



The Generative Lexicon: Studies on the Chinese Language

# 生成词库理论与 汉语研究

宋作艳 黄居仁 主编

商務印書館  
The Commercial Press



清华大学出版社

# 生成词库理论与 汉语研究

张其成 曹真真 主编

清华大学出版社

北京师范大学985工程二期项目资助

# 生成词库理论与 汉语研究

宋作艳 黄居仁 主编

 商务印书馆  
The Commercial Press

2018年·北京

## 图书在版编目(CIP)数据

生成词库理论与汉语研究/宋作艳,黄居仁主编. —  
北京:商务印书馆,2018  
ISBN 978-7-100-15610-3

I. ①生… II. ①宋… ②黄… III. ①现代汉语—  
研究 IV. ①H109.4

中国版本图书馆 CIP 数据核字(2017)第 297815 号

权利保留,侵权必究。

## 生成词库理论与汉语研究

宋作艳 黄居仁 主编

---

商务印书馆出版  
(北京王府井大街36号 邮政编码100710)  
商务印书馆发行  
北京市十月印刷有限公司印刷  
ISBN 978-7-100-15610-3

---

2018年1月第1版 开本 880×1230 1/32  
2018年1月北京第1次印刷 印张 16 1/4

定价:52.00元

## 学术顾问

袁毓林

王洪君

谢舒凯

## 作者译者

Anna Rumshisky

James Pustejovsky

Jessica L. Moszkowicz

Oľga Batiukova

马萨诸塞大学卢维尔分校计算机科学系

美国布兰代斯大学计算机科学系

美国布兰代斯大学计算机科学系

西班牙马德里自治大学语言文学系

安可思(Kathleen Ahrens)

陈圣怡

陈咏珊

丁晶

洪嘉麒

黄洁

黄居仁

李强

李智尧

林淑晏

林宗宏

刘琼怡

刘兆静

宋作艳

王洪君

王珊

魏雪

谢舒凯

熊佳娟

许展嘉

袁毓林

张爱玲

张秀松

张瑜芸

赵青青

周亚民

香港理工大学英文系

台湾“中研院”语言学研究所

香港理工大学中文及双语学系

华南理工大学国际教育学院

台湾师范大学应用华语文学系

华中科技大学外国语学院

香港理工大学中文及双语学系

上海大学文学院

台湾大学语言学研究所

台湾师范大学迈向顶尖大学计划

台湾清华大学语言学研究所

台湾交通大学

香港理工大学中文及双语学系

北京师范大学文学院

北京大学中文系

香港教育大学中国语言学系

北京大学中文系

台湾大学语言学研究所

西南财经大学经贸外语学院

台湾逢甲大学外语教学中心

北京大学中文系

江苏师范大学文学院

江苏师范大学文学院

台湾大学语言学研究所

香港理工大学中文及双语学系

台湾台北大学

# 前 言

宋作艳 黄居仁

生成词库(Generative Lexicon, GL)理论是美国布兰代斯大学 Pustejovsky 教授于 1991 年提出的,1995 年出版专著《生成词库》(*The Generative Lexicon*),理论框架基本成形。之后 10 年间不断改进,臻于完善。

生成词库理论是目前较新的一种语义学理论,主要有两大特色:1)首次把生成方法引入到词汇语义的研究中,提出了语义生成(组合)的机制。2)与传统的以动词为中心的理论模型不同,生成词库理论强调名词在语义组合中的重要性(Pustejovsky et al., 2013),像刻画动词一样对名词的语义进行了详细的刻画。尤其是物性结构(qualia structure)的引入,把词汇、语法、语义和百科知识结合在了一起。

近年来,生成词库理论迅速发展,已经在法语、意大利语、日语、韩语等语言研究中得到广泛应用,出版相关专著和论文集 10 余部,研究已颇具规模,成为目前较有影响力的语义理论之一。迄今为止,关于这一理论的国际性专题会议——生成词库国际研讨会(International Conference on Generative Approaches to the Lexicon)已经召

开了六届。第一届生成词库亚洲语言学专题研讨会(The First Workshop on Generative Lexicon for Asian Languages)也于2012年召开。生成词库理论中的一些重要思想,比如物性结构,已经被其他许多语言理论吸收,如概念语义学、认知语言学、构式语法、概念整合理论等。而且,有关生成词库理论的研究不只停留在理论层面,还包括资源建设,比较有代表性的是布兰代斯语义本体(Brandeis Semantic Ontology,BSO)(Pustejovsky et al.,2006)和面向12种欧洲语言的语义本体——简单词库(SIMPLE\* Lexicon)(Lenci et al.,2000)。

早在20世纪90年代,已有学者开始把生成词库引入汉语研究,不仅揭示了汉语中的一些规律,而且修正并改进了生成词库理论。历时十几年,相关研究成果已经比较可观,并在语言学界产生了一定影响。很多学者想了解这一理论在汉语中的应用情况,但这些成果多以论文的形式散落在各种期刊、论文集中,不易收集,尤其是港台学者的论文,多是英文版。因此,我们选编了这本论文集《生成词库理论与汉语研究》,旨在展现汉语生成词库理论研究的概貌和现状,促进这一理论在汉语研究中的应用。

本论文集主体部分主要有五个专题:生成词库理论介绍、句法语义研究、词汇语义研究、资源建设和跨领域研究。专题之前是生成词库理论创始人Pustejovsky教授的特邀论文,介绍基于生成词库理论的语义标注框架。

第一个专题“理论介绍”包括《生成词库》中译本(谢舒凯译)的节选和另外三篇介绍文章。这四篇文章各有侧重,互为补充,相得

---

\* SIMPLE是Semantic Information for Multifunctional Plurilingual Lexica的简缩形式。



益彰,勾勒了生成词库理论的框架及其发展变化。节选的是《生成词库》第五章《语意类型系统》,在这一章里,Pustejovsky 介绍了生成词库的四个表征层面:论元结构(argument structure)、事件结构(event structure)、物性结构(qualia structure)和词汇继承结构(lexical inheritance structure)。张秀松、张爱玲的《生成词库论简介》简单介绍了以 1995 年的《生成词库》为代表的生成词库理论的早期框架。宋作艳的《生成词库理论的最新发展》主要介绍了 1995 年之后生成词库理论框架的发展,主要表现在三分的语义类型体系——自然类(natural types)、人造类(artifactual types)和合成类(complex types),以及基于论元选择的语义生成机制(Generative Mechanisms of Argument Selection)。张秀松的《关于生成词库理论的争论》介绍了相关的学术讨论,可以加深我们对这一理论的认识。

第二个专题“句法语义研究”包括八篇文章,主要探讨语义的组合问题。Frege 的组合原则(the principle of compositionality)认为,句子的整体意义是它的组成部分语义的总和及其各部分的组合方式的函数。生成词库理论(Pustejovsky, 1995; Pustejovsky, 2011; Pustejovsky et al., 2013)也主张语言的意义是组合性的,但同时指出,很多表达形式的语义并非直接组合(direct compositionality),不能直接满足组合性原则,语义的组合是动态的、具有生成性,受上下文调制(contextual modulation)。具体表现为论元变换(argument alternation)、多义、意义变化等。生成词库理论认为可以通过丰富词项的词汇语义表征(尤其是语义类型体系)和组合机制来实现语义生成,保持组合性。词项的词汇语义表征包括论元结构、事件结构、物性结构和词汇类型结构(lexical typing structure),其中物性结构是生成性特征的核心,可以用来生成新的类型。Pustejovsky(1995)把组合机制分成了三类:类型强迫(type coercion)、选择约束(select-

tive binding)和协同组合(co-composition)。近年来,类型强迫纳入了语法上的论元选择机制,这样,根据论元的匹配程度,分为三种基于论元选择的生成机制(Pustejovsky, 2006):纯粹类型选择(pure selection)、类型调节(type accommodation)和类型强迫(type coercion)。其中最重要的就是类型强迫,这一机制可以把论元转换成符合函项要求的类型,从而解决类型错配(type mismatch)问题。强迫(coercion)这一术语不是生成词库理论特有的,最早源于编程语言,后来被借用到形式语义学和构式语法中(Lauwers and Willems, 2011),用来解决涉及句法-语义、词汇语义-语用等的接口问题,是目前语言研究中的热点问题。生成词库理论的特点在于把强迫机制纳入论元选择机制,并通过引入物性结构和丰富语义类型体系来解释类型移变(type shift)的词汇语义基础。这个专题的八篇文章中,前六篇是关于类型强迫的,后两篇与协同组合有关。

黄居仁、安可思的《“个体”“类”和“事件”:量词对名词的语义强迫》是较早利用类型强迫研究汉语的文章。这篇文章利用类型强迫机制分析了汉语量词对名词的强迫作用,证明了汉语量词可以强迫名词解读为个体、类或事件。接下来的两篇文章都是关于宾语强迫(complement coercion)的,即宾语位置的事件强迫现象,宾语的类型由实体转换为事件。林宗宏、刘琼怡的《强迫、事件结构与句法》通过英汉比较,发现汉语缺乏宾语强迫,比如 *John began the book* 不能直译成“约翰开始这本书”,必须出现动词“读”“写”等。他们认为这是汉语高解析性的一种表现,并进一步提出了一种假设:丰富的次词汇事件信息(sub-lexical event information)是事件强迫的基础,汉语之所以较少使用事件强迫,是因为汉语名词中缺乏次词汇事件信息。许展嘉、谢舒凯、林淑晏的《从实体到事件:基于语料库的汉语宾语类型强迫机制研究》先是利用网络语料库找到了汉语中大量的宾语强

迫动词,证明了这是一种普遍机制,在汉语中也起作用。然后对其中的 20 个动词做了更细致的分析,考察了它们在语料库中的实际表现,通过多变量分析的统计方法将这些动词分为两大类:第一类动词比第二类动词更常强迫其宾语,而第二类动词比第一类动词能够强迫更多的宾语类型。宋作艳的《逻辑转喻、事件强迫与名词动用》把事件强迫机制引入名词动用的研究,实际上是把解释类型误配 (typemismatch) 现象的强迫机制扩展到了范畴误配 (category mismatch) 现象,即一个句法范畴出现在了另一句法范畴出现的位置,句法范畴和语义功能之间出现了不匹配 (Francis and Michaelis, 2004)。与前人的研究视角不同,这一方法不是看源名词在名源动词的语义结构中充当什么语义角色,而是反过来,看名源动词在源名词的语义结构中充当什么物性角色。比如“百度一下”是“用百度搜索一下”的意思,“搜索”是“百度”的功用角色。研究结果证明,这一方法的解释更形式化、更概括、更准确。洪嘉馥、黄居仁、安可思的《两个摄取义动词的事件选择与强迫:基于 MARVS 理论》探讨了两个摄取义动词“吃”与“喝”在现代汉语中可能的事件强迫类型。李强的《现代汉语方位词语义强迫现象研究》主要探讨了不同方位词对名词语义的强迫,比如“间”会强迫名词做出复数义解读,“桌子间”中桌子的数量一定大于一。

逻辑性多义是生成词库理论要解决的主要问题之一,张秀松的《从生成词库论看汉语词的逻辑多义性》例举了汉语中部分名词和动词的逻辑性多义现象并利用生成词库理论做了解释,涉及强迫和协同组合机制。比如“我的书”在不同的语境中可能分别凸显施成角色、功用角色和书作为物体的一面。“烤”体现动词的逻辑性多义,根据宾语名词的不同,会呈现出两种语义:状态变化义(如“烤土豆”)和制作义(如“烤面包”)。这是动词和宾语名词协同组合的结果。张瑜

芸、谢舒凯的《协同组合机制下的动词宾语特性——以汉语微博语料库中的烘焙动词为例》利用微博语料库对“烤”做了更深入的个案分析,发现工具名词和量词有助于两种语义的区分。

第三个专题“词汇语义研究”包括五篇文章,主要利用物性结构和语义类型来分析汉语复合名词的构词和语义关系。王洪君的《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》认为,生成词库理论对于语义特征和语义场层级的描写抓住了词义在组合中的关联,因而语言学的系统价值很高,并在“锅”“碗”的组合搭配差异分析中利用了物性结构和自然类、人造类、合成类的区分。Johnson and Busa(1999)建议用物性修饰(*qualia modification*)来刻画偏正式名名复合词(词组)内部的语义关系,现在这一方法已经引入很多语言复合名词的研究。黄洁的《名名复合词内部语义关系多样性的认知理据》在名名复合词内部语义关系多样性的认知研究中引入了物性结构,认为不同复合词会激活名词<sub>i</sub>的不同物性角色。例如“花蕾、花蕊、花瓣、花冠”等激活的是花的整体-部分这一组成角色信息,而“花茶、花环、花圈、花饰”等激活的是花作为材料这一功用角色信息。不同名词凸显的物性角色也不一样,比如“电”最显著的本质结构是作为动力的功用角色,所以由“电”构成的名名复合词也多是表示由电作为动力创造的物品,如“电刀、电车、电灯、电炉、电扇”。李智尧、谢舒凯的《汉语双名词复合词中的物性修饰:跨语言的考察》从跨语言的角度例举了名名复合词中的四种物性修饰关系:形式(*formal*)、构成(*constitutive*)、功用(*telic*)和施成(*agentive*)。如“手表、玻璃门、菜刀、牛奶”中的定语名词分别说明了中心语名词所指事物的分类、材料、功用和来源,体现了上述四种关系。值得注意的是,上面两篇文章虽然都用物性角色来分析名词复合词,但关注点略有差异:李智尧、谢舒凯文章中的物性修饰关系关注名名复合词中两个名

词之间的物性修饰关系,即名词<sub>1</sub>激活的是名词<sub>2</sub>的哪种物性角色(如“电车”中“电”显示的是“车”的动力来源,因此激活的是“车”的施成角色);黄洁的文章关注名词<sub>1</sub>的何种物性角色被激活(如“电车”中激活的是“电”作为动力的功用角色)。宋作艳的《类词缀与事件强迫》发现部分“NP+类词缀”词中隐含着谓词,这些谓词基本上是名词的功用角色或施成角色。比如“面包师”中隐含的“做”是“面包”的施成角色,而“做面包”是“面包师”的功用角色;“钢琴家”中隐含的“弹”是“钢琴”的功用角色,“弹钢琴”是“钢琴家”的功用角色。

生成词库理论按其所代表的意义内容将名词分为自然类、人造类和合成类三类(Pustejovsky, 2001、2006)。自然类主要与形式角色和/或构成角色相关,而人造类主要与功用角色和/或施成角色有关,二者之间最大的区别是是否有“意向性”(intentionality),合成类包含至少两个类。比如“花”是自然类、“刀”是人造类,“书”是物体和信息合成类[phys·info],包含物质实体和信息两个类。这一分类很有创意,能解释语言中的一些现象,尤其是自然类与人造类的区分体现了语义上一种很重要的对立,两类词在语言中有系统性差异。但是这种三分的系统在方法上是有问题的,内部标准不统一,类与类之间不平行,会造成交叉。自然类与人造类的区分是偶分式的,是非此即彼的关系,但是“合成类”可能是人造类,也可能是自然类。比如“书”是合成类,但也是典型的人造物,也属于人造类。王珊、黄居仁的《生成词库理论的类型系统建构——以事件名词为例》通过一些句法格式和搭配关系证明了部分事件名词是合成类,而且由此提出合成类事件名词也有自然与人造之分,比如“雪”和“早餐”都是合成类(既指事件也指具体事物),但一个是自然类,一个是人造类。

第四个专题“资源建设”。生成词库理论是基于认知和计算的语义模型,是面向自然语言处理的,目前在中文信息处理方面的应用才

刚刚起步,本专题选的两篇文章具有代表性。袁毓林的《汉语名词物性结构的描写体系和运用案例》提出了一个本土化的汉语名词物性结构的描述体系。根据汉语名词在文本中基本的组合方式、搭配习惯和语义解释,定义了十种物性角色(形式、构成、单位、评价、施成、材料、功用、行为、处置和定位),最终结果相当于一个名词搭配词典。生成词库理论只是给出了一个名词语义描写框架,至于物性结构到底包含几个角色,以及哪些语义信息属于哪种角色,可以有所调整。正如 Jackendoff(2002)所言,这些调整没有本质上的区别。此物性结构的描写就增加了一些比较有特色的物性角色,比如“单位”(量词)就是汉语词汇语义中很重要的一个物性角色。文章通过具体的汉语例子说明这些名词的物性结构知识可以用来分析汉语复杂的句法、语义,比如隐含谓词、有价名词、中动句、话题结构、准定语句等。而且,从角色的有无以及具体角色值上可以看出名词语义的差异。比如人造类的“椅子”具有自然类“水”所不具有的“材料”角色。魏雪、袁毓林的《基于语义类和物性角色建构名名组合的释义模板》利用物性结构来解决网络中名名组合(如“摩托妈妈”)的自动释义问题,构建了名名组合的释义模板,建立了释义模板库。在此基础上,作者利用《知网》等语言资源进一步探索自动发现名词的施成角色和功能角色的方法,从而初步创建了一个名名组合的自动释义程序。

第五个专题“跨领域研究”介绍两篇非常有特色的文章。一篇是黄居仁等的《汉字所表达的知识系统:意符为基本概念导向的事件结构》。该文章认为汉字意符具有与生成词库相似的衍生能力,单一意符所代表的基本概念可以通过概念衍生而形成完整的知识体系。衍生的方向与物性角色密切相关,比如“睑”“睫”是“目”一部分,体现了组成关系,“睹”“瞻”是“目”的功能,体现了功用关系。文章以《说文解字》意符“目、耳、口、鼻、舌”为研究对象,将含有相关意符的

字与这个意符所表达的基本概念的关联依照分类表示出来,建构了由个别基本概念带领的意符知识系统架构。刘兆静、陈咏珊的《浅谈物性结构对汉语儿童语言习得的影响——以名词修饰结构为例》发现,儿童习得“修饰语+的+名词”结构的先后顺序受物性修饰关系的影响。儿童最早习得表形式关系的一类,然后是表构成的一类,最后才是功用和施成。例如,依次习得“大大的眼睛、兔子的尾巴、好玩的东西、宝宝的声音”。一方面是因为形式和构成表达的是基本属性,而功用和施成是派生的,多与事件有关;另一方面大概因为成人对儿童输入的前两类更多。

## 参考文献

- Francis, Elaine J. and Laura A. Michaelis (eds.) 2003. *Mismatch: Form-Function Incongruity and the Architecture of Grammar*. Stanford, CA: CSLI Publications.
- Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Johnston, Michael and Federica Busa. 1999. Qualia structure and the compositional interpretation of compounds. In Evelyn Viegas (ed.), *Breadth and Depth of Semantics Lexicons*, 167—187. Dordrecht: Kluwer.
- Lenci, Alessandro, Nuria Bel, Busa Federica et al. 2000. SIMPLE: A general framework for the development of multilingual lexicons. *International Journal of Lexicography* 13. 4: 249—263.
- Lauwers, Peter and Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49. 6: 1219—1235.
- Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics* 17. 4: 409—441.
- Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Pustejovsky, James. 2001. Type construction and the logic of concepts. In Pierrette Bouillon and Federica Busa (eds.), *The Language of Word Meaning*,

- 91—123. Cambridge: Cambridge University Press.
- Pustejovsky, James. 2006. Type theory and lexical decomposition. *Journal of Cognitive Science* 6; 39—76.
- Pustejovsky, James. 2011. Coercion in a general theory of argument selection. *Linguistics* 49. 6; 1401—1431.
- Pustejovsky James, Havasi Catherine, Littman Jessica, Rumshisky Anna and Verhagen Marc. 2006. Towards a generative lexical resource: The Brandeis semantic ontology. In *Proceedings of the fifth Language Resource and Evaluation Conference*, 1702—1705. European Language Resources Association (ELRA), Genoa, Italy Vol. 7.
- Pustejovsky, James, Pierrette Bouillon, Hitoshi Isahara, Kyoko Kanzaki & Chungmin Lee (eds. ). 2013. *Advances in Generative Lexicon Theory*. Berlin: Springer.



# 目 录

## GLML: 基于生成词库的语义标注语言

..... James Pustejovsky 等著 赵青青 宋作艳 译 1

## 一 理论介绍 ..... 33

语意类型系统 ..... James Pustejovsky 著 谢舒凯 译 34

生成词库理论简介 ..... 张秀松 张爱玲 68

生成词库理论的最新发展 ..... 宋作艳 99

关于生成词库理论的争论 ..... 张秀松 124

## 二 句法语义研究 ..... 147

“个体”“类”和“事件”:量词对名词的语义强迫

..... 黄居仁 安可思 著 熊佳娟 丁晶 译 148

强迫、事件结构与句法

..... 林宗宏 刘琼怡 著 赵青青 译 林宗宏 校订 174

从实体到事件:基于语料库的汉语宾语类型强迫机制研究

..... 许展嘉 谢舒凯 林淑晏 201

逻辑转喻、事件强迫与名词动用 ..... 宋作艳 220

两个摄取义动词的事件选择与强迫:基于 MARVS 理论

..... 洪嘉麒 黄居仁 安可思 244

现代汉语方位词语义强迫现象研究 ..... 李强 257

从生成词库论看汉语词的逻辑多义性 ..... 张秀松 288