

名词词义描写和研究需要什么样的语义学知识？*

李强 袁毓林

(北京大学 中文系/中国语言学研究/计算语言学教育部重点实验室, 北京 100871)

摘要: 该文主要讨论名词的词义描写和研究问题, 首先通过对几种主要的词汇语义学理论(包括结构主义语义学、生成主义语义学、概念语义学和自然语义元语言理论)进行介绍和评述, 指出它们在对名词进行语义刻画方面存在缺陷和不足。然后, 重点引入生成词库理论的物性结构描写方式, 阐明它与前几种理论的区别及其自身的特点。最后, 在生成词库理论的基础上, 展示物性结构知识在有关名词分析中的四个研究案例(词语缺省、隐喻义生成、供用句、中动句)和在自然语言处理中的可能性应用。

关键词: 词汇语义学理论; 生成词库理论; 物性结构; 名词分析; 自然语言处理

What Kind of Semantic Knowledge is Needed in the Description and Research on the Meaning of Nouns?

LI Qiang YUAN Yulin

(Dept. of Chinese Lang. & Lit., Peking University, Research Center of Chinese Linguistics,
Ministry of Education Key Laboratory for Computational Linguistics, Beijing 100871, China)

Abstract: This paper mainly discussed the description and research on the meaning of nouns. Firstly, several main lexical semantics theories (include structuralism semantics, generative semantics, conceptual semantics and natural semantic metalanguage) were introduced and reviewed, on the basis of which defects and deficiencies on semantic description were shown. Then, the qualia structure in generative lexicon theory was introduced, and its differences from the above mentioned theories and features were illustrated. Finally, based on generative lexicon theory, four examples of analysis on nouns using qualia structure (include word default, metaphor meaning generation, provide-using sentence, middle construction) were exhibited and its possible application in natural language processing was shown.

Keywords: lexical semantics theory; generative lexicon theory; qualia structure; nouns analysis; natural language processing

1 引言

句法语义研究中的“动词中心论”主张促使已有的多数词汇语义学流派将研究重心放在动词上, 因而产生了不少关于如何描述动词语义信息的理论和看法; 在关于动词和名词研究的天秤上, 指针总是向动词倾斜。而实际的研究情况显示, 作为数量最多、使用最广的两类词类范畴, 名词也具有同动词一样的研究价值, 对于名词词义结构的描写和刻画的精度在一定程度上制约了名词句法语义研究的深度; 有关名词的语义句法研究不是缺少课题, 而是缺乏一种有效的对名词词义结构进行描写的方式。本文就着力探讨这一问题, 在对几种不同的词汇语义学理论进行介绍和评述的基础上, 说明它们在对名词语义信息的描写上存在不足, 进而重点引入生成词库理论的词义描写方式, 阐明它与之前的几种理论的区别及其自身的特点。本文的相关研究意在指出: 生成词库理论在一定程度上弥补了现有词汇语义学理论在对名词进行语义刻画方面的缺陷和不足, 极大地拓宽了词汇语义学的研究领域和范围, 为语义学提供了一种新的研究范式; 同时, 生成词库论的形式化描写方案也对涉及名词的自然语言处理研究具有一定的启发性作用。

2 结构主义语义学对名词词义的描述

结构主义理论对词汇语义学的发展进程产生了非常重要的影响, 其核心思想是把语言看成一个秩序井然的系统, 而不仅仅是一堆散落杂乱的词语。它主要通过三种理论方法, 即词

* 收稿日期: 定稿日期:
基金项目: 国家自然科学基金项目(61375074, 61371129)。

汇场、义征分析和关系语义对名词词义进行分析。其中，词汇场理论和关系语义学研究的是词语之间的聚合关系，前者注重同义词的聚合，而后者不仅关注同义词，还关注词语之间的其他语义关系，如上下义关系、反义关系、整体部分关系等等；与这两种方法相比，义征分析为词义构成和分解提供了一种技术性较强的分析手段，能够对名词的语义信息进行描写，为人们经验性地判断不同词语是否属于同一词汇场，或是否存在语义关系提供可靠的分析途径。

结构主义的研究对象是“语言”，而不是“言语”，因此它只关注语言内部的结构，希望在对词义进行描写时能在词汇的语言知识和百科知识之间划出一道清晰的分界。但是，仅仅围绕构成语义知识的语言内部结构来划定界限并不是一件容易的事。无论是词汇场理论、关系语义学，还是义征分析方法都无法与百科知识划清界限。词汇场理论将不同词语纳入到词场结构之中，但不同词汇场之间，甚至在同一词汇场内部，都会出现模糊的情况，比如，一个词语可能同时属于两个词汇场，或者属于一个词汇场的边缘成分；面对这种情况，要对它进行判定就必然依赖于百科知识。关系语义学声称词汇意义就是词汇之间的关系，但正如 Murphy^[1]所指出的，意义关系在本质上是元语言的，并不是阐释词汇意义的基础，而是对词汇语义的一种认识。更为重要的是，意义关系并不存在于词语与指称物之间，而存在于词语与对其进行解读的具体语境之间，因为语境因素对词语的意义关系具有重要的影响；那么，既然需要语言系统内部之外的因素参与确定不同词语之间的意义关系，关系语义学认为意义关系可以对词语进行“内容描写”的观点就很难成立；同时，这也意味着，词语意义的解读是离不开百科知识帮助的。义征分析方法的基础是语义特征，而语义特征恰恰是从我们对所指事物的概念化认识中提取出来的，这种概念化认识无疑应该被纳入到百科知识的范畴之中。

简而言之，虽然结构主义希望能在语言知识和百科知识之间做出清晰的区分，但它所提出来的研究方法都无法避免不牵涉到百科知识，百科知识或多或少地都会参与到我们对名词语义的描写和识解之中。此外，在结构主义的背景下，语义学研究的重心仍然在单个词语的意义分解和词义的聚合性上，而关于词义的组合性特征则很少涉及，这也在一定程度上限制了语义学的研究范围。不过，在自然语言处理领域，结构主义所提出的这种词义特征分解和意义聚合的理念具有相当重要的价值，它直接对一些机器词典，如 WordNet 等语义知识资源的建造具有启发性作用。

3 生成主义语义学对名词词义的描述

在结构主义语义学的义征分析方法的基础上，Katz & Fodor^[2]创立了生成主义语义学（也可称为“卡茨语义学”）。生成主义语义学体现了对结构主义语义学的整合和超越，它将义征分析引入到生成语法之中，使得语义分析形式化为可供描写的框架。它不像结构主义那样通过语义特征对同一语义场的词语进行对比分析，而是尽量使单个词语的不同意义可以形式化地表征为形式语法的一部分。以名词 bachelor 的描写为例：

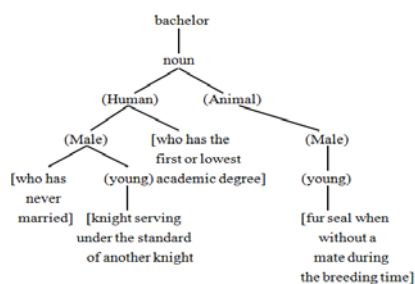


图1 生成主义语义学对名词 bachelor 的描写

在上面表征图中，除了词形和词类之外，主要有两类语义成分：语义标记（semantic

marker) 和辨义成分 (distinguisher)。前者由圆括号表示, 后者由方括号表示。而这两类语义成分决定了词语在句法和语义上组合的可能性。比如, “bachelor” 虽然有不同的意义, 但在 “The old bachelor finally died” 这个句子中, 它只能表示 “单身汉”, 而不能表示其他意义。这受到了语义标记的影响, 因为只有表示 “单身汉” 意义的语义标记才与 “old” 在语义上协调, 而表示其他意义的语义标记都与 “old” 相互排斥。由此可见, 生成主义的词义描写方法可以将词语的词汇意义组合投射到短语结构的意义之中, 从而能够对词项之间在句法和语义组合上的选择限制进行操作。相对于结构主义过分注重词义的聚合描写, 生成主义语义学的这种词义描写方法为词义的组合性研究打开了通道, 进一步拓宽了语义学研究的范围。更为重要的是, 这样一种词义分解的方法也为计算机处理自然语言提供了重要的理论支撑, 在词义消歧、机器翻译等方面能够发挥积极的作用。

不过, 在对待语言知识和百科知识的问题上, 生成主义语义学依旧遵循结构主义的基本倾向, 试图在语言知识和百科知识中做出区分。Katz & Fodor^[2]认为语言研究的重点应当集中在语言知识上, 而不是世界知识上, 对语义设置一个上限范围很有必要, 要把语义研究与情景因素等隔离开来。但是, 从生成语义学对词义的处理范式可以发现, 百科知识仍然在词义描写中留有痕迹, 上图 bachelor 的词义描写中的那些语义标记正是这样一种反映, 它们是从词义所指事物的百科知识中所抽取出来的语义标签。

此外, 从上图对 bachelor 的语义描写可以发现, 生成语义学对词语的多义性问题也给予了关注, 这体现了生成主义对于词义的原子论 (atom) 处理态度, 即把词义分解为多个不可再分的最小词义单位。然而, 生成主义对于多义词词义的处理是一种静态式的罗列, 因为这些最小词义单位之间是分割的、互不联系的, 它们之间的动态关系, 即多义词的不同义项是如何在句子中推导产生的, 这一点并不能通过生成语义学的描写方法展现出来。而对于多义词不同义项之间的内在逻辑关系以及词汇的创造性使用特点的研究, 可能更加有助于窥探人们是如何识别词义以及对于句子意义进行理解的, 同时对于词义消歧等计算机语言处理任务来说都会有一定的作用。

4 概念语义学对名词词义的描述

概念语义学由 Jackendoff 一手创立, 在他看来, 意义就是外部世界在人们心理所形成的一种表征方式, 这种表征方式可以被称为概念结构, 它充当了语言与世界的接口。概念结构由一些概念基元构成, 比如, EVENT (事件)、STATE (状态)、THING (物体)、PATH (路径)、PLACE (方位)、PROPERTY (属性) 等, 而这些概念基元在说不同语言的民族中都是共通的, 具有普适性。在这些概念基元的基础上, 可以形成各种简单或者复杂的概念结构, 再由概念结构投射到语言层面, 形成各种各样的句子。因此, 在概念语义学对词项的描写框架中, 我们可以看到词汇语义与句法框架的结合。以动词 “run” 的描写框架为例:

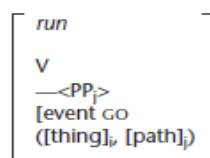


图 2 概念语义学对动词 run 的描写

词项下方的第一行是词类标记, 第二行是词项出现的句法环境, 第三行是词项的概念结构, 而概念结构实际上就对应了 run 所表达的语义: 一个事件中的某物 (主语) 沿着介词短语所表示的可选路径移动。从概念语义学对词项的描写方式可以发现, 它对于词义的处理基本上采取的是一种分解式的态度, 即把词义分解为若干个概念基元, 而究竟用什么样的概念基元进行描写则完全取决于词项本身在现实世界中的表征方式。因此, 从这个角度看, 概

概念语义学试图将语言知识与百科知识结合起来,而不是对立,从而界定两者之间的可能分工。从自然语言处理的角度看,概念语义学对于词条的描写框架与 Fillmore 所创立的“框架语义学”(Frame Semantics)^①具有相似之处,前者的“概念结构”其实对应的就是后者的“概念框架”,通过描述概念结构中的参与者元素在句法上的位置实现,建立特定词项的句法和语义框架;利用这样一种方法所建立的知识资源库(如 FrameNet^②)在机器翻译、人工智能等领域起到了重要的作用。

此外,概念语义学主要是对单一或复杂的事件进行描写,EVENT 和 STATE 是两个最基本的概念情景,而其它概念基元都是这两个情景的参与者。因此,对于由名词充当的 THING 这一概念基元来说,概念语义学只能以[±bound]和[±internal structure]这两对语义特征来加以界定^[3]。换言之,概念语义学对于名词的语义信息描写是非常不足的,这也是多数语义学理论的相同点,因此,发展出一套能对名词的语义信息进行描写的框架是非常有必要的。

5 自然语义元语言理论对名词词义的描述

由 Wierzbicka^[4]创立的自然语义元语言理论秉持着这样一种假设:词语的意义隐含在基元词中,基元词是基元概念的表征形式,基元概念本身无须界定。这样一种假设的认识论基础是:如果一种语言里的所有词语都用其他词语来定义,那么就可以保持在语言和世界之间不发生任何联系,语言内部就是一个纯净的系统。在该理论中,最为基本和核心的概念就是基元概念,它被认为具有普遍性,是不同种族人群都具有的,并且,这些概念在所有语言中都是以某个特定词语或至少某个特定形式来表达。根据 Goddard^[5]的统计,已经识别的语义基元有 62 个,这些语义基元在不同的语言中具有不同的表征形式。因此,根据自然语义元语言理论,对于词语意义的界定,就可以通过这些语义基元来加以说明。以动词 break 为例,它可以按照下面这种方式定义:

Person-X broke Y (e.g. Peter broke the window) =
X did something to Y
because of this, something happened to Y at this time
because of this, after this Y was not one thing any more

上面这个定义所用的所有词语都是从语义基元中抽取出来的,自然语义元语言理论正是试图通过这样有限的、为数不多的语义基元来对语言中的所有词语进行定义。

很显然,自然语义元语言理论的主体思想是从结构主义所提出的语义特征那里继承而来的;不过,与语义特征数目的无规定性和无限性不同,这种理论方法假定,只需要从数量固定的概念基元的集合中抽取其中的一小部分,再通过组合操作就可以对词义进行描写。这种假设类似于化学中的元素学说:世界上的事物千变万化,但都是由一百多个元素通过化学物理反应组合转化而成的。

但是,这样一种词义描写方式存在方法论上的问题。在对词语进行语义描写的过程中,选取什么样的概念基元和怎样对概念基元进行组合会在很大程度上影响词义描写的精确性;另外,概念基元本身是否可进行再分解,分解的标准是什么,不可分解的标准是什么,这些都不容易确定。最根本之处还在于:这些语义基元是通过对自然语言的意义进行描写的过程中所假定的,目前并没有任何证据显示这些语义基元具有普遍性,因此,也就不能确定这样一套词义描写方法对于所有语言来说具有普遍的效力。利用语义基元给词语所下的定义的准

^① 框架语义学是 Fillmore 提出的研究词语意义和句法结构意义的语义理论方法,该理论主张对于词语意义的描述必须联系特定的语义框架,因为框架是信仰、实践、制度、想象等概念结构或概念模式的图式表示,是言语社团中人们相互交流的概念基础^[6]。

^② FrameNet 是由加州大学伯克利分校所建立的框架网项目,是基于框架语义学、并以语料库为基础建立的在线英语词汇资源库。关于该资源库的具体介绍,可参见袁毓林、李强^[7]。

确性也就值得怀疑。比如，Wierzbicka^[8]在对“杯子”的意义进行说明时提到：

一只中国杯子，不管它大小、厚薄、优美与否，是否有把子或者托盘，只要它明显地适合人们喝热水，在正常环境下能够用一只手举到嘴边，都可以称为“杯子”。

从这段定义中可以发现，Wierzbicka 认为“杯子”的本质核心概念是“能够从中喝热的液体”和“可以供人们用一只手举到嘴边”。但是，这样的定义准确吗？如果抽屉里有一个装满夹子和铅笔的近似于杯子形状的容器，虽然它也具有上面两种核心概念，但和我们所说的“杯子”很显然不是同一个事物。由此可见，自然语义元语言理论采用概念基元对词语所下的定义还是存在一些问题。

此外，从语言与世界的关系来看，虽然该理论的倡导者希望用词语定义词语的描写方式来与世界知识划清界限，但它并没有解决语义基元是否与外部世界相联系的问题。正如 Geeraerts^[9]所言：“像 touch 之类的基元，感知体验必须加以考虑，以免我们把语言看作一个理想的独立王国，而与语言之外的世界完全隔绝开来”。

6 生成词库理论对名词词义的描述

由 Pustejovsky^{[10][11]}创立的生成词库理论被誉为“当代语义学中最精细的形式化分析范式”^[9]，在词汇语义学的研究中引起了极大的关注。生成词库论认为词义并不是静态地、缺乏联系地被列举出来，而是具有动态生成性特征，即词语在不同语境下的意义都可以通过一套语义生成机制被生发出来。生成词库理论所要达到的目标就是要挖掘出这种语义生成机制，这是该理论对词义研究所做出的最大贡献。Pustejovsky et al.^[12] 更为明确地将这种理论目标概括为：

开发生成词库的目的在于提供一种成分组合语义学，对源自现实语言使用中的意义进行语境调制。

以这种目标为导向，生成词库解释了语义研究的基本内容：

- (a) 解释语言的多样性本质；
- (b) 刻画自然语言话语的语义系统；
- (c) 解释词语在新语境下的创造性用法；
- (d) 发展出一套更为丰富的共同组合的语义表征体系。

伴随上述语义研究的目标和内容，生成词库理论首先着眼于构建一套崭新的词义描写方法，这套词义描写方法是由四个层面的语言知识所组成的一个层级体系；并且，由于生成词库论的创始人 Pustejovsky 教授的研究兴趣重点在自然语言处理方面，所以，这套词义描写方法自创立之始就深深地打上了自然语言处理的烙印，它以一种形式化的方式表征出来：

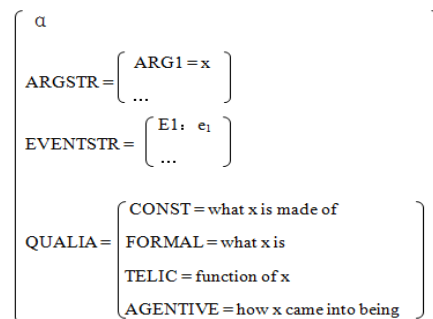


图 3 生成词库理论对词项的描写框架

在上图中，ARGSTR 代表论元结构，主要描写论元的数量、类型以及句法层面的实现；EVENTSTR 代表事件结构，主要描写动词语义中的事件信息，包括状态、过程、迁移等；QUALIA 代表物性结构，是一套关于词汇本体知识的描述体系，主要包括四种物性角色：形式角色 (formal role, FORMAL)、构成角色 (constitutive role, CONST)、施成角色 (agentive

role, AGENTIVE) 和功用角色 (telic role, TELIC)。此外, 未在上图中展示出来的词汇类型结构 (Lexical Typing Structure) 也是生成词库论对词项的描写内容之一, 主要用来说明一个词项在一个类型系统中的位置, 以及与其他词项的语义关联方式, 如上下位关系、整体部分关系等等。这里, 我们需要重点对物性结构进行说明, 因为它是生成词库理论最为核心的内容, 同时也是构建词义的基础, 为语言知识和百科知识的表征提供了接口。

物性角色的根基源于亚里士多德的“四因”说, 在亚氏看来, 世界上的万事万物都具有四个方面的属性, 即形式因、质料因、动力因和目的因, 这四方面属性汇聚了我们对于事物的一般性、常规性的认识和看法, 反映了我们对于事物最为朴素和深刻的概念化理解方式。换言之, 这四因其实就是一种百科知识, 在此基础上所创立的物性结构描写体系, 旨在以一种系统化、条理化的方式对词项 (主要是名词) 的语义结构进行描写, 以求能够更为直接地反映我们对于词项单位的基本感受和认识。Pustejovsky 希望通过这套物性结构体系, 能够将亚氏所总结的百科知识以形式化的方式直接和语言平面挂钩, 并使之参与到我们的言语理解和话语建构之中。对应于亚氏的“四因”, 上述四种物性角色的定义分别如下:

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、大小、形状、维度和颜色等;

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、重量、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;

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;

功用角色 (telic role): 描写对象的用途和功能。

这四种物性角色展示了名词指称事物所涉及的百科知识, 极大地丰富了名词的语义内涵, 对于深入刻画名词的语义结构和信息具有重要的指导作用; 同时, 它们都是名词所涉及的基本角色, 有着很强的心理现实性。更为重要的是, 它们将人们对于名词指称事物的基本感受和认识以形式化和表征化的方式展示出来, 让名词的语义信息变得不再飘忽不定从而难以把握。

从生成词库理论对词义的描写框架中, 可以明确以下几点: 1. 结构主义理论、生成主义理论和自然语义元语言理论认为语言知识和百科知识应当划清界限, 而生成词库理论的态度与它们截然不同, 强调百科知识与语言知识没有界限; 在生成词库的框架下, 语义问题不是局限于语言系统内部所讨论的对象, 而是牵涉到人们对于世界的了解和认识状态, 这种百科知识会对话语的理解和建构、话语形式的塑造产生重要的影响, 而物性角色为百科知识的描写提供了一种形式化很高、操作性很强的解决手段^①。2. 从 Katz & Fodor 所创立的生成语义学理论来看, 生成词库很明显地受到了它的影响, 也试图为词义提供一个可分解的模式, 不过, 生成词库的理论目标决定了它与生成语义对词义的处理策略是不同的: 生成主义对多义词的词义采取的是静态罗列的手段, 词义之间是分割、互不联系的, 而生成词库则认为多义词的词义之间具有内在的逻辑关系, 词汇具有创造性使用的特点, 词语在组合过程中所浮现的语义, 即词义在不同语境下的动态生成, 可以通过物性结构对词汇的描写实现。这一点也与结构主义仅仅关注单个词语的意义分解和词义的聚合性问题大不一样, 生成词库对于词义组合性问题的研究拓宽了语义学研究的范围和手段。3. 与概念语义学用两组语义特征对名词进行描写不同, 生成词库的描写更加全面和丰富, 名词所包含的语义结构和信息真实地反映了人们对名词所指事物的概念化认识, 因此, 与概念语义学将百科知识应用到动词 (事件)

^① 对于生成词库将百科知识纳入到名词词义描写的做法, 一些学者也提出了质疑和批评。比如, Fodor & Lepore^[15]就曾极端地认为: 如果词项能够表达意义, 那么词库中应该保留的是关于词项的信息, 而不是关于世界的信息; 词义的内容应该严格区分语言知识和世界知识, 语言之外的推理性因素 (指物性角色) 不应该被纳入到词义描写的框架之中。此外, 生成词库的做法也与以 Katz 为代表的语义分解派相抵触, 后者同样认为词义具有客观性, 在分析语义时必须将语言知识和人的百科知识分开。总之, 诸如语言知识和百科知识的区分、语言研究, 特别是语义研究是否应该引入百科知识等问题, 一直都在引起广泛的讨论。

描写的充分性相对，生成词库则是将百科知识充分地应用到对名词（事物）的描写，弥补了目前语义学理论描写名词的语义结构和词义信息的不足。4.自然语义元语言理论利用语义基元对词义进行描写，但一方面，正如上面所看到的它对于“杯子”的定义，这种描写方法给词语所下的定义不具有准确性；另一方面，语义基元并未通过验证，其可靠性值得怀疑。而生成词库的语义描写则是建立在人们对事物可感知、可把握的具体层面之上的，因而具有可靠、真实的概念基础，因此，相对于没有明确任何定义方式的语义基元，生成词库对于词义的描写信息才是实实在在的共存于不同民族之中的普遍概念内容。5.在自然语言处理领域，虽然结构主义的语义特征和语义场、生成语义的词义分解、概念语义的概念结构和自然语义元语言的概念基元等都是重要的语义自动处理的形式模型，但它们对于名词语义信息结构的刻画都存在或多或少的不足，不能以一种全面的形式化的描写方式被计算机学习和储存；而生成词库论的相关主张可以弥补这一不足，为计算机建立一个强可学性（learnable）的名词知识资源库提供基础的理论支撑。

7 基于生成词库理论的有关名词分析的案例及其对自然语言处理的启发

生成词库理论所提出的名词的物性结构知识及其描写框架，可以为解释汉语中的一些复杂问题，特别是涉及名词的句法、语义现象提供理论支持；并且，对于这些语言现象在自然语言处理方面的应用也能提供一定的启发性作用。

7.1 名名偏正结构中的词语缺省和语义蕴含

袁毓林^[13]提出“谓词隐含”的概念，以此说明“的”字结构中的名词（NP₁）和中心语名词（NP₂）之间存在某种述谓关系（predicative relation）；宋作艳^[14]也较为详细地考察了“的”字结构中谓词的隐现情况。可以说，“谓词隐含”现象的提出能够帮助我们还原特定语言结构所隐含的语义内容，例如：

- (1) a. 红木的家具→红木**制造**的家具
- b. 晚间的新闻→晚间**播放**的新闻
- c. 书架的参考书→书架**摆放**的参考书
- d. 左手的手套→左手**戴**的手套

囿于理论的匮乏，当时只能限于指出整个结构隐含谓词这一现象，还不能有效地预测和限定不同的结构具体隐含的是什么样的动词。但是，现在有了物性结构的理论支撑，我们就可以大胆地对此进行预测：“家具”的施成角色是“制造”，“新闻”的施成角色是“播放”，所以 a 和 b 中隐含的动词是“制造”和“播放”；“书架”的功用角色是“摆放”，“手套”的功用角色是“戴”，所以 c 和 d 中隐含的动词是“摆放”和“戴”。也就是说，名词的物性角色决定了上面这些名名偏正结构中能隐含且只能隐含特定的动词。与此相关，在物性角色的基础上，我们还可以进行系统的推广，以确定某些不同语义类型的名名偏正结构中所隐含的动词。比如，在以 a 为代表的表示质料（材料）义的结构中，隐含的动词是 NP₂ 的施成角色，如（2）所示；在以 b 为代表的表示时间义的结构中，隐含的动词可以是 NP₂ 的施成角色，如（3）所示；在以 c 为代表的表示处所义的结构中，隐含的动词可以是 NP₁ 的功用角色，如（4）所示；在以 d 为代表的表示非固有领属义的结构中，隐含的动词可以是 NP₂ 的功用角色，如（5）所示。

- | | |
|----------------------------|-----------------------------|
| (2) 木头的桌子→木头 制作 的桌子 | 缎子的被面→(用) 缎子 做/缝 的被面 |
| 羊肉馅的包子→羊肉馅 做 的包子 | 大理石的雕像→大理石 刻 的雕像 |
| (3) 昨天的报纸→昨天 出版 的报纸 | 中秋的福利→中秋 发放 的福利 |
| 冬季的运动会→冬季 举办 的运动会 | 兔年的火灾→兔年 发生 的火灾 |
| (4) 仓库的粮食→仓库 储存 的粮食 | 楼顶的房客→楼顶 住 的房客 |

- 抽屉里的卡片→抽屉里**放**的卡片 衣架上的衣服→衣架上**挂**的衣服
 (5) 小张的铅笔→小张**写**字的铅笔 小王的衣服→小王**穿**的衣服
 小李的飞机→小李**坐**的飞机 小陈的首饰→小陈**戴**的首饰

这样，通过确定名名偏正结构表达什么样的语义，就可以利用物性角色来对结构中隐含的动词进行补充；而如何还原出名名偏正结构中的动词也是计算语言学界研究的重要问题，因为它对信息检索、问答系统、机器翻译等诸多自然语言处理任务都有所帮助。现在有了物性角色这一层语义知识，我们就可以在名词语义类的基础上归纳不同语义类组合所形成的释义模板，而释义模板中的动词就可以通过物性角色来加以咬合，最终开发出一套名名偏正结构的释义系统。

值得一提的是，魏雪^[16]以及魏雪、袁毓林^[17]利用上述方法构建了汉语名名组合的释义模板库，并且在这种数据库的基础上初步实现了一个汉语名名组合的自动释义程序，其自动释义的准确率达到 94.23%。

此外，名词物性角色还可以对传统语法框架下不易说明的名名偏正结构的歧义现象进行统一的解释。例如：

- (6) a. 鲁迅的书——鲁迅**写**的书/鲁迅**藏**的书/**讲述**鲁迅的书
 b. 玻璃的碟子——玻璃**制造**的碟子/玻璃的窗户——*玻璃**制造**的窗户

a 例中，“鲁迅”除了作为一个个体的普通人之外，还有自己特殊的身份标签“作家”、“文化名人”，所以在“鲁迅的书”中，当“鲁迅”作为一个普通人时，他可以像很多人一样“藏书”，“藏”是“书”的处置角色^①；当他是一位作家时，他还可以“写书”，“写”是“书”的施成角色；当他是一位文化名人时，他还可以被别人“讲述”，“讲述”是“书”的功用角色。正是因为名词“书”具有不同的物性角色，再加上“鲁迅”身份的特殊性，能够使得上述结构产生歧义。b 例中，“碟子”的构成角色和施成角色分别是“玻璃”和“制造”，所以，“玻璃的碟子”可以还原为“玻璃制造的碟子”，与之相对，虽然“窗户”的施成角色是“制造”，但其构成角色不仅仅是“玻璃（面）”，还包括“木头（框）”，所以，“玻璃的窗户”不能还原为“玻璃制造的窗户”，而是“玻璃制造的窗户的面儿”。可见，NP₂的构成角色会对“玻璃的 NP₂”结构解读为“玻璃制造的 NP₂”还是“玻璃制造的 NP₂的 NP₃”产生限制。

同样地，面对这类有歧义的名名偏正结构，如何识别该结构的歧义性及消解歧义也是自然语言处理所面临的一个重要问题。不过，可以确信的是，要想真正解决好这类问题，就离不开对名词语义信息的掌握和描写，使计算机拥有一个关于名词所指事物属性特征的知识库，而物性角色在这个知识库的构建中则能发挥积极的作用。以(6a)为例，如果计算机知道“鲁迅”有个特殊的身份“作家”，有其功用角色“写”；是个普通的人，有其行为角色“藏”^②；是个文化名人，有其处置角色“讲述”；并且，在知识库中还储存着“书”的施成角色“写”，处置角色“藏”和功用角色“讲述”，那么，通过某种特征加权机制和以动词为中心的语义关联机制，就可以为“鲁迅的书”获得“写、藏、讲述”这些隐含的释义动词。最终，不仅可以帮助计算机完成歧义结构的识别，而且还可以获得歧义结构的多种语义解读。再以(6b)为例，如果计算机具有“碟子是玻璃做的”这一语义知识，即知道“碟子”的材料角色是“玻璃”^③，通过识别名名偏正结构中的第一个名词也是“玻璃”，就可以在“碟子”的施成角

^① 处置角色 (handle role) 是袁毓林^{[18][19]}对物性角色的六种扩展描述中的一种，意思是：人或事物对名词所指事物的惯常性的动作、行为、影响。比如，对“水”的处置是“打、各、取、蓄、洒、放、排、倒、喷、波、玩儿”等等。

^② 行为角色 (act role) 指名词所指事物的惯常性动作、行为、活动。比如，“水”的行为是“流、流动、奔腾、翻滚、滴、淌、流淌”等等。

^③ 材料角色 (material role) 指创造名词所指事物所用的材料。比如，“椅子”的材料是“木头、竹子、藤子、木、竹、藤、钢、铁、塑料、硬板”等等。

色上定义规则，解释该结构的语义是“玻璃制造的碟子”；相反，如果计算机知道“窗户”的材料角色是“玻璃和木头”，构成角色是“面和框”，“面”和“框”分别由“玻璃”和“木头”制成，通过识别该结构中的第一个名词是单独的“玻璃”，就可以定义规则排除“玻璃制造的窗户”这种语义解释，而生成“玻璃制造的窗户面”这一更为准确的语义。

总而言之，名词物性角色可以对名名偏正结构中能够隐含什么样的动词进行简单、整齐的限定和说明，也能对与之相关的一些语言学现象给出深层原因的解释；同时，也能够为名名偏正结构在自然语言处理中的释义问题提供一定的帮助。

7.2 名词隐喻义的语义生成

名词隐喻是语言中广泛存在的一种语言现象，关于隐喻义是如何产生的这个问题，我们一般都是诉诸于对于名词所指事物的百科知识，依靠经验性认识尝试在不同事物之间建立起相似性的关联，从而将其中一个事物的属性特征转嫁到另外一个事物之上。名词物性角色的提出为我们观察名词所指事物的内在联系，以及解释名词隐喻义的生成提供了一种方式。例如（引自李强^[20]）：

- (7) a. 工业是国民经济发展的核心
- b. 他是个猴子
- c. 造论文
- d. 树的臂膀

a 例想要表达的是工业对于国民经济发展的核心性，这个意义的生成是依靠将喻体名词“核心”的物性结构中的功能角色投射到本体名词“工业”这一过程实现的。因为“核心”是人体最重要的生命器官，其功用角色是维持生命的基本运作；同样，“工业”在整个生产领域也是处于绝对重要的地位，是经济发展的推动剂。这种功能上的相似性使得“工业”和“核心”之间的功用角色的映射容易被激活和提取的，因此，我们很容易将“核心”的功用角色转嫁到“工业”上，从而完成对“工业”的隐喻义识解。

b 例想要表达的是他聪明、灵活，这是依靠将“猴子”物性结构中的形式角色“机灵聪明”投射到本体名词“他”这一过程实现的，从而说明本体名词“他”也具有类似的特征。

c 例中，“论文”一般都是“写”出来的，但也可以用“造”，实际上是因为我们把论文当作是一个建筑，“建筑”的施成角色“造”投射到“论文”之上，形成“论文”比作“建筑”的隐喻义。

d 例中，“树”的构成角色是“枝干”，“枝干”与“树”构成领属和被领属的关系；“人”的构成角色是“臂膀”，“臂膀”和“人”构成领属和被领属的关系。基于“树”和“人”在构成角色上的相似性关系，我们会将“人”的特性赋予“树”上，形成“树比作人”的隐喻。

从上面的举例可以看出，名词物性角色对于我们建立两个名词之间的隐喻式理解起了非常重要的作用，正是在物性角色的基础上，我们可以在相关名词之间建立起内在的联系，从而确保我们能够轻松自如地理解一些非常规的组合形式。同时，深入细致地刻画名词的物性角色信息对于计算机理解隐喻这类问题也能起到一定的作用。王治敏^[21]在谈到汉语名词隐喻的语义映射分析和名词隐喻所隐含的思维模式时，就曾指出，隐喻的映射就是把那些已知领域物体的典型特征及相关属性映射到相对陌生的事物属性上。“n+n”隐喻映射是全方位的，不仅包含名词的名称，还包括属性、功用等宽泛的映射范围。同时事物所具有的“长、宽、高、点、线、面、体积”等几何特征在名词短语隐喻中都有所体现。最后，他认为建立一个汉语隐喻知识库，把名词隐喻重要的属性特征形式化到知识库中是十分有必要的，这对于计算机识别理解隐喻式表达非常有帮助。

可以预见，这个知识库的建立离不开物性结构的指导作用，只有把隐喻名词的物性结构，或称为属性特征刻画清楚了，才能归纳名词隐喻的基本类型。

7.3 供用句对中动词的选择约束

供用句是汉语中一种非常特殊的句式结构，表达“某些存在物被提供给某些人使用”的语义。关于这种句式的形成机制问题，已有的研究都存在一个无法解释的问题，即为什么供用句对其中的动词有特殊的选择要求，有些供用句可以成立，而有些则不行。请看下面两组句子的对比：

- | | |
|-------------|-------------|
| (8) 一锅饭吃十个人 | *一锅饭做/要十个人 |
| 一张沙发坐三个人 | *一张沙发抬/搬三个人 |
| 一只箱子装五件大衣 | *一只箱子盖五件大衣 |
| 两桶漆刷一面墙 | *两桶漆放一面墙 |

由这些例子可见，供用句中的动词是有其特殊选择限制的。现有研究虽然可以对上面左侧这些“能说”的供用句具有较好的覆盖面，但无法限定排除右侧这些“不能说”的供用句。因此，有必要从一个新的角度对供用句的上述差异做出解释，而名词物性角色则为解释这种差异提供了一种有效的手段。

观察上面这些例子，我们发现：主语名词物性角色中的功用角色与供用句中的动词具有高度的一致性，凡是动词为主语名词的功用角色，这样的供用句可以成立；相反，如果动词不是主语名词的功用角色，这样的供用句则不能成立。例如，“饭”的功用角色是“吃”，而不是“做/要”；“沙发”的功用角色是“坐”，而不是“抬/搬”；“箱子”的功用角色是“装”，而不是“盖”；“漆”的功用角色是“刷”，而不是“放”。这样，通过物性角色这一概念就可以简单地给供用句对动词的选择设置一条没有例外的规定，具有很强的解释力。

对于自然语言处理任务来说，供用句给计算机自动标注句子成分的语义角色带来了困难。汪昌松、靳玮^[22]指出，在对供用句进行自动语义标注过程中，以动词为中心，无论是常规句还是供用句，计算机很容易将动词前的成分标注成施事，而将动词后的成分标注为受事。但很显然，这是不正确的。常规句和供用句动词前后的语义角色刚好相反，在常规句中，施事成分在前，受事成分在后，而在供用句中，施事成分在后，受事成分在前。因此，这会对计算机自动识别语义角色加以干扰。现在，我们发现供用句中的动词都是句首名词的功用角色，这可以为计算机自动判定句式结构，进而进行自动语义角色标注任务提供帮助。如果计算机碰到一个句子中出现双数量词结构，并且它们之间通过动词连接，通过自身存储的有关名词物性角色的知识库来判定动词是否是句首名词的功用角色，如果是，则判定这个句子是供用句，如果不是，则判定这个句子是常规句；在此基础上再进行下一步的语义角色标注任务，就可以实现语义角色的准确标注。并且，在计算机判定句式结构的基础上，还可以通过一定的语义规则实现对供用句这一特殊句式的语义解释，比如，可以规定计算机碰到“一锅饭(NP1)吃(V)三个人(NP2)”这样的句式，就语义解释为“一锅饭(NP1)可以供/让三个人(NP2)吃(V)”，这对文本理解、机器翻译和信息抽取等方面都是有一定帮助的。

7.4 中动结构成立的合格性条件

中动结构是不同语言中广泛存在的一种现象，是“主动形式表被动意义”^[23]。在汉语中动句的研究中，一个引起学者们普遍关注的问题是中动句中动词的语义性质和用法特点，比如，Sung^[24]和曹宏^[25]都曾讨论过中动句对动词的选择限制条件。其中，曹宏^[25]将汉语中动句对动词的选择限制条件归纳为“只有自主动词中的及物动词才有可能进入中动句，非自主动词中的及物动词不能进入中动句”。应该说，动词的自主性和非自主性特征的确是中动句能否成句的一个非常重要的条件。但是，我们也发现有一些例子的合格与否是“自主性”不能完全地加以解决的。例如：

- (9) a. 这辆车开起来很容易。

- b. *这辆车踢起来很容易。
- (10) a. 这件衣服穿起来很舒服。
b. *这件衣服洗起来很舒服。
- (11) a. 这种苹果吃起来很酸。
b. *这种苹果吃起来很容易。
c. 这种核桃吃起来很容易。

在上面的(9)和(10)中, 每组两个句子除了动词不同外, 其余成分都相同, 所形成的中动句存在合格性上的差异。在(11)中, a句和b句的主语名词和动词都相同, 而形容词不同, 所形成的中动句在合格性上存在差异; 而b句与c句相比, 更换主语名词, 其余部分不变, 所形成的中动句又变得能说。这一系列的对比差异告诉我们: 虽然这些句子中的动词都是自主性的及物动词, 但有的中动句能说, 而有的则不能说; 其中的原因不可能在于动词是否具有“自主性”这一语义范畴, 否则上面这些句子都应该合格。

而从名词的物性角色角度出发就可以对“自主性”解释方案做一个很好的补充, 能够解释为什么上面的9a、10a和11a、c成立, 而9b、10b和11b不成立。在(9a)中, “车”的功用角色是供人们“开”的, 为了更加方便人们开车, 使得人们开起车来感觉更加容易轻松, 汽车生产商可以在汽车的构造设计上下功夫, 或者增加一些部件, 或者对车的内部座椅等设备进行改良等等, 以增加车的“易开性”; 也就是说, “车”本身的形式角色和构成角色与“开起来很容易”这一状态具有语义上的关联性和共构性, 所以, “这辆车开起来很容易”是成立的; 与之相对的是9b, 虽然“车”也可以被踢, 但“车”的基本功能是供人们开的, 它本身没有任何构成部件是为了供人们踢它而设计的, 也就是说, 在我们对“车”的概念化认识中, 它没有物性角色与“踢起来容易”在语义上相互关联, 所以, “这辆车踢起来很容易”也就不合格。在(10a)中, “衣服”的功用角色是供人们穿的, 衣服的布料、质地、柔软程度等都与人们穿它的体验感觉密切相关, 决定是否能够产生“舒服”的感觉, 也就是说, 衣服的构成角色和形式角色与“穿起来很舒服”具有语义关联性和共构性, 因此, “这件衣服穿起来很舒服”能够成立; 相反, 在(10b)中, “衣服”的功用角色不是用来“洗”的, 其形式角色和构成角色与“洗起来很舒服”这种状态之间也没有语义上的关联性, 所以, “这件衣服洗起来很舒服”就不能成立。同样的分析可应用到(11)中, “苹果”的功用角色是用来“吃”的, 口味是它的形式角色, 因此, “苹果”本身的形式角色与“吃起来很酸”具有语义关联性, a句成立; 在c句中, “核桃”具有坚硬的外壳, 如果不用工具就很难将外壳剥除, 核桃的这一形式角色决定了要想吃它就不是一件容易的事情, 换言之, 这种形式角色与“吃起来不容易”所表示的状态具有语义共构性和关联性, 所以c句是合格的; 相反, 在我们对“苹果”的概念化认识中, 它的皮很薄, 不具有核桃那种坚硬的外壳, 也不需要像核桃那样去除外壳才能吃到里面的果实, 所以, 苹果的这一形式角色与“吃起来不容易”所表示的状态不具有语义共构性和关联性, 因而所形成的b句也就不合格。

通过上面几例的分析, 我们可以发现: 从名词物性角色角度出发可以对中动句的成立与否做出解释和说明。在物性角色的基础上, 中动句的合格性条件可以总结为: 主语名词的物性角色与谓语成分在语义上具有共构性和关联性; 名词的物性角色能够在一定程度上促使谓语成分所表示状态的发生。

在自然语言处理领域, 尤其是在问答系统的设计中, 如何能让计算机自动生成上面这些合格的a句, 而避免出现像b这些不合格的句子? 通过观察中动句的生成方式, 发现: 人们在概念系统中对于名词的属性特征的认识在很大程度上决定了后面谓语部分应该采取什么样的动词和形容词, 以实现它们在语义上的融合。那么, 如果能将名词的这种属性特征表征为形式化的知识体系存储在计算机中, 就应该可以为计算机生成合格的中动句提供一定的帮助。以上面(9)为例, 如果计算机知道“车”的功能用途是“开”, 并且“车”的某些内

部构成部件的功能用途也是为了“提高性能，让驾驶者得到更加方便、容易的驾车体验”，通过大规模语料搜索经常与“开（车）”这一动作行为所关联的感受评价类形容词，将这样一种语义关联信息以指针链接的方式存储在计算机中，再辅之以“名词+动词+起来+形容词”这样一种语序规则，最终就可以按照语序规则生成“车开起来容易”这样合格的句子，而不会出现“车踢起来容易”；相反，如果计算机知道“足球”的功能用途是“踢”，通过文本搜索与“踢（球）”语义上相关的感受评价类状态，再通过线性层面上连接相关成分就可以生成“球踢起来容易”的中动句。这样一种语义关联方式可以图示化表示成：

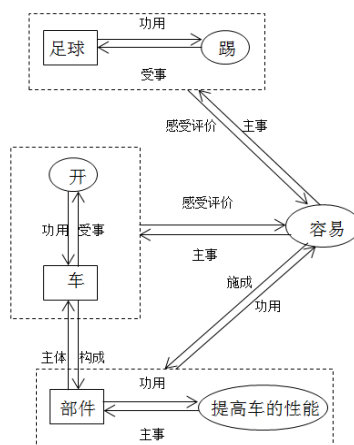


图4 “车”、“足球”、“开”、“踢”、“容易”等成分的语义关联

在调查大规模真实文本语料的基础上，通过名词物性角色、动词和形容词的论元结构描述，可以把上图中各个成分通过语义关系关联起来形成一个语义网络，从而为计算机自动抽取相关成分并生成合格的中动句提供条件。同时，也能为自然语言理解中的复杂的自动推理提供一种基于网络的匹配。

8. 结语

从以上四个案例的分析可见，名词在汉语语法分析中具有非常重要的作用，其地位并不亚于动词和形容词，对于名词词义的描写有助于相关语言现象的说明和解释。生成词库理论所提出的名词的物性结构知识不仅关注名词词义之间的纵向聚合关系，同时也关注名词和动词之间的横向组合关系，尤其是能够解释句子中词语之间根本性的语义组合方式及其组合语义的生成。对于词语缺省、词语限定等动词无法解释或解释不给力的情况，名词的物性结构知识更是起到枢纽和支点的作用，可以有效地帮助说明句子中词语之间的意合方式和运作机制。对于名词语义结构和词义信息的形式化刻画，有助于相关语言现象的分析，同时也能为自然语言处理的相关问题提供重要的语义知识资源支撑。

参考文献

- [1] Murphy M. Lynne. Semantic Relations and the Lexicon [M]. Cambridge: Cambridge University Press, 2003.
- [2] Katz J. Jerrold, Jerry A. Fodor. The structure of a semantic theory [J]. Language, 1963, 39(2):170-210.
- [3] Jackendoff, R. Parts and Boundaries[C]//Beth Levin and Steven Pinker. Lexical and Conceptual Semantics. Oxford: Blackwell, 1992:9-45.
- [4] Wierzbicka, Anna. Semantic Primitives [M]. Frankfurt: Athenaeum, 1972.

- [5] Goddard, Cliff. Ethnopragnatics: a new paradigm[C]//Goddard. Ethnopragnatics: Understanding Discourse in Cultural Context. Berlin: Mouton de Gruyter, 2006.
- [6] Fillmore C. J., Johnson C.R., Petruck M.R. Background to FrameNet [J]. International Journal of Lexicography, 2003, 16 (3): 236 - 250.
- [7] 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, 28(5):1-12.
- [8] Wierzbicka, Anna. Lexicography and Conceptual Analysis [M]. Ann Arbor, MI: Karoma, 1985.
- [9] Geeraerts, Dirk. Theories of Lexical Semantics [M]. New York: Oxford University Press Inc, 2010.
- [10] Pustejovsky, J. The Generative Lexicon [J]. Computational linguistics, 1991, 17(4):409-441.
- [11] Pustejovsky, J. The Generative Lexicon [M]. Cambridge: MIT Press, 1995.
- [12] Pustejovsky J., P. Bouillon, H. Isahara, et al. Advances in Generative Lexicon Theory [C]. Dordrecht: Springer, 2013.
- [13] 袁毓林. 谓词隐含及其句法后果——“的”字结构的称代规则和“的”的语法、语义功能[J]. 中国语文, 1995, 4:241-255.
- [14] 宋作艳. 也谈与“的”字结构有关的谓词隐含[J]. 汉语学习, 2014, 1:20-28.
- [15] Fodor A. Jerry, Lepore Ernie. The Emptiness of the Lexicon: Reflections on James Pustejovsky's "The Generative Lexicon" [J]. Linguistic Inquiry, 1998, 29(2): 269-288.
- [16] 魏雪. 面向语义搜索的汉语名名组合的自动释义研究[D]. 北京大学硕士学位论文, 2012.
- [17] 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, 27(2):172-181.
- [18] 袁毓林. 基于生成词库论和论元结构理论的语义知识体系研究[J]. 中文信息学报, 2013, 27(6):23-30。
- [19] 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, 16(1):31-48.
- [20] 李强. 基于物性结构和概念整合的名词隐喻现象分析[J]. 语言教学与研究, 2014, 6:44-53。
- [21] 王治敏. 汉语名词隐喻的语义映射分析[J]. 语言教学与研究, 2009, 3: 89-96.
- [22] 汪昌松, 靳玮. 从名词的物性结构看容纳句中动词的允准机制[R]. 第十六届汉语词汇语义学国际学术研讨会论文, 北京师范大学, 2015.
- [23] Quirk, R. A Grammar of Contemporary English [M]. London: Longman, 1972.
- [24] Sung Kuo-ming. Case Assignment under Incorporation [D]. University of California at Los Angeles, 1994.
- [25] 曹宏. 中动句对动词形容词的选择限制及其理据[J]. 语言科学, 2004, 3(1):11-28.

作者简介：作者一李强（1988—），博士研究生，主要研究领域为汉语句法学、语义学和语用学，也涉及计算语言学和中文信息处理。Email: leeqiang2222@163.com。作者二袁毓林（1962—），男，北京大学中文系教授、博士生导师，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；



作者一



作者二