

基于物性角色的同义名词辨析方法探讨^{*}

李 强 袁毓林

北京大学中文系/中国语言学研究中心/计算语言学教育部重点实验室

提要 同义名词辨析一直是研究词汇语义的学者们所关心的问题,但传统的辨析方式和手段在实际效果上很难令人满意,主要原因在于它们大体上还是沿袭一种基于意义的辨析路子,而意义是难以确切把握的。本文尝试利用生成词库论中的物性角色描写体系初步建立一个较为完整的同义名词描写框架;通过对 30 余组同义名词进行描写,我们总结了若干条带有普遍性的关于同义名词语义差异的规律,并展示了物性角色在同义名词辨析方面的优势;希望这种描写方法能最终应用到中文信息处理和国际汉语教学之中。

关键词 同义名词 词语辨析 物性角色 描写框架 语义规律

DOI:10.13724/j.cnki.ctiw.2014.04.010

一 引言

任何一种语言都有相当数量的同义词,对于母语学习者来说,他们能够比较清晰地掌握这些同义词之间的细微差别;可是对于第二语言学习者,如何能准确地辨析同义词成为他们在学习过程中的一个“拦路虎”,是一道难以逾越而又不得不跨越的屏障。所以,近些年来,无论是从事理论语言学研究的学者,还是处于教学第一线的教师,都对同义词给予了比较多的关注。比如,现有的几部较有影响力的同义词词典^①,收录的同义词规模都在上千组,基本上涵盖了较为常见的同义词组,是较为完备的工具书。国际汉语教学中对同义词的重视更不必多言,吴琳(2008)就指出:“同义词教学是中高级对外汉语词汇教学的重中之重,然而在教学实践中还存在着许多问题。”可见,如何探索并发展出一套比较完备的同义词辨析及教学方法是非常有必要的。

在同义词的整个系统中,同义名词又占据着举足轻重的地位,因为名词和动词是语言中

* 本课题的研究得到国家社科基金重大招标项目《汉语国际教育背景下的汉语意合特征研究与大型知识库和语料库建设》(批准号:12&·ZD175)的资助;《世界汉语教学》匿名审稿专家对本文初稿提出了很有价值的修改意见;硕士生刘彬、博士后研究员吴国向对文中的一些错误进行了修订。谨此一并致以诚挚的谢意。文中错误概由笔者本人负责。

① 《现代汉语同义词词典》(刘叔新主编,中国社会科学出版社,2011)、《同义词大词典》(程荣主编,上海辞书出版社,2010)、《现代汉语同义词词典》(朱景松主编,语文出版社,2009)。

最基本的词类范畴,名词作为一个开放的类,占据了词典里的大部分词项,在现代汉语词类的成员数量中,占了三分之二左右(郭锐,2002:210)^②。可惜的是,之前我们对同义词的关注大多集中在动词上,而对名词的关注力度则稍显不足。比如,刘春梅(2006)就曾指出:“像‘餐厅/食堂、波浪/波涛、船舶/船只’这样的表物同义词,至少教材和词典都还未曾问津。”而与目前研究现状恰好相反的是,同义名词的辨析对于国际汉语教学有着非常重要的作用。比如,第二语言学习者会经常发生这样的偏误^③:

(1)* 我们坐公共汽车到了一家宾馆的日本菜食堂,我吃了一杯拉面。(餐厅)

* 釜山挨着海洋。(大海)

* (那个城市)因为公路很宽,没有交通堵塞的情况。(马路)

很显然,之所以会发生这样的错误,是因为学习者并没有掌握“食堂—餐厅/海洋—大海/公路—马路”这三组同义名词之间的细微差别。

此外,同义名词在中文信息处理方面的应用也不可小觑。袁毓林(2008a、2008b)指出同义表达形式对于语义搜索、问题回答等信息检索系统有着非常重要的作用,因为对于机器来说,要理解自然语言的意义是非常困难的,而这又是自然语言处理过程中不可避免的问题;所以,如何形式化地、可操作地定义“意义”就成了一个非常重要的问题,而同义表达形式就是解决该问题的一个途径,只要机器能学会在若干同义形式之间自由地进行相互转换,就说明机器本身已经理解词语的意义;不过,同义表达形式不是等义表达形式,同样,同义词也不是等义词,同中有异、异中有同的特点决定了在自然语言处理过程中,我们不仅要关注同义词词义相同的一面,也要注意区分相异之处。例如:

(2)a. 那个厂家生产什么? b. 那个工厂生产啤酒。

如果计算机预先存储“厂家/工厂”这一对同义词,那么就可以从文本中找到跟问句 a 中的关键词虽然字面不同、但是意义相同的答句;但有时情况并非总是如此。比如:

(3)a. 那一带地区共有多少座工厂? b. ?那一带地区共有 30 座厂家。

如果文本中有“那一带地区共有 30 个厂家”这句话,计算机仅仅知道“个/座、工厂/厂家”是同义词恐怕还不够,它还要懂得“工厂”和“厂家”在词义上的区别,否则就会回答出 b 句这样并不太自然的句子,因为量词“座”一般并不修饰“厂家”。

本文主要探讨同义名词的词义辨析方法,以生成词库论中的物性角色框架为理论参照来多维度地考察同义名词之间的同和异。这样一种同义名词描写方式实质上是将词项放在具体的组合义场中,通过搭配来比较同义词的异同,相比于传统的辨析方法在深度和广度上都有所突破,并能对国际汉语教学和中文信息处理领域产生一定的影响。

二 传统的同义名词辨析中存在的问题

如何辨析同义名词一直是研究词汇语义的学者们所关心的问题^④。比如,张志毅、张庆云(2007:215)认为:名词同义词的主要差别表现在所指的人或事物的语义特征上,还常有义

② 在郭锐调查的 43330 个词中,名词有 27408 个,占 63%。

③ 例句取自刘春梅(2006)。

④ 之前的大部分文章著述讨论的都是辨析同义词的方法,涉及同义名词辨析的研究并不多。

域大小、普通和集合(词/词汇)、义位重点、陪义等方面的差别。周祖谟(1959:50、2006:40)指出辨析同义词,应从三方面入手:1.词义上的区别;2.风格色彩上的区别;3.用法上的区别。符淮青(2006:113-114)主张采用“种差+类词语”的分析模式对同义名词进行辨析,但这种分析模式依赖于对词语意义的正确把握。总体看来,传统的同义名词辨析虽然也涉及到词语搭配方面的差异,但没有交代应该从哪些方面对名词的搭配组合进行考察,缺乏一种可操作性的手段。它们更加注重从意义上对同义名词进行区分,但由于不同的词典编写者对于词义的认识有所不同,这就必然导致了同一个词条会出现不同的释义结果,所以,又可能会带来词义鉴别上的麻烦。符淮青(2006:114)对此就曾评论到:“从释义说明的内容很难看出这些词义的细微差别,甚至还可能有误解。”

另一方面,单纯的意义辨析虽然对于有一定汉语水平的母语者来说是有效的,但是,如果将这种意义辨析应用到二语学习者的语言学习之中,效果则不一定特别明显,因为对于二语学习者来说,词语之间意义上的细微差别是难以分辨和把握的。如果从词语的组合搭配方面入手,通过不断向他们灌输常见的组合搭配形式,使之成为心理词库(mental lexicon)的一部分,实际运用时只需要将特定的组合形式从词库中调动出来,这样一种基于分布的辨析路子对于二语学习者辨析同义名词可能更为快速和有效。

此外,即使是近些年在同义名词辨析中使用比较普遍的义素分析法(又称为“语义特征分析法”)也不断受到诟病。从事认知语言学研究的学者就曾对义素分析法产生了质疑(张敏,1998:60),比如,按照义素分析法,英语 bachelor(单身汉)的意义应该分解为[+成年]、[+男性]、[+未婚],但是,承担着抚养孩子义务的离婚男子、未娶正室却已纳妾的旧式中国男子、任何一个离婚或丧偶男子,操英语者在是否可以将他们称为 bachelor 的问题上并不确定。当然,我们可以通过增加语义特征来限定,但随之而来的问题是“按照这样的分析,需要有多少语义特征?语言分析真的需要数量如此多的语义特征吗?找出这么多的语义特征对描写语言真有什么好处吗?”(王洪君,2010)此外,王洪君(2010)还指出,区别性特征的语言学价值在于能够反映成分在组合搭配中的活动能力,特征分析绝不是为了分析而分析,不是为了更琐碎而分析,而是为了以少驭多,为了说明上级单位组合时活动方式的成组性。可见,义素分析法还是一种基于词义的描写方法,并不能很好地反映词语的组合搭配能力;如果将其应用到同义名词辨析之中,其结果和原来的基于意义的辨析并没有多少区别,只不过是有一些较为简洁的“二元对立组”呈现其语义差异,其实质仍然是“换汤不换药”,能起到的区别作用非常有限。

基于以上认识,我们认为同义名词的辨析,不仅需要关注意义方面的差别,更需要照顾到用法方面的差异;从语言运用的层面上看,后者显然更具有实际意义和可操作性,尤其是对二语学习者来说,在意义难以把握的情况下,将学习重心转移到词语的用法、搭配上,从搭配中发现同义名词之间的差异,进而再去探究产生这些差异的深层原因。这种方法可能更有利于二语学习者简单、快速地学习并掌握同义名词,同时也能为中文信息处理提供一定的便利。

三 生成词库论中的物性角色及其扩展描述

物性结构(Qualia Structure)是美国布兰代斯大学教授 Pustejovsky 等人倡导的生成词

库理论(Generative Lexicon Theory, GLT)中的核心内容,它旨在回答这样一个语言学问题:同一个词和不同的词组合会产生哪些不同的意义?这些不同的意义究竟是怎样产生的?比如,英语中“good”与“car/meal/knife”等组合形成“a good car/a good meal/a good knife”,若细细体味,我们会发现每一个词组中“good”的含义并不相同,分别为“开起来好、味道好、切东西好”。基于上面这种有趣的语言现象,同时受到亚里士多德的“四因说”(Aristotle's four causes of knowledge,即质料因、形式因、目的因和动力因)的启发,生成词库理论设计了一套词项物性结构的描写框架,主要包括四个层面的语义知识:

(1)构成角色(constitutive role):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)、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):描写对象在更大的认知域内区别于其他对象的属性,包括方位(orientation)、大小(magnitude)、形状(shape)、维度(dimensionality)和颜色(color)等;

(3)功用角色(telic role):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),人可以与某物发生直接联系,如 beer 的功用角色是 drink;另一种是间接功用角色(purpose telic),指某个事物可以用来协助完成某个活动,如 knife 的功用角色是 cut;

(4)施成角色(agentive role):描写对象怎样形成或产生的,如创造、因果关系等。

可以说,这四种物性角色展示了名词指称的实体所涉及的百科知识,极大地丰富了名词的语义内涵,较之以往的名词语义描写在广度和深度上都有所突破,对于深入刻画名词的语义有重要的指导作用;同时也反映了人类对名词的基本认识和所关切的内容。上述四种角色都是名词所涉及的基本角色,有着很强的心理现实性。受 Pustejovsky 的这套物性角色的启发,袁毓林(2013、2014)通过对大规模真实文本中的名词跟相关动词等的搭配关系和选择限制进行调查后发现,Pustejovsky 的四种物性角色并不足以全面反映名词跟动词等的搭配限制,于是对物性角色框架进行了更进一步的扩展描述,将物性角色的总数增加至 10 个,除了上面的四个物性角色外,其余六个分别是:单位、评价、行为、处置、材料、定位^⑤。通过这十种物性角色,我们可以对名词的语义进行更加全面和细致的描绘。

依据袁毓林(2014)对于十种物性角色的定义,我们可以通过检索语料库确定每一种物性角色所包含的对象和成员。一般来说,形式角色和构成角色是共有的、必须的,其余物性角色根据实际情况确定。也就是说,给出任意一个名词词项,我们都可以为它描写刻画相应的物性角色。这十种角色都是与名词自身密切相关的;换言之,要想对一个名词有全面的认识和掌握,仅了解名词本身的意义是远远不够的。掌握某个名词的意义可以为我们提供这个名词的使用环境,即在什么样的语言环境下用它;而究竟该怎么去用,则需要依靠该名词和其他词的搭配情况。我们认为,这样一种有条理的、系统的、完备的词汇描写方法可以应用到同义名词的辨析之中。

^⑤ 受篇幅限制,本文未给出每一种物性角色的定义,详细内容可以参看袁毓林(2014)。

四 关于同义名词描写框架的设想

一个较好的同义词的辨析框架,不仅有利于第二语言学习者快速、准确地掌握同义名词之间的差异,同时,还能够面向计算机,服务于中文信息处理。为此,我们参考了欧洲多语言句法—语义词库 LE-SIMPLE 的描写框架,并对其进行了一些修改,使之能够满足汉语同义名词辨析的需求。

SIMPLE 是由 EC DGXIII 资助的一个大型的语言工程项目,旨在用一个统一的模型来描写欧洲十二种语言的语义词库 (semantic lexicon) 信息^⑥,所涉及的内容包括本体类型 (ontological type)、论元结构 (argument structure)、语义框架 (semantic frame) 等。在 SIMPLE 所创建的模板结构中,共包括 17 个不同的条目,内容涵盖了词项的语义信息和句法配置信息,从而达到对词项较细的颗粒度描写,为下一步建立大规模的欧洲语言资源库和语言工程应用提供可靠的基础^⑦。

将上述面向语言工程的语义描写模型应用到同义词的辨析之中,会拓宽同义词研究的深度和应用性。根据汉语自身的特点,结合袁毓林 (2013、2014) 对于物性角色的扩展描述,我们对 SIMPLE 模型做出修改,建立了一个适用于描写汉语同义名词的框架体系,从而为国际汉语教学和中文信息处理提供服务。经过我们改动后的描写框架如下表所示:

表 1 同义名词词条描写框架

| 词条 | 需要标注的对象 |
|--------|--------------------------------------------------------------------------------------|
| 义项标号 | 《现代汉语词典》中相应义项的标号;如果仅有一个义项,则无需标号 |
| 拼音 | 标注词条的拼音 |
| 感情色彩 | 标注词条的感情色彩(褒义、积极、中性、消极、贬义) |
| 义域 | 标注该名词所在的语义场信息;可以参考《同义词词林》的三级分类体系 |
| 词典释义 | 标注《现代汉语词典》对于该词条的定义 |
| 述谓结构 | 描写该名词的降级述谓结构;如果无,则不描写 |
| 论元选择限制 | 描写具有降级述谓结构的名词所携带的论元的相关信息 |
| 形式角色 | 描写名词的分类属性、语义类型和本体层级特征,可以参考 Hownet、WordNet 语义资源以及北大计算语言所《现代汉语语义词典》等标注语义类 ^⑧ |
| 构成角色 | 描写事物在一个更大的范围内构成或组成其他事物;还可以描写根据若干属性特征对该名词所指称事物的分类 |
| 材料角色 | 标注构成事物的材料信息 |
| 单位角色 | 标注名词所指称事物的计量单位,包括度量词、名量词、动量词、时间量词等 |
| 评价角色 | 标注名词所指称事物的主观评价、情感色彩信息(多为形容词) |

⑥ 这十二种语言分别是加泰罗尼亚语、丹麦语、荷兰语、英语、芬兰语、法语、德语、希腊语、意大利语、葡萄牙语、西班牙语、瑞典语。

⑦ 关于 SIMPLE 的介绍及其所创建的模板结构,可以参看 Lenci et al. (2000)。

⑧ 《现代汉语语义词典》网址: http://klcl.pku.edu.cn:8080/seek/ccl_index.php; HowNet 网址: http://www.keenage.com/zhiwang/c_zhiwang.html; WordNet 网址: <http://wordnet.princeton.edu>。

| | |
|------|---------------------------------------------|
| 施成角色 | 标注名词所指事物的来源和形成途径(多为动词) |
| 功用角色 | 标注名词所指事物的用途(多为动词) |
| 行为角色 | 标注名词所指事物的行为动作(多为动词) |
| 处置角色 | 标注施加于名词所指事物的行为动作(多为动词) |
| 定位角色 | 标注以名词所指事物为坐标确定的方向和位置(多为介词和方位词) |
| 同义词集 | 标注与该名词同义的常用名词;可以通过查询相关的同义词词典获得 ^⑨ |
| 多义性 | 标注该名词可能存在的若干不同义项;可以参考《现代汉语词典》中的释义 |

可以看到,上述的描写体系涵盖了与名词关系密切的一些重要的语义信息,这可以服务于信息抽取、词义消歧、同义词辨析等自然语言信息处理。不过,与 SIMPLE 建立的描写框架既关注词条的语义信息、同时也囊括词条相关的句法配置信息不同的是,在上述构建的体系中,我们只对名词的语义信息进行了较为全面和完整的刻画,而不做句法配置上的描写。如果将来还有更进一步的需要,可以按照袁毓林(2013、2014)设计的描写框架,再把相关的句法配置增添进去,形成一个比较完整的句法—语义描写体系。此外,框架中还可以增加另外一些信息,这由实际需要而定。比如,可以增加与汉语词条相对应的英语翻译,可以增加词条的词频信息,还可以增加词条自身所携带的隐喻、转喻信息(如描写“北京”时,就可以增加“常常指代中国政府”这一条额外的转喻义)。

总之,基于上面的同义名词描写框架,我们可以为每一个名词建立一个内容丰富、信息详实的语义资源库,并利用名词描写框架对相关的同义名词进行比较,可以较为清晰地展示同义名词在相关语义栏目上的共性和差异。

五 从物性角色观察同义名词的语义差别

利用北大 CCL 现代汉语语料库和我们建立的同义名词描写框架,通过对 30 余组同义名词进行辨析,我们从中抽象出了一些高层的、带有普遍性的语义差别,列举如下。

(1)形式角色的不同反映同义名词具有不同的突显意义。比如,事业 vs. 产业,都是领域类名词,它们的意义之间有交叉部分,《现代汉语词典》(第 6 版)(以下简称《现汉》)对“事业”的释义是:“人所从事的,具有一定目标、规模和系统而对社会发展有影响的经常活动”,对“产业”的释义是:“构成国民经济的行业和部门”。相比较而言,“事业”的义域要宽于“产业”,“产业”的一些物性角色也适用于“事业”;同时,“事业”的语义类是一种社会领域,“产业”的语义类是一种经济领域,“事业”突出的是社会效益,具有公益性、福利性和非盈利性特征,而“产业”突出的是经济效益,具有盈利性、经营性和市场竞争性特征,这种差别导致了它们在物性角色上又存在不少区别。又比如,工厂 vs. 厂家,《现汉》对于“工厂”的释义为:“直接进行工业生产活动的单位,通常包括不同的车间”,对“厂家”的释义为:“指工厂”。可见,“工厂”和“厂家”的义域范围有交叉,它们在物性角色上有共同之处;同时,“工厂”的语义类是一种处所,突出自身作为建筑的实体属性,而“厂家”的语义类是一种机构,突出自身作为

^⑨ 比如:朱景松主编《现代汉语同义词词典》,语文出版社,2009;刘叔新主编《现代汉语同义词词典》(第四版),中国社会科学出版社,2011;程荣主编《同义词大词典》,上海辞书出版社,2010;梅家驹、竺一鸣等主编《同义词词林》,上海辞书出版社,1996。

组织单位的抽象属性,这导致了它们在物性角色上又存在诸多差别。类似的情况还包括公民 vs. 人民 vs. 人们、成果 vs. 后果、奥秘 vs. 秘密等同义名词组。

(2) 构成角色的不同反映同义名词不同的属性特征。名词的构成角色通常反映事物的构成和分类,比如,事业 vs. 产业,“事业”的构成角色“社会主义(事业)、民主(事业)、革命(事业)、改革开放(事业)”等具有鲜明的社会性、公共性和政治性,而“产业”的构成角色“工业(产业)、农业(产业)、石油(产业)、能源(产业)”等具有鲜明的经济性、生产性和盈利性,不过,这些词语仍能与“事业”搭配,是因为我们注重它们对社会发展的服务性功能。一些本身无所谓公共性质或生产性质的词语,如“电影、运输、旅游、保健”等,与“事业”搭配时强调非经济性,与“产业”搭配时注重经济性。以“电影事业/电影产业”为例,北大 CCL 语料库中与“电影产业”共现的名词一般是“集团、公司、市场、票房收入、资金、资本、营销策略、市场经济、竞争力、市场体系、文化产业、经济增长点”等,动词一般是“发行、开放、接轨、投资、商业化、操纵、融入、进军国际、合作”等,这些与之共现的名词和动词通常都出现于用来描述市场经济的句子中,具有极强的经济属性,这说明“电影产业”词语本身具有经济意味;而这些名词和动词无一与“电影事业”共现,说明“电影事业”词语本身具有非经济性。再比如,奥秘 vs. 秘密,“奥秘”的构成角色包括“宇宙(奥秘)、恒星(奥秘)、太阳(奥秘)、自然现象(奥秘)、发光(奥秘)、生理(奥秘)、微生物(奥秘)、遗传(奥秘)”等,而“秘密”的构成角色包括“经济(秘密)、军事(秘密)、外交(秘密)、政治(秘密)、商业(秘密)、核(秘密)、国家(秘密)、通信(秘密)”等,从构成角色可发现:“奥秘”具有描述自然界客观事物的特征,而“秘密”倾向于描述人为造成的结果。类似的情况还包括变故 vs. 事故、场合 vs. 场所、分界 vs. 分野等同义名词。

(3) 材料角色的不同反映名词所指对象和事物在所用材料方面的差异。比如,灰 vs. 灰尘,它们共有的材料角色都可以是“细土”,但“灰”还可以是一些粉末状的东西,它的材料角色包括“碳、烟、柴、纸、骨”,这些都不是“灰尘”的材料角色。又如,钱 vs. 钞票,“钞票”的材料角色是“纸”,但“钱”的材料角色可以包括“铜、金、银”等合金材料。

(4) 单位角色的不同反映同义名词所指对象和事物在构造、形状等方面的差异。比如,边疆 vs. 边界,“边界”可以受量词“段、条”修饰,是因为“边界”具有线性的特征,但“边疆”具有面的特征,所以不能受其修饰。再比如,工厂 vs. 厂家,“工厂”可以受“座、片、间”修饰,而“厂家”却不行,这是因为“工厂”具有处所、实体性质,而“厂家”是一个机构。类似的情况包括公民 vs. 人民、顾客 vs. 客户、木头 vs. 木材、思路 vs. 思绪等同义名词组。

(5) 评价角色的不同反映同义名词的不同感情色彩和性质。王治敏等(2005)指出:词语蕴涵的情感色彩对与其共现词语有所限制,与其共现词语往往也要求具有统一的情感色彩。比如,工厂 vs. 厂家,“工厂”的评价角色“锈迹斑斑、简陋、豪华、空荡、拥挤、废弃、花园式”都不能用来修饰“厂家”,是因为这些形容词都是用来描述物质实体的特征,而不能修饰抽象事物。又如,人民 vs. 人们,“人们”的评价角色“违法、险恶、不安好心、粗鲁、贪婪、不忠诚”等含有贬义色彩,不能用来描述含褒义色彩的“人民”,否则就会出现感情色彩的不协调。类似的情况还包括成果 vs. 后果、边疆 vs. 边境、场合 vs. 场所、气候 vs. 天气等同义名词组。

(6) 施成角色的不同反映同义名词所指对象和事物的不同来源途径。比如,公民 vs. 人民,“公民”作为一种身份,施成角色可以是“做、作为、身为、当、成为”,但“人民”不表示身份,因此没有这样的施成角色。又如,工厂 vs. 厂家,因为“工厂”是一个物质实体,所以它的施成

角色包括“建造、投资、兴建、盖”等,而这些角色都不能适用于表机构义的“厂家”。类似的情况还包括成果 vs. 后果、奥秘 vs. 秘密、处境 vs. 地步、高潮 vs. 热潮等同义名词组。

(7) 功用角色的不同反映同义名词所指对象和事物在功能和用途方面的差异。比如,事业 vs. 产业,虽然都是为了满足社会的物质文化需求,但它们各自的侧重点有所不同,“事业”更加强调对社会的政治、文化、生活等方面的影响,注重社会效益,以公益性为基本属性,以服务性为基本功能;而“产业”更加侧重对于经济发展的拉动和促进作用,注重经济效益,以盈利性为基本属性,以拉动经济为基本功能。这种差异导致“事业”和“产业”在功用角色方面存在一些差异,“产业”的功用角色集中在社会经济方面,在语料中一般与“增加收入、开辟财源、促进国民经济增长、增加社会财富”等含有经济性意味的词语搭配;而“事业”的功用角色集中在社会生活方面,在语料中常与“推动社会进步、促进社会发展、服务人民群众”等社会性意味较强的词语搭配。又如,契约 vs. 条约,“契约”的功用角色体现为经济领域的“确定租赁关系、证明出卖或抵押等关系”,而“条约”的功用角色则更多地体现为政治文化领域“促进人权、保护人权、确定合作关系”等。

(8) 行为角色的不同反映同义名词所指对象和事物在发展、变化方面的差异。比如,思路 vs. 思绪,虽然它们的意义几乎没有什么差别,但因为“思绪”强调人思考的头绪,是一种动态的过程,所以行为角色包括“停留、流动、飞向、集中、越过、驰骋、进入、回荡、纷至沓来、萦绕、遨游”等,而“思路”强调思考的条理脉络,是一种静态的状态,所以就没有这些行为角色。又如,言论 vs. 舆论,“舆论”因为强调的是一种新闻主体,行为角色包括“谴责、反对、监督、支持、赞誉、称赞、评论、认为、主张、猜测”等,而“言论”是主体发表的看法和意见,所以就缺乏这些行为角色。

(9) 处置角色的不同反映同义名词所指对象和事物在跟人的关系方面的差异。比如,工厂 vs. 厂家,因为“工厂”是物质实体,处置角色有“毁坏、没收、搬迁、捣毁、摧毁、轰炸、毁坏”等,这些都不能适用于“厂家”。又如,奥秘 vs. 秘密,因为“秘密”的人为性,其处置角色包括“保守、出售、侵犯、出卖、严守、保护”等,而“奥秘”倾向于描述自然界事物,所以没有这些处置角色。类似的情况还包括场合 vs. 场所、气候 vs. 天气、食物 vs. 食品、药材 vs. 药品等同义名词组。

(10) 定位角色的不同反映人或其他事物跟名词所指对象的不同位置和方向关系。比如,雨 vs. 雨水,“雨”的定位角色包括“里、前、后”等,“里”反映“雨”的空间方位特征,这是“雨水”也具有,而“前、后”反映“雨”的时间方位特征,这是“雨水”并不具备的。又如,饭 vs. 米饭^①,“饭”的定位角色包括“里、下(面)、上(面)、前、后”,“里、下(面)、上(面)”反映了“饭”的空间方位特征,这是“米饭”也具备的,但“前、后”不能与“米饭”搭配,说明“米饭”并不如“饭”那样具有时间方位特征。

符淮青(2000)提出:“同义词的同异,包括词位义、应用中的词义变化、用法的不同,都是在组合中实现的。在组合中分析同义词至关重要。”也就是说,仅仅通过观察同义词本身并不能清晰地洞察它们之间的同异,通过组合分布能够对词项的特征有更深刻的认识。基于

^① “饭”除了名词性义项外,还有动词性义项“吃饭”。我们这里比较的是它们作为名词的义项。感谢匿名审稿专家提醒作者注意这一点。

物性角色描写框架,我们可以列举同义名词在诸多方面的同和异,从而能更深层次地揭示同义名词在组合过程中的一些语义特点。

六 从物性角色归纳同义名词的语义特点

上一节主要是以物性角色为窗口,从微观上观察同义名词之间的语义差别,并抽象出若干条具有普遍性的语义规律,这种观察基本上属于一种横向的视角。本节将继续以物性角色为窗口,从宏观上纵向归纳同义名词的语义特点,同时反映同义名词在物性角色上的差别和联系。

(1)域概念的一致性是同义名词与不同词语进行组合搭配的基础。认知语言学认为句子中相关词语要保持“域的概念单一性”(conceptual unity of domain),即所有成分都必须放在一个域中来理解,否则就会发生隐喻(metaphor)和转喻(metonymy)现象。(Croft, 1993)在一个词汇或短语结构的内部,“域概念一致性”原则同样有效,比如,事业 vs. 产业,在构成角色中,“产业”的经济属性决定与它搭配的通常是在经济领域涉及商品生产的一些词语;能与“事业”搭配的一些词语虽然也涉及物质生产领域,但“事业”的公共性特征会使它们原本的经济属性弱化,比如,词语“工业产业”中的“工业”强调其对于社会、企业的经济拉动作用,但“工业事业”中的“工业”则注重其对于社会的公共服务性功能;在施成角色中,涉及商品生产行为的“开发”能与“产业”搭配,而很少与“事业”搭配。又如,工厂 vs. 厂家,因为“厂家”处于抽象的机构域中,所以要求与它搭配的物质角色也必须处于该域之中,如果是物质实体域的词语,就不能与之搭配,比如“*盖厂家、*修建厂家、*锈迹斑斑的厂家、*空荡的厂家、*毁坏厂家、*没收厂家”等。

(2)同义名词与其组合搭配的不同词语在属性上具有相似性。比如,产业 vs. 事业,“产业”的单位角色“串”,其语义和形状属性能够描述“产业”的发展模式和特征——呈现地理分布上的集聚或者传递性影响,而“事业”一般并不具备这样的特征,所以“串”很少与“事业”组合。又如,边疆 vs. 边界,“边界”的特征是线状,所以能被单位角色“段、条”、评价角色“清晰、漫长”修饰,但“边疆”的特征是面状,修饰它的评价角色通常是“辽阔、广大、广阔、广袤、宽广”。

(3)同义名词共有物性角色取决于它们高层语义类的一致性,而区别性物性角色取决于它们低层语义类的差异性。比如,“公民”和“人民”在高层语义类上都指人,但“公民”的低层语义类是人的身份,而“人民”的低层语义类是复数概念的人群,这决定了它们在单位角色和施成角色上存在差别。再比如,“财产”和“财富”的高层语义类都指钱财,但“财产”的低层语义类偏向于物质钱财,而“财富”的低层语义类偏向于精神钱财,这种差别决定了它们在构成角色和处置角色上有所不同。类似的情况还包括顾客 vs. 客户、分界 vs. 分野、规律 vs. 规则、过程 vs. 历程、时机 vs. 时刻等同义名词组。

(4)两个同义名词在很多物性角色上都体现出差异,有些物性角色的差异会蕴涵另一些差异。比如,形式角色的差异会蕴涵其他方面的差异,以“工厂”和“厂家”为例,“工厂”的形式角色是建筑,而“厂家”的形式角色是机构,这种差异导致了“工厂”的单位角色可以是“座、片、间”,施成角色可以是“盖、兴建、建造”,评价角色可以是“锈迹斑斑、空荡、简陋”,处置角色可以是“没收、搬迁、毁坏”,而这些都是“厂家”所不具备的。再如,施成角色的差异会蕴涵

其他方面的差异,以“药材”和“药品”为例,药材的施成角色包括“生长、盛产、种植”,“药品”的施成角色包括“制、研制、制成、发明、开发”,这种差别导致它们在单位角色上不同:“药材”的单位角色可以是“亩”,“药品”的单位角色可以是“盒、个、支”;也可以是处置角色的不同:“药材”的处置角色可以是“采集、采挖、采用、加工、粉碎”,“药品”的处置角色可以是“试验、验收”;也可以是评价角色的不同:“药材”的评价角色可以是“野生、天然、上等”,“药品”的评价角色可以是“合格、过期、假冒伪劣”。

(5)名词的类会影响同义名词的辨析,不同类的名词在物性角色上的差异显著度有所不同。比如,钱 vs. 钞票,“钱”的材料角色可以包括“铜、金、银”等合金材料,“钞票”的材料角色是“纸”,它们在材料角色上的差异相比于其他物性角色来说可能更为显著。又如,场合 vs. 场所,“场所”侧重于描述空间,它的施成角色包括“设立、建立”,处置角色包括“开放、恢复、封闭、捣毁、申请、登记、监控”,这些都不适用于“场合”;“场合”侧重于描述情景,它的评价角色包括“紧急、隆重、正式、死气沉沉、沉闷、喜庆”,这些也同样不适用于“场所”,它们在上述物性角色的差异更为显著。

总之,物性角色描写体系提供给我们的不仅是一种同义名词的描写方法,而且还是一种观察同义名词语义差别和联系的视角。在组合搭配中区分同义名词的原则体现了形式和意义之间的对应关系,相较于传统同义词辨析注重意义上的区别,形式分布无疑提供了更为可靠的依据和手段。

七 物性角色描写体系与传统词义描写方法在同义名词辨析方面的差异及优势

传统的词义描写更多地关注词语本身的意义和内涵,这对于同义名词辨析具有重要的指导作用,但还远远不够,因为“词典释义说明的只是词所代表的事物的一部分特征,其他特征,特别是那些对本族人来说是不言而喻的特征都省略了”。(符淮青,2006:64)若要达到对词语的精细化描写,就需要全方位地对词语的概念结构进行考察。总结现有的几部较有影响力的同义词词典,我们发现它们对于同义名词组描写的一般架构大致都是:1)标注词项的词性和共同语义;2)分别描述词项的不同语义并列举几个句子加以展示。比如,对于“场合/场所”的辨析,朱景松编《现代汉语同义词词典》的描写方式如下:

【同】名词。都表示特定的处所。

【辨】1. “场合”表示说话、办事的全部环境,包括处所、时间、在场的人物等。

(1)他自负最能适应环境,对什么人、在什么场合,说什么话。

(2)她挺机灵,知道要应付这种场合,她还缺乏经验。

“场所”表示特定的处所。

(3)这儿是医院,公共场所。

(4)茶馆又是人们交际应酬的场所。

2. “场合”多为泛指,甚至侧重指说话、办事的条件。

(5)在很多场合,他都会有上乘的表现。

(6)中国人自己的事情,中国人自己去办,没有必要在国际会议的场合安排解决。

“场所”表示供人活动的具体场所。

(7)希望大家今后常来,充分利用这个场所。

(8)他们办公有办公地点,娱乐有娱乐的场所,住家专做住家之用。

这样一种描写方式存在以下几个缺陷:1)对于同义词相同点的概括比较简单,仅仅关注词性和意义上的共性,缺乏对词语其他方面,尤其是组合能力的关注;2)相异点概括得不够简练,缺乏清晰性。比如,上面的【辨】中,1和2的内容其实很大程度上是相互重合的,只需指出一点即可;3)对于相异点大多从意义方面给予区别,而未能真正触及使用层面的差异,因此也未能清晰地指明不同词项的具体使用环境有何区别;4)对于词典的查询和使用者来说,这样的描写方式不直观、不方便使用;5)由于这种描写方式和布局的随意性,缺乏系统和规划,所以很难应用到自然语言处理领域。

物性角色描写体系旨在扩充词项的具体内涵,深化词项的知识结构,主张构建单个词项的“本体知识”(Ontology),从而为更加细致的语言描写和自然语言处理服务。正如第四节中同义名词描写框架所展示的,将这样一种较大规模的描写体系应用到同义名词的辨析之中,会较为清晰、全方位地展示同义词项之间的联系和差别,所涵盖的内容信息要远远超过一般同义词词典的描写;更重要的是,这样一条条的语义信息被安排在一个具有层级性的结构(hierarchical structure)之中,并且明确地展示了词项的全部内容,可以直接并且有选择性地服务于 NLP 应用。

此外,就同义词词典编纂工作来说,物性角色的描写体系也应能够发挥重要的作用。近些年来,关于二语习得中的词汇语义研究一直得到比较多的关注,不同学者从不同的角度都曾指出:要真正掌握一个词,需要掌握不同方面的词汇知识,这些不同方面的知识共同构成了一个“词汇知识框架”(Vocabulary Knowledge Framework);虽然不同学者对于该框架所应包含的具体内容看法不一,但一致认为应该包括:词的口语形式、词的书写形式、词的语法行为、词的搭配形式、词的意义等等。这种词汇知识框架为同义词典编纂提供了良好的依据,但现有的几部同义词词典在某些方面的工作依然还很薄弱,尤其是关键性的词语的搭配知识,它们无一能够详细而又全面地进行描写;所以,物性角色的描写框架提供了一个强大的工具,可以利用它来全面地描写词语的组合搭配。另外,通过对北大 CCL 语料库进行检索,可以得到相对较为完整、可信的词语使用情况。这样一种基于语料库、全面而细致的条目化描写方式可以为同义词词典的编纂工作提供一些宝贵的经验,从而为编出一本高质量的同义词词典提供相应的理论支撑。

八 服务于中文信息处理和国际汉语教学的同义名词辨析的 可视化计算机操作界面的设想

哈工大信息检索实验室在《同义词词林》的基础上发表了《哈工大信息检索研究室同义词词林扩展版》,并开发了一套在线的同义词查询系统,供用户查询与目标词同义的词语。但这套系统只是罗列出较多的同义词,用户并不能从该系统中获取同义词之间的相关差异,这不能不说是一个遗憾!而目前,市面上虽然已经出现了不少纸质本的同义词词典,且收词数量具有相当的规模;但在线版的同义词查询系统仍然寥寥无几,这可能一方面出于经费的缺乏,另一方面因为整理同义词的任务量比较繁重。我们认为,针对同义词,应该主动出击、各个击破,分别找到对名词、动词、形容词、副词等不同词类词语的最佳描写方法,从而将其利用到同义词的辨析之中。

具体到本文,生成词库理论中的物性角色描写体系就是这样一种能对同义名词进行有效辨析的方法,同时结合欧洲多语言句法—语义词库 SIMPLE 对于词条的描写,我们建立了如第四节表 1 所示的同义名词描写框架,基本上涵盖了名词所有的语义信息,能较为全面、多视角地对同义名词进行辨析。相对于传统的同义名词辨析方法,这样一种描写体系可以更加方便地应用到计算机上,从而为设计一个可视化的计算机查询界面奠定基础。具体来说,这种可视化界面的设计思路可以分为:1)当用户输入待查询词项时,计算机可以检索出如第四节表 1 所示的具体词项的描写框架的信息;2)在描写框架中的“同义词集”一栏中,用户可以点击其中一个同义词,计算机呈现该同义词的描写框架信息;3)在屏幕的侧栏可以添加标识为两个同义名词比较的标签,用户点击该标签即可获得用名词描写框架展示的相同点和不同点的图示化结构;4)设置相应的例句标签供用户点击查询同义名词所出现的真实文本中的句子。这样一种同义名词辨析的可视化界面会给用户,尤其是汉语学习者提供很大的便利,相较于传统的同义词词典,这种结构化的设计将词条的语义信息有规则地排列起来,使查询者不仅能对某一具体词项的句法语义信息有较为全面的了解,同时也能对同义名词之间的同和异有比较好的掌握;从使用者的角度来说,清晰、简洁、有条理的展示界面无疑能增加他们学习的效率和积极性。

综合上述内容,我们认为:物性角色的描写体系为我们提供了一个观察名词的窗口,利用该体系不仅可以对单个名词进行描写,也可以对多个同义名词进行比较;在充分描写的基础之上,进而将成果移植到计算机上,形成一个交互式的查询系统,可以最终服务于自然语言处理和国际汉语教学。

参考文献

- 程 荣主编(2010)《同义词大词典》,上海:上海辞书出版社。
- 符淮青(2000)同义词研究中的几个问题,《中国语文》第 3 期。
- 符淮青(2006)《词义的分析与描写》,北京:外语教学与研究出版社。
- 郭 锐(2002)《现代汉语词类研究》,北京:商务印书馆。
- 刘春梅(2006)现代汉语表物同义名词研究,《云南师范大学学报》第 2 期。
- 刘叔新主编(2011)《现代汉语同义词词典》(第四版),北京:中国社会出版社。
- 梅家驹、竺一鸣等编(1996)《同义词词林》,上海:上海辞书出版社。
- 王洪君(2010)从两个同级义场代表单字的搭配异同看语义特征和语义层级——以“锅”和“碗”为例,《世界汉语教学》第 2 期。
- 王治敏、朱学锋、俞士汶(2005)基于现代汉语语法信息词典的词语情感评价研究, *International Journal of Computational Linguistics and Chinese Language Processing (CLCLP)* 4: 581—592。
- 吴 琳(2008)系统化、程序化的对外汉语同义词教学,《语言教学与研究》第 1 期。
- 袁毓林(2008a)面向信息检索系统的语义资源规划,《语言科学》第 1 期。
- 袁毓林(2008b)用同义表达形式来扩充信息检索的查询语句例证研究——对于一种基于语义的搜索方式的若干设想,《语言文字应用》第 2 期。
- 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。
- 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。
- 张 敏(1998)《认知语言学与汉语名词短语》,北京:中国社会科学出版社。
- 张志毅、张庆云(2007)《词汇语义学与词典编纂》,北京:外语教学与研究出版社。

- 中国社会科学院语言研究所词典编辑室编 (2012) 《现代汉语词典》(第 6 版), 北京: 商务印书馆。
- 周祖谟 (1959) 《汉语词汇讲话》, 北京: 人民教育出版社。
- 周祖谟 (2006) 《汉语词汇讲话》, 北京: 外语教学与研究出版社。
- 朱景松 (2009) 《现代汉语同义词词典》, 北京: 语文出版社。
- Croft, William (1993) The role of domains in the interpretation of metaphors and metonymies. *Cognitive Linguistics* 4: 335–370. In Dirk Geeraerts (ed.), *Cognitive linguistics: Basic readings*, 269–302. Berlin: Mouton de Gruyter, 2006. (中译本, 《认知语言学基础》, 邵军航、杨波译, 上海: 上海译文出版社, 2012 年)
- Lenci, Alessandro, Nuria Bel, Federica Busa et al. (2000) SIMPLE: A general framework for the development of multilingual lexicons. *International Journal of Lexicography* 13: 249–263.
- Pustejovsky, James (1991) The generative lexicon. *Computational linguistics* 4: 409–441.
- Pustejovsky, James (1995) *The generative lexicon*. Cambridge: MIT Press.

A Qualia Role Based Approach to the Analysis of Chinese Synonymous Nouns

Li Qiang Yuan Yulin

Abstract Discrimination of synonymous nouns has been the concern of scholars studying lexical semantics. However, a traditional meaning-based analysis is unsatisfactory, because meaning itself is elusive. This paper adopts qualia roles in generative lexicon theory, with the aim to establish a relatively complete framework of synonymous noun description. Through describing more than 30 pairs of synonymous nouns, we come up with several common regularities in the semantic information about synonymous nouns. The analysis exhibits the advantages of qualia roles in differentiating synonymous nouns. We hope the framework as proposed here can be eventually applied to Chinese information processing and international Chinese teaching.

Keywords synonymous nouns, lexical differentiation, qualia role, description framework, semantic regularities

作者简介

李强, 男, 北京大学中文系博士生。主要研究兴趣为汉语句法学、语义学和语用学。[Email: leeqiang2222@163.com]

袁毓林, 男, 北京大学中文系教授、博士生导师。曾在《中国社会科学》、《中国语文》、《当代语言学》、《世界汉语教学》和《中文信息学报》等刊物上发表论文 100 余篇, 出版专著 8 部。主要研究理论语言学和汉语语言学, 特别是句法学、语义学、语用学, 也涉及计算语言学和中文信息处理等应用性领域的问题。[Email: yuanyl@pku.edu.cn]