

语料驱动的高校外语教学研究

宋月霜

黑龙江大学外国语学院(区域国别学院), 黑龙江 哈尔滨 150000

摘 要 : 本文基于英美两个大规模语料库 BNC 和 COCA, 分析探讨了语料驱动的词汇教学、语料驱动的语法教学、语料驱动的语篇教学和语料驱动的思政教学内容、路径和方法, 为高校外语教学实践提供了有益指导。

关 键 词 : 语料驱动; 词汇教学; 语法教学; 语篇教学; 思政教学

Study on Corpus-driven Approaches to Language Teaching

Song Yueshuang

School of Foreign Languages and Literature (School of Region and Country Studies), Heilongjiang University, Harbin, Heilongjiang 150000

Abstract : This article utilizes two large-scale corpora, BNC and COCA, to analyze and explore corpus-driven language teaching content and approaches, including vocabulary teaching, grammar teaching, discourse teaching, ideological and political education. The findings of this article offer valuable guidance for enhancing foreign language instruction in higher education institutions.

Keywords : corpus-driven; vocabulary teaching; grammar teaching; discourse teaching; ideological and political education

一、语料库与语料库语言学

语料库 (corpus) 是指以电子形式存储的大量自然发生语言实例的文本集合, 这些语言实例的汇总和整理有明确的目的和方法 (Gena R. Bennett, 2010: 2)。语料库语言学是以语料库为手段研究人类语言现象。语料库语言学通过收集、存储和分析大量来自于不同语域的自然语言文本, 揭示语言的结构、功能和使用模式。所谓语域是指语言使用的具体情境, 例如口语或书面, 书面还会细分为学术论文、小说、新闻报道等不同语域。在语料库研究中, 语域和体裁两个术语经常互换使用 (Gena R. Bennett, 2010: 11)。语料库语言学主要关注语言使用的实际模式、语域间的语言变体、语言的语义语用分析等。语料库能够为语言教学提供客观可靠的数据支撑和科学有效的数据分析, 是外语教师的得力助手, 为外语教学提供智力支持。

二、语料驱动的外语教学研究

1. 语料驱动的词汇教学研究

传统的课堂词汇教学模式是教师领读发音, 解释意义, 提供例句, 词汇知识的呈现是字典模式, 单一且固化, 无法体现目标词汇在不同语域中的使用表现。学生不了解目标词汇常用于书面语还是口语? 多见于学术语篇还是非学术语篇? 适用于正式文体还是非正式文体? 词汇搭配偏好? 同义词之间的细微差别? 学习语言, 就是要学习使用中的语言, 学习本族语者在现实世界交流中究竟怎样运用语言? 语言运用的规律和模式究竟有哪些? 语料

驱动的词汇教学能够基于统计数据提供上述问题答案, 语料库呈现的词汇信息是全面、系统、多元化的。

在词汇教学实践中, 基于语料库 (如 Corpus of Contemporary American English 当代美国英语语料库: COCA 包含有超过 10 亿个词, 涵盖了 20 世纪 90 年代到现在的美国英语) “WORD” 功能栏检索目标词汇, 如 achieve, 我们可以向学生呈现以下信息: achieve 在不同语域中出现的频次信息, 如它作为动词在学术体裁中使用频次最高; achieve 各种屈折形式, 如 achieves, achieved, achieving 使用次数统计信息; achieve 的定义、网络在线词典的定义、示意图片、发音视频 (YouGlish, PlayPhrase, Yarn) 以及汉译 (Google, Reverso, Linguee); achieve 同义词信息, 如 attain, complete, realize, succeed 等; achieve 使用的相关主题信息; achieve 搭配信息, 包括名词、动词、形容词以及副词的惯常搭配对象, 与 achieve 惯常搭配的名词可以是 goal, success, result, objective, level, effect, dream, status; achieve 的派生词汇, 如 achievement, achievable, achiever; 包含有 achieve 2-4 词的词群, 左右出现的搭配词情况; achieve 频繁出现的文本和网站; achieve 索引行信息, achieve 左侧右侧连续 3 词被标以不同颜色, 提供词性信息。基于索引行, 可以观察 achieve 一词在具体语境中的使用特征。

在词汇教学实践中, 基于语料库 (如 British National Corpus 英国国家语料库: BNC 包含有约 1 亿个词, 涵盖了 20 世纪后期的英国英语) “COMPARE” 功能栏, 可以衡量同义词的搭配偏好和分布差异, 例如输入 empty 和 vacant 两个词, 条件设定为右侧搭配 1 个名词的情形, 结果显示: 一是 empty 在该语料库

基金项目:

1. 黑龙江大学新世纪教育教学改革工程项目《语料库语言学视域下的高校外语教学理论与实践》研究成果, 项目编号 2022C65。

2. 黑龙江大学新世纪教育教学改革工程项目《大学英语教学中的课程思政途径探析》研究成果, 项目编号 2022C66。

作者信息: 宋月霜, 女, 1979 年生于大庆, 黑龙江大学外国语学院 (区域国别学院) 副教授, 硕士研究生, 研究方向: 语料库语言学、外语教学、特殊目的英语。

中使用5333次, vacant使用785次, 前者是后者的6.79倍, 说明 empty 使用频率远远高于 vacant; 二是基于两词的相对频率比率 (Relative Frequency Ration) 或者互为参照比率 (Mutual Reference Ration), 按照从高到低排序, 与 empty 搭配排在前三位的名词分别是 glass, stomach 和 bottle, 而 vacant 与上述三词搭配统计次数为0。与 vacant 搭配排在前三位的名词分别是 possession, post 和 title, 而 empty 与上述三词搭配统计次数为0, 能够揭示 empty 更倾向于描述物理状态的“空”, 与之搭配的多为有形容器, 而 vacant 更多表示未被占用, 多用于功能性空缺的语境中, 如房屋状态 (vacant possession)、职位状态 (vacant post) 以及头衔状态 (vacant title) 等, 多用于正式或专业语境中, 如在房地产领域, vacant possession 表示空置交付, 即房产没有人、物品或使用权占有权所形成的障碍; 三是有些名词如 house 既可以与 empty 搭配也可以与 vacant 搭配, empty/vacant 相对频率比率为 $83/4=20.8$, 说明 empty house 比 vacant house 更常用, 前者出现频率是后者的20.8倍。另外, 两种表达传递的语境意义不同, empty house 强调空无一物或短期无人居住, 而 vacant house 强调法律或用途上的空置或长期无人居住。基于语料库教师可以帮助学习者厘清同义词使用的具体语境和语义范围等细微差异, 基于真实语料 (如索引行) 创建同义词辨析习题, 加强巩固词汇使用偏好。

2. 语料驱动的语法教学研究

在语法教学层面, 首先来看时态的分布和功能。以过去时教学为例, 普遍遵循的讲授模式是: 解释时态表征的意义; 展现过去时的句子结构 (肯定句、否定句、疑问句等); 讲解英语动词过去时的屈折变化 (规则与不规则变化); 提示一般过去时常见的时间状语等。但是课堂教学实践中, 并没有向学生解释动词的时态偏好, 所谓时态偏好是指动词在不同的语境或语境中以哪种时态出现的频率更高, 以了解动词在现实世界中使用的时态分布特征和规律。基于语料库统计, know, matter 和 suppose 等动词现在时使用占比高, 而 smile, reply 和 pause 等动词过去时使用占比高 (Biber et al., 1999: 459)。基于 COCA 语料库, 查询 know 动词时态分布, 发现 know 使用 2110629 次, knew 使用 278027 次, known 使用 174757 次, knows 使用 138180 次, knowing 使用 59778 次, knowed 使用 257 次, 其中现在时占比 81.4%; 查询 pause 动词时态分布, 发现 paused 使用 13248 次, pause 使用 4093 次, pauses 使用 2951 次, pausing 使用 2513 次, 过去时占比 58.1%。不同的动词所出现的时态偏好受到语义、语境以及句法功能等因素影响: 状态动词、真理动词等倾向于一般现在时, 状态动词表达相对稳定的状态, 而真理动词用于描述普遍规律、科学事实, 不受时间的约束。而动作动词或叙述性动词等倾向于一般过去时, 因为动作动词经常用来表示已经完成的事件, 叙述性动词用于描述过去发生的事件, 因此受到时间的约束。从语域分布来看, 在学术语域一般现在时使用频率高, 而在小说语域一般过去时使用频率高。基于 BNC 或 COCA 语料库发现新闻语域的时态偏好受到报道内容影响和制约, 如标题现在时、正文事件过去时、动态事件进行时、未来走向将来时。时态教学应该着眼于应

用, 基于语料库的统计, 指导学生正确理解、运用和把握时态, 知其然知其所以然。

在语法教学层面, 其次看句法结构的分布和功能。我们以主语从句及其外置结构为例, 探讨它们在不同语境中的分布情况以及使用特征。所谓主语从句外置 (extrapose) 是指将主语从句移到句末, 句首添加形式主语 it。基于语料统计, 在所有的语境中 that 主语从句是非常罕见的。在学术文体中, that 主语从句每百万词仅出现 5-10 次, 外置结构 20 多次。在口语语料库中, that 主语从句几乎不存在。外置结构之所以数量居多, 源于两个原因: 一是句尾信息焦点原则, 确保关键信息得到突出; 二是基于认知加工省力原则, 外置结构能够使读者迅速捕捉句子主要谓语动词, 便于理解。但是基于以下两种情况, that 主语从句应置于句首, 而非外置: 一是句子的谓语部分是较长的复杂结构, 其中包含有另一个 that 从句; 二是 that 主语从句陈述已知信息—前文提及或已经确立的事实。基于语料库统计发现, 对教学的启示是: 一是帮助学生理解掌握 that 主语从句所适用的目标语境; 二是向学生解释在何种具体语境下使用 that 主语从句及其外置结构, 并设计相应练习进行操练, 以此帮助他们在日后的写作任务中适切地使用该从句结构 (Biber et al., 1998: 77-82)。

3. 语料驱动的语篇教学研究

语篇教学首先关注语步模式的体裁或语域间差异。BNC 和 COCA 语料库覆盖多种体裁或语域, 包含口语以及书面语, 如博客、网页、小说、杂志、新闻报道以及学术语篇等。基于语料库真实语篇的呈现, 可以研究目标体裁的语步结构。体裁是有步骤、有目的的活动类型和交际事件。日常教学中以学术论文为例, 借助学术英语语料库, 明确不同学科间的学术论文语步结构差异, 学生所在学科科研论文的语步模式, 科研论文各个组成部分的语步特征, 如引言、研究方法、研究结果、讨论甚至致谢的语步特征。提高学习者学术话语能力, 使其学术话语符合国际同行认可的体裁模式和语言模式, 是外语教学的重要使命。

语篇教学其次关注语篇性问题。衔接和连贯是语篇的重要属性, 也是定义语篇的重要特征。连贯涉及的是内容的隐性逻辑关联, 衔接涉及的是词汇语法的显性形式关联。Halliday 系统梳理过衔接的各种词汇和语法手段, 指称 (reference) 便是其中重要的语法衔接机制。基于语料库, 可以统计分析不同语境语篇间的指称类型以及存在的差异。Douglas Biber 等 (1998: 116-118) 比较了四种语境 (对话、讲演、新闻报道和学术写作) 三种指称类型 (外指代词指称、回指代词指称以及回指名词指称) 分布: 一是根据篇章中每 200 词出现的指称数量统计, 新闻中使用指称数量最多, 学术写作最少, 对话和讲演居中持平; 二是从指称代表的新旧信息分布来看, 对话中有超过 70% 的旧信息指称, 相反学术写作有超过 65% 的新信息指称; 三是从表达旧信息的指称类型来看, 新闻报道和学术写作外指代词指称十分罕见, 除作者偶尔使用 I 以外, 外指代词指称几乎是没的, 书面语域很大程度上依赖于回指。但书面和口语语境所偏好的回指名词形式是不同的: 对话和讲演使用回指代词和完整名词数量相当, 而书面语域绝大多数情况下使用完整名词进行回指。在学术英语写作教学中, 可

以基于语料库统计数据, 指导学生正确运用指称, 包括学术语篇指称运用的总体比率、学术语篇指称表达新信息的偏好以及学术语篇指称表达旧信息时回指名词的使用偏好等。帮助学生把握学术写作指称使用规律, 符合国际规范, 提升学术交际能力。

三、语料驱动的思政教学研究

作为外语教育工作者, 始终要思考培养什么人、为谁培养人、怎样培养人的问题。过去几十年外语教学一直秉持学习西方语言了解西方文化的宗旨, 而如今随着中国在世界崛起, 我们要让世界听到中国声音, 提升大学生运用英语讲述中国发展变化、讲述中国故事的能力, 赋予他们理论自信、道路自信和文化自信。上海外国语大学语料库研究院推出《习近平谈治国理政》多语种数据库综合平台, 为以《习近平谈治国理政》及其外译为教学内容的课程思政教学提供了极大便利和强大支撑。

《习近平谈治国理政》多语种数据库包含语料库、知识库和文献库。语料库教学模块不仅展示中英平行语篇, 而且提供翻译

策略与技巧查询。教师可以基于语料库平台提供的翻译实例, 讲解词性转换、归化、具体化、省译、视角转换、虚化、异化和增译等具体的翻译方法。语料库科研模块提供基础信息、词频、KWIC、索引行、词簇及搭配等语言数据的查询, 例如利用KWIC查询关键词 achieve, 会出现 achieve the great renewal, achieve socialist modernization, achieve new victories, achieve all-round development, achieve the rejuvenation, achieve prosperity 等着眼于中国国情、具有中国特色的表达, 在教学实践中, 基于语料数据教师为学生梳理总结 achieve 搭配组合模式, 形成词块记忆。知识库总结梳理了相关术语以及中华典故的英译, 译文权威, 方便学生查找使用。文献库模块支持按标题、作者、来源、出版时间等检索项进行单一检索或复合检索, 并支持查看相关中英文文本文献、影像资料、图片资料等, 此部分有助于扩展学习内容, 丰富语言接触。

语料库语言学研究方兴未艾, 外语教师要积极利用语料库以及语料库语言研究成果提升教育教学质量, 建立数字化、智能化课堂, 依托真实语言数据改进教学内容和方法。

参考文献

-
- [1]Gena R. Bennett. 2010. *Using Corpora in the Language Learning Classroom*. Michigan, USA: The University of Michigan Press.
- [2]Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. New York: Longman.
- [3]Douglas Biber, Susan Conrad & Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press.
- [4]宋月霜 2022 EFL学习者学术话语能力研究 中国社会科学报 .