

第五讲 语义分析

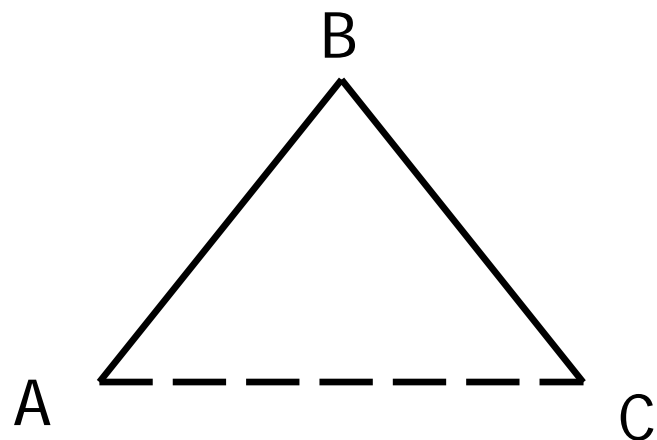
<http://ccl.pku.edu.cn/alcourse/nlp>



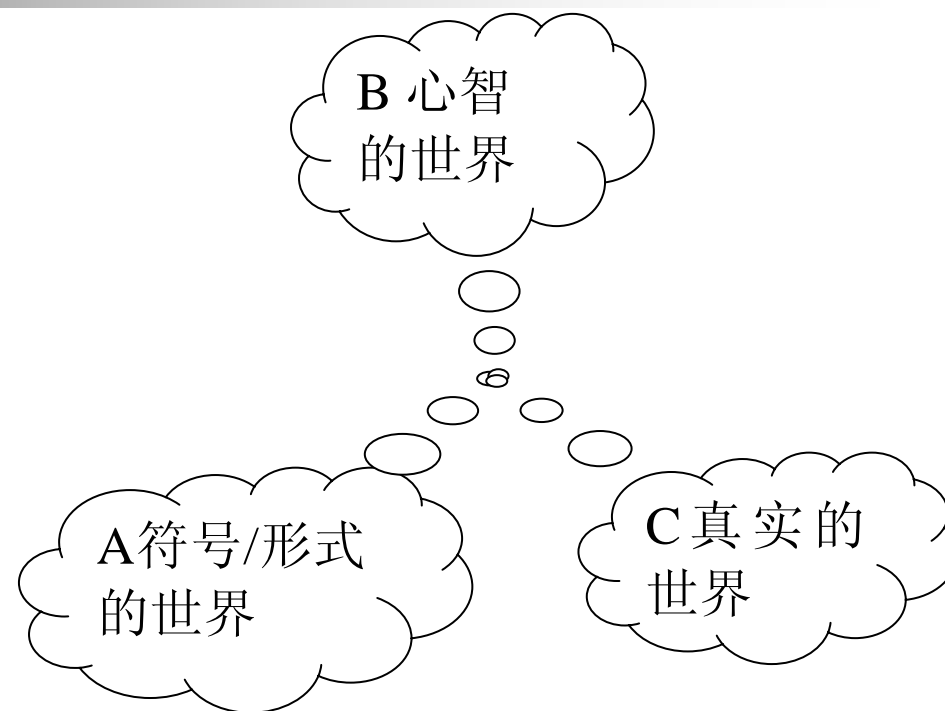
提纲

- 一 从形式到意义：引入语义知识的必要性
- 二 语义知识的类型及语义知识的获取
- 三 语义知识的应用
- 四 小结

1 从形式到意义：引入语义知识的必要性



A是对B的抽象，B是对C的抽象；
A通过B，与C发生间接的联系



转引自徐烈炯（1995）《语义学》pp.32-33，语文出版社

Ogden,C.K. & I.A.Richards, 1923, The Meaning of Meaning, Routledge & Kegan Paul, London.

在符号世界的内部看意义

意义 := 符号/形式变换

I. 在一种语言内进行的符号变换

A. 张三打了李四 \longrightarrow B. 李四被张三打了

II. 在不同语言之间进行的符号变换

A. 张三用手打了李四 \longrightarrow B. Zhang San hit Li Si with his hand

III. 在不同性质的符号系统之间进行的符号变换



\longrightarrow B. 这是残疾人通道

A 的意思是 B



Paraphrase – 解释意思

- If you paraphrase someone or paraphrase something that they have said or written, you express what they have said or written in a different way.

—— Collins COBUILD Dictionary

- para- “beside” + phrazein “to tell”, to tell in other words

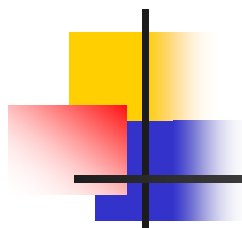
—— <http://www.etymonline.com/>



意义：符号之间无止尽的变换关系

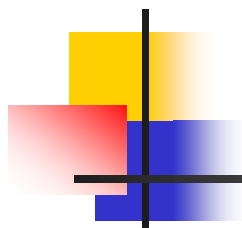
A	B
“我买了辆车”	→ 意思1：我付钱从某处购买了一辆车
	意思2：我拥有了一辆车
	意思3：我可以使用这辆车
	意思4：我的钱都花完了

A 意味着 B {B1,B2,B3,B4,...}



一个形式可以变换为其他多个形式

- | | |
|---------------------------|--------------|
| 例 1. 这个 <u>编辑</u> 很不错。 | → editor |
| 2. 他每年要 <u>编辑</u> 一百万字的书。 | → edit |
| 3. 这小伙子是干警察的好 <u>材料</u> | → 人选 |
| 4. 把这个小伙子的 <u>材料</u> 送人事部 | → 资料 |
| 5. 这是一种新型 <u>材料</u> | → 原料 |
| 6. 编辑部有许多读者。 | → 编辑部有许多人。 |
| 7. 这本书有许多读者。 | → 这本书有许多人。 ❌ |



一个形式可以变换为其他多个形式

例8

人是会死的

苏格拉底是人

苏格拉底是会死的

例8a 鲁迅的作品不适合小学生读

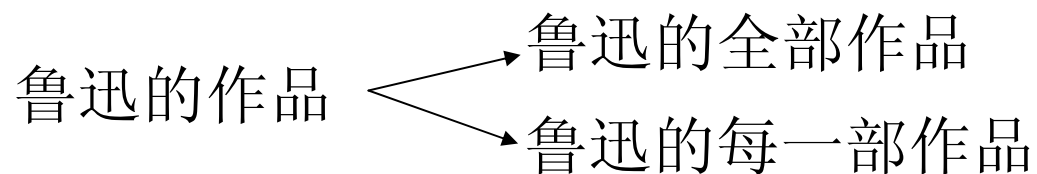
《祝福》是鲁迅的作品

《祝福》不适合小学生读

例8b 鲁迅的作品不可能在一天之内全部读完

《祝福》是鲁迅的作品

《祝福》不可能在一天之内全部读完

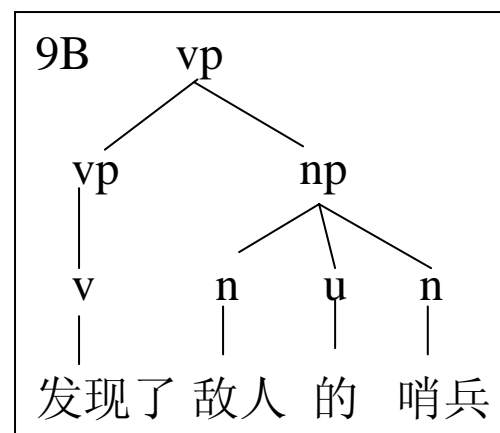
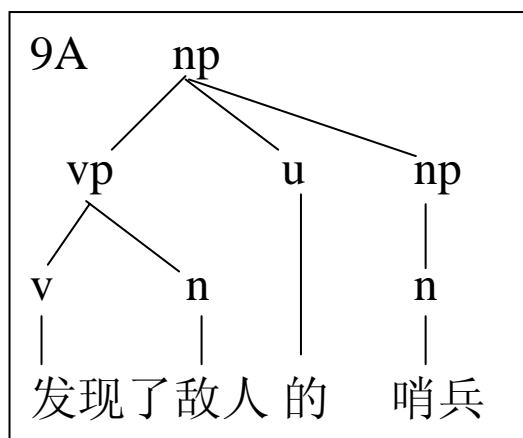


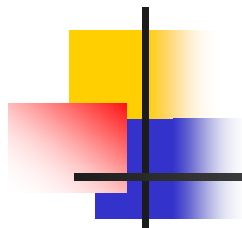


一个形式可以变换为其他多个形式

例9 发现了敌人的哨兵

vp np de np





多个符号形式可以变换为同一个形式

例1 学一食堂 供应 西餐 吗?
学一食堂 卖 西餐 吗?
学一食堂 提供 西餐 吗?
学一食堂 有 西餐 吗?
学一食堂 做 西餐 吗?
学一食堂 经营 西餐 吗?

供应(学一食堂, 西餐)

例2 学一食堂 抵制 西餐 吗?
学一食堂 反对 西餐 吗?
学一食堂 对西餐说不 吗?

¬供应(学一食堂, 西餐)



释义： 正确的形式变换

- Q: Which country is Athens in?
- A: Greece

city_table

City	Country	Population
Athens	Greece	1368
Bangkok	Thailand	1178
...

Q → SQL

Select Country From city_table Where City = "Athens"



二 语义知识的类型及知识的获取

1. 关于词义聚类关系的知识
 - 语义特征/义素描述
 - 语义分类树
 - 语义关系网
2. 关于词义组合关系的知识
 - 配价理论
 - 格语法/论元结构理论
 - 框架语义学理论

http://ccl.pku.edu.cn/doubtfire/lc_list.asp?folder=semantics
http://ccl.pku.edu.cn/doubtfire/Semantics/973_Beida/index.htm



2.1 关于词义聚类关系的知识

语义知识的类型	代表性的语义知识库/工程项目
语义特征（义素）描述	HowNet
语义分类树	905语义工程
语义关系网	HowNet WordNet / 北大ICL-CCD MindNet



2.1.1 语义特征（义素）描述

→ n个义素

词语 \ 义素	婚姻	成年	男性	人
光棍	—	+	+	+
寡妇	—	+	—	+
女童	—	—	—	+
男婴	—	—	+	+
单身汉	—	+	+	+

↓
2ⁿ个词语

语义特征（义素）描述

→ n个义素

	对象	接触	环境	工具	方式	...
炒	荤、素、面	间接	固体	
烤	荤、面、素	直接	固体	
炸	荤、面、素	间接	油	
煮	荤、面、素	间接	水	
蒸	面、荤、素	间接	汽	
炖			
...						

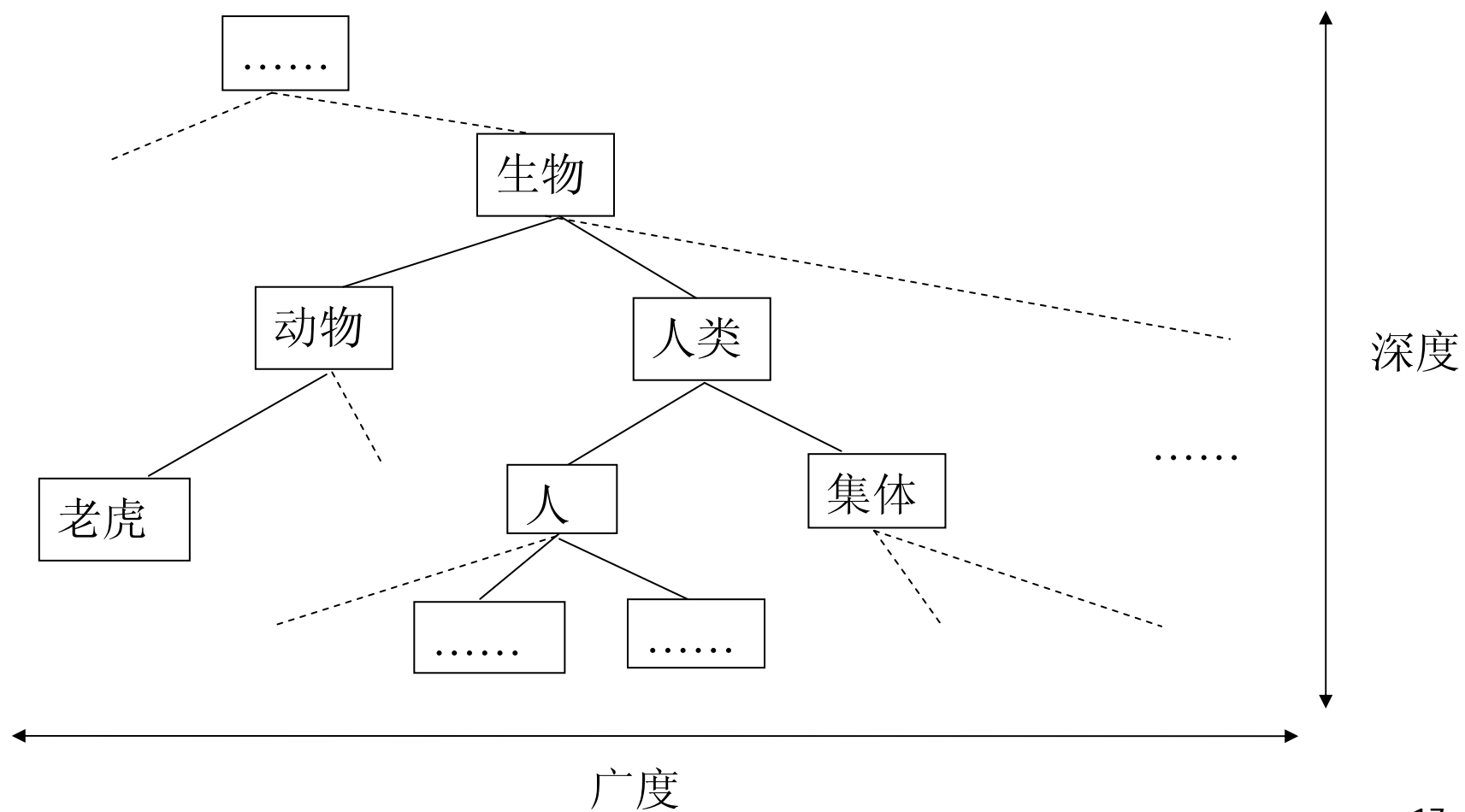
2ⁿ个词语

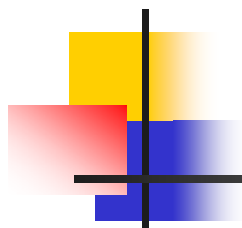


语义特征（义素）描述

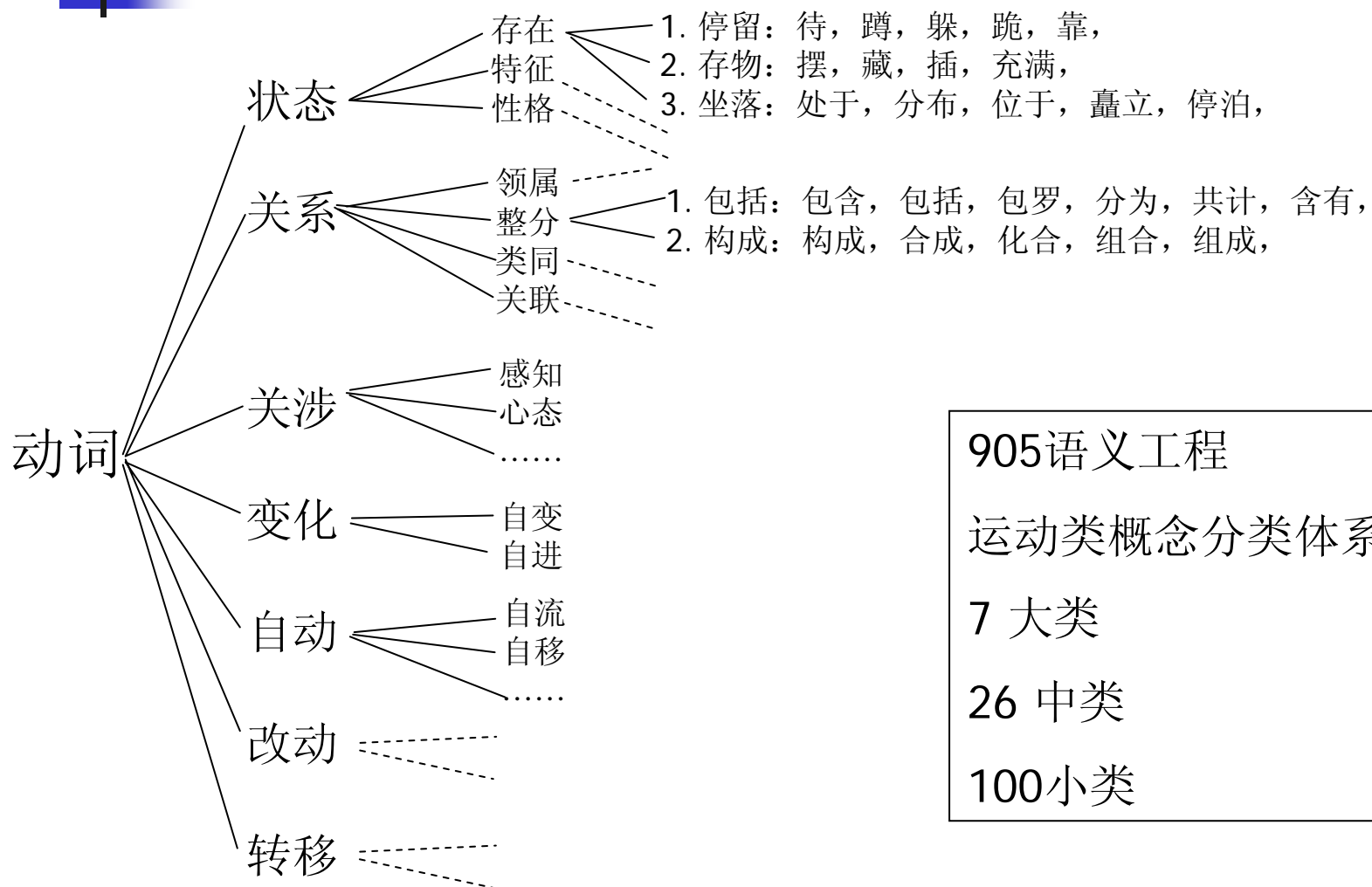
1. 语义特征的性质 $\left\{ \begin{array}{l} \text{概念义（客观义）} \\ \text{附加义（感情色彩义/主观义）} \end{array} \right.$
2. 获取语义特征（义素）的方式：
 - 2.1 参考**词典释义**，提取共同项/对立项特征；
 - 2.2 考察词语在**句法格式**中使用时的变换差异和对比差异，对具有相同/相反的句法行为的词语，提取共同项/对立项特征；

2.1.2 语义分类树





语义分类树



905语义工程

运动类概念分类体系：

7 大类

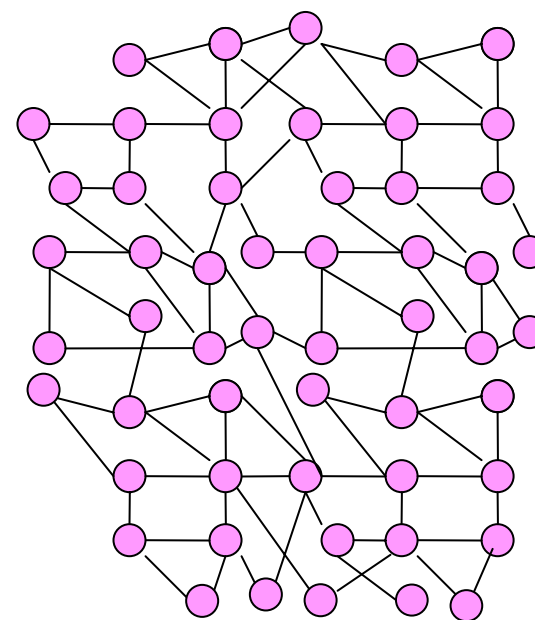
26 中类

100小类

2.1.3 语义关系网

Attribute属性	Goal目标	Possessor领有者
Cause原因	Hypernym上位	Purpose意图
Co-Agent联合施事	Location场所	Size大小
Color颜色	Manner方式	Source源点
Deep_Object深层宾语	Material材料	Subclass子类
Deep_Subject深层主语	Means方法	Synonym同义
Domain领域	Modifier修饰语	Time时间
Equivalent同位	Part部分	User使用者

MindNet





语义关系网的自动获取

car: a vehicle with 3 or usually 4 wheels and driven by a motor, esp. one for carrying people

```
car
|
|__Hyp>---- vehicle
|
|__Part>---- wheel
|
|__<Tobj ---- drive
|               |
|               |__Means>--- motor
|
|__Purp>----- carry
|               |
|               |__Tobj>--- people
```



2.2 关于词义组合关系的知识

理想目标：描述任意词语之间的组配关系

语义知识理论	代表性的语义知识库/工程项目
配价语法	ILD, 北大配价语义词典
格语法/论元结构理论	现代汉语述语动词机器词典, Propbank
框架语义学	FrameNet

http://ccl.pku.edu.cn/doubtfire/Course/Computational Linguistics/contents/SemDict_Overview_zwd20030723.pdf

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



配价语法

- 配价（valence）这一概念借自化学。目的是说明一个动词能跟多少个名词性成分发生组合关联。

游泳	[某人] 游泳	1价动词
吃	[某人] 吃 [某食物]	2价动词
送	[某甲] 送 [某乙] [某物]	3价动词



配价数

语义：动作行为跟 x 类事物有意义联系。

句法：动词周围有 x' 个空位安放跟它有意义联系的名词。

V

_____ **V**

_____ **V** _____

_____ **V** _____ _____

.....

零元（价）动词

一元（价）动词

二元（价）动词

三元（价）动词

X 元（价）动词



X 价动词

- _____ 为(向, 对) _____ V
 - 董永 总为 别人 着想
 - 大家 向 董永 看齐
 - 董永 对 七仙女 发火

- _____ V _____ VP
 - 那个犯人 企图 逃跑
 - 董永 觉得 不舒服

- _____ V _____ NP _____ VP
 - 董永 告诉 七仙女 晚上六点开会



从动词的配价到形容词、名词的配价

形 — 名 配价

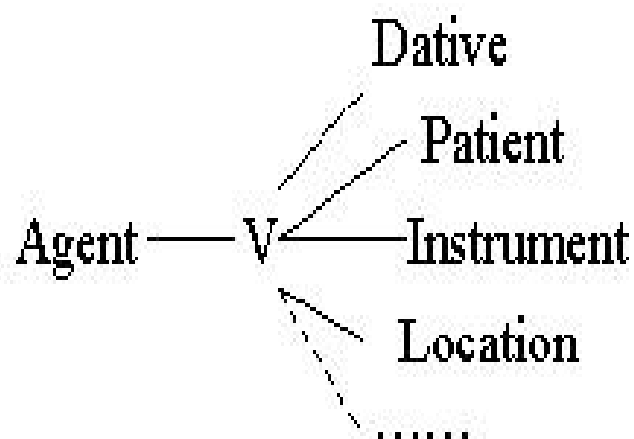
优秀	[某人] 优秀	1价形容词
友好	[某甲] 对 [某乙] 友好	2价形容词

名 — 名 配价

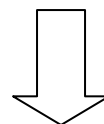
质量	[某物] 的 质量	1价名词
态度	[某甲] 对 [某乙] 的 态度	2价名词

论元类型 — 论旨角色

- 反映动词所能结合的名词的不同语义角色



主语 + 动词 + 宾语



施事 + 动作 + 受事

施事 + 动作 + 工具

处所 + 状态 + 当事

.....

— 句法模式

细化

— 语义模式



论旨角色的定义方式

■ 原型施事 (proto-typical agent)

语义特点 \ 题元	自 立	使 动	感 知	变 化
施事	+	+	+	—
感事	+	—	+	—
主事	+	—	—	+

■ 施事、感事、主事句法上的分布

句法特点 \ 题元	作基础句的主语	出现在“不 VP”之前	出现在“没有 VP”之前
施事	+	+	+
感事	+	+	—
主事	+	—	+



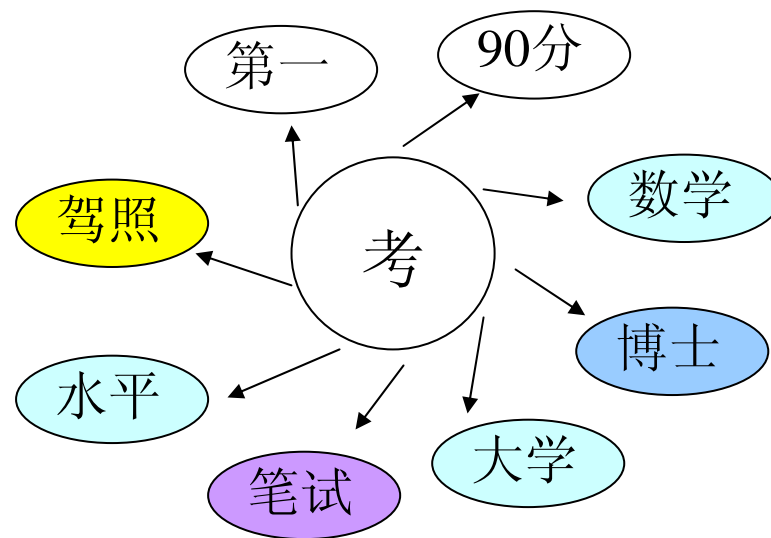
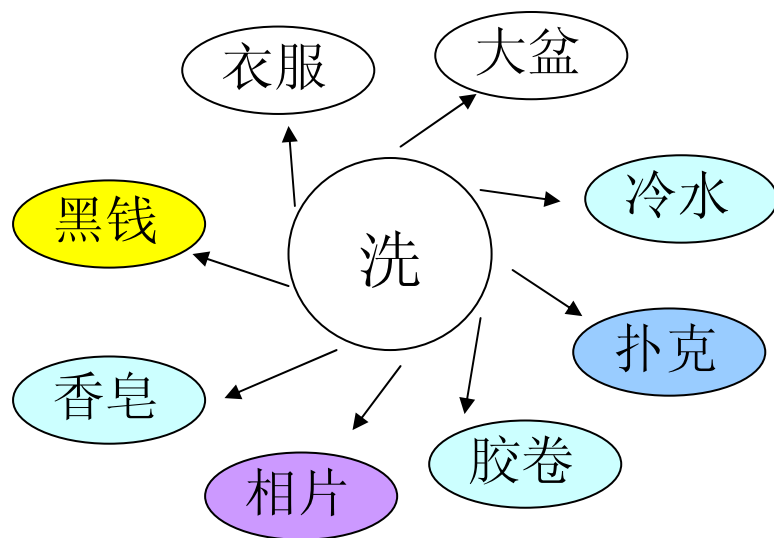
施事的句法定义示例

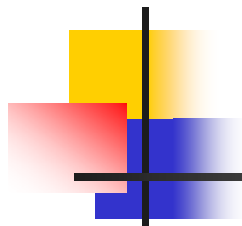
- 1) 作基础句的主语。例如“弟弟哭了”“小张吃了一碗汤面”;
- 2) 作“把”字句的主语。例如“哥哥把校长给得罪了”;
- 3) 作“被、由”等介词的宾语。例如“后勤工作由老刘负责”;
- 4) 作“V—V、V了V”等重叠形式的主语。例如“你试（一）试”;
- 5) 作“不VP”和“没有VP”等否定形式的主语。例如：“小沈不去”;
- 6) 不在宾语位置上出现，除非句首是处所性成分。

例如：“*笑了小明”;

袁毓林（2003）《一套汉语动词论元角色的语法指标》，《世界汉语教学》2003年第3期。

论旨角色的复杂性





论旨角色的复杂性

1. 我**飞**了你的马。
2. 一下午就**敲**了两百多字。
3. 今天老张干嘛去了？ — **蹬**煤去了。
4. 他这回总算**游**了一个冠军。
5. 他晚上又要来**催**稿子。
6. 他刚**拖**了地板，窗户还没有擦。

结构压缩与论元新增

原始的双事件结构				压缩后的句法结构	
V1	O1	V2	O2	V1	O2
飞	象	吃	马	飞	马
敲	键盘	输入	汉字	敲	字
蹬	三轮车	运送	煤块	蹬	煤
游	(泳)	获得	冠军	游	冠军
催	人	交	稿子	催	稿子
拖	拖把	擦洗	地板	拖	地板
参与	讨论	发表	意见	参与	意见
考	驾驶技术	获得	驾照	考	驾照
考	文化知识	进入	大学	考	大学
洗	胶片	显现出	照片	洗	照片

[+ 获得]
[+ 给予]
[+ 位移]



论旨角色占据句法位置的灵活性

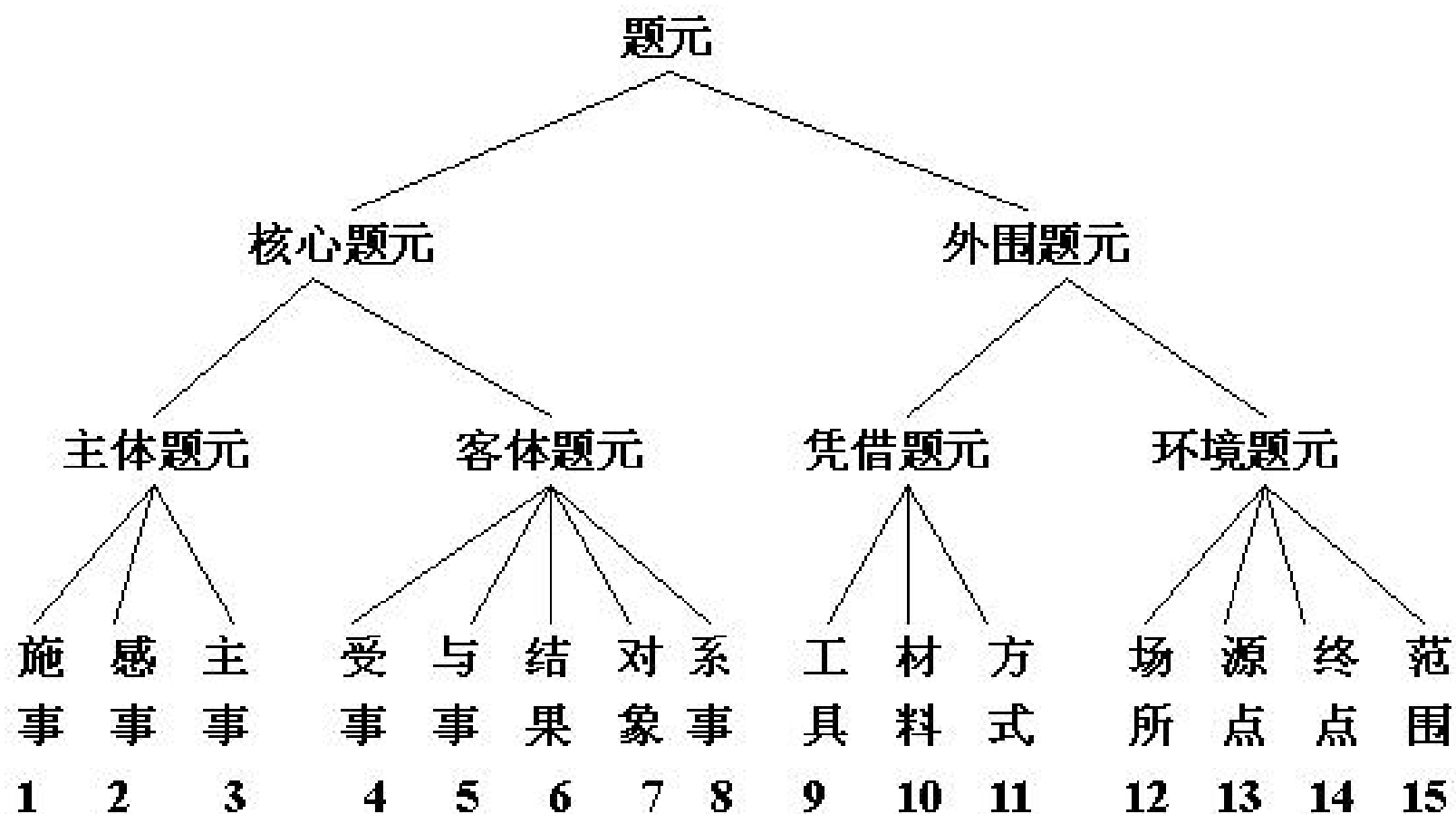
衣服 晾 阳台
便道 走 行人
前排 坐 领导

阳台 晾 衣服
行人 走 便道
领导 坐 前排

花 浇 水
红布 包 书
西红柿 炒 鸡蛋

水 浇 花
书 包 红布
鸡蛋 炒 西红柿

汉语动词的论旨角色层级系统





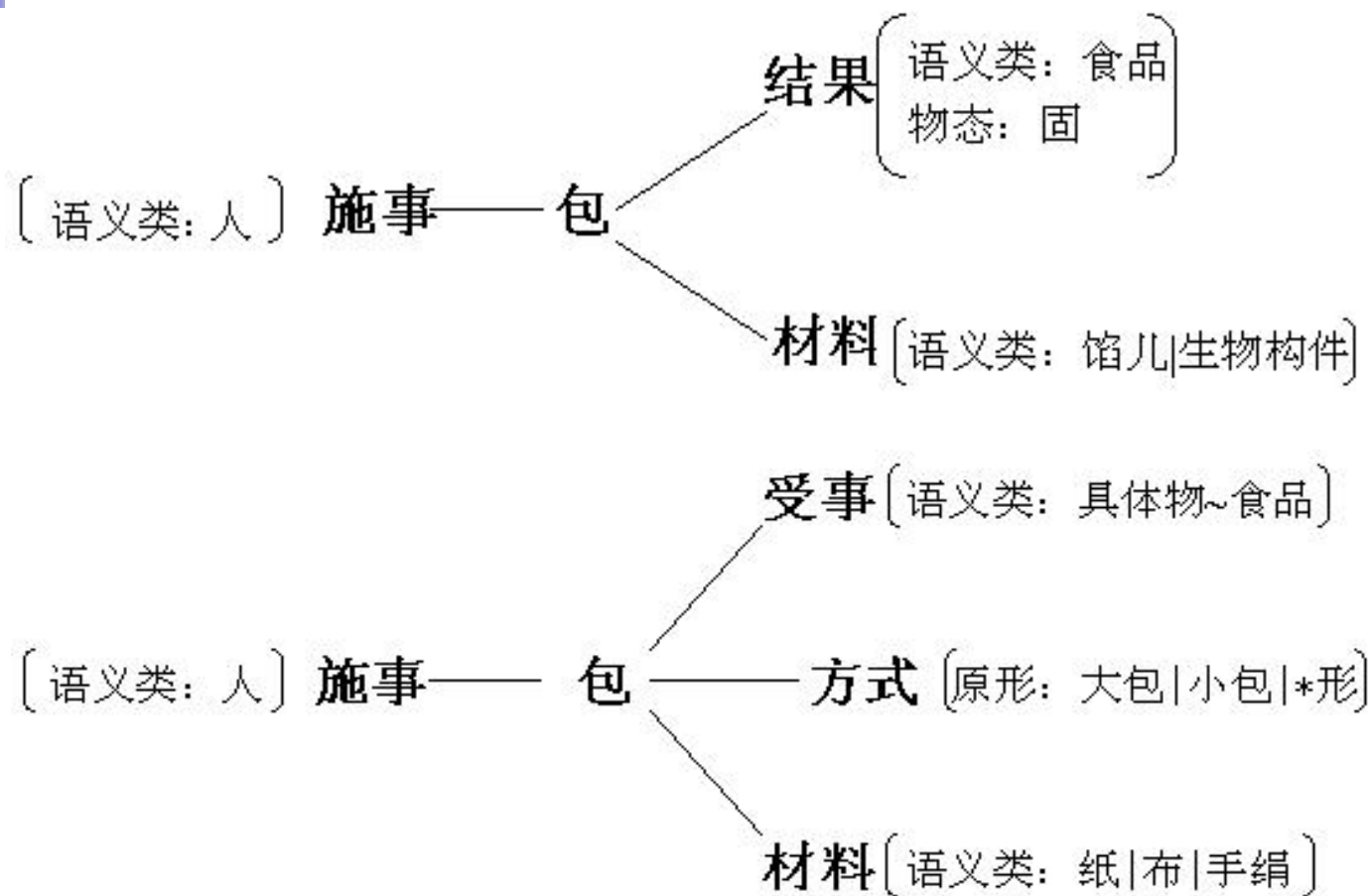
对论旨角色的选择限制(selectional restriction)

- 动词的各个论旨角色在语义上满足什么样的约束条件

吃	施事论元	[+动物]	张三吃西瓜	
	受事论元	[+固体][+食物][+药物]	* 张三吃思想 * 空气吃鱼	
想	对象论元	[+人]	想妈妈	思念
	对象论元	[+抽象物]	想答案	思索



选择限制（语义组配）





描述论旨角色选择限制的方式

- 句法属性层次
 - 商量: {当事: [...复数:Yes,...]}
- 语义属性层次
 - 吃: {施事:[语义类:人|动物],受事:[语义类:食物|药物]}
- 词例层次
 - 嗑:{施事:[语义类:人],受事:[词语:*瓜子]}
- 正面描述与反面描述
 - 变卖:{施事:[语义类:人|集体],受事:[语义类:~人]}



选择限制与特殊语境

1 否定语境

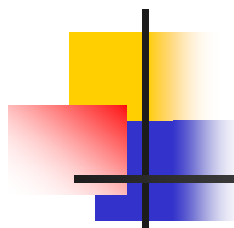
张三 不可能 吃思想
张三 怎么 可能 吃思想

2 超常现象

张三 跟李四 比赛 吃 玻璃

3 隐喻

只花了半天就吃掉了敌人两个营



小结

配价描述的内容	取值（具体的描述方式）
配价数	<ol style="list-style-type: none">1. 动词跟 x 类名词有语义联系；2. 动词周围有 x 个空位放置名词；（x 一般取值 0-3 ）
论旨角色	<ol style="list-style-type: none">1. 跟动词有语义关联的 x 类名词分属哪些类型（T_1, \dots, T_i）2. T_i 能够出现在动词周围的哪些空位上
动词对其论旨角色的选择限制	<p>充当T_i的名词要满足哪些条件？</p> <ol style="list-style-type: none">1. 语法（形式）特征 语义属性（类别）特征2. 包容性条件 排除性条件



三 语义知识的应用

- 帮助判断一个形式是否合语法；
- 帮助做句法结构分析（消歧）；
- 确定多义词的义项（Word Sense Disambiguation）；



3.1 帮助判断句子是否合乎语法

1a. 手枪比步枪更难使用。

1b. 手枪比武器更难使用。

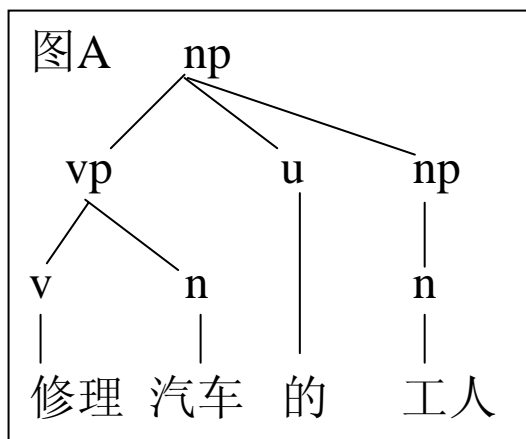
2a. 这酒不香，这酒是醇香。

2b. 这酒不是香，这酒是醇香。

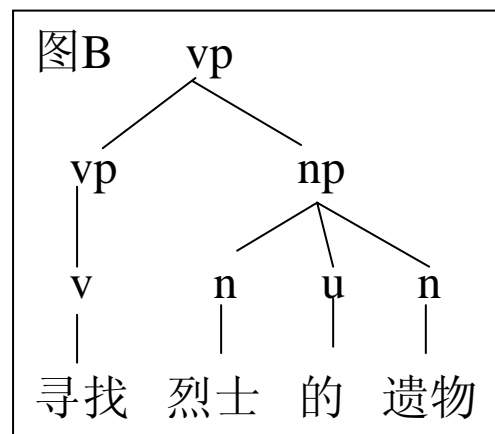
左边的例句哪些能说，哪些不能说？
有何规则？

3.2 帮助做句法结构分析

A 修理汽车的工人



B 寻找烈士的遗物



v + n + 的 + n

v + n[交通工具] + 的 + n[人]

v + n[人] + 的 + n[具体事物]

加入语义知识进行句法结构分析

修理 {[施事:人][受事:人工物]}

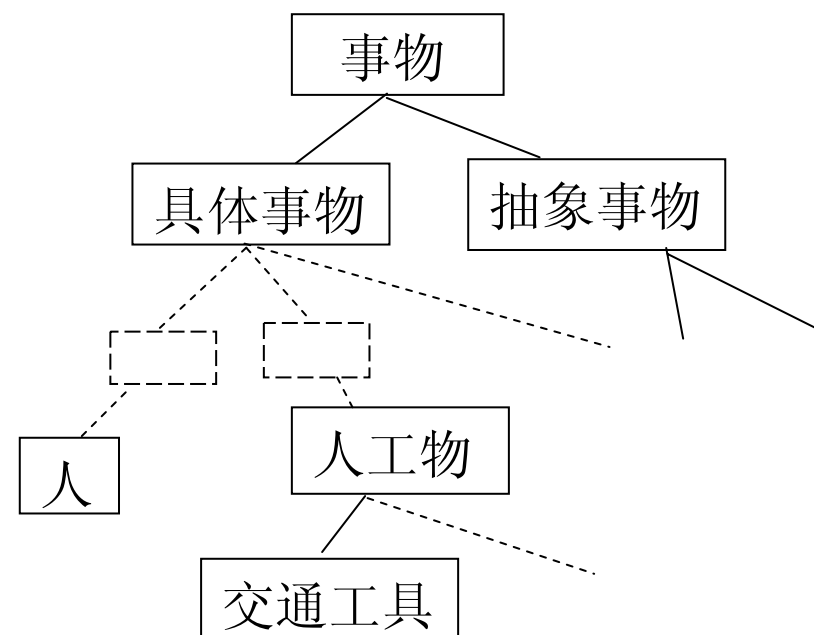
寻找 {[施事:人][受事:具体事物]}

工人 [语义类:人]

汽车 [语义类:交通工具]

遗物 [语义类:具体事物]

烈士 [语义类:人]

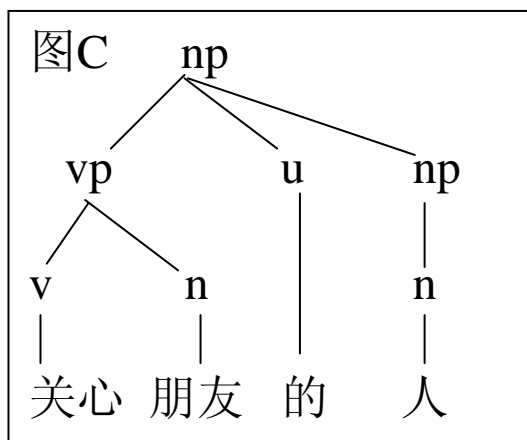


&& vp → vp np :: \$.内部结构=述宾, %vp.受事=%np

&& np → vp u<的> np :: IF %vp.内部结构=述宾 THEN %vp.施事=%np ENDIF

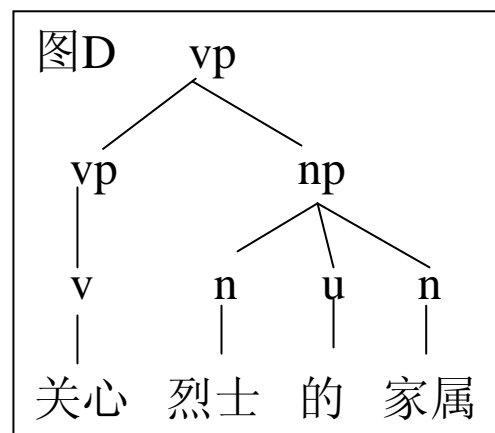
从语义类到语义特征

C. 关心 朋友 的 人



v + n + 的 + n
v + 人 + 的 + 人

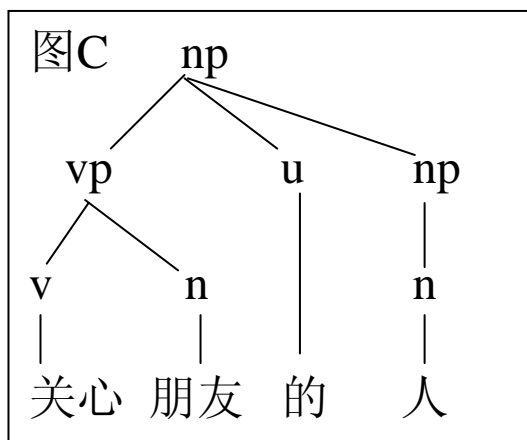
D. 关心 烈士 的 家属



v + n + 的 + n
v + 人 + 的 + 人

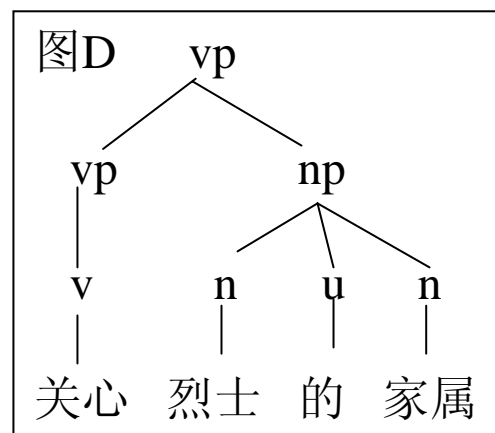
从语义类到语义特征

C. 关心 朋友 的 人



v + n + 的 + n
v + 人 + 的 + 人
[+亲属关系]

D. 关心 烈士 的 家属



v + n + 的 + n
v + 人 + 的 + 人
[+亲属关系]



从语义类到语义特征

烈士 [语义类:人, 配价数:0]

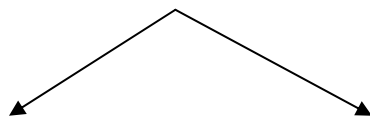
家属 [语义类:人, 语义特征:亲属关系, 配价数:1]

朋友 [语义类:人, 语义特征:亲属关系, 配价数:1]

v + n + 的 + n



v + 人 + 的 + 人



v + 人[+亲属关系] + 的 + 人

v + 人 + 的 + 人[+亲属关系]

从 词类 到 语义类 到 语义特征, 是不断对结构模式进行细化的一个过程



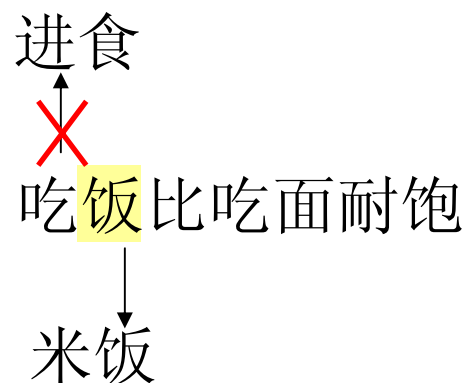
3.3 帮助确定多义词的义项

	多义词	义项	释义	英语译词
A	想主意	想 ₁	思索, 思考	think
B	想女儿	想 ₂	想念, 怀念	miss

利用V-N语义组合知识来判定动词义项



帮助确定多义词义项



吃饭: sense1 吃米饭
sense2 进食

利用语义聚合知识（上下位）来判定动词义项



四 小结

1. 语义知识跟句法知识的差别主要在于**知识颗粒度**的不同。
从某种意义上说，语义知识就是细化了的句法知识。
2. 跟语法范畴的提取是基于**聚合**和**组合**两种基本关系一样，人们也是从这两个角度出发去获取语义知识的。并且语义知识的形式化表达也可采用特征结构及合一的形式。



进一步阅读文献

- 冯志伟等译（2005）《自然语言处理综论》第14，15，16章。
- 詹卫东，2001，《确立语义范畴的原则及语义范畴的相对性》，载《世界汉语教育》2001年第2期。
- 徐烈炯，1995，《语义学》，语文出版社
- Geoffrey Leech, 1983, *Semantics: The Study of Meaning*, Penguin Books.
- 蒋严，潘海华，1998，《形式语义学引论》，中国社会科学出版社
- Shalom Lappin, ed., 1996, *The Handbook of Contemporary Semantic Theory*, Oxford: Blackwell.
- 姚天顺 等, 1995,《自然语言理解》，清华大学出版社，第四章，pp50-81,（语义网络）
- 张普, 1991,《信息处理用现代汉语语义分析的理论与方法》，载《中文信息学报》1991年第3期。
- VerbNet <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>
- FrameNet <http://framenet.icsi.berkeley.edu/>
- HowNet <http://www.keenage.com/>
- WordNet <http://wordnet.princeton.edu/>
- MindNet <http://research.microsoft.com/en-us/projects/mindnet/>



复习思考题

1. 请用配价理论分析下列动词的共性和差异。

见面 顶嘴 发火 撤职

2. 下面两个词组都是“n a 的 n”格式，但它们的句法结构不同，请给出基于语义知识的判断条件。

a. 文科优秀的教材

b. 颜色鲜艳的花朵