

# 国内生成词库理论研究的回顾与展望\*

李 强

(上海大学 中文系, 上海 200444)

**[摘 要]** 近些年来,生成词库论发展极为迅速,研究规模日益扩大。生成词库论视角下汉语研究的相关工作也在不断向前推进,取得了不少值得关注的研究成果。文章对目前汉语方面的研究成果进行回顾和总结,分别从语言本体、语言认知、语言计算和语言教学四个维度展示生成词库论在汉语上的研究和应用。对于汉语中一些不好描写和解释的旧问题,生成词库能够提供一种新的观察视角,同时对于扩大和深化汉语的相关研究课题也有一定的帮助。

**[关键词]** 生成词库论;语言本体;语言认知;语言计算;语言教学

**中图分类号:** H14   **文献标识码:** A   **文章编号:** 1672-1306(2018)01-0055-15

## 一、引 言

兴起于20世纪90年代的生成词库理论(generative lexicon theory, GLT)被认为是“当代语义学中最精细的形式化分析范式”<sup>①</sup>,由美国布兰代斯大学(Brandeis University)Pustejovsky教授创立。受亚里士多德关于事物“being”(存在)这一问题探讨的启发,他将目光聚焦到语言中的词语在动态语境下的使用情况:为什么不同的词语在不同的语言环境下会产生不同的意义?不同词语之间是靠什么组合在一起的?有没有一套普遍的描写和解释手段可以在人们的语言知识(language knowledge)和非语言知识(即世界知识, world knowledge)之间建立起联系的纽带?基于对以上这些问题的思考, Pustejovsky试图建立一套既能在语言理论方面建构模型,又能在语言应用方面解释实际问题的框架体系。

生成词库创立至今已经有将近20年的时间,英语 WordNet 的主要研制者 Fellbaum<sup>②</sup>曾给予生成词库如下的评价:“自从乔姆斯基的《句法结构》问世以来,语言的生成性已被广泛认识,但直到 Pustejovsky 的生成词库理论提出后,生成性观念才被运用到词库的研究中。”可见,生成词库在词汇研究这一领域开创了一片新天地,开垦出了一片前人都未曾涉猎过的新沃土,以至于在 Geeraerts<sup>③</sup>所著的《Theories of Lexical Semantics》一书中,专门用了11页(165~176)来对该理论进行介绍,足见该理论在词汇语义研究中的重要地位。

近年来生成词库论视角下的汉语研究取得了不少值得关注的研究成果。比如, Huang & Ahrens<sup>④</sup>

\* 作者简介:李 强,男,安徽巢湖人,上海大学讲师,研究方向为汉语语法学。

基金项目:国家重点基础研究973计划项目“语言认知的神经机制”(2014CB340502);教育部人文社会科学重点研究基地重大项目“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”。

① Geeraerts, Dirk. Theories of Lexical Semantics[M]. New York: Oxford University Press Inc, 2010.

② Fellbaum, C. The Generative Lexicon by James Pustejovsky[J]. Language, 1997, (3).

③ Geeraerts, Dirk. Theories of Lexical Semantics[M]. New York: Oxford University Press Inc, 2010.

④ Huang, C.-R. & Ahrens, K. Individuals, kinds and events: classifier coercion of nouns[J]. Language Sciences, 2003, (25).

以“强迫”为主线系统地描写了汉语中量词对名词的语义强迫现象;Liu et al.<sup>①</sup>、Lin & Liu<sup>②</sup> 讨论了汉语中的动词强迫现象;张秀松<sup>③</sup>从生成词库角度解释了词的逻辑多义性问题;宋作艳<sup>④</sup>最为系统地利用生成词库论来研究汉语中“事件强迫”(event coercion)。此外还包括袁野<sup>⑤</sup>、张秀松等<sup>⑥</sup>、李强<sup>⑦</sup>等的研究,都从不同的角度层面对汉语中的相关语言现象进行了探讨。本文根据近些年的研究热点和方向着重介绍生成词库论在汉语方面的研究,分别从语言本体、语言认知、语言计算和语言教学四个维度展示相关研究成果。对于汉语中一些不好描写和解释的旧问题,生成词库能够提供一种新的观察视角,同时对于扩大和深化汉语的相关研究课题也有一定的帮助。

## 二、生成词库论概述

生成词库论的产生最初源于 Pustejovsky 对常规多义关系的关注。比如:

1) a. I put a glass on the table.

b. The present is made of glass.

上面两句中, a 中的 glass(玻璃杯)是可数物体,而 b 中的 glass(玻璃)是不可数物体。所以, glass 的这两个意义就是一种常规的多义关系, Pustejovsky 称此为“逻辑多义关系”(logical polysemy)。

在处理这种多义关系上,传统的“词汇意义计算”(sense enumerative lexical, SEL)模式限定了每一个词项  $\alpha$  只有一个意义,从而排除了多义词的存在;也就是说,若词项  $\alpha$  具有不同的意义,那么,该模式就将  $\alpha$  分解为  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$ , 每一个  $\alpha$  对应一个不同的意义。对此, Pustejovsky 指出了其弊端,他认为,从方法论上看,这种词义处理模式能够较好地解决词项在不同语言分布环境下产生不同的意义这一问题,尤其是能够对同一个动词和不同的名词组合时为什么能形成不同论元结构给出解释。但是,从词库的角度来说,如果每一个不同的意义都负载于一个词项之上,那么该词库所包含的词项数量是非常巨大的。我们不禁要问:人们究竟该如何学习并掌握如此大规模的词库?

所以,出于经济性(economy)和可学习性(learnability)的考虑,可以把这些不同意义归结于词语的不同用法。那么,随即而来的一个问题是:词语如何在不同的语境下产生某一个特定的意义?甚至在某些语境下词语可能获得一个全新的意义,这又是怎么实现的?为此,生成词库论主张词库不仅要列出词语的意义,还要解释词语的动态性、创造性使用,主张词语的意义是组合的、动态的、生成的。作为个体的词项,它的意义是相对稳定的,但进入到句子层面后,在上下文中可以获得一些创新性用法,而这些创新性用法都可以通过若干程序获得,该程序提供了每个词项在该系统中的编码知识,主要是物性角色和语义类型以及与该知识相联系的一些组合生成机制。

物性角色(qualia role)的提出源于亚里士多德的“四因说”(Aristotle's four causes of knowledge): 质料因、形式因、目的因和动力因。它是生成词库论中最为核心的内容,反映了人们对于事物的基本认

① Liu, Mei-chun, Chu-Ren Huang, Ching-Yi Lee. Lexical information and beyond: constructional inferences in semantic representation[R]. Proceedings of the 13th Pacific Asia conference on language, information and computation, 27~38. Taipei, Taiwan, 1999.

② Lin T.-H., Jonah and Liu C.-Y., Cecilia. Coercion, Event structure, and syntax [J]. Nanzan linguistics, 2004, (2).

③ 张秀松. 从生成词库论看汉语词的逻辑多义性[J]. 北方论丛, 2008, (3).

④ 宋作艳. 生成词库理论与汉语事件强迫现象研究[M]. 北京: 北京大学出版社, 2015.

⑤ 袁野. 构式压制、转喻和广义转喻框架[J]. 外国语言文学, 2010, (3).

⑥ 张秀松, 袁毓林, 陈振宇, 周强. 基于情境网络的汉语亲属名词的语义描写体系[J]. 世界汉语教学, 2011, (3).

⑦ 李强. 从生成词库论看动词“写”与名词的组合[J]. 云南师范大学学报(对外汉语教学与研究版), 2015, (2).

识和感知,主要涉及4个层面的语言知识(Pustejovsky 1991,1995)<sup>①</sup>:

构成角色(constitutive role):描写一个物体与其组成部分之间的关系;

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;

功用角色(telic role):描写对象的用途和功能;

施成角色(agentive role):描写对象怎样形成或产生的。

这4种物性角色弥合了语言知识与百科知识之间的缺口,从而为语言的可计算特征提供了一套百科知识式的词汇描写方法。

就词语本身的性质而言,Pustejovsky(2001,2006)<sup>②</sup>根据词语所代表的意义内容构建了一个三分(tripartite)的知识分类体系,即语义类型,包括:

自然类(natural type):与物性结构中的形式角色和构成角色相关的概念,谓词来自物质域。比如:石头、水、兔子、天空、树;

人造类(artifactual type):与物性结构中的施成角色和功用角色相关的概念,谓词与这两个角色联系。比如:医生、妈妈、椅子、桌子、刀;

合成类(complex type):常被称为“点对象”(dot object),由自然类和人造类组成的概念。比如:书、报纸、杂志、音乐会、唱片。

这样一种三分体系有别于语言计算领域常见的知识本体(ontology),比如 WordNet<sup>③</sup>,尤其对于事物实体而言,构建上述分类系统能够将那些具有多义性的词语进行恰当的归类,从而将内在的不同义面(semantic facet)进行有效地表征。比如,“书”就是一个多义词,在“尽信书就不如无书”中,前一个“书”代表信息内容,后一个“书”代表物质实体;所以,“书”应该被划入到合成类之中。而这在 WordNet 是较难处理的,“书”对应于 WordNet 名词分类体系中的 artifactual;不过,这种分类很显然并没能将“书”的不同义面区分开来。

在物性角色和语义类型的基础上,生成词库阐述了词项组合的几种不同机制,主要包括:纯粹类型选择(pure type selection)、类型调节(type accommodation)和类型强迫(type coercion)(Pustejovsky 2001,2006,2011)。<sup>④</sup>其中,尤其值得注意的是“类型强迫”,它是指:“它是一种语义操作手段,能将论元转变成一种功能函项(谓词)所需要的类型,否则就会出现类型匹配错误。”

在此基础上,Pustejovsky<sup>⑤</sup>对“类型强迫”做了更为具体和细致地扩展描写和分类,使之具有更强的解释力。他将“类型强迫”分为两类,分别为“强迫引入”(coercion by introduction)和“强迫利用”(co-

① Pustejovsky,J.The Generative Lexicon [J].Computational linguistics,1991,(4);Pustejovsky,J.The Generative Lexicon [M].Cambridge: MIT Press,1995.

② Pustejovsky,J.Type Construction and the Logic of Concepts[A].In Pierrette Bouillon & F.Busa (Eds.),The Language of Word Meaning[C].Cambridge University Press,2001;Pustejovsky,J.Type theory and lexical decomposition [J].Journal of Cognitive Science,2006,(6).

③ 关于 WordNet 的相关内容,可参考 Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database [M]. Cambridge, Massachusetts: MIT Press,1998.

④ Pustejovsky,J.Type Construction and the Logic of Concepts[A].In Pierrette Bouillon & F.Busa (Eds.),The Language of Word Meaning[C].Cambridge University Press,2001;Pustejovsky,J.Type theory and lexical decomposition [J].Journal of Cognitive Science,2006,(6);Pustejovsky,J.Coercion in a general theory of argument selection[J].Linguistics,2011,(6).

⑤ Pustejovsky,J.Coercion in a general theory of argument selection[J].Linguistics,2011,(6).

ercion by exploitation)。

强迫引入(coercion by introduction):把论元包装成函项所要求的类型,这是通过引入名词的相关物性角色实现的。

强迫利用(coercion by exploitation):提取论元类型的一部分来满足函项的要求,这是通过利用名词的相关物性角色实现的。

上述两种强迫机制可以解释下面的语言现象:

2)a. John enjoyed the book.

b. Mary read the rumor.

在例2)a中,动词“enjoy”在语义上要求它的宾语成分必须是动作性的,但“book”并不满足该语义要求。于是,“enjoy”强迫“book”产生“read the book”的语义解读,而这是通过利用“book”的功用角色“read”实现的。这就是“强迫利用”。在例2)b中,动词“read”要求它的宾语必须具有 *phys*(物质实体)的属性特征;但是,名词 *rumor* 却只有 *info*(信息内容)的属性特征。于是,“read”可以直接将 *phys* 属性特征引入到名词 *rumor* 之上。这就是“强迫引入”。

不同组合机制有效地解释了词项为什么能够组合在一起以及组合之后意义的生成性问题,特别是对那些意义不完整(incomplete)或限定不足(under-specification)的句子而言,生成词库论提供了一种解释手段,有力地说明了动态语境下词语意义的生成和变化以及句子最终浮现出合适语义表达的过程。

### 三、语言本体研究

运用生成词库论对汉语本体进行研究是目前关注度最大、研究成果最多的一块领域,其研究课题涉及词语构造方式、语义组合与强迫和句法格式语义还原等内容。

#### (一)词语构造方式

Song & Zhao<sup>①</sup> 和 Song & Qiu<sup>②</sup> 利用生成词库论中的物性角色对汉语中的双音节和三音节复合词进行了语义关系及语义信息的标注,总体上包括3种不同的情况:

1. 复合词的一个语素是另一个语素的某种物性角色。比如,前一个语素是后一个语素的形式角色,包括“[玻璃]纸、[塑料]纸、[干]花、[喇叭]花”等;前一个语素是后一个语素的构成角色,包括“[油]纸、[砂]纸、[桃]花、[玫瑰]花”等;前一个语素(如果是名素则隐含了一个动素)是后一个语素的功用角色,包括“[墙]纸、[复印]纸、[蒸]锅、[救护]车”;前一个语素(如果是名素则隐含了一个动素)是后一个语素的施成角色,包括“[剪]纸、[道林]纸、[石]灰”等。<sup>③</sup>

2. 在含有动词性语素的复合词中,动词性语素是整个复合词的某种物性角色。比如“石雕、冰雕、石刻、木刻”中,动词性语素“雕、刻”是 NV 复合词的施成角色,“蒸锅、砍刀、救护车、饮用水”中的动词性语素“蒸、砍、救护、饮用”是 VN 复合词的功用角色,“作文、画像、炒面、来信”中的动词性语素“作、画、炒、来”是 VN 复合词的施成角色。

3. VN/NV/VV 动词性结构是 VN/NV/VV 式复合词的某个物性角色。比如,“管家、围墙、兜肚”,

① Song Zuoyan & Zhao Qingqing. Annotating Qualia Relations and Types in Chinese Compound nouns [J]. International Journal of Knowledge and Language Processing, 2013, (3).

② Song Zuoyan & Qiu Likun. Qualia Relations in Chinese Nominal Compounds Containing Verbal Elements [J]. International Journal of Knowledge and Language Processing, 2013, (1).

③ “道林”是最早生产道林纸的公司,体现了这种纸的来源;“石灰”是石头煅烧而产生的,“煅烧石头”产生“灰”。



VN就是复合名词VN的功用角色；“剪纸、剪报”，VN就是复合名词VN的施成角色；“天赋、心得、火烧”，NV就是复合名词NV的施成角色；“编辑、装饰、导演”，VV就是复合名词VV的功用角色，“雕刻、记录”，VV就是复合名词VV的施成角色。

Song & Zhao<sup>①</sup>更进一步地对隐喻式名名复合词进行了物性角色方面的考察。在名名复合词中，修饰语名词和中心语名词都可能会发生隐喻现象。比如，“楼”的形状像“塔”，所以称为“塔楼”；“耳朵”和“大脑”的关系近似于“侧房”和“正房”的关系，所以“侧房”被称为“耳房”；“房子”的功能是提供居住的处所，所以具备这一功能的车子被称为“房车”。这些名名复合词中的修饰语名词产生了隐喻义；并且，这种隐喻义分别是在形式角色、构成角色和功用角色的基础上产生的。此外，中心语名词也会发生隐喻。“雪”的形状像“花”，所以称为“雪花”；“脚”和“人”的关系近似于“页面底部”和“页面”的关系，所以“页面底部”称为“页脚”；小脑中有一块连接左右两侧的组织，因为功能类似于“桥”的连接作用，因此称为“脑桥”。在这些名名复合词中，都是中心语名词产生了隐喻义；并且，这种隐喻义也同样分别是在形式角色、构成角色和功能角色的基础上形成的。

以生成词库视角来探讨汉语复合词构词语素之间的关系，这属于构词法领域的研究。这种研究有别于我们传统上对于词语构造的一般性认识。过去的构词研究往往根据语素的数量将词语分为单纯词和合成词，合成词进一步分为复合式合成词和派生式合成词，复合式里面再根据语素之间的语义语法关系分为联合式、主谓式、述宾式等，派生式复合词再进一步研究词根和词缀之间的关系等。<sup>②</sup>这种关于词语的构造研究大体上还是基于形式或形态上的，对于词语中不同语素之间的关系，特别是语义关系，关注得还不够，而生成词库视角下的研究则在一定程度上弥补了这个方面的不足。

此外，按照物性角色可以对词语中的语素加以标注，进而将语素之间的语义关系清晰地表示出来；通过对词语大规模的语义信息标注，可以在此基础上抽象出一套语素组合的模板，最终实现机器对词语的自动语义分析和剖析。同时，在汉语作为第二语言的词汇教学领域，利用物性角色对词语所做的标注可以用来说明和解释词语的意义，以便汉语学习者能更为清晰和透彻地掌握词语的意义。这些都是传统的词义分析方法难以解决或者不好解决的问题。

## (二) 语义组合与强迫

语言成分的语义组合性问题一直以来都是语义学研究的重要课题。从生成词库论的角度来研究词语的语义之间是如何组合在一起的，这是一个颇为新颖的尝试。

Wang & Huang<sup>③</sup>结合名词物性角色、语义类型和组合机制分别对汉语中的动词和形容词与名词的组合进行了考察。在动名组合中，他们以动词“看”为例，尽可能多地搜集了“看NP”短语，然后根据“看”的不同意义分为不同的情况，以不同的组合机制为条目依次对“看NP”进行了细致的研究。在形名组合中，他们分两种情况：(1)说明形容词“长”可以修饰许多不同类型的名词，并且具有潜在的多义

① Song Zuoyan & Zhao Qingqing. Qualia Relations in Metaphorical Noun-Noun Compounds[R], Proceedings of the 6th International Conference on Generative Approaches to the Lexicon, September 24~25, Pisa, Italy, 2013.

② 张斌. 现代汉语描写语法[M]. 北京: 商务印书馆, 2010.

③ Wang, Shan & Huang Chu-Ren. Adjectival Modification to Nouns in Mandarin Chinese: Case Studies on “cháng+noun” and “adjective+tú shùguān”[R], Proceedings of The 24th Pacific Asia Conference on Language, Information and Computation (PACLIC24). November 4~7, Sendai, Japan, 2010a; Wang, Shan & Huang Chu-Ren. Compositional Operations of Mandarin Chinese Perception Verb “kàn”: A Generative Lexicon Approach[R]. Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC 24). November 4~7, Sendai, Japan, 2010b.

性。比如,“长隧道”中,“长”修饰作为事物的个体,说明的是“隧道”的形式特征;“长午餐”中,“长”修饰作为事件的行为,说明的是“午餐”所持续的时间。(2)以名词“图书馆”为核心,考察可以修饰它的形容词。不同的形容词与“图书馆”组合时,突显其不同的物性角色。比如,“大”突显形式角色,“古老”突显施成角色,“好”则可能突显形式、构成、功用和施成角色中的任何一个。<sup>①</sup>可见,动词、形容词与名词的语义组合中涉及名词的物性角色以及不同的组合机制。

李强<sup>②</sup>在上述研究的基础上对汉语其他动词与名词的组合进行了考察,包括动词“读、写、听”。通过对这些动词和名词所形成的动宾结构的研究,一方面揭示了它们在语义组合上的不同机制;另一方面指出了宾语名词受到动词的语义强迫作用发生转喻现象,名词转喻义的生成是通过物性角色和组合机制共同实现的。

在语义组合性的探讨中,最为人所熟知的无疑是“弗雷格原则”(Frege's Principle),它给出了语义组合与生成所遵循的最基本的规律,即“一个语言表达式的意义是其组成成分的意义及它们句法组合方式的功能函项”。但是,并不是所有的语言现象都遵守这条原则。很多情况下,整个结构的语义超出而并不等同于其组成成分的语义之和。其中,最为典型的就如例2)所示的“类型强迫”现象。

很多语言都存在语义强迫,汉语自然也不例外,并且体现得更为充分和明显。从汉语词汇层面的语义强迫来看,主要包括:名词(包括名词性词缀)强迫、动词强迫、形容词强迫、副词强迫、方位词强迫和量词强迫。

名词强迫现象可以分为两种情况:(1)体现在名词和名词性词缀构成的词语内部;(2)体现在名词和名词构成的复合词内部。在词语或复合词内部,其中的一个名词(名词性词缀)强迫利用另一个名词的一个物性角色使整个结构的语义得到还原和充盈。前一种情况如宋作艳<sup>③</sup>所讨论的由名词性词缀“家、夫、手、员、民、计、场、热、风”等构成的词语。比如,“钢琴家”表示“[弹]<sub>功用</sub>钢琴的人”;“车夫”表示“[拉]<sub>处置</sub>车的人”;“小号手”表示“[吹]<sub>功用</sub>小号的人”;“程序员”表示“[开发]<sub>施成</sub>程序的人”;“股民”表示“[炒]<sub>功用</sub>股的人”;“血压计”表示“[测量]<sub>处置</sub>血压的仪器”;“篮球场”表示“[打]<sub>功用</sub>篮球的场地”;“汉语热”表示“[学习]<sub>处置</sub>汉语的热潮”;“音乐风”表示“[创作]<sub>施成</sub>音乐的风气”。后一种情况是,在名词构成的复合词内部,名词<sub>1</sub>强迫构建与名词<sub>2</sub>相关的事件性语义解读。比如,“书桌”表示“[放]<sub>处置</sub>书的桌子”;“台球桌”表示“[打]<sub>功用</sub>台球的桌子”;“水果刀”表示“[切]<sub>功用</sub>水果的刀”;“服装厂”表示“[生产]<sub>施成</sub>服装的工厂”。

动词强迫现象,这方面较早的研究有 Lin & Liu<sup>④</sup> 和 Liu et al.<sup>⑤</sup>。Lin & Liu 认为汉语并不存在类似英语的动词强迫现象,因为英语普遍存在诸如“John began(reading) a book”这种动词强迫宾语名词发生语义转变(由事物指代事件)的现象,而汉语却没有“张三开始一本书”这种说法。并且,他们把产生这种语言现象的差异归结为:英语名词由于词汇化过程(lexicalization)具有事件性特征,而汉语由于不

① “好图书馆”一般面积大、设施和书类齐全丰富、功能完善、建造质量好。

② 李强.从生成词库论看动宾组合及其语义转喻[J].语言教学与研究,2017,(6).

③ 宋作艳.类词缀与事件强迫[J].世界汉语教学,2010,(4).

④ Lin T.-H.,Jonah and Liu C.-Y.,Cecilia.Coercion,Event structure,and syntax [J].Nanzan linguistics,2004,(2).

⑤ Liu, Mei-chun, Chu-Ren Huang, Ching-Yi Lee. Lexical information and beyond: constructional inferences in semantic representation[R]. Proceedings of the 13th Pacific Asia Conference on Language, Information and Computation, 27~38. Taipei, Taiwan, 1999.

经历词汇化,所以汉语名词就不包含事件性特征,名词不能直接跟在要求宾语为事件性语义的动词之后,因而动词语义强迫名词的现象就不会发生。但是,关于英汉名词这种事件性特征上的差异看上去更像是一种为了迎合特定解释需要而特设的假定;更何况,汉语中也存在动词语义强迫名词现象。

Liu et al.以汉语中的“赶”为例,认为“赶”的语义中包含时间性(temporal)特征,所以它会强迫宾语名词也必须具有时间性特征,而时间性往往是过程性和动作性的体现。“赶+NP”的语义通常是在较短的时间之内通过特定的行为完成一定的任务;所以,“赶”的宾语名词就必须表现动作性,而这是通过名词的功用角色或施成角色实现的。比如,“赶比赛”既可能是“赶看<sub>[Telic]</sub>比赛”,也可能是“赶参加<sub>[Agentive]</sub>比赛”。Hsu & Hsieh<sup>①</sup>同样进行了类似的研究。而更为系统的研究则见于宋作艳<sup>②</sup>,她对能够产生强迫作用的动词进行了分类,包括“喜欢”类、“推迟”类、“学习”类等,说明动词强迫现象在汉语中具有普遍性;并且根据宾语名词和隐含动词的语义特点,她把动词强迫现象分成了6类,分别是活动隐含、事件隐含、句法补位、词汇化、事件实体和轻动词补位。

形容词强迫现象可以分为两种不同的情况:(1)形容词和名词的语义类型不匹配(type-mismatch),形容词强迫构建一个与名词所指事物相关的事件。<sup>③</sup>比如,“舒服的椅子”表示“[坐起来]<sub>功用</sub>舒服的椅子”;“伤心的诗歌”表示“[读起来]<sub>功用</sub>伤心的诗歌”;“轻松的小品”表示“[看起来]<sub>功用</sub>轻松的小品”;“很快的打字员”表示“[打字]<sub>功用</sub>很快的打字员”。(2)形容词在语义上并不要求事件性的语义特征,因此并不强迫名词产生与之相关的事件义,而是强迫整个结构在语义解读过程中产生另外一个名词。比如,“俊俏的女子”表示“[容貌]<sub>构成</sub>俊俏的女子”;“高大的小张”表示“[身材]<sub>构成</sub>高大的小张”;“聪明的战士”表示“[头脑]<sub>构成</sub>聪明的战士”。<sup>④</sup>

副词强迫现象是指,原本用来修饰动词和形容词的副词,也可以用来修饰名词,它们可以强迫名词产生与该名词所指称事物相关的性质义。比如,“他很女人”中,“很”强迫“女人”转指女人的贤惠、体贴等特性;“这人太奶油”中,“太”强迫“奶油”转指奶油的油膩等特性;“他这人可不书生”中,“不”强迫“书生”转指书生的不善交际、柔弱等特性;“她变得非常城市”中,“非常”强迫“城市”转指城市的新潮、时尚等特性。

方位词强迫现象是指,方位词可以对前面名词的语义属性特征产生强迫。Li(2013)<sup>⑤</sup>将这种现象分为7种不同的类型,分别是:(1)对名词数量义的强迫,如“嘴唇之间”,“之间”强迫“嘴唇子”必须是两张;(2)对事件名词语义的强迫<sup>⑥</sup>,如“雨前备伞/风里雨里”,“前”强迫“雨”表示事件,“里”强迫“雨”表示事物;(3)对名动词语义的强迫,如“演讲前/演讲里”,“前”强迫“演讲”表示动作行为,“里”强迫“演讲”表

① Hsu Chan-Chia & Hsieh Shu-Kai. To Coerce or Not to Coerce: A Corpus-based Exploration of Some Complement Coercion Verbs in Chinese[R], Proceedings of the 6th International Conference on Generative Approaches to the Lexicon, September 24~25, Pisa, Italy, 2013.

② 宋作艳.生成词库理论与汉语事件强迫现象研究[M].北京:北京大学出版社,2015.

③ 宋作艳.生成词库理论与汉语事件强迫现象研究[M].北京:北京大学出版社,2015.

④ 李强.汉语形名组合的语义分析与识解——基于物性结构的探讨[J].汉语学习,2014,(5).

⑤ Li, Q. Coercion of Locatives in Mandarin Chinese [A]. in Chinese Lexical Semantics [C]. Verlag Berlin Heidelberg: Springer, 2013.

⑥ 对于“事件名词”的介绍,可参考:韩蕾.现代汉语事件名词分析[J].华东师范大学学报(哲学社会科学版),2004,(5); Wang, Shan & Huang Chu-Ren. Type Construction of Event Nouns in Mandarin Chinese[R]. Papers of the 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26): 2012.

示内容信息;(4)对载体名词语义的强迫,如“书下/书上”,“下”强迫“书”表示物质实体,“上”强迫“书”表示内容信息或物质实体;(5)对容器名词语义的强迫,如“椅子上/椅子里”,“上”强迫“椅子”表示平面,“里”强迫“椅子”表示空间;(6)对名词所指事物形状的强迫,如“锅上/锅里”,“上”强迫“锅”表示平底锅,“里”强迫“锅”表示凹形锅;(7)对多义名词语义的强迫,如“鸡中/鸡上/鸡里”,“中”强迫“鸡”表示家禽,“上”强迫“鸡”表示肉类食品,而“里”强迫“鸡”表示家禽或肉类食品。

量词强迫现象是指,量词可以对名词的语义属性产生强迫。Huang & Ahrens<sup>①</sup>将汉语的量词分为:个体量词(individual classifiers)、种类量词(kind classifiers)和事件量词(event classifiers),每一类量词都存在对名词的强迫。比如,“条”强迫后面的名词具有细长的特征,“张”强迫后面的名词具有扁平的特征;所以,“一条蛇/一张纸”可以说,而“一条纸/一张蛇”就不能说;“一款新车/一辆新车”中,“款”强迫“新车”做类指义解读,“辆”强迫“新车”做个体义解读;“一架飞机/一班飞机”中,“架”强迫“飞机”做个体事物解读,而“班”则强迫“飞机”做事件性解读。此外,李强、袁毓林<sup>②</sup>还对汉语中一个常见的量词“场”进行了个案研究,从生成词库论考察了“场”跟相关论元成分的组合情况,指出“场”具有事件量化的功能,要求与之搭配的论元成分必须具有时间性和事件性的语义特征,并通过相关的类型强迫机制生成恰当的语义。

语义强迫是自然语言中一种非常常见的现象,它是对那些语义类型不匹配(mismatch)、不完整(incomplete)或限定不足(under-specification)等情况的一种调节机制。语义组合中的强迫具有充分的心理学基础<sup>③</sup>,它是人们处理组合性意义的一种内化(innate)机制。对于语言中强迫现象的研究有助于我们探讨意义的生成性问题,帮助我们了解大脑处理组合性意义的运作过程。

### (三)句法格式语义还原

生成词库可以被应用到特定语言结构的语义和句法研究之中。这里,我们以“的”字结构为例加以说明。

宋作艳<sup>④</sup>考察了“VP+的+NP”和“NP<sub>1</sub>+的+NP<sub>2</sub>”结构中隐含的谓词及其隐现情况。对于“NP<sub>1</sub>+的+NP<sub>2</sub>”,她总体上分为领属关系和属性关系两大类,每一大类下面再划分为若干小类。她指出,领属关系这一类都可以转换成“NP<sub>1</sub>有NP<sub>2</sub>”,属性关系这一类根据NP<sub>1</sub>和NP<sub>2</sub>的不同语义关系可以隐含不同的动词。“VP+的+NP”同样也可以分为领属和属性两类,并且,隐含谓词“推动、造成、用于”等是比较抽象的动词,出现与否几乎不增添任何具体信息,主要是标明NP与VP之间的施成或功用关系,所以强调时才出现。至于谓词隐现的规律,她从动词的分类看,语义越实在的动词越容易显现;从动词义与构式义之间的关系看,越是与构式义不完全匹配的谓词越容易显现。最后,“NP<sub>1</sub>+的+NP<sub>2</sub>”结构的多种解释主要是因为结构的歧义性和构式义的概括性。

李强<sup>⑤</sup>同样在生成词库论的基础上探讨“的”字这类特殊的结构,与宋作艳<sup>⑥</sup>讨论“VP+的+NP”结

① Huang, C.-R. & Ahrens, K. Individuals, kinds and events: classifier coercion of nouns [J]. Language Sciences, 2003, (25).

② 李强,袁毓林.量词“场”的事件量化功能分析[J].语言研究,2016,(4).

③ Pylkkänen, L. & McElree, B. An MEG study of silent meaning. Journal of Cognitive Neuroscience, 2007, (19).

④ 宋作艳.也谈与“的”字结构有关的谓词隐含[J].汉语学习,2014,(1).

⑤ 李强.谓词隐含、物性角色和“NP<sub>1</sub>+的+NP<sub>2</sub>”结构[J].语言研究,2014,(4).

⑥ 宋作艳.也谈与“的”字结构有关的谓词隐含[J].汉语学习,2014,(1).



构不同,该文主要讨论“NP<sub>1</sub> + 的 + NP<sub>2</sub>”结构隐含的谓词、谓词隐现的规律以及结构歧义等问题。从高层平面上,通过对 NP<sub>1</sub> 和 NP<sub>2</sub> 语义关系的考察把“的”字结构中谓词隐含的现象分为两大类:a.谓词可隐可现;b.谓词不可隐。其中,b类又可以划分为两个小类:b1.补不出谓词;b2.必须补出谓词。每一大类下面还可以划分成若干不同的小类。至于谓词为什么能够隐含,这是由“的”字结构中 NP<sub>1</sub> 和 NP<sub>2</sub> 所代表概念或事物之间的紧密程度不同造成的。NP<sub>1</sub> 和 NP<sub>2</sub> 之间缺乏规约化的联系,谓词必须显现;NP<sub>1</sub> 和 NP<sub>2</sub> 之间具有必然的联系,谓词必须隐含;NP<sub>1</sub> 和 NP<sub>2</sub> 之间具有自然的联系,谓词可隐可现。最后,“NP<sub>1</sub> + 的 + NP<sub>2</sub>”结构的歧义性源于 NP<sub>1</sub> 和 NP<sub>2</sub> 隐含谓词的多样性,而谓词的多样性又是因为 NP(包括 NP<sub>1</sub> 和 NP<sub>2</sub>)有多种物性角色。

Wang & Huang<sup>①</sup>以汉语领属关系的“的”字结构为研究对象,主要探讨了该结构的两个问题:一是领属者 NP<sub>1</sub> 和领有物 NP<sub>2</sub> 各自的语义类型及其物性角色关系;二是 NP<sub>1</sub> 和 NP<sub>2</sub> 之间语义的相互制约。其中,第二个问题又可以分为几种不同的情况,比如,“象棋”原本有两个义面:物质实体和游戏事件,但在“象棋的形状”中,“形状”制约了“象棋”只能表示物质义,而事件义则被压制了,“早餐的牛奶”“演唱会的音乐”“报纸的文字”也都是类似的情况;“姚明的篮球”存在歧义,因为“篮球”有两个义面:物质实体和运动事件,所以它既可以表示“姚明买的篮球”,也可以表示“姚明打篮球”,“柳传志的联想”“大楼的窗户”也都是类似的情况;“白宫”有两个义面:机构和建筑,“门”也有两个义面:物质实体和缝隙,但在“白宫的门”中,“白宫”只能表示建筑义,而“门”则可表示任意一个意思。

从上述成果来看,应用生成词库来研究汉语特定格式的句法和语义的成果目前来说还不是很多,研究对象和范围还有待扩大;特别是对于那些能够浮现出动词义的句法格式来说,生成词库还大有用武之地。

#### 四、语言认知研究

隐喻(metaphor)和转喻(metonymy)是非常普遍的语言现象,同时也是语言学研究的重要议题。随着认知语言学的不断深入,人们越来越认识到它们不仅仅是一种修辞手段,更是一种重要的认知机制。<sup>②</sup>生成词库论为隐喻和转喻的研究提供了一个新的角度,尤其是物性角色对于名词的描写,作为储存在心理词库(mental lexicon)中的词汇知识的一部分,它可以更加细致地描写和刻画隐喻和转喻的语义理解与生成途径。

##### (一)隐喻义的理解与生成

隐喻是不同的高级体验域(superordinate experiential domain)中两个概念之间的映射,是将源域(source domain)的特征映射到目标域(target domain),或者说目标域的概念可以通过源域来理解。隐喻通常要牵涉两个事物,并且这两个事物之间存在一种类比关系,即相似性。所以,事物所存在的某些属性特征就成为构造隐喻式表达的关键。

李强<sup>③</sup>主张建立物性结构和概念整合的糅合网络模型对汉语中的隐喻表达进行解释,详细说明了

① Wang, Shan & Huang Chu-Ren. A Generative Lexicon Approach to Possessive Relation of Mandarin Chinese[R]. Paper presented at The Symposium Commemorating the 90th Anniversary of the Birth of Prof. Dexi Zhu and the 50th Anniversary of Prof. Jianming Lu's Teaching, Peking University, Beijing, China, 2010.

② Lakoff, G. & Johnson, M. Metaphors we live by [M]. Chicago: The University of Chicago Press, 1980.

③ 李强. 基于物性结构和概念整合的名词隐喻现象研究[J]. 语言教学与研究, 2014, (6).

这种糅合网络模型对于解释隐喻现象所发挥的作用。比如,在“工业是国民经济发展的心脏”中,“工业”和“心脏”各自都能建立一个物性结构的描写框架,分别形成各自的概念空间,概念空间中的元素即物性角色;这些不同的物性角色能够建立起一一对应的关系,因而会被带入到整合空间中。在整合空间中,“心脏”和“工业”的功用角色得到激活(activate),“心脏”维持生命的功能角色会被利用进而投射到“工业”上,使得“工业”的重要性得到凸显。正是由于物性角色和概念整合,人们能够最终理解这句话所要表达的意义。名词隐喻义的生成与理解涉及一个心理加工模型,这个心理模型可以分为5个步骤:刻画本体名词和喻体名词的物性结构>构建名词的心理空间,建立物性角色的映射关系>本体名词和喻体名词进行组合,使之能够共现于一个句子>本体名词和喻体名词被带入整合空间,喻体名词的物性角色会投射到本体名词之上>由整合空间的名词压缩和物性角色投射驱动表层语言形式的生成。

总之,物性角色在名词隐喻义的理解过程中起到了非常关键的作用,更为重要的是,生成词库一开始就为物性结构设计了一套形式化的描写和表征方式<sup>①</sup>,以便于计算机对于信息的存储,这样可以直接被利用到语言计算的相关领域之中。名词隐喻义的识别与理解也是自然语言处理的一道难题,物性角色的形式化表征方式可以为这道难题的解决提供一些帮助。

## (二)转喻义的理解与生成

与隐喻相同,转喻也被看作是人的认知能力的一部分。关于转喻,之前也有不少研究成果,生成词库为这一问题提供了新的思考角度。

宋作艳<sup>②</sup>从事件强迫对汉语中名词动用这类逻辑转喻现象进行了深入的探讨。名词动用是指,在某一特定的语境下,名词起到了动词的作用,名词做了事件性的语义解读,而这种现象通常被认为是一种转喻,是以一个事件中的参与者来转指事件、以部分转喻整体的概念现象。比如,“我还没有博客过”中的“博客”,实际意义表示“写博客”;“百度一下”中的“百度”,实际意义表示“用百度搜索”;“把白菜窖上”中的“窖”,实际意义表示“用窖储藏”。她对10种不同类型的名词动用现象进行了物性角色方面的考察,比如,上面几例中,“写”是“博客”的施成角色,“搜索”是“百度”的功用角色,“储藏”是“窖”的功用角色。在此基础上总结出了两条关于名词动用生成的规律:(1)具有明确、单一的功用角色或施成角色的名词更容易作动词;(2)具有间接功用角色的名词比具有直接功用角色的名词更容易作动词。此外,还从性质、半能产性、多种解释和汉英差异等角度对名词动用和逻辑转喻之间的异同进行了梳理和比较。该文从生成词库论的角度解释了名词动用这一语言现象的产生过程,逆向观察名源动词在源名词的语义结构中所充当的物性角色,从而发现了一些非常独特和新颖的事实和规律。

除了名词动用这类现象外,汉语中还存在名名互转的情况。从生成词库论的角度对汉语名名互转这类概念转喻现象进行研究是个新的尝试。物性角色对于形成概念转喻所依赖的认知框架起了一定的作用,常见的认知框架基本上都可以用物性角色来加以概括和说明。比如,“见到的尽是新面孔”中,“新面孔”转指“新人”,传统上认为它们属于“整体-部分”认知框架,所以能够形成转指,但从物性角色的角度看,这种转喻的形成基本上依赖于物性结构中的构成角色(“面孔”是“人”的一部分);“穿的确良”中,“的确良”转指“衣服”,传统上认为它们属于“材料-物体”认知框架,所以能够形成转指,但从物性角色的

① 可以参考 Pustejovsky(1995:82)的图表。

② 宋作艳.逻辑转喻、事件强迫与名词动用[J].语言科学,2013,(2).

角度看,这种转喻的形成基本上依赖于物性结构中的材料角色(“的确良”是“衣服”的材料)<sup>①</sup>;“喝一杯”中,“杯”转指“杯子里的水”,传统上认为它们属于“容器-内容”认知框架,所以能够形成转指,但从物性角色的角度看,这种转喻的形成基本上依赖于物性结构中的功用角色[“杯子”的功用角色是“装(水)”]。此外,类型强迫还可以对概念转喻的发生过程进行解释。在概念转喻中,名词转指现象的产生离不开语义强迫,这种强迫力量来自于动词、副词,句式或者会话语境和常识性知识。逻辑转喻和概念转喻之间的异同可以从压缩结构、语义所指、转指义固化、半能产性和多种解释的角度加以说明。

物性角色和类型强迫具有很强的心理现实性,它们可以被看作是人的认知能力的一部分。从生成词库论的角度去研究诸如隐转喻这类语言认知领域的问题,可以帮助我们打开新的研究视角,发现新的语言事实。

## 五、语言计算研究

生成词库论的创始人是美国布兰代斯大学计算机科学系和 Volen 国家综合系统中心教授,主要从事自然语言的理论和计算研究。所以,从生成词库论创立之始,它就带有了浓厚的语言计算和应用背景;并且,试图要解决机器理解自然语言语义的问题。可以说,语言计算是生成词库非常重要的一块应用领域。就汉语而言,近些年来尝试把生成词库论应用到汉语语言信息处理方面的努力也一直在进行;不过,这个领域的研究还有待深化和拓展。

袁毓林<sup>②</sup>较早地主张将生成词库论应用到汉语语义资源建设。随着自然语言处理研究的向前推进,语义资源建设成为制约该领域继续发展的一个重要瓶颈。一个好的语义资源对于信息检索(information retrieval)、问题问答(question answer)、文本推理(text inference)、词义消歧(sense disambiguation)等都具有至关重要的作用。而这种语义资源至少包括3种:汉语词网(Chinese WordNet, CWN)、汉语谓词网(Chinese VerbNet, CVN)和汉语命题库(Chinese PropBank, CPB)。其中,汉语词网的建设应该是基础中的基础。而面对汉语词网的建设,有没有一种较为完备和系统的理论可以参考和借鉴?生成词库论可以为建设CWN提供科学的方法论指导和词义描写与组织的理论框架。更为重要的是,生成词库所提出的形式化的词义描写框架和与之相配套的可计算的词义表达式,可以服务并应用到面向内容计算的语义资源的建设工程上。

袁毓林<sup>③</sup>在(2008)基础上做了进一步延伸和拓展。面对如何建设合适的汉语语义知识库,从而为文本语义的计算机自动分析提供可靠的基础这一问题,该文确定了重点的研究内容是:如何描写汉语名词、动词、形容词等实词的语义结构以及它们之间的关联关系;明确了以生成词库论和论元结构理论为指导,以刻画名词、动词和形容词的相关语义结构为基础的一种揭示词语之间搭配连接和选择限制关系的语言计算路子。值得提出的是,该文对生成词库论中4种物性角色进行了拓展,将物性角色的总数扩展到10种以满足汉语名词的语义搭配要求;并且为名词添置了相应的句法框架的描写信息。这样,名词语义和句法信息能和动词、形容词语义和句法信息实现相互关联,从而能够将事物和跟事物相关的事

① 材料角色(material role)指的是创造名词所指事物所用的材料。该角色是袁毓林(2014)所扩展的6种物性角色中的其中一种。

② 袁毓林.语义资源建设的最新趋势和长远目标——通过影射对比、走向统一联合、实现自动推理[J].中文信息学报,2008,(3).

③ 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).

件的有关世界知识及其语言表达形式表示出来,这为最终形成以名词(实体)为检索核心的、面向对象(object orientation)的语义知识库奠定了充分的基础。

袁毓林、李强<sup>①</sup>所做的是对以上两个文献所包含的主体思想的一种细化工作。围绕着 Roger Chaffin 在他跟 Christiane Fellbaum 的私人交流中所提出来的“网球问题”(Tennis Problem)<sup>②</sup>,该文对目前比较主流的几种语言词汇和概念知识库系统(包括 WordNet、VerbNet、FrameNet、ConceptNet 等)进行了检讨,指出它们在解决“网球问题”上还都存在一定的局限性,进而指出基于生成词库论的名词物性结构知识描写体系可以解决“网球问题”,主张用名词的物性结构知识和相关的句法组合知识来构建一种以名词(实体)为核心的词汇概念网络,以弥补上述几种知识库系统的不足,为自然语言处理提供一种可资参考的词汇概念知识库体系。以“网球、网球拍、网球场”和“生日、蛋糕、蜡烛”这两组词语为例,分别构建了概念语义关系网络图。在网络图中,由任意一个节点可以通过指针性的链接关联到另一个节点,而建立指针性链接的依据就是生成词库论中的物性角色,依靠这些物性角色,不同概念之间的关系可以被触发(trigger)和激活(activate),有效地模拟了一个储存在人脑中的复杂的认知概念网络,从而为实现计算机的自然语言理解提供了理论依据和技术支撑。

魏雪<sup>③</sup>主要以现代汉语(特别是网络搜索词)中的名名组合的语义关系为主要研究对象,通过对将近 900 个名名组合中的名词语义类型进行归纳,得出了 370 个左右的语义类组合模式及其相应的释义模板,在此基础上搭建了名名搭配数据库,并直接服务于名名组合释义短语的程序开发。其中,在释义模板的总结和归纳中,名词物性角色中的施成角色和功用角色发挥了重要的作用。比如,有关处所的语义模式中有一类是“处所/空间+事件”(河南矿难、雅典奥运会、森林大火、江苏霍乱),它们的释义中相应地都隐含了事件名词的施成角色(在河南[发生]的矿难、在雅典[举行]的奥运会、在森林[出现]的大火、在江苏[出现]的霍乱);再比如,有关“受事+施事”的语义模式中有一类是“人工物/身体构件/食物+用具”(煤气炉子、羊毛剪子、香烟筒),它们的释义中相应地都隐含了用具名词的功用角色([烧]煤气的炉子、[剪]羊毛的剪子、[装]香烟的筒)。通过对名名组合语义类的归纳,得到名名组合的释义模板库,而在模板库中可以规定其中隐含动词是名词的功用角色还是施成角色。这样,当释义系统遇到任意一个名名组合时,通过分析名词的语义类就可以选择什么样的释义模板对它进行释义,再利用名词的物性角色就可以最终实现对名名组合进行自动地、准确地释义。可以预见,这样一种名名组合的自动释义程序对于自然语言处理,尤其是机器翻译、语义搜索来说是非常有用的。

总之,生成词库论所倡导的一种形式化、可供操作的语义描写手段为语言的可计算提供了可能性,自然语言处理中供生成词库发挥作用的领域应当还有不少,这有待于我们继续深入挖掘。值得一提的是,北京大学中文系已经在开始着手进行以生成词库论为核心的大规模语义资源的研究和建设,目前正在建设名词物性角色和句法配置数据库,再辅之动词配价数据库、形容词配价数据库以及动词蕴涵式库的建设,希望能够最终形成一个对信息抽取、内容计算、舆情分析、情感评价等多种自然语言处

① 袁毓林,李强.怎样用物性结构知识解决“网球问题”? [J].中文信息学报,2014,(5).

② “网球问题”指怎样把 racquet(网球拍)、ball(网球)和 net(球网)之类具有情境联想关系的词汇概念联系起来、发现它们之间的语义和推理关系。这是自然语言处理领域中一个非常棘手的问题。详见 Fellbaum, C.(ed.) WordNet: An Electronic Lexical Database [M].Cambridge, Massachusetts: MIT Press, 1998.

③ 魏雪.面向语义搜索的汉语名名组合的自动释义研究[D].北京大学硕士研究生学位论文,2012.



理和应用任务都有用的多层次的语义知识体系。

## 六、语言教学研究

伴随汉语教学在全球范围内的火热开展,越来越多的学者开始从语言教学的角度来对汉语进行研究:一方面是研究那些对于汉语学习者来说非常重要,而在之前的研究中常被忽视的语言学基础知识;另一方面是致力于开发新的教学方法和手段,利用新技术来辅佐对外汉语教学。在这两方面,生成词库论都能发挥一定的作用。

李强、袁毓林<sup>①</sup>从生成词库论的名词语义类型和物性角色描写体系对词典的名词释义进行研究。汉语中有一类特殊的名词,它们包含两个不同的义面,但在具体的语境中往往只实现为其中一个义面。比如,“篮球、足球”等既可以表示球类运动,又可以表示由材料制成的球体;“电视、电话”等既可以表示具有某种用途的设备装置,又可以表示通过设备发出的图案、声音、文字等信息;“决定、通知”等既可以表示行为所代表的事件,又可以表示该事件的具体内容信息。但是,词典对于这类词的释义还有不太完善的地方,主要表现为只收录这些词的其中一个义面,而忽略了另外一个义面。比如,“早饭、象棋、雨”等,《现代汉语词典》(第6版,简称《现汉》)只释义了作为物质实体的义项,而忽略了作为事件的义项;“唱片、书、屏幕”等,《现汉》只释义了作为物质实体的义项,而忽略了内容信息的义项;“演讲、调查、分析”等,《现汉》只释义了作为事件的义项,而忽略了内容信息的义项。通过利用物性角色描写体系对这些词语进行分布式的验证说明,这些词语的不同义面也应当作为独立的义项分列出来以体现其用法上的差异。不难发现,对于名词词义的这种分化和细化处理是有利于汉语学习者对于词义的理解和运用的。词义的不同能够反映其用法的差异,学习者通过掌握词语的不同意义之间的细微差别,才能更加正确和规范地运用这些词语。

李强、袁毓林<sup>②</sup>把关注的视角转移到同义名词辨析问题之上。同义名词词义之间的细微差别一直都是汉语教学中的重点和难点问题,如何发展出一套比较理想的辨析方法是汉语理论研究者 and 从事汉语教学的一线教师都在思考的问题。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。现在有了物性角色,可以为同义名词辨析这个难题提供一种可能的解决方案。借鉴欧洲多语言句法语义词库的描写模板,在此基础上结合物性角色描写体系,确定了汉语同义名词辨析的思路、方式及框架;更为重要的是,在这种理论的指导下开发了一套在线的可供查询的同义名词辨析系统,该辨析系统不仅可以查询某个名词的意义和用法信息,还可以查询两个同义名词之间在意义和用法上的共性和差异。这样一种同义名词辨析系统,以一种简单、友好的界面使同义名词之间的细微差异能够一目了然地呈现在汉语学习者面前,从而使同义名词辨析具有更强的可操作性和实用性。与传统的主要基于意义的同义词辨析相比较,基于物性角色的同义名词辨析顺应了分布语义学的相关主张,即词语的意义差别是通过其分布上的差异表现出来的。这种基于搭配组合的辨析方式可能会比单纯基于意义的辨析路子更便于汉语学习者理解、掌握同义名词之间的差异,对于汉语国际教学能够起到一定的推动作用。

① 李强,袁毓林.从生成词库论看名词的词典释义[J].辞书研究,2016,(4).

② 李强,袁毓林.基于物性角色的同义名词辨析方法探讨[J].世界汉语教学,2014,(4);李强,袁毓林.服务于国际汉语教学的同义名词辨析查询系统建设[J].对外汉语研究,2015.

汉语教学方面的研究和应用是对原本旨在解决词义变化性问题的生成词库论的拓展和延伸,一定程度上丰富了生成词库的研究内容和视角。目前,生成词库背景下汉语教学方面的探讨还是初步的,已有研究显示了该理论所具有的宽阔的研究领域和应用价值。

## 七、小结

近些年来,生成词库发展极为迅速,研究规模日益扩大,除了英语研究之外,还已经广泛应用于法语、意大利语、韩语和日语等语言的研究之中(Godard & Jayez; Johnston & Busa; Im & Lee; Nishiguchi)<sup>①</sup>。相较于其他语言的生成词库论研究,汉语研究开展的时间还并不是很长,有待深入探讨和挖掘的课题内容还有不少。

本文对目前这一研究趋势和成果做了回顾和总结,分别从语言本体、语言认知、语言计算和语言教学4个方面展示了生成词库丰富的研究内容和应用范围。其中,语言本体和语言认知侧重于理论性,而语言计算和语言教学侧重于应用性。作为一种全新的理论和形式化的操作方式,生成词库能够解释汉语中的一些语言现象,或是对旧问题进行新角度的诠释。

当然,生成词库还有不少有争议的地方,比如 Jayez<sup>②</sup>所指出的动词对于宾语名词的强迫所引发的生成过度 and 生成不足问题;说话人的一般背景知识在形式上应该如何被纳入到生成词库论之中,如何区分哪些是背景知识,哪些不是背景知识;“物质实体”这个概念是否是清晰的、恒定的。未来汉语研究方面还可以深入思考的问题包括:

第一,概念术语的界定。生成词库将“类型强迫”分为两类:“强迫引入”和“强迫利用”,但“强迫利用”这一概念是否合适值得推敲。因为“强迫”的本意是词义中原本没有某个意义,而最终在邻近词语的影响下被迫产生这个新的意义。但“强迫利用”的情况是,这个新的意义本身已经存在于这个词义当中,只不过并没有突显。汉语的相关语言事实是否能够厘清概念的本质,这值得讨论研究。

第二,语篇背景知识的融合。生成词库论所提出的词汇表征结构和各种不同的组合机制,基本上是在以句子为基础的语境上对词义变化现象进行解释;也就是说,它是脱离语篇背景知识的一种解释策略。但是,句子的意义是在语篇环境中浮现出动态意义的,所以,句子平面上的解释虽然能够说明一定数量的情况,但无法解释更为复杂的超单句情况。如何将语篇知识的形式化表征以及与词汇概念结构的形式化表征相结合,这或许是未来可以进一步研究的课题。

第三,超词组平面的研究。应用生成词库论研究汉语,除了可以继续深化词语组合性这一课题外,还可以深入讨论特定短语结构的语义浮现问题,也可以研究特定句式对于词语,特别是动词的选择制约

<sup>①</sup> Godard D. & Jayez J. Towards a Proper Treatment of Coercion Phenomena[A]. in Proceedings of the 6th Conference of the European Chapter of the ACL[C]. Utrecht: OTS Utrecht, 1993; Johnston, M. & Busa, F. Qualia Structure and the Compositional Interpretation of Compounds[A]. In E. Viegas (ed.) Breadth and Depth of Semantic Lexicons[C]. Dordrecht: Kluwer Academic, 1999; Im, Seohyun & Lee, Chungmin. Type Construction of Nouns with the Verb ha- 'do[A]. In Language, Information, and Computation: Proceedings of the 16th Pacific Asia Conference[C]. Jeju, Korea, 2002; Nishiguchi, Sumiyo. Application of Event Semantics to Qualia Structure of Nouns for Disambiguation of Japanese Postposition No[A]. In Proceedings of Conference of the Pacific Association for Computational Linguistics (PACLING2009)[C]. Sapporo, Japan, 2009.

<sup>②</sup> Jayez, Jacques. Underspecification, context selection, and generativity[A]. In Bouillon & Busa (ed.) The Language of word meaning[C]. Cambridge: Cambridge University Press, 2001.

作用以及句式语义的生成性问题。在自然语言处理上,可以考虑如何将名词物性角色与动词、形容词的论元结构相结合,让机器自动生成合格的语言表达形式。在汉语句式的句法语义构造以及自然语义理解方面,还可以继续利用生成词库深入挖掘。

## Research on the generative lexicon theory in China: Retrospect and prospect

LI Qiang

(Department of Chinese Language and Literature, Shanghai University, Shanghai 200444, China)

**Abstract:** In recent years, advances in the generative lexicon theory have been obvious, and, thanks to it, the research on Chinese has also made much progress with praiseworthy achievements. This paper gives a summary of the research findings concerning the Chinese language through the application of the generative lexicon theory, covering the four aspects of language ontology, language cognition, language computation and language teaching. The generative lexicon theory can provide a new perspective for solving some puzzling problems in Chinese studies and help further the research on the Chinese language.

**Key Words:** generative lexicon theory; language ontology; language cognition; language computation; language teaching

[责任编辑:李德鹏]