

生成词库理论概要

詹卫东

北京大学中文系

zwd@pku.edu.cn

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

提纲

- 1 基本思想
- 2 理论特点
- 3 分析实例

1 基本思想

生成词库是一种成分组合语义学，目标是对词语的多义、意义模糊、意义不明确、意义变化等做出有效的描写和解释，特别是对语言使用中词义的有规则的多义（regular polysemy）做出合理的分析。Pustejovsky把这种多义现象称为“逻辑性多义”（logical polysemy），指一个词在不同语境中意思不完全相同，但又明显关联，基本意义相同，关联方式有一定的模式。

生成词库理论的理想是对新语境中的词语的创造性用法做出解释（the creative uses of word in novel contexts, Pustejovsky, 1995）。

多义词 与 传统词义描写体系的问题

- | | | | |
|---|--------------------------|-----------|---------------|
| 1 | a. a fast typist | 一个很快的打字员 | 一个打字速度很快的打字员 |
| | b. a fast car | 一辆很快的轿车 | 一辆运行速度很快的轿车 |
| | c. a fast highway | 一条很快的高速公路 | 一条车流速度很快的高速公路 |
| | | | |

- | | | |
|---|-----------------------------------|---------|
| 2 | a. Mary baked the potatoes | 玛丽烤了马铃薯 |
| | b. Mary baked a cake | 玛丽烤了蛋糕 |

- | | | |
|---|---------------------------------------|----------|
| 3 | a. John began to read the book | 玛丽开始写这本书 |
| | b. John began the book. | 玛丽开始这本书 |

- fast, bake, began 是否需要立多个义项?
- 如何解释 fast, bake, began在不同语境中的意义（用法）差异?

生成词库论的基本想法

- 不采取词义分解 (decomposition) 的做法
- 相反, 重视 (1) 词语跟其他词语组合时表现的意义
(2) 词语在不同组合环境中意义的变化

试图通过对词语的基本意义结构的有效组织, 来预测它可能的语境 (临时/新) 义

2 理论体系

- a. LEXICAL TYPING STRUCTURE 一个词语在语言词汇类型系统中的位置
词汇类型结构

- b. ARGUMENT STRUCTURE 谓词的论元数量、性质
论元结构

- c. EVENT STRUCTURE 语言表达的事件类型以及可能包括的子事件结构
事件结构

- d. QUALIA STRUCTURE 描述一个词项（主要是名词）的述谓意义的结构，包括该词项所指对象的构成情况、来源、用途和功能等
物性结构

物性结构

(1) 构成角色
(constitutive role)
CONST

描写一个物体与其组成部分之间的关系，包括材料 (material)、重量(weight)、组成部分与组成成分等；也指物体在一个更大的范围内构成或组成其他物体；

(2) 形式角色
(formal role)
FORMAL

描写对象在更大的认知域内区别于其他对象的属性，包括大小 (magnitude)、形状 (shape)、维度 (dimensionality)、方位 (orientation) 和颜色 (color) 等

(3) 功用角色
(telic role)
TELIC

描写对象的用途 (purpose) 和功能 (function)。主要包括两种，一种是直接功用角色 (direct telic)，如beer的功用角色是drink；另一种是间接功用角色 (purpose telic)，如knife的功用角色是cut

(4) 施成角色
(agentive role)
AGENTIVE

描写对象怎样形成或产生的，即来源。比如，“饺子”的施成角色是“包”这个动作。

物性结构示例

book

ARGSTR =

ARG1 = x: info

ARG2 = y: physobj

QUALIA =

info·physobj _lcp

FORMAL = hold (y, x)

TELIC = read (e, w, x. y)

AGENT = write (e, v, x. y)

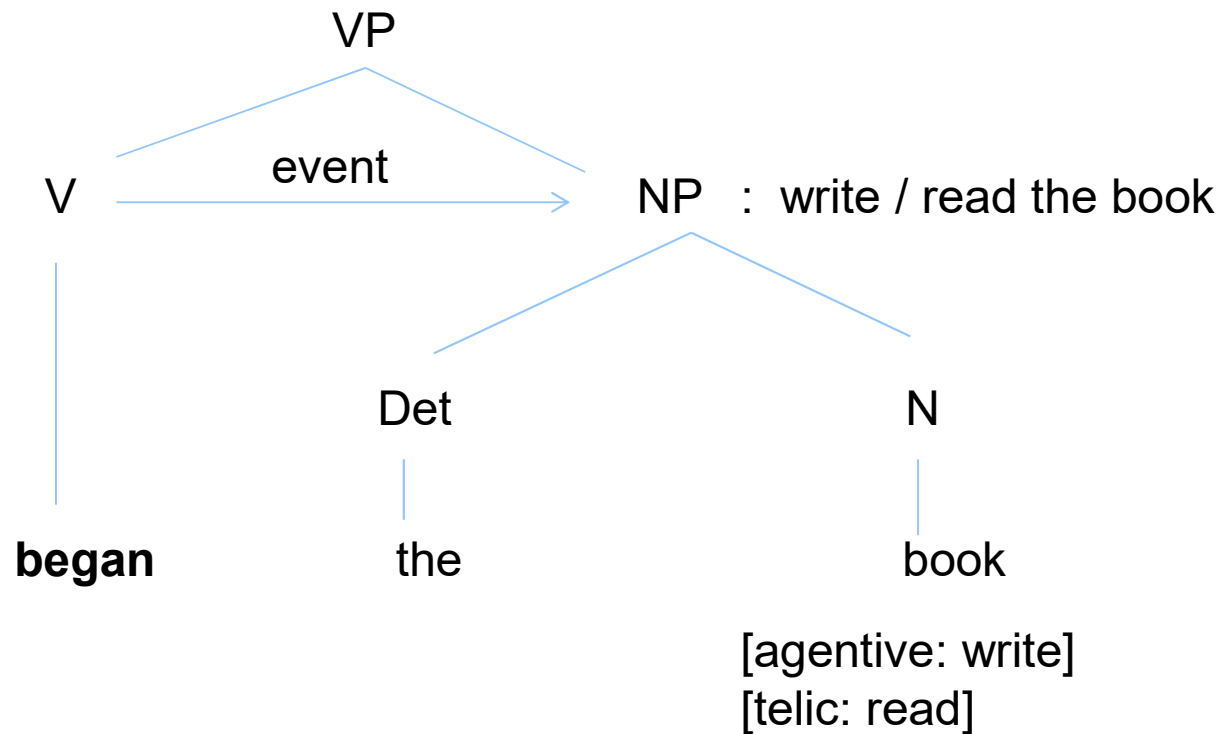
- (1) **单位**
unit, UNI
名词的计量单位, 即跟名词组配的量词
- (2) **评价**
evaluation, EVA
对名词所指事物的评价。如对“战士”的评价有“伟大、出色、聪明、笨拙、……”
- (3) **行为**
action, ACT
名词所指事物惯常发生的动作、行为。如“老师”的行为是“教书、上课、写作、研究……”等等
- (4) **处置**
handle, HAN
名词所指事物惯常受到的动作、行为。如“学生”的处置是“录取、表扬、批评、处罚、开除……”等等
- (5) **材料**
material, MAT
名词所指事物生产时所用的材料。如“椅子”的材料是“木头、竹子、钢铁、石头、塑料、……”等等
- (6) **定位**
orientation, ORI
名词所指事物在表达空间或时间时, 用什么介词来确定其与其他人或事物之间的时空关系性质。如“学校”的“定位”是“在、从、到、向……”等等。

物性结构的意义

- Pustejovsky设计的语义知识表示框架中，物性结构，或称为物性角色描写体系占有很重要的地位。可以理解为是语言学知识和非语言学知识的表征接口。物性结构是一套关于词汇本体知识（ontological lexical knowledge）的描述体系，是生成词库理论中最为核心的内容。
- 借助物性角色知识，可以对语言的创造性使用，尤其是对词项在不同的语言环境下浮现出（emerge）不同的意义这一动态性词义变化现象进行说明和解释。
- 物性结构是静态的知识描述。要对词项的语境义进行解释（生成），还需要“强迫机制”（Coercion mechanism），即“类型转换操作”（type shift operation）来实现。

事件强迫示例

- a. John **began** to read the book
- b. John **began** the book.



3 实例分析

- 补充结构中的隐形成分，帮助得到句法结构的语义解释

赶 论文 → 赶写论文

木头 桌子 → 木头 做的 桌子

托福热 → 考托福 热

汉语热 → 学习汉语 热

- 帮助分析句法结构合语法性的条件

他打篮球 打 得 好 → 他的篮球打得好

他打篮球 打 得 穷困潦倒 → * 他的篮球打得穷困潦倒

他当老师 当 得 不错 → 他的老师 当得不错

他当老师 当 得 很灰心 → * 他的老师 当得很灰心

宋作艳（2015）

袁毓林（2014）

困难的问题

体操 奶奶

成人 英语

大学 英语

少儿 英语

哑巴 英语

大嘴 英语

职称 英语

专业 英语

广告 英语

科技 英语

学术 英语

会计 英语

导游 英语

华尔街 英语

.....

赶 火车 → 赶乘火车

赶 出租车 → ?

赶 地铁

赶 末班地铁

赶 早班车

赶 午饭 → ?

赶 场子

参考文献

- Pustejovsky, James (1991) The Generative Lexicon. Computational linguistics (4) : 409 - 441.
- Pustejovsky, James (1995) The Generative Lexicon. Mass: MIT Press.
- Pustejovsky, James (2011) Coercion in a general theory of argument selection , Linguistics 49 - 6 , 1401-1431.
- Pustejovsky J., P. Bouillon, H. Isahara, K. Kanzaki & C. Lee (eds.) (2013) Advances in Generative Lexicon Theory. Dordrecht: Springer.
- Pierrette Bouillon, Federica Busa, 2001, *The Language of Word Meaning*, Cambridge University Press.
- 袁毓林 (2013) 基于生成词库论和论元结构理论的语义知识体系研究 , 《中文信息学报》第6期 , 第23 - 30页。
- 袁毓林 (2014) 汉语名词物性结构的描写体系和运用案例 , 《当代语言学》第1期 , 第31 - 48页。
- 宋作艳 (2015) 《生成词库理论与汉语事件强迫现象研究》 , 北京大学出版社2015年版。