

文章编号: 1003-0077 (2011) 00-0000-00

## 机器词典释义模版的建构过程和运用案例\*

王恩旭<sup>1,2</sup>, 袁毓林<sup>1</sup>

(1. 北京大学, 北京 100871; 2. 曲阜师范大学, 山东 曲阜 273165)

**提要:** 机器理解词主要借助于词典, 但目前的词典释义还不准确, 也不完备。对于这些问题, 通过分析词的语义结构、建构词的释义模版来解决。通过分析词的语义结构, 弄清词义中包含着哪些语义成分、语义关系, 确定哪些是必有成分、必有关系, 哪些是可有成分、可有关系。然后, 结合实例讨论释义模版的建构过程、原则与方法。最后, 通过释义模版, 解决词典释义不完备、语义联系不明显(以多义词为例)、循环释义(以同义词为例)、新词释义(以新造合成词为例)等问题。

**关键词:** 生成词库论; 释义模版; 多义词; 同义词; 新词

中图分类号: TP391

文献标识码: A

## The construction process and application case of Chinese words interpretation template

Wang Enxu<sup>1,2</sup>, Yuan Yulin<sup>1</sup>

(1. Peking University, Beijing 100871, China ; 2. Qufu Normal University, Shandong, Qufu 100871, China)

**Abstract:** machine understanding words are mainly based on dictionaries, but the present dictionary interpretation is inaccurate and imperfect. this paper tries to solve these problems by the way of analyzing the words semantic structure and constructing the words interpretation templates. By analyzing the words semantic structure, we should find out what semantic components and semantic relations that words contain, and determine which of them are necessary and which of them are unnecessary. Then, with examples, this paper discusses the process, principles and methods of constructing interpretation templates. Last, this paper shows that constructing the interpretation template is helpful to solve the following problems: the polysemous words interpretation problem, the synonym words interpretation problem, the new words interpretation problem, etc.

**Key words:** Generative Lexicon Theory; interpretation templates; polysemy; synonyms; new words

### 1 引言

理解词义是理解句义的前提与基础。没有对词义理解的准确和完备, 就谈不到对句义理解的准确和完备。从这个意义上说, 研究并解决词的释义问题意义重大。

为了提高机器理解词语的准确率, 国内外学者做了大量的工作, 形成了一批有代表性的资源或词典, 如 WordNet、VerbNet、FrameNet、ConceptNet、HowNet、《现代汉语语义词典》等。总的来说, 成绩是显著的, 通过研究, 提高了词义消歧、信息检索、机器翻译的准确率。但缺欠也是明显的, 比如, WordNet 忽视了论元语义角色信息、Verbnet 忽视了论元名词间的关系信息、ConceptNet 忽视了词语间的组合关系信息等<sup>[1]</sup>。缺欠的存在, 限制了这些词典或资源在自然语言处理中的作用。

机器理解词义主要借助于词典, 但目前的词典释义还不准确, 也不完备。突出地表现在: 1) 文献<sup>[2]</sup>考察词典对名词的释义时发现, 释义不完备的情况随处可见。比如“书”“唱片”等名词本应包含“物质”“信息”两个义位, 但词典只收录“物质”一个义位, 漏掉了

\* 收稿日期: 定稿日期:

**基金项目:** 教育部人文社科基地重大项目“汉语意合语法框架下的词汇语义知识表示及其计算系统研究”; 973 计划项目(2014CB340502); 中国博士后科学基金项目(2016M600004); 山东省社科基金项目(16CZJW31); 国家语委科研项目“现代汉语词语释义的原则与方法研究”

**作者简介:** 王恩旭(1978—), 男, 博士后, 主要研究方向为词汇语义学、计算语言学; 袁毓林(1962—), 男, 博士, 教授, 博士生导师, 主要研究方向为理论语言学 and 汉语语言学、计算语言学和中文信息处理。

“信息”义位。2) 文献<sup>[3]</sup>调查了《现代汉语词典》(2005版, 简称《现汉》)、《现代汉语规范词典》(2004版, 简称《规范》)、《新华字典》(2001版)等权威辞书后发现, 辞书中“本义-引申义”的联系普遍不够明显; 通过辞书, 读者不能自然地本义推出引申义。3) 在当前的词典释义中, 循环释义的情况还比较常见。比如“漂亮”, 《现汉》(2012版)的释义就是循环的:

【漂亮】好看; 美观。

【好看】看着舒服; 美观; 精彩。

【美观】好看; 漂亮。

而且, 据文献<sup>[4]</sup>对《现汉》(2012版)的考察, 2094个双音复合词中循环释义的有246个, 占总量的11.7%。循环释义或同义词释义的方式, 对以汉语为母语的人来说, 问题不大, 但对机器或外国学习者来说, 却容易出问题。因为, 任何两个词的意义都不完全对等。比如“漂亮”, 可用于人、物, 也可用于事; 而“好看”只用于人、物, 不用于事; “美观”只用于物, 不用于人或事(不说“这个人/这件事很美观”)。此时, 用一个同义词来解释另一个同义词, 就容易让读者忽略同义词之间的差别, 造成词语使用上的混乱。4) 不论词典的规模有多大、更新的速度有多快, 通常都无法穷尽所有的新词, 总有一些是机器无法识别的, 而这些新词往往就是机器理解词的主要障碍。<sup>1</sup>

本文的目标: 通过分析词的语义结构、构建词的释义模版, 解决多义词、同义词、新造合成词等释义问题, 为自然语言处理及机器词典编纂服务。

## 2 分析词的语义结构

分析词的语义结构, 本文采用生成词库论的方法。<sup>2</sup>根据文献<sup>[5]</sup>, 词的概念结构中可能包括以下几方面的信息:

(1) **语义类型 (TYPESTR)**: 每个词都属于一个特定的语义类。确定了一个词属于哪个语义类, 也就确定了它和哪类词有蕴涵关系或语义继承关系。

(2) **论元结构 (ARGSTR)**: 确定谓词的论元性质、数量及其句法实现情况。

(3) **事件结构 (EVESTR)**: 确定表达式的事件类型、事件成分及事件关系。

(4) **物性结构 (QUALIA)**: 包括一个词的表现形式是什么 (Formal)、由什么构成 (Constitutive)、有哪些功能 (Telic)、怎样产生 (Agentive) 四种物性角色。

词类不同, 词的语义结构分析方法往往便不同。分析名词的语义结构, 适合从物性结构入手。和其他词类相比, 名词的语义信息更为复杂, 除了文献<sup>[5]</sup>的四种物性角色外, 还可能包含材料、单位、处置、行为、单位、定位六种物性角色<sup>[1]</sup>。分析名词的语义结构, 先要弄清名词包含了哪些物性信息; 然后再分析它的语义成分和语义关系。以“椅子(有腿有靠背的坐具)”为例, 分析它的物性结构, 可以得到以下信息:

语义类信息: 人造物; 家具类

物性结构信息: 功用角色(供人坐); 构成角色(由靠背、四条腿等构成)

在此基础上, 确定“椅子”的词义成分和词义关系:

词义成分: 3个(人/家具/靠背和四条腿)。

词义关系: 3种。一种是**语义类关系**, “家具”和“椅子”是上下位关系; 一种是**功用**

<sup>1</sup> 以分词为例, 研究表明, 60%的分词错误是新词和未登录词造成的<sup>[6]</sup>。

<sup>2</sup> 生成词库论是近年兴起的一个重要的词汇语义学理论。和已有的理论相比(如语素分析法、语义场理论、题元角色理论、论元结构理论、格语法、概念结构理论、框架语义学、认知语义学等), 生成词库论提供了一个更为丰富、完备的词义描述框架, 被广泛应用于多种语言的词典编纂和语义资源建设之中。比如, 欧盟的两个重大项目——12种语言语义词典项目(SIMPLUE)和多语数据库项目(EuroWordNet)就是在生成词库论的指导下建设的。但在汉语中, 生成词库论还限于解释一些个别的语言现象, 系统的面向计算的词汇语义分析尚未展开<sup>[7]</sup>。本文将在生成词库论的指导下分析汉语词的语义结构。

关系，“供人坐”或“用来坐”是“椅子”的功用；一种是**构成关系**，“靠背和四条腿”是“椅子”的构成成分。

分析动词的语义结构，适合从论元结构入手。论元结构的基本语义成分是动词和论元，基本语义关系是动词和论元的关系。论元有两种：一种是必有论元，一种是非必有论元。前者可以进一步细分为施事、感事、经事、致事、主事、受事、与事、结果、对象、系事等语义角色，后者可以进一步细分为工具、材料、方式、原因、目的、时间、处所、源点、终点、量幅、范围等语义角色<sup>[8]</sup>。根据必有论元数量的多少，可将动词分为三类：一价动词（带一个必有论元）、二价动词（带两个必有论元）和三价动词（带三个必有论元）。必有论元是论元结构的必有构成成分，往往不可或缺，缺少这些论元，论元结构便不再完整；非必有论元是论元结构的必要构成成分，往往可以缺省，但为了释义的准确、充分，有时需要出现一些非必有的论元。分析动词的论元结构，先要弄清动词需要跟带哪些论元，哪些是必有论元，哪些是非必有论元；然后弄清动词和论元之间的关系。以“走（人或鸟兽用双脚在地上交替向前移动）”为例，分析它的论元结构，可以得到以下信息：

语义类信息：身体活动；位移类

必有论元：1个（人或鸟兽）；非必有论元：3个（双脚 / 前 / 地上）

语义关系：6种（一个是**语义类关系**，“移动”和“走”是上下位关系；一个是**施事-动作关系**，“人或鸟兽”是“移动”的施事；一个是**工具-动作关系**，“（用）双脚”是“移动”的工具；一个是**处所-动作关系**，“（在）地上”是“移动”的处所；一个是**方式-动作关系**，“（双脚）交替”是“移动”的方式；一个是**动作-方向关系**，“移动”的方向是“（向）前”）

分析形容词的语义结构，适合从搭配入手。选择从搭配入手，基于以下几方面的考虑：1) 形容词的意义虚泛、空灵，从其自身入手不容易把握，最好的办法莫过于看它的搭配对象。2) 从形式上看，形容词常和两类词搭配：一类是名词，如“白纸”“薄皮”“明月”“甜枣”“低空”等；一类是动词，如“明说”“慢走”“乱跑”“高唱”“紧握”等。从意义上看，形容词并不直接和实体名词搭配，而是先和属性名词搭配然后再和实体名词搭配，即“颜色白的纸 / 厚度小的皮 / 光线明的月亮 / 味道甜的枣 / 高度低的天空”；同样，形容词也并不直接和动词搭配，而是先和属性名词搭配然后再和动词搭配，即“以公开的方式说 / 以较慢的速度走 / 以较高的声音唱 / 秩序杂乱的跑 / 以较大的力度握”。这说明，形容词无论和实体名词还是和动词搭配，本质是一样的：都是先和属性名词搭配然后再和实体名词或动词搭配的。从这个意义上说，形容词的真正搭配对象是属性名词。3) 形容词和属性名词之间存在着一种稳定的关系。如果形容词和属性名词之间的关系已经规约化，那么即便不出现属性名词，也不会影响到对句子的理解。比如“这双鞋大”“那座山高”，其中的属性名词没有出现，但根据常识可以推知：“鞋大”不是说鞋的开口大或鞋帮大，而是说它的尺码大；“山高”不是说山的价值高或等级高，而是说它的海拔高（从山顶到山脚的垂直距离大）。如果形容词和属性名词之间的关系尚未规约化，那么一定要出现属性名词，否则句子的意思便无法理解。比如“\*这种油漆大”“\*那种电器高”无法理解，就是因为属性名词没出现；补出属性名词后（比如“这种油漆味道大”“那种电器价格高”），句子就可以理解了。可见，形容词对属性名词有一种特别的依赖关系。类似的关系，在实体名词和属性名词之间也存在：实体名词包含了多种复杂的属性，一种属性用一个属性名词来描述，比如“尺寸 / 颜色 / 味道 / 形状 / 距离 / 态度 / 品格 / 长短 / 面积 / 价格 / 重量 / 脾气 / 风格 / 形式 / 性能 / 动力 / 用途”等。基于这样的认识，文献<sup>[9]、[10]</sup>指出，实体名词、属性名词和形容词之间存在配价上的关系，实体名词可以看作是属性名词的论元，形容词是属性的值，即“实体名词（属性名词）=形容词”。比如“味道 / 价格”是属性名词，它们从属于实体名词“油漆 / 电器”，后者是前者的配价成分，形容词“大 / 高”是配价的值，由此可以得到“油漆味道大 / 电器价格高”。

词的意义和搭配之间存在着一种内在的联系<sup>[11]</sup>。知道了一个词有什么样的搭配，也就大体知道了这个词的使用环境，进而知道了这个词的意义和用法。对于形容词来说，主要和属性名词搭配。分析形容词的语义结构，就是要弄清形容词和什么样的属性名词搭配，这些属性名词反映的是名词的哪些属性。弄清了这些，也就大体弄清了形容词的搭配情况和使用情景，进而知道了形容词的意义和用法<sup>[10]</sup>。以“薄（扁平物体的厚度小）”为例，分析它的语义结构，可以得到以下信息：

语义类信息：物性的值；小类

语义成分：3个（实体名词：扁平物体；属性名词：厚度；属性值：小）

语义关系：3种（一种是**语义类关系**，“薄”和“小”之间有上下位关系；一种是**领属关系**，“扁平物体”包含着多种属性，“厚度”是其中的一种，两者存在着领属关系；第三种是**属性和属性值关系**，“厚度”是属性，“小”是该属性的值，两者是属性和属性值的关系）

属性名词和属性值是形容词意义的基本构成成分。解释形容词的意义，通常都会用到这两个成分。实体名词是形容词释义的必要补足成分，当属性名词不足时，可以用实体名词进行补充。比如“淡”有7个意义：①味道不浓；②特指盐分少；③泛指液体或气体中所含的某种成分少；④感情、兴趣、印象、关系等不深；态度不热情；⑤颜色浅；⑥生意少；⑦内容少（释义源自《规范》第3版）。其中，义项①②④⑤⑦直接用“属性名词+属性值”的释义方式；义项⑥找不到合适的属性名词，用“实体名词+属性值”进行释义；义项③除了“属性名词+属性值”的释义方式外，还引入了和实体名词相关的信息“液体或气体中所含的”，使得释义更加准确。

### 3 建构词的释义模版

基于词的语义结构分析，构建词的释义模版。构建释义模版时，遵循三个原则：

1) 不同的词类之间，释义模版可能是不同的。比如，构建名词的释义模版，适合从物性结构入手；构建动词释义模版，适合从论元结构入手；构建形容词的释义模版，适合从搭配对象（属性名词）入手；构建连词的释义模版，适合从使用条件入手。比较：<sup>3</sup>

【椅子】X是……，X是用来……，X由……构成。（X表被释词，下同）

（语义类）（功用角色）（构成角色）

【走】当人或鸟兽X时，他/它用…在…以…方式向…移动。

（施事）（工具）（处所）（方式）（方向）（语义类）

【薄】如果说一个物体X，那么意味着它的……（方面）……。

（实体）（属性）（属性值）

【跟】当……时用X。

（使用条件）

2) 同一词类内部、不同的语义类之间<sup>4</sup>，释义模版也可能是不同的。比如名词“纸（可供写字、绘画、印刷、包装等用的薄片状的东西，多用植物纤维制成）”和“质地（材料的结构性质）”词类相同，但由于语义类别不同（前者属于“具体事物-办公用品”类，后者属于“抽象事物-属性”类），释义模版便不同。前者的语义结构中包含着功用、形式、施成等

<sup>3</sup> 为增强词义信息的完整性和典型性，本文在释义时采用了文献<sup>[1]</sup>的方法，将被释词嵌入释义之中。以“椅子”为例：

《柯林斯》【chair】A chair is a piece of furniture for one person to sit on, with a back and four legs.

《现汉》【椅子】有腿有靠背的坐具。

《本文》【椅子】椅子是一种供人坐的家具，它有靠背和四条腿。

和《现汉》相比，本文释义的语义信息更加丰富。（进一步讨论，参见第5部分）

<sup>4</sup> 本文的语义分类主要参照《现代汉语语义分类词典》<sup>[13]</sup>、《现代汉语语义词典》（北京大学计算语言学研究所），个别分类吸收了现代语言学的研究成果。比如，将动词分为自主动词和非自主动词是吸收了文献<sup>[14]</sup>的成果。

物性角色，需要构建一个包含多种物性角色的释义模版；后者的语义结构只包含构成角色，只需要构建一个包含构成角色的释义模版即可。试比较：

【纸】X是……，X的形状是……，用X来……，X由……制成。  
(语义类)(形式角色) (功用角色)(施成角色)

【质地】X表示……方面的性质。  
(构成角色)(语义类)

类似的情况在其他词类中也存在。以动词为例，自主动词和非自主动词的语义类别不同，语义结构不同，释义模版也不同。自主动词是动作发出者做主、自由决定、自由支配的动作行为，动作行为具有“有意性”和“可控性”，比如“打、跑、吃、开、发球、生产”等；非自主动词是动作发出者无法自由支配的动作行为，动作行为具有“无意性”和“非可控性”，比如“病、死、跌、塌、丢、发芽、生根”等<sup>[14]</sup>。自主动词的动作发出者可以有意地利用某种工具、通过某种方式、为了某种目的发出动作行为，或通过某种动作有意地影响某人(某物)。因此自主动词的语义结构，除了包含动词和必有论元信息外，还包含意愿、工具、方式、依据、目的、处置等方面的信息。相对而言，非自主动词的语义结构，通常只包含包含动词和必有论元信息，不包含意愿、工具、方式、依据、目的、处置、影响等信息。试比较：

**自主动词：【打】**如果你X某人或某物，那么你有意用……大力击打他/它。  
(施事)(受事) (意愿)(工具)(方式)(语义类)

【发球】②如果你X球，那么你依照…把球发给……。  
(施事)(受事) (依据)(处置)(语义类)(与事)

**非自主动词：【丢】**如果你X某物，那么你失去了它。  
(施事)(受事) (语义类)

【发芽】如果种子X芽，那么它长出幼芽。  
(施事)(受事) (语义类)

3) 在同一词类、同一语义类内部，释义模板要相同或尽可能相同。比如：

- “笔、墨、纸、砚”等属于办公用品类名词，应该具有相同的释义模版；
  - “紫檀、白桦、红柳、黄蘗”等属于乔木类名词，应该具有相同的释义模版；
  - “打、击、敲、捶、揍”等属于打击类动词，应该具有相同的释义模版；
  - “煎、炒、烹、炸、煮”等属于烹饪类动词，应该具有相同的释义模版；
  - “高、长、宽、薄、深”等表示度量类属性的值，应该具有相同的释义模版；
  - “好看、漂亮、美观、俊秀”等表示评价属性的值，应该有相同的释义模版；
  - “稍、较、很、极、太”等属于程度类副词，应该具有相同的释义模版；
  - “朝、向、对、往、迎、逆”等属于方向类介词，应该具有相同的释义模版；
  - “自、打、从、起、到、至”等属于定位类介词，应该具有相同的释义模版；
- ……

有关同一语义类词语的释义模版，后面还会进一步讨论，这里暂略。

#### 4 利用释义模版解决词的释义问题

文献<sup>[15]、[16]</sup>利用释义模版解决了名名短语的自动释义问题(准确率为94.23%)，本文利用释义模版解决词的释义问题。根据本文的观察，利用释义模版可以解决以下四方面的问题：

1) 解决多义词的释义问题。多义词的义项之间存在语义上的联系，利用这种联系，可以“把错综复杂的一词多义的现象理出头绪，并串联起来，从而更好地掌握一个词的全部意义”<sup>[17:39-40]</sup>。为了增强义项间的语义联系，《新牛津英语词典》(1998)、《规范》做了大量的研究工作，但义项间的语义联系依然不够明显<sup>[3]</sup>。本文通过释义模版来解决这一问题。以形容词“好”为例，《规范》(第3版)概括了6个义项，分别是：①美；优点多的；令人满意的。②友爱；和睦。④(身体)健康。⑥表示赞同、答应、结束或不满、警告等语气。⑦容

易。⑨用在某些动词前面，表示效果好。通过考察，本文发现这6个义项都是在义项①的基础上发展出来的，区别在于搭配对象。搭配对象不同，“好”的意义便不同。因此，从搭配入手可以为“好”建构一个统一的释义模版，即“如果说……好，那么他或它……方面优点多；令人满意”。利用这一模版，可以将“好”的各项意义关联起来。试比较：（为了节约篇幅，义项①用完整释义模式“如果……，那么……”，其他义项用简洁释义模式。下同。）

表1 多义词“好”的释义比较

	《规范》的释义		本文的释义
①	优点多的；令人满意的。	好+N	如果说一个人/事/物好，那么他或它的某方面 <u>优点多</u> ； <u>令人满意</u> 。跟“坏”相对。
②	友爱；和睦	好+N	（人际关系） <u>令人满意</u> 。跟“友爱”相近。
④	（身体）健康	好+N	（身体状况） <u>令人满意</u> 。跟“健康”相近。
⑥	表示赞同、答应等语气。	好，S	（某话题） <u>令人满意</u> 。跟“赞同”等相近。
⑦	容易。	好+V	（做某事的 <u>优点</u> ）容易。
⑨	用在某些动词前面，表示效果好。	好+V	（做某事的效果） <u>令人满意</u> 。

和《规范》相比，本文的释义有三方面优点：1）前者侧重于形容词的属性值，对形容词的搭配对象（比如属性名词）、使用环境关注不够；本文不仅关注形容词的属性值，也关注形容词的搭配对象和使用语境，通过设置典型语境凸显形容词的典型意义和典型用法，便于读者理解形容词的意义。2）用同一释义模版解释多义词的多项意义，有利于增强义项间的语义联系。3）兼用“跟……相近”或“跟……相对”的释义方式，有利于增强被释词与其相关同义词、反义词的联系。

利用释义模版可以解决多义形容词的释义问题，也可以解决其他多义词的释义问题。比如动词“传”，《规范》概括了6个义项：①一方交给另一方；上代交给下代。②把知识、技能等教给别人。③广泛散布；宣扬。④命令别人来。⑤表达；流露。⑥热或电在导体中流通。仔细观察这6项意义，不难发现它们的核心意义是一致的，都表示“转移”。利用这一共性，可以建立统一的释义模版：“如果一方X……给另一方，那么……从一方转移到另一方”。利用这一模版，可以将“传”的各项意义重新解释如下：

**【传】释义模版：如果一方X……给另一方，那么……从一方转移到另一方。**

- ①如果一方（上一代）X物品或技艺给另一方（下一代），那么物品或技艺从一方（上一代）转移到另一方（下一代）。跟“传递、遗传”相近。
- ②如果一方X知识、技能等给另一方，那么知识、技能等从一方转移到另一方。跟“传授”相近。
- ③如果一方X言论、信息等给另一方，那么言论、信息等广泛地从一方转移到另一方。跟“传播”相近。
- ④如果说话人X命令给听话人，那么命令从说话人这里转移到听话人那里，并使听话人来说话人这里。跟“传唤”相近。
- ⑤如果一方X思想、感情等给另一方，那么思想、感情等从一方转移到另一方。跟“传情”相近。
- ⑥如果某物体X电或热，那么电或热从该物体的一端转移到另一端。跟“传导”相近。

利用释义模版解决多义词的释义问题，需要有语言学研究的支持。可喜的是，学术界已经做了一些相关的研究工作。文献<sup>[18]</sup>发现，多义词有一个共同的核心义。比如“发”，核心义表示“（使）…离开…”，利用核心义可以将“发”的各项意义联系起来：义项①“发射”表示“使箭或炮弹离开”；义项②“产生；发生”表示“使某种事物脱离本体”；义项③“派

遣；出发”表示“(使)人有目的地离开”；义项④“打开”表示“使覆盖的部分离开”；义项⑤“启发”表示“使蒙蔽思想的部分离开”。如果说文献<sup>[18]</sup>的研究还是个案式的、无法窥测核心义全貌的话，那么文献<sup>[19]</sup>对“按、拔、报、保、北、备、比、标、并、波、驳”等200多个多义词核心义的研究则完善了“核心义的概念体系”，凸显了核心义对多义词释义的价值。虽然核心义研究和释义模版研究在研究角度、方法、材料、表现方式等方面有着众多的不同，但两者的终极目的是是一致的：都是为了找到一种语义线索，将多义词的各项意义关联起来。因此，建构释义模版时，可以充分吸收核心义的研究成果，促进语言学研究向机器词典编纂、机器自动释义的方面转化。

2) 解决同义词的释义问题。同义词的广泛分布，让人们习惯于用一个同义词来解释另一个同义词。比如，解释“好看”用“漂亮；美观”；解释“漂亮”用“好看”，解释“美观”用“好看；漂亮”。这样做，有两方面的缺欠：1) 造成了大量的循环释义，这方面的研究以WordNet为代表。2) 难以凸显同义词之间的细微语义差别。同义词之间的差别是细微的，有些需要通过细粒度(fined-grained)的语义分析后才能凸显出来；不进行细粒度的语义分析，往往难以发现同义词之间的细微语义差别，也难以真正地解决同义词内部的释义问题。利用释义模版则可以解决这一问题。由于同义词以名、动、形为绝大多数<sup>5</sup>，因此，下面的讨论将围绕着名、动、形同义词展开。

先来看名词，以“椅子”类同义词(如“椅子”“竹椅”“躺椅”“长椅”“摇椅”等)为例。前面解决了“椅子”的释义问题：椅子是……，椅子是用来……，椅子由……构成。解决了“椅子”的释义问题，和“椅子”同一语义类词的释义问题也就解决了。因为，这些词和“椅子”共用一个释义模板。稍有不同的是，前者在后者的基础上增加或替换了一些语义角色(下划线部分)。比如：

**释义模版：X是一种……的椅子。**

【竹椅】X是一种 由竹子做的 椅子。 (增加了施成角色)

【长椅】X是一种 长条形的 椅子。 (增加了形式角色)

【躺椅】X是一种 用来躺卧的 椅子。 (替换了功用角色)

【摇椅】X是一种 有枕部和扶手、可以前后摇晃的 椅子。(替换了构成、功用角色)

接下来看动词，以“打击”类动词为例。文献<sup>[20]</sup>用了“工具、方式、力度、部位/对象、频率、目的/结果”六种语义成分区分“打击”类动词。为了凸现动词典型语境中的典型用法，本文解释“打击”类动词时用了“施事、受事、意愿、工具、力度、频率、语义类、部位、目的/结果”九种语义成分。释义之前，先构建“打击”类动词的释义模版：

如果 施 X 受,那么 施 有意 用… 以…力度或频率 击打 受或受的…部位,以/使…。

(施事)(受事) (意愿)(工具)(力度或频率)(语义类) (部位) (目的/结果)

然后，利用这一模版，解释“打击”类动词的意义。比如：

【打】如果你 X 某人或某物，那么你 有意 用手或器物 大力 击打 他(它)。

【敲】如果你 X 某物(门、鼓等)，那么你 有意 用手或器物 连续 击打 它，以引起某人的注意。

【捶】如果你 X 某人或某物，那么你 有意 用拳头或棒槌 连续 击打 他(它)。

【拍】如果你 X 某人或某物，那么你 有意 用手掌或片状物 击打 他(它)。

【掴】如果你 X 某人，那么你 有意 用手掌 击打 他(脸/耳光)。

【扑】如果你 X 某物，那么你 有意 用手或器物 轻轻 击打 它。

【揍】如果你 X 某人，那么你 有意 用手或器物 击打 他。

在这九种语义成分中，施事、意愿、语义类是“打击”类动词的共有语义成分，击打部位、

<sup>5</sup> 根据文献<sup>[21]</sup>对《新华同义词词典》(2005版)的统计，名词、动词、形容词占同义词的绝大多数(占92%)。其中，动词数量最多(占45%)，形容词次之(占25%)，名词再次之(占22%)。

击打工具、击打力度、频率等是每个“打击”动词特有的语义成分（下划线部分）。解释“打击”类动词，除了解释它们共有的语义成分外，还要解释每一个动词特有的语义成分。解释前者，是为了凸显“打击”类动词的共性；解释后者，是为了凸显每一个“打击”动词的个性，或者说，是为了凸显各“打击”类动词之间的细微语义差别。

最后看形容词，以“漂亮”“好看”“美观”为例。解释这些词语，《现汉》《规范》的释义是循环的，本文的释义则不会出现这样的问题。试比较：

表 2 同义词“漂亮”“好看”“美观”的释义比较

	《规范》释义	本文的释义
【漂亮】	好看。	如果说一个人（事、物、动作）X，那么他（他）某方面优点多，看起来令人满意。跟“好看”相近。
【好看】	漂亮；美观。	如果说一个人（物、动作）X，那么他（它）看起来令人满意。跟“漂亮”相近。
【美观】	好看；漂亮。	如果说一个景（物）X，那么它看起来令人满意。跟“好看”相近。

通过释义模版，“漂亮”“好看”“美观”的差别就很明显了：“漂亮”可以评价人、事、物或动作的整体（如“这一仗打得真漂亮”评价打仗的整体），也可以评价人、事、物或动作的某一方面（如外观）；“好看”通常评价人、物、动作的某一方面（如外观）；“美观”通常评价景、物的某一方面（如外观、款式等）。

3) 解决新造合成词的释义问题。新词虽然词形、词义是“新的”，但词法结构、语义结构往往是“旧的”。因此，可以用“旧的”释义模板解释新词。比如“裸X”，分析它的语义结构，可以建构如下释义模板：

【裸X】（相对……来说，）裸X是一种除了其自身外、什么都没有或不附带的X。

利用这一模板，可以解释像“裸机”这样已收入《现汉》的新词，也可以解释像“裸官”“裸考”“裸油价”这样未收入《现汉》的新词：

【裸机】裸机是一种除了其自身外、什么（网络、系统等）都不附带的机器。

【裸官】（相对家属、财产在国内的官员来说，）裸官是一种除了其自身外、什么（家属、财产）都不在国内的官员。

【裸考】（相对有加分的考试来说，）裸考是一种除了其自身（的分数）外、什么（加分）都没有的考试。

【裸油价】裸油价是一种除了油自身的价格外、什么（附加税）都不附带的价格。

利用释义模版，可以解决两类新词的释义问题：

一类是构词形式相似的新词，以仿拟词为代表。通过仿拟产生的新词，在构词形式、表意特点方面有很多相似之处。利用这些相似之处，可以为仿拟词建构统一的释义模版。比如：“裸X”表示“一种除了其自身外、什么都没有或不附带的X”；“XX门”表示“一种和……有关的丑闻”（如“艳照门/伊朗门/拉链门/甲醛门/认证门/爆炸门/婚庆门”等）；“XX控”表示“一种近乎极端地喜欢……的人”（如“微博控/职场控/镜子控/冰山控/肌肉控”等）；“X姐”表示“一种从事……（工作）的年轻女性”（如“空姐/海姐/吧姐/网姐”等）；“微X”表示“一种……（方面）比同类事物小的X”（如“微博/微信/微客/微商/微电影/微消费”等）；“互联网+”表示“一种互联网和……相结合的行业”（如“互联网+政务服务/互联网+手机/互联网+杂货店/互联网+社区/互联网+金融/互联网+农业”等）。

一类是语义上密切相关的新词，以新造同义词为代表。这些词在产生新意义之前，概念意义可能并不相同；进入某一特定的领域并被赋予了某种特定的意义后，概念意义才变得相同或基本相同。比如“奔驰”“宝马”“大众”“标致”“现代”，没有描述汽车品牌之前，不是同义词；进入汽车行业并用来描述汽车品牌后，才变成了同义词。不是同义词的时候，不



能用释义模版来解释；变成同义词后，才可以用释义模版来解释：（X 表被释词，下同）

X 是……汽车品牌，X 始创于……，X 的标志是……，X 的特色是……，X 的车型有……。

（语义类）（施成角色）（形式角色）（评价角色）（构成角色）

有些新造同义词的使用条件受限，只能在特定的语境下使用。解释这些新词，需要设置一个特定的语境。比如“潜水”“冒泡”“吐槽”“活跃”“话唠”“传说”，作为同义词，通常只在 QQ 群中使用，用来描述 QQ 群成员的活跃程度。其中，“潜水”的活跃程度最低，发言次数最少，甚至长期不发言；“冒泡”的活跃程度较低，偶尔发几次言；“吐槽”的活跃程度中等，针对某些话题发言较多；“活跃”的活跃程度较高，发言次数超出了平均水平；“话唠”的活跃程度很高，发言次数超出平均水平很多；“传说”的活跃程度最高，发言次数远远超出了平均水平，只有少数人能达到。基于以上分析，可为“潜水”类新词构建如下释义模版：

如果某人在 QQ 群中处于 X 等级，那么意味着他在该群的活跃程度…，发言次数…。

4) 有利于解决句子或篇章的理解问题。释义模版不是凭空产生的，而是在大量真实语料的基础上提炼、概括而成的。释义模版概括得越全面、越准确，越有利于理解词在真实语境中的意义和用法。以“食堂”为例，《规范》将其解释为“机关、团体里向本单位人员供应伙食的部门和处所”，还不够全面。因为，该释义只在“食堂”和“创办单位（机关、团体）”“就餐人员（本单位人员）”“食物（伙食）”之间建立起了联系，有利于理解例 a 类的句子；没有在“食堂”和“柴米油盐”“炊事员”“筷子”之间建立起联系，不利于理解例 b 类的句子。（语料源自北京大学 CCL 语料库）

a. 工人们在食堂打了饭就出来，……。（就餐人员、食物）

在本世纪四十年代初，县城的中学没有食堂，……。（创办单位）

她……，在医院食堂买了饭菜带回来让我们吃了以后，又匆匆地去上班了。（创办单位、食物、就餐人员）

b. 我们学校食堂用蒸箱蒸馒头，……。（创办单位、食物、炊具）

村里办起了食堂，家中的米盐柴什么的也全被村里没收了，……（创办单位、食材）

我看见了……这条船的食堂，炊事员正在洗筷子，……。（创办者、炊事人员、餐具）

在物性结构理论的指导下，本文将和“食堂”有关的各项信息（如“创办单位”“就餐人员”“食物”“食材”“餐具”“炊具”“炊事员”等）都关联到了一起，形成了新的“食堂”释义：

**【食堂】**X 是一个由食材、食物、炊具、餐具、炊事人员等构成（构成角色）的非盈利性场所（语义类），主要由机关、学校等企事业单位创办（施成角色），用于制作和供应企事业单位内部职工、学生的饮食（功用角色）。

有了这一释义，就不仅可以理解例 a 类的句子，也可以理解例 b 类的句子。

以上说了释义模版对句子理解的作用，下面再说一下释义模版对语篇理解的作用。以汽车品牌“奔驰”为例，对于这一品牌，百度百科用了大量的篇幅（14000 多字）来描述，内容涉及“奔驰创始人”“奔驰标志”“品牌特色”“发展历程”“车型列表”等各个方面。

图 1 百度百科对“奔驰”的语篇描述目录

## 目录

### 1 奔驰创始人

### 2 奔驰标志

### 3 品牌特色

- 几经变迁，锐气不减
- 名牌本色
- 贵族产品
- 文化精髓轿车
- 紧凑的SLK
- 内外配置
- 科技新星——幻形F200
- 驰骋疆场控群雄

- 超级跑车CLK-GTR

### 4 发展历程

### 5 进入中国

### 6 奔驰汽车车型列表

- 乘用车系
- 商用车系
- 轻型商用车

### 7 商旅车(MPV)

- 外观
- 内饰

- 座椅

- 动力
- 安全性

### 8 经营理念

### 9 aero trailer

### 10 奔驰品牌创始人

### 11 易发故障及解决方法

#### 方法

- 四轮定位
- 更换火花塞
- 节气门体脏污后的症状

- 免拆清洗喷油器
- 转向机漏油
- 水泵损坏渗漏冷却液
- 燃油泵的故障现象
- 制动开关故障
- 空气流量计的故障现象
- 下摆臂球头损坏后症状

解释“奔驰”的意义时，不可能将这些信息全部都收入词典，只能收入和“奔驰”相关的最重要的信息。基于上文的汽车释义模版，参照百度百科，本文将“奔驰”的意义概括如下：

【奔驰】奔驰是德国汽车品牌（语义美），始创于1926年（施成角色），标志是三叉星（形式角色），因外貌出众、性能卓越、售后服务良好（评价角色）而闻名，主要车型有乘用车系、商用车系、轻型商务车系（构成角色）。

由于“奔驰”的释义源于“奔驰”品牌的语篇描述，是“奔驰”语篇描述的一种精炼化表示。因此，可以将“奔驰”的释义看作是“奔驰”语篇描述的一个关键词集合。弄清了“奔驰”的意义是什么，也就大体弄清了“奔驰”语篇描述的关键词有哪些。从这个意义上说，做好词语“奔驰”的释义工作，显然对于理解语篇中的“奔驰”也是有帮助的。

## 5 结语

机器词典和词汇语义资源是自然语言理解中最重要、最基础性的语言资源（冯志伟2011），机器学习、数据挖掘、单词相似度计算、词汇搭配关系自动获取、词汇联想关系自动获取、动词次范畴框架的自动获取等，通常都要在机器词典和词汇语义资源的基础上进行。但目前的机器词典和词汇语义资源的建设还不尽如人意，词的语义知识描述、表示等问题依旧是制约词汇语义分析深入的一个关键问题<sup>[22]</sup>。

为了解决词义知识的描述问题，本团队成员做了大量的研究工作。包括：构建了汉语亲属关系自动推理模型<sup>[23]、[24]</sup>、名名组合的释义模板<sup>[15]、[16]</sup>、汉语名词物性结构的描写体系<sup>[1]</sup>、基于物性结构的形容词语义分析框架<sup>[25]</sup>，揭示了复合名词词义中的物性角色分布规律<sup>[26]</sup>。在前期研究的基础上，本文进一步推进，尝试建构系统的汉语词的释义模版。建构释义模版时，本文对释义方法做了两方面的改进：

1) 受柯林斯词典（Collins COBUILD Advanced Learner's English Dictionary 2006年第5版）的启发，放弃了传统的、描述式的释义方法，改用了基于用法的、例证式的释义方法。以“打”为例，试比较：

《柯林斯》【hit】①If you hit someone or something, you deliberately touch them with a lot of force, with your hand or an object held in your hand.

《现汉》【打】①用手或器具撞击物体。

《本课题》【打】①如果你打某人或某物，那么你会故意用手或器物大力击打它。

以往的释义方法，被释词和释义是分离的，加以缺少必要的语境信息，容易造成词义信息的缺失，不利于学习者理解和使用；例证式的释义方法，被释词和释义是一体的，被释词被嵌入到了释义之中。例证式释义方法的优点，诚如《柯林斯》词典前言所说：“所有义项均采用整句释义，亲切自然，深入浅出，成功凸显了词语在典型语境中的典型用法，释义本身就是绝佳的例证”。另外，在释义中加入语境信息，除了可以凸显“词语在典型语境中的典型用法”外，还最大限度地补足了词义信息，有利于解决词义信息不完备、不完整的问题。

2) 放弃按词、按义项逐一释义的方法，改用了模版统一释义的方法。建构释义模版之前，先要分析词的语义结构，通过分析词的语义结构，弄清词义的构成成分和构成关系，确

定哪些是词义的必有成分和必有关系、哪些是可有成分和可有关系。然后，再建构词的释义模版。建构释义模版时，遵循三个基本原则：不同的词类之间，释义模板可能是不同的；同一词类内部、不同的语义类之间，释义模板也可能是不同的；但在同一词类、同一语义类内部，释义模板要相同或尽可能相同。

研究表明，建构释义模版有助于解决词语释义不完备问题、多义词释义问题、同义词释义问题、新造合成词释义问题以及句子、语篇的理解问题。如果本文的方法是正确的，那么有望能为解决词汇语义知识描述、表示难题提供一个新思路。

## 参考文献

- [1] 袁毓林. 汉语名词物性结构的描写体系和运用案例[J].当代语言学, 2014, (1):31-48.
- [2] 李强, 袁毓林. 生成词库理论和名词语义的结构描述与概念解释[A].词汇学国际学术会议暨第十一届全国汉语词汇学学术研讨会论文, 北京: 北京大学, 2016.
- [3] 施春宏. 词义结构的认知基础及释义原则[J].中国语文, 2012, (2): 114-127.
- [4] 马英新. “动+名”偏正式双音复合词的结构义及其释义研究[D].石家庄: 河北师范大学博士学位论文, 2013.
- [5] Pustejovsky J. The generative lexicon[M]. Cambridge: MIT Press, 1995.
- [6] Sprout R., Emerson T. The First International Chinese Word Segmentation Bakeoff [C]//Proc of the 2nd SIGHAN Work-shop on Chinese Language Processing, 2003: 11-17.
- [7] Pierrette Bouillon, Federica Busa (eds.), 黄居仁, 苏祺(导读), 2001/2010, 词义的语言:词义理论的跨学科讨论[C].北京: 北京大学出版社.
- [8] 袁毓林. 汉语配价语法研究[M].北京: 商务印书馆, 2010.
- [9] Miller, G.A. Nouns in WordNet[A]. In C. Fellbaum(ed.). WordNet: An Electronic Lexical Database [C]. Cambridge & MA: The MIT Press, 1998.
- [10] 袁毓林. 形容词的语义特征和句式特点之间的关系[J].汉藏语学报, 2013, (7): 147-165.
- [11] Lyons, J. Semantics[M]. Cambridge: Cambridge University Press, 1977.
- [12] 柯林斯出版公司. Collins COBUILD advanced learner's English dictionary: new edition[M]. 北京: 外语教学与研究出版社, 2006.
- [13] 苏新春(主编). 现代汉语语义分类词典[Z].北京: 商务印书馆, 2014.
- [14] 马庆株. 自主动词和非自主动词[J].中国语言学报, 北京: 商务印书馆, 1988, (3):157-180.
- [15] 魏雪, 袁毓林. 基于语义类和物性角色建构名名组合的释义模板[J].世界汉语教学, 2013, (2):172-181.
- [16] 魏雪, 袁毓林. 基于规则的汉语名名组合的自动释义研究[J].中文信息学报, 2014, (3):1-10.
- [17] 郭在贻. 训诂学(修订本)[M].北京: 中华书局, 1986/2005.
- [18] 张联荣. 谈词的核心义[J]. 语文研究, 1995, (3):31-36.
- [19] 王云路, 王诚. 汉语词汇核心义研究[J].北京: 北京大学出版社, 2014.
- [20] 蒋绍愚. 打击义动词的词义分析[J].中国语文, 2007, (5):387-401.
- [21] 欧阳丽文, 李仕春. 从词性的角度看现代汉语同义词的分布规律[J].广西教育学院学报, 2013, (2):82-85.
- [22] 中文信息学会. 中文信息处理发展报告[Z]. www.cipsc.org.cn, 2016.
- [23] 陈振宇, 袁毓林, 张秀松, 等. 一种基于大知识库的亲属关系自动推理模型[J]. 中文信息学报, 2010a, 24(3):117-123.
- [24] 陈振宇, 袁毓林. 汉语亲属关系的语义表示和自动推理[J]. 中国语文, 2010b, (1):44-56.
- [25] Wang Enxu, Yuan Yunlin. The Meaning of the Chinese Polysemous Adjective “Hao” (Good). Chinese Lexical Semantics. Wu Yunfang et al. (Eds), CLSW2017, Springer Verlag, Berlin Heidelberg, 2017. (forthcoming)
- [26] 王恩旭, 袁毓林. 词义中物性角色的分布及其对词语释义的影响[J].外国语, .2017. (已录用, 待发表)

作者联系方式:

王恩旭 北京大学承泽园 130 楼 207 室 100871  
电话:18813141663  
电子邮箱: [wangbush000@126.com](mailto:wangbush000@126.com)

袁毓林 北京大学中文系 100871  
电子邮箱: [yuanyl@pku.edu.cn](mailto:yuanyl@pku.edu.cn)