

文章编号: 1003-0077(2018)01-0068-07

机器词典释义模版的建构和运用

王恩旭^{1,2}, 袁毓林¹

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

摘要: 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, 该文通过分析词的语义结构和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、循环释义、新词释义等问题。

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 and Application of Chinese Words Interpretation Template

WANG Enxu^{1,2}, YUAN Yulin¹

(1. Department of Chinese Language and Literature, Peking University, Beijing 100871, China;

2. School of Liberal Arts Qufu Normal University, Qufu, Shandong 273165, China)

Abstract: Machine understanding of words is mainly based on dictionaries, but the present dictionary interpretation is inaccurate and imperfect. This paper investigates this issue by analyzing the semantic structure words of and constructing interpretation templates of the words. By analyzing the words semantic structure, we try to discover what semantic components and semantic relations 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. Finally, this paper shows that constructing the interpretation template is helpful to solve the following issues: the polysemous words interpretation, the synonym words interpretation, the new words interpretation problem, etc.

Key words: generative lexicon theory; interpretation templates; polysemy; synonyms; new words

0 引言

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。突出表现在:

(1) 文献[1]、[2]考察词典的名词释义时发现, 释义不完备的情况随处可见。比如“书”“唱片”等本该包含“物质”“信息”两个义位, 但词典只收录“物质”一个义位, 漏掉了“信息”义位。

(2) 文献[3]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) 等辞书后发现, “本义-引申义”的联系普遍不够明显。

(3) 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。

从文献[4]对《现汉》(2012 版) 的调查来看, 2 094 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

收稿日期: 2017-09-29 定稿日期: 2017-12-05

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(2016M600004); 山东省社科基金(16CZWJ31); 国家语委科研项目(YB135-45)

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。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) 不论词典的规模多大、更新的速度多快,通常都无法穷尽所有的新词,而这些新词往往就是机器理解词的主要障碍。^①

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。

1 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。^② 根据文献[5],词的概念结构中可能包括以下四方面的信息:

(1) 语义类型(TYPESTR): 每个词都属于一个特定的语义类;

(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[5]的四种物性角色外,还可能包含材料、单位、处置、行为、定位五种物性角色^[1]。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(语义类关系,“家具”和“椅子”是上下位关系;功用关系,“供人坐”或“用来坐”是“椅子”的功用;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^[8]。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)。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:

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(语义类关系,“移动”和“走”是上下位关系;施事-动作关系,“人或鸟、兽”是“移动”的施事;工具-动作关系,“(用)双脚”是“移动”的工具;处所-动作关系,“(在)地上”是“移动”的处所;方式-动作关系,“(双脚)交替”是“移动”的方式;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:另一类是名词,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”;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”。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,其真正的搭配对象是属性名词。(3)文献[9-10]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”。

词的意义和搭配之间存在着一种内在的联

① 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^[6]。

② 和已有的理论相比,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。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^[7]。

系^[11]。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。弄清了这些,也就大体弄清了形容词的意义和用法^[2]。以“薄(扁平物体的厚度小)”为例:

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(语义类关系,“薄”和“小”之间有上下位关系;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;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)

2 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;构建动词的释义模版,适合从论元结构入手;构建形容词的释义模版,适合从搭配对象入手。比较:^①

【椅子】X是……(语义类),X是用来……(功用角色),X由……构成(构成角色)。(X表被释词,下同)

【走】当人或鸟兽(施事)X时,他/它用……(工具)在……(处所)以……方式(方式)向……(方向)移动(语义类)。

【薄】如果说一个物体(实体)X,那么意味着它的……方面(属性)……(属性值)。

(2) 同一词类内部、不同的语义类之间^②,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。试比较:

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

【质地】X表示……方面的(构成角色)性质(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,释义模版也不同。自主动词具有“有意性”和“可控性”,比如“打、跑、吃、开、发球、生产”等;非自主动词具有“无意性”和“非可控性”,比如“病、死、跌、塌、丢、发芽、生根”等^[14]。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、目的等信息。相对而言,非自主动词的语义结构,通常只包含动词和必有论元信息,不包含意愿、工具、方式、依据、目的等信息。试比较:

自主动词:

【打】如果你(施事)X某人或某物(受事),那么你有意(意愿)用……(工具)大力(方式)击打(语义类)他/它。

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

【丢】如果你(施事)X某物(受事),那么你失去了(语义类)它。

【发芽】如果种子(施事)X芽(受事),那么它长出(语义类)幼芽。

(3) 在同一词类、同一语义类内部,释义模版要相同或尽可能相同。比如:

“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;

“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;

……

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。

3 利用释义模版解决词的释义问题

文献[15-16]利用释义模版解决了名名短语的

① 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[1]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)

② 本文的语义分类主要参照《现代汉语语义分类词典》^[13]、《现代汉语语义词典》(北京大学计算语言学研究),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[14]的成果。

自动释义问题(准确率为 94.23%)。在词语层面,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”^[17: 39-40]。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^[3]。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)概括了六个义项,本文从搭配入手为“好”建构了统一的释义模版,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,将“好”的各项意义关联了起来,如表1所示。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> ;令人满意。跟“坏”相对
②	友爱;和睦	好+N	(人际关系)令人满意。跟“友爱”相近
④	(身体)健康	好+N	(身体状况)令人满意。跟“健康”相近
⑥	表示赞同、答应等语气	好,S	(某话题)令人满意。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)令人满意

和《规范》相比,本文的释义有三方面优点:①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》概括了六个义项: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。仔细观察这六项意义,不难发现它们都表示“转移”。利用这一共性,可以建立统一的释义模版。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,需要语言学研究的支持。文献[18]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[18]的研究还是个案式的、无法窥测核心义全貌的话,那么文献[19]对“按、拔、报、保、北、备”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。

(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,不仅造成了大量的循环释义,也难以凸显同义词之间的细微语义差别。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^①。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”类同义词为例。前面解决了“椅子”的释义问题,基于这一释义,可以解决和“椅子”同一语义类词的释义问题。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。解释“打击”类动词,文献[21]用了“工具、方式、力度、部位/对象、频率、目的/结果”六种语义成分,本文用了“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。释义前,先构建释义模版:

如果 施(施事) X 受(受事),那么 施有意(意愿) 用……(工具) 以……(力度或频率) 击打(语义类) 受或受的……部位(部位),以/使……(目的/结果)。

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。

【掴】如果你 X 某人,那么你 有意 用手掌 击打他(脸/耳光)。

【揍】如果你 X 某人,那么你 有意 用手或器物 击打他。

施事、意愿、语义类是共有的语义成分,击打部位、击打工具、击打力度等是每个“打击”动词特有的语义成分(下划线部分)。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,即凸显“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,本文则不会出现这样的问题,二者的比较如表 2 所示。

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”),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版:

【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。

利用这一模版,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,可以解决以下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

^① 根据文献[20]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%)。

链门”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”等)。

另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后，概念意义才变得相同或基本相同。比如“奔驰”“宝马”“大众”等，没有描述汽车品牌之前，不是同义词；进入汽车行业后，才变成了同义词。变成同义词后，可以用释义模版来解释：

X 是……汽车品牌_(语义类)，X 始创于……_(施成角色)，X 的标志是……_(形式角色)，X 的特色是……_(评价角色)，X 的车型有……_(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”，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；“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。基于以上分析，可构建“潜水”类新词的释义模版：

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”“就餐人员”“食物”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“食材”“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)

a1. 工人们在食堂打了饭就出来，……。 (就餐人员、食物)

a2. 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)

b1. 我们学校食堂用蒸箱蒸馒头，……。 (创办单位、食物、炊具)

b2. 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)

b3. 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)

通过物性结构，本文将和“食堂”有关的重要信息都关联到了一起，形成新的“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，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 14 000 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”等方面，如图 1 所示。

目录	1 奔驰创始人	超级跑车 CLK-GTR	座椅	免拆清洗喷油器
	2 奔驰标志	4 发展历程	动力	转向机漏油
	3 品牌特色	5 进入中国	安全性	水泵损坏渗漏冷却液
	▪ 几经变迁，锐气不减	6 奔驰汽车车型列表	8 经营理念	燃油泵的故障现象
	▪ 名牌本色	▪ 乘用车系	9 aero trailer	制动力开关故障
	▪ 贵族产品	▪ 商用车系	10 奔驰品牌创始人	空气流量计的故障现象
	▪ 文化精髓轿车	▪ 轻型商用车	11 易发故障及解决方法	下摆臂球头损坏后症状
	▪ 紧凑的 SLK	7 商旅车(MPV)	▪ 四轮定位	
	▪ 内外配置	▪ 外观	▪ 更换火花塞	
	▪ 科技新星——幻形 F200	▪ 内饰	▪ 节气门体脏污后的症状	
	▪ 驰骋疆场挫群雄			

图 1 百度百科对“奔驰”的语篇描述目录

利用释义模版，可将以上的描述概括如下：

【奔驰】奔驰是德国汽车品牌_(语义类)，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)。

词语释义是语篇描述的一个关键词集合，弄清

了“奔驰”的词汇意义是什么,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“奔驰”的释义工作,对于理解语篇中的“奔驰”也是有帮助的。

4 结语

机器词典和词汇语义资源是自然语言理解中的基础性资源,机器学习、数据挖掘、单词相似度计算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^[22]。

为了解决词义知识的描述问题,本团队构建了亲属关系自动推理模型^[23-24]、名名组合释义模板^[15-16]、汉语名词物性结构描写体系^[1]、形容词语义分析框架^[25],揭示了复合名词词义中的物性角色分布规律^[26]。在此基础上,本文进一步推进,建构了汉语词的释义模版。建构释义模版时,本文做了两方面的改进:

(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】① If 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;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。”

(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。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、句子理解及语篇理解等问题。

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[C]. 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] Sproat R, Emerson T. The first international Chinese word segmentation bakeoff [C]//Proceedings of the 2nd SIGHAN work-shop on Chinese language processing, 2003: 11-17.
- [7] Pierrette B, Federica B (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[M]. 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 [M]. WordNet: An Electronic Lexical Database. 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. 现代汉语语义分类词典[Z]. 北京: 商务印书馆, 2014.
- [14] 马庆株. 自动词和非自动词[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] 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.

(下转第 95 页)

- tion[C]//Proceedings of the Proceedings of SPIE - The International Society for Optical Engineering, 2012, 8297(1):82970Y-82970Y-8.
- [11] Moysset B, Bluche T, Knibbe M, et al. The A2iA Multi-lingual text recognition system at the second maurdor evaluation[C]//Proceedings of the Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on. IEEE, 2014:297-302.
- [12] Thomas S, Chatelain C, Heutte L, et al. A deep HMM model for multiple keywords spotting in handwritten documents[J]. Pattern Analysis and Applications, 2015, 18(4):1003-1015.
- [13] Roy P P, Bhunia A K, Das A, et al. HMM-based Indic handwritten word recognition using zone segmentation[J]. Pattern Recognition, 2016(60):1057-1075.
- [14] Li J, Yu D, Huang J T, et al. Improving wideband speech recognition using mixed-bandwidth training data in CD-DNN-HMM[C]//Proceedings of the Spoken Language Technology Workshop. IEEE, 2012: 131-136.
- [15] Españañoquera S, Castrobleda M J, Gorbemoya J, et al. Improving offline handwritten text recognition with hybrid HMM/ANN models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 33(4):767-773.
- [16] Graves A, Ndez S, Gomez F, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]//Proceedings of the International Conference on Machine Learning. ACM, 2006:369-376.
- [17] Povey D, Ghoshal A, Boulianne G, et al. The Kaldi speech recognition toolkit[J]. Idiap, 2012.
- [18] Miao Y. Kaldi + PDNN: Building DNN-based ASR systems with Kaldi and PDNN[J]. Computer Science, 2014.



范道尔吉(1980—),博士研究生,讲师,主要研究领域为人工智能、模式识别与蒙古文信息处理。
E-mail: fandaoerji@126.com



高光来(1964—),硕士,教授,主要研究领域为人工智能、模式识别与蒙古文信息处理。
E-mail: csddl@imu.edu.cn



武慧娟(1981—),硕士,讲师,主要研究领域为模式识别、蒙古文信息处理。
E-mail: wuhj@imu.edu.cn

(上接第 74 页)

- [21] 蒋绍恩. 打击义动词的词义分析[J]. 中国语文, 2007(5): 387-401.
- [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)



王恩旭(1978—),博士后,主要研究领域为词汇语义学、计算语言学。
E-mail: wangbush000@126.com



袁毓林(1962—),博士,教授,博士生导师,主要研究领域为理论语言学和汉语语言学、计算语言学和中文信息处理。
E-mail: yuanyl@pku.edu.cn