



英语专业写作教学 语料库建设与研究

邹申 主编

上海外国语大学“211”三期建设重点学科资助项目“英语专业写作教学语料库建设与研究”

项目编号:SISU211-3-1-1-032

英语专业写作教学语料库 建设与研究

主 编 邹 申

副主编 潘鸣威

邓 杰

严 华

復旦大學 出版社

图书在版编目(CIP)数据

英语专业写作教学语料库建设与研究/邹申主编. —上海:
复旦大学出版社, 2011. 11
ISBN 978-7-309-08506-8

I. 英… II. 邹… III. 英语-写作-教学研究 IV. H315

中国版本图书馆 CIP 数据核字(2011)第 203722 号

英语专业写作教学语料库建设与研究

邹 申 主编

责任编辑/倪琴芬

复旦大学出版社有限公司出版发行

上海市国权路 579 号 邮编:200433

网址: fupnet@fudanpress.com <http://www.fudanpress.com>

门市零售:86-21-65642857 团体订购:86-21-65118853

外埠邮购:86-21-65109143

上海第二教育学院印刷厂

开本 850 × 1168 1/32 印张 9.75 字数 240 千

2011 年 11 月第 1 版第 1 次印刷

ISBN 978-7-309-08506-8/H · 1797

定价: 30.00 元

如有印装质量问题, 请向复旦大学出版社有限公司发行部调换。

版权所有 侵权必究

序

《英语专业写作教学语料库建设与研究》一书是在上海外国语大学“211”三期建设重点学科资助项目“英语专业写作教学语料库建设与研究”(项目编号:SISU211-3-1-1-032)成果基础上完成的。该项目起始于2008年,历时三年。项目的目标是以英语写作教学理论和语料库语言学为基础,系统地收集中国英语专业学生的写作语料,采用计算机辅助人工标注的方式,初步建立英语专业写作教学语料库,为深入探索英语学习者的写作能力发展规律、写作教学策略,完善外语教学理论与实践提供客观依据。此外,鉴于英语专业写作教学语料库为教学专门用途语料库,所采用的标注体系体现当今写作教材编写理论和原则,可为写作教师提供鲜活的学生习作范例,以丰富课堂教学实例。经过三年的辛勤努力,项目组已初步建成一个约45万词次,包括1500余篇不同年级学生英语作文的语料库。与此同时,项目组开发了一个赋码体系和与之配套的计算机标注软件,并在此基础上完成了一系列研究文章。

在此必须一提的是,项目的顺利完成离不开各位成员的智慧和努力,以及团队协作精神。以下各位成员在不同阶段参与了项目的研究工作:潘鸣威起草了写作教学语料库赋码方案,为写作教学语料库赋码体系的设计和完善起了关键性作用;邓杰专门设计了语料库标注软件和检索/分析软件,为提高标注工作的效率和精准率作了重要贡献;严华曾多次对话料进行全面、系统的整理,为后期的分析和研究打下良好基础;李清华参与了第一部分的撰写;张继东、王俭、陈建林、姚涓涓、泰中华、史光孝、郑佩芸、朱玉山等在不同阶段参与了项目讨论、语料标注、语料

分析和报告撰写等项工作。可以这样说,作为项目的成果,本书是全体人员集体智慧的结晶,是团队合作的产物。在此,我作为项目负责人向各位成员表示由衷的谢意。

最后,我们期望写作教学语料库的诞生能够对深入了解英语专业学生写作能力的发展特征,对完善教学大纲、充实写作教学理论、更新写作教学模式、提高写作教材编写的科学性、针对性和实用性起到积极推动作用。

邹申

2011年8月28日

目 录

序.....	1
第一部分 前言.....	1
第二部分 建库原则及方案	16
第三部分 写作教学语料库总体信息报告	46
第四部分 相关研究文章	68
附录.....	265

第一部分 前言

1.1 语料库的价值及作用

语料库的建立及研究是 20 世纪中后期语言学及应用语言学研究领域的一大发展特点,迄今为止已有近 50 年的历史。在近半个世纪的岁月里,语料库的功能及其对语言学及应用语言学的价值在不断地被人们挖掘和发现,并已逐步成为应用语言学领域的一个独立分支,即语料库语言学。纵观语料库的发展历史,我们可以看出以下五大特征:

第一,库容量大。语料库的建立与研究在不断完善、不断更新。新一代的语料库容量大,如 BNC 已达到 2 亿词次。21 世纪初已实现通过互联网检索的在线语料库 (WEBCORP) 更是达到了 10 亿—50 亿词次(何安平,2004, p. 3)。

第二,种类多样化。比如,原始语料库(raw corpus)按照一定规则(语域、语体、历时或共时)收集现实中使用的口、笔语资料。赋码语料库(annotated corpus)是对原始语料进行词性、语法、语音,语义或语篇进行标记的语料库。比如,已作词性赋码的 BROWN 和 LOB,已作语法赋码的 ICE-GB 语料库等。平行语料库(parallel corpus)则是在两种或多种语言之间在句子或短语层面上实现同步对译的语料库。学习者语料库(learner corpus)旨在收集非母语学习者的口、笔语料。比如,由比利时鲁汶大学 Granger 教授等建立的 ICLE(国际英语学习者书面语料库)和 LINDSEI(国际英语学习者口语语料库)。

第三,自动赋码系统研究的深化。自 20 世纪 90 年代以来,

语料库自动赋码系统从单纯的词性赋码向着句法、语义和语篇赋码的方向发展(何安平, 2004, p. 4)。

第四,语料库的建立为探究语言在使用过程中呈现出来的某种规律性提供了宝贵的数据。通过语料库研究人们可以发现:1)某些词汇在特定语域中的实际使用频数;2)反复出现的习语或搭配模式;3)某些词汇与特定语法结构同现倾向;4)与某种语义场相关联的词汇或语法结构;5)与某类语篇相关联的语法结构和语篇结构的使用特征等(何安平, 2004, p. 5)。

第五,学习者语料库为深化外语教学研究提供新的途径。学习者语料库的建立,为研究者提供了大量新鲜实例,为了解和总结学习规律和学习特征提供大量可靠性高的数据,从而使研究更具有说服力和概括性。这将对二语习得和外语教学理论和实践研究有着重大意义。从目前学习者语料库的使用来看,其功能主要体现在两个方面:1)辅助词典和教材的编写;2)利用现代化手段改进外语课堂教学的模式和手段。比如,在计算机辅助教学的平台上已出现了数据驱动式教学(data-driven learning)。此外,学习者语料库的数据资源可作为课程设置、教学大纲、教学目标制定的决策依据。另外,语料库资源也可为语言测试的设计和开发提供大量客观的语言应用依据。

1.2 国内学习者语料库概述

我国外语语料库研究起始于20世纪80年代。上海交通大学的JDEST语料库由杨惠中教授为首的项目组于1980年代中期建成。该语料库收录了理工科的英文学术论文等内容,是一个专用英语语料库,也是中国和亚洲地区的第一个英语语料库。该库语料涵盖30多个学科领域的多种文类,初建库容为100万词次,后扩容为400万词次。JDEST提供了学术英语领域通用词汇、技术词汇、次技术词汇的应用信息,为中国大学英语教学大纲制定提供了可靠的科学量化依据,为大学英语教学改革发

挥了重要的作用,曾获得国家教育委员会优秀教学成果奖。JDEST 不失为研究学术英语特征的重要资源。JDEST 项目的最重大意义在于,它开创了我国外语界进行语料库研究的先河,为世界范围内的第一代语料库建设,尤其是专门语料库建设提供了一套经典的原则、方法和技术范式。该语料库主要收集大学英语学习者的语料,现容量为 500 万词次。

之后,上海交通大学和广东外语外贸大学联合建设的中国学习者英语语料库(CLEC)是一个包括中学生、大学英语 4 级和 6 级、英语专业低年级和高年级学生在内的书面英语语料库,总词次为 100 万,以书面语为主(桂诗春、杨惠中,2003)。2005 年上海交通大学、河南师范大学和解放军外国语学院专家、学者共同建立了一个大学英语口语语料库,总词次为 75 万(杨惠中、卫乃兴,2005)。中国外语教育研究中心依据英语专业学生口试语料,已建成一个英语专业口语语料库(北京外语教学与研究出版社)。国内最新建立的语料库是《中国高校外语专业多语种语料库建设与研究——英语语料库子项目(CEM)》(上海外语教育出版社,2008)。该语料库预计收集约 400 万词次的英语专业学生语料,包括考试作文和翻译、平时作文和翻译。语料为标注语料,主要标注语料中出现的失误。

上述语料库的建设和研究对我国的外语教学都发挥了积极作用。例如,上海交通大学当时建立语料库的直接目的是为大学英语教学改革,尤其是为制定新的大学英语教学大纲提供决策依据。该语料库通过语料比较、统计和筛选等方法,为中国大学英语教学提供通用词汇和技术词汇的应用信息,为确定大学英语教学大纲的词表提供了可靠的量化依据。20 世纪初中国学习者英语语料库(CLEC)的建成,也大大推动了以中国英语学习者为对象的研究以及英语教学改革。

从建库目的上看,上述国内外语料库主要为语言/语言习得研究、外语教学研究、教学改革提供实证依据,而并非直接用于教学实践,故属于通用语料库。为专门用途(教学目的)建立的

语料库在国内也有例子,如华南师范大学何安平教授主持的中国英语教育语料库包含了多个子库,如英语教材、课堂教学会话和学生书面和口头英语(<http://sfs.scnu.edu.cn/corpus/>)。该语料库的建立为中学教师进行教学研究,为理论学习与教学实践相结合创造了良好条件。

语料库作为一种辅助性工具,其价值体现在研究者可以通过分析和归纳大量鲜活语料,提高研究结果的解释力度和概括性。同理,教师如果能在教学中利用语料库中的语言运用实例,对提高学生的语感和语言运用能力将会大有益处。鉴于此,研究和建立针对教学用途的语料库是语料库研究的一个发展方向。

1.3 英语作为第二语言(ESL) 写作文本特征研究概述

文本特征研究大都集中在文本的某些方面,对反映写作能力的文本特征的全面研究尚不多见。迄今为止,对 ESL 写作文本特征的研究分为两个层次:一是宏观层面(macro-level):语篇组织和思想表达;二是微观层面(micro-level):词汇,句法等(Cumming, *et al.*, 2000)。Cohen (1994)认为写作测试研究是一个巨大的挑战,因为写作测试应涵盖的内容太多,具体说来应包括下列方面:(1)内容:深度和广度(content: depth and breadth of coverage);(2)思想表达:观点明确(rhetorical structure: clarity and unity of the thesis);(3)结构:有条理,符合规范(organization: sense of pattern for the development of ideas);(4)语域:文体恰当(register: appropriateness of level of formality);(5)风格:优雅得体(style: sense of control and grace);(6)简洁性:语言使用的有效性(economy: efficiency of language use);(7)意义的准确性:选词确切(accuracy of meaning: selec-

tion and use of vocabulary); (8) 符合语言规范: 语法、拼写和标点符号正确 (appropriateness of language conventions: grammar, spelling, punctuation); (9) 易于读者理解: 表达清楚 (reader's understanding: inclusion of sufficient information to allow meaning to be conveyed); (10) 读者的认同 (acceptance): 努力提高文章的说服力 (reader's acceptance: efforts made in the text to solicit the reader's agreement, if so desired)。

研究文本特征的常用方法是语篇分析。尽管研究目的、过程和结果不尽相同, 我们还是能够从以下文本研究中观察到它们之间的共同关注点。

Sasaki 和 Hirose (1999) 的调查研究旨在了解日语教师评价日本大学生日语说明文的依据。研究结果分析得到 6 个次维度: 1) 主题的鲜明性 (clarity of the theme); 2) 对读者的吸引力 (appeal to the readers); 3) 思想表达 (expression); 4) 结构 (organization); 5) 语言知识 (knowledge of language forms); 6) 社会意识 (social awareness)。

Milanovic, *et al.* (1996) 的研究数据显示 11 种文本特征; 其中包括: 语法; 结构: 句子层面 (符合语法规则), 段落层面 (句子的排列和衔接), 论证结构 (思想表达、思路 [development of ideas]); 词汇: 用词准确、恰当, 词汇丰富; 拼写等项。

Sakyi (2000) 的研究发现, 评分员所关注的文本特征主要是: (1) 内容和组织; (2) 语法和规范类错误 (mechanical error); (3) 句子结构和词汇。研究者进一步把文本特征分为优秀作文的“积极特征” (positive features) 和较差作文的“消极特征” (negative features)。其中积极特征包括:

- (1) 主题鲜明 (good thesis statement);
- (2) 思路清晰 (good/clear thesis development);
- (3) 结构严谨 (过渡恰当、衔接连贯) (good organization/use of transitions or coherence);
- (4) 论证充分 (providing support/examples);

- (5) 逻辑性强 (good logic);
- (6) 语法错误少 (very few or no grammatical errors);
- (7) 词汇丰富 (sophisticated vocabulary);
- (8) 句子结构复杂、正确 (complex/correct sentence structure)。

而消极特征表现为:

- (1) 论点不明确;
- (2) 结构松散或混乱;
- (3) 论证不严密;
- (4) 不连贯;
- (5) 例证不充分或不相关;
- (6) 句子结构简单;
- (7) 语法错误多。

Cumming, Kantor 和 Powers (2001, 2002) 的研究发现评分员在评阅 TOEFL 作文时所关注的文本特征包括 11 种:

- (1) 卷面 (layout): 书写、可读性等;
- (2) 文本整体 (total written production): 篇幅长度、每段的长度等;
- (3) 词汇 (vocabulary): 选词的准确性、多样性等;
- (4) 组织结构 (rhetorical organization): 篇章结构、段落排列、衔接等;
- (5) 风格 (style): 清晰度、口语化程度、地道性等;
- (6) 句子 (sentences): 句子结构的完整性、多样性、复杂性等;
- (7) 语法 (grammar): 时态、语态、常用词类、主谓一致等;
- (8) 思想表达 (topic development): 论证的逻辑性、连贯性、说服力、任务的完成程度等;
- (9) 错误 (errors): 错误的多少、严重程度等;
- (10) 标点符号 (punctuation): 常用标点符号和大小写等;
- (11) 受试个人特征 (personal situation): 年龄、性别、种族、

母语等(详见表 1.1)。

表 1.1 TOEFL 作文的文本特征

卷面(Layout)	a. Handwriting b. Typing c. Spacing d. Lineation and paragraphing e. Legibility f. Editing or self-corrections g. Upper/lower case script
文本整体 (Total written production)	a. Length of essay b. Length of paragraph c. Length of sentence
词汇 (Vocabulary)	a. Choice b. Range c. Variety d. Precision e. Morphology f. Idioms g. Missing words
组织结构 (Rhetorical organization)	a. Essay structure • Introduction • Title • Topic orientation • Thesis statement • Sequencing • Body • Conclusion b. Paragraphing • Number of paragraphs • Development of paragraphs • Format of paragraphs • Topic sentence c. Cohesion d. Narration

续表

	e. Dialogue f. First person g. Second person h. Third person i. Reader involvement
风格 (Style)	a. Clarity b. Colloquialness c. Communicativeness d. Comprehensibility e. Formulaic f. Sophistication g. Idiomaticity h. Literalness i. Interestingness j. Conciseness k. Specificity l. Strangeness m. Convincing n. Effectiveness o. Exaggeration p. Clichés q. Repetition r. Register s. Fluency t. Literacy u. Creativity v. Ambiguity w. Awkwardness x. Presumptuousness y. Triviality z. Moralistic aa. Consistency bb. Avoidance strategies

续表

句子 (Sentences)	a. Sentence structure b. Sentence complexity c. Number of sentences d. Itemized lists e. Sentence variety f. Completeness g. Run on h. Questions
语法 (Grammar)	a. Preposition b. Pronoun c. Subject-verb agreement d. Tense e. Auxiliary f. Article g. Number (singular/plural) h. Adverbial i. Adjective j. Relative clause k. Adjective clause l. Gerund m. Participle n. Verb o. Subject p. Object q. Passive/active voice r. Noun phrase s. Conjunction t. Conditional u. Comparatives and superlatives

续表

思想表达 (Topic Development)	a. Argument • Logic • Reasoning • Development • Examples • Analogies or metaphors • Supporting evidence b. Task completion c. Topical focus d. Relevance e. Coherence f. Persuasiveness g. Misunderstanding
错误 (Errors)	a. Quantity b. Gravity c. Frequency d. Typographical errors
标点符号 (Punctuation)	a. Comma b. Period c. Capitalization d. Question mark e. Apostrophe f. Quotation marks g. Colon h. Semicolon i. Hyphen
受试个人特征 (Personal situation)	a. Age b. Gender c. Ethnicity d. Mother tongue e. Location f. Beliefs g. Motivation h. Prior education or writing experience i. Potential or readiness for learning or university studies

Cumming 等人(2001)把最受关注的文本特征归纳为以下3种:

(1) 组织结构(rhetorical organization): 引语(introductory statements); 衔接; 任务完成的程度(fulfillment of the writing task)等。

(2) 思想表达(expression of ideas): 论证的逻辑性; 清晰性(clarity)等。

(3) 语法和词汇的准确度与流利度。

根据所得出的文本特征, Cumming 等人(2001)提出 TOEFL 的写作评分标准应包括两大方面: 语篇质量和语言使用能力(详见表 1.2)。

表 1.2 TOEFL 写作评分标准的两大方面

1 Discourse
(a) Organization
(b) Coherence
(c) Progression
(d) Development, specificity, and quality of information and ideas
(e) For integrated tasks, ability to integrate/summarize/report
2 Language use
(a) Appropriacy and range of vocabulary and idiom used
(b) Appropriacy and effectiveness of language used to manage discourse connections
(i) Rhetorical functions, illocutionary markers
(ii) Paraphrasing, restating, citing, quoting source material
(c) Accuracy, appropriacy, and range of phrase-level syntax and morphological markers
(i) Subordinate clauses, participle phrases
(ii) Grammatical features such as articles, tense, adverbials
(d) Spelling, punctuation, other orthographic or typographic conventions

关于中国大学生英语作为外语(EFL)写作能力的文本表现, 现有文献仅见两项研究。

根据中国大学生限时议论文作文思维过程的基本环节(包