理性主义、经验主义与语料库语言学¹

梁茂成

(北京外国语大学 中国外语教育研究中心, 北京 100089)

提要: 二十世纪的语言学发展历程是理性主义和经验主义对立与融合的过程。理性主义者主张语言学家依靠内省来解释语言, 而经验主义者则主张从实践出发, 以行为数据 (performance data) 为研究对象, 对语言进行科学、系统的描写, 并从中归纳出语言理论。当今, 语言学界的理性主义和经验主义正在进一步融合, 语言学研究必将由以实证、统计与语言学家的理性思考相结合的研究方法所主导。语料库将成为语言学家验证直觉、完善理论的重要途径, 并将引起语言学领域的一系列变革。

关键词: 理性主义, 经验主义, 语料库语言学

中图分类号: H030 文献标识码: A

语言学研究方法的不同源自哲学领域的思想纷争。理性主义和经验主义是哲学领域的两大重要思潮, 多个世纪以来一直深深地影响着语言学研究方法。本文回顾二十世纪语言学领域主要变革中的理性主义与经验主义之争, 在此基础上分析不同语言学研究方法间的差异以及语料库语言学内部不同方法之间的差异, 预测语言学研究方法的发展趋势。

1. 二十世纪语言学中研究对象的变化

二十世纪是语言学发展的辉煌时期。在这个世纪初期, 瑞士语言学家索绪尔创建了影响后来几乎所有语言理论乃至人文学科理论的结构主义思想, 其对语言与言语、所指与能指、历时语言学与共时语言学、横向组合关系与纵向聚合关系等基本概念的区分至今仍然被语言学界所广泛沿用, 影响着一代又一代的语言学家。这些二元对立关系构成了现代语言学最基本的概念。对于索绪尔来说, 语言是同一言语社团 (speech community) 所共同拥有的、同质的、不依赖于个人的东西, 语言学家无法对形态各异的个人言语进行研究, 因而语言才是语言学研究的真正对象 (Saussure 1966)。然而, 索绪尔也意识到, 语言和言语是紧密相关且互为前提、相互渗透的, 但在索绪尔所处的那个年代, 技术的发展状况还不能为语言学家观察和收集大量的言语事实提供足够的方便, 这注定言语事实只能被摆在相对次要的位置。

正当结构主义思想在欧洲一统天下的时候, 为二十世纪的语言学作出重要贡献的另一人物——乔姆斯基所倡导的转换生成语法理论在美国异军突起。乔姆斯基采用形式化的方法, 借助短语结构规则 (phrase structure rules), 提出普遍语法 (universal grammar, UG) 的理论, 认为人类具有一种与生俱来的习得任何语言的能力。和索绪尔不同的是, 乔氏认为, 语言存在于人们的大脑之中, 语言学家的任务不是去研究大量杂乱无章的言语事实等行为数据, 而是凭借语言学家自身的直觉 (intuition) 和内省 (introspection) 构建能够生成语言的语法 (Chomsky 1965: 20)。对于乔氏对语言能力和语言行为的区分, 后来有的语言学家欣然接受, 但也有有的语言学家 (如 Alien & Seidenberg 1999) 则认为两者不能截然分开。乔姆斯基的数理、认知等语言思想影响到自然语言处理、心理学乃至整个认知科学领域, 被人们称为“乔姆斯基革命”。

就在国际语言学界浪潮迭起、结构主义和转换生成语法交替主宰语言学历史舞台的同时, 另一批语言学家在悄无声息地进行着语言调查和描写。早在世纪之初, Vilem Mathesius (1882-1946) 和 John R. Firth (1890-1960) 就在不懈地进行“言语”的调查 (黄国营 1999), 而且此类调查从来没有停止过。随着各种技术的进步和人们对语言认识的提高, 这种以语言描写为主要目的的语言学理论与方法取得了很大进步 (见桂诗春 2004), 语言调查规模渐大, 并与计算机技术紧密地结合了起来, 建成了一个又一个大型计算机语料库。1959年,

¹ 本研究得到国家社科基金一般项目 (批准号: 09BYY033) 的支持, 特致谢忱。

伦敦大学 Randolph Quirk 组织发起了“英语用法调查”(The Survey of English, SEU)项目,有计划地收集大量英语语料。1964年,就在转换生成语法诞生地美国,布朗大学(Brown University)的一批学者建成了容量达到100万词的Brown Corpus。70年代,经G. Leech, S. Johansson和H. Hofland三方协同,英国的Lancaster大学与挪威的Oslo大学和Bergen大学合作建成了与Brown Corpus同等规模的LOB语料库,并且所收集的语料兼顾英语口语。1975年,以J. Svartvik为首的一批语言学家经过艰辛的努力,终于建成了第一个英语口语语料库London-Lund Corpus(LLC)。直到九十年代,Leech(1991: 9)仍然认为,就英语口语研究的语料源而言,LLC“至今仍无与伦比”。至此,语料收集进入笔语与口语并重阶段,随后建立的英国国家语料库(British National Corpus)、英语文库(Bank of English)等语料库兼收口语和书面语语料,且规模不断扩大,达到亿词以上,形成了二十世纪语言学界的又一次伟大革命——语料库语言学。经过数代语言学家的努力,以大量“言语”为研究对象、采用科学手段、口语和书面语并重的语言学研究方法终于从后台走到了前台,成为新时期描写语言学的主要方法。一时间,语料库语言学的大潮被众多语言学流派所接纳,使得只有“几百个词条和数十条语法规则的受限语言分析系统”被戏称为“玩具”(黄昌宁 2002)。

与结构主义和转换生成语法相比,语料库语言学不像前两者那样有一个划时代的伟大人物作为本学科或学派的领军人物。语料库语言学从根本上讲只是一种研究方法(Leech 1991),一种正在得到越来越广泛注意和接受的语言观。Svartvik(1996)指出,语料库正在成为主流,它不但是一种方法,而且也是一种新的哲学思维方式。著名辞典编撰家Michael Rundell(1996)认为,“大规模真实语料已从语言学研究的边缘地带走到了中心地带,成为语言学研究的基本资源”。Čermák(2003)认为,“语言学界的最高期待与语料库语言学紧密相关”。系统功能语法的代表人物Halliday(1993: 24)也承认,“语料库语言学把数据收集和理论概括结合了起来”,使我们对语言的理解发生了一个“质的变化”。

纵观二十世纪语言学发展的历程,在结构主义时期,语言学的研究对象是同一语言社团中人们所共同拥有的、同质的“语言”;在转换生成语法时期,语言学的研究对象是“理想的语言使用者”的“语言能力”;而到了世纪末,随着众多大型语料库的建立,语言学以大量的、人们实际使用的语言(即语言行为)为研究对象。乔姆斯基认为这种行为数据缺乏同质性因而没有研究的价值,但“把孩子和洗澡水一起倒掉”(McEnery & Wilson 1996)的方法似乎并不可取。撇开种种理论争议不谈,语料库已经成为“语言研究中默认的数据源”(Teubert 2005)已经成为不争的事实。

2. 理性主义与经验主义的分歧

关于语言与言语之间的关系,索绪尔认为,语言是社会的、同质的、抽象的,而言语是个人的、异质的、具体的,语言存在于言语之中,是社会集团的一种规约。索绪尔还认为,对言语活动的研究包含两个部分,一部分以语言为研究对象,另一部分以言语为研究对象。而且,“言语的事实总是在前的”(转引自黄国营 1999)。关于书面语和口语之间的关系,索绪尔认为,口语是第一位的、主要的,而书面语是第二位的、次要的(Saussure 1966)。但在索绪尔的《教程》中,言语的地位后来被抹杀了(黄国营 1999),其重要原因之一可能是限于当时的研究手段,语言学家无法在言语发生的第一现场对足够的言语事实进行记录和观察。由此可见,索绪尔的语言研究体系中并不排斥对言语这种行为数据的研究。

转换生成语法以笛卡尔的理性主义思想为基础,认为一切语言知识都来源于理性,语言实质上指的是“理想的语言使用者”的内在能力,因此不可能通过分析外在的言语事实而获得。基于这样的思想,乔姆斯基认为,人与生俱来就有一种习得语言的能力,语言学家的任务是要发现存在于人们大脑内部的、可以用来生成言语事实的那些规则。乔姆斯基反对通过收集语料的方法来解释语言,认为“如果要把语言学作为一门真正的学科”,我们所“观察到的语言使用情况(observed use of language)”“绝不可成为语言学研究的主体(surely cannot constitute the subject-matter of linguistics)”;从理论的角度看,“我们观察到的实际使用的言语,其大部分都是各种各样支离破碎、不规范的表达法(much of the actual speech observed consists of fragments and deviant expressions of a variety of sorts)”(Chomsky 1965: 4; 201)。换言之,人们大脑中储存的是井然有序的语言,但一旦由口中说出,顷刻间

就变成了杂乱无章的言语；然而，这些杂乱无章的言语一旦进入到听者的耳中，顿时又恢复到了原先的秩序。

转换生成语法学家置内省证据于首要地位，注重研究者自己的“直觉”。他们声称“操母语者的直觉”是“确凿素材”。凭借这些素材，他们可以对语言“进行描写甚至作出解释”（Chomsky 1965: 20）。为了说服人们，这些语言学家自己成了“完全同质的语言群体中的、完全了解该语言的、理想的语言使用者”（ibid: 4）的代言人。显而易见，转换生成语法难以自圆其说，他们一方面认为言语事实混沌无序，个体差异太大，不能成为语言学的研究对象，而另一方面，在缺乏研究素材的情况下，只好依靠自己的直觉编造一些句子，并根据这些句子演绎出一套套复杂的短语结构规则。在这其中，自相矛盾的论点是，乔姆斯基一方面认为普通人的言语不可靠，另一方面把自身的地位凌驾于普通人之上，编造出一些其他的语言使用者未必会说出来的言语作为例证，并依此构建语法规则。正如 Beaugrande (2002) 所说的那样，转换生成语法学家“自己充当理想语言使用者的模型这种做法靠不住，除非他们的地位和所接受的专门学术训练使得他们具备超人的内省能力”，因此他们很难称职地担当“完全同质的语言群体”的典型代言人。这些语言学家对自己编造出来的素材进行一番描述，并由这些句子推及到语言的一般本质，从而得出所谓的具有普遍意义的结论（Beaugrande 2002）。换言之，室内语言学家（Beaugrande 语）把自己凌驾于普通的语言使用者之上，仅从自己熟悉的语言出发，根据自己的直觉和内省，省去实地语言学家们辛苦的劳动，得出的却是针对抽象含义的语言理论，难免有闭门造车之嫌。他们一方面对普通人的日常言语行为不屑一顾，认为普通人的言语杂乱无章，另一方面又认为语言学家作为“理想的语言使用者”具有超越常人的、可以信赖的语言直觉。试问，如果自然语料像乔姆斯基所说的那样存在偏态，因而无法代表语言的全貌，那么我们有什么理由相信理性主义语言学家们编造的语料就不存在偏态呢？

可见，以理性主义为哲学基础的转换生成语法尽管声称语言研究不应该基于人们日常使用的言语事实，但并没有真正地拒绝言语事实，而是依靠自己编造的少量言语素材构建普遍语法的规则。

与理性主义相对立的另一种哲学思想是经验主义（empiricism）。正如英国语言学家罗宾斯（1997: 145）所说的那样，“经验主义和唯理论的对立，以不同的形式，贯穿整个语言学史”。早期的经验主义主张经验是人的一切知识和观念的惟一来源，认为经验能够提供对事物最权威的认识。他们片面地强调经验或感性认识的作用，贬低乃至否定理性认识的作用。面对千差万别的语言现象，经验主义认为对语言所下的任何结论都应该依据客观事实。

经验主义和理性主义在语言学界的相互对立由来已久。直到今天，尽管乔姆斯基的一些追随者已经发现语料库对研究语言的巨大价值，但乔姆斯基本人似乎仍然没有改变他原来对语料库的看法（Čermák 2003），语言学领域理性主义与经验主义的对立仍然存在。Beaugrande (2002) 认为语言学家常常可以分为实地语言学家（field linguists）和室内语言学家（homework linguists），而 Fillmore (1992) 则以沙发中的语言学家（armchair linguists）和实地语言学家（field linguists）来区分这两种截然不同的语言观。Lager (1995) 分别以“语料库语言学家”和“理论语言学家”冠名这两类语言学家，并对他们间若干个不同方面作出了如下总结：

表 1: 理论语言学家与语料库语言学家特性对照（引自 Lager 1995）

语料库语言学家	理论语言学家
研究对象为 言语/语言行为	研究对象为 语言/语言能力/语言系统
以 描述 语言事实及语料库中的语言使用为目标	以 解释 语言事实和语言使用为目标
研究方法为 数据驱动 ，常常提供很长的词表	研究方法为 理论驱动
满足于简单的测量工具，因为这些工具能够快速处理长文本；能够容忍简单方法不能处理的数据噪音（noise）	舍简单而取精密，舍速度而取准确；如果对数据加以验证，数据的样本量会很小
采用 定量 研究方法	采用 定性 研究方法
沿袭 经验主义 的传统	沿袭 理性主义 的传统
把文本视作 物理产品	把文本视作 理论实体

主要对特定语言的语法感兴趣	以构建普遍语法 (Universal Grammar) 为目的
只研究语言形式	研究涉及语言形式及意义
把握文本的整体属性, 通揽整个文本	把握文本的局部属性, 观察文本的细微之处
研究涵盖无限文本 (可能只涉及表面现象)	对有限的 (编造的) 文本作深入分析
采用概率论和统计学方法	采用基于规则 (基于知识, 基于推理, 基于逻辑) 的方法
依赖于对显性语言行为的观察	依靠直觉/内省获取资料
分析从语境中获取的真实语料	孤立地、脱离语境地分析“玩具式语例”
培根式思维, 认为归纳是科学的实质	认为假设/演绎是科学的实质
对发现程序 (Discovery procedures) 深信不疑	对合理性/验证的方法深信不疑

显然, 作为贯穿语言学发展历程的两大哲学思想, 理性主义和经验主义对语言学研究及其方法的认识过于极端, 不利于语言学的发展。

首先, 语言 (语言能力) 和言语 (语言行为) 之间的关系并非泾渭分明, 而是一个连续统。脱离语言去谈论言语或脱离言语去谈论语言都是不可操作的。语言学理论如果得不到言语事实的验证、对言语事实没有解释能力, 必将成为空洞而没有实际意义的理论。同样, 一味地强调言语事实而不对言语事实加以概括抽象, 研究便会失去意义。正如 Leech (1992) 所说的那样, “长期以来, 人们假想语言能力与语言行为之间存在一道鸿沟, 并过分地夸大了这一道鸿沟。语言可以理解为一种心理现象, 也可以理解为母语使用者的语言行为, 但因为后者是前者的产物, 故而两种含义的语法之间也必然存在着亲密的姻缘关系”。因此, 孤立地研究语言或言语, 不利于我们科学地认识语言学的本质。语言学不只是语言的语言学, 也不只是言语的语言学。

其次, 完全不依赖内省的唯事实方法和完全脱离语言事实的内省法都具有显而易见的局限性。观察和内省是我们获取数据的两种方法, 且两者之间存在着很大的互补性 (许余龙 2000), 两者间的矛盾也并非不可调和。引用 Johansson (1992: 334) 的话, “语法并不能奇迹般地从语料中自然产生, 而需要智能的语法学家对以往研究的成果、自身的直觉和对文本的观察加以分析和归纳”。理性主义忽视言语事实, 将整个语法体系建立在为数不多的、并非自然的语料上, 注定不能反映自然语言的全貌 (李葆嘉 2002)。

第三, 统计方法并非语言学研究的目的, 而只是进行定性分析的起点, 而且, 统计分析往往需要建立在一些初步认识的基础上。关于定量研究和定性研究方法, Bunge (1995) 曾作过这样的总结: “定量分析和定性分析之间并非水火不容, 两者之间存在着很大的互补性。量所统计的是具有同一质的样本的量, 或质的强度。因此, 在形成概念认识时, 质先于量。”可见, 将经验主义的定量分析和理性主义的定性分析结合起来, 将更有利于语言学的发展。

3. 理性主义与经验主义的融合

令人欣慰的是, 到了近代和现代, 哲学领域传统的理性主义逐渐发展为科学的理性主义, 不断用实验对理论进行修正, 表现出经验和实验与演绎推理的密切结合与相互容纳的趋势 (宋春阳 2003)。经验主义发展到中世纪以后也越来越多地带有理性主义的色彩, 到了二十世纪, 经验主义与理性主义并存和融合的现象日趋明显 (陈勇 2003)。

受这种融合趋势的影响, 现、当代的经验主义语言观显著不同于古代的经验主义语言观, 它注重经验, 但同时汲取理性主义合理的方面, 在方法上既注重观察和收集语言素材, 并把这些言语事实作为语言分析的第一手数据, 进行进一步的归纳, 以推及一般, 促进语言学理论的发展和更新。同时, 现代的经验主义也不排斥语言学家的内省和直觉。Grishman (1986) 认为, 语料库语言学的一个自然的功能是对语言学家提出的各种语法进行检验。著名语料库语言学家 J. Aarts (1992) 认为“语料库是语言学家们验证假设的试验台”, 主张语言学家基于直觉构建一套语法规则, 并在语料库中不断验证这一套语法规则, 通过这样的途径把基于直觉而构建的语法改造成为基于观察而构建的语法。显然, 这种设想试图把语言学家的直觉与来自言语事实的大量数据结合起来。从哲学的角度看, 这种方法体现了理性主义与经验主

义的融合。Beaugrande (2002) 把语言本身视作为理论, 把言语视作为高度理论化的实践, 强调语言和言语之间存在着相互驱动的关系, 主张实地语言学与室内语言学合二为一。Fillmore (1992) 认为, 沙龙中的语言学家与实地语言学家应该取长补短, 语言研究中应该把语言学家的直觉和经验结合起来。McEnery & Wilson (1996) 也认为, 基于语料库的研究与基于直觉的研究两者之间并不排斥, 有着较强的互补性。

在国内语言学界, 徐烈炯 (1997) 强调我们应该正视和解决语言学理论与语言事实间的张力问题, 杜绝“对明明同理论相悖的反例听而不闻、视而不见”的“唯理论”做法, 同时反对“对理论不屑一顾, 认为搞理论是雕虫小技, 唯有拿出事实才过硬”的“唯实论”的观点, 主张语言学理论与语言事实相互结合。褚孝泉 (2003) 认为从经验到理论是一个互动复杂的过程, 理论观念从一开始就是所采用的经验材料的一部分, 反对在语言学领域把经验和理论割离开来。许余龙 (2000) 强调大型语料库在语言理论和语言事实发掘中的作用, 主张各语言学流派、各种研究方法之间互相借鉴, 甚至互相结合, 以促进语言学的发展。

综上所述, 国内外的语言学家已经充分认识到理性主义和经验主义的对立十分不利于语言学的发展, 因此两者间的关系应该从过去的对立转化为相互融合, 只有这样才能更有力地推动我们对语言的科学认识。没有数据支持的语言学研究显然是有问题的 (Čermák 2003), 也必将成为过去。由于理性主义方法与经验主义方法并非水火不容 (Klavans & Resnik 1996), 两者间的相互结合预示着语言学研究的的发展趋势。

4. 语料库语言学展望

由于大型语料库包含了大量的语言事实, 无论是从事理论研究的理论语言学家, 还是从事实证研究的经验语言学家, 语料库都将成为最重要的数据源。语料库语言学将更好地综合理性主义和经验主义, 扬长避短。在克服理性主义和经验主义各自缺点的同时, 两者的长处也将得到长足的发展。理性主义和经验主义的方法将在语言学领域更多、更深入地相互渗透, 并对语言学的众多方面产生很大的影响。具体表现在:

1) 语料库语言学的发展将迫使我们重新认识语言学的一些基本问题, 特别是语言与言语之间的关系问题。

结构主义语言学把语言与言语之间的关系割离开来, 乔姆斯基语言能力和语言行为概念的提出更加深化了语言与言语之间关系的断裂 (桂诗春 2004)。Aitchison 认为把语言能力和语言运用截然分开难以接受, 认为“心理语言学家对语言运用和语言知识同样感兴趣; 两者密切相关, 任何人只注意其中一个因素, 而忽略另一个因素, 是很奇怪的” (转引自桂诗春 2004)。

如今, 随着语料库语言学的发展, 语言和言语间的对立将不断趋于淡化, 人们更加清楚地认识到语言和言语其实是一个连续统。随着理性主义地位的削弱和语料库语言学方法的发展, 取而代之的将是言语事实和从言语事实中归纳出来的语言理论。

2) 大多语言学研究将在不同程度上依赖语料库。

随着语料库重要地位的确立, 大多语言学研究将在不同程度上依赖语料库。从对语料库的依赖程度上看, 研究方法大致可以归为三类, 即语料库驱动的方法 (corpus-driven approach)、基于语料库的方法 (corpus-based approach) 和语料库支持的方法 (corpus-supported approach) (参见 Tognini-Bonelli 2002)。

语料库驱动的方法更多地受传统的经验主义思想的影响, 主张理论来源于语料库, 研究者在研究之前应该没有任何理论预设。基于语料库的方法主张研究者 (依据直觉和理论) 提出自己的假设, 然后利用语料库对假设进行验证或修正。以上两种方法是语料库语言学领域的主流方法, 由此所产生的研究成果很多。

诚然, 由于语料库中的数据是产出性数据, 依靠这种数据很难对语言产生和语言运用的过程加以直接分析 (如心理语言学研究)。一些研究者主张采用三角验证 (triangulation) 的

方法,一方面从语料库中获取数据,同时从其他途径(如心理实验、大规模问卷等)获取实验数据,两种数据共同指向相同的结论。这种语料库支持的方法依靠语料库数据和其他类型的的数据共同回答研究问题。例如, Gilquin & Th. Gries (2009)专门撰文,在对多项研究进行综述的基础上,强调应该把语料库语言学研究方法与心理语言学等实验方法相结合; Ellis & Simpson-Vlach (2009)和 Wulff (2009)以语料库语言学和心理语言学相结合的方法分析学术英语中的熟语性(idiomacity)。

3) 自然语言处理技术发生变化。

自然语言处理技术中将不断地结合理性主义和经验主义的理念,以达到更好的处理效果。

仅以词性赋码器(POS tagger)为例。根据运行原理的不同,词性赋码器常可分为基于规则的(rule-based)和基于统计的(statistically-based)两大主要类型(Jurafsky & Martin 2000)。基于规则的分析方法来源于转换生成语法对语言生成规则的认识,这种方法通过人工编写的一系列规则来确定文本中词汇的语法属性;基于统计的分析方法来源于经验主义的思想,首先对部分训练语料进行人工词性赋码,然后统计各词汇在特定语法环境中出现的概率。在进行自动赋码的过程中,赋码工具将首先分析词汇的语法环境,根据语法环境的不同,为相应词汇选择概率最大的词性赋码。

在当代语料库语言学中,人们已经开始把来源于理性主义的基于规则的分析方法与来源于经验主义的基于统计的分析方法结合起来(Čermák 2003)。比如, Brill 词性赋码器本是一个基于转换规则的赋码器(Brill 1992),如今也融入了基于概率分析词性的方法(Brill & Wu 1998)。黄国营(1999)也认为,计算机自然语言处理系统从“基于规则”发展到“基于统计”,再发展到“基于规则”和“基于统计”相结合,实现了理性主义和经验主义的融合。这种融合还在不断加深,预示着自然语言处理工具的发展方向。Brill & Mooney (1997)也认为,语言学界人们越发注重把经验主义的方法和理性主义的独到之处相互结合起来,以达到更高的准确率。黄昌宁和张小凤(2002)也指出,复杂特征集和合一语法、词汇主义语法等理性主义框架下的自然语言处理方法与经验主义框架下的语料库和统计方法的结合,“是经验主义方法和理性主义方法相互融合的可喜开端”。

4) 定性研究和定量研究相结合的研究方法成为主流的研究方法。

传统的理性主义不主张通过分析语言事实的方法来研究语言,提倡对省内语料进行深度分析,而传统的经验主义则强调依靠定量方法从观察的数据中归纳语言理论。

语言学家需要考虑的是语料库的深度(即语言描写的精度)和广度(即能够描写多少语言素材),并依此进行定性研究和定量研究(Beaugrande 2002)。定性研究可以帮助人们更加深刻地认识语言问题,而定量研究可以让人们更加直观而广泛地认识语言问题。因此,定性研究方法和定量研究方法在揭示问题的深度和广度上各有所长,前者从深度挖掘语言理论,而后者则从广度对语言诸方面加以描写。定量研究以大量的数据表明语言发展的趋势,而定性研究则对定量方法所揭示的语言发展趋势作深度分析,两种研究方法相辅相成。在以往的语言学研究中,语言学的部分分支学科,如语义学、心理语言学、语用学、认知语言学等,主要甚至全部采用定性分析的方法。近年来的语言研究中,定性研究和定量研究的相互结合的趋势日趋明显。在未来的语言学研究中,理性主义与经验主义的融合将使得定性研究方法和定量研究方法在语言学领域的众多分支学科的研究中同时得到广泛的应用(潘永樑 2001)。定性分析和定量分析相结合是理性主义与经验主义相互融合的必然产物。

5) 对口语和网络语言的研究将得到很大的加强。

人们逐渐意识到,以往的语言学研究素材主要是书面语,而书面语并非人们的主要交际手段。口语是第一位的,它产生于书面语之前,而且世界上的许多语言至今只有口语形式,这充分表明了口语在语言研究中应有的地位。在将来的语言学研究中,将更多地收集口语语料,并对口语进行转写后进行多层级的标注并加以分析。目前,大部分大型语料库(如英国

国家语料库、英语文库、美国英语语料库等)中已经包含了口语语料,但由于口语语料的收集远远难于书面语料,大部分语料库中口语所占的比重远远小于书面语,这一问题将在未来的语言学研究中逐渐克服。

由于互联网是人类使用语言的重要环境之一,且从网络上可以快捷地收集到大量的语言素材,建立网络语言语料库、研究网络语言的特点已经成为语言学研究的热点之一(参见 Cabrera 等 2009; Fletcher 2001; Renouf 2003)。

鉴于以上种种变化, Lager (1995)对理论语言学家和语料库语言学家所做的种种区分已经不复存在,当代语言学研究所体现出来的是理性主义和经验主义之间更多的融合。

4. 结论

二十世纪的语言学发展史是理性主义与经验主义斗争的历史,斗争的结果是两种哲学观的融合,语料库语言学将发展成为新时期语言学研究中的主流方法。语料库语言学的发展将极大地推动人们对语言的认识,对语言学研究方法和研究视角将产生深远的影响。

参考文献

- [1] Aarts, J. Intuition-based and observation-based grammars. [A] K. Aijmer & B. Altenberg. (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik* [C]. London: Longman, 1992.
- [2] Alien, J. & M. S. Seidenberg. The emergence of grammaticality in connectionist networks [A]. In B. MacWhinney (ed.). *The Emergence of Language* [C]. NJ: Lawrence Erlbaum Associates, Inc., 1999: 115-152.
- [3] Barnden, J., M. Lee, and K. Markert. (eds.). *Proceedings of the workshop on corpus-based and processing approaches to figurative language* [C]. University Centre for Computer Corpus Research on Language technical papers; Vol. 14, special issue. Lancaster: UCREL, 2001.
- [4] Beaugrande, R. Descriptive Linguistics at the Millennium: Corpus Data as Authentic Language [J]. *Journal of Language and Linguistics*, 2002.1 (2): 91-131.
- [5] Brill, E. A simple rule-based part-of-speech tagger [A]. *Proceedings of the 3rd Conference on Applied Natural Language Processing* [C]. Trento, Italy, 1992: 152-155.
- [6] Brill, E. and R. Mooney. An overview of empirical natural language processing - Natural Language Processing [J]. *AI Magazine*, 1997,18(4): 13-24.
- [7] Brill, E. and J. Wu. Classifier combination for improved lexical disambiguation [A]. *Proceedings of the 17th international conference on Computational linguistics, Vol 1* [C]. Association for Computational Linguistics, 1998: 191-195.
- [8] Bunge, M. Quality, Quantity, Pseudoquantity and Measurement in Social Science [J]. *Journal of Quantitative Linguistics*, 1995,2 (1): 1-10.
- [9] Cabrera, R. G., M. Montes, P. Rosso & nor-Pineda L. Villase. Using the Web as corpus for self-training text categorization [J]. *Information Retrieval Archive*, 2009, 12 (3): 400 - 415
- [10] Čermák, F. Today's Corpus Linguistics: Some Open Questions [J]. *International Journal of Corpus Linguistics*, 2003, 7 (2): 265-282.
- [11] Chomsky, N. *Aspects of the theory of syntax* [M]. Cambridge: MIT,1965.
- [12] Ellis, N. & R. Simpson-Vlach. Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education [J]. *Corpus Linguistics and Linguistic Theory*, 2009. 5 (1): 61-78.
- [13] Fillmore, C. J. "Corpus linguistics" or "Computer-aided armchair linguistics" [A]. In J. Svartvik (ed.). *Directions in Corpus Linguistics* [C]. Berlin: Mouton de Gruyter, 1992.
- [14] Fletcher, W. Concordancing the Web with KwiCFinder [A]. In *Proceedings of The American Association for Applied Corpus Linguistics Third North American Symposium on Corpus Linguistics and Language Teaching* [C], 2001. (Available at <http://www.kwicfinder.com>)
- [15] Gilquin, G. & Th. Gries, S. Corpora and experimental methods: A state-of-the-art review1 [J]. *Corpus Linguistics and Linguistic Theory*, 2009. 5 (1): 1-26.
- [16] Grishman, R. *Computational Linguistics: an introduction* [M]. Cambridge: Cambridge University Press. 1986.
- [17] Halliday, M. A. K. Quantitative Studies and Probabilities in Grammar [A]. In M. Hoey (ed.) *Data, Description, Discourse: Papers on English Language in Honour of John Sinclair (on his sixtieth birthday)* [C]. London: Harper Collins,1993.
- [18] Johansson, S. Comments on a paper by Staffan Hellberg [A]. In J. Svartvik (ed.). *Directions in Corpus Linguistics* [C]. Berlin: Mouton de Gruyter,1992.
- [19] Jurafsky, D. and J. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition* [M]. NJ: Prentice Hall. 2000.
- [20] Klavans, J. & P. Resnik. *The Balancing Act* [M]. Cambridge : MIT Press. 1996.

- [21] Lager, T. *A logical approach to computational corpus linguistics* [D]. Gothenburg Monograph in Linguistics 14. Unpublished Doctoral Dissertation, University of Gothenburg, 1995.
- [22] Leech, G. The state of the art in corpus linguistics [A]. In K. Aijmer & B. Altenberg. (eds.) *English corpus linguistics: studies in honour of Jan Svartvik* [C]. London: Longman, 1991.
- [23] Leech G. Corpora and theories of linguistic performance [A]. In J. Svartvik (ed.). *Directions in Corpus Linguistics* [C]. Berlin: Mouton de Gruyter, 1992.
- [24] McEnery, T. & A. Wilson. *Corpus linguistics* [M]. Edinburgh: Edinburgh University Press. 1996.
- [25] Renouf, A. 'WebCorp: providing a renewable data source for corpus linguists' [A]. In S. Granger & S. Petch-Tyson (eds.). *Extending the scope of corpus-based research: new applications, new challenges* [C]. Amsterdam: Rodopi. 2003.
- [26] Rundell, M. The corpus of the future and the future of the corpus [OL]. Paper delivered at special conference on "New Trends in Reference Science", University of Exeter, 1996. (<http://www.ruf.rice.edu/~barlow/futcrp.html>)
- [27] Saussure, F. de. *Course in general linguistics* [M]. Translated by Wade Baskin. New York: McGraw-Hill, 1966.
- [28] Svartvik, J. Corpora are becoming mainstream [A]. In J. Thomas. & M. Short (eds.). *Using Corpora for Language Research* [C]. London: Longman, 1996.
- [29] Teubert, W. My version of corpus linguistics [J]. *International Journal of Corpus Linguistics*, 2005, 10 (1): 1-13.
- [30] Tognini-Bonelli, E. *Corpus Linguistics at Work* [M]. Amsterdam/Philadelphia: John Benjamins, 2002.
- [31] Wulff, S. Converging evidence from corpus and experimental data to capture idiomaticity [J]. *Corpus Linguistics and Linguistic Theory*. Volume, 2009, 5 (1): 131-159.
- [32] 陈勇, 论经验主义和理性主义之争——关于西方语言学研究中的认识论 [J]. 外语学刊, 2003 (3): 57-62.
- [33] 褚孝泉, 试论语言科学的经验性基础 [J]. 学海, 2003(6): 45-55.
- [34] 黄昌宁, 中文信息处理的主流技术是什么? [J]. 计算机世界报, 2002(24): 7-8.
- [35] 黄昌宁、张小凤, 自然语言处理技术的三个里程碑 [J]. 外语教学与研究, 2002 (3): 180-187.
- [36] 黄国营, 二十世纪理论语言学的起始和终结 [J]. 清华大学学报 (哲学社会科学版), 1999 (1): 83-85.
- [37] 桂诗春, 以概率为基础的语言研究 [J]. 外语教学与研究, 2004(1): 3-9.
- [38] 李葆嘉, 语言学的渊源、流派及其学科性质的变迁 [J]. 江苏社会科学, 2000(5): 126-131.
- [39] 罗宾斯, 简明语言学史[M], 许德宝等译. 北京: 中国社会科学出版社, 1997, .
- [40] 潘永樑, 语料库语言学的目的和方法 [J]. 解放军外国语学院学报, 2001(2): 1-5.
- [41] 宋春阳, 科学理性时代的语言学研究——理性主义的发展与语言学研究 [J]. 北方论丛, 2003(2): 89-92.
- [42] 徐烈炯, 语言学理论与语言事实 [J]. 现代外语, 1997(3): 27-32.
- [43] 许余龙, 也谈语言学理论与语言事实 [J]. 外国语, 2000(3): 2-9
- [44] 许余龙, 对比语言学发展新趋势、新思考——第五届世界对比语言学大学评述 [J], 外语教学与研究, 2009 (4): 279-283.

Rationalism, Empiricism and Corpus Linguistics

LIANG Mao-cheng

(National Research Center for Foreign Language Education, Beijing Foreign Studies University, Beijing 100089, China)

Abstract: The history of linguistics in the 20th century witnessed the opposition and convergence of Rationalism and Empiricism. While Rationalists draw on their introspection to interpret linguistic phenomenon, Empiricists advocate that language be described and theories be formulated on the basis of performance data drawn from the language in actual use. Today, Rationalism and Empiricism in linguistics are converging to an ever more intense extent, and research methodology in linguistics is inevitably to be dominated by a mixed approach combining empirical methods and linguists' introspection. Language corpora have become testbeds for linguists' intuitions and important resources to improve theories with. This tendency will lead to a series of changes in linguistics.

Key words: Rationalism, Empiricism, corpus linguistics

收稿日期: 2009-12-16

作者简介: 梁茂成, 北京外国语大学中国外语教育研究中心研究员, 博士生导师; 主要研究领域为应用语言学、语料库语言学。Email: frankliang0086@yahoo.com.cn