



清华大学认知科学研究系列丛书

Series of Cognitive Science Research, Tsinghua University

自然语言 形式理论研究

FORMAL THEORIES OF NATURAL LANGUAGES

蔡曙山 邹崇理◇著

 人民教育出版社



清华大学认知科学研究系列丛书

Series of Cognitive Science Research, Tsinghua University

自然语言 形式理论研究

FORMAL THEORIES OF NATURAL LANGUAGES

蔡曙山 邹崇理◇著

人民出版社

责任编辑:陈亚明

图书在版编目(CIP)数据

自然语言形式理论研究/蔡曙山 邹崇理著.

-北京:人民出版社,2010.2

ISBN 978-7-01-007798-7

I. 自… II. ①蔡…②邹… III. 自然语言-理论研究 IV. TP18

中国版本图书馆 CIP 数据核字(2009)第 36451 号

自然语言形式理论研究

ZIRAN YUYAN XINGSHI LILUN YANJIU

蔡曙山 邹崇理 著

人民出版社 出版发行

(100706 北京朝阳门内大街 166 号)

北京龙之冉印务有限公司印刷 新华书店经销

2010 年 2 月第 1 版 2010 年 2 月北京第 1 次印刷

开本:710 毫米×1000 毫米 1/16 印张:39

字数:574 千字 印数:0,001-3,000 册

ISBN 978-7-01-007798-7 定价:76.00 元

邮购地址 100706 北京朝阳门内大街 166 号

人民东方图书销售中心 电话 (010)65250042 65289539

献给
我们共同的导师
周礼全先生

——作者

清华大学认知科学研究系列丛书及本项成果

受以下基金资助：

教育部“985工程”哲学社会科学创新基地项目
清华大学认知科学创新基地项目；

教育部认知科学重大理论与应用研究重大课题攻关
项目（批准号：07JZD0005）。

特此致谢！

绪 言

基于形式结构的自然语言理解

—

2000 年进入清华大学时，我即开始构思本书，原因很简单，因为我需要为研究生系统开设一门自然语言逻辑的课程。这门课程自 2002 年开始在清华大学讲授，每年一次，至今已经讲了 6 遍。

以什么样的理论框架和研究方法来写这本书是我从一开始就考虑的问题。6 年来，我们一直不断完善这个研究构架和研究方法。

语言，特别是自然语言，是我们最亲近的东西。但对自然语言，语言学家、逻辑学家、哲学家、人类学家、心理学家、计算机科学家和神经科学家有完全不同的理解，也有完全不同的研究方法。

20 世纪中期以来现代语言学、语言哲学和语言逻辑的发展，特别是近 30 年来认知科学的发展为我们研究自然语言提供了一种新的视界，这是一种综合的眼光；它还为我们提供了一种新的方法，这是一种多学科交叉的整合的研究框架。具体地说，我们选择句法学、语义学和语用学的三分框架，并使用现代逻辑的形式化分析方法来研究自然语言。

因教学和研究工作的需要，我选择我的同门师兄邹崇理研究员作为合作伙伴，共同写作此书。邹崇理的学术专长在语义学、特别是形式语义学研究领域，他在这个领域耕耘十多年，在蒙太格语法和形式语义学研究中收获了丰富的成果。我的专长在言语行为理论和语用学领域，在这个领域也工作了十多年。几年前在美国哈佛大学和麻省理工学院做访问研究时，

师从乔姆斯基做了形式句法学方面的研究。可以说，本书的两位作者是长期从事自然语言逻辑研究的专业人员。因此，我们选择形式化的语言逻辑研究方法，同时注意吸收认知科学众多分支学科的研究成果，这些学科包括语言哲学、心智哲学、现代逻辑、现代语言学特别是理论语言学(包括句法学、语义学和语用学)、心理语言学和语言心理学、认知心理学、认知人类学、人工智能特别是自然语言理解、认知神经科学等。

二

本书所依据的理论背景和研究框架是：

1. 自然语言的丰富性和生动性

应该从进化的角度来理解自然语言，理解人类语言与动物语言的本质差异，理解自然语言与人工语言、自然语言与形式语言的关系。

1) 自然语言是人类长期进化的产物

宇宙的历史有 150 亿年，地球的历史有 45 亿年至 47 亿年，生命的历史 21 亿年至 43 亿年，人类的历史 160 万年，声音语言的历史有一百多万年，而文字即书面语言的历史至今不过 3000 多年。以上事实说明，自然语言是宇宙、自然、人类长期进化的产物。

自然语言是在人类进化中产生的、约定俗成的语言。

人类语言是一种特殊的符号语言。这种语言区别于动物的身体语言和声音语言。符号语言使得人类和经验可以积累为知识和文化，并可以跨越时空传播并代代相传。

符号语言既是人类长期进化的产物，又反过来促进了人类的发展和进化。

2) 20 世纪西方哲学存在从自然语言到形式语言再回归自然语言的两 次转向

自然语言是在自然群体中流通的语言，如汉语、英语、法语、德语、俄语、西班牙语等。在语言的使用过程中，人类又创造出一种只属于某些群体所使用的非自然语言，即人工语言。人工语言不是在人的自然群体中

流通的语言,而只在非自然形成的特定人群中使用的语言。人工语言就是人们为实现某种目的,人为创造的符号语言,如数学语言、逻辑语言、物理学语言、化学语言、世界语等。

人工语言分为非形式的人工语言和形式的人工语言,前者如世界语,后者如现代逻辑使用的一阶语言和高阶语言等等。

形式语言是用于形式推理的人工语言。本书要特别研究自然语言与常用的逻辑语言如一阶语言和高阶语言的关系。

20世纪西方哲学存在从自然语言到形式语言再回归自然语言的两次转向。第一次转向是20世纪初发生的从自然语言到理想语言的转向,这次转向的结果是分析哲学的产生;第二次转向是20世纪40年代开始的从理想语言回归于自然语言的转向,这次转向的结果是日常语言哲学的产生,以及20世纪50年代开始在更加广阔背景下产生了语言哲学的各个分支学科,它使得对自然语言的语形学、语义学和语用学的研究渐次成为西方哲学的主流。

西方哲学的两次转向对语言学和逻辑学产生了重大的影响。当代西方语言学、逻辑学都更加重视自然语言的研究。本书指出,当前语言逻辑的语言基础是回归的自然语言。所谓回归的自然语言,是在“自然语言—形式语言—自然语言”的辩证运动中,在吸收自然语言研究和形式语言研究积极成果基础上无比丰富的自然语言。在此基础之上建立的自然语言逻辑,其语言基础是回归的自然语言,其方法是现代语言学和现代逻辑学所使用的语言分析和逻辑分析的方法。

3) 注重口头语言是现代语言学和语言逻辑的重要特征

书面语言(written language)是用文字来记录和表达的语言。口头语言(spoken language)也称为言语(speech)。

索绪尔区分了语言(langue)和言语(parole)。他认为,言语是一种个人的行为,是一种个人的活动,但个人的言语活动受到社会的制约。语言是言语活动中的社会部分,它不受个人意志的支配,是社会成员共有的,是一种社会心理现象。索绪尔的符号学以语音为中心,他侧重处理“语音模式”,把书面语看做是分离的、第二位的、依赖语音并能与之比较的记

号系统。但对索绪尔来说,书面语与口语的关系,就如同能指与所指的关系,前者依赖于后者。因此,在索绪尔的语言学中,口语和书面语是同样重要的。

现代语言学更加重视口头语言而不是书面语言。现代语言学家认为,声音语言或口头语言是第一位的,书面语言不过是对口头语言的记录。尽管有的民族没有书面语言,但任何民族都有口头语言。因此,现代语言学研究的着眼点是言语,是通过言语来研究语言,而不是相反。

乔姆斯基的句法结构理论非常重视语音。在他的句法结构模型和标准模型中,语音形式 PF 和逻辑形式 LF 是表层结构的两种表现形式。尽管后来乔姆斯基的理论有很多变化,但他始终关注语音表达式,这一点是不变的。在乔姆斯基看来,书面语言不过是口头语言或者说语音的记录而已。奥斯汀的言语行为理论也非常重视口头语言,所谓言语就是口语。言语行为理论的核心思想是“通过说话来做事”(Doing something in saying something),口语作为言语行为的载体,是言语行为理论研究的对象。所谓言语行为理论,就是研究用来做事的这种口头语言即言语的形式、内容和规律的。

4) 自然语言与人类进化

在宇宙、地球、人类漫长的历史中,以表意符号为特征的自然语言的历史不过是短短的一瞬,然而,它却改变了人类进化的方向,促进人类社会日新月异的发展。

使用符号语言是改变人类进化方向的三件大事之一,其他两件事是直立行走和火的使用。

根据迪肯(Deacon,T.W.) 1997 年的研究,人类的语言和大脑是长期进化的产物;语言反映了一种新的思维模式,即符号思维的模式;符号思维引发了自古人类诞生以来在过去二百多万年间语言与大脑的双重交叉的进化——“区分人类身体和大脑的许多生理特征最终被那种代代相传的思想所继承”。

语言与进化是认知人类学从而也是整个认知科学的重要研究领域,它的一些重要结果使我们对语言学、逻辑学、哲学的认识产生了根本的改变。

2. 句法学、语义学和语用学的三分研究框架

1) 符号学的三分框架

句法学(syntax)、语义学(semantics)和语用学(pragmatics)原本是符号学(semiotics)的分类法,它们将符号学的研究对象分为三个互相包含、逐步深入的研究领域。符号学的这个三分法对20世纪西方语言哲学和语言逻辑的发展具有根本性的影响。

瑞士语言学家索绪尔(Ferdinand de Saussure, 1857—1913)和美国哲学家皮尔斯(Charles Sanders Peirce, 1839—1914)几乎同时在大西洋两岸独立地创建了符号学。此后,符号学一直有两个名称,一个是Semiology,它用指索绪尔传统的符号学,即欧洲的符号学;另一个是Semiotics,它用指皮尔斯的传统的符号学,即美国的符号学。

2) 符号学三分研究框架的影响

欧洲符号学与美国的符号学有完全不同的传统和研究方式。欧洲符号学试图将符号学作为整个人文艺术学科的共同平台,而美国的符号学更接近逻辑学。

符号学的三分法对其后的学术研究影响深远。这是因为,人文艺术学科的研究对象,一般都可以看做是某种特殊的记号,所以,这些学科的研究也就可以纳入符号学的框架。但是,在符号学的构架下,以句法学、语义学和语用学的方法来研究某一门学科,又不完全等同于这门具体学科的研究,它带有“元理论”(meta theory)的特点,即把某一理论作为自己的研究对象。

首先受到这种影响的是语言学。在现代语言学中,按照句法学、语义学以及语用学的框架和方法来研究语言理论的学科被称为理论语言学(theoretical linguistics)。另外一门专门研究语言理论的学科叫作逻辑语言学(logical linguistics),它又分为结构理论(形态学)、意义理论和有效性理论三个部门。^①语言符号学(linguistic semiotics)则是将语言学作为符号学的一个分支来进行研究,它是具有元理论性质的语言学。

^① Rescher, N. (1968) *Topics in Philosophical Logic*. D. Reidel Publishing Company, p.7.

逻辑学的发展也受到这种三分法的影响。雷歇尔(N. Rescher)在《哲学逻辑》(1968)一书中,将逻辑句法学(logical syntax)、逻辑语义学(logical semantics)、逻辑语用学(logical pragmatics)和逻辑语言学(logical linguistics)合称为元逻辑(metalogic)。他在著名的逻辑分类图(A Map of Logic)中,将基本逻辑、元逻辑、数学逻辑、科学逻辑和哲学逻辑并列,作为现代逻辑的基本理论。此外,语言逻辑亦称逻辑符号学,也是按照语形学、语义学和语用学的构架来研究自然语言的逻辑理论的。

作为20世纪西方哲学主流的语言哲学,更是受到符号学三分法的明显影响,它是完全按照语形学、语义学和语用学的构架来研究言语哲学的。施太格缪勒(Wolfgang Stegmüller)的《当代哲学主流》(1978/1986)、马蒂尼奇(A. P. Martinich)的《语言哲学》(1985)都是按照句法学、语义学和语用学的框架来阐述语言哲学的。

3) 语言逻辑的三分研究框架

语言逻辑使用现代逻辑的方法来研究自然语言的逻辑问题,包括自然语言的结构、意义和使用的问题,也就是自然语言的语形、语义和语用问题。因此,语言逻辑也称作自然语言逻辑。

关于这个学科的名称,英文文献中有三种主要的提法:logic of language; logic of natural language; logic and language。中文“语言逻辑”这个名称,可以对应于这三种提法。

语言逻辑已形成丰富的文献资料。如詹姆士·麦考利(James D. McCawley)的《语言逻辑:语言学家总想知道而又羞于启齿的东西》;克赖安(D. Cryan)、沙蒂尔(S. Shatil)和梅布林(B. Mayblin)2004合编的《逻辑导论》;甘乌特(L. T. F. Gamut)的上下两卷本《逻辑导论》、范·本瑟姆(J. van Benthem)和缪伦(A. ter Meulen)合编的《逻辑和语言手册》;胡安·尤里安格瑞卡(Juan Uriagereka)的《韵律和推理——简约论的语形学导论》等。这些经典的文献都采用句法学、语义学和语用学的三分构架来研究语言逻辑。

值得注意的是,20世纪50年代以后,句法学、语义学和语用学的研

究中相继出现了形式化的倾向,形式句法学、形式语义学和形式语用学的著作相继问世。乔姆斯基关于句法结构理论的一系列重要著作都是形式化的。此外还有卡得科弗(Peter W. Culicover)、沃索(Thomas Wasow)和阿克玛吉安(A. Akmajian)等合编的《形式句法学》,这是一本论文集,收录了乔姆斯基、贝克(C. L. Baker)、巴赫(Emmon Back)、布雷斯南(Joan Bresnan)、帕蒂(Barbara Hall Partee)、莱特富特(David Lightfoot)、埃蒙兹(Joseph Emonds)、杰肯朵夫(Ray Jckendoff)、塞尔柯克(Elisabeth Selkirk)、奥厄尔(Richard T. Oehrle)、沃索、安德森(Stephen R. Anderson)、黑尔(Kenneth Hale)、斯蒂尔(Susan Steele)、阿克玛吉安和吉(James Paul Gee)等十多位语言学家句法学研究的论文。按照他们的理解,这些人的工作都属于形式语法学的范畴。形式语义学的研究由蒙太格开其端,代表作是托马森(Richard H. Thomason) 1974 所编的《形式哲学:理查德·蒙太格论文选集》。此后,形式语义学的著作和资料可以说是汗牛充栋。如道蒂(David R. Dowty) 1979 年的《语词意义和蒙太格语法》;巴赫(Emmon Back) 1989 年的《形式语义学的非形式教程》;卡恩(Ronnie Cann) 1993 年的《形式语义学》;波特纳(Paul Portner)和帕蒂 2002 年所编的《形式语言学:基本读物》;波特纳 2005 年的《何谓意义:形式语义学基础》等等。我国学者、本书作者之一邹崇理长期从事蒙太格语法和形式语义学研究,著有《逻辑、语言和蒙太格语法》(1995)、《自然语言逻辑研究》(2000)、《逻辑、语言和信息》(2002)等。此外还有我国学者蒋严、潘海华《形式语义学引论》(1998)等。

在句法学、语义学和语用学三个领域中,语用学涉及语言的使用者即人的因素,是最难形式化的,其开展也最晚。塞尔和范德维克(John R. Searle and Daniel Vanderveken) 1985 年的著作《语用逻辑基础》(*Foundations of Illocutionary Logic*)是第一本以形式化的方法来分析语用行为和语用力量的专著,作者还建立了语用逻辑(illocutionary logic)的分析系统。笔者在《言语行为和语用逻辑》(1998)一书中,对塞尔和范德维克的这项工作做了深入的分析,建立了语用逻辑的形式系统,包括命题

的语用逻辑、量化的语用逻辑和模态的语用逻辑，并将它应用于计算机言语行为和人工智能的分析。阿萨·卡谢(Asa Kasher) 1998年编的6卷本大部头《语用学》(*Pragmatics: Critical Concepts*)介绍了语用学的核心思想，该书的一些部分如间接言语行为、语用学和心理学也使用了部分形式化的分析方法。最新也是最有影响的一部形式语用学的专著是妮瑞特·卡德蒙(Nirit Kadmon) 2001年的《形式语用学：语义学、语用学、预设和注意》(*Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*)是第一本以“形式语用学”为名的书，本书在一个非常广阔背景下，对语用学的一些领域如动态语义学、预设、注意等问题进行了形式化的分析处理。特别值得提出的是，该书用了近200页，即将近一半的篇幅，详细分析了属于认知语言学领域的注意问题。例如，注意的同一性、注意—诱导解释、“论域选择”或“自由参数”分析、要做出选择的注意诱导理论、注意和范围、带有“对比主题”的复合注意结构、注意的预设及其与其他预设的相互作用等等。对这些问题，作者都做出非常专业的形式化分析。

从这些发展我们可以看到，符号学的三分法对语言逻辑的形成和发展具有至关重要的影响，自然语言的形式化研究是循着句法学—语义学—语用学的路径展开的，语言逻辑的学科构架就是句法逻辑、语义逻辑和语用逻辑。本书就是按照句法学、语义学和语用学三分法来构造语言逻辑的理论体系的。

3. 为什么要对自然语言的形式化进行研究

为什么要对自然语言的形式化进行研究？这是一个非常严肃而深刻的问题。笔者认为，对这个问题可以从以下三个方面来回答。

1) 从形式化方法的特征和作用看

何谓形式化？形式化是以形式语言为基础，以形式公理系统为工具，从希尔伯特方案的提出到哥德尔定理的证明所发展起来的一种重要的数学证明和推理方法。简言之，形式化是这样一种特殊的方法，它使用形式语言和形式推理方法。形式化方法具有两大特点：其一，它使用的语言是形式语言，这种语言与自然语言不同，它没有预先赋予的意义，它是表意的，但它的意义是经过解释后获得的。其二，它使用的方法是形式推理方

法,只要推理的前提真,规则保证推理的结论也真。

形式化方法被广泛应用到数学、逻辑学、语言学、物理学、计算机科学等众多学科的研究之中。如果我们想要了解一个理论的性质,了解它与其他理论的关系,除了形式化,我们没有别的路可走。因为当我们想要把某一个理论作为研究对象时,唯一可行的方法是用另外一套语言系统来指代它。正如我们想要使用计算机来处理文字、图像处理和网络通信时,我们只有使用一套没有预先的意义、经解释后又能代表文字、图像和其他信息的表意符号即 0 和 1 的系统来实现这样的目的。因此,当我们想要把自然语言作为研究对象时,自然语言本身并不能完成这个目的,而符号学的方法特别是形式化的符号学方法却可能方便地承担起这个任务。

2) 从自然语言理解看

在人工智能的自然语言理解中,采用模块化的方法来进行自然语言处理,其由低到高的处理模块是:音位学模块→形态学模块→词汇学模块→句法学模块→语义学模块→语用学模块。

在每一个模块中,都需要建立若干计算机模型,而这就需要建立自然语言的形式化理论。例如,计算机句法加工采用乔姆斯基的句法结构理论来建模。自然语言理解中句法模型做得比较成功,原因是句法结构具有比较成熟的理论,而这种理论成熟的标志就是其形式化程度较高。目前自然语言理解的语义模型取得了很多积极的成果,这得益于近几十年来形式语义学的发展。自然语言理解中语用模型的建立是开展得最晚,也是难度最大的一项工作,因为语用学涉及的因素更多,更难做形式化的处理。因此,本书在语义学和语用学的形式化研究方面所做的尝试对推进自然语言理解是有益的。^①

3) 从认知科学的观点看

近 30 年来认知科学的发展从根本上改变了人们对脑、精神、语言、文化以及人的认知能力的看法。例如,认知神经科学的研究表明,人脑的自然语言加工具有形式结构,也分为句法加工、语义加工和语用加工,它

^① 本人所指导的博士研究生王志梅在这方面做了有益的工作。他的博士学位论文“语用要点的模型研究”被评为 2008 年清华大学优秀博士学位论文。

们在脑的不同区域进行，具有不同的特征。

人的大脑如何对语言进行加工？在做 ERP 或 fMRI 的研究时就需要建模，因此要分析自然语言的形式结构。

三

1. 本书的主要内容和特点

本书的主要内容和特点表现在以下几个方面：

1) 阐明了自然语言与形式语言的联系和区别，提出在 20 世纪的语言哲学和语言逻辑的发展过程中，有一个从自然语言到形式语言，再回归于自然语言的辩证发展过程。

2) 分析语言逻辑的发展背景，明确语言逻辑的研究对象和方法。以“回归的自然语言”为基本的概念和范畴，详细分析现代语言学和现代逻辑学的发展和成就，提出语言逻辑是现代语言学与现代逻辑学的交叉学科。明确了语言逻辑的研究对象和方法。

3) 使用句法学、语义学和语用学三分法框架研究语言逻辑。国外同类研究多用单一方式研究语言逻辑三个部分的某一个方面，如乔姆斯基从 20 世纪 50 年代开始，半个多世纪以来仅仅研究句法学，尽管他前后期的理论有很大的发展和变化，但仍然属于句法学的范畴，他拒绝接受语义学，更不用说语用学。施太格缪勒在《当代哲学主流》中将三分法用于语言哲学的框架，但该书是评介性质的哲学史著作，不是系统的理论研究。本书将句法学、语义学和语用学作为语言逻辑的研究框架，这种研究路线和方法是具有独特性的。

4) 句法学、语义学和语用学三个部分分别以乔姆斯基、蒙太格、奥斯汀和塞尔作为代表性人物和基本理论的主体，因为他们是当代西方语言学、语言哲学和语言逻辑领域中最具权威性的人物。在此基础上，吸收国外学者在形式句法学、形式语义学和形式语用学方面的最新研究成果。例如，形式语用学的研究在欧美 20 世纪末才发展起来的。本书的重要创新是采用形式化方法来做句法学、语义学和语用学研究，形成了形式句法

学、形式语义学和形式语用学的研究领域和研究方法，深化了语言逻辑研究。因此，这种研究方法在国内语言学、逻辑学和哲学研究领域具有借鉴的意义。

5) 在整体结构上，将句法学、语义学和语用学结合起来研究，并弄清楚它们之间的关系，这种研究方法不仅在国内，就是在国际上也有独特之处。在形式句法学上，将乔姆斯基的句法结构与逻辑句法分析加以对比，并特别研究乔姆斯基句法理论是如何导向心智与认知的；在形式语义学上，分析蒙太格形式语义学的发展，并将形式语义学方法应用于汉语的语义分析；在形式语用学上，分析从奥斯汀到塞尔的理论发展，包括奥斯汀的言语行为理论、塞尔对言语行为理论的发展、塞尔的间接言语行为理论，着重分析塞尔的语用逻辑系统，介绍作者在语用逻辑形式化研究方面的独立创新工作，分析形式语用学的最新理论预设、焦点和心理。

总之，本书确定了语言逻辑的三分法研究框架，明确了语言逻辑的研究对象和方法，系统阐述了语言逻辑三个部分句法学、语义学和语用学的代表性人物及其理论，采用形式句法学、形式语义学和形式语用学的方法深化了语言逻辑的研究。

2. 本书的适用范围

本书关于形式句法学、形式语义学、形式语用学的基础理论研究，在语言学、逻辑学、心理学、哲学、人工智能和神经科学的研究方面有一定的应用价值。本书可供语言学、逻辑学、哲学、心理学、计算机科学、神经科学和认知科学及其他相关领域的教师和研究生用作教材、教学参考书或研究工作的参考读物。

本书分工如下：

蔡曙山：绪言、第一章、第二章、第三章、第四章、第七章、第八章，以及参考文献和后记。

邹崇理：第五章、第六章。

张玲(博士生)负责编辑本书术语索引和人名索引。

目 录

绪 言 基于形式结构的自然语言理解.....	1
第一章 自然语言和形式语言	1
第一节 自然语言	1
第二节 形式语言	17
第三节 书面语言和口头语言	25
第四节 对象语言和元语言	30
第二章 自然语言逻辑	40
第一节 回归自然语言	40
第二节 现代语言学	49
第三节 现代逻辑学	71
第四节 语言逻辑的对象和方法	95
第五节 语言逻辑的意义	129
第三章 乔姆斯基的生成语法	141
第一节 乔姆斯基的革命	141
第二节 短语结构理论	165
第三节 词库理论	187
第四节 转换理论	195
第四章 形式句法理论的新进展	218
第一节 原则和参数理论	218
第二节 最简方案	261
第三节 乔姆斯基语言理论中的逻辑问题	276
第四节 语言、心智、认知	287