



A Cognitive Study and Computational
Analysis of Language (Revised)

语言的认知研究和计算分析

(增订本)

袁毓林 著

商务印书馆
The Commercial Press

语言的认知研究和计算分析

(增订本)

袁毓林



商务印书馆
The Commercial Press

2014年·北京

图书在版编目(CIP)数据

语言的认知研究和计算分析/袁毓林著. —增订本—北京:商务印书馆, 2014

ISBN 978-7-100-10741-9

I. ①语… II. ①袁… III. ①认知科学—语言学—研究 IV. ①H0

中国版本图书馆 CIP 数据核字(2014)第 217819 号

所有权利保留。

未经许可,不得以任何方式使用。

语言的认知研究和计算分析(增订本)

袁毓林 著

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

北京市艺辉印刷有限公司印刷

ISBN 978-7-100-10741-9

2014年10月第1版

开本 850×1168 1/32

2014年10月北京第1次印刷 印张 18 7/8

定价: 52.00元

目 录

一、认知科学和语言研究

认知科学背景上的语言研究	3
语言的认知研究和计算分析	35
关于认知语言学的理论思考	69

二、名词配价和语义计算

现代汉语二价名词研究	107
一价名词的认知研究	147
简论计算机理解自然语言	182

三、谓词隐含和构式转换

句法空位和成分提取	221
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 ...	241
“者”的语法功能及其历史演变	277

四、原型理论和语法范畴

词类范畴的家族相似性	321
基于原型的汉语词类分析	356
怎样判定语法结构的类型?	393

五、语言结构和心理机制

语言学范畴的心理现实性	435
话题化及相关的语法过程	462
语言信息的编码和生物信息的编码之比较	504

附录:学术随笔

学术评价与引文指数	539
现代汉语语法研究的回顾(2005 - 2009 年)与展望 (2010 - 2014 年)	550
中国语言学在信息科学上的话语权 ——建设语言资源,争取有话可说	559
徜徉在应用需求和学理探究之间 ——数字化时代语言研究的困惑和蛻迁	575
后记	589

一、 认知科学和语言研究

认知科学背景上的语言研究*

本文介绍和评论广泛流行于哲学、心理学、语言学、计算机科学和神经生理学中的认知主义思潮,分析在此基础上形成的认知科学的学科背景和学术源流。其中涉及:(1)哲学上的心脑二元论和物理主义思潮,神经生理学对大脑中神经元之间的通信原理的研究;(2)智能起源上的天赋论和建构论,以及各自对语言能力的解释力量;(3)基于信息加工观点的认知心理学,以及跟皮亚杰发生心理学的比较;(4)作为人工智能研究和认知心理学的理论基础物理符号系统假设,以及与之抗衡的神经网络思想;(5)哲学家对认知主义的批判,以及我们的反批判;(6)语言研究上的两种认知主义的操作路线,以及另外一种可能的研究途径。

0. 引言: 认知主义和结构主义

自乔姆斯基革命以来,当代语言学呈现出范式纷纭、学派林立的景象。如果不计较技术和细节上的巨大差异,勉强概括各种

* 原载《国外语言学》1996年第2期,此次做了增订。

研究路子背后的共同特征;那么,当代一些有影响的语言学流派大都可以归在认知主义这个名目之下。而此前的强调按照语言的内部结构去理解和研究语言的思潮便是结构主义。从学术背景上看,结构主义语言学的哲学基础是广泛流行于当时知识界的逻辑实证主义,心理学基础是摒弃心灵的行为主义心理学。因此,结构主义语言学只注意对现实的话语素材进行切分和描写,拒绝讨论难于摆脱心灵主义纠缠的意义问题。认知主义语言学是对结构主义语言学及其学术背景的反动,其哲学基础是当代的分析哲学(包括语言哲学),心理学基础是注重内省的各种心灵主义心理学。事实上,认知主义语言学的不同流派在研究取向上是十分不同的。比如,乔姆斯基的生成语法是一种注重形式的认知主义,它深受数学和逻辑原理的形式化研究的影响;R. Langacker的认知语法是一种注重概念的认知主义,它强调作为语言核心的意义是一种心理现象(因此必须按心理现象来描写);戴浩一的认知功能语法是一种注重语言的信息传递功能的认知主义,它强调语法结构来自对现实的象征(所以语言学必须研究语法结构在观念方面的理据);我们则倡导一种注重可计算性的认知主义,摸索一种基于认知并面向计算的语言研究的途径(toward a cognition-based & computation-oriented approach to linguistic study)。

认知主义是当代广泛流行于哲学、心理学、语言学、神经科学和计算机科学等研究领域的一种思潮,要想对这样一种源远流长又旁涉广泛的学术潮流做出全面而适切的介绍是不可能的。我

们打算紧扣认知、智能、心智、知识、计算和表达等认知主义的核心概念,来讨论它们的学科背景和学术源流,希望能以点带面地展示认知主义的研究纲领和具体成果。

1. 心脑二元论和物理主义思潮

认知科学(cognitive science)是一门以智能系统(包括自然的和人工的)的工作原理为研究对象的新兴学科,它是从哲学、心理学、语言学、计算机科学和神经生理学的交叉领域中发展起来的前沿性学科。^① 在认知科学中,智能(intelligence)是一个核心概念。但是,哲学家对于智能的看法有很大分歧。比如,笛卡尔(René Descartes, 1596 - 1650 年,法国哲学家、生理学家,解析几何的创始人)提出了著名的二元论,认为心智(mind)跟身体是两种完全不同并可以分离的实体;心智的本质是思想,能够脱离身体而存在。^② 与此相对的是物理主义思潮,^③物理主义(physicalism)强调一切心智事件(精神现象和智力活动)最终都可以还

① 1977 年期刊 *Cognitive Science* (《认知科学》) 出版, 认知科学学会成立。1979 年, 认知科学学会召开第一次正式年会。这一系列事件标志着认知科学在 20 世纪 70 年代正式诞生并登上世界科学的殿堂。参考章士嵘(1992), 第 1、38 页。

② 荷兰哲学家斯宾诺莎(Baruch Spinoza, 1632 - 1677 年) 曾用所谓几何学的方式对此做了一些证明, 详见斯宾诺莎(1991), 第 51 - 61、69 - 71、180 - 187 页。英国哲学家赖尔(G. Ryle, 1900 - 1976 年) 对这种心脑二元论的“笛卡尔神话”进行了饶有趣味又不乏同情的批判, 详见赖尔(1988), 第 5 - 19 页。

③ 详见艾耶尔(1987), 第 193 - 216 页。

原成大脑的生理过程和物理过程。这种观点把智能看作是人脑这种特殊物质的一种功能,相信心理过程以脑过程为基础并可以还原为脑过程。随之而来的问题是:如何用脑过程这种低层次的现象去说明心理过程这种高层次的现象?

为了回答这个问题,神经生理学从物理、化学角度分析大脑神经元(neuron,即神经细胞)传递信息的电学和化学过程,证明神经细胞的电学性质与其化学性质是互相联系、互相影响的。神经生物学家设想精神活动可能跟脑子里的神经脉冲(impulse)的图型相关。为了充分评价这一假设的含义,他们研究了神经元是怎样工作的、它们相互之间是怎样通信的、它们是怎样组成局部的或分布的网络的、它们之间的连接是怎样随经验而改变的。

当代神经科学证明,大脑的结构是非常复杂的。表现为:(1)大脑中至少有 10^{10} 个神经细胞,大体相当于银河系中星球的数量级;(2)神经细胞的多样性,有的细胞具有能跟相邻细胞通信的短轴突,有的细胞具有伸向其他区域的长轴突。当然,神经细胞也有其简单性的一面。功能相似的神经元都集成柱状或板状而穿过皮层,并且所有的神经元都是以非常相似的方式传递信息的——信息沿着轴突(axon)以动作电位(action potential)的短脉冲的形式传播。具体地说,一个受刺激的神经元产生一种叫动作电位的脉冲而把信号输送给另一个神经元。这些信号像波浪一样沿着神经细胞的单个轴突(即神经纤维)传播,并在突触(dendrite,即神经细胞之间的接触点)处转化成化学信号。神经

元静止时,它的外膜保持大约-70毫伏的电位差(内表面相对于外表面是负的)。静止时,钾离子(K)比钠离子(Na)更易于透过膜。细胞受刺激后,膜对钠的渗透性加强,从而导致正电荷的流入而启动一个脉冲,于是膜电位瞬即逆转。脉冲起始于细胞体和轴突的接合部并引离细胞体。脉冲到达突触前神经元的轴突末端后就诱导释放递质分子。递质通过狭窄的突触间隙而和突触前膜上的受体相结合。这样的结合作用会打开离子通道,而其本身又往往会导致在突触后神经元里产生动作电位。^①这就是现代神经生理学对神经元之间的通信原理解释。

总之,神经科学家相信:大脑结构的复杂性足以解释人的记忆、情绪和想象力等心智奥秘,精神活动可以还原为大脑神经网络上的电信号。^②

2. 人类智能的天赋论和建构论

在智能的来源问题上,当代心智哲学和心理学中有天赋论(innatism)和建构论(constructivism)两大理论派别。天赋论可以乔姆斯基为代表,他深受笛卡尔哲学的影响,信奉源于柏拉图(Plato)的天赋论。乔姆斯基反对经验论把人类的心智看作是一

① 见《科学》,第5页,更详细的讨论见邓斯和平森(1983),第99-103页。

② 参考《科学》,第1-11页;章士嵘(1992),第3-5页。

块“白板”^①、一切知识都是后天经验习得的想法，他反对行为主义心理学“刺激-反应”的模式。他认为人的智能结构和认知能力是人类这个物种的大脑生物学结构所固有的，这种潜在的结构和能力一旦受到外部诱因的驱动就能被激活，产生出观念和知识。乔姆斯基强调：(1)人具有生成和理解无限多新句子的语言能力，它是人类心智的组成部分，它是人的一种固有的机制，正是这种机制把人类的经验映射为语法；(2)语言中存在成体系的普遍现象(universals)，但缺少证据说明它们是经过学习或经验而习得的，可能的解释是作为生物本能的普遍语法(universal grammar)使然。语言作为生物本能地具有部分确定的结构(表现为普遍语法)，正如生理器官的普遍性对人类来说是确定的一样。乔姆斯基据此证明：人类的所有知识都可以从这个物种天赋的心智特征中推导出来。^②

建构论可以瑞士心理学家和哲学家皮亚杰(Jean Piaget, 1896-1980年)为代表。皮亚杰深受从洛克(John Locke)、休谟(David Hume, 1711-1776年)到斯宾塞(Herbert Spencer,

① “白板说”是一种主张认识来源于经验的哲学思想。白板，拉丁文是 *tabula rasa*，原指一种洁白无瑕的状态。西方哲学家用它比喻人类心灵的本来状态像白纸一样没有任何印迹。古希腊哲学家亚里士多德在其《论灵魂》中认为，灵魂如同蜡块一样，从外物接受印纹。17世纪英国经验论哲学家洛克(J. John Locke, 1632-1704年)继承和发展了这个思想，批判了天赋观念说，认为人出生时心灵犹如白纸或白板，对任何事物均无印象；人的一切观念和知识都是外界事物在白板上留下的痕迹，最终都导源于经验。参考谷歌互动百科的“白板说”条目。

② 详见 Chomsky(1975)和(1977)。

1820-1903年)的古典经验论的影响,在几十年的生物学、哲学、逻辑学和心理学(特别是儿童实验心理学)研究的基础上,他提出:认知的结构既不是在客体中预先形成了的,因为这些客体总是被同化(assimilate)到那些超越于客体之上的逻辑数学框架中去;也不是在必须不断地进行重新组织的主体中预先形成的。认识起因于主客体之间的相互作用,有机体自我调节(self-regulation)的适应机制使客体被同化到主体的图式(scheme)中。图式来源于动作,作用于新客体的活动反复多次并通过一般化而形成图式。^①

皮亚杰认为,跟所有的生物的发展一样,在智力的成长过程中起作用的主要有两条原则:适应和组织。适应(adaptation)指我们的心理越来越有效地对环境的需要做出反应,它有同化和顺应两个双向建构的过程,其中起协调作用的机制是调节(accommodation)。同化指认知主体将客体纳入主体的图式之中,顺应指认知主体调整原有的图式并创立新的图式以适应新的客体。调节则指认知主体如何控制同化和顺应的双重建构以达到平衡,即图式之间的分化和协调问题。组织指我们的心理以越来越整合的方式来结构或组织,其中最简单的水平是图式,它是能作用于物质的某种行动(物质的或心理的)的心理表征(psychological representation)。对于新生儿来说,吸吮、抓握和看都是图式,它

^① 详见皮亚杰(1995),第15-16、21-22、58、67-70页,章士嵘(1992),第74-76、157-161页。

们是新生儿得以认识世界的方式——通过作用于世界的动作来认识世界。这种在遗传的反射图式的基础上通过同化和顺应而建立起来的图式叫动作图式,这是在感觉运动的基础上形成的认知的第一图式。动作图式一经形成便会在同化、顺应和调节的不断循环中走向完善。接着,儿童在用某种方式和手段取得某种结果的过程中,逐步建构起有目的控制的图式——目的图式,这便是智能的萌芽。在此基础上,儿童便会形成一系列能应付新情况的经验图式,并进一步发展成为抽象思维——形式运算图式。这些图式有规律地变得更加整合和协调,最终产生出成人的心理。^① 因此,智能的发展就是人的认知结构连续地建构和再建构的过程,是一个从建构初级图式到高级图式的实践的过程。

综上所述,乔姆斯基是预先决定论者(predeterminist),强调作为人类最基本的智能要素的语言能力的认知结构是天赋的,是在人的器官中甚至基因中就早已预先编制好了的(preprogramming),后天的发育和环境因素只不过是促使这种结构成熟而已。相反,皮亚杰是相互作用论者(interactionist),强调认知结构是后天建构的,智能、知识只能来自后天的活动、操作和实践。但是,乔姆斯基的天赋论在生物学和心理学上都是很难完全确证或否证的。因为人种突变从生物学上看是一种随机的、不好解释的现象。为什么这种随机的突变会给人提供了一种表达力很强的语言能力呢?怎么能把一种合理的语言结构归因于这种随机

^① 参考 Solso (1979), 中译本第 370-381 页, 章士嵘(1992), 第 159 页。

的突变呢?^① 不过,皮亚杰源于感觉运动的智能建构也说明不了普遍语法不因个体和语种而异这一事实。^②

3. 信息加工观点和认知心理学

在智能的研究上,认知心理学采取计算机模拟以及其他实验方法来探索人类心智的一般的工作原理。认知心理学认为人的动机在表现为意识之前要经历一系列的转变、改造和处理,就像通信过程需要将信号进行编码和译码一样。因此,可以把人看作是一个信息加工系统,人与环境和其他认知主体之间有着信息交流的关系。信息在进入认知主体之后又会经历被转换、简约、合成、存储、重建、再现和使用等加工过程。通过考察信息加工的链条,人们就可以对系统的行为做出预测,并分析这种行为与环境的关系。这样,认知心理学为研究人的行为提供了一个全面而一贯的工作模型。由于认知心理学把人类的心智过程看作是一个信息加工(information processing)的过程,因而可以利用信息加工的概念和方法来研究人类内在的认知过程。比如,认知心理学运用信息和编码等概念来研究记忆,用存储、检索、控制、缓冲器

① 其实,这个问题可难不倒乔老;记得他曾经雄辩地说过,像动物的眼睛这种极其精细的器官,也是一种随机的突变的产物,很难从生物学上给出合理的解释。(出处失记——笔者)

② 关于乔姆斯基和皮亚杰 1975 年在巴黎会议上的当面辩论,请看章士嵘(1992),第 80—83 页。

和网络图等概念对记忆的结构和动态过程进行比喻性的描写,使记忆等一些过去比较含糊的概念转化为可操作的术语,从而为建立形式化的认知模型奠定了基础。^①

认知心理学的理论目标是要解释人在完成认知活动时是怎样进行信息加工的,研究内容包括:

- (1) 人是怎样从世界中获取信息的;
- (2) 这些信息是怎样表征的,即以什么样的表达形式存储在人的大脑中;
- (3) 这些信息是怎样转化为知识的;
- (4) 这些知识在大脑中是怎样表征的;
- (5) 这些知识在人解决问题时是怎样被调用的;
- (6) 大脑中的各种信息和知识是怎样指导人的注意和行为的。^②

其中,(1)讨论感觉信号的检测和加工、模式识别、知觉广度和选择性注意等问题;(2)-(4)讨论记忆的结构和过程、记忆模型、短时记忆及其容量、长时记忆的组织 and 信息回收、记忆组织及其记忆模型等问题;(5)(6)讨论在概念形成、问题求解、语言理解等高级认知活动中,人怎样调用知识和信息加工的策略等问题。可见,认知心理学一方面受惠于计算机科学,另一方面又为计算机模拟人类智能提供了可操作的认知模型。

① 参考章士嵘(1992),第25-28页;塞尔(1991),第33-46页。

② 详见 Solso (1979),中译本第1-22页。