

# 语料库语言学研究的两种范式： 渊源、分歧及前景<sup>\*</sup>

北京外国语大学 梁茂成

**提要：**语料库语言学领域的两种研究范式之争由来已久。“语料库驱动”的研究范式以“激进的经验主义”为哲学基础，以 Firth 的语境论为语言学基础；而“基于语料库”的研究范式以“温和的经验主义”为哲学基础，发展过程中受到了美国结构主义语言学理论的重要影响。哲学基础和语言学理论基础的不同导致了研究范式的差异。本文系统梳理语料库语言学两种研究范式的渊源及其基本思想的形成原因，分析两种研究范式的主要分歧，并展望语料库语言学的发展前景。

**关键词：**语料库、语料库驱动、基于语料库、研究范式

[中图分类号] H0-06 [文献标识码] A [文章编号] 1000-0429(2012)03-0323-13

## 1. 引言

自从美国布朗大学 Henry Kučera 和 Nelson Francis 于 1964 年创建第一个电子语料库——布朗语料库(Brown Corpus)以来，语料库语言学已经获得了长足的发展，并就若干问题达成了共识。但与此同时，语料库语言学阵营内部分歧也日益明显，出现了基于语料库的(corpus-based)研究和语料库驱动的(corpus-driven)研究两种主要范式。

本研究系统梳理语料库语言学渊源，试图厘清两种研究范式的理论基础，进而分析两者的分歧与得失，并展望语料库语言学的发展前景。

---

<sup>\*</sup> 本研究是王克非教授主持的国家社科基金重大项目“大规模英汉平行语料库的建立与加工”(10ZD&127)和本文作者主持的国家社科基金项目“中国学生英语树库建设与研究”(09BY033)的部分研究成果。在本文写作过程中，卫乃兴教授和许家金博士曾提出宝贵意见，特此致谢。

## 2. 语料库语言学渊源之一:基于语料库的研究范式

### 2.1 布朗语料库和美国的语料库建设与研究

1950 年代是美国结构主义语言学的鼎盛时期,也是转换生成语法的萌芽时期。时任美国语言学会主席的布朗大学教授 W. F. Twaddell (1906-1982) 看到了语料库对语言研究的潜在价值,先后引进了 Henry Kučera (1925-2010) 和 Nelson Francis (1911-2002) 两位学者。Kučera 于 1952 年获哈佛大学博士学位,师从著名结构主义语言学家 Roman Jakobson (1896-1982)。Francis 也是一位美国结构主义语言学家,对方言调查有着浓厚的兴趣。在 Twaddell 的策划下,Francis 和 Kučera 于 1963 年着手建设布朗语料库,并于 1964 年完成了这个容量为一百万词的第一个电子语料库(Léon 2005)。

遗憾的是,布朗语料库在美国本土并未产生太大影响,主要原因是转换生成语法的兴起使得当时的美国失去了语料库语言学发展的土壤。由于语料库语言学与转换生成语法所主张的理性主义注重“语言能力”、排斥“语言行为”的思想背道而驰,因而很难得到认可。Leech & Johansson (2009)、Francis (1982) 曾对布朗语料库的建设者们在语料库建设过程中所遇到的来自转换生成语法的质疑做过描述,他们把布朗语料库视为“a waste of time”(Leech & Johansson 2009: 6)。可想而知,语料库语言学在美国很难得到发展。

然而,布朗语料库对欧洲的计算机语料库建设和语料库语言学的发展起到了重要的引领和催化作用,特别是其语料取样方法对欧洲一系列语料库的建设提供了重要的借鉴作用(Leech & Johansson 2009)。

布朗语料库建成之后,语料库语言学在美国沉寂了若干年,直至 1980 年代,以 Douglas Biber 为代表的一些语言学家另辟蹊径,采用 MF/MD 分析法,利用语料库进行语域变异(register variation)分析,在国际语料库语言学界产生了重要影响。北亚利桑那大学、密歇根大学等高校分别建立了语料库语言学团队,密歇根大学还创建了 MICASE、MICUSP 等语料库。

### 2.2 Randolph Quirk 和“英语用法调查”

英国是最具经验主义哲学传统的国家之一,Francis Bacon、John Locke、David Hume 等哲学家的经验主义思想影响着一代又一代英国人。伦敦大学是英国当代语言学的发源地,也是英国语料库语言学的发源地。20 世纪初期至上半叶,伦敦大学学院(University College London, UCL)和伦敦大学亚非学院(School of Oriental and African Studies, SOAS)聚集了 Daniel Jones、J. R. Firth、Randolph Quirk、Michael Halliday、Sidney Greenbaum、Jan Svartvik、Geoffrey Leech、David Crystal 等著名语言学家,其中 Randolph Quirk 无疑是

当时语料库语言学领域的领军人物。

Quirk 语言学思想的形成深受美国结构主义的影响。Quirk 曾获英联邦基金会资助,于 1951-1952 年间留学美国。据 Quirk(2002)的描述,访美期间,他的思想受到 Charles Fries 的很大影响。Charles Fries 虽是美国结构主义语言学的代表人物之一,他的研究方法 with 当今的语料库语言学研究方法却颇为相似。早在 1925 年,他曾自建一个专门的小型语料库,并依此进行研究(Marckwardt 1968: 206)。1940 年,Charles Fries 曾主持出版过一部与当今语料库语言学十分相关的著作,名为 *English Word Lists: A Study of Their Adaptability for Instruction*。此外,Peter Fries(2010)曾专门探讨 Charles Fries 的语言学思想与当今语料库语言学思想的相似之处,认为他是语料库语言学领域的先驱之一。Sampson & McCarthy(2004)收录了 Charles Fries 的“*From The Structure of English*”一文,并在导言中明确指出,Charles Fries 才是“真正开启现代语料库语言学研究传统的人物”(同上: 9)。

Quirk 主张研究方法的多元化。1970 年,Quirk 曾与 Sidney Greenbaum 合作出版 *Elicitation Experiments in English*, 系统介绍诱发数据的获取和分析方法。Léon(2010: 430)指出,Quirk 甚至不拒绝 Chomsky 的转换生成语法,只是反对 Chomsky 只注重语言能力,排斥语言行为的做法。

Quirk 于 1959 年开始“英语用法调查”(the Survey of English Usage, SEU)这一语料库建设项目,在项目建设过程中培养了一批著名学者,在很大程度上影响了语料库语言学的发展。学界常常将 1970 年代的 Quirk 团队称为“四人帮”,包括 Randolph Quirk、Sidney Greenbaum、Jan Svartvik 和 Geoffrey Leech,将 1980 年代增加了 David Crystal 的 Quirk 团队称为“五人帮”。

基于“英语用法调查”,UCL 的学者们开展了多项研究,其中最具有代表性的是 *A Grammar of Contemporary English* (1972)和 *A Comprehensive Grammar of the English Language* (1985)。这两部语法至少在两方面是以前的任何一部英语语法都不可比拟的(Hudson & Walmsley 2005; Mukherjee 2006): a) 是真正含义的描写语法(descriptive grammar),改变了规定语法(prescriptive grammar)的传统; b) 改变了只注重书面语(特别是文学语言)、忽视口语的语法传统。这两方面特征恰恰反映了美国结构主义语言学思想对 Quirk 的影响。

继“英语用法调查”之后,1990 年代,“英语用法调查”项目组建成了国际英语语料库中的英国英语部分(International Corpus of English - Great Britain, ICE-GB),并对语料库进行了细致的句法标注。ICE Corpus Utilities Program (ICE-CUP)的开发使得 ICE 的使用更为便捷,成为国际学界的常用语料库之一。此外,项目组还于近年完成了当代英语口语历时语料库(the Diachronic

Corpus of Present-Day Spoken English)。

### 2.3 Geoffrey Leech 和兰卡斯特大学的语料库语言学研究

Geoffrey Leech 早年就读并任职于 UCL。Quirk 对他影响很大,从 1962 到 1969 年间,“英语用法调查”一直是他当时研究工作的一个重要方面。1969 年他调入兰卡斯特大学之后,仍继续基于“英语用法调查”的语法研究。

兰卡斯特大学的语料库语言学团队在 Leech 的领导之下,承传并发展了 Quirk 的思想。Leech 调入建校不久的兰卡斯特大学之后,开始组建自己的语料库语言学团队。经过几十年的发展,该团队取得了卓越的成就:

首先,他们建成了一系列语料库,总容量约一亿零六百万词 (McEnery & Hardie 2012: 78)。其中特别值得一提的是,1991 年,在牛津大学出版社、朗文出版公司、英国国家图书馆、英国科学院等部门的倡导和支持下,设计容量为一亿词(包含英语口语和笔语)的英国国家语料库 (British National Corpus, BNC) 的建设工作正式启动,项目由 Leech 等人负责。1994 年语料库建成之后,他们又对整个语料库进行了词性标注。如今,BNC 已经成为语言研究领域最常用的大型语料库之一。

其次,在自然语言处理方面,他们也取得了骄人的成就,主要表现在自动标注技术方面。经过多年的努力,他们开发了 CLAWS 词性标注系统 (part-of-speech tagger)、句法标注系统 (syntactic parser)、USAS 语义标注系统 (semantic tagger)、韵律标注 (prosodic annotation) 系统、指代关系标注 (anaphoric annotation) 系统等多个语料库标注系统。

在语言研究方面,他们积极拓宽研究领域,利用语料库进行多种交叉学科的研究,涉及理论语言学、话语分析、语用学、文体学、语言教学、语言测试、社会语言学等多个学科领域。

### 2.4 布朗语料库团队、Quirk 团队、Leech 团队间的相互交流和影响

以上三个团队在语料库建设和研究过程中彼此交流、相互影响,逐渐形成了“基于语料库”的研究范式。

Quirk 留美期间,曾与布朗语料库的策划者 Twaddell 有过密切接触 (Quirk 2002),Léon (2005) 甚至认为建立布朗语料库是 Twaddell 和 Quirk 两个人的主意。在布朗语料库项目启动之前曾召开过一次小型研讨会,讨论语料库的建设方案,Quirk 应邀参加了会议,并且担任布朗语料库的项目组顾问 (Francis 1982; Leech 2002; Léon 2005)。不仅如此,“英语用法调查”项目建设期间,布朗语料库的项目负责人 Francis 曾于 1962 年担任 Quirk 项目组的成员,负责方言调查 (Kučera 2002; Léon 2005; Quirk 2002),而 Leech 更是项目组的重要成员。

此外,在兰卡斯特大学筹建语料库的过程中,Leech 也邀请 Francis 担任顾问。据 Leech(2002)描述,在 ICAME 的建设过程中,Francis 也曾给出很多绝好的主意,比如建议沿用布朗语料库的采样方法建设 LOB 语料库,这样可以方便语料库间的对比。在 ICAME 组织里,Leech 和 Francis 更是共同的策划者和组织者。

可以说,布朗语料库、“英语用法调查”及兰卡斯特大学所建设的大量语料库都是同根同源的。频繁的交流与合作增进了团队间的理解和共识,逐渐形成了“基于语料库”的研究范式。

### 3. 语料库语言学渊源之二:语料库驱动的模式

#### 3.1 Firth 和他的语境论

J. R. Firth (1890-1960)是伦敦学派<sup>1</sup>的创始人,于 1954-1957 年间担任英国语文学会(the Philological Society of Great Britain)会长,是英国历史上第一位语言学教授(1944 年)(Anderson 1985: 178; Plug 2008)。

文献(Chapman & Routledge 2005: 80-81; Langendoen 1968; Robins 1964: 27)中普遍认为,Firth 在语言学领域主要有两大贡献,即 1)提出“意义”是语言学研究中的重点,而“意义”是由“语境”决定的。这一思想是 Firth 的“语境论”(the theory of context of situation)的基础;2)在音系学领域的研究和“韵律分析”方法的提出。Palmer(1968: 1)认为 Firth 是英国语言学的唯一开拓者。

Firth 的“语境论”是在波兰人类学家 Malinowski 的理论基础上发展而来的。在弗斯看来,语言是一种社会行为,语言研究的主要目的就是分析意义(Chapman & Routledge 2005; Firth 1957)。意义由语境决定,“任何词使用在一个新的语境之中,就会成为一个新词”(each word when used in a new context is a new word)(Firth 1957: 190),因此研究意义需要从语境入手。Firth 反对割离语境研究语言,强调文本的整体性,提倡着眼于语境中的文本。

为了更好地说明意义的分析方法,Firth 提出了搭配关系(collocation)这一概念<sup>2</sup>。在“Modes of meaning”(1957)一文中,弗斯以 dark 和 night 为例,指出 night 的意义之一表现为它可以与 dark 搭配,同样,dark 的意义之一表现在它

<sup>1</sup> “伦敦学派”并不是按照地缘划分的一个学派,而专门指 Firth 创立的一个语言学流派,时间跨度从 1940 年代末至 1960 年 Firth 去世时为止,成员主要包括当时在伦敦大学亚非学院工作或学习的一批学者(Langendoen 1968)。Firth 去世后,“伦敦学派”的学者继续发展他的理论,语言学史上常称之为“新弗斯学派”(the neo-Firthian School)(Mitchell 1975)。

<sup>2</sup> 在 Firth 之前,Harold E. Palmer(1877-1949)等人曾对搭配进行过研究,但 Firth 提倡通过搭配关系来研究意义,这一思想对后来学者影响很大。

可以与 night 搭配。因此,搭配是一种组合层面(syntagmatic)上词项之间的相互关系,搭配词彼此双方相互期待。Firth 将这一种从搭配关系入手分析意义的方法称为“通过搭配研究语义”(meaning by collocation)。这种分析方法成为后来兴起的短语学(phraseology)研究的理论基础。

Firth 的作品并不多,但争议颇多(Dinneen 1995: 321)。有学者认为 Firth 的学术观点许多情况下表述得不很清楚,常常晦涩难懂(Anderson 1985: 178)。尽管学界对 Firth 的评价褒贬不一,但他对语言学的贡献却是不可否认的。Firth 的思想影响了英国几代人的语言学研究。他去世后,伦敦大学亚非学院的一大批下属和学生成了英国语言学界的重要人物。

### 3.2 “新弗斯学派”和新一代“新弗斯学派”

Firth 于 1960 年去世。次年, Halliday 发表“Categories of the theory of grammar”一文。Mathews(1999: 280)将 Halliday 的这篇论文称为“新弗斯学派”的开山之作。1966 年,为了纪念导师, Firth 的几个下属和学生共同出版了文集 *In Memory of J. R. Firth*, 文集中包括 Halliday 的“Lexis as a linguistic level”和 Sinclair 的“Beginning the study of lexis”。至此,以 Halliday 为核心人物的“新弗斯学派”(the neo-Firthian School)形成(Langendoen 1968; Robins 1969)。在 *In Memory of J. R. Firth* 这一文集中,语音问题仍然是研究重点(Robins 1969),以词语(lexis)为中心的思想得到了进一步发展。

Halliday 对 Sinclair 学术思想的形成产生过重要影响。据 Sinclair(2004b: vii)回忆, Halliday 一直是他的“良师益友”,从一开始就告诫他要“相信文本”(trust the text)。从 1958 到 1963 年的五年间, Halliday 在多方面对 Sinclair 给予指导和帮助,使他受用一生。期间, Halliday 极力鼓励 Sinclair 利用计算机研究语言中的搭配(Léon 2007: 411)。

作为“新弗斯学派”领军人物的 Halliday 后来移居美国和澳大利亚,提出并逐渐完善了“系统功能语法”体系,对“搭配”问题不再关注(Léon 2007: 411)。与此同时,以 Sinclair 为代表的一批学者却在坚定地延续着对词语和搭配的研究。他们不断发展 Firth 有关搭配的理论,并与 Collins 出版公司合作建立了大型语料库 Bank of English,试图基于大型语料库<sup>3</sup>来研究搭配。在这一过程中, Sinclair 与 Halliday 的学术取向也逐渐产生了分歧。Halliday (1992: 63)认为,“Sinclair 是一位词典学家,他的目标是从词典中构建语法;而我是一个语法学家,目标是从语法中构建词典”。Tucker(1998: 146)对 Halliday 和 Sinclair 的

<sup>3</sup> Tognini-Bonelli (2001: 157)认为, Firth 或许不会像 Sinclair 那样用语料库来发展他的思想。



研究方法差异进行了分析,他把 Sinclair 的方法称为 lexis-driven grammar,把 Halliday 的方法称为 grammar-driven lexis。在评价 Hunston & Francis(2000)和 Tognini-Bonelli(2001)时,Halliday(2005: 174)也毫不掩饰地指出,“语料库驱动的语法并未做到不依赖任何理论”。

基于以上事实,笔者认为,Sinclair 与 Halliday 学术取向的分歧标志着新一代“新弗斯学派”(the New neo-Firthian School)的形成。这一学派以 Sinclair 的思想为理论基础,采用“语料库驱动”的研究范式,主张从语料库中建构理论、以全新视角对语言进行描写;他们关注搭配,同时也关注意义,并且认为搭配和意义是密不可分的。新一代“新弗斯学派”以伯明翰大学为中心,后来又辐射到利兹、利物浦等校,但他们与 ICAME 等语料库语言学国际组织相对独立(McEney & Hardie 2012: 81)。

新一代“新弗斯学派”的研究极大地推动了语料库语言学的发展,伯明翰大学也因此成了世界上最重要的语料库语言学研究中心之一。这一学派丰富了 Firth 关于搭配的理论 and 思想,建成了迄今为止世界上最大的语料库,并开创性地将语料库信息和辞书编纂结合起来,与 Collins 出版公司联合推出了 CoBuild 系列词典。不仅如此,Michael Hoey 的词汇触发 (lexical priming) 理论、Susan Hunston 和 Gill Francis 的型式语法 (pattern grammar) 以及 Bill Louw 的语义韵 (semantic prosody) 学说更加强化了词语为中心的语言学思想,并将搭配、类联接的研究与意义结合起来。

## 4. 两种研究范式及其分歧

### 4.1 两种研究范式的提出及其哲学基础

2000 年,Sinclair 退休后,协助妻子 Tognini-Bonelli 在意大利经营非盈利性的培训机构 the Tuscan Word Centre,主要目的是培训语言教师和语言研究者使用语料库。在 Sinclair 的影响下,Tognini-Bonelli (2001: 99)首次对“基于语料库”和“语料库驱动”的研究范式进行了区分。在她看来,“基于语料库”的研究范式利用语料库对已有的理论或假设进行探索,目的在于验证或修正已有理论,而“语料库驱动”的研究范式则以语料库作为出发点和唯一观察对象,对语言中的各类现象进行全新的界定和描述。

不难看出,对已有语言分析体系应该抱何种态度,这是“基于语料库”和“语料库驱动”这两种研究范式间的根本差异。前者并不试图推翻已有的语言分析体系,研究中只是将语料库视为众多数据种类中的一种,不排斥在必要时使用其他类型的数据(如诱发数据,甚至内省数据),而后者主张一切源自语料库。从哲学角度看,前者属于“温和的经验主义”(moderate empiricism),而后者属于“激

进的经验主义”(radical empiricism)(参见 Gast 2006; Tognini-Bonelli 2001)。

以 Sinclair 为代表的新一代“新弗斯学派”信奉“激进的经验主义”,极力主张“相信文本”,主要集中在伯明翰大学、利兹大学、利物浦大学、香港理工大学等,而绝大多数语言研究者在学术取向上则更乐于信奉“温和的经验主义”,采用“基于语料库”的研究范式(Gast 2006: 115-116),将语料库语言学视为一种研究方法(如 Gries 2010; McEnery & Hardie 2012 等)。

#### 4.2 两种研究范式的分歧

由于渊源不同,信奉的哲学思想有别,语料库语言学的两种范式之间存在一些重要差异,这些差异在“训练营大讨论”(Bootcamp discussion)中得到了充分的体现。

2008年8月12日, *Corpus Linguistics and Linguistic Theory* 杂志主编、认知语言学家 Stefan Gries 在 Corpora List 论坛发帖,宣布他们即将开办一个名为 Quantitative Corpus Linguistics with R 的培训班。当日, Bill Louw 针锋相对地发帖,指责培训班课程中既无搭配研究,也不包含语义韵,这是课程设计的重大失误,劝大家谨慎报名。14日, Wolfgang Teubert 也加入讨论,并指出,在“我这样过时的语料库语言学者”(old-fashioned corpus linguists like myself)看来,过去的十多年里,“语料库语言学”这一术语被人“绑架了”(hijacked),暗示在认知语言学框架下谈论语料库,背离了语料库(驱动)语言学研究的准则(Teubert 2010: 356)。至此,“训练营大讨论”正式拉开帷幕,几十位学者先后加入,讨论涉及语料库语言学的学科性质、研究目的、研究领域、研究方法等问题,而且言辞十分激烈。后来, *International Journal of Corpus Linguistics* 在 2010 年第 3 期将部分影响较大的帖子收集整理并出版了专辑,以澄清语料库语言学中的一些重要问题和两种研究范式之间的主要差异。

其实,两种研究范式之间的差异早就存在,“训练营大讨论”只是把这些差异较为集中地展示了出来。笔者现综合所掌握的文献,对语料库语言学两种研究范式的主要分歧做以下归纳:

##### 1) 关于语料库语言学学科属性的分歧

“语料库驱动”的研究范式认为,语料库语言学是一门独立的学科,摆脱任何已有的语言分类体系和研究框架,从真实数据出发,对语言进行全新的描写(Tognini-Bonelli 2001: 99)。“基于语料库”的研究范式认为,语料库语言学并非一门独立的学科,而是一种研究方法,可以用来验证已有假说和理论(Gries 2010; McEnery & Hardie 2012; McEnery & Wilson 2001)。



## 2) 关于研究目的和分析步骤的分歧

“语料库驱动”的研究范式主要目的在于语言描写,主张从观察词语入手,反对利用语料库之外(corpus-external)的任何理论前提。研究中主要采用“词项-环境法”(Item-Environment method)(Francis 1993: 146,转引自卫乃兴 2011: 14),即利用索引技术(KWIC),对词项及其左右语境进行逐个分析和归纳。分析中所涉及的主要方面包括节点词、搭配、类联接、语义倾向和语义韵等(详见 Sinclair 1996b)。对索引行的细致解读和分类从来就是“语料库驱动”研究范式中的最大重点。

“基于语料库”的研究范式不排斥外部理论,目的在于验证已有假设。在这种研究范式中,词语不是研究的唯一切入点,研究可以涉及性别、语域等多种非语言因素。因此,这种研究范式与语用学、社会语言学、批评话语分析、认知语言学甚至心理语言学都有接口。“基于语料库”的研究范式采用普通实证研究的步骤,即“提出假设-分析数据-验证假设”的步骤。

## 3) 关于语料库标注的分歧

“语料库驱动”的研究范式主张“干净文本原则”(clean text policy)(Sinclair 1991: 21-22),反对采用传统的语言分析体系对文本进行词性标注和句法标注(Sinclair 2004a),而“基于语料库”的研究范式认为语料库的标注可以为语料库带来“增值”,方便从语料库中提取各种语言信息。

## 4) 关于语料库建设中文本取样方法的分歧

话语分析是 Sinclair 关注的重要问题之一。在他看来,任何文本都有完整的内部结构,文本中的所有句子通过衔接和连贯等手段构成一个整体(Sinclair 2005)。Sinclair (1996a: 9)明确指出,将文本切割成大小相同的片段,会破坏文本的整体性。

相反,在“基于语料库”的研究范式中,为了尽可能保证语料库的代表性,语料库中文本的长短大致相等。McEnery & Hardie(2012: 152)认为,Sinclair 的说法确有道理,但整体文本的版权问题很难解决,而且如今的语言分析技术(如索引行)只能观察文本的局部,难以观察文本的整体。

## 5) 关于语料库分析中统计学方法的差异

由于“语料库驱动”的研究范式主要目的在于语言描写,常常不需要对比,主要采用“词项-环境法”,分析技术主要依赖语料库语言学中的索引技术。利用软件工具析出索引行后,通过人工观察、归类的方法对各种语言现象进行概括,最终形成结论。由于这种方法主要靠人工完成,常常不需要任何统计学检验,McEnery & Hardie(同上)称之为“手眼并用技术”(hand and eye techniques)。虽有不少学者主张通过计算 MI 的方法来确定搭配关系的强弱,通过卡方检验

或对数似然比的方法来确定频数差异,但大部分研究中所使用的都是描述统计学(descriptive statistics),包括频数、百分比等。

相反,“基于语料库”的研究范式沿用了普通实证研究的方法,即提出假设、以数据验证假设。假设的验证常常需要对照组,比较和统计检验是研究中不可缺少的环节。在验证假设的过程中,所使用的大部分统计分析方法是推断统计学(inferential statistics)。

#### 6) 对非语料库语言学研究方法接受程度的分歧

“语料库驱动”的研究范式主张“相信文本”,视语料库为语言研究的唯一数据源,反对使用其他形式的数据。相反,“基于语料库”的研究范式并不反对使用语料库之外其他形式的数据(如通过各种诱发手段获取的数据)。随着近年来语料库在各类研究中的广泛运用,越来越多的研究者主张综合多种研究方法的结果,进行“三角验证”(triangulation)。

### 5. 语料库语言学的发展趋势

毋庸置疑,“语料库驱动”的研究范式在语言学界产生了重要影响,在语言描写、辞书编纂、搭配分析等方面,提出了一些重要理论,也产出了一些重要成果。但这种研究范式具有一定的局限性:

1) 这种研究范式以“激进的经验主义”为哲学基础,完全采用自下而上的方法,否定前人的所有研究,但在观察索引行、对语言现象进行分析和归纳时,又难以摆脱现有知识体系的影响。因此,许多学者(如 Gries 2010; Xiao 2009)认为真正含义的“语料库驱动”是不可能的,是“神话”。

2) 这种研究范式主张一切源自语料库,而有些语言现象是语料库中无法获取的(如语言习得的心理过程),因而研究具有不可避免的局限性。“语料库驱动”的研究范式充其量也只能解决部分语言相关问题。

3) 这种研究范式完全排斥诱发数据等非语料库数据,而通过诱发手段获取数据已经成为研究中十分常见的方法。“语料库驱动”研究范式的排他性很难被各领域的研究者所接纳。

4) 这种研究范式反对语料库标注,而从生语料库中所能获取的语言信息十分有限。所谓的“词项-环境法”依赖人工甄别,更限制了研究的规模和信度。

诚然,“语料库驱动”的研究范式在辞书编纂和词语描写方面有着无可比拟的优势,加之人工解读索引行对技术要求不高,在较长一段时间内,这种研究范式仍然可能会受到一些研究者的追捧,也必将在某些领域取得更大的研究成果。但由于以上种种局限, Sinclair 去世之后,“语料库驱动”研究范式的许多特点正在逐渐淡化。例如,就连 Sinclair 和 Tognini-Bonelli 创办的 the Tuscan Word

Centre 也将词性标注和句法标注列为其培训课程的内容之一。随着各领域内越来越多的语言研究者意识到语料库的优势,在自己的研究领域(如心理语言学、社会语言学等)中不断使用各种类型的语料库,“基于语料库”的研究范式在不知不觉中得到了推广。或许,若干年之后,语料库语言学领域的研究范式之争会不复存在,语料库语言学也将大踏步地走出英国,走向全世界。

## 6. 结语

语料库语言学研究中的两种范式之争从语料库语言学诞生之日起就不同程度地存在。其中,“语料库驱动”的研究范式以“激进的经验主义”为哲学基础,以 Firth 的语境论为语言学理论基础,而“基于语料库”的研究范式以“温和的经验主义”为哲学基础,发展过程中受到了美国结构主义语言学理论的重要影响。哲学基础和语言学理论的不同导致了两种研究范式之间的种种差异。随着语料库技术的不断进步,语言学众多领域的学者逐渐地意识到语料库的价值,在自己领域的研究框架下越来越多地使用语料库来丰富研究手段,进行“三角验证”。这种趋势的持续会使得“基于语料库”的研究范式逐渐成为语料库语言学研究的主流范式和默认范式,语料库语言学方法论的特点会得到不断的加强。

### 参考文献

- Anderson, S. 1985. *Phonology in the Twentieth Century* [M]. Chicago: The University of Chicago Press.
- Bazell, C., J. Catford, M. Halliday & R. Robins (eds.). 1966. *In Memory of J. R. Firth* [C]. London: Longman.
- Brown, K. & V. Law (eds.). 2002. *Linguistics in Britain: Personal Histories* [C]. Oxford: Blackwell.
- Chapman, S. & P. Routledge (eds.). 2005. *Key Thinkers in Linguistics and the Philosophy of Language* [C]. Edinburgh: Edinburgh University Press.
- Dinneen, F. 1995. *General Linguistics* [M]. Washington, D. C.: Georgetown University Press.
- Firth, J. 1957. Modes of meaning [A]. In J. Firth (ed.). *Papers in Linguistics 1934-51* [C]. Oxford: OUP. 190-215.
- Francis, N. 1982. Problems of assembling and computerizing large corpora [A]. In S. Johansson (ed.). *Computer Corpora in English Language Research* [C]. Bergen: Norwegian Computing Centre for the Humanities. 7-24.
- Fries, P. 2010. Charles C. Fries, linguistics and corpus linguistics [J]. *ICAME Journal* 34: 89-121.
- Fries, C. & A. Traver. 1940. *English Word Lists: A Study of their Adaptability for Instruction* [M]. Ann Arbor, MI.: George Wahr.
- Gast, V. 2006. *The Scope and Limits of Corpus Linguistics* [M]. Tübingen: Königshausen &

- Neumann.
- Fries, C. 2004. From *The Structure of English* [A]. In G. Sampson & D. McCarthy (eds.). 2004. 9-26.
- Gries, S. 2010. Corpus linguistics and theoretical linguistics; A love-hate relationship? Not necessarily... [J]. *International Journal of Corpus Linguistics* 15; 327-343.
- Halliday, M. 1961. Categories of the theory of grammar [J]. *Word* 17; 241-292.
- Halliday, M. 1966. Lexis as a linguistic level [C]. In Bazell *et al.* (eds.). 1966. 148-162.
- Halliday, M. 1992. Language as system and language as instance: The corpus as a theoretical construct [A]. In J. Svartvik (ed.). *Directions in Corpus Linguistics* [C]. Stockholm: Walter de Gruyter. 61-77.
- Halliday, M. 2005. *Computational and Quantitative Studies* [M]. London: Continuum.
- Hudson, R. & J. Walmsley. 2005. The English Patient: English grammar and teaching in the twentieth century [J]. *Journal of Linguistics* 43; 593-622.
- Hunston, S. 2002. *Corpora in Applied Linguistics* [M]. Cambridge: CUP.
- Hunston, S. & G. Francis. 2000. *Pattern Grammar* [M]. Amsterdam: John Benjamins.
- Kučera, H. 2002. Obituary for W. Nelson Francis [J]. *Journal of English Linguistics* 30; 306-309.
- Langendoen, D. 1968. *The London School of Linguistics* [M]. Cambridge, MA.: The MIT Press.
- Leech, G. 1991. The state of the art in corpus linguistics [A]. In K. Aijmer & B. Altenberg (eds.). *English Corpus Linguistics* [C]. London: Longman. 8-29.
- Leech, G. 2002. Geoffrey Leech [C]. In K. Brown & V. Law (eds.). 2002. 155-169.
- Leech, G. & S. Johansson. 2009. The coming of ICAME [J]. *ICAME Journal* 33; 5-20.
- Léon, J. 2005. Claimed and unclaimed sources of corpus linguistics [J]. *Henry Sweet Society Bulletin* 44; 34-48.
- Léon, J. 2007. Meaning by collocation. The Firthian filiation of Corpus Linguistics [A]. In D. Kibbee (ed.). *Proceedings of ICHoLS X* [C]. Amsterdam: John Benjamins. 404-415.
- Léon, J. 2010. British empiricism and Transformational Grammar: A current debate [A]. In D. Kibbee (ed.). *Chomskyan (r)evolutions* [C]. Amsterdam: John Benjamins. 423-444.
- Marckwardt, A. 1968. Charles C. Fries [J]. *Language* 44; 205-210.
- Mathews, P. 1999. Linguistic theory and the philological society from the 1930s to the 1970s [J]. *Transactions of the Philological Society* 97; 259-285.
- McEnery, T. & A. Hardie. 2012. *Corpus Linguistics; Method, Theory and Practice* [M]. Cambridge: CUP.
- McEnery, T. & A. Wilson. 2001. *Corpus Linguistics* [M]. Edinburgh: Edinburgh University Press.
- Mitchell, T. 1975. *Principles of Firthian Linguistics* [M]. London: Longman.
- Mukherjee, J. 2006. Corpus linguistics and English reference grammars [A]. In A. Renouf & A. Kehoe (eds.). *The Changing Face of Corpus Linguistics* [C]. Amsterdam: Rodopi. 337-354.

- Palmer, F. 1968. (ed.). *Papers of J. R. Firth, 1952-1959* [C]. Indiana: Indiana University Press.
- Plug, L. 2008. J. R. Firth: A new biography [J]. *Transactions of the Philological Society* 106; 337-374.
- Quirk, R, S. Greenbaum, G. Leech & J. Svartvik. 1985. *A Comprehensive Grammar of the English Language* [M]. London: Longman.
- Quirk, R, S. Greenbaum, G. Leech & J. Svartvik. 1972. *A Grammar of Contemporary English* [M]. London: Longman.
- Robins, R. 1964. *General Linguistics: An Introductory Survey* [M]. London: Longman.
- Robins, R. 1969. Review of *The London School of Linguistics* by D. Terence Langendoen [J]. *Language* 45(1): 109-116.
- Sampson, G. & D. McCarthy (eds.). 2004. *Corpus Linguistics: Readings in a Widening Discipline* [C]. London: Continuum.
- Sinclair, J. 1966. Beginning the study of lexis [C]. In Bazell *et al.* (eds.). 1966. 410-430.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation* [M]. Oxford: OUP.
- Sinclair, J. 1996a. Preliminary recommendations on Corpus Typology [OL]. <http://www.ilc.cnr.it/EAGLES/pub/eagles/corpora/corpusstyp.ps.gz> (accessed 13/03/2012).
- Sinclair, J. 1996b. The search for units of meaning [J]. *Textus* 9; 75-106.
- Sinclair, J. 2004a. Intuition and annotation: The discussion continues [A]. In K. Aijmer & B. Altenberg (eds.). *Advances in Corpus Linguistics* [C]. Amsterdam: Rodopi. 39-59.
- Sinclair, J. 2004b. *Trust the Text* [M]. London: Routledge.
- Sinclair, J. 2005. Corpus and text: Basic principles [A]. In M. Wynne (ed.). *Developing Linguistic Corpora* [C]. Oxford: Oxbow Books. 1-16.
- Teubert, W. 2010. Our brave new world? [J]. *International Journal of Corpus Linguistics* 15; 354-358.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work* [M]. Amsterdam: John Benjamins.
- Tucker, G. 1998. *The Lexicogrammar of Adjectives* [M]. London: Continuum.
- Xiao, R. 2009. Theory-driven corpus research: Using corpora to inform aspect theory [A]. In A. Lüdeling & M. Kytö (eds.). *Corpus Linguistics: An International Handbook* [C]. Berlin: Mouton de Gruyter. 987-1008.
- 卫乃兴, 2011, 《词语学要义》[M]. 上海: 上海外语教育出版社.

收稿日期: 2012-02-28

通讯地址: 100089 北京市 北京外国语大学中国外语教育研究中心