

语言学中的形式语义学

李可胜

【提要】形式语义学是介于语言学和逻辑学之间的交叉学科,它以理论语言学为语言研究的理论依据,以数理逻辑的方法为语言研究的工具,目标是对自然语言的语义进行形式化描述,从而实现机器对自然语言的自动理解。最近的30年,形式语义学在国际上发展迅猛,新的理论不断涌现,并且与理论语言学、计算语言学交叉互动。但是作为理论语言学的实践应用领域之一,形式语义学在我国的发展却严重落后于整体语言学的发展进程。

【关键词】语言学 形式语义学 自然语言处理 语义模型

【中图分类号】HO **【文献标识码】**A **【文章编号】**1000-2952(2009)02-0112-06

自然语言的形式语义学(Formal Semantics)又被称为逻辑语义学(Logical Semantics),以上个世纪70年代的蒙太格语法为开端。经过近40多年的发展,在欧美发达国家已成为当今的热点课题,理论分支异彩纷呈,发展势头迅猛。^①其内在的原因:形式语义学是语言研究与逻辑学的交叉领域,为计算语言学的自然语言处理提供了语言学理论支持和计算方法论的支持。但令人遗憾的是,相对于我国庞大的语言学研究群体,形式语义学的发展严重滞后,其理论和实践价值远未被人们所充分认识。

一、形式语义学和理论语言学的关系

形式语义学是语言学和逻辑学的交叉学科,其研究人员主要来自于这两大学科,研究方向也有语言学偏向和逻辑学偏向之分,本文主要针对前者而言。语言学中的形式语义学是直接利用理论语言学的研究成果(主要体现在语义研究以及语义、句法和语用互动界面的研究上),用严格的形式化方法刻画自然语言的语

义。计算语言学家冯志伟教授曾指出:“计算机对自然语言的研究和处理,一般应该经过如下三个方面的过程:第一,把需要研究的问题在语言学上加以形式化(linguistic formalism),使之能以一定的数学形式严密而规整地表示出来;第二,把这种严密而规整的数学形式表示为算法(algorithm),使之在计算上形式化(computational formalism);第三,根据算法编写计算机程序,使之在计算机上加以实现(computer implementation)”。^②形式语义学做的就是第一个方面的工作,因此,可以说形式语义是理论语言学与计算语言学之间的桥梁。没有形式语义学做中介,理论语言学很难直接与计算语言学中的自然语言语义处理做“无缝对接”,这使得形式语义学在今天这个信息技术时代显得尤为重要。

(一)众所周知,计算语言学中的语义分析处理要比语形分析处理复杂得多(本文中的语

^① 参见邹崇理《逻辑、语言和信息——逻辑语法研究·引论》,人民出版社2002年版。

^② 冯志伟:《自然语言的计算机处理》,上海外语教育出版社1996年版,第1页。

形是指语音的物理属性和文字的句法结构属性)。自然语言的语义可以随着所指、处所、时间以及语境的不同而发生变化,而且语义本身也具有不确定性。如“母亲”这个词在人们头脑中唤起的概念要远远超出女性、成年、已婚、有子女这样一些义项,它还应包括抚育子女的社会责任、疼爱的个人亲情感觉、牺牲奉献的原型形象、繁衍后代的象征等,而且对不同的人,这个词唤起的概念也不尽相同。

语义所指本身也具有模糊性。比如当我们说“三班的同学睡觉了”时,可以指“三班的所有同学都睡觉了”这一情形,但如果有一两个同学没有睡觉,这句话也能被接受,因为“三班的同学”可以看成是一个整体。就如同“三班的同学搭起了帐篷”并不一定要求每一个三班的同学都参与到搭帐篷的活动中,甚至不一定要求三班的大多数同学参与,只要这个帐篷是某些三班同学搭起来的,这句话是可以接受的。

(二) 由于语义的复杂性,用形式化方法刻画语义的难度大大超出人们最初的预期。也正是这一原因,一些语言学家甚至放弃了对语义做形式化的努力。如乔姆斯基在自己的句法理论中就放弃了语义部分,认为句法是独立于语义而存在的。但是随着语言学的发展,尤其是近年来,认知语言学和语用学的成熟以及构式语法理论的兴起等,语言学对语义问题以及语义、语用和句法交互界面的研究已经是今非昔比。语义结构不再被看成是杂乱无序的,人们对语义和句法二者的关系也有了新的认识。

按照认知语言学的理论,语言系统并不是直接将世界知识投射到语言系统内形成语义的。相反,当客观世界的信息反映到人脑时,要受到人脑认知机制的操控,如基于原型的范畴分类,隐喻和转喻机制作用等,从而形成概念。当这些概念对应于特定语言符号时,就是自然语言的语义。^① 不仅如此,语言推理也有人脑中概念结构和认知模式的参与。由于人脑的认知具有特定的模式,语义的结构也就有章可循。而且认知语言学认为,不仅词汇的指称是语义

系统的一部分,句法系统本质上也是语义系统的一部分,语言的词汇和语法结构都是不同层次的语义单位。^②

语言学近年来的热点之一就是在语义、句法和语用三者交互界面上的研究,即将语义、句法和语用结合起来考虑,解释复杂的语言现象,近十年来兴起的构式语法(Construction Grammar)就是其中的一个典型。构式语法中的构式(Construction),是指语言中相对固定的形式与语义的结合体。词汇是一种构式,而句法结构也可以像词一样表示某种独立的意义,称为构式语义。换句话说,构式和词汇一样,也是一种意义与形式对应的实体。从广义的角度讲,任何语言表达式,只要它的形式或意义的某个方面严格地说不能从其组成部分推导出来,就都属于构式。

(三) 认知语言学、构式语法都是当今理论语言学所取得的重要成果,也对其理论产生了重要影响。但也都存在着共同的问题,即没有形成像转换生成语法那样具有严密结构的形式化体系,因而不具有可计算性。也许正是出于这个原因,这些涉及语义的语言学理论没有对计算语言学产生很大影响。与此形成鲜明对照的是,当年具有形式化体系的转换生成语法一经出现,便得到计算机界的青睐。时至今日,乔氏的一些思想,如句法改写规则、递归构造、句法树分析等,在各种计算语言学理论中,如广义短语结构句法,词汇功能语法,中心语驱动语法等,都或多或少能看到一些影子。

而认知语言学、构式语法等虽然能对许多语言现象给出满意的解释,有些还在局部问题上形成一定的体系(如Fauconnier的概念空间理论等),但却始终缺乏一个完整的理论体系,更谈不上形成严密的形式系统。因此,虽然这些理论研究在语言学内开展得如火如荼,但计

① 参见赵艳芳《认知语言学概论》,上海外语教育出版社2001年版,第3章。

② 参见R. W. 兰盖克《认知语法基础(1):理论前提》,北京大学出版社2004年版,第1章。

算语言学界一方却保持着冷清的状态。据笔者统计,以我国最主要的自然语言处理杂志《中文信息处理》为例,从2000年到2007年发表的200多篇研究论文中,关键词中带认知两个字的仅有4篇,这与语言学内部的认知语言学热潮形成鲜明的对比。

形式语义学的目标是将自然语言的语义转换成一个个体系严密的形式化系统。本质上看,语言学中的形式语义学并不是理论语言学的一个流派;相反,它是理论语言学的一种实践,其宗旨是为理论语言学提供一种形式化方法,以适用于自然语言分析的可计算性要求。从这个意义上讲,形式语义学是理论语言学与计算语言学之间的一座桥梁。

二、发展形式语义学的迫切性

自然语言处理是信息化时代对理论语言学提出的新课题。自然语言是人类知识的主要载体,人类用语言来表达、贮存、传播、继承知识,并进行知识的创新。在今天这个生产知识化、经济全球化、文化多元化、信息网络化的时代,对自然语言信息的分析处理,已成为社会持续发展的必然要求。尤其在进入1990年代以后,互联网技术的飞速发展,全球语言信息量呈几何倍数增长,人们迫切需要用自动化手段来处理海量的语言信息。因此,语言信息处理超越计算语言学的范畴,成为计算科学、语言学、逻辑学和人工智能等多种学科共同关注的领域,是一个不足为奇的现象。

自然语言信息处理的目标是直接借助自然语言,在人与计算机之间实现信息交流。这项任务分为两个部分:一为应用部分,即建立各种处理自然语言的计算机应用软件系统和建设语言信息处理的基础资源,如电子词典和术语数据库等等。这一部分主要是计算语言学的任务,也是有计算科学背景的研究人员的长项。而另一部分则是理论研究,目的是发现语言的内在规律、探索语言理解和生成的计算方法。这一部分任务无疑应该

由语言学家担当,尤其应由有形式语义学背景的研究人员担当。

然而现状颇为令人担忧,语言学研究与自然语言处理存在脱节现象。清华大学孙茂松教授在2008年12月首都师范大学召开的语言学学科建设讨论会上曾指出:我国的语言学家和计算语言学家在上个世纪八九十年代曾有过一段甜蜜的合作阶段,但现在双方在一起合作的现象少了,并委婉地指出原因之一是双方的共同语言少了。

回顾过去30年,计算语言学主要停留在语形处理上,而理论语言学可以为句法结构分析提供成熟的理论支持,因而语言学和计算语言学合作很多。随着语形处理技术日趋成熟,其技术潜力也基本被挖掘殆尽,比如词性标注、汉语词的切分、机器翻译等,近20年以来,进展缓慢。互联网上的信息搜索,主要采用的是语词匹配技术;语音识别与合成软件,也多是基于对语音的物理属性分析。这些技术同它们刚出现时候相比,仅仅是在处理速度上有很大改进(这里面还有一部分是硬件技术改进的功劳),并无实质上的技术提升。这种现象很大程度上应归咎于自然语言形式语义分析技术的滞后。当然,考虑到语义的复杂性,这种滞后也是历史的必然。

据笔者对2000年到2007年的《中文信息处理》杂志发表的全部文章做的分析,除了少数几篇文章对20世纪80年代以后的形式语义学有所涉及外,如张威和周昌乐的《汉语隐喻理解的逻辑描述初探》和方立的《动态意义理论:逻辑语义学的继续发展》等。其他涉及语义处理的论文所依据的,基本都是80年代以前提出的(有些经过了改良)语言学理论,如格语法、语义网络、概念依存等。而涉及构造语义模型的论文,则是少之又少。2007年,中科院院士张钹在其发表在《中文信息处理》上的专稿《自然语言处理的计算模型》中指出:“在自然语言自动处理过程中,计算机处理的直接对象并不是实际的自然语言,而是它的计算模型,因此要真正理解自然语言自动处理的问题,并找出解决的办法,还需要从语言处理建模的角

度来探讨这些问题，可惜目前这方面的探讨还不多”。

从计算语言学的历史发展来看，解决语义问题主要靠构建语义本体（Ontology）的方式。这些工作的共同特点是，它们都是基于世界知识的语义本体，即用形式的方法表征世界知识，从而为自然语言配备语义。具体的做法包括构造语义网络或者建设带标词库，即词库中的每个词汇标上该词汇的语义特征。通过这些方法（包括二者混合方式）构造的语义本体在自然语言处理中确实起到一定作用，但显然尚有不足。

仍以“母亲”一词为例。自然语言语义比世界知识更加丰富、更加复杂。比如用女性、成年、已婚、有子女来定义“母亲”一词，则“我生来就没有母亲”（表示对不负责任的母亲的怨恨）和“姐姐就是我的母亲”（表示对姐姐养育之恩的感激之情）等完全可以接受的自然语言语句，在语义模型中就得不到满意的解释。也许有人会认为这个问题应该归因于语用或语境因素，但当代语言学的研究更多地是把句法、语义和语用结合起来考虑。

认知语言学认为：语言的意义不限于语言内部，而是根植于人与客观世界的互动认知，根植于使用者对世界的理解和信念。因此语义知识和语用知识是不可分的，语言形式是认知、语义和语用等因素共同促成的结果。语言学研究中，句法、语义和语用的交互界面研究是当前的一大热点。因此，把世界知识当作语义，并没有语言学的理论支持，只能是自然语言处理中的“权宜之计”，而非解决之道。

三、形式语义学的优势和潜力

形式语义学将数学和逻辑手段引入到对自然语言的语义刻画中，这对传统的语言学研究是一种挑战。在形式语义学看来，自然语言的句法部分在本质上与人工语言相同，是一个形式系统，只是结构更为复杂；自然语言的语义部分是这个形式系统的语义解释。形式语义学

的核心任务之一是构造人工逻辑语言，以便把有歧义、语义不自足的自然语言语句翻译成无歧义、语义自足的逻辑语言表达式；任务之二是构造语义模型，用于对逻辑语言进行模型论解释，从而间接地实现自然语言的计算机的“理解”。从数理逻辑的角度讲，一个语义模型就是一个语义解释系统，因此“模型概念是形式语义学的核心，其基础是所谓的代数结构”。^①

与目前计算语言学中的语义本体有所不同，形式语义学的语义模型是纯数学意义上的模型。它具有公理性，即构造模型的材料是无需定义的原于概念——个体，然后通过集合论和递归的方法构造出语义模型。而现在大多数语义本体都不具有公理性，比如用来定义“母亲”的“女性”、“成年”、“已婚”、“有子女”等义素，本身均非原子概念。正是由于形式语义学语义模型的公理性，在理论上，我们是可以构造出刻画所有语义内容的模型，当然在实践中，这项工作极其复杂的，还需要长时间的努力和探索。

形式语义学语义模型的另一个重要特点，就是从自然语言本身的语义出发，遵循自然语言语义的形成规律，按照自然语言语义的需要来构造语义模型。这点很重要，因为如前所述，自然语言的语义并不是将世界知识简单地映射到人脑中就形成了，而是人脑与客观世界互动的产物。客观世界将信息传达到人脑，人脑对所接受的信息进行加工，从而形成语义。而且这个语义的加工过程不是一劳永逸的，而是随着人对客观世界的不同认知而不断变化。形式语义学的语义模型可以是一个动态的构建过程，因而可以满足自然语言语义的这一特点。

40多年来，形式语义学发展很快。虽然以今天的眼光看，蒙太格在20世纪70年代所构造的语义模型和英语部分语句系统非常简单，只能算是一种“实验品”。它能解释的句子仅限于

^① 邹崇理：《自然语言逻辑研究》，北京大学出版社2000年版，第21页。

类似于“*John loves Mary* (约翰爱玛丽)”和“*John seeks a unicorn* (约翰寻找独角兽)”以及“*Every man loves a woman* (每一个男子都爱一个妇女)”这样简单的句子,但意义重大。蒙太格的语义模型实际上开创了用数学方法为自然语言构造语义模型的先例。而且这个模型所解释的自然语言语义还涉及到不同的方面。比如针对 *John loves Mary*, 只需要对外延语义加以解释就可以得到语句的语义;但对于 *John seeks a unicorn* 而言,则需要对 *unicorn* 的内涵语义进行解释,因为“约翰爱玛丽”蕴含着玛丽存在,或曰玛丽是实际存在的,而“约翰寻找独角兽”并不蕴含着有独角兽的存在,因为人寻找某样东西并不意味着一定存在这样东西。所以 *seek* 这样的内涵性动词和 *love* 这样的外延性动词需要做不同的模型论解释;而 *Every man loves a woman* 涉及到 *every* 和 *a* 的量化辖域的问题,如果 *a* 的辖域在 *every* 的辖域之内,则表示“每一个男子都爱一个女子,但不同男子爱的女子可能不一样”。反之,如果 *a* 的辖域在 *every* 的辖域之外,则表示“存在一个特定女子,她被所有男子爱”。这些问题在蒙太格的形式系统和语义模型中都得到了很好的翻译和解释。

蒙太格语法之后,各种理论层出不穷。如广义量词理论、代数语义结构、事件语义学、话语表现理论、动态语义学、情境语义学等。这些理论及其实践都大大提高了语义模型的解释力。比如广义量词理论发展了蒙太格语法对自然语言量化现象的模型解释;代数语义结构可以刻画自然语言中的复数词项和物质词项的所指;事件语义学将事件当作个体处理,使得形式语义学处理抽象意义词汇的能力大大提高;而话语表现理论和动态语义学等则使得形式语义学的分析跨越了句子界限,向话语篇章延展。而情境语义学是一种具有反传统性质的形式语义学理论,它将语义理解成一种信息结构。从认知角度考虑,人对外部世界的认知具有个体化的特点,即倾向于将抽象的、无形的关系或性质看成是一种实体存在。由此人对外部世界最基本的认知就是个体和关系,情境语义学把

关系看成是另一种个体,这种个体与关系组成的集合就是最基本的可处理数据,也就是一个信息条目,而情境就是信息条目的聚合。情境语义学为自然语言构造了由情境所构成的语义模型,语句的外延所指就是它所描述的情境,而语句的意义则是通过几种不同特征的情境类型之间的关系体现出来。

形式语义学除了提供日益丰富的语义模型之外,还提供将自然语言翻译成逻辑语言的句法工具,如荷兰斯蒂德曼(M. Steedman)和范本瑟姆(J. van Benthem)等人发展和完善的范畴类型语法,英国的肯普森(R. Kempson)提出的动态句法理论。这些理论的目标是克服自然语言句法语义不对称现象,将有歧义的自然语言语句翻译成无歧义的逻辑语言,以便在语义模型上进行解释。近年来,形式语义学的句法翻译工作在国内也有人做过尝试,并取得了一定的成绩。如赵章界和白硕发表在2005年《中文信息处理》上的《短语结构制导的范畴表达式演算》等。

四、形式语义学对理论语言学的影响

形式语义学作为语言学与数理逻辑学的交叉学科,本身并不是一种纯粹的语言学理论流派,而是应用各种数理逻辑方法来分析刻画自然语言。鉴于自然语言语义的复杂性,形式语义学注定需要语言学家和数理逻辑学家的通力合作。同时,随着形式语义学的成长,形式语义学本身也对语言学和逻辑学的发展产生了很大影响。

在西方发达国家,尤其是美国,很多具有形式语义学背景的学者成为语言学领域的知名学者。实际上,形式语义学的创始人蒙太格是数理逻辑学家,而他的主要合作者,被麻省理工学院的逻辑学家斯塔内克(R. Stalnaker)称为形式语义学之母的帕蒂(B. H. Partee),其博士论文导师就是语言学家乔姆斯基,而且她本人也曾担任过美国语言学会的会长,是形式语义研究中最具有代表性

的语言学家。其他在语言学界有影响的形式语义学家还有林克 (G. Link)、道蒂 (D. Dowty)、柯尔克亚 (G. Chierchia)、兰德曼 (F. Landman)、克里夫卡 (A. Krifka)、露斯坦 (S. Rothstein) 等人。

形式语义学家在语言研究中涉及的方面很宽泛，几乎涵盖了自然语言语义的所有方面。他们的研究采用了集合论和模型论的方法，这使他们得以发现传统语言学方法不容易发现的问题，讨论用非数理逻辑的方法无法深入讨论的问题。从目前出版的文献看，在形式语义学所讨论的语言学问题中，形式语义学家具有明显优势的领域包括：复数词项和物资词项的指称、时态时制的刻画、性质和事件的讨论、自然语言的量化研究等。比较明显的例证就是，国际上非综合性的语言学杂志中，上述领域内的研究论文基本都发表在以语义为主题的学术杂志上，而且采用的研究方法都是形式语义学的方法。这些杂志包括《语言学与哲学》(Linguistics and Philosophy)、《语义学学刊》(Journal of Semantics)、《自然语言语义》(Natural Language Semantics) 等。

同时，语言学研究也是形式语义学不可分离的部分。同样关注自然语言语义问题，偏向逻辑学的形式语义学家们往往喜欢将复杂的语

义问题简单化，从而可以比较“干净地”讨论逻辑系统问题，比如广义量词理论、情景语义理论、语义自动机等。他们的工作是为形式语义学提供更多更有效的逻辑工具。而偏向语言学的形式语义学家则不同，他们一般有着很好的语言学理论基础，因而在讨论自然语言的语义问题时，往往比前者更深入。他们不回避自然语言语义的复杂性，努力使逻辑系统更加贴近自然语言，致力于解决自然语言中的现实问题。

除此之外，也有一部分形式语义学家试图将形式语义学与形式句法学相结合，如麻省理工学院的海姆 (I. Heim) 和马萨诸塞州大学的克瑞策尔 (A. Kratzer) 等人，他们的研究也产生了一定的影响。

然而，在国内的学术界，相对于庞大的语言学群体而言，形式语义学的发展严重滞后，真正运用形式语义学进行语言研究的学者寥寥无几；反过来，这种情况也必然会影响到理论语言学的持续发展，是一个亟待解决的问题。

本文作者：合肥师范学院外语系讲师、中国社会科学院研究生院哲学系现代逻辑2007级博士研究生
责任编辑：马光

Formal Semantics in the Linguistics

Li Kesheng

Abstract: Formal semantics is interdisciplinary between linguistics and logic. It takes mathematic logic approaches to the language studies, aiming at formalizing the semantics of natural language, which is a necessity to the natural language understanding. During the past 30 years, with various new theories put forward, formal semantics has witnessed a remarkable improvement in the western counties, and a large quantity of interaction between theoretic linguistics and computational linguistics. However, as one of the practice field of theoretic linguistics, formal semantics has lagged behind the development of the other areas of linguistics.

Key words: linguistics; formal semantics; natural language processing; semantic model