

专题·多因素语言研究

栏目主持人: 许家金

语料库语言学注重对语言事实的细颗粒度描写,并将这种深度描写视作对语言运作机制的理论解释。语料库语言学领域的多因素分析(multifactorial analysis)和 John Sinclair 开创的短语学,为理论语言学作出独特贡献。本专栏五篇文章旨在展示多因素分析在我国英汉语研究中的最新进展。许家金一文综述了多因素分析的“概率语境共选”理念,以及相关多因素研究议题。黄莹、任伟一文采用多分类逻辑斯蒂回归建模和多重对应分析,考察了 let、allow、permit 构成的英语允让构式的致使倾向。田笑语、张炜炜一文借助随机森林和多分类逻辑斯蒂回归,比较了“使”、“令”、“让”构成的三种分析型致使构式在中国大陆、中国台湾、新加坡三地汉语变体中的变异情况。房印杰、梁茂成一文综合了二分类逻辑斯蒂回归和可接受性判断两种方法,探讨了中国英语学习者中介语关系代词 that 取舍的使用行为。徐秀玲一文利用条件推断树的可视化方法展示了翻译汉语回指语何种情形下选用显性和隐性主语回指语的机制。多因素分析虽有技术门槛,但确是揭示语言中形义选择机制的有效手段,可谓:语言选择难归一因,多元综览方得全豹。

多因素语境共选: 语料库语言学新进展*

许家金

(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 北京 100089)

摘 要: 概率性多因素语境共选是语料库语言学的新近热点。它致力于探讨语言形式和语义之间的对应关系,主要包括“一形多义”和“一义多形”。词典学和社会语言学较早采用多因素方法,认知语言学则是近年多因素分析的主阵地。多因素分析广泛运用于词语义项区分、近义词辨析、近义构式选择、语法标记选择、语序选择等方面。多因素分析既是语言描写技术,也有很强的理论解释力。

关键词: 多因素分析; 语境共选; 形义对应; 多元统计; 语料库语言学

中图分类号: H0 文献标识码: A 文章编号: 1004-6038(2020)03-0001-11

1. 引言

本文旨在重新探讨语料库语言学的学科属性。当前多数国内外学者都将语料库语言学视为方法论,认为语料库语言学是一堆语言材料、一组统计算法、一套分析工具。如此一来,语料库语言学仅是“器”,无关“道”,空有“语言学”之名。语料库语言学的发展史表明,事实并非如此(许家金 2014)。“概率性多因素语境共选”是本文对语料库语言学学科

* 基金项目: 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”(项目编号: 17JJD740003) 阶段性成果。多因素分析主要选题一节的文献梳理得到张懂博士的协助,特此致谢。

作者简介: 许家金 教授,博士,博士生导师,研究方向: 话语研究,二语习得,语言对比与翻译,语料库语言学。作者邮箱: xujiajin@bfsu.edu.cn

特性的初步概况。

Leech(2011:158)曾指出语料库语言学是有很强“方法论导向的语言学分支”(a methodologically-oriented branch of linguistics)。这一论断兼顾了方法和学科。然而,这一说法并未指明作为语言学分支,除了方法优势外,语料库语言学何以与社会语言学、心理语言学、认知语言学等相提并论?

以下本文将讨论语料库语言学对语言本体的认识,及其描写和阐释语言的方法和路径。一般认为语料库语言学重描写,轻阐释。这种观点将描写和阐释对立起来,显然有些过于简单化。我们认为深入的描写强过肤浅、附会的理论阐释,同时深入描写本身也有一定程度的阐释性(胡建华,2018:476-477)。

语料库语言学内部也存在方法导向和理论导向两派,即所谓“基于语料库的范式”和“语料库驱动范式”。本文将综合各派理念的共核部分,从而勾勒语料库语言学的核心语言学价值。

2. 语料库语言学的语言观

语料库语言学主张用法本位(usage-based)(参见Kemmer & Barlow,2007)。用法观强调语言使用者的主体性和使用场景在语言形成中的作用。语法是言语社团经年累月交际活动的动态浮现特征(Hopper,1987)。因此学界早有“语法即用法”(Bybee,2006)的认识。从语言习得角度看,用法观认为掌握语言不是人的天性(nature)而是习性(nurture)。

自上世纪后半叶以来,除生成语言学外,其他诸多语言学流派大都可以归到基于用法的语言观(usage-based linguistics)之下。这一视角:(1)尊重语用;(2)强调功能;(3)依赖语境;(4)关注变异。据此,社会语言学、语用学、话语研究、(系统)功能语言学、历史语言学,乃至认知语言学,都是用法本位的。

尊重语用,从哲学上反映了用法语言观的经验主义基础,即作为抽象概念的“语言”以及人们语言知识的获得不是先天的,而是长期浸染于某种语言环境,不断与他人接触而形成的。语言是动态浮现语法现象。语言知识是认知固化的产物。

强调功能,指的是对语言形式和结构的探讨,都与意义和功能不可分割。形式和意义/功能是一币两面。甚至于说,意义和功能先于语言形式和结构。

依赖语境,指的是语用无不存在于特定语境。语境中的语言形式都反映一定的话语功能。充足的语境足可以消歧,足可以陈述事实,明示意图,还可以联结读者和作者,说话人和听话人。

关注变异,谈的是语境变化。比如英语变体研究、历时语言研究、中介语研究、翻译研究,都是某种形式的语言变异研究。

语境在语言使用行为上,体现为共选关系,这是一种依存式的关联。语境共选有广有狭。譬如社会语言学常常关注语言与社会语境的共选机制。语境化是分析语言使用的核心方法。语境是将用法语言观各语言学分支串联起来的重要线索。

在语用基础上浮现出的语法有其规约性,体现为“群言”(communal language);同时,语

言具有丰富的个体特色和语境依存性(contextual dependency)。语言学理论既要能解释群言的规律性,也要解释个人语言的创造性和特定群体语言的变异性。规律使人达成共识,创造和变异令交际饶有趣味。这可以用“性相近、习相远”来加以概括,即人类有共享的语言规律,同时语言又千差万别。语料库语言学揭示语言规律性和创造性的解决方案是:共选和概率。这两者合而为一,共同驱动着语言的运作。共选理论包括狭义的语言特征共选,以及语言特征和语境变量的共变关系。共选更多解释性相近,共变更多解释习相远。

在语料库语言学内部,共选研究先前主要体现为以搭配分析为主要手段的短语学研究,包括扩展意义单位分析框架、局部语法等;另一种则包括语言特征与语境特征的共选研究,例如,语料库话语研究、基于语料库的社会语言学研究、基于语料库的认知语言学研究。前者是所谓的语料库驱动的路径,后者是所谓基于语料库的路径。其实两者都是对语境变量的综合考察。从这个意义上说,两个具有竞争意味的流派,本是同根生,都是这里讲的概率性语境共选视角下的多因素分析。

概率性多因素分析,首先强调语言形式的选择受限于若干语境特征。这些特征的影响力有强有弱,且有可能共同作用。

受制于不同语境条件的语言形式选择,决定了语言形式会有不同的意义潜势。这种意义潜势的选择很大程度上是概率性的。语言研究中的概率并不完全具有随机性,而是具有条件概率性(conditional probability)。

3. 概率性多因素语境共选研究的起源

多因素观念日常处处可见。例如人体健康由遗传、饮食、锻炼、环境、医疗等多方面因素决定。某种具体疾病也可能有多重诱发因素,因此在医学研究文献中经常采用多元统计方法便不足为奇。

在语言学相关领域,多因素理念也绝非新鲜事物。早在英语教学研究发轫之初,Fries就强调“语境取向”(contextual orientation)的必要性。他指出“若要全面掌握语言,需要系统观察和记录语句在确切语境中使用的多项特征”(1945: 57)。全面掌握多项语境特征,正是多因素分析的核心思想。Firth(1951)的“语境意义观”(contextual theory of meaning)和Harris(1954)的“分布假说”(distributional hypothesis)反映的也都是这种取向。

从概率视角进行多因素语境分析,始于Hanks(1986; 1996)、Atkins(1987)。Hanks(1986: 43)将词典词条描述为一个“方程式”(equation)。方程式左边是“条首词”(head-word,又译作“词目”),右边则是用来解释条首词的语境特征。Hanks主张应通过真实语料将典型的(typical、norm)和拓展的(exploited)构型识别出来,以便合理安排义项。Hanks(1996)以动词语义刻画为例,分析了urge、indict、bother、abandon等动词的句法语义环境。例如,这些动词的上下文中一般会出现什么样的主语(名词还是代词,人还是物,具体还是抽象语义等)。Hanks指出不同“意义潜势”的实现取决于不同词汇、句法、语义、语用的组合(1996: 78, 94),他认为这一思想与Firth所倡导的“识词于其所友”(You shall know a word by the company it keeps)十分吻合。Hanks(1996: 79)把这种语境因素的“选择制约/

倾向”(selectional restrictions/preferences)称为“行为概貌”(behavioral profile)。这一思路被命名为“语料库构型分析”(Corpus Pattern Analysis,简称CPA)并进一步升级为“规范与拓展理论”(The Theory of Norms and Exploitations,简称TNE)(Hanks,2013)。而“行为概貌”这一概念和分析思路被语料库语言学沿用。

Atkins(1987)是另一项重要的早期概率性多因素研究。作者基于COBUILD语料库,识别出数十个可以区别danger一词3个不同义项的词汇语法使用特征。Atkins称其为“语义识别标记”(semantic ID tags)。例如,义项一“危险”,danger用作不可数。其语义识别标记包括:后可接for/from/in/to/with,而不接that;in+(冠词/量词/形容词)+danger;物主代词+danger;out of danger等。义项二“危害”,danger用作可数名词。典型语义识别标记包括:danger(s)后接具体事物,例如dangers of smoking/pregnancy等。义项三“风险”,其典型语义识别标记包括:限定词+danger+of;danger后接不期待发生的事,例如danger of collapse/disintegration等。

Sinclair(2004)的扩展意义单位分析和局部语法研究,在一定程度上采用的也是多因素语境共选理念。只是主张围绕意义单位的研究,建构一套原创的语言理论。

早期的概率性多因素分析聚焦于词典编纂,特别是义项的确立。它们关注的“一形多义”(polysemy/semasiology)现象,即同一词形有多个含义,需要借助多重词汇、句法、语义等局部语境加以解析。

另一类概率性多因素研究关注的是“一义多形”(synonymy/onomasiology)现象。这方面研究在Labov开创的变异社会语言学领域开展较早。人们在表达相同或相近语义时,因不同社会情境而会采用不同的语言形式。例如,汉语中“你”和“您”的使用,都是第二人称代词,但使用场合不同。这种研究的一大特点是因变量是两个或多个选项(比如,这里谈到的“你”和“您”)。Labov(1966)的经典案例是分析纽约几家商场不同说话人在发/r/音是否卷舌的情况。他考察了r在词中间(fourth)或词尾(floor)、随口表达或强调表达、高中低档商场、说话人所在楼层、职业、种族、性别、年龄等多个语言因素和场景因素对r发音的影响。这是典型的多因素研究设计。Labov当时只采用了描述性统计,没有使用多元统计方法。上世纪60年代David Sankoff开发了“变项规则程序”(variable rule program),即Varbrul,用于对Labov(1966)类似的多因素研究进行多元统计分析,该软件后改名GoldVarb,现可通过R软件包Rbrul加以实现。该软件的核心统计方法为逻辑斯蒂回归建模(logistic regression modelling)。类似的研究在社会语言学和二语习得领域已开展约半个多世纪。

近一二十年,认知语言学家重新引发了一波多因素研究热潮。其中两位代表性认知语言学家是Dirk Geeraerts和Stefan Gries。前者更关注语言变异,后者注重研究结论的认知解释。Geeraerts关注言内和言外变量的综合考察;Gries关注的主要是言内因素对词义或构式选择的制约程度(两者研究取向的异同,可参见Heylen et al.,2008)。

概括来看,言内语境共选的多因素探索从词典学开始;言外语境共选的多因素探索从社会语言学开始。相关研究关注的重点为语言形式和意义之间的对应关系。比如,一形多义(同一个词汇语法形式的不同含义)或一义多形(同一含义的不同词汇语法变体)。

4. 多因素分析的主要选题

总体来说,一形多义的研究数量明显少于一义多形。一形多义的研究多半与 Hanks 和 Atkins 的研究相似,往往围绕某个特定单词,识别判定该词不同含义的词汇、句法、语义等制约因素。例如,Gries(2006: 82)利用层次聚类法区分了动词 run 的不同义项。Glynn(2014: 133,140)复制了 Gries(2006)关于 run 的研究,并增加了英语变体和语域两个变量,在统计方法上采用了对应分析和逻辑斯蒂回归建模。

一义多形的多因素分析广泛运用于词汇和语法研究,少量用于话语语用研究。相关研究可归为以下几类:1)近义词辨析;2)近义构式选择;3)语法标记选择;4)语序选择。

(1) 近义词辨析。所谓“辨析”,从语言使用者角度看即是对近义词语加以区分,然后正确选用。类似研究常围绕近义形容词、副词、名词和动词展开。例如 Gries(2010)就英语大小类形容词(big、large、great; little、small、tiny)做了多因素区分。Desagulier(2014)对美国英语中 rather、quite、fairly、pretty 四个副词进行了区分。Janda 和 Solovyev(2009)对俄语中近十个表示欢乐和悲伤的名词进行了多因素辨析。另有英俄跨语言开始类动词(英语 begin、start; 俄语 nač inat' / nač at nač inat' sja / nač at' sja、stat)的对比分析(Divjak & Gries, 2009)。这里介绍的研究,看似关注点为单词,而在具体分析中落实的语言形式多半是短语(或构式)。在分析相关词语的搭配情况时,还时常会考虑到搭配成分的语义和概念含义。因而,上述这些研究被归为认知语义研究。

(2) 近义构式选择。这类研究选题关注动词构式较多,这可能与构式语法中比较关注论元结构有关。相关构式的选择一般称作构式交替(alternation),即某一具有共同概念义的上位构式通常有两种或更多构式变体。在实际语用中,选用某种变体由多个因素综合决定。

5

较常见的研究包括对(1)与格构式(如 Mary gave John a book 与 Mary gave a book to John 之间的选择)的研究。因为是动词相关的构式,这方面的研究通常涉及的因素包括主语、宾语、介词宾语的名词/代词属性、名词短语的长度、生命度、具体/抽象等指标。除了关于英国英语和美国英语(Bresnan & Ford, 2010)的研究外,还有南亚英语(Gries & Bernaisch, 2016)中的与格构式选择倾向研究。英语之外的其他语言(如德语,Geleyn, 2017)以及跨语言(张懂 许家金, 2019)与格交替对比研究也不在少数。

其他动词构式交替研究,包括:(2)动补构式。例如 Gries(1999; 2003)关于动词加小品词构式交替选择(pick up the book 与 pick the book up)的制约因素研究。Deshors(2015)关于动词后接 to do 还是 ing 的选择问题;(3)致使构式。比如 Speelman 和 Geeraerts(2010)关于荷兰语 doen(类似英语中的 do 或 make)以及 laten(类似英语中的 let 或 make)等致使构式的研究。

名词性构式交替中,探讨最多的是属格构式(名词加's 及 of 结构的选择问题)。例如,Rosenbach(2003)关于英语属格选择问题。Gries 和 Wulff(2013)还分析了中国和德国英语学习者的属格交替问题。其他的语言现象还有 more 加形容词/副词和形容词/副词词尾加 er 的比较结构研究(Hilpert, 2008)等。

(3) 语法标记选择。此外一大类多因素研究关注的是语法标记的选用。此类研究中

数量最多的是有关于宾语从句引导关系代词 *that*/零形式的选择倾向。Shank 等(2016)分析了 *think* 等动词后接 *that* 或省略 *that* 的各因素单独影响及其交互作用。

Deshors(2015)对比了中国和法国英语学习者在使用 *can* 和 *may* 这两个极高频情态动词的选择困难。类似的现象,比如 *be going to* 与 *will/shall* 的选择问题,也有不少研究(如 Szmrecsanyi 2003)。

(4) 语序选择。这类研究涉及论元先后顺序、从句和主句先后顺序等,例如 Diessel(2008)谈及时间状语从句出现在主句前后的影响因素。Wiechmann 和 Kertz(2013)关注的是 *although* 和 *whereas* 引导的让步状语从句在主句前后的制约条件。Wulff(2003)专门考察了英语中多个形容词排序的多个制约因素的影响程度。

除此之外,还有少量多因素研究涉及话语和社会语用层面的现象。例如, Han 等(2013)关于上海话话题标记的选择问题。Gries 和 Adelman(2014)研究了日语话题主语出现和不出现的制约因素。Levshina(2017)比较了欧洲多国语言中敬语标记(例如法语中的 *tu* 和 *vous*、德语中的 *du* 和 *Sie*、俄语中的 *ты* 和 *вы*)的选择问题。

总体来说,相关研究涉及词汇和语法现象选择的情况较多,只有极少量研究关注话语语用现象。多数研究着重考察的是言内多因素,有一部分研究加入了语域、变体和历史时期等话语和社会语言学变量。

5. 多因素分析主要统计方法

从上面有关多因素研究选题的概述可知,所谈现象并非全新。多因素分析可以促成旧题新做和旧题深做。在目前技术条件下,多元统计分析所用的基础数据,很大程度上依赖研究人员自行标注。语料的手工标注极为耗时,当然也最能体现研究的语言学价值。语言属性的精细标注加上多元统计,使得多因素研究可以兼顾理论深度和方法优势。

多元统计在多因素分析中的价值,是要协助研究者回答语言特征选择的概率性理据,即在多大程度上,人们在何种语境下,会更倾向选择语言特征 A 而非 B。用于多因素分析的常见多元统计方法有线性分类器、聚类分析、对应分析、逻辑斯蒂回归、多维尺度、条件推断树、随机森林等(这些方法还存在相应的变体)。简而言之,这些方法解决的都是语言特征选择的制约条件,我们称为“形义聚分”问题。以下将从用于多因素分析的数据格式、常见统计方法做一介绍。

数据格式。以往语料库语言学的典型数据格式主要为词表、多词词表、搭配词表、主题词表和索引行等。多因素分析所采用的典型数据格式为“表格型数据”(tabular data),这种数据在 R 语言中被称为“数据框”(data frame),见表 1。

其中第一列 Instance 是关系代词使用 *that* 或采用零形式的情况。后面 Subject、Mat-Polar、VerbSema、MatLengt 几列数据体现的是主语、主句极性、动词语义、主句长度属性等多个因素。若有更多因素,则依次添加列即可。每一横行代表一个具体的语言使用实例。表 1 中为五句话中关系代词使用 *that* 或采用零形式的情况。根据研究实际,往往需要如此标注数千行。语料标注通常在 Excel 中进行,常存储为 csv 格式,也可以把 Excel 工作表中

的数据复制粘贴到*.txt 格式纯文本文件中。在 R 程序中可通过 read.table(file = file.choose()) 命令读取该标注文件,进行多元统计。Glynn(2014: 325 – 327) 给出的多因素统计方法一览表是个有效的导引。

表 1 多因素分析标注样例(两种数据格式)

表格型格式					csv(逗号分隔值) 格式				
Instance	Subject	MatPolar	VerbSema	MatLengt	Instance	Subject	MatPolar	VerbSema	MatLengt
that	n	p	cognition	2	that	n	p	cognition	2
zero	pron	n	speech	5	zero	pron	n	speech	5
zero	pron	n	activity	4	zero	pron	n	activity	4
that	pron	p	cognition	3	that	pron	p	cognition	3
that	n	p	cognition	2	that	n	p	cognition	2

常见多元统计方法。多元统计方法可分为探索性和验证性两种。聚类分析、对应分析等更具探索性,即研究者不预设语言属性(如时态、抽象/具体、语域)与语言特征(如 that 的使用与否)的关联性,而由数据根据语言特征使用概率自行计算出其互相吸引的程度。以英语程度副词辨析为例,根据与之搭配的形容词情况,聚类分析可以用 pvclust 函数及软件包将 fairly 和 quite、a bit 和 somewhat、completely 和 totally 聚到一起(Desagulier 2014: 163)。再比如,我们可以利用 ca 函数及软件包以对应分析方法将动词 talk 和具体话题、suppose 和抽象事件等在二维图形上对应到一起,从而直观看出这些动词使用的语义倾向(Glynn 2014: 458)。而逻辑斯蒂回归则是典型的验证性多元统计。它是基于前人文献,将可能有关的语言属性都尽可能囊括到分析之中。统计运算会赋予每个属性/变量一定的权重,并剔除权重较低的变量。例如,我们可以利用 MASS 软件包中的 glm 函数,对制约动词后接 to do 还是 ing 的句法、语义、语用等多个因素的影响程度及其相关变量之间的交互影响进行量化(许家金 陈哲 2018)。

多因素分析的结果解读起来并不容易,常常需要进行可视化。典型的可视化方法有聚类分析树形图(dendrogram)、对应分析二维图、多维尺度图、条件推断树图和动态图(motion chart)等。

需要指出,多元统计方法是基于概率性语境共选的语言观,并非刻意要将研究方法变得如此复杂。以往基于简单频数的统计,将语言视作随机性现象。目前看来,这种认识有必要更新。语言使用中体现出的概率属于条件概率。例如,主语位置如果是第一人称单数,那么其后的谓语动词的选择就不那么随意了,而必须考虑到相应的性、数、格的一致问题。这一定意义上说明了为何先前基于均值比较和正态分布的一些统计方法并不完全有效。

再有,多因素分析中经常采用的逻辑斯蒂回归建模,而不是通常的线性回归,就是考虑到我们将语言特征选择作为反应变量,这种选择往往有两种或两种以上的可能。通常的线性回归因变量只能是一个变量,且必须是数值型变量,如词汇判断任务中的反应时。另外,逻辑斯蒂回归模型中还可加入随机效应变量(其他自变量为固定效应变量,此类模型称“混合效应逻辑斯蒂回归”)。具体到某个语法范畴变量,可能会包含多个具体形式,例如认知动词会包含 think、assume、suppose 等等,这些必须要作为随机效应变量来处理。如果语料中将具体的作者或说话人作为变量,则也要作为随机效应变量来处理(参见 Baayen et al. 2008)。

总之,采用多因素分析法,就是为了反映真实的复杂语境互动,从而更好地表述、呈现语言选择的机制。一定程度上,这克服了之前语料库语言学分析中过于依赖“多用”和“少用”的浅层描写问题,在描写、解释和预测语言选择机制方面都有所建树。

6. 我国多因素分析的现状与前景

社会语言学传统的多因素分析在我国起步较早。例如,徐大明(1999)曾介绍新加坡华社双语调查中采用的变项规则分析法,即逻辑斯蒂回归。在词典编纂传统的多因素研究方面,方子纯和陈坚林(2014)对表示“巨大”的六个词(immense、enormous、huge、massive、tremendous和vast)的行为概貌进行了多因素分析,考察的语言属性包括它们的名词搭配词、表语用法和语域分布模式,采用的多因素统计方法是“层级配置频次分析”。

认知语言学传统的多因素分析,由张炜炜和Liu(2015)、张炜炜和刘念(2016)、房印杰(2016)等撰文介绍多因素分析的理念及方法。我国学者关于此类多因素实证研究也逐步兴起,包括邵健和张建理(2017)、张炜炜和王芳(2017)、许家金和陈哲(2018)、张懂和许家金(2019)、房印杰和梁茂成(2019)、徐晶凝和郝雪(2019)等人以中文完成的研究成果以及Guo和Chow(2014)、Zhang(2016)、Yao和Collins(2019)、Shao等(2019)等人以英文发表的成果。

概率性语境共选视角下的多因素语言研究,以用法本位为语言观,可以更好地整合语言学内部的不同研究范式。它所希望解决的核心语言学问题是形义映射,具体包括“一形多义”和“一义多形”。多因素分析既关涉语言构型的原型性和规约性,也十分关注变异性和创造性。

语境共选视角下的多因素分析实际上是对用法观的实证化和操作化。同时,多因素研究较好地处理了以往学界对语料库语言学“计算复杂”(computationally sophisticated),但“语言理论方面幼稚”(linguistically naïve)的问题(Murakami et al. 2017)。与单因素研究不同,多因素研究的核心是语境特征的概率性共选,可以说是对Firth语境论的新发展。多因素分析既是描写,其语言属性的共选机制 also 具有很强的阐释力,是有依据可验证的阐释。同时,相关结论还可以预测今后出现的语言现象。

综合来看,我国学者开展的多因素研究,相当数量是对国外成果的复制。我们应更多关注本土化的研究选题,例如汉语研究、中国英语学习者中介语研究、外国留学生汉语中介语研究、英汉对比与翻译研究等等。同时在研究方法上,可尝试将多因素语料分析和心理语言学实验结合,对研究结果进行交叉验证。

参考文献:

- [1] Atkins, B. 1987. Semantic ID tags: Corpus evidence for dictionary senses [A]. In *Proceedings of the Third Annual Conference of the New OED Centre* [C]. Waterloo: University of Waterloo Press.
- [2] Baayen, R. D., Davidson, D. & Bates, E. 2008. Mixed-effects modeling with crossed random effects for subjects and items [J]. *Journal of Memory and Language* (4): 390–412.
- [3] Bresnan, J. & M. Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English [J]. *Language* (1): 168–213.

- [4] Bybee J. 2006. From usage to grammar: The mind's response to repetition [J]. *Language* (4) : 711 – 733.
- [5] Desagulier G. 2014. Visualizing distances in a set of near synonyms: *Rather , quite , fairly , and pretty* [A]. In D. Glynn & J. Robinson (eds.) . *Corpus Methods for Semantics* [C]. Amsterdam: John Benjamins.
- [6] Deshors S. 2015. A multifactorial approach to gerundial and *to*-infinitival verb-complementation patterns in native and non-native English [J]. *English Text Construction* (2) : 207 – 235.
- [7] Diessel H. 2008. Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English [J]. *Cognitive Linguistics* (3) : 465 – 490.
- [8] Divjak S. & S. Gries. 2009. Corpus-based cognitive semantics: A contrastive study of phrasal verbs in English and Russian [A]. In K. Dziwirek & B. Lewandowska-Tomaszczyk (eds.) . *Studies in Cognitive Corpus Linguistics* [C]. Frankfurt: Peter Lang.
- [9] Firth J. 1951. Modes of meaning [A]. In G. Tilletson (ed.) . *Essays and Studies* [C]. London: John Murray Publishers Ltd.
- [10] Fries C. 1945. *Teaching and Learning English as a Foreign Language* [M]. Ann Arbor: The University of Michigan Press.
- [11] Geleyn T. 2017. Syntactic variation and diachrony: The case of the Dutch dative alternation [J]. *Corpus Linguistics and Linguistic Theory* (1) : 65 – 96.
- [12] Glynn D. 2014. Techniques and tools: Corpus methods and statistics for semantics [A]. In D. Glynn & J. Robinson (eds.) . *Corpus Methods for Semantics* [C]. Amsterdam: John Benjamins.
- [13] Gries S. 1999. Particle movement: A cognitive and functional approach [J]. *Cognitive Linguistics* (2) : 105 – 145.
- [14] Gries S. 2003. *Multifactorial Analysis in Corpus Linguistics* [M]. New York: Continuum.
- [15] Gries S. 2006. Corpus-based methods and cognitive semantics: The many senses of *to run* [A]. In S. Gries & A. Stefanowitsch (eds.) . *Corpora in Cognitive Linguistics* [C]. Berlin: Mouton de Gruyter.
- [16] Gries S. 2010. Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics [J]. *The Mental Lexicon* (3) : 323 – 346.
- [17] Gries S. & A. Adelman. 2014. Subject realization in Japanese conversation by native and non-native speakers [A]. In J. Romero-Trillo (ed.) . *Yearbook of Corpus Linguistics and Pragmatics* [C]. Cham: Springer.
- [18] Gries S. & T. Bernaisch. 2016. Exploring epicenters empirically: Focus on South Asian Englishes [J]. *English World-Wide* (1) : 1 – 25.
- [19] Gries S. & S. Wulff. 2013. The genitive alternation in Chinese and German ESL learners: Towards a multifactorial notion of context in learner corpus research [J]. *International Journal of Corpus Linguistics* (3) : 327 – 356.
- [20] Han W. , A. Arppe & J. Newman. 2013. Topic-marking in a Shanghainese corpus: From observation to prediction [J]. *Corpus Linguistics and Linguistic Theory* (2) : 1 – 29.
- [21] Guo H. & D. Chow. 2014. A corpus-based variationist approach to *bei* passives in Mandarin Chinese [J]. *Corpus Linguistics and Linguistic Theory* (1) : 139 – 173.
- [22] Hanks P. 1986. Typicality and meaning potentials [A]. In M. Snell-Hornby (ed.) . *Proceedings of the 2nd EURALEX International Congress* [C]. Tübingen: A. Francke Verlag.
- [23] Hanks P. 1996. Contextual dependency and lexical sets [J]. *International Journal of Corpus Linguistics* , (1) : 75 – 98.
- [24] Hanks P. 2013. *Lexical Analysis: Norms and Exploitations* [M]. Cambridge: The MIT Press.

- [25] Harris Z. 1954. Distributional structure[J]. *Word* (2 - 3) : 146 - 162.
- [26] Heylen K. J. Tummers & D. Geeraerts. 2008. Methodological issues in corpus-based Cognitive Linguistics [A]. In G. Kristiansen & R. Dirven (eds.) . *Cognitive Sociolinguistics* [C]. Berlin: Mouton de Gruyter.
- [27] Hopper P. 1987. Emergent grammar[A]. In J. Aske N. Beery J. Michaelis & H. Filip (eds.) . *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* [C]. Berkeley: Berkeley Linguistics Society.
- [28] Hilpert M. 2008. The English comparative: Language structure and language use [J]. *English Language and Linguistics* (3) : 395 - 417.
- [29] Janda L. & V. Solovyev. 2009. What constructional profiles reveal about synonymy: A case study of the Russian words for sadness and happiness [J]. *Cognitive Linguistics* (2) : 367 - 393.
- [30] Kemmer S. & M. Barlow (eds.) . 2007. *Usage Based Models of Language* [C]. Stanford: CSLI.
- [31] Labov W. 1966. *The Social Stratification of English in New York City* [M]. Washington D. C. : Center for Applied Linguistics.
- [32] Leech G. 2011. Principles and applications of Corpus Linguistics [A]. In V. Viana S. Zyngier & G. Barnbrook (eds.) . *Perspectives on Corpus Linguistics* [C]. Amsterdam: John Benjamins.
- [33] Levshina N. 2017. A multivariate study of T/V forms in European languages based on a parallel corpus of film subtitles [J]. *Research in Language* (2) : 153 - 172.
- [34] Murakami A. P. Thompson S. Hunston & D. Vajn. 2017. "What is this corpus about?": Using topic modelling to explore a specialised corpus [J]. *Corpora* (2) : 243 - 277.
- [35] Rosenbach A. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English [A]. In G. Rohdenburg & B. Mondorf (eds.) . *Determinants of Grammatical Variation in English* [C]. Berlin: Mouton de Gruyter.
- [36] Shank C. J. Van Bogaert & K. Plevoets. 2016. The diachronic development of zero complementation: A multifactorial analysis of the *that*/zero alternation with *think*, *suppose* and *believe* [J]. *Corpus Linguistics and Linguistic Theory* (1) : 31 - 72.
- [37] 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* (2) : 150 - 174.
- [38] Sinclair J. 2004. *Trust the Text: Language Corpus and Discourse* [M]. London: Routledge.
- [39] Speelman D. & D. Geeraerts. 2010. Causes for causatives: The case of Dutch "doen" and "laten" [A]. In T. Sanders & E. Sweetser (eds.) . *Causal Categories in Discourse and Cognition* [C]. Berlin: Mouton de Gruyter.
- [40] Szmrecsanyi B. 2003. *Be going to* versus *will/shall*: Does syntax matter? [J]. *Journal of English Linguistics* (4) : 295 - 323.
- [41] Wiechmann D. & E. Kertz. 2013. The positioning of concessive adverbial clauses in English: Assessing the importance of discourse-pragmatic and processing-based constraints [J]. *English Language and Linguistics* , (1) : 1 - 23.
- [42] Wulff S. 2003. A multifactorial corpus analysis of adjective order in English [J]. *International Journal of Corpus Linguistics* (2) : 245 - 282.
- [43] 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* (2) : 120 - 149.

(下转第 21 页)



- In G. Kristiansen & R. Dirven (eds.). *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems* [C]. Amsterdam: John Benjamins.
- [23] Talmy, L. 2000. Force dynamics in language and cognition [J]. *Cognitive Science* 24 (1): 49 – 100.
- [24] Verhagen, A. & S. Kemmer. 1997. Interaction and causation: Causative constructions in modern standard Dutch [J]. *Journal of Pragmatics* 27 (1): 61 – 82.
- [25] Wolff, P. 2003. Direct causation in the linguistic coding and individuation of causal events [J]. *Cognition*, 87 (1): 1 – 48.
- [26] 李临定. 1986. 现代汉语句型 [M]. 北京: 商务印书馆.
- [27] 刘婧 李福印. 2017a. 力动态视角下汉语“使”字句的语义构建——一项基于语料的调查 [J]. *语言学研究* 41 (1): 122 – 135.
- [28] 刘婧 李福印. 2017b. 致使义视角下的“使”字句及其英语表达形式——一项基于平行语料库的调查 [J]. *西安外国语大学学报* 47 (1): 39 – 45.
- [29] 陆军. 2010. 基于语料库的学习者英语近义词搭配行为与语义韵研究 [J]. *现代外语* 33 (3): 276 – 286.
- [30] 杨江锋. 2016. 基于语料库的原型理论视域下汉语迂回致使结构研究 [J]. *外语教学* 37 (2): 7 – 11.
- [31] 张佳易 胡健. 2011. 基于语料库的中国大学生英语迂回使役构式的习得研究 [J]. *合肥工业大学学报 (社会科学版)* 35 (1): 13 – 22.
- [32] 张炜炜 王芳. 2017. 从基于样例的概念空间看构式交替——以“让”和“给”的被动用法为例 [J]. *外语与外语教学* 38 (6): 22 – 33.
-
- (上接第 10 页)
- [44] Zhang, W. 2016. *Variation in Metonymy* [M]. Berlin: Mouton de Gruyter.
- [45] 方子纯 陈坚林. 2014. 基于语料库的同义形容词行为特征研究 [J]. *外语教学与研究* 42 (6): 842 – 852.
- [46] 房印杰. 2016. 语言学研究中的多因素分析 [J]. *语料库语言学* 1 (1): 82 – 92.
- [47] 房印杰 梁茂成. 2019. 可比语境下关系代词取舍的多因素分析 [J]. *外语教学与研究* 47 (3): 435 – 446.
- [48] 胡建华. 2018. 什么是新描写主义 [J]. *当代语言学* 40 (4): 475 – 477.
- [49] 邵健 张建理. 2017. 单宾语句次类判定及聚类的多维度量化研究 [J]. *解放军外国语学院学报* 40 (1): 61 – 67.
- [50] 徐大明. 1999. 新加坡华社双语调查——变项规则分析法在宏观社会语言学中的应用 [J]. *当代语言学* 1 (3): 25 – 35.
- [51] 许家金. 2014. 许家金谈语料库语言学的本体与方法 [J]. *语料库语言学* 1 (2): 35 – 44.
- [52] 许家金 陈哲. 2018. 中国大学生英语口语中不定式和动名词述补行为的多因素分析 [J]. *外语教学与研究* 46 (3): 380 – 392.
- [53] 徐晶凝 郝雪. 2019. 建议言语行为内部调节手段的语用调控 [J]. *世界汉语教学* 43 (3): 353 – 368.
- [54] 张懂 许家金. 2019. 英汉与格交替现象的多因素研究 [J]. *外国语* 72 (2): 24 – 33.
- [55] 张炜炜 N. Liu. 2015. 认知语言学研究的“实证周期” [J]. *外语研究* 33 (3): 18 – 23.
- [56] 张炜炜 刘念. 2016. 认知语言学定量研究的几种新方法 [J]. *外国语* 69 (1): 71 – 79.
- [57] 张炜炜 王芳. 2017. 从基于样例的概念空间看构式交替——以“让”和“给”的被动用法为例 [J]. *外语与外语教学* 38 (6): 22 – 33.