

基于语料库的历时语言研究述评^{*}

北京外国语大学 许家金

提要:本文旨在梳理近年兴起于各语言学分支领域的历时研究动向。语料库理念和方法是这一研究趋势的重要推动力。得益于众多千年和百年历时语料库,以及多元统计和可视化方法,词汇、短语、句法、话语、语用、认知等诸多语言维度的长期演变和短期渐变成果叠出。语料库语言学的概率方法,以及语言特征和语境因素共现/共变的理论思维,使得历时语言变化研究在描写和阐释方面均有建树。

关键词:语言演变、语言渐变、历时语言学、多因素分析、语料库

[中图分类号]H0-09 [文献标识码]A [文章编号]1000-0429(2020)02-0200-13

1. 引言

近年,语言的历时研究呈激增趋势,出现了诸如“历时/历史语料库语用学”(Diachronic/Historical Corpus Pragmatics, Kohnen 2009; Taavitsainen *et al.* 2014)、“历史认知语言学”(Historical Cognitive Linguistics, Winters *et al.* 2010)、“(英语)历史语用学”((English) Historical Pragmatics, Jucker & Taavitsainen 2010, 2013)、“历史社会语言学”(Historical Sociolinguistics, Nevalainen & Raumolin-Brunberg 2014)、“历时构式语法”(Diachronic Construction Grammar, Barðdal *et al.* 2015)、“量化历史语言学”(Quantitative Historical Linguistics, Jensen & McGillivray 2017)等新兴子学科,且有扩展和蔓延之势。

历时语言研究属于用法本位语言学(usage-based linguistics),因此重视真实语料的运用。历时研究发展迅猛,很大程度上得益于不断建成的历时电子语料库。在语言学文献中,历时语言研究一般称为“历史语言学”(Historical

^{*} 本文系教育部人文社会科学重点研究基地重大项目子课题“大数据视野下的外语及外语学习研究”(17JJD740003)的阶段性成果。

Linguistics)或“历时语言学”(Diachronic Linguistics),其焦点是语言演变机制(Campbell 1999: xiv)。相关研究包括词汇演变历程的考据(即词源学)、语言家族谱系的构拟(即对比语文学)、实词虚化为功能词机制的探究(即语法化¹研究)等。此外,历时语言研究与社会语言学交集甚广。变异社会语言学中的历时变异,方言学中的古语用法,以及从语言接触视角解读当今语言,都具有历时研究属性。

专门从事历时语言研究的学者人数不多,然而历时语言研究之于语言学绝非可有可无。相反,许多语言学问题,其答案正蕴含在历时语言学之中。譬如,在我国古汉语学界,何乐士(1984: 261)对先秦两汉语言分期问题的研究,就提出以西汉为界划分汉语的解决方案。何乐士通过系统量化语法对比发现,《史记》与《左传》有显著不同,在汉语史分期问题上,先秦与两汉应划为两个时期。

语言时刻在变,语言变化是其本体属性,因此开展历时语言研究是语言学题中应有之义。

2. 历时语言研究语料及分析方法

为表述方便,以下我们将历时语料库分为“千年历时语料库”和“百年历时语料库”²,前者包含历时几百年乃至上千年的语料,后者则只涵盖百年以内的文本。理论上,两者是连续体,难以截然分开。

2.1 千年历时语料库

由百年以前文本构成的语料库数量远超想象。其核心原因是此类文本不受版权制约,可自由使用。另一重要推动因素是,很多互联网巨头、出版企业、学术机构注入巨资,促成海量古籍文献的高品质电子化。谷歌图书(Google Books)、archive.org、HathiTrust 数字图书馆(HathiTrust Digital Library)、EEBO 早期英语图书在线项目(Early English Books Online)、古登堡工程(Project Gutenberg)是这类电子文本库的代表。这些项目的主要特点是语料量巨大,库容可超百亿词次,同时它们致力于图书的文本化,即全文可检索。

此外,还有很多由语言学家创建的千年历时语料库。其中最有影响的当属芬兰赫尔辛基大学的系列语料库。赫尔辛基英文语料库(The Helsinki Corpus of English Texts)是最早的通用型电子化英语历时语料库。该库由 Matti Rissanen 在 1991 年主持建成,规模为 150 万词左右。其语料跨度从公元 750 年

1 本文介绍的历时语言变化,有些处于语法化的某个阶段,或具备语法化的某些特征,并不都是严格意义上的语法化研究。

2 本节相关语料库的基础信息可参见: <http://www.helsinki.fi/varieng/CoRD/corpora/>。

前后到 1700 年前后,涉及古英语、中古英语和早期现代英语。赫尔辛基大学还创建了一些专题历时语料库。例如,“早期英语书信语料库家族”(The Corpora of Early English Correspondence, 1998 年初步建成并处于持续扩建中,目前规模已超 500 万词次,时间跨度为公元 1403 年至 1800 年)、“医学早期英语书面语语料库”(The Corpus of Early English Medical Writing, 375 万词次,时间跨度为公元 1375 年至 1800 年)、“1560-1760 英语对话语料库”(The Corpus of English Dialogues 1560-1760, 2006 年建成,120 万词次,时间跨度为公元 1560 年至 1760 年)等。瑞士苏黎世大学创建了“ZEN 英语新闻历时语料库”(Zurich English Newspaper Corpus, 简称 ZEN, 2004 年建成,160 万词次,时间跨度为公元 1661 年至 1791 年)。

欧洲之外,美国学者也致力于创建其独具特色的历时语料库。ARCHER 语料库(A Representative Corpus of Historical English Registers)和 COHA 语料库(Corpus of Historical American English)是其中突出代表。ARCHER 语料库由 Douglas Biber 于 1993 年初步建成,后由英国曼彻斯特大学维护更新。最新的 ARCHER 3.2 版本约有 330 万词次,同时收录英国和美国英语。时间跨度为公元 1650 年至 1999 年。ARCHER 语料库包含广告、戏剧、小说、教会布道辞、期刊、法律、医学、新闻、早期散文(early prose)、科学、书信及日记共 12 个语域。COHA 语料库于 2010 年由 Mark Davis 建成(Davis 2012)。其中包含 1810 年至 2009 年间出版的小说、通俗杂志、新闻、非虚构类图书共计 4 亿词次。不难看出,美国历时语料库的特色是多语域取样。这也是美国语料库语言学的重要特色(许家金 2019: 5)。

2.2 百年历时语料库

收集几百年间的英语文本并将其电子化,终究不是易事。近些年新增的一类历时语料库,即我们这里谈的百年历时语料库,其中很多是在当代共时语料库基础上自然延伸而来。这类语料库的突出代表是布朗家族语料库(Brown Family Corpora)。这一族语料库以 Francis & Kučera (1964)建成的 1961 年 Brown 美国英语平衡书面语语料库为起点。后有 1961 年 LOB 英国英语语料库、1992 年的 Frown 美国英语语料库、1991 年的 FLOB 英国英语语料库、2006 年的 AmE06 美国英语语料库和 BE06 英国英语语料库、2006 年建成的 BLOB-1931 英国英语语料库,以及 2009 年的 Crown 美国英语语料库和 CLOB 英国英语语料库。这些库合在一起,构成了英美英语百年历时语料库。这些语料库之所以能够合并使用,原因在于它们都采用相同的取样方案,即语料库包括 500 个文本,每个文本 2000 单词,分为 15 个文类、4 大体裁,因而各库之间高度

兼容且可比。

在口语历时语料库建设方面,DCPSE语料库(The Diachronic Corpus of Present-Day Spoken English)抽取了上世纪60年代到80年代的伦敦-隆德英语口语语料库(London-Lund Corpus,简称LLC)中的40万词,以及上世纪90年代ICE-GB语料库中的40万词英语口语语料组合而成。

2018年发布的BNC2014语料库与1994年建成的经典BNC语料库,构成了20年间短期平衡历时语料库。两者分别包含9,000万词书面语和1,000万词口语,且取样方案十分接近。

美国学者则在大数据历时库方面占得先机,Mark Davis创建的COCA语料库,以每年2,000万词规模,按口语、小说、通俗杂志、新闻、学术5类体裁等比扩充。该库现包含1990年迄今6亿词次语料,成为名副其实的短期通用型美国英语历时库。在专题历时库方面,Mark Davis还创建了“《时代周刊》语料库”(Time Magazine Corpus),收录了《时代周刊》1923年创刊至2006年所有年份的期刊文本,可用于分析80多年间美国通俗杂志中语言使用的历时变化。

2.3 历时语言研究方法

本节将概述历时语言研究的量化方法。简而言之,共时与历时语言研究的区别在于,前者只需做一次数据采集和分析即可;后者则需采集至少两个时间点上的语言数据,并加以对比。根据研究需要,如果采集了多个时间点的语料,就可以将前后的静态语言切片联结成更全面的图景,这好比地质学中通过岩层断面推断地球演变史。

目前很多在线语料库或文本库都提供检索结果的历时展示。例如,Google N-gram viewer、COHA语料库能够分别以折线图或柱状图显示某些语言特征的历时频数变化。一些单机版语料库软件,如WordSmith Tools(2012年发布的第6版开始)提供按时间呈现检索结果的“时间轴”(Time-line)功能。若读入语料文本包含年份信息,则检索结果可以折线图及直方图的可视化方式显示检索项的历时分布。BFSU PowerConc软件也可设定文本年份,从而按时间分布呈现频数情况。Hans Rosling开创的动态图(motion chart),可以很好地将不同时期具有突出特点的语言特征,以及多个语言特征之间的相互关系,以动画效果连续呈现出来。

在历时语言研究的统计方法方面,比较能代表相关统计方法最新进展的是聚类分析、对应分析和混合效应逻辑回归建模等。这些方法属于多因素分析。它们可以将不同历史时期进行切分,以判断历时语言演变的节奏快慢,也可以综合考察多个语言特征与历史时期之间的对应关系,以及哪些因素更能反映相应

时期的语言特点等等。

下文将介绍基于语料库的长期语言演变和短期语言渐变研究概况,其中涉及对词汇、短语、语法、话语、语用、认知等多个方面的历时语言研究。

3. 语言演变研究

虽说历时语言研究近年呈现集中发展趋势,但语言历时演变从来都是语言研究的重点。例如,国际最著名的 ICAME 语料库研究会,1996 年在 International Computer Archive of Modern English 的基础上增加“中世纪英语”,改名为 International Computer Archive of Modern and Medieval English,就是一个明证 (Leech & Johansson 2009: 18)。基于语料库的语言演变研究涉及单个语言特征(往往以词汇化形式出现的语法范畴,如 *that*、*be* 动词、*thou* 等)、特定短语项(如 *in terms of*、*frankly speaking*),以及对多个语法范畴的综合考察(如语法复杂度;属格标记与体裁、历史年代、说话人性别等因素的交互影响)等一些研究选题。

在较早开展的历时语料库语言学研究中, Matti Rissanen 等人的工作最具影响。Rissanen (1991: 272, 287-288) 曾基于赫尔辛基历时语料库开展了宾语从句引导词 *that* 及其零形式演变趋势的研究。在该研究中, Rissanen 对各历史时期口语和书面语中 *that*/ 零形式分布、引导动词(如 *say*、*tell*、*know*、*think*) 对 *that*/ 零形式的选择倾向、主句主语为名词或代词、引导动词与连接标记之间是否有插入成分、引导动词是否是限定形式,以及体裁对 *that*/ 零形式的选择偏好,进行了全面的量化统计。得出以下结论:宾语从句的零形式引导语并非 *that* 的省略,在古英语中 *that*/ 零形式两者并存。我们不能说其中一种形式是另一种形式的省略。只不过,在口语风格英语语料中,零形式自古至今都是宾语从句连接成分的非标记形式。在书面体裁英语中, *that* 历来就是默认形式。就具体历史阶段来看, 17 世纪是英语宾语从句 *that* 省略的高峰期,而 18 世纪人们更倾向于循规蹈矩保留 *that*。就引导动词来看,动词 *say* 后面接 *that* 更倾向于出现在正式语体中。而从 15 世纪至今,动词 *think* 后会更多采用零形式引导宾语从句。新进入英语的动词,其引导的宾语从句起初都采用显性 *that* 标记,经过若干历史时期后,才会逐步出现零形式的情况。可见, *that*/ 零形式选择使用的历时发展受体裁、引导动词等多方面因素的影响。在近年的语法范畴演变研究中, Sommerer (2018) 关于古英语中冠词形成的构式研究,也体现出很强的理论和方法价值。

在专就短语层面的英语历时演变研究方面, Hoffman (2005: 132, 138) 是一项代表性成果。他集中讨论了一组“介词 + 名词 + 介词”型复杂介词短语

(如 on behalf of、with regard to、in search of、in relation to 等)的演变问题。这里以 in terms of 为例,对 Hoffman 的研究加以概述。Hoffman 基于古登堡工程以及牛津英语词典中的例句,构建了其历时英语语料库,同时整合了当代英语语料。经对英国国家语料库调查发现, in terms of 是最高频的复杂介词短语。In terms of 的零星使用始见于 19 世纪末的语料, 20 世纪初叶使用逐渐增多,且在英语口语中最为常见。即便在口语中, in terms of 也是在相对正式的会谈或会议中使用更多,而在普通人的会话中出现并不多。In terms of 的演进过程经历了明显的语用化。语料统计显示, in terms of 经常与填充停顿(13%)、话语片段重复(9%)等非流利现象共现,有时 in terms of 本身也是非流利话语的一部分,属于即席话语中的迟疑标记,有时可起到转换话题的作用,或者说是一种话轮管理的交际策略标记。历时语料可以有效揭示 in terms of 这样的语法片段如何逐步词汇化及语用化。短语这一语言学范畴与认知语言学中的构式有很大交叠。近年也有不少针对某些构式进行历时考察的案例(另见 Shao *et al.* 2019)。

随着语料库语言学方法的深化,同时对多个语法特征进行综合考察已十分常见。Biber & Gray (2016) 基于 1700-2005 年间各类学术文本,对学术英语句法复杂度的历时演变做了深入探究。其中早期语料源自 ARCHER 历时语料库,另外他们还自建了 20 世纪学术英语库,同时结合了当代英语语料资源。所有语料都充分考虑到小说、新闻、学术(甚至文科、理科等子学科)的语域细分,以便深入考察语域与语法特征之间的关联。该研究得到一些重要发现,一改以往对学术英语的固有认识。先前通常认为学术英语的突出特点是句式复杂(特别是嵌套句型较多)、多用被动态和名物化;并且,相对于口语、小说、新闻,学术语体随着时间推移,在语法上相对稳定,变化不大。但 Biber 和 Gray 的历时研究表明,传统意义上的学术英语特征并未显著高于其他语域,而学术英语在名词短语的构成上,经历了极具语域特色的突出变化。学术语域的句法复杂性主要表现为名词短语复杂性,而非主从句嵌套方面的复杂性。具体而言,几百年间,学术英语由多用主从复杂句式逐步发展为多用“形容词 + 名词”、“名词 + 名词”、“名词 + 名词同位语”、“名词 + 分词短语”、“名词 + 介词短语”等构造。这些名词构造使得当代学术英语结构更为紧致(structurally compressed),信息更为浓缩。相应的名词短语事实上起到了关系从句和属格('s 和 of 结构)等作用。这反映了学科发展不断精细化,学科领域不断分化。科学中新的现象、概念、方法不断涌现,而这些主要通过名词性成分来承载。比如当代学术文本中高频使用的 N-based、N-related 一类的合成词,起到的作用便是交代新兴概念和方法等相关背景,或者说明不同概念之间的关系。

Biber 和 Gray 的研究除了发现学术英语语法演变主要围绕名词,体现出浓

缩化演变趋势外,与之相对,他们还发现非正式语体的语法演变,较多围绕与动词相关的语法范畴展开。后者的演变在近 100 年也存在显著变化。不过,从千年的语言演变历史来看,这样的变化只能算作渐变。

4. 语言渐变研究

近 100 年左右的语言渐变,总体表现为“通俗化”(popularization)或者“口语化”(colloquialization)的倾向,具体表现为缩略形式(如's、't)增多、半情态动词(如 have to)增多、进行体增多等。

Leech *et al.* (2009) 基于两代布朗家族语料库,即上世纪 60 年代的 Brown 和 LOB 语料库,以及上世纪 90 年代的 Frown 和 FLOB 语料库,从历时和英美变体两个维度对多项语法特征进行了对比。这些语法范畴包括:虚拟语气、(半)情态动词、进行体、被动语态、轻动词、非限定小句、名词短语。以情态动词的渐变为例,四个百万词级语料库中呈现出相当一致的趋势,即在 30 年间,所有情态动词(would、will、can、could、may、should、must 等)均有不同幅度的减少。而在情态动词的语义表现方面,原先有多重含义的 may、should 逐渐只保留其中最常用的一个含义,这印证了语法化理论里的“专门化”(specialization)过程。而就 must 而言,在总体频数降低的同时,其表示强制义务情态(deontic obligation)含义的用法出现下降。从 must 的主语来看,第一人称和第二人称的数量在 30 年间明显减少。Leech 等人将情态动词总数的降低和强制意味情态义项的减少解释为英语的“平易化”(democratization)(同上:49,88)。与此同时,其他一些英语表达顺势而升。例如,与 must 的少用形成对照的是 be able to、be going to(包括 gonna)、have to、need to、want to(包括 wanna)的多用。除语法化、平易化外,近 30 年英国英语还出现“美国化”(Americanization)倾向。例如,与上世纪 60 年代相比,英国英语中 help...do 明显增多,而 help...to do 的使用显著减少,这与美国英语 help...do 的使用模式越发接近。可见,即便在很短的历史时期内,也会发生较为明显的语法变化。语料库证据告诉我们,语法并非像以往认为的那样相对稳定,几十年之内不会发生显著变化。

除了像 Leech 等人对 30 年间两组语料库进行对比外,研究中若能更细密地收集历时语料,则可对语言渐进做精细考察。比如 Leech (2011) 在 1961 年和 1991/1992 年语料基础上增加了 1901 年、1930 年、2006 年数据,从而达到 5 个英国英语数据点,据此更清晰地揭示出一个世纪以来,英国英语整体呈现少用情态表达的趋势。Leech 还利用 COCA 和 COHA 两个美国英语语料库,证明了美国英语同样出现情态动词减少的情况。此类研究中,均衡取样的历时语料不易获得,而不同时期语料取样的可比性是研究信度和效度的保证。不应出现

Millar (2009) 以《时代周刊》这一特定杂志语料库中的情态用法, 得出整个英语走势的错误做法。

在语言渐变研究领域, 也如长期语言演变一样, 出现了综合数十个语法特征考察英语演变的研究思路。Yao & Collins (2019) 运用分层聚类 (hierarchical cluster analysis) 和多维尺度 (multidimensional scaling) 等多元统计方法, 围绕 69 个语法特征的分布情况对澳大利亚英语四个不同时期 (1931、1961、1991/1992、2006) 的演进进行了综合讨论。在分析过程中, 该研究将澳大利亚英语语料库 (AusBrown) 与不同时期的美国英语布朗家族语料库和英国英语布朗家族语料库进行了对比。作者从与英国的宗主关系和全球化等社会历史因素角度解释了澳大利亚英语追英、自创、学美的语法变化特点。同时, 在澳大利亚英语中也出现了前文所谈的口语化和信息浓缩的演变特征。

像 Yao & Collins (2019) 这样, 以多特征多变量的方法考察文本, 与 Biber (1988) 的语域变异研究较为类似, 换言之, 已进入话语研究层面。

5. 话语、语用、认知历时研究

前文介绍的历时语言研究, 多在既有语言现象基础上, 增加不同时期语料数据点, 通过前后对比观察历时变化。而近些年基于语料库的历时语言研究中, 在语言现象选择方面也有所拓展, 主要体现为对话语、语用、认知现象的考察, 相关成果数量增长迅速。

语法化过程有时历经千年, 而话语、语用、认知现象的演变, 往往通过对近百年, 甚至几十年间的语料对比便可观察到。

基于语料库的历时话语研究并非都是最近 10 年左右新增而来。Biber & Finegan (1989: 515) 基于 4 个多世纪的小说、杂文和信函语料, 考察了 400 多年来三个语域中 67 项词汇语法特征所呈现的语域特征。总体上, 三个体裁都随着时间推移, 逐步表现得更加口语化。这种口语化在 18 世纪时经历了一些反复。这跟前文 Leech 等人谈到的英语平易化十分类似。Atkinson (1999: 142-145) 按照 Biber 和 Finegan 的方法系统研究了 1675-1975 年间的《伦敦皇家哲学学会会刊》(*The Philosophical Transactions of the Royal Society of London*), 发现学者们开始变得更为客观, 较少以作者本人作为主语; 抽象表述不断增加; 话语的叙述性减弱。

在话语研究领域, 针对特定概念以及事件的话语建构研究, 都或多或少地涉及历时语料的使用。例如, 1586 年以来英国詈骂语的演变史 (McEnery 2006)、 “9·11 事件” 前后新闻报道方式的变化 (Qian 2010)、 半个世纪以来英美博物馆宣传话语的演变 (Lazzeretti 2016)、 1987-2014 年《纽约时报》中“危机” 概念的

表述 (Zinn & McDonald 2018)、学术话语的历时变化 (Hyland & Jiang 2019) 等。还有些课题则限于历史上的话语现象。例如, 历史文本中的会话互动 (Culpeper & Kytö 2010)、17 世纪娼妓文化 (McEnery & Baker 2017) 等。历时视角之所以在话语研究中不可或缺, 原因在于有关现象或事件都会经历酝酿、发展、关键节点及后续影响这样的演变路径。在批评话语研究领域, 更有“话语 - 历时分析法” (Discourse-historical Approach, Reisigl 2017), 主张将历时分析作为话语研究必不可少的维度。

语用现象的历时研究以 Andreas Jucker 及其创办的《历史语用学学刊》 (*Journal of Historical Pragmatics*) 最具代表性。在语料库历时语用研究方面, Jucker 成果最为丰硕。Jucker & Taavitsainen (2014) 基于 COHA 美国英语历史语料库以及 COCA 美国当代英语语料库, 考察了 1810 年至 2010 年间“赞誉”这一言语行为的演变。除了发现语料中 compliment 一词出现频率存在明显减少趋势外, 作者还对赞誉类型、赞誉发出者、赞誉受话者、赞誉目标、赞誉的回应都做了手工标注。研究发现与 Holmes (1988) 的结论不同, 200 年来, 美国英语中以 compliment 一词发出的赞誉言语行为由男性占主导 (75% 以上)。相关的赞誉主要关注个人特质 / 友谊、能力, 其次才是对外貌和物品的赞美。另外, 七成以上美国人都倾向于接受他人赞誉。Su (forthcoming) 利用局部语法对美国英语中的道歉进行了历时描写。另一类重要的历时语料库语用研究是有关语用标记的形成, 这一过程往往跟词汇化、语法化交互影响 (参见 Kong & Qin 2017)。

相关研究中涉及的认知现象, 主要包括概念隐喻、概念转喻以及构式的历时演变。这一研究取向多由西欧学者主导, 特别是 Dirk Geeraerts 及其团队成员。例如, Glynn (2014) 对比了 19 世纪和 20 世纪的通俗文学作品中“家” (home) 的概念隐喻演变。他对“家”的概念范畴 (即 place ‘处所’、house ‘住宅’、land ‘土地’、nation ‘国家’) 及语义特征 (包括 lodging ‘起居’、shelter ‘庇护所’、comfort ‘舒适’、security ‘安全’、origins ‘源泉’、belonging ‘归属’、possession ‘拥有’、struggle ‘奋斗’、building ‘建筑’) 进行了详细标注。Glynn 通过层次聚类分析发现, 19 世纪到 20 世纪的 200 年间, home 表达住宅和国家这两个概念的情况始终比较稳定。多重对应分析则显示, 在概念、语义、时代三个变量的交互中, 可以观察到从 19 世纪到 20 世纪, house 都与“舒适”、“起居”关联密切, 而在 20 世纪 house 很明显与“奋斗”、“拥有”关系比较近。可见 20 世纪人们背负着更重的住房压力。Zhang (2016: 181-285) 运用多维尺度法对我国周朝至民国汉语中女性概念的转喻用法做了历时考察, 揭示了女性概念在不同目标域范畴以及不同题材语料中的演变规律。

隐喻、转喻之外,构式是历时认知语料库研究中最重要话题。比如, Delorge *et al.* (2014) 运用对应分析法,对 19 世纪和当代荷兰语中转移类构式的使用倾向进行了描绘。Shank *et al.* (2014) 运用逻辑回归法对 1560 年至 2010 年的英语口语和书面语中 I think that 和 I think [零形式] 的选择制约做了多因素考察。这与上文介绍的 Rissanen (1991) 研究几乎同题,然而考察的因素更为细致,作者们针对 9,720 例含 think 的语句,深入分析探讨了历史时期、语体、主句成分、从句长度、主语代词使用、主从句时态一致性、主从句指代一致性、主从句极性一致性等因素。另外,在研究方法上,Shank *et al.* (2014) 与 Rissanen (1991) 的主要不同在于,前者不但分析了各因素的单独影响,还探讨了因素间的交互作用。更重要的是,每个因素在 I think that/ 零形式这一构式演变中的影响程度都通过量化指标得以表示。Shank *et al.* (2014) 总体上验证了前人研究,但在部分指标上也更新了此前的结论。例如,不像 Rissanen (1991) 发现零形式用法越来越多,相反,随着历史的推移,that 作为导句词的用法实际上是有所增加的。

在词汇、语法、话语、语用、认知这些语言本体层面的历时研究之外,利用语料库开展语言接触、翻译对原创语言的影响研究,也散见于文献中(参见 House 2016: 96-108),值得关注。

6. 结语

用法本位语言观强调语法的动态浮现性,这一理念本身便蕴含“时间性”(temporal) (Hopper 1987: 141)。每个语言特征,如同生命体,都有其历史。在历时语料库支撑下,学者们可以对传统的词源学、对比语文学、语法化进行更为系统的考察。概而言之,在大数据、多元统计、共现与共选等新思维的指导下,我们可以将任一语言学分支都冠以“历时/历史……研究”,本文开始处已列出很多此类学科融合,另有“历史语料库文体学”(Historical Corpus Stylistics, Studer 2008)、“历史话语分析”(Historical Discourse Analysis, Brinton 2015 等)的例子。这也与国际学界学科深化、细化的趋势一致。近年历时语言研究子学科的层出不穷之势,也是学科专门化的自然体现。虽说可以对所有语言现象都进行历时语料库的探究,然而,这终究不过是一种研究视角和路径。热衷于打造新兴子学科,有舍本求末之嫌。我们主张应当多谈实际语言问题,少提新兴学科领域。

在研究选题上,基于历时语料的词汇语法(如情态动词及 that 从句等)研究,可靠性明显更高;然而有关语用现象(如言语行为)的研究,仅从个别词汇短语入手,以形索义,终因形义并非一一对应,仍然显得捉襟见肘。

在历时语料库研究的分析方法上,多维度分析、多因素分析被越来越多的研究采用。这些方法不仅对语言现象的描写更加细致、全面;更重要的是,它们可以考察多个语言特征以及语言特征与语境变量之间的相互影响。这在很大程度上已进入到现象的解释层面。

从上文综述可见,在这股新兴的研究热潮之中,中国学者,特别是一批海外归来的青年学者,已将我国历时语料库研究成果推向国际。只可惜相关研究更多是英语本体研究,也多半是在西方学者框架内的复制或延展性研究,缺少对我国本土历时研究选题的关注。在国内,赵秋荣、王克非(2013)、秦洪武、夏云(2017)、王克非、秦洪武(2017)、庞双子(2018)、朱一凡(2018)立足汉语,关注了外译汉以及翻译对现当代汉语形成的影响,获得了很多有价值的发现。

参考文献

- Atkinson, D. 1999. *Scientific Discourse in Sociohistorical Context* [M]. Mahwah, N.J.: Lawrence Erlbaum.
- Barðdal, J., E. Smirnova, L. Sommerer & S. Gildea (eds.). 2015. *Diachronic Construction Grammar* [C]. Amsterdam: John Benjamins.
- Biber, D. 1988. *Variation across Speech and Writing* [M]. Cambridge: CUP.
- Biber, D. & E. Finegan. 1989. Drift and the evolution of English style: A history of three genres [J]. *Language* 65: 487-517.
- Biber, D. & B. Gray. 2016. *Grammatical Complexity in Academic English: Linguistic Change in Writing* [M]. Cambridge: CUP.
- Brinton, L. 2015. Historical discourse analysis [A]. In D. Tannen, H. Hamilton & D. Schiffrin (eds.). *The Handbook of Discourse Analysis* (2nd edition) [C]. Oxford: Wiley Blackwell. 222-243.
- Campbell, L. 1999. *Historical Linguistics: An Introduction* [M]. Edinburgh: Edinburgh University Press.
- Culpeper, J. & M. Kytö. 2010. *Early Modern English Dialogues* [M]. Cambridge: CUP.
- Davis, M. 2012. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English [J]. *Corpora* 7: 121-157.
- Delorge, M., K. Pleviets & T. Coleman. 2014. Competing 'transfer' constructions in Dutch: The case of *ont*-verbs [A]. In D. Glynn & J. Robinson (eds.). *Corpus Methods for Semantics* [C]. Amsterdam: John Benjamins. 39-57.
- Francis, W. & H. Kučera. 1964. *Manual of Information to Accompany a Standard Corpus of Present-day Edited American English* [M]. Providence, Rhode Island: Brown University.
- Glynn, D. 2014. The conceptual profile of the lexeme *home*: A multifactorial diachronic analysis [A]. In J. Díaz-Vera (ed.). *Metaphor and Metonymy across Time and Cultures* [C]. Berlin: Walter de Gruyter. 265-293.
- Hoffman, S. 2005. *Grammaticalization and English Complex Prepositions: A Corpus-based Study* [M]. London: Routledge.
- Holmes, J. 1988. *Women, Men and Politeness* [M]. London: Longman.
- Hopper, P. 1987. Emergent grammar [A]. In J. Aske, N. Beery, L. Michaelis & H. Filip (eds.).

- Papers of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* [C]. University of California, Berkeley: Berkeley Linguistic Society. 139-157.
- House, J. 2016. *Translation as Communication across Languages and Cultures* [M]. London: Routledge.
- Hyland, K. & F. Jiang. 2019. *Academic Discourse and Global Publishing: Disciplinary Persuasion in Changing Times* [M]. London: Routledge.
- Jenset, G. & B. McGillivray. 2017. *Quantitative Historical Linguistics* [M]. Oxford: OUP.
- Jucker, A. & I. Taavitsainen (eds.). 2010. *Historical Pragmatics* [C]. Berlin: Mouton de Gruyter.
- Jucker, A. & I. Taavitsainen. 2013. *English Historical Pragmatics* [M]. Edinburgh: Edinburgh University Press.
- Jucker, A. & I. Taavitsainen. 2014. Complimenting in the history of American English: A metacommunicative expression analysis [A]. In I. Taavitsainen, A. Jucker & J. Tuominen (eds.), *Diachronic Corpus Pragmatics* [C]. Amsterdam: John Benjamins. 257-276.
- Kohnen, T. 2009. Historical corpus pragmatics: Focus on speech acts and texts [A]. In A. Jucker, D. Schreier & M. Hundt (eds.), *Corpora: Pragmatics and Discourse* [C]. Amsterdam: Rodopi. 11-36.
- Kong, L. & H. Qin. 2017. The development of manner of speaking markers in English and Chinese: Pragmaticalization, grammaticalization and lexicalization [J]. *Journal of Pragmatics* 107: 16-30.
- Lazzeretti, C. 2016. *The Language of Museum Communication: A Diachronic Perspective* [M]. Cham: Palgrave Macmillan.
- Leech, G. 2011. The modals ARE declining: Reply to Neil Millar's "Modal verbs in TIME: Frequency changes 1923-2006" [J]. *International Journal of Corpus Linguistics* 16: 547-564.
- Leech, G. & S. Johansson. 2009. The coming of ICAME [J]. *ICAME Journal* 33: 5-20.
- Leech, G., M. Hundt, C. Mair & N. Smith. 2009. *Change in Contemporary English* [M]. Cambridge: CUP.
- McEnery, T. 2006. *Swearing in English: Bad Language, Purity and Power from 1586 to the Present* [M]. London: Routledge.
- McEnery, T. & H. Baker. 2017. *Corpus Linguistics and 17th-Century Prostitution* [M]. London: Bloomsbury.
- Millar, N. 2009. Modal verbs in TIME: Frequency changes 1923-2006 [J]. *International Journal of Corpus Linguistics* 14(2): 191-220.
- Nevalainen, T. & H. Raumolin-Brunberg. 2014. *Historical Sociolinguistics* [M]. London: Routledge.
- Qian, Y. 2010. *Discursive Constructions around Terrorism in The People's Daily (China) and The Sun (UK) before and after 9.11* [M]. New York: Peter Lang.
- Reisigl, M. 2017. The Discourse-historical Approach [A]. In J. Flowerdew & J. Richardson (eds.), *The Routledge Handbook of Critical Discourse Studies* [C]. London: Routledge. 44-59.
- Rissanen, M. 1991. On the history of that/zero object clause links in English [A]. In K. Aijmer & B. Altenberg (eds.), *English Corpus Linguistics* [C]. London: Longman. 272-289.
- Shank, C., K. Plevoets & H. Cuyckens. 2014. A diachronic corpus-based multivariate analysis of "I think that" vs. "I think zero" [A]. In D. Glynn & J. Robinson (eds.), *Corpus Methods for*

- Semantics* [C]. Amsterdam: John Benjamins. 279-303.
- Shao, B., Y. Cai & G. Trousdale. 2019. A multivariate analysis of diachronic variation in *a bunch of NOUN*: A Construction Grammar account [J]. *Journal of English Linguistics* 47(2): 150-174.
- Sommerer, L. 2018. *Article Emergence in Old English: A Constructionalist Perspective* [M]. Berlin: Walter de Gruyter.
- Studer, P. 2008. *Historical Corpus Stylistics* [M]. London: Continuum.
- Su, H. (forthcoming). Local grammars and diachronic speech act analysis: A case study of apology in the history of American English [J]. *Journal of Historical Pragmatics* 21(1).
- Taavitsainen, I., A. Jucker & J. Tuominen (eds.). 2014. *Diachronic Corpus Pragmatics* [C]. Amsterdam: John Benjamins.
- Winters, M., H. Tissari & K. Allan (eds.). 2010. *Historical Cognitive Linguistics* [C]. Berlin: Walter de Gruyter.
- Yao, X. & P. Collins. 2019. Developments in Australian, British, and American English grammar from 1931 to 2006: An aggregate, comparative approach to dialectal variation and change [J]. *Journal of English Linguistics* 47(2): 120-149.
- Zhang, W. 2016. *Variation in Metonymy: Cross-linguistic, Historical and Lectal Perspectives* [M]. Berlin: Walter de Gruyter.
- Zinn, J. & D. McDonald. 2018. *Risk in The New York Times (1987-2014): A Corpus-based Exploration of Sociological Theories* [M]. Cham: Palgrave Macmillan.
- 何乐士, 1984, 《史记》语法特点研究——从《左传》与《史记》的比较看《史记》语法的若干特点 [A]。载程湘清(编), 《两汉汉语研究》[C]。济南: 山东教育出版社。1-261。
- 庞双子, 2018, 基于英汉历时平行语料库的源语透过性研究——以物称代词 it 为例 [J], 《外国语》(6): 91-101。
- 秦洪武、夏云, 2019, 《基于历时语料的翻译与现代汉语互动研究》[M]。上海: 上海交通大学出版社。
- 王克非、秦洪武, 2017, 基于历时复合语料库的翻译与现代汉语变化考察 [J], 《外语教学与研究》(1): 37-50。
- 许家金, 2019, 美国语料库语言学百年 [J], 《外语研究》(4): 1-6。
- 赵秋荣、王克非, 2013, 英译汉翻译语言的阶段性特点——基于历时类比语料库的考察 [J], 《中国翻译》(3): 15-19。
- 朱一凡, 2018, 《基于语料库的英汉翻译对当代汉语影响的研究》[M]。上海: 上海交通大学出版社。

收稿日期: 2019-08-09; 修改稿, 2020-01-10

通讯地址: 100089 北京市 北京外国语大学中国外语与教育研究中心 / 人工智能与人类语言重点实验室