体系的合理性、科学性。由于其研究是面向人的，又受限于当时的技术条件等原因，没有任何一位语言学家完成过数以万计的词语的归类。《现代汉语词典》只将数量很少的词（主要是虚词）归了类。
- 在汉语语法研究史上，本词典第一次完成了**7**万多词语的归类。到目前为止，还没有第二家。



词语的属性描述 (1)

从理论上来说，对于集合 S 上的一个分类 P 恒对应集合 S 上的一个属性表 A ，两者定义同样的等价关系。因此，分类与属性描述是可以相互转换的。

如果为所描述的对象确立 m 个二值属性（ $m \geq 1$ ），则最多可将对象的集合划分为 2^m 个不相交的子集。反之，若将对象的集合划分为 M （ $M \geq 2$ ）个不相交子集，则至少要确立 $[\log_2(M-1) + 1]$ 个不同的二值属性（这里的方括号代表取整运算）。



词语的属性描述 (2)

由于划归同一类的词语仍有相互区别的属性特征，继续细分会造成分类体系庞杂，难以适应不断发现新的属性特征的研究过程，属性描述是恰当的策略。

《现代汉语语法信息词典》最重要的设计思想是在分类的基础上详细描述属于同一类的每个词语的详细的语法属性。

这个设计思想也成为北大计算语言所后来开发其他语言知识库建设的指导原则。

从动词库中抽取的部分语法属性字段

词语	同形	义项	系词	助动	趋向	体谓准	双宾	单作补	复数主	后名	很	着了过	重叠	离合	兼类
保存						体				可		着了过	ABAB		
成为			系			体									
得到						体准						了过			
告诉						体谓	双					了过			
协商						体谓			复	可		了			
加以						准									
冒险										可		过	VVO	离	a
去	A1	除掉				体						了过	VV		
去	A2	~上海			趋	体		可				了过	VV		
去	B	扮演				体						了过			
应	A	答应						可				了			
应	B	应该		助		谓									
支持	1	支撑				体						着了过			
支持	2	鼓励并帮助				体谓准					很	着了过	ABAB		
指挥						体谓				可		着了过	ABAB		n

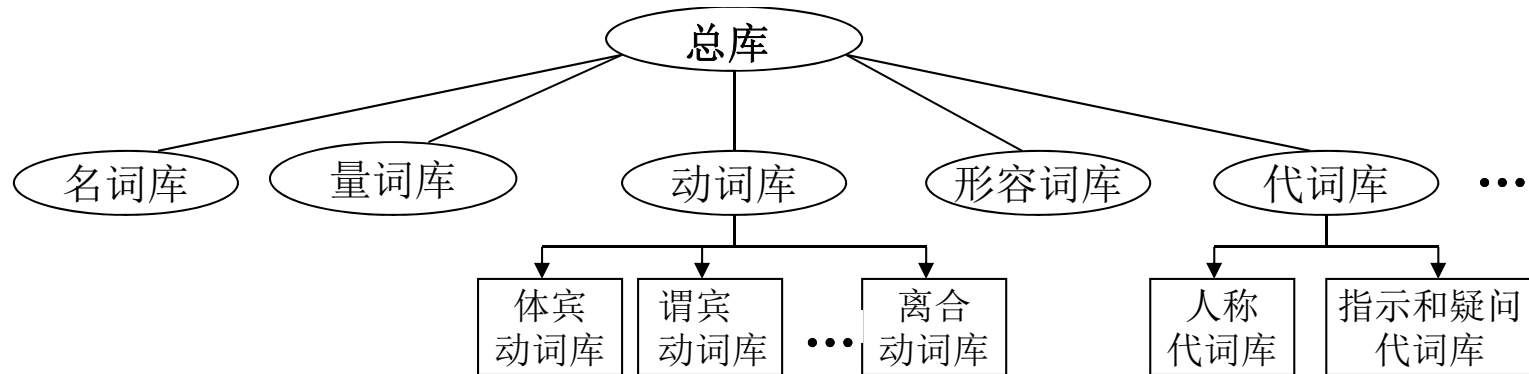
数据库的总体结构与规模

每一个数据库文件都刻画了词语及其属性的二维关系。

词典中共有34个数据库文件。总库1个，各类词库25个。其中，代词库下又设2个分库，动词库下设6个分库。

各类词的特有属性填在各类词库中。所有词的共同属性则容纳在总库中。总库中的属性包括读音、词类、拼音、虚实、体谓等，共计约20项。

所有的库都可以进行连结（JOIN），连结条件可以用“词语”+“词类”+“同形”主关键字表达。这样，34个库文件构成有上下位关系的“树”，子节点继承父节点的全部信息，或者说，将父节点与子节点连结起来就可以得到词语的更全面的信息。



库名	记录数	属性 字段数
总库	73877	13
名词	35201	31
时间词	565	16
处所词	183	15
方位词	194	21
数词	165	26
量词	456	24
区别词	757	13
代词	205	19
人称代词分库	49	8
指示代词分库	157	15
动词	14496	47
体宾动词分库	7630	27
谓宾动词分库	1321	8
双宾动词分库	185	12
动结式分库	3178	10

库名	记录数	属性 字段数
动趋式分库	6195	32
离合词分库	3420	8
形容词	2857	33
状态词	986	18
副词	1174	22
介词	108	28
连词	203	15
助词	38	12
语气词	53	13
前接成分	11	9
后接成分	43	9
成语	5264	15
简称略语	400	14
习用语	3031	15
语素	7223	14
标点符号	52	17
总计		579



工作量

73,874个词语的总信息量为3,369,828。这些信息所需存储空间为29,000,686字节。

最后粗略地计算一下这项工程所需要的人力。假设一个人确定一个属性信息只需要2分钟，填写3,369,828个信息需要6,739,656分钟，约等于112327.6小时。如果每天工作8小时，则需要14040.95天。按现在每年的工作日约为250天计算，折合的结果为56年，即完成这项工程至少需要投入56个人年的工作量(16年，每年4个人)。