

# 汉语构词法概要

---

詹卫东

北京大学中文系

[zwd@pku.edu.cn](mailto:zwd@pku.edu.cn)

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

# 词的构成单位：语素

- 词是由语素构成的。语素是语言中最小的音义结合体。

morpheme

- 语素可从三个角度分类：

## 1) 语素在词中的地位

词根：表示基本词汇意义的语素

老师 挑战者 歌手 同学们

词缀：表示附加意义的语素

孩子 鞋子 杯子 椅子

词尾：表示同一个词的变化形式，构形词缀

玩儿 朵儿 份儿 桃儿

## 2) 语素的独立程度

自由语素：可以独立

粘着语素：依赖其他成分

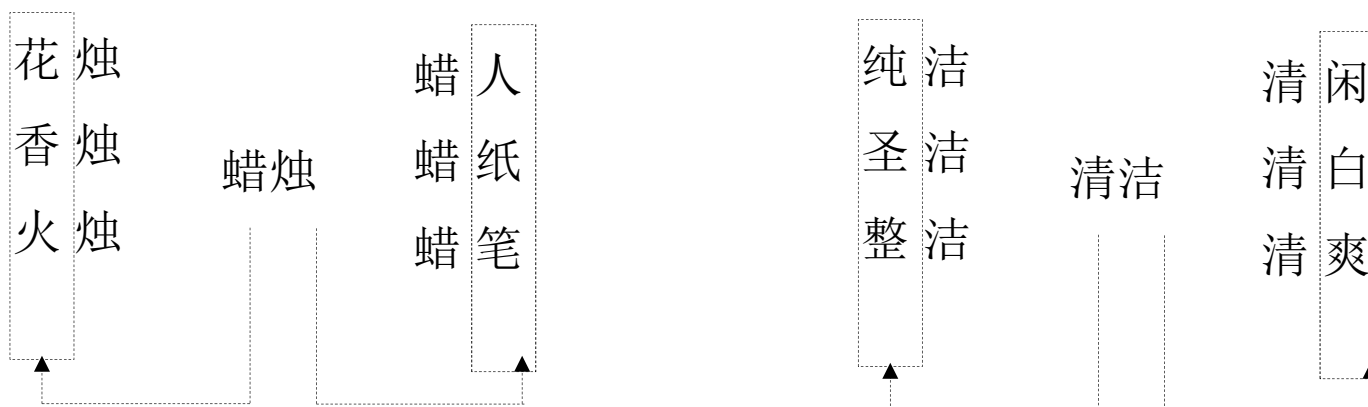
## 3) 语素的位置

定位语素：在语法结构体中位置固定

不定位语素：在语法结构体中位置不固定，可前，可后

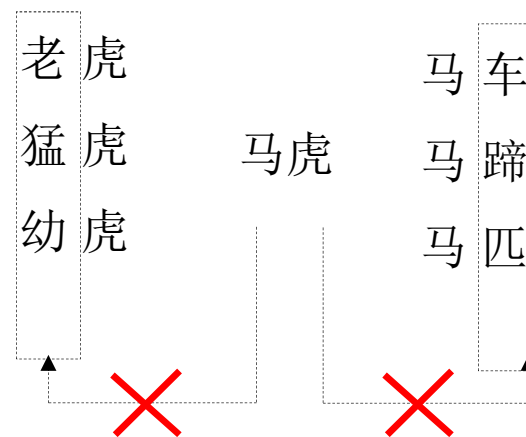
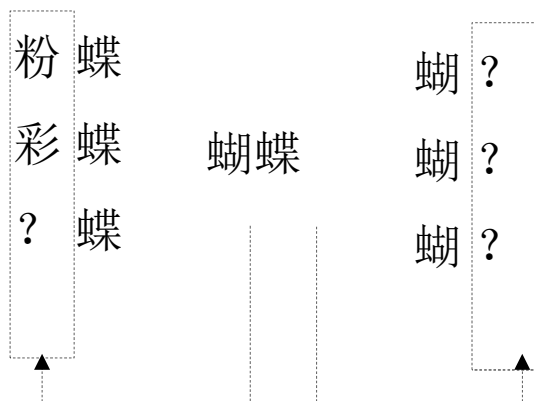
# 语素和词的关系

- 语素是语言的备用单位 vs. 词是语言的使用单位
- 语素是构词单位



用替换法来鉴定语素

# 用替换法鉴定语素



老虎 = 马糊

一个语言单位x能被替换，表示x具有一定的活动能力，可以“备用”

# 音节和语素

- 单音节语素 vs. 多音节语素

人 民 风 险 严 峻

徘徊 雷达 (radar) 盘尼西林 (penicillin)

- 英格兰 (England) → 英国 → 英
- 巴士 (bus) → 大巴 中巴 小巴  
的士 (taxi) → 面的 摩的 马的 打的 的哥 的姐
- 蝴蝶 → 蝶泳  
蚂蚁 → 工蚁 蚁兵 蚁后 蚁巢  
骆驼 → 驼毛 驼峰

有些纯粹的“音节”，由于活动能力增强，可“升格”成为“语素”。

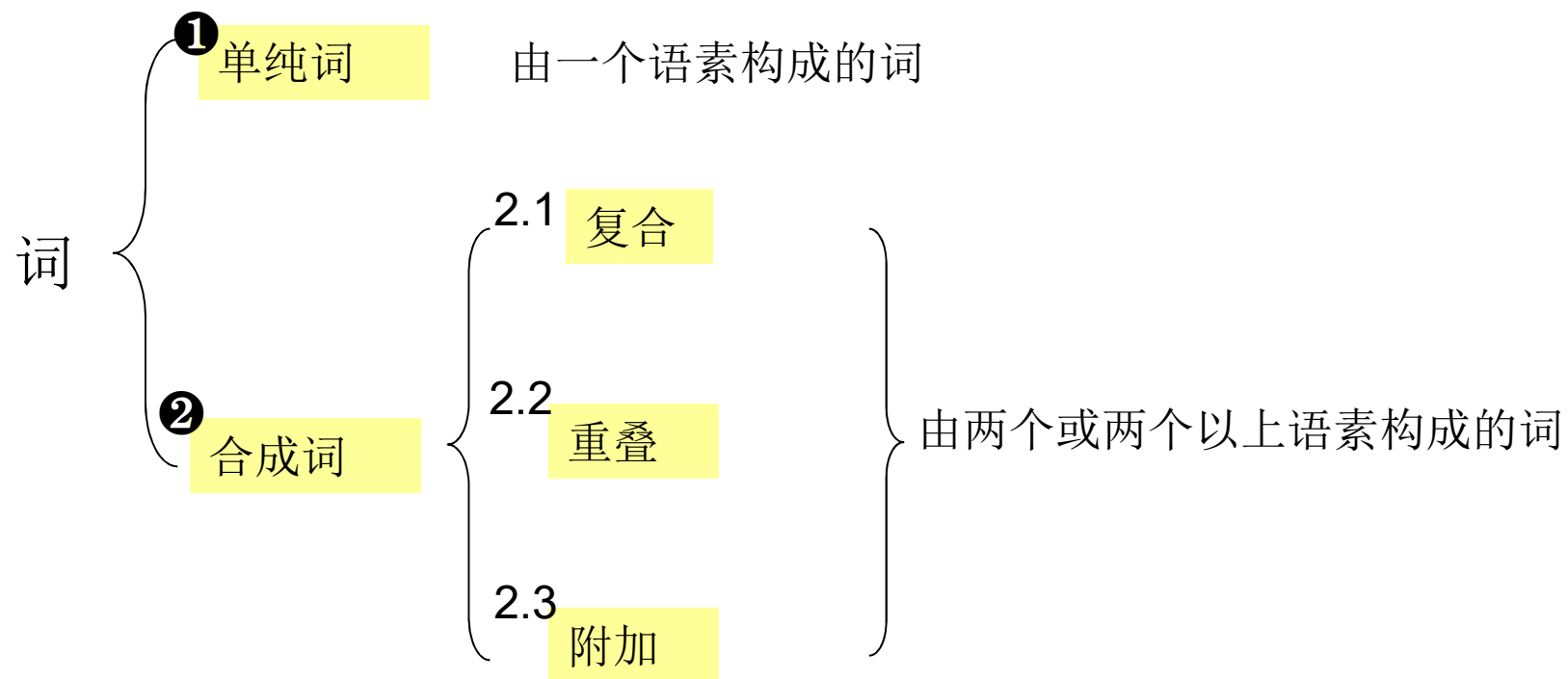
# 语素、词、词组

- 不成词语素 + 不成词语素 → 词  
民众 思索 历史 观赏
- 不成词语素 + 成词语素 → 词  
民兵 观看 朗读 高耸 瞎子 人格
- 成词语素 + 成词语素 → 词 / 词组  
马路 建设 迟到 大人  
大雨 吃饭 吃饱 站稳
- 实义语素 vs. 偏义语素  
窗户 人物 忘记 国家

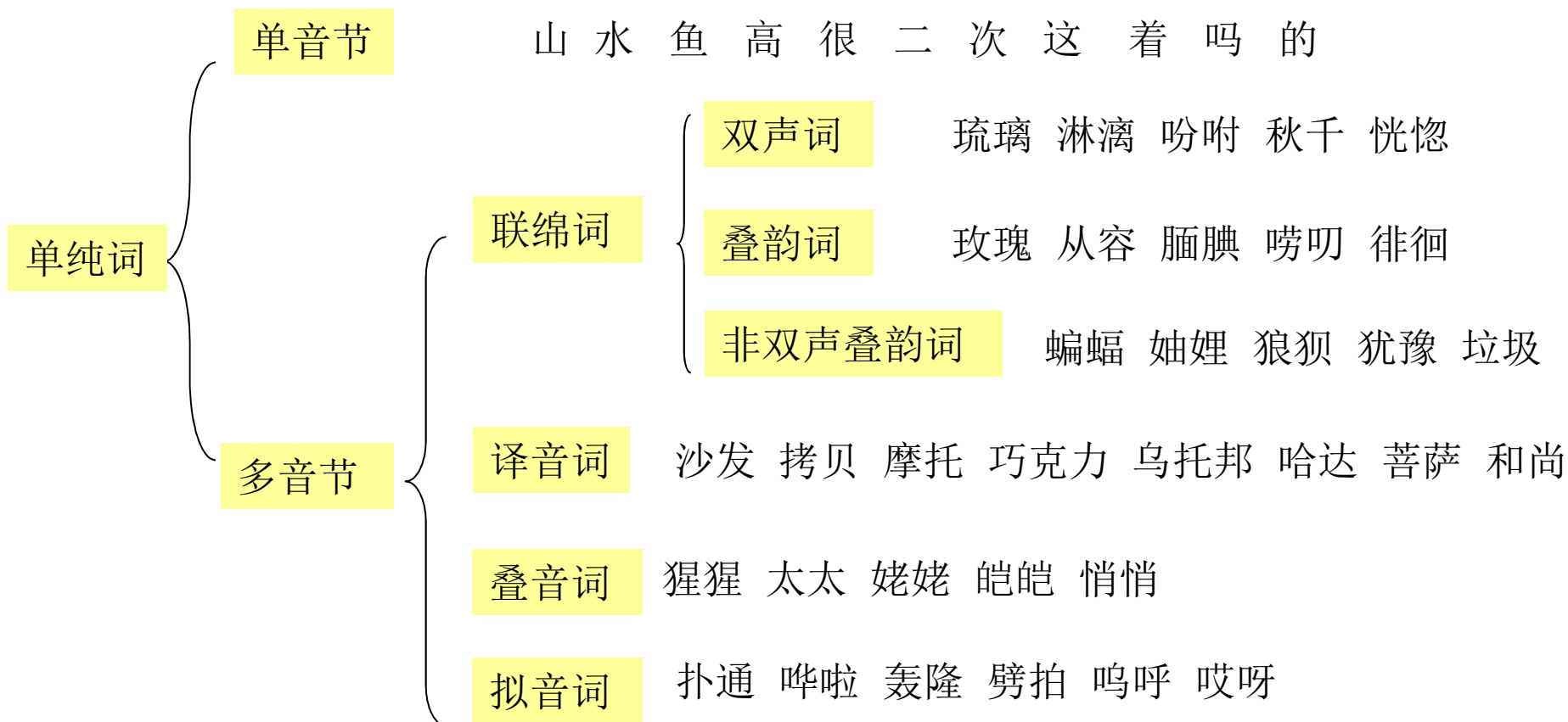
词组可扩展，内部成分易替换

词义和语素义的关系

# 词的构造类型



# 单纯词





# 合成词



# 并列式复合词

	名 + 名	动 + 动	形 + 形
同义	声音 仓库 头绪 根本	制造 休息 计算 喜欢	孤独 完全 奇怪 干净
反义	天地 矛盾 彼此 师生	裁缝 教学 开关 来往	高低 早晚 轻重 缓急
类义	江山 眉目 国家 窗户	飞跃 整理 忘记 考试	聪明 广大 细软 小巧

# 连动式复合词

V1+V2: 先后续接

查封 抽调 割让 借用 认领 领养 扮演 退休  
撤换 叫卖

例：查封报馆：  
查报馆 + 封报馆

V1+V2: 中间嫁接

逼供 请教 促进 逗笑 逗乐 遣返 引见 召见  
召集 劝退 劝降 迫降

例：对他进行逼供：  
逼他（招）供

# 偏正式复合词

前偏后正

正 偏	名	动	形
名	草帽 货车 手表 电灯	席卷 瓦解 龟缩 意译	火红 雪白 肤浅 漆黑
形	黑板 新房 红旗 平台	重视 轻信 热爱 小看	鲜红 浅黄 狂热 嫩绿
动	试卷 考场 住宅 摇篮	跳投 飞扑	飞快 透明 滚圆 喷香

名词

动词

形容词

## 偏正式复合词：名 + 动

- 比况：像N一样地V

鱼跃 蝉联 蚕食 虎视 鸟瞰 鼠窜 粉碎  
鼎立 蜂拥 狐疑 云集

- 凭借：以/用N来V

笔谈 笔试 面试 口试 言传 身教 目送  
目击 力争 力荐 法治

- 处所：在/向/从N处V

中立 空袭 左倾 右倾 上传 下载

- 时间：在N时V

春耕 秋收 晨练 午休 午睡 夜游

# 陈述式复合词

后 前	动	形
名	海啸 事变 眼花 地震 耳鸣 口吃	性急 手软 肉麻 心虚 心疼 胆怯

不同于“名+动”偏正式复合词，V说明N的性状变化

偏正式复合词：物产 口罩 木刻 球拍

NV → VN

# 支配式复合词

后 前	名	形	动
动	关心 留意 惊人 吓人 开幕 闭幕 司机 司仪	举重 入迷 失明 讲和	怀疑 抱怨 催眠 效劳

# 补充式复合词

后 前	形	动
动	改善 纠正 证明 冻僵 降低 说明	看见 打倒 撤回 展开 推翻 放开

后 前	量	名
名	船只 车辆 枪支 书本 房间 花朵	雪花 脑海 耳朵?

前主  
后次



## 真词缀 附加

前缀 + 词根	老师 老虎 老鼠 老百姓 阿姨 阿婆 阿爸 阿哥 阿妹 小丑 小偷 小姐 小青年
词根 + 后缀	桌子 椅子 骗子 矮子 凳子 刀儿 皮儿 画儿 零碎儿 破烂儿 圆乎乎 胖乎乎 傻乎乎 粘乎乎

# 类词缀 附加

类前缀 + 词根	半自动 半元音 半封建 超音速 超音质 超短波 非金属 非晶体 非卖品
词根 + 类后缀	学员 教员 会员 指挥员 电工 技工 木工 瓦工 临时工 绿化 美化 僵化 恶化 扩大化 自由主义 虚无主义 社会主义

# 重叠词

## AA

爸爸 哥哥 姐姐 妈妈 舅舅 叔叔 娃娃 爷爷  
星星  
偏偏 常常 刚刚 仅仅 单单 万万 稍稍  
好好 久久 远远 慢慢

## AABB

骂骂咧咧 婆婆妈妈 形形色色 风风火火  
兢兢业业 密密麻麻 鬼鬼祟祟 堂堂正正

一个阔人说要读经，喻的一阵一群狭人也说要读经。岂但‘读’而已矣哉，据说还可以‘救国’哩。

（鲁迅 《这个与那个》）

正如“水感”特好的人有可能成为世界级游泳运动员一样，让有“球感”的人去打球踢球，有“生意感”的人去担任厂长经理，有“新闻感”的人去当记者，“群众感”特强的人当干部，这于本人于国家于事业都大有好处。

（《文汇报》）

# 词的缩略形式

AB → A

美国→美    澳门→澳    眼睛→眼

AB → B

香港→港    历史→史    学校→校

ABCD→AC

北京大学→北大    初级中学→初中

ABCD→AD

高等学校→高校    归国华侨→归侨

ABCD→BC

香港大学→港大    人民警察→民警

ABCD→BD

香港小姐→港姐

政治协商会议 → 政协

人民代表大会 → 人大

香港出产的影片 → 港产片

人工影响天气办公室 → 人影办

ABCD→AB

清华大学→清华

ABCD→CD

人民公社→公社

紧缩

减缩

# 词的缩略形式

出入境 技战术 中西医 军烈属 离退休

高中低档 老少边穷地区 中小学

大到暴雨 小到中雪

三好 三峡 三通 三包

七情六欲 七窍 五脏六腑 五讲四美

合  
缩

缩略的基本原则：

- (1) 保留区别性特征成分      美国-> 美      美国-/-> 国
- (2) 避免缩略后成分造成混淆      澳门-/-> 门      香港-/-> 香

# 词的变形

- 喜不喜欢 相不相信 认不认识 小不小便
- 学习不学 考试不考 游泳不游
  
- 连头都不抬一下 连澡都不洗 连个屁都不敢放一个
- 帮什么忙 什么忙都帮不上
- 帮了几次忙 帮不帮得上忙
- 帮了大忙
- 他帮了我 的大忙
- 拜托你帮帮忙
  
- 大人：你这衣服袖口怎么这么脏啊？
- 孩子：幼儿园升国旗啊
- 大人：升国旗把你衣服升脏了？

傻里傻气 糊里糊涂 古里古怪 慌里慌张

王海峰，2011，《现代汉语离合词离析形式功能研究》，北京：北京大学出版社。

# 网络新词语

- |                                    |    |
|------------------------------------|----|
| □ 酱紫（这样子），造（知道），表（不要）              | 合音 |
| □ 尼玛、有木有、肿么、童鞋、妹纸、内牛满面、灰常、杯具、洗具、餐具 | 谐音 |
| □ 分特、粉丝                            | 译音 |
| □ 哥屋恩（滚）                           | 拼音 |
| □ 高富帅，累觉不爱，人艰不拆 白骨精 无知少女           | 缩略 |
| □ 减肥控、微博控、香水控 X控                   |    |
| □ 刷屏、高级黑                           |    |
| □ 二逼、屌丝                            |    |
| □ 顶、汗、囧                            |    |
| □ .....                            |    |



## 网络新词语（续）：注音心声体/虚伪体

例1：这种有（qí）趣（guài）的「注音心声体」最初是怎么来的？（知乎问答）

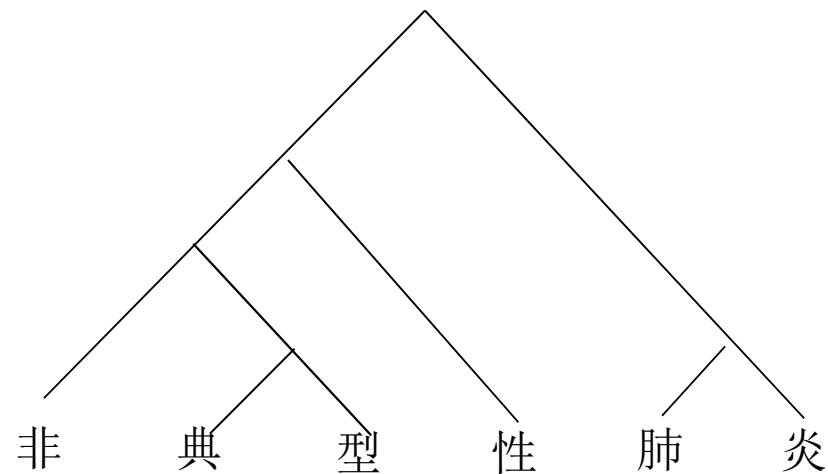
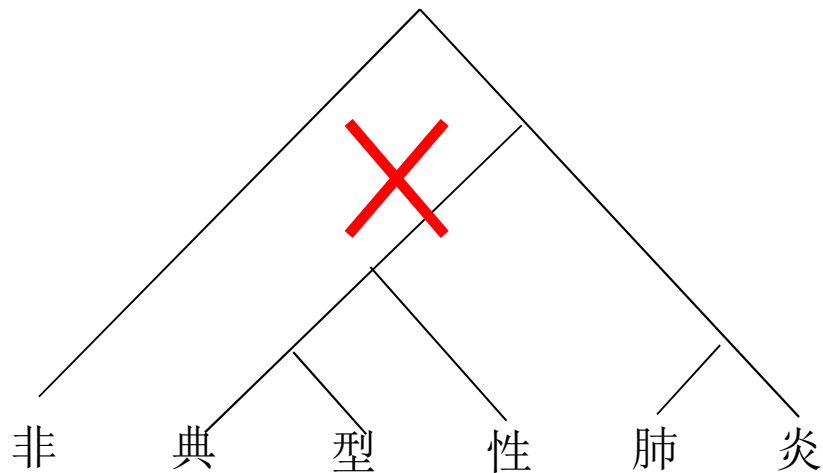
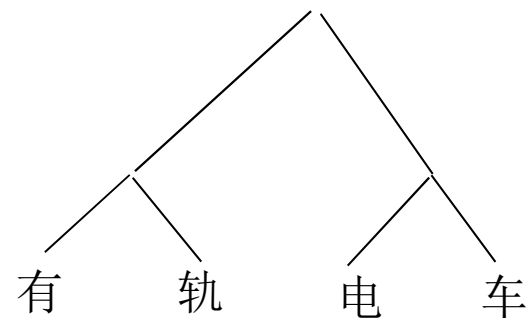
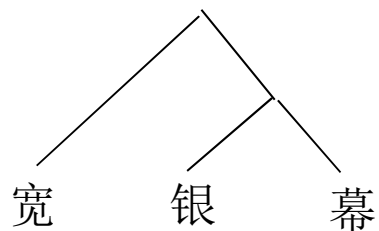
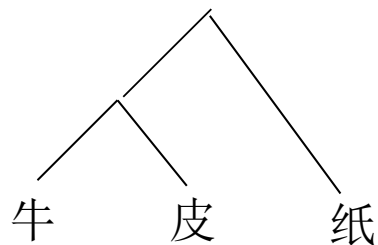
例2：A：我做的菜好吃吗？  
B：嗯，厨（tai）艺（nan）不（chi）错（le）。

例3：迎奥运，树（tree）新（new）风（bee）

例4：A: Do I act like the Big Knowledge Woman?  
B: No.  
A: Is that "no" spelled y-e-s?  
B: S-o-r-t of.

# 词的内部层次

语素 → 语素组 → 词



# 构词法知识的应用

- 郑家恒、李文花，2002，基于构词法的网络新词自动识别初探，《山西大学学报》2002年第2期，115-119页。
- 徐艳华，2011，面向自动分词的三音节新词语构词法研究，孙茂松、陈群秀主编《中国计算语言学研究前沿进展（2009-2011）》（第十一届全国计算语言学学术会议CCL-2011会议论文集），清华大学出版社2011年版，58-63页。
- 邹刚、刘洋、刘群、孟遥、于浩、西野文人、亢世勇，2004，面向Internet的中文新词语检测，《中文信息学报》2004年第6期，1-9页。
- Jian-Yun Nie, Marie-Louise Hannan, Wanying Jin, 1995, Unknown Word Detection and Segmentation of Chinese using Statistical and heuristic Knowledge, Communication of COLIPS 1995; 5(1-2): pp. 47-57.
- Tao Liu, Bing-Quan Liu, Xiao-Long Wang, Ming-Hui Li, 2007, The Effectiveness Study of Local Maximum Feature for Chinese Unknown Word Identification, Journal of Chinese Language and Computing 17 (1): 15-26.

① 自动发现文本中的新词，辅助编纂新词语词典

② 自动分词系统中识别未登录词

# 基于构词法的一些新词过滤规则示例

1.  $[a-z]^*d$  : 所有以副词结尾的词性序列;
2.  $u[a-z]^*$  : 所有以助词开头的词性序列;
3.  $[a-z]^*u$  : 所有以助词结尾的词性序列;
4.  $[a-z]^*c[a-z]^*$  : 所有包含连词的词性序列。
5.  $q[a-z]^*$  : 所有以量词起始的词性序列。
6.  $[m]^+$  : 单纯的数字;
7.  $[t]^+$  : 所有的日期;

邹刚 等 (2004)

例: “岁/qt 就读/vi 于/p”

# 基于构词法的一些新词过滤规则示例

郑家恒、李文花（2002）

- 常规构词规则     if wordtype (A B ) = 'N 'and wordtype (C) = 'N 'then new (A B C) = 'N'  
如果汉字串ABC的构成形式为A B + C 型，且满足A B 、C 分别为名词，  
则将A B C做标记，认为是新词。
- 特殊构词规则     if firstwordtype (A) = 'N ' and secondword (B) = '吧' then new (A B) = 'N'  
若二元组中的第二个字是“吧”，且第一个单字词为名词，则认为该二元组为新词。
- “互斥性字符串”  
过滤规则     if wordtype (A) = 'P 'and wordtype (B) < > 'N ' then delword  
若A 为介词，B 为非名词类，则将字符串“A B”去除。