

基于语料库的学术英语程式语 研究与教学应用*

陆小飞¹ 刘颖颖²

摘要: 程式语是学术话语的关键特征之一,在学术英语的教学中起着至关重要的作用。本文首先从基于使用的语言习得理论出发,简要讨论心理语言学和语料库语言学研究揭示的程式语在语言习得中的基本作用,梳理不同种类程式语的操作化定义与提取方法,进而聚焦基于语料库的学术英语程式语研究,总结学术英语程式语教学资源建设成果及其潜在应用。文章最后重点阐述未来学术英语研究与教学中程式语形式分析、功能分析有机结合的重要性与可行性。

关键词: 学术英语话语; 程式语; 语料库分析

Abstract: Formulaic language, a prominent characteristic of academic discourse, plays a critical role in English for Academic Purposes (EAP) pedagogy. Drawing on the usage-based theory, this paper starts with a brief discussion of the fundamental importance of formulaic language in language acquisition as supported by psycholinguistic and corpus linguistic research. It then teases out various operational definitions and extraction methods of formulaic language, reviews corpus-based studies of formulaic language in academic English, and summarizes lists of formulaic expressions and their potential pedagogical applications. Finally, the paper highlights the importance and feasibility of the synergy of form-first and function-first approaches to formulaic language analysis in future EAP research and pedagogy.

Key words: academic English discourse; formulaic language; corpus-based analysis

中图分类号: H319 文献标识码: A 文章编号: 1004-5112(2019)05-0030-09

1. 引言

随着国际学术交流日益频密,各学科领域越来越多的研究人员需要具备英语文献阅读能力、学术英语写作和会话能力,高校和科研机构对学术英语教学与研究的需求持续增长。程式语(formulaic language)是学术英语话语的重要组构单位,学术英语程式语的掌握使用利于理解和构建语篇,提升语言的流利性、准确性和地道性。

在学术英语程式语研究中,基于语料库的程式语研究不仅考察了语言水平、语言使用经验、一语背景、体裁、学科等因素与程式语使用之间的关系(如 Cortes 2004; Römer 2009; Simpson-Vlach & Ellis 2010; O'Donnell *et al.* 2013),而且编制了各类程式语列表(如 Durrant 2009; Ackermann & Chen 2013; Cunningham 2017; Lei & Liu 2018),筛选出了具有较高教学价值的程式语。梳理和总结学术英语程式语研究成果有助于研究者更深入地了解程式语研究进展,也有助于学术英语教师根据教学需求选用合适的程式语教学资源。本文通过回顾和剖析基于语料库的程式语研究,发现已有研究在程式语修辞功能分析上仍有不足,进而讨论了学术

* 本文系国家社会科学基金一般项目“学术英语阅读的内容难度特征与智能模型研究”(编号 18BYY110)的阶段性成果之一。感谢《外语界》编辑部与匿名审稿专家的反馈意见。

英语研究与教学中程式语的形式分析、功能分析有机结合的重要性与可行性。

2. 程式语的作用、操作化定义与提取方法

2.1 程式语的作用

Wray (2002) 将程式语定义为连续或非连续的语言序列。心理语言学实验发现一语和二语学习者程式语的加工速度明显快于非程式语,因而将程式语视作大脑中整体存储和提取的语言单元(Jiang & Nekrasova 2007; Conklin & Schmitt 2008)。但是,近年来部分学者对这一观点提出了质疑。Sivanova-Chanturia (2015) 认为程式语较快的加工速度并不能证明程式语是整存整取、未经分析的语言单元。Conklin & Schmitt (2008) 发现低水平二语学习者加工程式语时,仍然采用逐词加工的方式。虽然程式语的“整存整取观”尚存争议,程式语在语言习得中的重要作用已经得到普遍认可。基于使用的语言习得理论认为,一语学习本质上是学习者在语言使用过程中逐渐积累语言实例(包括大量程式语),并从语言实例中总结归纳图式构式(schematic construction)的过程,遵循“程式语—低域模式—构式”的发展路径(Ellis 2002)。二语学习是否也遵循这一发展路径尚无定论,但程式语在二语习得中同样具有重要作用,因为程式语知识是学习归纳性构式的基础,程式语学习是基于语法的语言学习的重要补充,记忆并使用程式语能使学习者运用一些结构较为复杂、超出他们现有语法知识水平的语言形式(Myles 2004; Wulff 2019)。语料库语言学研究也证明了程式语在语言使用中占据重要地位。Erman & Warren (2000) 发现,程式语在口语和书面语语料库中各占 58.6% 和 52.3% 的比例。简而言之,程式语既是语言知识的重要组成部分,也是促进语言发展的重要条件。

心理语言学研究发现语言习得过程受语言形式的使用频率、突显性(salience)、语言标记(cue)和结果之间关联强度(contingency)等因素的影响(Ellis & Cadierno 2009; Ellis *et al.* 2015)。例如,一个语言单元只有被学习者注意到且在语言输入中的频率积累到一定程度时,才能被学习者习得;语言学习者很容易忽视突显性较低的语言特征(如第三人称单数形式);语言形式与功能之间的关联性越强,语言单元就越容易习得。基于类似研究成果,程式语教学应该尽可能提高程式语的输入频率,突出显示目标程式语,优先选择教授具有典型功能的程式语等。

2.2 程式语的操作化定义与提取方法

基于语料库的学术英语程式语研究提出了很多程式语操作化定义,包括搭配(collocation)、词束(lexical bundle)、程式(formula)、短语框架(phrase frame)等,不同操作化定义促成了不同研究成果。搭配指两个高于随机频率结伴同现的词。两个搭配词只要在一定上下文范围内同时出现即可,并不一定构成不间断的两词序列。Durrant (2009)、Ackermann & Chen (2013)、Lei & Liu (2018) 先后根据不同语料库编制了一般学术英语常用搭配列表。这三个搭配列表由于侧重不同的目标搭配,采用不同的搭配筛选方法,最终各自的内容存在差异。从搭配类型来看,Durrant (2009) 列表中的搭配大多是一个功能词加一个实义词的语法搭配,如 *this study, based on* 等。Ackermann & Chen (2013) 的列表限于词汇搭配,即形容词与名词或名词与名词搭配、动词与名词或动词与形容词搭配、动词与副词搭配等。Lei & Liu (2018) 的列表与 Ackermann & Chen (2013) 一样也是词汇搭配列表,但他们特意保证了各种搭配类型数量的均衡性。从语料来源来看,Durrant (2009) 的语料来源是五个学科领域的学术论文,Ackermann & Chen (2013) 的语料包括 28 个学科的学术论文和课本章节,而 Lei & Liu (2018)

选用的语料主要是期刊论文、书评以及英美大学应用语言学专业英语母语博士生的毕业论文。在搭配的认识与提取标准上,常用的语料库参数包括频率(frequency)、互信息值(MI)、t 值(t-value)、离散率(dispersion)、关键性(keyness)等。Durrant(2009)的研究以 BNC 语料库中的非学术语料作为比较基准,通过关键性选取学术论文库中的 1000 个关键搭配,同时筛选出五个学科领域中高频率、高互信息值的搭配。与 Durrant(2009)的研究相同,Ackermann & Chen(2013)在筛选搭配时也采用频率和互信息值,并额外增加了离散率和较高的 t 值参数($t \geq 4$),因为他们发现低于设定 t 值标准的搭配通常是非目标搭配的名词介词搭配或较长词组的片段。Lei & Liu(2018)采用的参数类型与 Ackermann & Chen(2013)一致,但提高了搭配的频率标准,降低了 t 值标准($t \geq 2$)。

除了两词搭配,多词短语也是基于语料库的学术英语程式语研究的重点之一,可被称为词束、程式、短语框架等。Biber *et al.* (1999)将高频连续词语序列称为词束。词束识别与提取的主要标准是形符频率和文本覆盖率(或离散率),因此词束可能含有两个意义单元(如 *a lot of the, in this essay I*),结构不一定完整。但是,Biber 等研究者认为词束是语篇的重要组成部分,具有语用功能,符合反复出现的交际需要。词束按照所含词语数量可以分为不同类型,以三词到六词长度词束的研究居多。不同研究提取词束的频率标准并不一致。以四词英语词束为例,频率标准从每百万词 10 次到 40 次不等(Biber *et al.* 1999; Biber *et al.* 2004)。识别词束的频率标准受语料库规模、词束长度、语域、学科等的影响。语料库规模越大,频率的选取标准越高。词束越长,频率标准越低。学术口语中词束的频率比书面语高(Biber *et al.* 1999; Biber *et al.* 2004),Biber & Barbieri(2007)认为这一差异根本上是由于学术口语和书面语需要表达的交际功能不同。针对不同学科领域,词束提取可以选用不同的频率标准,因为使用相同频率标准从不同学科文本中提取词束时,得到的词束数量会有差异。例如,Hyland(2008:9)发现同样使用每百万词 20 次的频率标准,在他构建的电子工程学术论文语料库中有 213 个高于这一频率标准的词束,而在相似规模的生物学学术论文语料库中高于这一频率标准的词束只有 131 个。Simpson-Vlach & Ellis(2010)对程式的定义接近词束,但他们识别程式时采用了频率和互信息值两个参数标准相结合的方式,创造性地使用了多元线性回归模型方法,采用频率和互信息值作为自变量,教师的教学价值分值作为因变量,并最终通过模型预测提取得到的所有程式的教学价值,根据教学价值对程式进行排序。他们指出,如果使用频率作为程式语的单一识别标准,会忽略部分由低频词组成、具有明确语篇功能和教学价值的程式(如 *due to the fact that*),却过度提取由高频词组成、没有明确语篇功能的程式(如 *it has been*)。

短语框架最初由 Stubbs 的研究生 Isabel Barth 提出(Fletcher 2002-2007),与搭配框架(collocational framework)概念相关(Stubbs 2007:90-91)。搭配框架指语法功能词包围实义词形成的非连续词组(如 *a * of*);短语框架是指存在一个变量位的非连续词语序列,涵盖范围显然比搭配框架更加广泛(Stubbs 2007:90)。搭配框架对词性有限制,短语框架并没有预先限制非变量位和变量位的词性。不同于词束和程式,短语框架可以是非连续的词语序列。比如,“*plays a * part in*”代表一个短语框架,其中星号(*)代表的变量位可用“*large, significant, major, vital*”等形容词(Stubbs 2007:90-91)填充,构成同一个短语框架类别。Fletcher(2002-2007)经 Stubbs 建议在 kfNgram 软件中引入了短语框架分析的相关功能,能够提取文本中含有一个变量位且长度不低于两词的短语框架,显示每个短语框架的总形符频次,显示变量位可填入的词语以及涵盖的具体短语的形符频次。这一分析工具在短语框架的语料库研究中得到广

泛应用(如 Römer 2009; Cunningham 2017; Lu, Yoon *et al.* 2018)。Römer(2009)揭示了不同水平学术语言使用者在短语框架使用上存在差异, Cunningham(2017)、Lu, Yoon *et al.* (2018)分别研制了针对数学学科和科研论文引言的学术英语短语框架列表。值得一提的是,已有研究尝试探讨含有一到两个变量位的短语框架并开发相关分析工具,如 AntGram 程序(Anthony 2018)。

由于程式语操作化定义的多样性,研究者关注了不同操作化定义及识别参数对程式语研究的影响。O'Donnell *et al.* (2013)在不同操作化定义下对比考察语言使用经验(分本科生、研究生、专家组别)和一语/二语背景对程式语使用的影响,分析结果是否存在差异。他们采用目标程式语(Simpson-Vlach & Ellis 2010)、n 元短语(分别使用频率和语料库外部互信息值参数两种识别方式)、短语框架等操作化定义,发现不同的程式语操作化定义导致本科生、研究生和专家组别的不同分析结果:采用目标程式语频率,专家组和本科生/研究生组之间存在显著性差异;采用 n 元短语频率,研究生/专家组和本科生组之间存在显著性差异;采用短语框架类型数,任意两组之间都不存在差异。Lu, Kisselev *et al.* (2018)重复了 O'Donnell *et al.* (2013)的研究,并分析了使用语料库内部互信息值识别所得的 n 元短语和短语框架的类形符比参数,发现分析结果与原研究基于语料库外部互信息值的结果一致,而短语框架的类形符比揭示了三个组别两两之间存在差异。Lu, Kisselev *et al.* (2018)也进一步分析了各参数代表的意义,指出语料库内部互信息值更适合识别语言使用者自己认定的程式化语言单位,而语料库外部互信息值是利用外部标准评价语言使用者多大程度上使用了程式化语言;短语框架的类形符比体现程式语的多样性,而频率、互信息值等参数与程式语使用频次有关。

3. 基于语料库的特定体裁和学科学术英语程式语研究

搭配、词束和程式列表都是基于多体裁、多学科的综合语料库研制而得,为一般学术英语教学和研究服务。部分学者对一般学术英语概念提出了质疑,指出不同体裁、不同学科需要实现不同的交际功能(Hyland & Tse 2007; Hyland 2008),实现交际功能的语言形式也各不相同,因此更应关注不同体裁、不同学科领域的学术英语(Flowerdew 2005)。作为学术话语的重要组成部分,程式语在不同体裁和学科领域中的使用必然会有差异。

Hyland(2016:120)将体裁简要界定为“抽象的、被社会认可的语言使用方式,且代表写作者对某些重复出现情景的典型反应”。每种体裁不仅含有特定语步和语阶组成的语篇结构,为读者、作者等提供一个共同的参考框架,而且含有实现语步和语阶的惯用语言形式(Swales 1990),包括惯用程式语。不同体裁使用的程式语存在差异。比如, Simpson(2004)对比了密歇根学术口语语料库中独白和会话两种学术口语体裁中的词束频率,发现有些词束在一个体裁中的频率显著高于另一体裁,而且学术会话体裁中高频词束的类符数量是学术独白体裁的两倍多。

除了体裁之间的程式语差异,各个学科都建立了各自约定俗成的学术话语表达惯例,具有不同的修辞结构和惯用语言形式。Durrant(2015)分析了英国学术英语笔语语料库(British Academic Written English Corpus)中的四词词束,发现通过词束差异就可区分四个主要的学科群。鉴于不同体裁和学科之间存在程式语使用差异,成功的学术语言使用者不仅要有体裁意识,了解特定体裁的表达范式,掌握符合目标体裁特征的程式语,而且还要了解特定学科的学术写作规范和惯例,才能使用恰当的语言形式有效传递意义(Hyland 2003, 2008)。

越来越多的研究者开始关注特定体裁和学科中的学术英语程式语。学术英语搭配列表的编者 Ackermann & Chen(2013)曾指出,学术英语搭配列表研究应关注各学科学术英语的高频搭配。一些学科专用语料库及学术英语程式语列表相继问世,程式语列表大多基于学术论文体裁研制而成,涉及数学、电子工程、商务研究、应用语言学、生物学、电子通讯、心理学等学科领域,采用词束、短语框架等不同程式语操作化定义(Cortes 2004; Hyland 2008; Grabowski 2015; Cunningham 2017)。这些资源呈现了特定体裁和学科中的重要程式语,能够帮助学术英语教师了解各个学科学术论文的语言和体裁特征,针对性地设计教学活动,引导学生注意、学习和使用程式语。学术英语程式语列表也可作为课外补充学习资源提供给不同学科背景的学习者,使他们了解相应学科常用程式语的形式和功能,练习使用程式语,提高学术英语阅读和写作能力。

4. 程式语形式与功能分析相结合的重要性和可行性

4.1 程式语形式与功能分析相结合的重要性

基于语料库的学术英语程式语研究成果丰硕,但很少揭示相关程式语在特定学术体裁中的修辞功能,导致程式语资源不能直接有效地运用于学术英语教学。具体而言,两词搭配的整体修辞功能往往不明确,因此搭配列表一般并不关注搭配的语义功能,条目通常依据参数值(如关键性)或搭配类型呈现。程式语的功能分类也较为宽泛,不够细致。例如, Biber *et al.* (2004)的词束功能分类法将词束分为立场类、语篇组织类和指称类,每类下面又包含一到两层子分类,但即使是最低一层类别的修辞功能范围仍然较广,如指称类词束最低一层的修辞功能类别包括地点指称、时间指称、文本索引、多功能指称四类。Simpson-Vlach & Ellis(2010)的程式语研究主要借鉴了 Biber *et al.* (2004)的词束功能分类法,只是合并、增加、扩充了一些类别。Hyland(2008)把从学术论文语料库中识别提取的词束分为面向研究、面向文本和面向参与者三类,每一类下又分为二至五个小类,如面向研究的词束分为时间或空间位置、过程、量化、描述和主题词束。这些程式语功能分类研究都采用自下而上、形式优先的分析方法,利用不同参数自动识别程式语,然后在脱离语境的情况下对程式语进行功能分析,并没有真正触及程式语在语篇中的具体修辞功能。与基于语料库的学术英语研究相反,专门用途英语领域尤其是体裁分析(genre analysis)领域研究一般采用自上而下、功能优先的分析方法,通过人工标注语步、语阶的方式描述和分析体裁结构,将语言形式的分析置于次要地位。究其原因,体裁分析理论认为体裁知识不仅是语言形式的知识,而且是关于语言形式和功能之间映射的知识(Swales 1990)。学术英语学习者在学术写作或口语表达过程中往往从交际功能出发,根据所需传递的意义和功能来选择合适的语言形式(包括程式语),因此体裁分析更为注重语言功能。

形式优先和功能优先两种研究范式产生了很多对学术英语教学具有重要价值的成果,然而对形式或功能的单一关注则会导致“功能—形式”空白(function-form gap)(Moreno & Swales 2018:41)。Flowerdew(2005:325)呼吁采用体裁分析和语料库分析相结合的方法研究学术英语话语,我们认为这种方法也应运用于程式语教学和研究,以揭示程式语的“形式—功能”映射,帮助学习者习得相关知识,在学术英语话语中有效使用程式语。在学术英语教学方面,程式语形式和功能分析相结合有助于学习者习得程式语。目前,二语学习者在学术英语程式语学习上仍有困难,他们的搭配知识并不一定随着词汇知识的增长而增长,即使是高水平二语学

习者也有可能缺乏搭配知识并产出不合规范的词语搭配(Yoon 2016)。在词束习得方面,Cortes(2004)对生物和历史学科领域的词束研究发现,母语英语研究生很少使用期刊论文中的常用高频词束,所用的某些高频词束在研究生学术写作中的语篇功能也与期刊论文中的不同。在短语框架习得方面,Römer(2009)考察了专家、英语本族语大四学生、英语非本族语大四及研一学生的学术英语写作,比较了不同组别最常用的三元和四元短语框架,结果发现专家语料库与学生语料库中的高频短语框架存在明显差异,这说明学生并未完全掌握学术英语中的常用短语框架。为了帮助学生掌握程式语形式,理解程式语的语篇修辞功能,习得程式语形式与功能之间的映射知识,学术英语教学有必要兼顾程式语形式和功能。

4.2 程式语形式与功能分析相结合的可行性

近年来,研究者为了填补“形式—功能”研究空白作出了有益尝试,证实了程式语形式和修辞功能分析相结合在学术英语研究中的可行性(Durrant & Mathews-Aydınlı 2011; Cortes 2013; Cunningham 2017)。Cortes(2013)研究了学术论文引言部分的词束,结合语步、语阶将词束的修辞功能分为明确领域相关性、话题推论、回顾以往研究等类别。Cunningham(2017)提取了数学学科学术论文中的180个五元和六元短语框架,并把其中38个核心短语框架按修辞功能分为概述、完成证明、连贯/组织、设置证明及变量等。这两项研究虽仍采用形式优先的方法,但都对程式语的修辞功能开展了分析。Durrant & Mathews-Aydınlı(2011)使用功能优先的方法,手动标注了一名研究生写作语料库中的修辞语步和语阶,识别每个语阶中的语法结构和程式语。程式语的形式和功能分析有机结合在学术英语教学中也具有可行性。

首先,已有研究提出了将程式语形式和功能分析相结合的实用教学方案,并验证了教学方案的有效性(Nattinger & DeCarrico 1992; AlHassan & Wood 2015; Chen & Flowerdew 2018)。Nattinger & DeCarrico(1992:169-170)的教学方案兼顾程式语的形式和功能,比如在学术写作教学中教师先教授给学生一些短语及其变体,然后学生熟悉目标体裁的语篇结构,再逐一学习不同语篇修辞功能涉及的主要短语(完成语步、语阶所需的主要短语,如 *it is/has been asserted/believed/noted that X* 用于“提出研究主题”, *this paper will compare/contrast/describe/demonstrate that X* 用于“指明研究题目”等),最后练习连接组织短语、句子写出结构完整的较长语篇(如引言部分)。AlHassan & Wood(2015)的实证研究证实了兼顾形式和功能的程式语教学法的有效性,他们从以往研究和美国当代英语语料库中选取了经济学科的目标程式语(包括搭配、词束和短语),实施了为期10周共计900分钟的辅导教学,通过加强对程式语的关注、理解程式语功能与重要性、练习使用目标程式语的句子写作等教学活动,促使学生学术写作中的程式语频率和种类显著增加。

需要注意的是,形式与功能相结合的程式语教学法可能会因教学时长、教学活动不同而产生不同教学效果。Cortes(2006)对母语英语的历史系本科生开展了五次(每次20分钟,共100分钟)课堂教学实验,实验活动主要是分析从历史学学术期刊论文中提取的35个高频词束的语篇功能,但是实验前后学生学术写作中的目标词束使用频率并没有显著变化。因此,她指出短期显性词束教学可能只能增强学生对词束频率和功能的意识,而程式语使用水平提高需要开展更长周期的教学项目。基于使用的语言习得理论也认为语言使用经验在语言习得中具有重要作用,某个语言形式在语言输入中需达到一定频次才可能被学生习得。除了教学时间较短致使输入频次较低,Cortes(2006)采用的教学活动忽视了语言输出练习,没有注重语言输出对二语习得的作用,这也可能是导致教学效果不显著的原因之一,当然教学效果的主要影响因

素尚需考察验证。

其次,基于语料库的学术英语程式语研究为学术英语教师和学生筛选出跨学科及多个特定学科的目标程式语,学术英语教师可以根据教学需求选择合适的教学资源,满足不同专业学生的学术英语学习需求。教师在教学活动中使用学术英语程式语列表时,可进一步细化程式语的功能分析,让学生真正理解并学会运用程式语。

第三,语料库资源可与数据驱动的教学法结合使用。语料库资源可以作为数据驱动学习的语言数据,已有功能标注的语料库能够更好地满足教学需求。Chen & Flowerdew (2018)在学术英语讲座中采用了数据驱动的教学法,引导学生使用 BNCWeb、学科语料库和自建语料库三种语料库来解决语言形式和修辞功能的相关问题,得到了较好的教学反馈。Charles (2007)也证实使用自建语料库能够有效帮助学生学术英语写作中的多词短语、句式等语言形式及其修辞功能。Cotos *et al.* (2017)同样采用了数据驱动教学法,但与其他大多数研究不同,他们使用的是计算机辅助写作软件和已标注语篇修辞功能的学术英语语料库,学生需要检索语料库来回答研究者设计的语言形式和功能问题,并提交学术写作文章,研究结果同样支持数据驱动的学术英语教学法。值得注意的是,有的研究特别强调教师控制在数据驱动教学法中的重要性,指出教学活动应由教师提前设计、控制和引导,并非完全倚赖学生自主性(Charles 2007)。

5. 结语

学术英语程式语使用不仅需要掌握程式语的语言形式,还要掌握语言形式与修辞功能之间的映射、学术语篇的修辞结构、体裁的形式与功能特征、学科惯例等,因此学术英语程式语教学与研究需要兼顾程式语的形式与语篇功能。基于语料库的学术英语程式语研究已经获得丰富成果,但自下而上、形式优先的研究方法导致程式语功能分析的不足,标注语篇修辞功能的程式语教学资源 and 程式语教学相关实证研究仍然较为匮乏,学生在程式语形式和功能习得方面还存在困难。今后学术英语研究者与教师应该更加注重程式语形式与功能分析的有机结合,帮助学生充分理解和有效习得程式语,推动学术英语教学与研究深入发展。

参考文献

- [1] Ackermann K & Chen Y-H. Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach [J]. *Journal of English for Academic Purposes*, 2013, 12(4): 235–247.
- [2] AlHassan L & Wood D. The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study [J]. *Journal of English for Academic Purposes*, 2015, 17: 51–62.
- [3] Anthony L. *AntGram* (Version 1.0.0) [CP]. Tokyo: Waseda University, 2018.
- [4] Biber D & Barbieri F. Lexical bundles in university spoken and written registers [J]. *English for Specific Purposes*, 2007, 26(3): 263–286.
- [5] Biber D, Conrad S & Cortes V. If you look at ...: Lexical bundles in university teaching and textbooks [J]. *Applied Linguistics*, 2004, 25(3): 371–405.
- [6] Biber D *et al.* *Longman Grammar of Spoken and Written English* [M]. London: Longman, 1999.
- [7] Charles M. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions [J]. *Journal of English for Academic Purposes*, 2007, 6(4): 289–302.

- [8] Chen M & Flowerdew J. Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong [J]. *English for Specific Purposes*, 2018, 50: 97–112.
- [9] Conklin K & Schmitt N. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? [J]. *Applied Linguistics*, 2008, 29(1): 72–89.
- [10] Cortes V. Lexical bundles in published and student disciplinary writing: Examples from history and biology [J]. *English for Specific Purposes*, 2004, 23(4): 397–423.
- [11] Cortes V. Teaching lexical bundles in the disciplines: An example from a writing intensive history class [J]. *Linguistics and Education*, 2006, 17(4): 391–406.
- [12] Cortes V. The purpose of this study is to: Connecting lexical bundles and moves in research article introductions [J]. *Journal of English for Academic Purposes*, 2013, 12(1): 33–43.
- [13] Cotos E, Huffman S & Link S. A move/step model for methods sections: Demonstrating rigour and credibility [J]. *English for Specific Purposes*, 2017, 46(1): 90–106.
- [14] Cunningham K J. A phraseological exploration of recent mathematics research articles through key phrase frames [J]. *Journal of English for Academic Purposes*, 2017, 25: 71–83.
- [15] Durrant P. Investigating the viability of a collocation list for students of English for academic purposes [J]. *English for Specific Purposes*, 2009, 28(3): 157–169.
- [16] Durrant P. Lexical bundles and disciplinary variation in university students' writing: Mapping the territories [J]. *Applied Linguistics*, 2015, 38(2): 165–193.
- [17] Durrant P & Mathews-Aydinli J. A function-first approach to identifying formulaic language in academic writing [J]. *English for Specific Purposes*, 2011, 30(1): 58–72.
- [18] Ellis N C. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition [J]. *Studies in Second Language Acquisition*, 2002, 24(2): 143–188.
- [19] Ellis N C & Cadierno T. Constructing a second language [J]. *Annual Review of Cognitive Linguistics*, 2009, 7: 111–139.
- [20] Ellis N C, O'Donnell M B & Römer U. Usage-Based language learning [A]. In MacWhinney B & O'Grady W (eds.). *The Handbook of Language Emergence* [C]. Hoboken, NJ: Wiley-Blackwell, 2015. 163–180.
- [21] Erman B & Warren B. The idiom principle and the open choice principle [J]. *Text*, 2000, 20(1): 29–62.
- [22] Fletcher W H. *Kj/Ngram* [CP]. Annapolis, MD: USNA, 2002–2007.
- [23] Flowerdew L. An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies [J]. *English for Specific Purposes*, 2005, 24(3): 321–332.
- [24] Grabowski Ł. Phrase frames in English pharmaceutical discourse: A corpus-driven study of intra-disciplinary register variation [J]. *Research in Language*, 2015, 13(3): 266–291.
- [25] Hyland K. Genre-based pedagogies: A social response to process [J]. *Journal of Second Language Writing*, 2003, 12(1): 17–29.
- [26] Hyland K. As can be seen: Lexical bundles and disciplinary variation [J]. *English for Specific Purposes*, 2008, 27(1): 4–21.
- [27] Hyland K. Methods and methodologies in second language writing [J]. *System*, 2016, 59: 116–125.
- [28] Hyland K & Tse P. Is there an “academic vocabulary”? [J]. *TESOL Quarterly*, 2007, 41(2): 235–253.
- [29] Jiang N & Nekrasova T M. The processing of formulaic sequences by second language speakers [J]. *The Modern Language Journal*, 2007, 91(3): 433–445.
- [30] Lei L & Liu D. The academic English collocation list: A corpus-driven study [J]. *International Journal of Corpus Linguistics*, 2018, 23(2): 216–243.

- [31] Lu X, Kisselev O, Yoon J & Amory M D. Investigating effects of criterial consistency, the diversity dimension, and threshold variation in formulaic language research: Extending the methodological considerations of O'Donnell *et al.* (2013) [J]. *International Journal of Corpus Linguistics*, 2018, 23(2): 158-182.
- [32] Lu X, Yoon J & Kisselev O. A phrase-frame list for social science research article introductions [J]. *Journal of English for Academic Purposes*, 2018, 36: 76-85.
- [33] Moreno A I & Swales J M. Strengthening move analysis methodology towards bridging the function-form gap [J]. *English for Specific Purposes*, 2018, 50: 40-63.
- [34] Myles F. From data to theory: The over-representation of linguistic knowledge in SLA [J]. *Transactions of the Philological Society*, 2004, 102(2): 139-168.
- [35] Nattinger J R & DeCarrico J S. *Lexical Phrases and Language Teaching* [M]. Oxford: Oxford University Press, 1992.
- [36] O'Donnell M, Römer U & Ellis N C. The development of formulaic sequences in first and second language writing: Investigating effects of frequency, association, and native norm [J]. *International Journal of Corpus Linguistics*, 2013, 18(1): 83-108.
- [37] Römer U. English in academia: Does nativeness matter? [J]. *Anglistik: International Journal of English Studies*, 2009, 20(2): 89-100.
- [38] Simpson R C. Stylistic features of academic speech: The role of formulaic expressions [A]. In Connor U & Upton T A (eds.). *Discourse in the Professions: Perspectives from Corpus Linguistics* [C]. Amsterdam: John Benjamins, 2004. 37-64.
- [39] Simpson-Vlach R & Ellis N C. An Academic Formulas List: New methods in phraseology research [J]. *Applied Linguistics*, 2010, 31(4): 487-512.
- [40] Siyanova-Chanturia M. On the "holistic" nature of formulaic language [J]. *Corpus Linguistics and Linguistic Theory*, 2015, 11(2): 285-301.
- [41] Stubbs M. An example of frequent English phraseology: Distribution, structures and functions [A]. In Facchinetti R (ed.). *Corpus Linguistics 25 Years On* [C]. Amsterdam: Rodopi, 2007. 89-105.
- [42] Swales J M. *Genre Analysis: English in Academic and Research Settings* [M]. Cambridge: Cambridge University Press, 1990.
- [43] Wray A. *Formulaic Language and the Lexicon* [M]. Cambridge: Cambridge University Press, 2002.
- [44] Wulff S. Acquisition of formulaic language from a usage-based perspective [A]. In Siyanova-Chanturia A & Pellicer-Sánchez A (eds.). *Understanding Formulaic Language: A Second Language Acquisition Perspective* [C]. New York: Routledge, 2019. 19-37.
- [45] Yoon H-J. Association strength of verb-noun combinations in experienced NS and less experienced NNS writing: Longitudinal and cross-sectional findings [J]. *Journal of Second Language Writing*, 2016, 34: 42-57.

作者单位: 1. 广东外语外贸大学英语语言文化学院, 广东 广州 510420; 宾夕法尼亚州立大学应用语言学系, 宾夕法尼亚 帕克 16802
2. 宾夕法尼亚州立大学应用语言学系, 宾夕法尼亚 帕克 16802