

科技英语语料库的建设与应用

——一项基于航空英语语料库的研究

赵丽珠, 王爱国

(中国民航大学 外国语学院, 天津 300300)

摘要: 目前我国出版了大量科技英语教材, 但其编写一般是基于知识体系。学生常常会误认为科技英语是以知识为导向的课程。科技英语语料库的引入可以作为教材的有效补充, 为以学生为中心的教学活动提供有力支持。基于航空英语语料库的案例研究, 探索了语料库在科技英语词汇、语法教学中的实际应用, 指出语料库确实能够帮助学生将关注点落实到语言上。此外, 对基于语料库开展科技英语教学的远景和挑战进行了探讨。

关键词: 语料库; 科技英语; 航空英语

中图分类号: H 319.3

文献标志码: A

文章编号: 1009-895X(2014)03-0201-06

DOI: 10.13256/j.cnki.jusst.sse.2014.03.001

On the Establishment and Application of English for Science and Technology Corpora

—A Research Based on an EST (Aeronautics) Corpus

Zhao Lizhu, Wang Aiguo

(College of Foreign Languages, Civil Aviation University of China, Tianjin 300300, China)

Abstract: A large amount of EST teaching materials have been currently published in China. Generally, these materials are organized based on the framework of professional knowledge. Therefore, students often take it for granted that EST courses are knowledge-oriented. However, the author thinks that an introduction of corpora can be an effective supplement to EST teaching materials and contribute to student-centered activities in classroom. A case study based on EST (Aeronautics) corpus explores the application of corpora in EST vocabulary and grammar pedagogy. It concludes that EST corpora can indeed help students to focus more on language itself. Moreover, prospects and challenges of applying corpora in EST teaching are discussed in this paper.

Key words: corpus; English for science and technology; English for Aeronautics

语料库研究兴起于20世纪60年代, 其在语言学各领域所引发的革命, 引起了语言学家、教育学家的极大关注。语料库研究以及基于语料库的语言研究, 正逐渐成为语言学研究领域的话语中心, 其

涵盖范围及应用领域也日益广泛^[1]。

语言教学是语料库主要应用领域之一。语料库应用于语言教学的研究课题集中在: 语料库指导的方法(根据词语、搭配、结构使用的场合、出现

收稿日期: 2013-11-10

基金项目: 中央高校基本科研项目“基于语料库的航空英语教学模式研究”资助(ZXH2012F008)

作者简介: 赵丽珠(1982-), 女, 讲师。研究方向: 专门用途英语教学、语料库语言学。E-mail: judithzhao666@163.com

的频率和语体等信息编写合适的教学材料)、基于语料库的语言描写(利用语料库验证研究者所提出的假设)及语料库驱动的语料分析(通过观察得出语言使用情况的假设乃至结论)^[2]。

专门用途英语(English for Specific Purposes, 以下简称 ESP), 作为高等教育语言教学中重要的一个分支, 是一种“语言学习为目的服务的教学”^[3]。其教学目标是培养学生在特定的专业语境下具备运用英语交际的能力。在 ESP 众多分支中, 科技英语(English for Science and Technology, 以下简称 EST) 又是规模最大、发展相当活跃的一支。Peter Strevens 就曾强调要注意总结 EST 的特点, 以便将其应用于语言教学^[4]。而在 EST 教学实施过程中, 因为缺乏素材, 教师倾向于以个人经验为基础进行教学, 而实际上这样的“经验之谈”很容易对学生产生误导。学生对 EST 语言特征敏感性不足, 将符合特定交流机制的 EST 应用于实际就更无从谈起。

此时, 将 EST 语料库引入课堂就显得尤为重要了。但由于 EST 涵盖领域甚广(如航空英语、力学英语、水利英语等等), EST 教师常常尴尬地发现, 无法找到可供其使用的特定学科的语料库, 也就无法将其应用于教学。实际上, 教师若能了解语料库建库的原则和方法, 便可自建小型语料库用于特定语域 EST 的教学。在某些领域内搜集的真实语料, 其语言机制既符合科技英语的普遍规律和共同特点, 又能体现出该领域特有的词汇分布。本文将基于自建小型航空英语语料库, 探讨语料库在 EST 教学中的实际应用, 并对这种教学模式的优势与局限性进行探讨。

一、语料库的构建

(一) 建库原则

EST 语料库的建立既需考虑到 EST 的教学特点, 又需遵守语料库的建库基本原则。

(1) 代表性: 应当在对学生的语言运用进行需求分析的基础上建立 EST 语料库, 根据学习者未来职业背景下展开有效交际的需要, 选择语料; 所选取的语料应当在内容、体裁上最大程度地涵盖教学的各个方面。

(2) 真实性: 使用真实的语料作为教材是 EST 教学基本原则之一, 练习设计和课内外活动都应体现专用英语的社会文化情境^[5]。因此, EST

语料库应当选择特定主题领域(如: 航空英语)的真实语料。

(3) 操作性: EST 语料库的使用对象是普通的教师和学生, 因此应当具备较强的可操作性, 方能有效地应用于 EST 教学。

(二) 航空英语语料库的建立

航空英语教学的对象今后主要从事工程师、科学研究等技术相关的工作, 根据其将要面对的工作环境, 对其职业背景下交际的需要进行分析: 包括可以阅读专业期刊的文献; 能够用英文撰写技术报告和科学论文; 能够阅读专业手册和说明书; 能够听懂专业学术报告; 可以用英语参加学术会议的讨论等。

基于以上需求分析, 选择航空英语语域、不同体裁的语料, 建立小型语料库。各体裁包括文本篇数及形符数见表 1。

表 1 自建航空英语语料库各类体裁篇数及形符数

Tab.1 Genres and tokens of the self-built aeronautics corpus

体裁	册/篇数	形符数
原版教材	3	156 280
航空法律法规	3	42 092
专业期刊文献摘要	16	2 809
航空新闻	20	6 372
操作手册	1	144 906
总计	43	352 459

二、自建小型语料库在 EST 教学中的应用

目前, 学术界对于 EST 的研究积累了大量的成果, 但许多研究仍然停留在理论层面, 很多 EST 教师对教什么和如何教普遍感觉茫然。EST 教材的局限性是造成这种状况的重要原因之一。目前我国 EST 教材的编写依然以知识体系为主要框架结构。以飞机发动机为例, 教材一般分为活塞式发动机和喷气式发动机, 再对两种类型发动机的主要组成部分、工作原理、流程进行详细介绍。完全依靠这样的 EST 教材进行课堂教学, 会使学生产生 EST 的教学是以知识为导向的错误认识, 如何令学生对 EST 的关注落实到语言上, 是 EST 教师在教学中亟待解决的问题。

自建语料库为 EST 教学提供了大量的素材, 可以作为目前 EST 教材的有效补充。教师引导学生针对教材中的某一话题、词汇或语法现象, 基于语料库进行检索, 可大大拓展阅读范围、深入学习词汇及其译法、加强对词语搭配的关注, 并最终提高对 EST 语言交流机制的敏感性。这种“发现式学习”的方法将学生从传统的教学模式及 EST 教材的桎梏中解放出来, 极大地激发了学生的学习兴趣, 从而形成自主学习的模式。

(一) 语料库驱动下的词汇教学

1. 检索平行语料库, 学习词汇及其译法

Averil Coxhead 指出, (对于学术英语) 教师应当教授与学习者相关的词汇, 并最先教授出现频数最高的词汇^[6]。“与学习者相关”, 也即满足学习者需求的, 在语料选择、语料库构建的过程中已经得以实现。而对于出现频数最高的词汇, 则可以利用语料库的频率分析功能, 列出典型词汇和短语。基于自建航空英语语料库制作的高频词表, 可以看到频数最高的词汇为冠词、介词等功能词, 这与基于英国国家语料库 (British National Corpus, 以下简称 BNC) 统计的高频词十分吻合, 也说明在一定程度上 EST 体现出与通用英语一致的词汇分布。但这样的功能词对于词汇教学意义有限 (但可能揭示某种语体特点), 因此可以通过与通用英语语料库的对比、剔除高频词表中的功能词, 从而得到特定学术领域的专属词表。以自建航空英语语料库为例, 按照上述方法可以得到航空英语的专属词表, 因篇幅限制只列出词表中的前 10 位 (见表 2)。

表 2 自建航空英语语料库专属词表 (前 10 位)

Tab. 2 Aeronautics word list based on the self-built corpus (the most frequent 10 words)

排序	词汇	汉译	频数
1	flight	飞行, 航班	2 536
2	airplane	飞机	1 599
3	altitude	高度, 海拔	1 180
4	operation	操作, 运行, 经营	1 049
5	aircraft	航空器	1 039
6	speed	速度, 迅速	1 016
7	engine	发动机	952
8	pressure	压力, 气压, 紧急	821
9	system	系统, 体制, 秩序	820
10	display	显示, 展开	797
.....			
20	fail	发生故障, 失败	69

从制作的专属词表可以看出, 其中的高频词以名词居多, 这些高频词也同时体现了航空领域的焦点。该词表中出现的词汇, 应当作为教师课堂词汇教学的优先和重点。

以该词表中位于第 20 位的, 表示“故障”的 fail 为例。教师从语料库中选取合适的句子, 给学生布置任务, 以引起学生对该词译法的思考: 是否所有的“故障”都译为 fail?

例: “如果两个安定面配平/方向舵比率组件都发生故障, 就无法启用自动驾驶仪。”学生分小组讨论, 将讨论结果提交。对于“故障”一词, 学生普遍译为“fail”, “failure”, “break down”, 占全部译文的近 90%, 另外还有一组译为“malfunction”。这引起了学生们的兴趣, 究竟在真实语料中“故障”所对应的译法有哪些, 哪些译法出现的频率更高。通过对平行语料库的检索, 学生们进行了总结 (见表 3)。

表 3 平行语料库“故障”对应翻译

Tab. 3 Corresponding translation for “something going wrong” in the parallel corpus

故障	词性	频数	例句
fail/failed/fails	动词	217	One or more stages of rotor blades fail to pass air smoothly to succeeding stages.
failure/failures	名词	157	Deviations from the recommendations can result in structural damage or even complete failure of the airplane.
fault	名词	110	A system fault can result in the shutoff of all engine bleed air.
inoperative	形容词	101	A list developed for larger aircraft outlines equipment that can be inoperative for various types of flight.
malfunction	动词/名词	21	This can be caused by a malfunction in the pressurization system.
break down	动词词组	1	The low energy laminar flow tends to break down more suddenly than the turbulent layer.

通过以上实例,学生可以了解到语言使用的灵活性,两种语言并不存在“一对一”的绝对对等关系。在此过程中,学生丰富了词汇的积累,学习了高频词汇的不同译法,同时也提高了其自主学习的能力。

2. 利用语料库,探索语境共现

在 EST 教学中,词汇本身的积累十分重要。但同时要注意,词汇教学不可“断章取义”,词汇应当放在真实语料的交流情境中,结合上下文学习,这里就有必要引入词项的概念。词项的概念大于词汇,可以是最小的书写形式单词,也可以是连续的词语或词簇,还可以是非连续的结构或构式^[7]。因为语言“词以类聚”的特点,教师有必要引导学习者改变以往脱离语境阅读、翻译的习惯,更多地关注词语搭配。通过确定主题词制作词簇表,学习者可以观察、研究主题词的共现关系,进而尝试进行主题分析。这对于了解特定语域的话题分布、语言习惯、语言特征都大有裨益。

仍以上文中“故障”一词为例,提取所有含有“故障”的句子,构建一个子语料库。以“故障”对应的 fail, failure, malfunction, inoperative, break down, fault 为主题词,选择词簇长度为 3,对最低词频限制为 5,对词簇频率进行统计,结果见表 4。

表 4 “故障”对应的词簇

Tab. 4 Clusters parallel with “something going wrong”

故障	词簇	频数
failure	an engine failure	15
	a structural failure	11
fault	for datalink fault	6
fail	both FMCs fail	5
malfunction	a system malfunction	5

通过对共现频率的统计,学生发现与“故障”高度相关、强烈共现的词汇包括“发动机”“数据链”“系统”等词,用以描述具体发生故障的部件。此时,教师可进一步引导学生观察围绕“故障”主题词的高频词簇的特点,可以看到用名词作定语修饰名词的结构在以上所列高频词簇中所占比例达 62%。教师可就此引入语言学中名词定语的话题,这样的语言理论以语料库数据为依据,针对性强,有说服力,很容易引起学生的兴趣和重视,并将之

运用于日常的写作、翻译实践中。

(二) 语料库驱动下的语法教学

1. 利用统计学方法,探索 EST 语言结构特点

在科技发展日新月异、科技创新成为一个国家的核心竞争力的今天,担负着学术交流重任的科技英语,其重要性日益凸显。学生在学习 EST 的过程中常常感到高技术性词汇是学习的难点,但现实的情况是“技术性”越高的词汇,使用、翻译的难度越小。首先,高技术性词汇在科技英语中所占比例很小,根据 Inman 的统计,只占总词汇量的 9%^[8];另外高技术性词汇的显著特征就是确定性、唯一性,几乎不存在歧义。高技术词汇对学生理解、翻译造成的“难”题,很大程度上可以依靠专门词典的帮助来解决。相反,教师应当引导学生,更多关注 EST 的句法、文体、甚至语篇,因为这些层面也都体现了 EST 的专业性。

仍以上文提到的句子“如果两个安定面配平/方向舵比率组件都发生故障,就无法启用自动驾驶仪”为例。学生通过查阅专门词典解决了高技术性词汇(安定面配平/方向舵比率组件)的翻译问题,反而是在句子结构方面,学生的翻译产生了很大的分歧。比较典型的翻译为“If the two SRMs fail/malfunction, autopilot can't be used/applied/engaged.”“With the fault/failure of both SRMs, autopilot can't be used/applied/engaged.”将学生的翻译进行对比统计发现,近 70% 的学生选取了第一种翻译方法,也即从句中使用主谓结构,另外约 20% 的学生选择使用 fail 名词化结构 failure (另有 10% 的学生使用 fault) 来描述具体部件的故障。那么与通用英语相比,航空英语更倾向使用哪种结构呢?单纯比较 fail 与 failure 出现的频数似乎很难说明问题,这里有必要引入包含各语域的大型语料库进行对比,使用统计学方法来探索 EST 的语言结构特点。

统计 fail 和 failure 在自建航空英语语料库和 BNC 中的使用情况。为避免语域重复,BNC 语料库中选取了除科技英语外的语类进行对比研究。分别在自建航空英语语料库中和 BNC 中搜索 fail 及其屈折形式 fails, failed, 以及 failure 及其屈折形式 failures, 结果如表 5 所示。

表5 自建语料库与 BNC 中 fail 和 failure 频数

Tab.5 Frequency of "fail" and "failure" in the self-built corpus and BNC

频数	自建航空英语语料库	BNC
fail	69	3 250
fails	87	1 242
failed	61	7 613
fail 类总计	217	12 105
failure	140	5 405
failures	17	836
failure 类总计	157	6 241
总计	374	18 346

现将 fail 类和 failure 类在自建和 BNC 两个语料库中出现的总频数 374, 18 346 分别视为两个子语料库包含的词数, 利用卡方检验计算器计算这两个词及其屈折形式在这两个子语料库中的使用频率, 从而帮助分析航空英语中是否倾向使用以 failure 为代表的动词名词化结构。卡方检验计算结果如表 6 所示。

表6 fail 及 failure 卡方值

Tab.6 Chi-square for "fail" and "failure"

语料库形符数		自建语料库	BNC		
		374	18,346		
单词	自建语料库中词频	BNC 中词频	卡方值	P 值	
failure	157	6 241	10,3241	0.001 **	+
fail	217	12 105	10,3241	0.001 **	-

注:(该卡方计算软件由梁茂成制作)

由此可见, 在自建航空英语语料库中 failure 类使用频数更高, 其卡方值在 0.01 的显著性水平上存在统计学意义。因此得出结论: 与通用英语相比, 航空英语更倾向于使用动词名词化结构。

2. 教师引导对统计结果进行分析讨论

航空英语, 作为 EST 的特定语域, 其显示出的语言结构特征与 EST 的语言普遍规律一致。因此, 将以上的对比研究推而广之, 可以使学生意识到在 EST 的交流机制中, 主要的描述对象为现象、事物、状态等, 对个人情感和日常生活却鲜少涉及, 文本中会使用更多的名词和名词词组。与通用英语相比, 显示出较强的名词性特征。

教师可基于以上的统计学结果, 进一步引导学生分析动词名词化在 EST 中起到的表达效果。名

词化是指“从其他某个词类形成名词的过程或指从一个底层小句得出一个名词短语的派生过程”^[9]。这类现象属于 Halliday 提出的概念隐喻范畴。可鼓励学生对于相关语言学、翻译理论进行进一步了解, 但更为重要的是了解这样的结构在语言表达中产生的效果, 然后在撰写论文、报告或是进行翻译时尽量多使用类似的结构。

Halliday 认为名词化的使用与语体相关。书面语更多呈现名词化趋势是由于书面语在呈现现实世界的现象和经验时把它们作为物体来看待, 因而多使用名词。相反, 非正式语体特别是口语体, 是在用语言描述所发生的事情或所做的事情, 把现象和经验当作动作来描述, 因此多采用动词^[10]。在动词名词化的过程中, 用来修饰动词的语气词、情态动词往往被省略掉了, 因而模棱两可的情况大大减少, 这使得句子的意义更为客观、确定。这也正是动词名词化结构在科技英语语体中出现频率高于通用英语其他语体的原因之一^[11]。

名词短语的另一特点是可以包含一个甚至多个修饰成分, 也即一个名词短语可以用尽量少的词汇量传达尽量多的信息量, 这是动词短语所不能企及的。例如: Deviations from the recommendations can result in complete failure of the airplane. 如果用 fail 改写这样一个句子, 就不得不使用从句: If recommendations are not strictly followed, the airplane may fail completely. 相较之下, 使用了动词名词化的 failure 的句子简明扼要的特点就十分明显了。动词名词化结构可以高度浓缩信息, 使表达简洁、高效, 这正符合 EST 的交际需要。

此外, 动词或动词短语往往都需有施事出现, 而名词短语则可不用人称代词修饰, 对施事进行虚化, 从而使得文本更为客观、简洁。例如: Since data on at least one uncontained failure indicates the oil tank was overfilled at engine start, accomplish the following check until GE SB 72-648 is incorporated. 句中通过 failure 的使用省略了施事, 但同样客观、准确地表达了句意。

三、自建小型语料库应用于 EST 教学的远景与挑战

自建 EST 语料库的引入极大地缓解了 EST 教学素材缺乏的窘境。然而, 目前自建的航空英语语料库形符数刚刚超过 35 万, 只能算是小型语料库,

依然存在一定的局限性。特别是在运用统计学方法探索 EST 语言特点时,容量偏小往往会在一定程度上影响统计结果的准确性。但语料库的开放性决定了可以对其进行不断的扩容。教师可引导学生,使其加入语料库的构建过程,不但能使语料库的规模不断扩大,统计数据更为准确客观,还可激发学生的兴趣。

此外,可将学生的论文、报告、翻译习作,进行收集、整理,单独建立学习者语料库,以便教师发现学习者在语言使用过程中的闪光点和问题,同时也有利于学习者语料库的容量的积累和功能的动态增长。

基于语料库的 EST 翻译教学模式研究是语料库应用研究的新兴课题,具有广阔的研究前景。黎士旺指出语料库应用于翻译教学具有“数据直观化、分析多元化、结论科学化、成果有机化”的优势^[1]。相较于传统的翻译教学模式,教师介绍翻译理论、技巧→教师通过自省选取教材中例子进行演示→学生根据教师指定例题进行练习、语料库的引入可谓带来了一场变革。王克非教授在2007年语料库与译学研究国际学术研讨会上指出语料库可以“改变传统翻译教学中规定性方法”,“让学习者成为教学活动主体”^[12]。

此外,利用自建的语料库,可对 EST 特定语域(如航空英语)语料进行定量分析,总结其词汇分布、语法、甚至语篇的特点,帮助学生了解、掌握这些特点,以便今后在专业背景下交际时,遵循这些语言规律。但教师应注意到小型语料库因为规模的关系在代表性与说服力方面与大型语料库相比还存在一定的差距,此时应当注意多与大型语料库进行对比研究,以便得到更精确的数据和结论。

四、结 论

语料库的引入为 EST 传统教学模式带来了一场变革,使得以学生为中心的教学活动得以实现,EST 教学因缺乏素材遇到的困难也迎刃而解。本文通过搜集 EST 特定语域航空英语的语料,自建小型语料库,通过制作高频词表可有重点地进行词汇教

学,启发学生围绕高频词汇检索平行语料库,学习不同译法。亦可利用语料库探索语境共现,避免断章取义。引导学生使用语料库驱动的方法探索 EST 语言结构特点,达成语法教学的目的。教师应注意到小型语料库因为规模的关系在代表性与说服力方面尚存欠缺,但由于语料库具有开放性的特质,可不断进行扩容,亦可鼓励学生参加语料库的构建。

参考文献:

- [1] 黎士旺. 语料库与翻译教学[J]. 中国科技翻译, 2007, 20(3): 46-49.
- [2] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010.
- [3] Robinson P. EST Today: A Practitioner's Guide[M]. New York: Prentice Hall International(UK) Ltd, 1991.
- [4] Strevens P. EST after twenty years: a re-appraisedk [C] // Tickon M. EST: State of the Art. Singapore: SEAMEO Regional Center, 1988: 1-13.
- [5] 秦秀白. ESP 的性质、范畴和教学原则——兼谈在我国高校开展多种类型英语教学的可行性[J]. 华南理工大学学报(社会科学版), 2003, 5(4): 79-83.
- [6] Coxhead A. The academic word list: a corpus-based word list for academic purposes [C] // Kettemann B, Marko G. Teaching and Learning by Doing Corpus Analysis. 北京: 世界图书出版社, 2009: 76.
- [7] [英]戴维·克里斯特尔. 现代语言学词典[M]. 沈家煊, 译. 北京: 商务印书馆, 2000.
- [8] 于洋. 在搭配中学习科技英语次专业术语的实证研究[D]. 上海: 上海外国语大学, 2008.
- [9] Halliday M A K. Spoken and Written Language [M]. Victoria: Deakin University Press, 1985.
- [10] 杨信彰. 名词化在语体中的作用——基于小型语料库的一项分析[J]. 外语电化教学, 2006(2): 3-7.
- [11] 赵丽珠. 小型语料库在专门用途英语教学中的应用——以航空英语教学为例[J]. 北京航空航天大学学报(社会科学版), 2014, 27(2): 89-92.
- [12] 胡开宝, 吴勇, 陶庆. 语料库与译学研究: 趋势与问题——2007 语料库与译学研究国际学术研讨会综述[J]. 外国语, 2007(5): 64-69.

(编辑: 巩红晓)