

语言习得的联结主义模式

李 平 Richmond 大学

提要 语言学是认知科学的一个重要分支。本文探讨近年来对认知科学产生了重大影响的联结主义理论及方法,介绍联结主义的基本概念,在语言学及语言习得中的应用,以及它给语言研究提供的新思路。

关键词 语言习得 计算机模拟 联结主义

1. 引言: 认知科学与语言学

认知科学的发展日新月异。从上世纪五十年代到今天虽然只有短短的几十年,科学家们对人脑的构造及功能已经有了比较深入的认识。语言学在这个认识过程中起了十分重要的作用。特别是心理语言学,由于它跨学科的特征,使我们能通过对人们使用语言和学习语言的心理机制来透视人脑处理信息的普遍特征。本文拟从语言习得的角度来探讨目前风靡一时的联结主义模式(connectionist models)^①,并以此讨论认知科学及语言学的一般性问题。

从上世纪五十年代末期到今天,Chomsky的理论一直在语言学中占主导地位。Chomsky对传统的语言学理论提出了挑战,认为语言知识从根本上是一种心理机制,而这种机制的根本又是形式语法系统。也就是说,人脑是通过一个内存的规则系统(形式语法)来反映语言的。过去几十年中,Chomsky不断更新他对形式语法系统的描述,从原有的“转换生成语法”到今天的“最简方案”,虽然其间有不少变化,但不离其对规则的基本诉求。心理学家和心理语言学家们同样对规则系统深信不疑,认为只有规则系统才能够有效地反映人脑的高级抽象活动。这种认识乃是基于认知科学家的一个基本假设:人脑是处理符号系统(symbol system)的机器(Newell 1980)。这个假设对认知科学起了很大的影响:一旦我们将人脑当作符号系统,我们就可以很方便地描述这个机器对符号加工与处理的方式。从某个角度来看,我们可以拿这部机器与计算机作比较:描述人脑的过程跟描述计算机的软件操作过程一样。

这种将人脑看作符号系统的观点与心理学的模块理论(modular theory)有着密不可分的关系。18世纪Franz Gall提出了模块理论的最初假设。但那时的假设强调人的性格特征与脑骨骼的外型特征的关系,因而缺乏科学根据。现代心理学对模块理论表述最完备的莫过于Jerry Fodor(1983)。Fodor认为人脑的认知系统是由许许多多的模块组成的。这些模块有的负责语法、有的负责视觉、有的负责听觉,任务专一(domain-specific),互相独立(autonomous)。对于心理语言学来说,最重要的是这些模块在语言的加工过程中不能同时互动

^① Connectionism又称为neural networks(神经网络),国内有学者译作“连接主义”。但笔者认为“联结主义”能够更好地反映这个理论的特征。

(parallel interaction)。例如,当你听到“小明和小张在切蛋糕”这句话时,模块理论假设,我们是由语音系统开始,然后对词汇,再对语法,最后对语义进行加工。这是由低层到高层的一个过程(bottom-up process),次序严谨,不能打乱。再者,在对语法加工的同时,语音和语义都不能起作用:每一层面的信息都是自给自足的(informationally encapsulated)。模块理论的线性次序,及其分明的层次,对认知科学家具有极强的吸引力。但是,近十几年来它也受到了强烈的挑战。对模块理论及其在大脑中的表征,读者可参看 Uttal(2001)较系统的阐述及批判。对其挑战的主要理论当属联结主义了。

我们知道,符号系统的观点及模块理论的假设是建立在将人脑比作电脑的基础之上的。这种比拟的优点是,我们能够有效地讨论人脑在信息处理时的操作过程及加工特征(如线性次序,模块结构,加工流程图等等)。但它最大的缺点是难以在生物及神经学上找到对应的关系(neurally implausible)。人脑内有上千亿神经元,而且这些神经元之间的联结关系比起电脑中几百或上千的电极管要复杂得多。还有,电极管每秒可以进行几百万或几千万次运算,而神经元每秒则只可以发送或接收几百次电子化学的脉冲。因此,如果人脑是按线性次序来操作,每秒不过能计算一百次左右(100-step rule, Feldman and Ballard 1982)。显而易见,每秒一百个操作步骤是不能够完成复杂的认知过程的。例如,词语的加工过程至少精确到十分之一秒。最后,数字电脑只能接收单一的、清楚的符号信号(all or none),没有所谓的中介状态(partial status)。这与人脑的灵活性及可塑性有极大的差别。所有这些原因都给联结主义的观点铺下了基石。

联结主义的一些初期理论就已经与模块理论的基本假设针锋相对了。最著名的要算“互动激活”(interactive activation)理论。Rumelhart 和 McClelland(1981)提出了互动激活的基本假说。根据这个假说,语言加工的过程既包含从下至上的过程(bottom-up process),也包含从上至下的过程(top-down process)。与模块理论的假说相反,这两种过程可以在同一时间互动。举例来说,当你听到“小明和小张在切蛋糕”这句话时,既可有语音至词汇至语法至语义的过程,也可以有语境的作用由上至下帮助听者理解语义、语法、词汇及语音。这两种过程可以从听者在噪音的干扰下仍能完整理解句子的情况中看出来。如果“蛋糕”的“糕”字突然受到干扰(例如在电话交谈中),听者的理解系统可以自动修补并添加“糕”的语音。Rumelhart 和 McClelland 还举例说明,如果英文字母 R 或 K 的右上角被遮盖(类似 h),读者可以根据词的周围语境(WOR-)自动修补,达到理解 K 而不是 R。这种语境效应或词优效应(word-superiority effect)对互动激活的理论提供了有力的支持。

2. 联结主义的基本特征

互动激活的假说给联结主义用于语言分析中打下了基础。但严格地说,它还不能算是联结主义的模型。按照 Rumelhart, McClelland, 和 PDP(parallel distributed processing) Group (1986)的 PDP 理论,联结主义有以下两个基本特征。首先,在知识的表征(representation)方面,它强调“分布表征”(distributed representation)。分布表征与传统认知理论对知识的表征有很大的不同。上面我们提到,传统认知理论将人脑看作是符号处理系统,因而它采用的是“方位表征”法(localist representation)。这种表征的基本特点是一个信息加工的单位(或单

元)只表达一个概念(例如语素、字或词),而一个概念也只由一个单位来表达。这样,表达单位不能进一步分解为更小的单位,因为它与概念间有清楚的一对一的关系。分布表征与此不同:它强调一个概念由多个单元互相作用的关系来表达。例如,英文大写字母 F 和 E 之间的不同在于后者多了一横。照方位表征法, F 和 E 是分别由两个不同的单元来表达的。照分布表征法, F 和 E 可以由多个同样的单元来表达,所不同的某些单元在表达 E 时被激活,但在表达 F 时被抑制。这样一来,我们如果仅看这些个别的单元,它们既不表达 F,也不表达 E。F 和 E 的知识是由多个单元之间激活的关系来表达的。

联结主义区别传统认知理论的第二个基本特征在于它对知识学习的看法。这也是本文需详细介绍的。长期以来,心理语言学家认为,学习语言就是一个学习规则的过程。这种观点,如前所述,是与 Chomsky 的语言学理论密不可分。联结主义则认为,Chomsky 理论提供了有效的规则系统来描述语言本身,但这个系统不能描述学习的过程。由于联结主义采用分布表征,它认为知识学习的过程就是学习分布表征的过程。换句话说,学习是经过调节单元与单元之间的关系来完成的,而调节单元与单元之间的关系又是经过改变单元与单元之间的权值(weight)来完成的。那么什么是权值呢?权值是表达单元与单元之间联结的强度。权值数越高,单元之间的联结就越强。一旦联结网络中相应的单元都由适当的权值联系好了,知识的表达和学习的过程也就完成了。以上述简化的例子而言,如果我们已经学会了 F 这个字,那么学习 E 时只需要将最下部分的单元激活并给予高强度权值,将其与网络中其它单元联结起来,我们便学会了 E。

很显然,联结主义的这些特点,与传统的认知理论相比,有较强的“生理可解性”(biological plausibility)。单元、激活、抑制,以及联结强度等概念,都能在人脑中找到直接的对应。反观传统的认知理论,符号、规则、语言树形图等概念则相当地抽象,难以简单地对应于特定的生物机制。联结主义的目的就是要通过对前类概念的描述达到对后类概念的解释。

联结主义的思想早在上世纪四十年代初期就出现了(McCulloch and Pitts 1943)。McCulloch 等人认为神经网络的结构可以解释数理逻辑的功能。与当今的联结主义网络不同的是,他们的网络的输出只能是二进制的(on 或 off),且单元间联结的强度不能通过学习而改变。在 McCulloch 和 Pitts 之后有许多人对联结主义的思想加以改进,其间以五十年代末期的视觉网络(perceptron)最引人注目(Rosenblatt 1958)。视觉网络虽然克服了许多 McCulloch-Pitts 网络的问题(如不限于二进制输出,也可以在学习中改变单元的联结强度),但与所有早期的联结主义模型一样,都只能解决简单的“线性可分”(linearly separable)的问题,如逻辑上的“和”(and)与“或”(or)问题。对于“线性不可分”(或称“非线性”)问题,比如逻辑上的“排它或”(exclusive or, 简称 xor)问题,它们则是一筹莫展。例如:

- (1) ○ ○ → A (箭头表示“归类为”)
- (2) ○ □ → B
- (3) □ ○ → B
- (4) □ □ → A

在这些例子中，(1)和(4)之间的差别最大(两个圆形对两个方块)，(2)和(3)则都是由一个圆形加一个方块组成，只是次序不同。之所以说这个问题是“线性不可分”，是因为它要求将差别最大的单位归为相同的范畴(A)。这种分类法不能简单地在问题的平面上用直线切开，但对于人来说，我们能够灵活地使用非线性方法解决此类问题。从上世纪六十年代起，研究者们就开始考虑如何能使联结主义网络解决非线性问题。Rumelhart, McClelland 和 PDP Group(1986)的 PDP 理论对解决这类问题提出了有效的方法。

PDP 理论的联结主义网络一般由三个层次组成：输入层(input layer)、内隐层(hidden layer)和输出层(output layer)。输入层接受输入的代表(如汉字的字形)，输出层提供输出应有的代表(如汉字的分类)，而内隐层则存储网络所学到的知识代表(如汉字在各个不同学习阶段的形体)。网络学习由输入层开始，至内隐层，再达到输出层。这个学习过程是一个调节网络中各单元的激活程度及单元之间的联结强度的过程。PDP 理论对解决非线性问题最大的贡献在于它对内隐层与其它层次之间的调节方法(或称算法)。联结主义网络中至今最有影响的算法可能要推“反馈学习法”(back-propagation, 简称 BP 算法, Rumelhart Hinton and Williams 1986)。按照 BP 算法，网络每次学习输入与输出的关系时，同时也接受一个“指导信号”(teacher)。这个指导信号乃是网络应该提供的正确的输出。如果网络所产生的输出信号与指导信号有差别，那么这个差别的大小就会计算为网络的误差率。误差率然后反馈至网络，使相关的单元与单元之间的权值得到改变。这样不断改变的结果使网络能最后正确地产生所有的输出。最重要的是在这个不断地调节过程中，单元间的权值及内隐层单元的激活能够最有效地反映输出与输入间的关系，从而有效地反映输入层单位间的内在关系^②。

综上所述，联结主义理论与传统的认知理论有很大的区别。它所运用的基本概念都与人的生物机制有一定程度上的对应关系，例如单元对应于神经元，单元的联结对应于神经元的联结，权值对应于联结强度，激活与抑制对应于神经元间电生理活动的方式。如何能够利用单元、联结、权值、激活与抑制这些概念去更好地解释传统认知理论中的重大问题乃是联结主义理论成功的关键。由于语言习得是认知科学中的重大课题，下面我们讨论联结主义是如何解释语言习得的。

3. 联结主义网络在语言学及语言习得中的运用

由于内隐层和 BP 算法的出现，联结主义网络不单能够解决简单的非线性问题，如 xor，而且能够原则上解决任何非线性问题。语言现象也许是最复杂的和最有代表性的非线性问题之一。联结主义网络打开了语言研究的一扇新大门。Rumelhart 等人 1986 年的两卷 PDP 论著为联结主义作出了划时代的贡献，而语言又是其中讨论得最多的一个环节。在联结主义看来，我们现有的语法规则及语义范畴都只能作为有效的语言学理论描述，但不能作为心理表征的机制。换句话说，语言学理论有实用价值，但没有心理现实性。这种观点自

^② 由于篇幅及内容的限制，我们在这里撇开许多技术上的细节，着眼于提供与语言学有关的理论描述。想进一步了解 PDP 理论的读者可阅读 Rumelhart 等人 1986 年的两卷 PDP 论著。对联结主义与认知科学有兴趣的读者可阅读 Bechtel and Abrahamsen(1991), Ellis and Humphreys(1999), Spitzer(1999)的论文。对技术细节或数学模型有兴趣的读者可阅读 Andersen(1995), Dayhoff(1990), Fausett(1994), 以及 Hertz, Krogh and Palmer(1991)的论文。

1986年由 Rumelhart 等人在 PDP 一书中提出后,引起了极大争论。这个争论一直持续到今天,并无最后定论。

那么 PDP 理论怎样看待语法规则及语义范畴呢? Rumelhart 等人在 PDP 一书中有许多章节涉及到语言学的问题,包括言语感知,句子理解,语言习得等等,我们不能在这里具体一一加以讨论。但其中对语言学最有影响的一章就是 Rumelhart 与 McClelland 提出的英语过去时态的 PDP 模型。下面我们简单地介绍一下这个模型。

3.1 联结主义网络对语法规则的学习

众所周知,Chomsky 批判传统的行为主义心理学最有力的证据,就是儿童并非简单模仿成人语言,而是利用对规则的掌握进行类推。比如,儿童学习到一定阶段时,会说 broken 作为 break 的过去式,而不是说正确的不规则形态 broke (Brown 1973; Bybee and Slobin 1982; Kuczaj 1977)。broken 在成人语言中根本不存在,模仿学说的理论显然难以自圆其说。根据 Chomsky 的理论,语言习得研究者一直认为,最有效地解释儿童“泛化”(overgeneralization)的方法就是假定儿童在学习的某一阶段已经掌握了一个抽象的内在规则,如“在任何动词后加 -ed 成为该动词的过去式”,或“在任何名词后加 -s 成为该名词的复数形式”。由于内在规则的普遍适用性,儿童便把不规则的动词当作规则动词来处理(regularization),产生 broken, comed, 或 failed 等错误。要纠正这些错误,儿童必须逐字学习,加以校正。这个逐字学习的过程与规则的掌握过程完全不同。所以, Pinker 等人认为儿童在掌握英语过去时态时,有两种不同的学习机制在起作用(Pinker 1991, 1999; Pinker and Prince 1988):一种是学习一般性的形态规则,由此能产生泛化的结果。另一种则是“联想学习”(associative learning),将不规则动词的形态与其基本形式逐个对应起来。前者负责一般规律,后者负责单个例外。

PDP 理论与这种观点截然相反。Rumelhart 与 McClelland 的英语过去时态模型强调儿童学习过去式只有一种机制在起作用,那就是联结主义的机制。Rumelhart 等人使用了一个简单的联结主义网络来模拟儿童的学习过程,发现该网络能产生“U-形学习效应”。所谓 U-形学习效应是指儿童在早期的学习过程中基本不犯语法错误,如正确地使用 broke, came 或 fell。在中期的学习阶段,错误大量出现,如不正确地使用 broken, comed 或 failed。儿童在后期的学习阶段才逐步将错误消除(Bowerman 1982)。这种效应以前人们一直借助于规则的学习来解释:儿童早期没有学到规则,中期学到规则以后泛化使用规则,后期逐字调节,对规则的使用范围加以改正。在 Rumelhart 等人的联结主义网络中并无任何规则的表征,但网络却显现出规则的效应。这个网络是怎样达到这种效应的呢?在这个模拟中,网络收到的是每个动词词根的语音特征,然后与它的过去式的语音特征加以匹配。每次匹配的同时,网络中的联结权值得以改变。正是这些联结权值使网络对动词的基本形态与它的过去式之间的关系有了详尽的了解。这些关系反映过去式形态变化的基本规律(flow, glow, slow 都是带 -ed 作为过去式),从而指导网络在学习新动词时的类推行为(如 blow 也应带 -ed 作为过去式)。在这种学习过程中,网络能有效地将规则动词与不规则动词区别对待。但在同时,这个过程所产生的结果既有将不规则动词当作规则动词的情况(regularization, 如 blowed),也

有将规则动词当作不规则动词的情况(irregularization, 如 *ment* 作为 *mend* 的过去式)。后一种情况的产生是由于网络学到了一些不规则动词中的“次规律”(sub-regularities), 比如 *lend*, *send*, *spend* 的过去式分别是 *lent*, *sent*, *spent*。这种情况似乎难以用上述 Pinker 等人提出的“规则与例外”的双机制来解释。

Rumelhart 与 McClelland 的模型的一个核心的思想就是语言学规则是“浮现特征”(emergent properties)。也就是说, 联结主义网络通过单元、激活、抑制, 与联结等特征能够有效地表达语言行为, 而这种表达的有效程度仿佛其背后有语言学规则在支配。由上述所见, 单一的联结主义机制既能反映儿童对规则过去式的掌握, 也能反映其对不规则过去式的掌握。规则本身不需要在系统中明确表征, 但却通过网络学习浮现而出。我们可以通过一个简单的例子来了解浮现特征这个概念(Bates 1984)。联结主义中的规则行为可以与蜂窝的六角形状来加以比较。从单个的蜜蜂的行为来看, 蜂窝的六角形状似乎不可思议。但如果我们分析其动态物理的特征, 那么六角形则是恰到好处。每个蜜蜂在构造蜂窝时都只需要一小点蜜, 但当多个蜜蜂从多个角度将蜜一点一滴地挤入蜂窝, 当许多柔软的小圆形的蜜受到多角度同时挤压时, 整体蜂窝的形状便自然而然地成为六角形。在这种情况下, 我们说六角形是浮现特征, 而不需要假设蜜蜂拥有一个制造六角形的规则系统。最近, 语言学家和心理语言学家对浮现特征从多个角度给予了讨论, 一些相关的论点在 MacWhinney(1999)一书中有所介绍。

Rumelhart 等人的 PDP 模型的出现引发了一系列的争论, 尤其是它与 Pinker 等人的双机制的争论直到今天仍僵持不下。Pinker 等人对 Rumelhart 与 McClelland 的模型提出了许多问题, 尤其是认为它在词汇的表征上, 在模拟的程序上, 以及在语音语义的关系上都不能反映儿童学习过去式中的许多细节。后人对 Rumelhart 与 McClelland 的模型作了较大的修改(包括结构上的, 表征上的及训练程序上的修改, Plunkett and Marchman 1991, 1993; MacWhinney and Leinbach 1991), 发现虽然原有的模型确有缺陷, 但扩充后的模型仍支持原有模型的基本观点。从这些争论中我们可以回到本文开头讨论的问题而看到一个基本的对立, 那就是应该怎样看待人脑的构造与功能: 人脑到底是一个模块的符号处理系统呢, 还是一个多元的分布处理系统?

3.2 联结主义网络对语义范畴的学习

自 Rumelhart 和 McClelland 模型问世以来, 联结主义在语言习得中的研究主要注重在对语法规则和语音结构的表征上, 很少在语义方面下功夫。理由很简单: 语义太复杂。因此即使偶尔有涉及语义的联结主义网络, 也只是随机抽取语义特征, 而后加以轻描淡写。但是联结主义的分布表征及学习的特点其实对解决语义方面的问题有极大的帮助。有鉴于此, 笔者在上世纪九十年代开始研究如何用联结主义网络来学习语义范畴。

Li(1993)及 Li 和 MacWhinney(1996)从“隐型范畴”(cryptotype)着手研究语义的习得问题。隐型范畴在语义学中是个棘手的问题。Whorf(1956)在 1936 年对隐型范畴作了如下的“定义”: 隐型范畴是微妙的, 看不见也摸不着的, 不能以一个简单的标志加以命名的。这样的定义似乎叫人对隐型范畴最好敬而远之。以英语的前缀 *un-* 为例, 很多动词可以带 *un-*

(如 unbuckle, undress, unfasten, untie), 但也有很多动词不能带 un-(如 *unbuild, *unkick, *unmove, *unpush)。Whorf 认为有一个隐型范畴在支配 un-的使用。问题就在于语言学家不能清楚地描述隐型范畴。隐型范畴必须通过其它的型态标记(如前缀 un-)来负面定义。

Bowerman(1982)对 Whorf 提出的隐型范畴在语言习得中的作用做了探讨。她认为儿童在学习动词前缀 un-时经历一个与学习过去时态一样的 U-型效应。儿童在第一阶段正确地使用带 un-的动词, 因为他们尚未将动词词根与前缀区分开来。在第二阶段时大量的泛化使用 un-错误开始出现(如 *unhold, *unpress, *unsqueeze等)。在这个阶段重要的是儿童已经认识到了 un-的隐型范畴, 因此与隐型范畴相似的动词都被用来带 un-。最后阶段儿童才纠正错误。Bowerman 这样的解释十分合理, 但最大的问题是没有说明儿童是怎样获得 un-的隐型范畴的。

Li (1993)及 Li 和 MacWhinney(1996)模拟了联结主义网络学习隐型范畴的过程。网络的任务是按照能否带 un-给动词加以分类。我们的假设是隐型范畴之所以“隐型”, 乃是由于(a)隐型范畴涉及复杂的语义关系;(b)隐型范畴涉及动词的许多语义特征;(c)不同的语义特征在隐型范畴中有不同的激活程度;(d)语义特征之间存在着不是互相排斥而是相互交叉的情况。联结主义网络所使用的分布表征及非线性学习给我们研究隐型范畴提供了最理想的工具。我们的模拟结果表明, 当网络学到一定的词汇量时, 隐型范畴在网络的内隐层浮现而出。更重要的是, 当网络继续学习新词时, 隐型范畴指导它进行类推, 产生类似儿童在第二阶段时泛化使用 un- 的错误。这些结果表明, 联结主义网络可以通过学习带 un- 动词的语义特征之间的复杂关系以及这些特征与前缀共现的规律来形成隐型范畴的表征。通过对网络内隐层的统计分析, 我们可以看到带 un- 的动词有一定的特点, 而不带 un- 的动词有另外的特点。这些结果进一步说明学习隐型范畴或 un-不是一个简单的规则学习过程, 而是逐步累计相关特征的计算过程。这个计算过程考察词义, 词型, 以及词缀之间在所学语料中共现的频率与规律。我们的结果与前面讨论的联结主义网络学习语法规则的结果十分一致。两者都说明联结主义模型的单一机制能学习语言中的许多现象。

3.3 联结主义网络对语言先天性的看法

在前面我们提到 Chomsky 的理论对语言学产生了深远的影响。Chomsky 对于规则系统的阐述可谓尽善尽美。但其理论的另一核心是规则的“先天性”(innateness)。这个问题在语言学中有很多详尽的讨论(李行德 1992), 本文不多加赘述。与 Chomsky 理论相反, 联结主义理论强调学习的重要性, 强调网络从语言素材中抽取规律的能力。但与简单的经验主义(empiricism)不同, 联结主义并不否定先天性。这一点在 Elman 及 Bates 等人的《对先天性的再思考》(*Rethinking Innateness*)一书中有详细的讨论。在这里我们只简略地介绍一下 Elman(1990)等人的观点。

Elman (1996)等人认为, 前人对先天性的认识局限于单个层次, 但先天性本身有三个层次值得研究。第一个层次是表征上的层次(representational)。这个层次的先天性是指人脑具有先天固有的神经系统, 而且这个系统中的神经元之间的关系早已确定为表达特定的范畴与

概念。后天的经验或学习对这个系统的影响甚微。第二个层次是结构上的层次 (architectural)。这个层次的先天性是指人脑的构造对信息的加工或问题的解决有什么样的限制。人脑在局部或整体都有一些构造特征, 比如单个神经元的信息处理速度限制在每秒 100 个步骤左右, 比现有的数字计算机慢了许多(如前所述)。第三个层次是发展速度上的层次 (timing of maturational events)。这个层次的先天性是指人脑的各个区域有不同的发展进程, 如脑功能侧化 (hemispheric lateralization) 及神经元的再生 (neurogenesis)。语言习得的“关键期” (critical period) 就可能是由于人脑可塑性的降低而导致的, 反映发展速度上的先天性。这三个层次上的先天性都在前人的讨论之列, 但在语言学家眼中(从 Chomsky 到 Pinker, 再到 Bickerton), 先天性大多停留在第一个层次上。有趣的是, Elman 等人从神经生物学出发, 特地反驳第一层次上的先天性。他们指出, 人脑的 DNA 本身并无足够用来表达人类所需的多如牛毛的具体概念与范畴, 况且人脑的后天可塑性也与固有神经系统的看法不一致。因此, Elman 等人认为结构上和发展速度上的先天性更为合理及有效, 而且这两个层次上的先天性可以直接在联结主义网络中得到反映与表达(如网络的结构、关系及学习速度等)。

先天与后天, 自然与哺育的争论, 自古希腊哲学家开始一直到今天都没有完整的答案。语言学家、心理学家、及认知科学家现在开始寻找新的角度来探讨这个问题。包括 Elman 等人在内的一些学者认为, 单靠内在机制或外在因素都不足以解答人与环境之间的复杂且丰富的相互作用关系。因此我们应该仔细研究人与环境之间相互作用下所产生的“浮现特征”。这些浮现特征从联结主义的角度来看正是网络与学习材料之间相互作用的结果。Nelson (1999) 将这种观点推到一个新的层次, 认为人的神经系统本身会随着学习经验的增加而加以改变或得到发展。也就是说, 内在的神经机制本身也不是一成不变的。显而易见, 在这种情况下再坚持谈内在与外在或先天与后天谁更重要就显得毫无意义了。

4. 自组联结主义网络与语言习得

联结主义自 Rumulhart 等人的 PDP 论著问世以来已经在语言学、心理语言学、神经语言学以及语言习得中引起了一波又一波的研究高潮 (Ellis and Humphreys 1999)。但迄今为止这些研究大都局限于以下三个方面。首先, 大部分涉及语言的联结主义模型都只探讨语法或语音等语言形态方面的特征 (formal properties of language), 而很少研究语义或语用方面的特征。这一点在前面我们已经提到, 主要原因是后者的研究难度较大。第二, 以前的研究大都只使用极少数量的语言素材, 从几十到几百词汇不等。最著名的联结主义网络之一, Elman (1990) 的“简单回馈网络” (simple recurrent network) 只用了 29 个名词和动词。但这些网络能否适用于广泛的、大量的语言素材则是个问题 (所谓 scalability 的问题)。这如同语言学家用几个例句能否解释大量语言学现象一样。第三, 大部分研究语言的神经网络都只采用了典型的反馈学习法 (BP 算法)。BP 算法网络, 如前所述, 有特定的指导信号反馈网络, 使相关的权值加以改变。它是一种属于“有指导学习” (supervised learning) 的网络。这种网络在研究语言习得方面的可行性很值得怀疑 (Li 1999)。虽然儿童学习语言时也有成人指导和儿童模仿的成分, 但自从 Chomsky 批判行为主义的语言学说以来, 语言学家们大都认为儿童学习语言时不需要或不接受“错误反馈” (negative evidence, Bowerman 1988)。换句话说, 语言习

得基本上是一个无需指导的学习过程。

最近几年笔者及合作研究者试图突破以上几方面的限制,研究一种无需指导的自组联结主义网络(self-organizing connectionist network)来探讨语言习得(Li 1999, 2000; Li and Shirai 2000; Li and Farkas 2001)。这种网络属于非指导学习(unsupervised learning)的神经网络。自组联结主义网络相比传统的BP网络对语言习得而言有更大的心理现实性及生物有效性。在这种网络中,学习通常是在二维平面图中进行的(又称“自组网图”, self-organizing maps 或简称 SOM; Kohonen 1982, 1989, 1995)。网图中的每个单元都能对一个或多个输入单位加以反射。在学习的最初阶段,输入单位随机激活网图中的一个单元,这个单元就成为该输入单位的反射代表。随着网络的不断学习,该单元及其周围的单元对权值不断加以调节,使网图在下次处理同样的输入时能够激活同样的或邻近的单元。这样不断调节的过程就使网图上的每个单元只对某些特征相似的输入加以反射,从而使得网图能够利用有限的二维平面来表达多维的输入特征。

Miikkulainen(1993, 1997)将多个网图连接起来,用以学习语音、语义及字型的关系。每个网图本身只表达语音、语义或字型,但网图与网图之间通过赫伯学习法(Hebbian learning)来联结,以模拟各语言层面可能产生的相互作用。赫伯学习法(Hebb 1949)是一种有生物基础的规则。它的主要原则是两个神经元如果同时激活,它们之间的联结强度就会相应提高。笔者与实验室的研究人员近年利用这种多重网图模型来模拟语言习得中的一些具体问题。我们的模型一个最大的特征就是它能通过自组学习,对大量的语言素材进行加工,从词与词在句中中共现的机率中提取语法语义范畴。这种提取是根据最近自然语言处理中对大语料库加工的相关理论而产生的。Burgess 和 Lund(1997, 1999)提出了 hyperspace analogue to language(HAL)的理论,认为自然语言素材中词与词之间的关系提供了足够的语义信息。Landauer 和 Dumais(1997)也提出了类似的理论(Latent semantic analysis),认为语义可从词与篇章的关系中提取。在一系列的研究中,我们发现如果儿童分析成人话语中词与词的共现关系及其频率,可以获得词的语义及语法关系(Li, Burgess and Lund 2000)。这个结论与最近研究幼儿切分话语单位的结论是一致的(Saffran *et al.*, 1996, 1997)。同时,我们还提出了词汇的发展模型(DevLex),用以不断学习新的词汇表征(Farkas and Li 2001)。DevLex 不限于固定的词汇,而是通过语料的增加而相应地增加新词,并可以不断增加网络中的单元数目及网图数目(Farkas and Li 2002)。这种逐步增加的过程可以更适当地反映儿童语言学习或成人外语学习的过程。

我们将 DevLex 模型运用到语言习得中的几个具体问题上,比如前面提到的语义隐型范畴与前缀 un-的关系(Li 1999),英语时态的学习(Li 2000; Li and Shirai 2000),中英双语的词汇表征(Li and Farkas 2001; Li 2001 及 Li 的综述, 2002)。结果表明模型能有效地提取及表达语法语义范畴。在中英双语的模拟中,两种语言的词汇及语音都被网络自然地分离开来。在前缀与时态的模拟中,语义范畴的出现指导着形态标记的使用,从而产生儿童语言中类推或泛化的现象。总而言之,我们的模型克服了传统联结主义模型的局限:它利用自组而非反馈网络,学习大量自然语料,解决语义语法问题,从而达到更自然地反映语言习得本质

的目的。

5. 结语

从以上四个部分的讨论中,读者可以看到联结主义近十几年来对西方语言学、心理学及认知科学产生的巨大影响。可惜的是,联结主义应用在中文上的研究寥寥无几。除了陈鹰和彭聃龄(1994)对汉字认知以及笔者对语言习得的研究外,基本上找不到其它的文献。这与中国语言文字科学的发展是极不相称的。笔者希望通过本文起到抛砖引玉的作用,使国内学人将语言学与联结主义的研究推向一个高峰。

参考文献

- Anderson, J. 1995. *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.
- Bates, E. 1984. Bioprograms and the innateness hypothesis. *Behavioral and Brain Sciences*, 7, 188-190.
- Bechtel, W. and Abrahamsen, A. 1991. *Connectionism and the Mind*. Cambridge, MA: Blackwell.
- Bowerman, M. 1982. Reorganizational processes in lexical and syntactic development. In E. Wanner and L. Gleitman, eds., *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press.
- . 1988. The "no negative evidence" problem: how do children avoid constructing an overly general grammar? In J. Hawkins, ed., *Explaining Language Universals*. New York: Basil Blackwell.
- Brown, R. 1973. *A First Language*. Cambridge, MA: Harvard University Press.
- Burgess, C. and Lund, K. 1997. Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes*, 12, 1-34.
- . 1999. The dynamics of meaning in memory. In E. Dietrich and A. Markman, eds., *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines* (pp. 17-56). Mahwah, NJ: Erlbaum.
- Bybee, J. and Slobin, D. 1982. Rules and schemes in the development and use of the English past tense. *Language* 58: 265-289.
- Dayhoff, Judith. 1991. *Neural Network Architecture: An Introduction*. New York: Van Nostrand Reinhold.
- Ellis, R. and Humphreys, G. 1999. *Connectionist Psychology: A Text with Readings*. Psychology Press: Taylor and Francis.
- Elman, J. 1990. Finding structure in time. *Cognitive Science*, 14, 179-211.
- Elman, J. Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Farkas, I. and Li, P. 2001. A self-organizing neural network model of the acquisition of word meaning. In E.M. Altmann, A. Cleeremans, C.D. Schunn, and W.D. Gray, eds., *Proceedings of the Fourth International Conference on Cognitive Modeling*, pp.67-72. Mahwah, NJ: Lawrence Erlbaum.
- . 2002. Modeling the development of the lexicon with a growing self-organizing map. In H. J. Caulfield et al., eds., *Proceedings of the Sixth Joint Conference on Information Science*, pp.553-556. Association for Intelligent Machinery, Inc.
- Fausett, L. 1994. *Fundamentals of Neural Networks*. Englewood Cliffs, NJ: Prentice Hall.
- Feldman, J. A. and Ballard, D. 1982. Connectionist models and their properties. *Cognitive Science*, 6, 205-254.
- Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Hebb, D. 1949. *The Organization of Behavior: A Neuropsychological Theory*. New York, NY: Wiley.
- Hertz, J., Krogh, A. and Palmer, R. 1991. *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley.

- Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
- . 1989. *Self-organization and Associative Memory*. Heidelberg: Springer-Verlag.
- . 1995. *Self-organizing Maps*. Heidelberg: Springer-Verlag.
- Kuczaj, S. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior* 16: 589–600.
- Landauer, T., Dumais, S. 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211–240.
- Li, P. 1993. Cryptotypes, form-meaning mappings, and overgeneralizations. In E. V. Clark, ed., *Proceedings of the 24th Child Language Research Forum* pp. 162–178. Center for the Study of Language and Information, Stanford University.
- . 1999. Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. In M. Hahn and S.C. Stoness, eds., *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* pp. 308–313. Mahwah, NJ: Lawrence Erlbaum.
- . 2000. The acquisition of lexical and grammatical aspect in a self-organizing feature-map model. In L. Gleitman and Aravind K. Joshi, eds., *Proceedings of the Twenty Second Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- . 2001. Language acquisition in a self-organizing neural network model. In P. Quinlan, ed., *Connectionism and Developmental Theory*. Philadelphia and Brighton: Psychology Press.
- . 2002. Emergent semantic structures and language acquisition: A Dynamic Perspective. In H. Kao, C.K. Leong, and G.D., Guo, eds., *Cognitive Neuroscience Studies of the Chinese Language*. Hong Kong, China: Hong Kong University Press.
- Li, P., Burgess, C. and Lund, K. 2000. The acquisition of word meaning through global lexical co-occurrences. In E. Clark, ed., *Proceedings of the Thirtieth Stanford Child Language Research Forum*, Cambridge, MA: Cambridge University Press.
- Li, P. and Farkas, I. 2001. A self-organizing connectionist model of bilingual processing. In R. Heredia and J. Altarriba, eds., *Bilingual Sentence Processing*. North-Holland: Elsevier Science Publisher.
- Li, P. and MacWhinney, B. 1996. Cryptotype, overgeneralization, and competition: A connectionist model of the learning of English reversible prefixes. *Connection Science*, 8, 1–28.
- Li, P. and Shirai, Y. 2000. *The Acquisition of Lexical and Grammatical Aspect*. Berlin and New York: Mouton de Gruyter.
- MacWhinney, B. 1999. *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B. and Leinbach, J. 1991. Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, 40, 121–157.
- McCulloch, W. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7, 115–133
- Miikkulainen, R. 1993. *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*. Cambridge, MA: MIT Press.
- . 1997. Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59, 334–366.
- Nelson, C. 1999. Neural plasticity and human development. *Current Directions in Psychological Science* 8, 42–45.
- Newell, A. 1980. Physical symbol systems. *Cognitive Science*, 4, 135–183.
- Pinker, S. 1991. Rules of language. *Science*, 253: 530–535.
- . 1999. Out of the minds of babies. *Science*, 283: 40–41.
- Pinker, S., Prince, A. 1988. On language and connectionism: analysis of a parallel distributed processing

- model of language acquisition. *Cognition*, 28, 73-193.
- Plunkett, K. and Marchman, V. 1991. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 38, 43-102.
- . 1993. From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 21-69.
- Rosenblatt, F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386-408.
- Rumelhart, D., Hinton, G. and Williams, R. 1986. Learning internal representations by error propagation. In: David E. Rumelhart, James L. McClelland and the PDP Research Group, eds., *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.
- Rumelhart, D., James L. McClelland and the PDP Research Group, eds. 1986. *Parallel Distributed Processing. Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge, MA: MIT Press.
- Rumelhart, D. and McClelland, J. 1986. On learning the past tenses of English verbs. In: James L. McClelland, David E. Rumelhart and the PDP Research Group, eds., *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Vol. 2: Psychological and Biological Models*. Cambridge, MA: MIT Press.
- Saffran, J., Aslin, R. and Newport, E. 1996. Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Saffran, J., Newport, E., Aslin, R., Tunick, R. and Barrueco, S. 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, 8, 101-105.
- Spitzer, M. 1999. *The Mind within the Net*. Cambridge, MA: MIT Press.
- Uttal, W. 2001. *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: MIT Press.
- Whorf, B. 1956. *Language, Thought, and Reality* (edited by John Carroll). Cambridge, MA: MIT Press.
- 陈鹰、彭聃龄, 1994, 汉字识别和认知的连接主义模型。In H.-W. Chang, J.-T. Huang, C.-W. Hue, and O. Tzeng, eds., *Advances in the Study of Chinese Language Processing*. Vol. 1, Taipei: National Taiwan University Press, 211-240.
- 李行德, 1992, 语法的心理现实性。《国外语言学》第3期, 25-34页。

作者通讯地址: Li Ping
 Dept. of Psychology
 University of Richmond
 Richmond, Virginia 23173
 USA
 E-mail: pli@richmond.edu